# Framework for Comparison of Multi-Fidelity Approaches for Military Vehicle Design

**Philip S. Beran, Dean E. Bryson**
Air Force Research Laboratory
Wright-Patterson Air Force Base
UNITED STATES

**Andrew S. Thelen**
University of Dayton Research Institute
Applied Mechanics Division
UNITED STATES

**Matteo Diez, Andrea Serani**
Consiglio Nazionale delle Ricerche
Istituto di Ingegneria del Mare
ITALY

**Laura Mainini**
Politecnico di Torino and
Massachusetts Institute of Technology
ITALY

**Keywords**: Multi-Fidelity Methods, Multidisciplinary Design Optimization, Surrogate Modeling, Design Space Exploration

***ABSTRACT***

This paper overviews the efforts of the AVT-331 technical team under the NATO Applied Vehicle Technology Panel, which is studying multi-fidelity methods through application to vehicle design. The objectives of the team are to understand the potential benefits of multi-fidelity methods in vehicle design and to document the relative strengths and weaknesses of different multi-fidelity methods using a common benchmark suite developed by the team. The benchmark suite has multiple levels of complexity, beginning with analytic functions with well-understood properties and culminating with representative air and sea vehicle benchmarks. The suite also includes intermediate complexity benchmarks of relevance to air, sea, and space vehicles. Benchmark features are catalogued herein, including descriptions of benchmark objectives and enabling software needed to reproduce vehicle-level results, as well as experimental setup for analytic benchmarks. The principal multi-fidelity methods utilized across the team are summarized and the extensive array of papers published by AVT-331 are tabulated. This tabulation documents the mapping between method employed and problem studied. Lastly, the process is described by which the relative strengths and weaknesses of different multi-fidelity methods is assessed over eight assessment categories. Multi-fidelity results are not included in this paper, but are separately detailed in four workshop papers. Team assessments will be documented in the AVT-331 final report.

## 1 BACKGROUND AND MOTIVATION

One of the long-term challenges of the multi-disciplinary design optimization of vehicles is the efficient increase of modeling fidelity, when it is needed, to capture the critical physics that constrain or enable particular vehicle concepts. Throughout much of a target design space, lower levels of modeling fidelity may be sufficient to

accurately guide the optimization process. However, relying on physics-lean models for analysis throughout the entire design space may lead to designs that are infeasible, or significantly sub-optimal, when the physics corresponding to the design point are incorrectly modeled. Simply replacing these models with higher fidelity models during optimization is not a practical strategy, because of the higher computational cost of these more informative techniques. Multi-fidelity (MF) methods offer the conceptual framework to efficiently optimize vehicles by judiciously using a limited number of high-fidelity (HF) analyses while leveraging the information provided by low-fidelity (LF) methods. Of interest to the community is whether these methods, which have found utility for problems of modest complexity [1], can scale up to vehicle-level problems.

To explore the challenge of "scaling up", the NATO Applied Vehicle Technology (AVT) Panel chartered a Research Task Group (RTG), AVT-331 (*Goal-Driven, Multi-Fidelity Approaches for Military Vehicle System-Level Design*), to explore the application of multi-fidelity approaches to the system-level design of military vehicles. The goals of this RTG have been to understand the potential benefits of MF methods in vehicle design and to assess the relative strengths and weaknesses of different MF methods applied to a common benchmark suite. The RTG has had two principal challenges: first, the construction of vehicle-level benchmark problems that are representative while being widely distributable, and second, structuring group activities in a manner conducive to method comparison over a relatively short time-span.

## 2 PROBLEM STATEMENT, OBJECTIVES, AND APPROACH OF AVT-331

### 2.1 Problem Statement

Explore the potential benefits and detriments of applying multi-fidelity approaches to the system-level design of military vehicles.

### 2.2 Objectives

1. Develop and distribute a suite of benchmark problems of varying complexity for method comparison (including vehicle level benchmarks).
2. Collect results of different MF methods as applied to the benchmark suite.
3. Assess the relative strengths and weaknesses of different MF methods and the potential for MF methods to accelerate vehicle design.
4. Document benchmarks and method comparisons in a manner that will be useful to the broader scientific community, potentially serving as the basis for future study.

### 2.3 Approach of AVT-331

AVT-331 is structured in a manner to promote cooperation, create standard benchmarks, and assess MF methods. The RTG is a ten-nation partnership with members drawn from government, industry, and academia. Products of

**Table 1: Benchmark complexity.**

| Complexity | Key Characteristics |
|---|---|
| L1 | QoIs are analytic functions<br>evaluations exact<br>negligible computational cost<br>little engineering utility |
| L2 | QoIs computed from ODE or PDE solutions<br>evaluations have small to moderate numerical errors<br>moderate computational cost<br>engineering idealizations |
| L3 | QoIs computed from PDE solutions<br>evaluations have moderate to large numerical errors<br>large computational cost<br>vehicle-level engineering models |

the team are intended to be shared with the broader community, including the final report, which will be available for distribution to the public and which may form the basis for future cooperation.

The AVT-331 team has defined and is investigating a number of different benchmark problems at varying levels of complexity. These levels are denoted by L1, L2, and L3, in order of increasing complexity, as described in Table 1. Qualitative measures of complexity include the mathematical nature of the problem quantities of interest (QoIs), the anticipated accuracy of numerical solutions, the cost of these calculations, and the degree to which the benchmark reflects a real-world problem. These benchmarks provide the AVT-331 team a common foundation by which the effectiveness of different MF methods can be compared. L1 benchmarks are analytic in nature and can be evaluated exactly at minimal cost, sometimes with the addition of prescribed noise. At the other end of the complexity spectrum are L3 benchmarks, whose general characteristics are of significant engineering relevance to the organizations participating in AVT-331. L2 benchmarks represent a compromise in complexity. They may exhibit dimensionality and numerical problems (e.g., convergence) to challenge various MF methods, but have a computational cost that is sufficiently low to study these challenges to a level of detail perhaps not possible for the L3 problems.

AVT-331 benchmark models are non-proprietary and leverage existing benchmarks widely distributed in the air, space, and sea communities, as well as analytical problems popular in the applied mathematics and scientific computing communities. The open approach maximizes sharing within the AVT-331 team, provides lasting benefit to the air, space and sea communities following project completion, and encourages comment form the members of these communities not directly participating in AVT-331.

L1 benchmarks are used to stress test and evaluate the applicability of MF methods for design-space exploration and design optimization. Owing to the relatively low cost of sampling analytic functions, L1 benchmarks facilitate thorough design-space exploration and highly accurate (or exact) prediction of the global optima used to validate methods. In contrast, the high cost of samples in L3 benchmarks prevents thorough design-space exploration, thus shifting focus to design optimization.

The AVT-331 team is decomposed into four sub-groups:

1. an air-group studying an airfoil in compressible flow (L2) and a high-aspect-ratio, air-transport configuration, optionally with structural coupling (L3);
2. a sea-group studying resistance of a destroyer configuration in inviscid (L2) and viscous flow (L3) in calm and regular seas with rigid-body coupling;
3. a space-group studying shape optimization of a re-entry body (L2) for maximum lift to drag ratio in inviscid flow, free molecular flow, and transition regimes, and
4. an analytic-benchmarks group defining and studying a set of L1 problems to learn how different MF methods behave and perform in the face of specific mathematical properties of models and objective functions.

Decomposition of the team in this manner enables each sub-group to cooperatively develop and analyze models requiring domain expertise as well as to prepare needed reports and domain-focused documentation.

In addition to being non-proprietary, further requirements are set for team benchmarks. First, L3 benchmarks are formulated in a multidisciplinary manner: the transport configuration exhibits aeroelastic coupling and the destroyer configuration includes rigid-body coupling to model sea handling. Second, more than one organization is expected to study each benchmark to enhance cooperation, stimulate comparison of MF methods, and to increase model reliability and maturity. Third, L1 benchmarks are formulated to stress test MF methods over a variety of mathematical properties that are individually or jointly exhibited by more complex physics-based applications whose analytical solutions are rarely available.

The activities of AVT-331 are described in the following sections. First, the AVT-331 concept of MF methods is defined in Section 3 to highlight the technical scope within which this team operates. The L1, L2, and L3 benchmarks are summarized in Section 4, including information about levels of fidelity, types of physics, and contributing organizations. Methods are outlined and categorized in Section 5, with particular attention given to the identification of benchmarks to which various methods are applied. The process by which the strengths and weaknesses of contributed methods is assessed is described in Section 6. Finally, the products of AVT-331-related activities are summarized in Section 7 and concluding remarks provided in Section 8.

## 3 DEFINITION OF MULTI-FIDELITY METHODS

AVT-331 takes a functional approach to defining design-oriented MF methods, considering them to algorithmically accomplish the following:

1. synergistic use of multiple information sources; i.e., different physical models and/or numerical approximations,
2. acceleration of the design process while achieving reliable design solutions.

A good example of a design-oriented MF method is the exploration of a design space using a surrogate model that is constructed with data from multiple information sources and which adapts to the data acquired during the exploration of the space. This kind of method mixes data obtained from different information sources, and combines this data in a synergistic fashion to accelerate the design process, generally by leveraging useful information that can be gleaned from information sources that are cheaper to query.

Examples of approaches and capabilities considered outside of the AVT-331 scope include the following:

1. approaches that truncate design spaces based on data from one information source (dimensionality reduction) while exploring the space using other information sources (or more broadly, data from different information sources, drawn from a common discipline, used in a decoupled fashion);
2. techniques limited to one information source (e.g., surrogates of an information source are used as a low-fidelity model of that same information source), and
3. multi-objective approaches that evaluate objectives with different information sources (another example of data produced by different information sources used together but in a decoupled fashion).

The term "information source" is broadly interpreted here as an instance of a model with different possible approximations, either mathematical (e.g., different equation set), computational (e.g., different mesh), or some combination thereof. Experimental data sources are of high value in representing "truth", but since an experimental dimension is not present in AVT-331 and since we are considering design space exploration prior to the introduction of hardware, this source of approximation is not explicitly addressed (although particular MF formulations may be readily amenable to incorporation of this kind of data). While AVT-331 is focused on the algorithms that promote synergistic use of different information sources, it recognizes the critical importance of computational frameworks that enable different information sources to be queried in a single, flexible process.

AVT-331 uses the term fidelity in a manner consistent with that of Fernandez-Godino et al. [2], who treat fidelity of an information source as a "level of accuracy or complexity." In this sense, fidelity is a property of the information source, which may be known qualitatively but not quantitatively (unless a truth source can be queried) and which is generally dependent on mathematical and numerical approximations employed by the information source as well as design parameters. Table 2 summarizes a variety of ways in which the accuracy and complexity of an information source can be varied. Three of the categories correspond to adaptations to elements of physical fidelity: governing equations (changing assumed physics), model geometry (e.g., topological complexity), and coupling (e.g., decoupled versus fully coupled). Two of the categories correspond to adaptations to elements of numerical fidelity: model discretization (e.g., mesh refinement) and convergence (e.g., incomplete convergence). Lastly, the variation of data also represents a level of accuracy or complexity. This category can be exemplified by the retention of different numbers of vibratory modes extracted from a structural model.

In the AVT-331 vehicle benchmarks to be described, there are several MF studies involving variation of physical fidelity (primarily changes to governing equations from low to high fidelity, with one study adding or removing a trim constraint) and some MF studies involving variation of numerical quantities (primarily numerical discretization). Another set of benchmarks explore geometric refinements (scaling) through reduced truncation of a statistical representation of geometric quantities.

Other surveys of MF methods include those of Peherstorfer et al. [3] and Clark et al. [4]. Perspective derived from early AVT-331 activities is reported in [5]. Peherstorfer et al. [3] reviewed MF approaches within the context of uncertainty quantification (UQ) and defined three categories of methods: adaptation, fusion, and filtering. These categories broadly aggregate different strategies for leveraging and/or mixing predictions of quantities of interest from models of different fidelity, e.g., HF and LF models. The categories share one common feature: black-box evaluation of QoIs. While the team is interested in and currently exploring non-black-box approaches (beyond extraction of gradient information; e.g., [6]–[9]) the benchmarks employed in AVT-331

**Table 2: Key types of information-source adaptation.**

| Adaptation To | Type |
|---|---|
| Governing Equations | Physical fidelity |
| Coupling | Physical fidelity |
| Geometry | Physical fidelity |
| Discretization | Numerical fidelity |
| Convergence | Numerical fidelity |
| Data | Data coverage |

promote a data-centric philosophy of black-box sampling readily generalizable to different problems and well suited to coordinated analysis for method comparison.

## 4 SUMMARY OF DESIGN OPTIMIZATION BENCHMARKS

Several criteria are considered by AVT-331 in the selection of benchmarks. L3 benchmarks capture certain features (e.g., geometric) essential to a military vehicle, have a multidisciplinary component reflective of a system-level attribute, and address certain methodological challenges relevant to multi-fidelity methods. The lighter weight L2 benchmarks address certain physics and numerical challenges of relevance to an L3 study. And an L1 benchmark addresses certain methodological challenges of broad interest, even if the equation is not representative of behaviors of selected L2 or L3 benchmarks. To be useful as a basis of comparison, candidate problems are considered benchmarks when adopted by at least two groups, different levels of fidelity are identified, and preferably when the benchmark developer provides documented analysis tools with the benchmark, which are considered to be part of the benchmark definition.

L1 benchmarks are generally analytic in nature and can be evaluated exactly at minimal cost, potentially with prescribed noise. At the other end of the complexity spectrum are L3 benchmarks, whose general characteristics are of significant engineering relevance to the organizations participating in AVT-331. L2 benchmarks represent a compromise in complexity. They may exhibit dimensionality and numerical problems (e.g., convergence) to challenge various MF methods, but have a computational cost that is sufficiently low to study these challenges to a level of detail perhaps not possible for the L3 problems.

Most benchmarks have different configurations owing to variability of benchmark goals (optimization or global accuracy), the cost and complexity of the L2 and L3 benchmarks (e.g., different fidelity selections), or other simple benchmarks options (e.g., a different resource limitation). For quick reference, the different benchmark configurations are designated by the code $\mathrm{L_{ID}}$-$\mathrm{d_{ID}}$-$n$, where $\mathrm{L_{ID}} \in [\mathrm{L1}, \mathrm{L2}, \mathrm{L3}]$, $\mathrm{d_{ID}} \in [\mathrm{air}, \mathrm{sea}, \mathrm{sp}, \mathrm{eqn\ code}]$, and $n$ is a non-negative integer case number for each vehicle domain. For example, the fifth air benchmark of L3 complexity is designated L3-Air-5. The designation of L1 benchmark configurations replaces the domain code with an appropriate equation code; for example, L1-Rosen-1 for the Rosenbrock equation. The set of benchmarks selected, defined, and studied by AVT-331 is summarized in two places: Table 3 for L2 and L3 problems and Table 4 for L1 problems. Included in Table 3 are definitions of different physical and numerical fidelity levels used in the vehicle benchmark, goals of the benchmark study, lists of software needed to execute the benchmark, and information about AVT-331 groups supplying and using the benchmark.

**Table 3: AVT-331 L2 and L3 benchmarks.**

| D | C | FLs (physics) | FLs (numerics) | Essential Feature(s) | MD Attributes | Goal | Parameters | Major Dependencies | Challenges | Groups Using | Distr. Limitations | Comments |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Air | L3 | • Vortex Lattice<br>• Euler | Baseline | • Parametric wing-body-tail<br>• Rib-spare structure (option) | Aeroelasticy, steady (option) | • Min fuel burn, drag (option)<br>• Constraints | • OML shape<br>• Structural size (option) | AVL, SU2, NASTRAN, CAPS, ESP | High dimension | AFRL*, ITU, UDRI | None | • uCRM-13.5 derivative<br>• Ref. [10] |
| Air | L2 | • Potential + boundary layer<br>• Euler<br>• RANS | Grid refinement | Parametric wing sectional | None | • Min drag<br>• Constraints<br>• Global accuracy | Sectional shape | XFOIL, SU2 | • Noise from adaptive meshing<br>• Moderate dimension | AFRL, CIRA*, PoliTO | None | • Baseline 2822 airfoil<br>• Ref. [11] |
| Space | L2 | • LPI | Panel detection | Surface shape morphing | • Aerodynamics<br>• Aerothermo-dynamics (constraint) | • Min efficiency<br>• Constraints | Morpher control points | Morpher MIMMO | • Moderate dimension<br>• Noise for lower FLs | ITU, UoS* | Morpher not supplied (but it is free and open source) | • Baseline<br>• Ref. [12] |
| Sea | L3 | • Potential flow<br>• Strip theory<br>• RANS | Baseline | Parametric hull form | Rigid-body coupling | • Min resistance in calm water<br>• Min resistance in regular waves<br>• Max seakeeping in regular waves<br>• Constraints | Hull form | None | • Multi-obj<br>• Moderate dimension | CNR-INM, ITU, MARIN, NTUA | Solvers not supplied | • Baseline DTMB 5415<br>• Ref. [13] |
| Sea | L2 | • Potential flow | Grid refinement | Parametric hull form | None | • Min resistance in calm water<br>• Constraints | Hull form | None | Moderate dimension | CNR-INM*, MARIN, NTUA, PoliTo | None | • Baseline DTMB 5415<br>• Ref. [14] |

D = Domain; C = Complexity; FLs = Fidelity Levels; MD = Multi-disciplinary
AVL (vortex lattice); LPI (local panel inclination); XFOIL (potential aerodynamics); SU2 (Euler and RANS aerodynamics); NASTRAN (structure if aeroelastic); CAPS, EPS (geometry and mesh)
*Group Supplying

Two L3 vehicle-level problems are studied by groups within the AVT-331 team: the air group studies a high-aspect-ratio, air-transport configuration and the sea group studies a destroyer-type configuration. These non-proprietary models leverage existing benchmarks widely distributed in the air and sea communities to maximize sharing within the AVT-331 team, to provide lasting benefit to the air and sea communities following project completion, and to encourage comment from members of these communities not directly participating in AVT-331. As one goal of AVT-331 is to evaluate MF methods in the context of system-level optimization, each problem is formulated in a multidisciplinary manner: the transport configuration exhibits aeroelastic coupling and the destroyer configuration includes rigid-body coupling to model sea handling.

### 4.1 Analytic Benchmarks

The L1 benchmarks are analytic problems involving the evaluation of explicitly known and simple-to-evaluate functions with the optional addition of noise. These benchmarks are essentially verification tests for MF methods, allowing for (a) widespread use within the AVT-331 team, (b) the elimination of numerical errors, and (c) a high degree of convergence via the low computational cost of function evaluation. Fidelity discrepancies are manufactured by imposing defects onto an otherwise 'perfect' truth model. The availability of perfect reference models distinguishes these benchmarks from real-world applications, but allows MF methods to be validated over the specific mathematical features exemplified by each L1 problem. L1 benchmarks include suitably formulated problems based on widely known optimization tests, such as the Forrester, Rosenbrock, Rastrigin, and Paciorek functions, as well as *ad hoc* analytic functions formulated for testing MF modeling methods. An example is shown in Figure 1 for the Rastrigin-based benchmark which has been shifted and rotated to obtain a more challenging problem.

The use of such analytic functions has certain downsides. First, the cost of function evaluations across the fidelity spectrum may not vary nearly as much as in L2 and L3 problems due to the simplicity of the models. To address this problem, cost functions are synthesized in an attempt to model the naturally higher cost of higher fidelity models. Also, the use of conventional analytic functions naturally does not capture the system-level attributes of multidisciplinary problems. In AVT-331, the study of L2 benchmarks attempts to address issues of problem complexity, but with model evaluations of lower cost. A summary of L1 tests and their main features is provided in Table 4. A more detailed summary with problem dimensionality, computational budget, and artificial computational cost associated to each fidelity is reported in Table 5. Further details on the L1 benchmarks definition are presented in [15], [16].

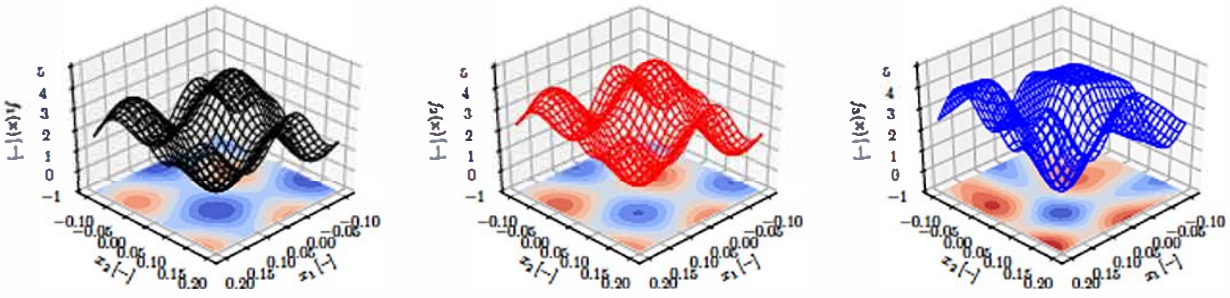


**Figure 1: Shifted-rotated Rastrigin Function: from left to right, $f_1$ (highest-fidelity), $f_2$, and $f_3$ (lowest-fidelity) [15].**

**Table 4: Analytical benchmarks main features [15].**

| ID | Name | Behaviors | Scalability | Discrepancy | Noise |
|---|---|---|---|---|---|
| L1-For | Forrester | Local / (Dis)continous | - | (non)linear | no |
| L1-Ros | Rosenbrock | Local | Parametric | nonlinear | no |
| L1-Ras | Shifted-rotated Rastrigin | Multi-modal | Parametric / Fidelity | nonlinear | no |
| L1-Het | Heterogeneous | Local / Multi-modal | Parametric | nonlinear | no |
| L1-Spr | Spring-Mass system | Multi-modal | Parametric / Fidelity | nonlinear | no |
| L1-Pac | Paciorek | Multi-modal | Fidelity | nonlinear | yes |

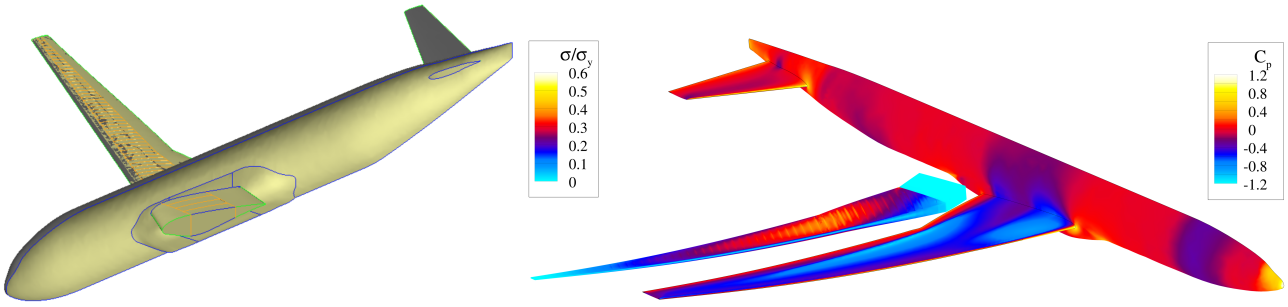
**Table 5: Experiments setup summary [15].**

| Function/Problem | Benchmark ID | $D$ | Budget | Fidelity cost | | | |
|---|---|---|---|---|---|---|---|
| | | | | $f_1$ | $f_2$ | $f_3$ | $f_4$ |
| Forrester | L1-For-1 | 1 | 100 | 1.00000E-0 | 5.00000E-1 | 1.00000E-1 | 5.00000E-2 |
| Jump Forrester | L1-For-2 | 1 | 100 | 1.00000E-0 | 2.00000E-1 | - | - |
| Rosenbrock | L1-Ros-1 | 2 | 200 | 1.00000E-0 | 5.00000E-1 | 1.00000E-1 | - |
| | L1-Ros-2 | 5 | 500 | 1.00000E-0 | 5.00000E-1 | 1.00000E-1 | - |
| | L1-Ros-3 | 10 | 1000 | 1.00000E-0 | 5.00000E-1 | 1.00000E-1 | - |
| Shifted-rotated Rastrigin | L1-Ras-1 | 2 | 200 | 1.00000E-0 | 6.25000E-2 | 3.90625E-3 | - |
| | L1-Ras-2 | 5 | 500 | 1.00000E-0 | 6.25000E-2 | 3.90625E-3 | - |
| | L1-Ras-3 | 10 | 1000 | 1.00000E-0 | 6.25000E-2 | 3.90625E-3 | - |
| Heterogeneous | L1-Het-1 | 1 | 100 | 1.00000E-0 | 2.00000E-1 | - | - |
| | L1-Het-2 | 2 | 200 | 1.00000E-0 | 2.00000E-1 | - | - |
| | L1-Het-3 | 3 | 300 | 1.00000E-0 | 2.00000E-1 | - | - |
| Springs | L1-Spr-1 | 2 | 200 | 1.00000E-0 | 1.66667E-2 | - | - |
| Springs-masses | L1-Spr-2 | 4 | 400 | 1.00000E-0 | 1.66667E-2 | - | - |
| Pacioreck | L1-Pac-1 | 2 | 200 | 1.00000E-0 | 2.00000E-1 | - | - |

### 4.2 Air Vehicle Bechmarks

The L3 air class of benchmark problems is based on an undeflected high aspect ratio variant of the NASA Common Research Model (CRM), coined the uCRM-13.5 by Brooks et al. [17] in the University of Michigan (UM) MDO group. The geometry definition was scripted in Engineering Sketch Pad (ESP) using UM's outer moldline and inner moldline CAD models as references.

High- and low-fidelity analyses, which may be rigid or aeroelastic, are prepared by a python interface to Computational Aircraft Prototype Syntheses (CAPS), which drives aerodynamic and structural meshing (via AFLR and EGADS, respectively) and input file generation, among other processes. The resulting physics analyses are based on linear panel aerodynamics (usually augmented with empirical wave drag estimates) or nonlinear CFD (primarily inviscid Euler), both of which are coupled to linear shell structures of the wingbox for aeroelastic cases. Figure 2(a) shows a sample geometry taken from the ESP graphical user interface.




*(a)* *(b)*

**Figure 2: Sample L3 air design: (a) IML and OML geometries, (b) sample static aeroelastic solution showing stress and pressure contours.**

**Table 6: Solvers for the L3 air vehicle problems [10]. Atmospheric conditions and other details are given later in Table 7.**

| Response | Software Set 1 | | Software Set 2 | |
|---|---|---|---|---|
| | Low-Fidelity | High-Fidelity | Low-Fidelity | High-Fidelity |
| Rigid | AVL [18] + Korn wave drag | SU2 [19] Euler | AVL [18] + Korn wave drag | FUN3D [20] Euler |
| Aeroelastic | Nastran [21], AVL [18] + Korn wave drag | Nastran [21] + SU2 [19] via pyCAPS | VLM + TACS [22] via Mphys [23] | FUN3D [20] Euler + TACS [22] via FUNtoFEM [24], [25] |

Open-source (or otherwise publicly available) software options for these physics analyses include SU2 and AVL for rigid high- and low-fidelity analyses. For aeroelastic analyses, SU2 and Nastran are loosely coupled using pyCAPS FSI coupling, while a combination of Nastran and AVL are used for low-fidelity. This set of widely available tools is described by Software Set 1 in Table 6. Other analysis tools, described by Software Set 2 in Table 6, include FUN3D for rigid high-fidelity analyses (which generally provides comparable information as SU2), FUN3D coupled to TACS via FUNtoFEM for aeroelastic high-fidelity analyses, and a NASA-developed VLM coupled to TACS via Mphys for aeroelastic low-fidelity analyses. While only available to a subset of the task group, Software Set 2 has the benefit of providing analytical design derivatives for high- and low-fidelity aeroelastic analyses [10]. Figure 2(b) shows a sample static aeroelastic solution using the FUN3D/TACS/FUNtoFEM analysis.

Using these geometry and analysis tools, two sets of objectives and constraints have been developed. Firstly, a simple, low-dimensional case considers lift-constrained drag minimization (L3-Air-1). As a more representative air vehicle design case, the second design problem considers fuel burn minimization subject to load factor and moment constraints during cruise (L3-Air-2, L3-Air-3); load factor and stress constraints are also enforced for a pull-up maneuver (L3-Air-4, L3-Air-5). Presumed parameters for these more complex problems are given in Table 7. Design variables may include planform, airfoil shape, structural thicknesses, and flow angles of attack, with specific cases described in Tables 8 and 9. Details of the L3 air benchmark are presented in [26].

Finally, the L2 air vehicle problem is the shape optimization of an RAE-2822 airfoil under transonic conditions, where the baseline configuration is the classical airfoil working in fully turbulent flow conditions modeled using

**Table 7: Design specifications for the L3 air fuel burn minimization problems. [10].**

| Description | Value | Units |
|---|---|---|
| Design range | 7725 | nmi |
| Specific fuel consumption | 0.53 | $lb_m/(lb_f*h)$ |
| Engine diameter | 2.85 | m |
| Takeoff fuel burn | 2.5 | % Maximum takeoff weight |
| Landing fuel burn | 1.0 | % Mass at end of cruise |
| Fixed mass | 91,250 | kg |
| Payload mass | 34,000 | kg |
| Reserve fuel mass | 15,000 | kg |
| Engine mass | 16,564 | kg |
| Structural mass for rigid cases | 27,250 | kg |
| Fixed mass center of gravity x-coordinate | 0.569 | Fraction of fuselage length |
| Payload mass center of gravity x-coordinate | 0.569 | Fraction of fuselage length |
| Maneuver load factor | 2.5 | g |
| KS stress factor of safety | 1.5 | - |
| Cruise Mach number | 0.85 | - |
| Cruise altitude | 37,000 | ft |
| Maneuver Mach number | 0.64 | - |
| Maneuver altitude | Sea level | - |

**Table 8: Design variable designations for the different L3 air parameterizations.**

| | Shape | | | | |
|---|---|---|---|---|---|
| | Planform | Airfoil | Twist | Structural | Aerodynamic |
| Vector name | $\mathbf{x_{s_p}}$ | $\mathbf{x_{s_a}}$ | $\mathbf{x_{s_t}}$ | $\mathbf{x_t}$ | $\mathbf{x_a}$ |
| Dimensionality | 2 or 5 | 14, 34, or 85 | 2 or 5 | 283 | 1 or 2 |

the Spalart-Allmaras turbulence model [27]. The open-source software modules for this benchmark include (a) airfoil shape parameterization, (b) automatic meshing, and (c) multi-fidelity aerodynamic solvers. These building blocks allow the resolution of an aerodynamic shape design optimization problem allowing a standardized and common set of objective function evaluators. Details on the problem definition are reported in [11], [28] including airfoil parameterization, mesh generation and flow field computation.

### 4.3 Space Vehicle Benchmarks

The design optimization of re-entering space vehicles is addressed by an L2 problem (L2-Space-1), regarding the shape optimization of the IXV technology demonstrator vehicle [29], aiming at maximizing aerodynamic efficiency at altitude $H = 70$ km, velocity $V = 6500$ m/s, angle of attach $\alpha = 10$ deg. The optimization is constrained by minimum lift requirement, maximum heat flux, and geometrical requirements to accommodate a cylindrical bay; see Figure 3(a). The vehicle shape is parameterized via radial basis functions, whose control

**Table 9: Optimization problem descriptions for the L3 air cases. Note that problems L3-Air-1 and L3-Air-2 use reduced and full sets of planform variables.**

| Case | DVs | Airfoil DoF | Designable Airfoils | # DVs | Objective | Constraints |
|---|---|---|---|---|---|---|
| L3-Air-1 | $\{\mathbf{x_{s_p}}, \mathbf{x_a}\}$ | 7 | 0 | 3 | Drag coefficient | Lift coefficient |
| L3-Air-2 | $\{\mathbf{x_{s_p}}, \mathbf{x_a}\}$ | 17 | 0 | 6 | Fuel burn | Cruise load factor & moment |
| L3-Air-3a | $\{\mathbf{x_{s_a}}, \mathbf{x_{s_t}}, \mathbf{x_a}\}$ | 7 | 2 | 17 | Fuel burn | Cruise load factor & moment |
| L3-Air-3b | $\{\mathbf{x_{s_a}}, \mathbf{x_{s_t}}, \mathbf{x_a}\}$ | 17 | 2 | 37 | Fuel burn | Cruise load factor & moment |
| L3-Air-3c | $\{\mathbf{x_{s_a}}, \mathbf{x_{s_t}}, \mathbf{x_a}\}$ | 17 | 5 | 91 | Fuel burn | Cruise load factor & moment |
| L3-Air-4a | $\{\mathbf{x_t}, \mathbf{x_a}\}$ | 7 | 0 | 285 | Fuel burn | Cruise load factor & moment<br>Maneuver load factor & stress |
| L3-Air-4b | $\{\mathbf{x_t}, \mathbf{x_a}\}$ | 17 | 0 | 285 | Fuel burn | Cruise load factor & moment<br>Maneuver load factor & stress |
| L3-Air-5a | $\{\mathbf{x_{s_a}}, \mathbf{x_{s_t}}, \mathbf{x_t}, \mathbf{x_a}\}$ | 7 | 2 | 301 | Fuel burn | Cruise load factor & moment<br>Maneuver load factor & stress |
| L3-Air-5b | $\{\mathbf{x_{s_a}}, \mathbf{x_{s_t}}, \mathbf{x_t}, \mathbf{x_a}\}$ | 17 | 2 | 321 | Fuel burn | Cruise load factor & moment<br>Maneuver load factor & stress |

points are also shown in Figure 3(a).

The L2 space benchmark is built up by combining (a) an open source mesh morpher [30], and (b) a MATLAB module that implements a multifidelity, panel based, aerothermodynamic solver for re-entry analyses [31], [32]. The aerodynamic module implements the Modified Newton Theory for the continuum regime and the the Schaaf and Chambre inclined flat-plate model in the free molecular flow regime. A bridging function approximates the aerothermodynamic characteristics in the transition regime. The aerothermal module implements several semi-empirical models, and for this test case the Kemp-Rose-Detra method has been used for the continuum regime, while the aerothermal Schaaf and Chambre has been used for the rarefied regime. Again a bridging function allows the generalization within the transition regime. Details for the L2 space problem are given in [28].

### 4.4 Sea Vehicle Benchmarks

The L3 sea benchmark is the hull-form optimization of the DTMB 5415 model (Figure 5). This is an open-to-public concept model used during the early development of the DDG-51, a USS Arleigh Burke-class destroyer. The model is widely used for towing tank experiments [33], CFD studies [34], and hull-form optimization [35]. The DTMB 5415 model (both bare hull and its appended variants) have been used also as a test case for AVT-204 [36] and AVT-252 [13] respectively on deterministic and stochastic design optimization methods for military vehicles and AVT-280 [37] on prediction capabilities of ships' large amplitude motions in heavy weather. The model main particulars (bare hull) are summarized in Table 10, where the length parameter $L_{\mathrm{pp}}$ is calculated from the fore perpendicular to the transom bottom edge.

Two L3 benchmark problems are solved within AVT-331. These are a modified version of the problem solved

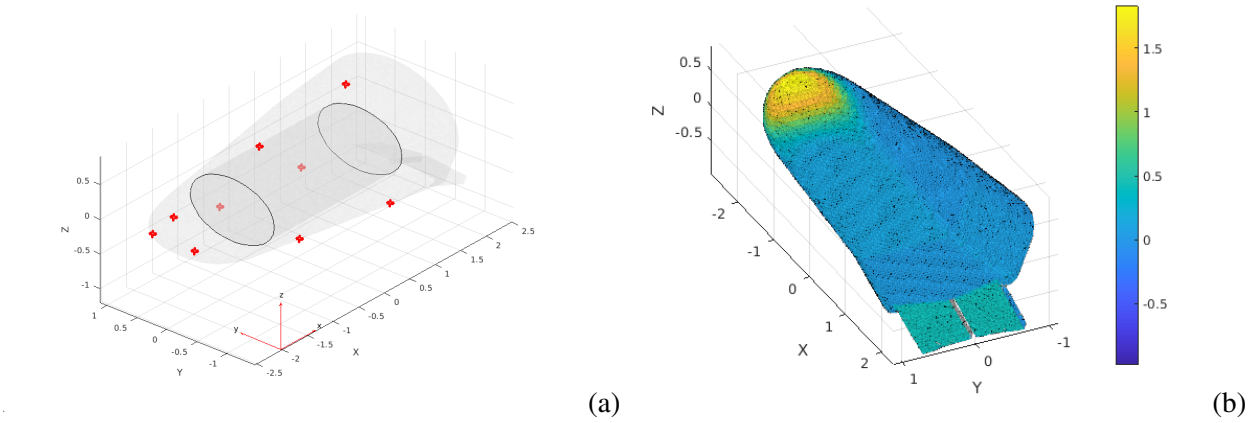



**Figure 3: Space benchmark: (a) baseline shape with control points and internal cylindrical bay, and (b) aerodynamic computation results for pressure coefficient.**

in AVT-204 and to some extent a simplified version of the problem tackled in AVT-252. Two sets of conditions are considered, namely calm water and regular waves. Design objectives include merit factors associated to both resistance (in calm water and waves) and seakeeping (motions in waves). Physical coupling is considered of hydrodynamics with rigid body equations of motion. All computations are assumed captive (self-propulsion not considered) with two degrees of freedom (heave and pitch motions); see Figure 4. Geometrical equality constraints include fixed length between perpendiculars and displacement, whereas geometrical inequality constraints include limited variation of beam and draught and reserved volume for the sonar in the dome; see e.g. [36]. No functional constraints are used in current test cases. The L3 benchmark problems are summarized in Table 3.

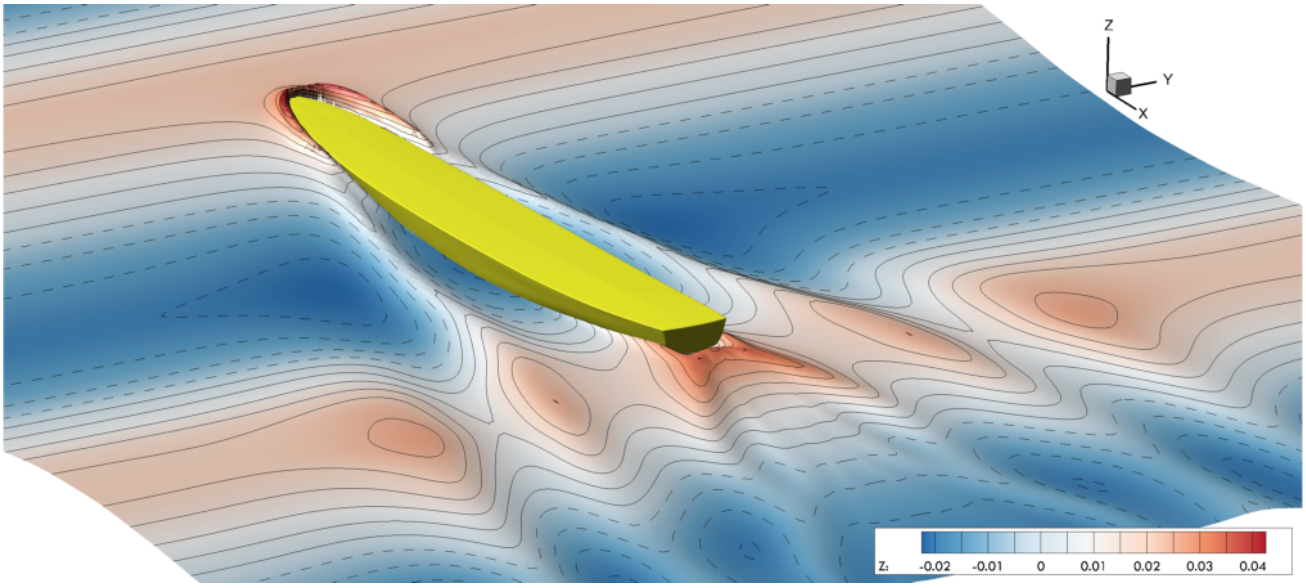



**Figure 4: DTMB 5415 in regular waves at Fr=0.28 by CFDShip-Iowa (RANS).**

**Table 10: DTMB 5415 main particulars.**

| Description | Symbol | Unit | Full scale | Model scale |
|---|---|---|---|---|
| Displacement | $\nabla$ | tonnes | 8,437 | 0.549 |
| Length between perpendiculars | $L_{\text{pp}}$ | m | 142.0 | 5.720 |
| Beam | $B$ | m | 18.90 | 0.760 |
| Draft | $T$ | m | 6.160 | 0.248 |
| Longitudinal center of gravity | LCG | m | 71.60 | 2.884 |
| Vertical center of gravity | VCG | m | 1.390$^\star$ | 0.056$^\star$ |
| Bridge longitudinal location | $\text{B}_x$ | m | 44.00$^\dagger$ | 1.772$^\dagger$ |
| Bridge vertical location | $\text{B}_z$ | m | 24.75$^\ddagger$ | 0.997$^\ddagger$ |
| Flight deck longitudinal location | $\text{D}_x$ | m | 132.0$^\dagger$ | 5.317$^\dagger$ |
| Flight deck vertical location | $\text{D}_z$ | m | 13.00$^\ddagger$ | 0.524$^\ddagger$ |
| Roll radius of gyration | $K_{xx}$ | – | $0.40B$ | |
| Pitch radius of gyration | $K_{yy}$ | – | $0.25L_{\text{pp}}$ | |
| Yaw radius of gyration | $K_{zz}$ | – | $0.25L_{\text{pp}}$ | |

$^\star$ Above the water line; $^\dagger$ From bow; $^\ddagger$ Above the keel

Three approaches are considered for the parameterization of the hull-form variations. Specifically, CAD-based models [36], free-from deformation (FFD) [38], and Akima surfaces [39]. The number of design variables spans from 10 to 22. Additionally in order to tackle the so-called Curse of Dimensionality (CoD), reduced-dimensionality parameterizations are considered based on the Karhunen-Loève expansion (KLE) – which is equivalent to proper orthogonal decomposition (POD) and at the discrete level principal component analysis (PCA) – of the shape modification vector provided by FFD. Methodological details and applications of KLE/POD/PCA parameterizations of airfoil sections and hull forms may be found in [40]–[42] and [43], respectively. A novel approach based on a reduced-dimensionality parametric model embedding [44] is used in AVT-331.

Different physical models (governing equations) and solvers with different computational grid sizes are considered for the multi-fidelity solution of design problems L3-Sea-1 and L3-Sea-2, which are summarized in Table 11. Computational tools span from Reynolds-averaged Navier-Stokes to nonlinear 3D potential flow solvers, from linear 3D potential flow to strip theory models. Figure 5 shows the pressure distribution and wave elevation as evaluated by CFDShip-Iowa (RANS) and WARP (potential flow) at Fr=0.28 in calm water. In addition to physical models, tight/loose/none coupling is considered of hydrodynamics with rigid body equation of motions for resistance evaluation. Details of the L3 sea benchmark are presented in [45].

Finally, the L2 sea benchmark (L2-Sea-1) investigates the hull-form optimization of the same DTMB 5415 model. To facilitate portability and enhance ease of use, the L2 problem is solved with the freely distributable potential flow solver, WARP. (See Table 12, with different spatial discretization levels, thereby providing seven levels of fidelity.) Conditions, objectives, and constraints are the same used for the L3-Sea-1 benchmark (Table 11). The shape parameterization stems from the dimensionality reduction procedure presented in [13], [46] and uses a superposition of 14 global modification functions. All elements required to run the L2 sea benchmark are managed by a wrapper and freely distributed. Details of the L2 sea benchmark are presented in [28].

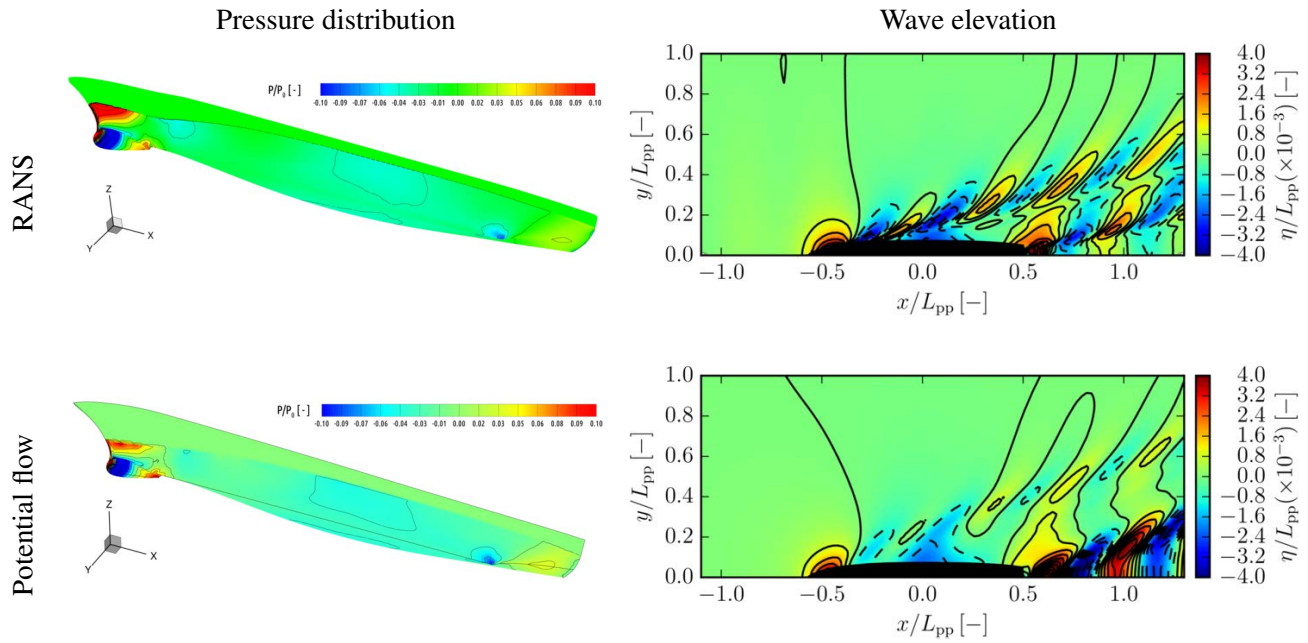



**Figure 5: Pressure distribution and wave elevation of the DTMB 5415 evaluated in calm water at Fr=0.28 by CFDShip-Iowa (RANS) and WARP (potential flow).**

## 5 SUMMARY OF METHODS AND CATEGORIZATION

The AVT-331 team is studying a number of MF methods for application to military vehicle design. The team applied these methods to a variety of benchmark problems to evaluate the potential benefits of multi-fidelity approaches for system-level design. General methodological categories include local and global methods, gradient-based and gradient-free approaches, surrogate-based (or machine-learning based) and surrogate-free methods. Methods are generally developed and tested for finding local/global design optima, or providing globally accurate models, or a suitable combination of both objectives. Some methods of key interest to AVT-331 are briefly described in this section. Although not comprehensive, this summary provides insights into the diversity of team analytical capabilities. Information about which methods are applied to which benchmarks is captured later in Table 14 (each method was typically applied to all the L1 benchmarks, but often only to one L3 vehicle benchmark).

### 5.1 Multi-Fidelity Sparse Polynomial Chaos Expansion (MFSPCE) Surrogate Models with Gradient Enhancement and Local Optimization

The Air Force Research Laboratory (AFRL) and University of Dayton (UD) have focused on the construction of surrogate models through Polynomial Chaos Expansion (PCE) and kriging approaches and have applied these models to design space exploration of L1 through L3 problems and gradient-based optimization of an L3 aircraft problem [60]. The MF approach involves computation of an additive-multiplicative bridge function

**Table 11: Summary of L3 sea benchmark problems.**

| Id. | Conditions* | Objectives | Constraints |
|---|---|---|---|
| L3-Sea-1 | • Calm water<br>• Fr = 0.28 | • Resistance | • Constant length between perpendiculars<br>• Displacement greater than or equal to the original<br>• Limited variation (5%) of beam over all and draft<br>• Reserved volume for the sonar in the dome |
| L3-Sea-2 | • Regular head waves<br>• $\lambda/L_{pp}$ = 1.2<br>• $H/\lambda$ = 1/30<br>• Fr = 0.28 | • Resistance mean value<br>• Seakeeping merit factor (considering RMS of vertical velocity at flight deck, vertical acceleration at bridge, pitch angle) | • Constant length between perpendiculars<br>• Displacement greater than or equal to the original<br>• Limited variation (5%) of beam over all and draft<br>• Reserved volume for the sonar in the dome |

* $\lambda$ is the wavelength and $H$ is the wave height

in an all-at-once fashion using sampled data [61]. The approach was tailored to a PCE basis and was integrated into a gradient-based optimization process in manner similar to trust-region model management [62] (see UMF method in Section IV.B), where the MF model is constructed and exploited for the line search of each step of

**Table 12: Solvers for the L3 sea benchmark problems.**

| Fidelity level | Solver type | Code | Reference | Proposed use in AVT-331 |
|---|---|---|---|---|
| High | RANS | $\chi$navis | [47] | Resistance and seakeeping |
| | | CFDShip-Iowa | [48] | |
| | | Fluent | [49] | |
| | | Fine Marine | [50] | |
| | | Star-CCM+ | [51] | |
| | | ISIS-CFD | [52] | |
| | | ReFresco | [53] | |
| Medium | Nonlinear PF (3D) | Rapid | [54] | Resistance |
| | | Shipflow (Basic, Motions) | [55] | Resistance and seakeeping |
| Low | Linear PF (3D) | WARP | [56] | Resistance |
| | | ITU-DAWSON | [57] | |
| | | SWAN (linear version) | [58] | Resistance and seakeeping |
| | Strip theory | SMP | [59] | Seakeeping |

a quasi-Newton optimization procedure. Rumpfkeil et al. have improved the efficiency and scalability of the underlying PCE methodology using compressed sensing [63], [64] and have extended the framework to permit local enrichments to capture behaviors not well described by global models [65]. Benchmarking activity reported by Rumpfkeil and Beran [65] will serve as a basis for some of the L1 and L2 benchmark studies in AVT-331.

### 5.2 Unified Multi-Fidelity (UMF) Quasi-Newton Method

The UMF optimization approach developed by Bryson [66] at AFRL and UD leverages the MFSPCE formulation just described (without compressed sensing) and successfully reduced the cost of unconstrained optimization of a lambda wing with a modest number of aerodynamic and structural design variables [67]. In some cases, UMF found designs more optimal than designs identified through HF analysis. The UMF procedure relies on the computation of objective derivatives to establish line search direction and to construct the line-search MF model in a manner that mitigates the CoD. First-order consistency is enforced to ensure consistency with conventional gradient-based optimization for small design steps, although the construction of the MF model over a broader space enables better decisions with larger design steps [66]. Of interest with this methodology is the incorporation of analytic derivatives to improve the reliability of the derivative estimate, improvement in method implementation for scaling to large numbers of design variables, and the extension of the methodology to include constraint functions (e.g., aircraft stress constraints or flutter constraints).

### 5.3 Multi-Fidelity Stochastic Radial Basis Functions (MF-SRBF)

The Institute of Marine Engineering (INM) in collaboration with the Ecole Centrale de Nantes (ECN) has focused on the development of MF stochastic radial basis function (MF-SRBF) surrogate models for design space exploration, global optimization, and uncertainty quantification. RBF with stochastic kernels are used to provide both the QoI prediction for unseen designs/conditions and the uncertainty associated with this prediction [68]. Additive corrections (bridge functions) are used between $M$ fidelity levels, which are assumed to be hierarchical with arbitrary $M \geq 2$. The prediction uncertainty is used to drive the adaptive selection of new training points. The desired fidelity level is also automatically selected based on the prediction uncertainty and the computational cost. The number of RBF centers for each fidelity level is identified based on a cross-validation metric. This allows the model to self tune, going from exact interpolation to regression if noisy training data is encountered. The method is able to handle adaptation to governing equatons, coupling, discretization, and convergence. Recent applications include the design optimization of a SWATH (small-waterplane-area twin hull) using RANS and PF solvers [69] and a destroyer hull using RANS with adaptive grid refinement and diverse adaptive sampling methods [70]. Recent extensions are presented in [71] including the capability of managing an arbitrary number of fidelity levels including noisy CFD outputs. Interestingly, experience shows that adding intermediate fidelity levels improves the handling of noisy functions [71]. MF-SRBF models are closely related to MF Bayesian Optimization.

### 5.4 Multi-Fidelity Bayesian Optimization (MF-GP, MFEI, MFPI, MES)

The Politecnico di Torino (PoliTO), Istanbul Technical University (ITU), Martime Research Institute Netherlands (MARIN), and INM have focused on the development and testing of MF methods based on Bayesian

optimization frameworks learning from MF sources. These methods are surrogate based optimization approaches that differ for either the formulation adopted for the surrogate model, or for the formulation of the acquisition function, or both. Surrogate models include linear and nonlinear autoregressive formulations of MF Gaussian Processes (MF-GPs). Those are combined with a variety of formulations for the MF acquisition function based on expected improvement (MFEI), probability of improvement (MFPI), and max value entropy search (MES). Both established implementations and original methods developed by the contributing teams are considered and compared. The interested reader is referred to [72]–[77] for further details on the individual approaches.

### 5.5 Asymmetric Multi-Objective Genetic Algorithm (AMOGA and CMA-ES)

The Italian Aerospace Research Centre (CIRA) and University of Strathclyde (UniSrath) have focused on development and testing of genetic and evolutionary algorithms specifically tailored for MF problems. The asymmetric multi-objective genetic algorithm is developed as a MF approach that relies on an asymmetric and elitist multi-objective genetic algorithm [78]. In the MF context, the optimization procedure uses the relation between the objective function and its approximation within a multi-objective and multi-level genetic algorithm. The first objective is the "exact" one, and it is usually computationally costly; the second one, instead, is usually much cheaper to evaluate, but it is also less precise. The multi-objective genetic algorithm takes advantage of this by introducing asymmetry in the evaluation of the objectives so that the cheap LF approximation is used many more times than the HF evaluation. A hybrid optimization technique is also proposed that couples an evolutionary multi-objective algorithm with a direct transcription method for optimal control problems [79].

### 5.6 Artificial Neural Networks (ANN)

Istanbul Technical University (ITU), and National Technical University of Athens (NTUA) have focused on the development and testing of machine-learning approaches based on MF artificial neural networks (ANN) methodologies [36]. These leverage multiple information sources for the construction of surrogate models, on which a global optimization method, such as the non-dominated sorting algorithm (NSGA-II; [80]), is applied to solve the global design optimization problem. ANN approaches have been successfully applied to global shape optimization of ships [36].

## 6 ASSESSMENT PROCESS

The goal of the assessment process is to collect and evaluate MF results so as to inform design communities of the potential benefits and drawbacks of MF methods, as well as to evaluate relative strengths and weaknesses of specific methods. By investigating L1, L2, and L3 benchmarks with MF methods across different vehicle types, the authors expect ample findings with which to document the expected properties of MF methods as applied to a wide variety of problems.

However, given practical resource and time limitations, MF methods studied by this RTG were not generally applied to all benchmark problems. Thus, head-to-head comparisons of specific methods are limited, which challenges the assessment of the relative strengths and weaknesses of individual methods to be expected for applications beyond those studied herein. Two aspects of this RTG's approach assists relative comparisons.

First, head-to-head comparisons are plentiful for L1 benchmarks, since all method contributors were asked to gather results for the application of their MF techniques to these simplified problems. This extensive data greatly increases the value of L1 studies. Second, the decomposition of the team into multiple sub-groups (L1, air, sea, space) provides an organizational structure enabling relative comparisons within sub-groups.

There are three pillars of the assessment process: the mathematical characteristics of the selected benchmarks, the categories of method assessment, and the quantification of method performance. Each of these are now described.

### 6.1 Benchmark Mathematical Characteristics

The AVT-331 L1 sub-group defined a number of mathematical characteristics relevant to the selection of analytic functions used to assess MF methods. These were based on discussions during meetings of AVT-331's exploratory team as well as subsequent discussions of the L1 sub-group. The characteristics targeted are described in detail elsewhere [15] and elaborated on here in terms of relevance to vehicle-level benchmarks:

1. Parametric Scalability: This characteristic indicates that the input dimension of the objective function can be increased to help evaluate method scalability. This can reflect the input complexity of L3 benchmarks.
2. Localized Behaviors: This characteristic indicates that localized behaviors of the objective function can differ significantly from the global behavior (e.g., frequency content). This can reflect vehicle-level physics localized to certain regions of the input parameter space.
3. Multimodality: This characteristic indicates that the objective function has numerous optima, making the search for global optima with local optimizers challenging. Multimodality is a challenge for L3 benchmarks whose efficient optimization may rely on local optimizers.
4. Discrepancy Type: This characteristic indicates that the objective function across different fidelity levels has different type (annotated here by the form of the lowest fidelity level). Type differences may be desirable for L3 benchmarks to increase computational efficiency in practice.
5. Fidelity Scalability: This characteristic indicates that the benchmark function is defined with a rich range of fidelity levels. This is not well studied with the current L3 benchmarks, but may become more important as the number of fidelity options increases (potential combinatorial growth with number of disciplines).
6. Discontinuous Response: This characteristic indicates that the objective is discontinuous in one or more parameters (jump behavior). This is not studied with the current L2 or L3 benchmarks, but the robust treatment of such features is an anticipated long-term challenge.
7. Noise: This characteristic indicates that the objective function is noisy (non-smooth). This can reflect, for example, noise produced by mesh regeneration in L2 and L3 benchmarks.

The degree to which each mathematical characteristic is reflected in a particular benchmark is summarized in Table 13. A characteristic not present or not well represented by a benchmark is marked "✗", whereas "✓" designates a characteristic well captured by a benchmark. The symbol "○" designates "not observed"; i.e., the behaviors are potentially present in the engineering benchmarks and cannot be ruled out *a priori* (as L2 and L3 benchmark results continue to be gathered, the team will attempt to better judge the features present in the benchmarks to evaluate their suitability for exhibiting localized features and/or multi-modality). Benchmarks

in general do not represent all characteristics, but tend to emphasize certain features, which motivates use of a broad benchmark suite.

The air benchmarks emphasize high dimensionality and the ability to evaluate parametric scalability. Even the L2-Air problem has up to 20 parameters, and the most complex L3 problem had 321 parameters. None of the L3 air and L2 space and sea vehicle benchmarks exhibit localized behaviors or multiple optima. The L3 sea vehicle problem in waves potentially include multiple optima due to the combination of resistance and seakeeping objectives. Localized behaviors or multiple optima are expected to be relevant in future studies as complexity increases, but perhaps was less evident in the current studies given the need to develop, distribute, and co-analyze a reliable benchmark. The L2 sea and space benchmarks each have more than two fidelity levels arising from numerical refinements, which enable the methods applied to those problems to be evaluated for fidelity scalability. Finally, all the engineering benchmarks (L2 and L3) exhibit noise owing to re-meshing.

**Table 13: Mathematical features exemplified in AVT-331 benchmark problems.**

| Benchmark | Parametric scalability | Localized behaviors | Multi-modality | Discrepancy type | Fidelity scalability | Discontinuous response | Noise |
|---|---|---|---|---|---|---|---|
| L1-For-1 | ✗ | ✓ | ✗ | linear | ✓ | ✗ | ✗ |
| L1-For-2 | ✗ | ✓ | ✗ | nonlinear | ✗ | ✓ | ✗ |
| L1-Ros-1 | ✓ | ✓ | ✗ | nonlinear | ✓ | ✗ | ✗ |
| L1-Ros-2 | ✓ | ✓ | ✗ | nonlinear | ✓ | ✗ | ✗ |
| L1-Ros-3 | ✓ | ✓ | ✗ | nonlinear | ✓ | ✗ | ✗ |
| L1-Ras-1 | ✓ | ✗ | ✓ | nonlinear | ✓ | ✗ | ✗ |
| L1-Ras-2 | ✓ | ✗ | ✓ | nonlinear | ✓ | ✗ | ✗ |
| L1-Ras-3 | ✓ | ✗ | ✓ | nonlinear | ✓ | ✗ | ✗ |
| L1-Het-1 | ✓ | ✓ | ✓ | nonlinear | ✗ | ✗ | ✗ |
| L1-Het-2 | ✓ | ✓ | ✓ | nonlinear | ✗ | ✗ | ✗ |
| L1-Het-3 | ✓ | ✓ | ✓ | nonlinear | ✗ | ✗ | ✗ |
| L1-Spr-1 | ✓ | ✗ | ✓ | nonlinear | ✓ | ✗ | ✗ |
| L1-Spr-1 | ✓ | ✗ | ✓ | nonlinear | ✓ | ✗ | ✗ |
| L1-Pac-1 | ✗ | ✗ | ✓ | nonlinear | ✓ | ✗ | ✓ |
| L2-Air-1 | ✓ | ○ | ✗ | nonlinear | ✗ | ✗ | ✓ |
| L3-Air-1 | ✗ | ○ | ✗ | nonlinear | ✓ | ✗ | ✓ |
| L3-Air-2 | ✗ | ○ | ✗ | nonlinear | ✓ | ✗ | ✓ |
| L3-Air-3* | ✓ | ○ | ✗ | nonlinear | ✓ | ✗ | ✓ |
| L3-Air-4* | ✗ | ○ | ✗ | nonlinear | ✓ | ✗ | ✓ |
| L3-Air-5* | ✓ | ○ | ✗ | nonlinear | ✓ | ✗ | ✓ |
| L2-Space-1 | ✗ | ○ | ✗ | nonlinear | ✓ | ✗ | ✓ |
| L2-Sea-1 | ✓ | ○ | ○ | nonlinear | ✓ | ✗ | ✓ |
| L3-Sea-1 | ✓ | ○ | ○ | nonlinear | ✓ | ✗ | ✓ |
| L3-Sea-2 | ✓ | ○ | ○ | nonlinear | ✓ | ✗ | ✓ |

## 6.2 Method Assessment Categories and Evaluation

This sub-section describes the process by which methods and their implementations are assessed against team benchmarks. The description has two parts: first, a discussion of method assessment categories and second, an outline of how methods are evaluated by the team in these categories.

The assessment categories (ACs) mirror the mathematical characteristics described above and modify what has been proposed elsewhere [5]:

1. Algorithmic efficiency (AC1): evaluates the efficiency of the MF method in terms of computational costs associated with sampling and error relative to an identified optimum. The cost metric is a percentage of a specified computational budget and is a function of number of samples at different fidelities and cost factors pre-assigned to those fidelity levels (following the L1 study [15]). Ideally, these costs are hardware independent. The error metric is normalized with respect to the identified optimum and a factor of error improvement is identified at set breakpoint percentages of computational budget (as prescribed by the benchmark sub-group; e.g., at 2%, 5%, 10%, and 100%) using the best design available up to the breakpoint cost level. For design-space exploration the error metric is computed from a pre-defined control set (following the L1 study [15]); also see Section 6.3.
2. Total cost efficiency (AC2): evaluates the efficiency of the MF method in terms of ratio of sampling cost to total computational cost. This metric seeks to evaluate overhead costs not directly related to sampling, which may be due to either algorithmic limitations or implementation inefficiencies. For optimization, this ratio is computed at each breakpoint cost level to track efficiency changes over the course of design optimization. For design-space exploration, this ratio is computed when the computational budget is expended.
3. Parametric scalability (AC3): evaluates the total computational cost growth of a MF method implementation with respect to increased dimensionality of a benchmark parameter space. Total computational cost growth may be driven by sampling requirements, overhead costs, or some combination thereof.
4. Fidelity scalability (AC4): evaluates the readiness of a MF method implementation to flexibly manage more than two fidelity levels. Independence from a predetermined fidelity hierarchy is preferred.
5. Effectiveness for localized behavior (AC5): evaluates the ability of the MF method to function when the local behavior of the objective varies significantly from a global trend.
6. Effectiveness for multi-modal behavior (AC6): evaluates the ability of the MF method to function when the objective function possesses numerous optima. (It is currently unclear to what extent this category is applicable to the current L2 and L3 benchmarks.)
7. Noise tolerance (AC7): evaluates the ability of the MF method to tolerate noise, which may be of a physical or numerical nature.
8. Effectiveness for discontinuous functions (AC8): evaluates the ability of the MF method to function when the objective function possesses a discontinuous feature (e.g., a step). (This category is only applicable to one L1 benchmark.)

### 6.3 Performance Quantification

Assessment falls into two categories: assessment of design-space exploration and assessment of design optimization. Tailored metrics are constructed for these assessments. L1 and L2 benchmarks generally reflect design-space exploration and design optimization, whereas L3 benchmarks generally revolve around optimization.

Method performance is judged according to two kinds of metrics described in detail in [15], goal-insensitive and goal-sensitive metrics. A goal-insensitive metric does not require knowledge of the global optimum and is used to characterize the effectiveness of MF methods in design exploration. A single metric is used, the root-mean-squared error between $f$ and $\hat{f}$ over a sample set, where $f$ is the design objective evaluation of highest fidelity ("truth") and $\hat{f}$ is the MF approximation to $f$.

A quasi-random sampling plan is introduced to compute the error in a manner that balances computational cost with accuracy (samples are distributed in an efficient manner) and consistency with low-bias (method evaluations use the same sampling plan but avoid foreknowledge of this plan). Owing to the cost of computing samples of $f$ at highest fidelity, the use of this metric is limited to assessment of design-space exploration for L1 and L2 benchmarks, where sampling cost is relatively low.

To assess the application of different MF methods to design optimization for a particular benchmark, a set of goal-sensitive metrics are defined, again as described in [15]. These metrics characterize the location of the computed optimum, $\mathbf{x}^*$, where $\mathbf{x}$ is the set of design parameters, and the value of the objective at that design location. For L1 benchmarks the true optimum is known [15] and the location metric representative of an error. For L2 and L3 benchmarks the true optimum is generally not known, in which case, the location metric is expressed relative to a reference point pre-defined for that benchmark.

### 6.4 Process

This section briefly describes team plans for how assessment will be carried out for the different MF methods contributed by team members. At the time of document preparation, method assessment had not yet been initiated and results from various benchmarks far from complete. As such, these plans are considered preliminary and subject to revision as additional information becomes available.

Methods will be assessed in two stages. First, the methods applied to L1 benchmarks will be evaluated against the assessment criteria described in 6.2 by the members of that sub-group. A quantitative scoring system [5] will be used in that evaluation, as well as the collection of qualitative evaluation information. The purpose of the L1 assessment is two-fold: (1) understanding the applicability of MF methods generally to the L1 benchmarks, and (2) revealing the relative strengths and weaknesses of those methods. In parallel, the domain-based sub-groups will compare MF methods applied to their benchmarks, primarily with regards to category AC1. In the second stage of method evaluation, L1 assessments will be validated using findings drawn from L2 and L3 benchmark comparisons. The purpose of this second step is to understand the degree to which L1 findings, highlighting method strengths and weaknesses, are relevant to applications of moderate to high engineering complexity.

## 7 OVERVIEW OF AVT-331 RELATED PAPERS AND OTHER PRODUCTS

The AVT-331 team has produced papers and software packages to facilitate dissemination of methods and approaches, and use by the broader community of L1, L2, and L3 benchmark problems for multi-fidelity approaches. A subset of these papers will be presented during the AVT-354 workshop and will serve as a basis for discussion on current gaps, research directions, and follow on activities on multi-fidelity methods for military vehicle design. Table 14 summarizes some of the AVT-331 papers to date.

**Table 14: AVT-331 related papers.**

| Reference | Benchmark | Method | Software | Presented at AVT-354 | Organization |
|---|---|---|---|---|---|
| Mainini et al. [15] | L1 (all) | Various | L1 packages | ✓ | PoliTO, INM, UD, UniStrath, CIRA, ITU, AFRL |
| Rumpfkeil et al. [81] | L1 (all) | MFSPCE | L1 packages | | UD, AFRL |
| Quagliarella et al. [28] | L2 (all) | CMA-ES | L2 packages | ✓ | CIRA, UD, AFRL, ITU, PoliTO, UniStrath, EPFL, INM |
| Quagliarella et al. [11] | L2-Air | CMA-ES | L2-Air package | | CIRA, INM |
| Bryson et al. [26] | L3-Air | MFSPCE, UMF, MF-GP | L3-Air package | ✓ | AFRL, UD, ITU |
| Thelen et al. [10] | L3-Air | UMF | L3-Air package | ✓ | UD, AFRL, NASA |
| Thelen et al. [82] | L3-Air | UMF | L3-Air package | ✓ | UD, AFRL, NASA |
| Serani et al. [45] | L3-Sea | MF-SRBF, MF-GP, ANN | L3-Sea FFD package | ✓ | INM, NTUA, ITU, MARIN |
| Scholcz and Kinkenberg [77] | L2-Sea | MF-GP | | ✓ | MARIN |
| Pellegrini et al. [83] | | MF-SRBF | | ✓ | INM, ECN |
| Pellegrini et al. [84] | | MF-SRBF | | | INM, ECN |
| Wackers et al. [71] | | MF-SRBF | | | INM, ECN |
| Ficini et al. [75] | | MF-GP | | | INM, UniRomaTre |
| Pellegrini et al. [85] | | MF-SRBF | | | INM, ECN |
| Liuzzi et al. [86] | L2-Sea | MF linesearch | L2-Sea package | | UniSapienza, INM |
| Pellegrini et al. [87] | L2-Sea | MF linesearch | L2-Sea package | | UniSapienza, INM |
| Tekaslan et al. [88] | | MF UQ approach | | | ITU |
| Pirlepeli [89] | | MF surrogate model | | | ITU, UniGlasgow |
| Di Fiore et al. [72] | L2-Space | DA MF-GP | | | PoliTO |
| Di Fiore and Mainini [73] | L1 | NM MF-GP | | | PoliTO |
| Di Fiore and Mainini [74] | L2-Air | DA-NM MF-GP | | | PoliTO |
| Di Fiore and Mainini [90] | L2-Air | DA MF-GP | | | PoliTO |

## 8 CONCLUSIONS

This paper overviews the objectives and activities of the AVT-331 technical team. AVT-331 is an international partnership to study multi-fidelity methods and their applicability to the system-level optimization of military vehicles. To conduct this study, the team has developed a suite of benchmarks drawn from the air, sea, and space domains, as well as analytical benchmarks of broad utility. These benchmarks target different mathematical characteristics of objective functions thought to be challenging in the exploration and optimization of design spaces. The benchmarks are of varying complexity (L1, L2, and L3), and feature two published, vehicle-level (L3) benchmarks studied by two or more organizations. The team intends to make benchmark models and software tools publicly available.

The team is evaluating several multi-fidelity surrogate modeling approaches with adaptive sampling strategies for increasing accuracy subject to a computational budget. Some methods have focused on optimization and others exploration. Some methods are gradient enhanced, and others not. At this time, results of the application of these methods to the benchmark suite are being collected and will be the subject of different AVT-354 workshop papers and will then be reported in extended form in the AVT-331 final report. A variety of assessment categories are defined to evaluate MF method strengths and weaknesses. Given resource limitations, methods are not assessed against all benchmarks, but are applied to all L1 (analytic) benchmarks. The AVT-331 team expects that the application of methods to the L1 benchmarks, coupled with application of certain methods to particular L3 benchmarks, will provide ample outcomes with which to judge the efficacy of MF methods.

## 9 ACKNOWLEDGMENTS

P. Beran and A. Thelen acknowledge support of the US Air Force Office of Scientific Research (grant 20RQCOR055, Dr. Fariba Fahroo, Computational Mathematics Program Officer). M. Diez and A. Serani acknowledge support of the Office of Naval Research (NICOP grants N62909-18-1-2033 and N62909-21-1-2042, administered by Dr. Woei-Min Lin, Dr. Elena McCarthy, and Dr. Salahuddin Ahmed). L. Mainini acknowledges support of the Visiting Professor Program of Politecnico di Torino. Distribution Statement A: approved for public release; distribution unlimited (AFRL-2022-2618).

## REFERENCES

[1] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, vol. 45, no. 1, pp. 50–79, 2009.

[2] G. Fernandez, C. Park, N. Kim, and R. Haftka, "Review of multi-fidelity models," *arXiv:1609.07196v3*, Mar. 21, 2017.

[3] B. Peherstorfer, K. Willcox, and M. Gunzburger, "Survey of multifidelity methods in uncertainty propagation, inference, and optimization," *SIAM Review*, vol. 60, no. 3, pp. 550–591, 2018.

[4] D. L. Clark, A. Makas, and R. V. Grandhi, "Status of multifidelity model management strategies in aircraft design," in *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2017, p. 4431.

[5] P. S. Beran, D. Bryson, A. S. Thelen, M. Diez, and A. Serani, "Comparison of multi-fidelity approaches for military vehicle design," in *AIAA AVIATION 2020 FORUM*, 2020, p. 3158.

[6] A. S. Thelen, L. T. Leifsson, and P. S. Beran, "Aeroelastic flutter prediction using multi-fidelity modeling of the aerodynamic influence coefficients," in *AIAA Scitech 2019 Forum*, 2019, p. 0609.

[7] A. Thelen, L. Leifsson, and P. Beran, "Multifidelity flutter prediction using local corrections to the generalized AIC," in *International Forum on Aeroelasticity and Structural Dynamics (IFASD 2019)*, 2019.

[8] K. Fuchi, E. M. Wolf, D. Makhija, N. A. Wukie, C. R. Schrock, and P. S. Beran, "Enhancement of low fidelity fluid simulations using machine learning," in *AIAA Scitech 2020 Forum*, 2020, p. 1409.

[9] C. Lupp and C. Cesnik, "Multi-fidelity aerostructural optimization of transport aircraft including a geometrically nonlinear flutter constraint," in *AIAA AVIATION Forum*, 2020.

[10] 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.

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

[12] G. Tumino, E. Angelino, F. Leleu, R. Angelini, P. Plotard, and J. Sommer, "The IXV project, the ESA re-entry system and technologies demonstrator paving the way to european autonomous space transportation and exploration endeavours," in *3rd Future Launchers Preparatory Programme Industrial Workshop, Glasgow*, 2008.

[13] A. Serani, F. Stern, E. F. Campana, and M. Diez, "Hull-form stochastic optimization via computational-cost reduction methods," *Engineering with Computers*, vol. 38, pp. 2245–2269, 2022.

[14] R. Pellegrini, A. Serani, G. Liuzzi, F. Rinaldi, S. Lucidi, and M. Diez, "A derivative-free line-search algorithm for simulation-driven design optimization using multi-fidelity computations," *Mathematics*, vol. 10, no. 3, p. 481, 2022.

[15] L. Mainini, A. Serani, M. Rumpfkeil, *et al.*, "Analytical benchmark problems for multifidelity optimization methods," *arXiv preprint arXiv:2204.07867*, 2022.

[16] L. Mainini, F. Di Fiore, A. Serani, *et al.*, "Analytical benchmark problems for multifidelity optimization methods," in *AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[17] T. R. Brooks, G. K. Kenway, and J. R. Martins, "Benchmark aerostructural models for the study of transonic aircraft wings," *AIAA Journal*, vol. 56, no. 7, pp. 2840–2855, 2018.

[18] M. Drela, "AVL overview," *available at https://web.mit.edu/drela/Public/web/avl/*, Feb. 2017.

[19] *SU2 foundation*, available at https://su2code.github.io, Dec. 2021.

[20] E. J. Nielsen and B. Diskin, "Discrete adjoint-based design for unsteady turbulent flows on dynamic overset unstructured grids," *AIAA journal*, vol. 51, no. 6, pp. 1355–1373, 2013.

[21] W. Rodden, R. Harder, and E. Bellinger, "Aeroelastic addition to NASTRAN," in *NASA Contractor Report 3094*, 1979.

[22] G. J. Kennedy and J. R. Martins, "A parallel finite-element framework for large-scale gradient-based design optimization of high-performance structures," *Finite Elements in Analysis and Design*, vol. 87, pp. 56–73, 2014.

[23] K. Jacobson and B. Stanford, "Flutter-constrained optimization with the linearized frequency-domain approach," in *AIAA SCITECH 2022 Forum*, 2022, p. 2242.

[24] J. F. Kiviaho, K. Jacobson, M. J. Smith, and G. Kennedy, "A robust and flexible coupling framework for aeroelastic analysis and optimization," in *18th AIAA/ISSMO Multidisciplinary analysis and optimization conference*, 2017, p. 4144.

[25] K. Jacobson, J. F. Kiviaho, M. J. Smith, and G. Kennedy, "An aeroelastic coupling framework for time-accurate analysis and optimization," in *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2018, p. 0100.

[26] 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.

[27] P. Spalart and S. Allmaras, "A one-equation turbulence model for aerodynamic flows," in *30th Aerospace Sciences Meeting and Exhibit*, 1992, p. 439.

[28] D. Quagliarella, P. S. Beran, D. Bryson, *et al.*, "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.

[29] G. Tumino, E. Angelino, F. Leleu, R. Angelini, P. Plotard, and J. Sommer, "The IXV project: The ESA re-entry system and technologies demonstrator paving the way to european autonomous space transportation and exploration endeavours," in *IAC 2008*, 2008.

[30] *MIMMO: Surface manipulation and mesh morphing library*, https://github.com/optimad/mimmo, OPTIMAD Engineering Srl, 2015.

[31] P. Mehta, E. Minisci, M. Vasile, A. C. Walker, and M. Brown, "An open source hypersonic aerodynamic and aerothermodynamic modelling tool," in *8th European Symposium on Aerothermodynamics for Space Vehicles*, 2015.

[32] A. Falchi, V. Renato, E. Minisci, and M. Vasile, "FOSTRAD: An advanced open source tool for re-entry analysis," in *15th Reinventing Space Conference*, 2017.

[33] A. Olivieri, F. Pistani, A. Avanzini, F. Stern, and R. Penna, "Towing tank, sinkage and trim, boundary layer, wake, and free surface flow around a naval combatant INSEAN 2340 model," DTIC, Tech. Rep., 2001.

[34] H. Sadat-Hosseini, D. H. Kim, S. Toxopeus, M. Diez, and F. Stern, "CFD and potential flow simulations of fully appended free running 5415m in irregular waves," in *World Maritime Technology Conference, Providence, RI, Nov*, 2015, pp. 3–7.

[35] A. Serani, G. Fasano, G. Liuzzi, *et al.*, "Ship hydrodynamic optimization by local hybridization of deterministic derivative-free global algorithms," *Applied Ocean Research*, vol. 59, pp. 115–128, 2016.

[36] G. Grigoropoulos, E. F. Campana, M. Diez, *et al.*, "Mission-based hull-form and propeller optimization of a transom stern destroyer for best performance in the sea environment," in *Proceedings of the VII International Congress on Computational Methods in Marine Engineering-MARINE*, 2017.

[37] A. Serani, M. Diez, F. van Walree, and F. Stern, "URANS analysis of a free-running destroyer sailing in irregular stern-quartering waves at sea state 7," *Ocean Engineering*, vol. 237, p. 109 600, 2021.

[38] D. D'Agostino, A. Serani, and M. Diez, "Design-space assessment and dimensionality reduction: An off-line method for shape reparameterization in simulation-based optimization," *Ocean Engineering*, vol. 197, p. 106 852, 2020, ISSN: 0029-8018.

[39] H. Akima, "A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points," *ACM Transactions on Mathematical Software (TOMS)*, vol. 4, no. 2, pp. 148–159, 1978.

[40] G. Robinson and A. Keane, "Concise orthogonal representation of supercritical airfoils," *Journal of Aircraft*, vol. 38, no. 3, pp. 580–583, 2001.

[41] D. J. Toal, N. W. Bressloff, A. J. Keane, and C. M. Holden, "Geometric filtration using proper orthogonal decomposition for aerodynamic design optimization," *AIAA journal*, vol. 48, no. 5, pp. 916–928, 2010.

[42] D. J. Poole, C. B. Allen, and T. C. Rendall, "Metric-based mathematical derivation of efficient airfoil design variables," *AIAA Journal*, vol. 53, no. 5, pp. 1349–1361, 2015.

[43] M. Diez, E. F. Campana, and F. Stern, "Design-space dimensionality reduction in shape optimization by Karhunen–Loève expansion," *Computer Methods in Applied Mechanics and Engineering*, vol. 283, pp. 1525–1544, 2015.

[44] A. Serani and M. Diez, "Parametric model embedding," *arXiv preprint arXiv:2204.05371*, 2022.

[45] A. Serani, S. Ficini, R. Broglia, *et al.*, "Resistance and seakeeping optimization of a naval destroyer by multi-fidelity methods," in *AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[46] A. Serani, M. Diez, J. Wackers, M. Visonneau, and F. Stern, "Stochastic shape optimization via design-space augmented dimensionality reduction and RANS computations," in *57th AIAA Aerospace Sciences Meeting, SciTech 2019*, 2019, p. 2218.

[47] R. Broglia and D. Durante, "Accurate prediction of complex free surface flow around a high speed craft using a single-phase level set method," *Computational Mechanics*, vol. 62, no. 3, pp. 421–437, 2018.

[48] J. Huang, P. M. Carrica, and F. Stern, "Semi-coupled air/water immersed boundary approach for curvilinear dynamic overset grids with application to ship hydrodynamics," *International Journal for Numerical Methods in Fluids*, vol. 58, no. 6, pp. 591–624, 2008.

[49] *ANSYS Fluent CFD Software*, available at http://www.ansys.com/products/fluids/ansys-fluent, 2016.

[50] *NUMECA Intl. FINE*$^{\text{TM}}$*/Marine CFD Software*, available at https://www.numeca.com/en_eu/product/finemarine, 2019.

[51] *CD-Adapco STAR-CCM+ 11.0 User Guide*, CD-Adapco Inc., Melville, NY, USA, 2016.

[52] P. Queutey and M. Visonneau, "An interface capturing method for free-surface hydrodynamic flows," *Computers & Fluids*, vol. 36, no. 9, pp. 1481–1510, 2007.

[53] G. Vaz, F. Jaouen, and M. Hoekstra, "Free-surface viscous flow computations: Validation of URANS code FRESCO," in *International Conference on Offshore Mechanics and Arctic Engineering*, vol. 43451, 2009, pp. 425–437.

[54] H. Raven, "Inviscid calculations of ship wavemaking-capabilities, limitations and prospects," in *22nd Symposium Naval Hydrodynamics, Washington, DC, USA, 1998*, 1998.

[55] L. Broberg, B. Regnström, and M. Östberg, *Shipflow user manual and theoretical manual*, FLOWTECH International AB, Gothenburg, Sweden, 2007.

[56] P. Bassanini, U. Bulgarelli, E. F. Campana, and F. Lalli, "The wave resistance problem in a boundary integral formulation," *Surveys on Mathematics for Industry*, vol. 4, pp. 151–194, 1994.

[57] C. Dawson, "A practical computer method for solving ship-wave problems," in *Proceedings of Second International Conference on Numerical Ship Hydrodynamics*, 1977, pp. 30–38.

[58] *SWAN2 User Manual, ship flow simulation in Calm Water and in Waves*, Boston Marine Consulting Inc., Boston, MA 02116, USA, 2002.

[59] W. G. Meyers and A. E. Baitis, "SMP84: Improvements to capability and prediction accuracy of the standard ship motion program SMP81," David Taylor Naval Ship Research and Development Center, Tech. Rep. SPD-0936-04, Sep. 1985.

[60] D. E. Bryson and M. P. Rumpfkeil, "Variable-fidelity surrogate modeling of lambda wing transonic aerodynamic performance," in *54th AIAA Aerospace Sciences Meeting*, 2016, p. 0294.

[61] D. Bryson and M. Rumpfkeil, "All-at-once approach to multifidelity polynomial chaos expansion surrogate modeling," *Aerospace Science and Technology*, vol. 70, pp. 121–136, 2017.

[62] D. E. Bryson and M. P. Rumpfkeil, "Multifidelity quasi-newton method for design optimization," *AIAA Journal*, vol. 56, no. 10, pp. 4074–4086, 2018.

[63] M. P. Rumpfkeil, D. E. Bryson, and P. S. Beran, "Multi-fidelity sparse polynomial chaos surrogate models for flutter database generation," in *AIAA Scitech 2019 forum*, 2019, p. 1998.

[64] M. P. Rumpfkeil and P. S. Beran, "Multifidelity sparse polynomial chaos surrogate models applied to flutter databases," *AIAA Journal*, vol. 58, no. 3, pp. 1292–1303, 2020.

[65] M. Rumpfkeil and P. Beran, "Multi-fidelity, gradient-enhanced, and locally optimized sparse polynomial chaos and kriging surrogate models applied to benchmark problems," in *AIAA Scitech 2020 Forum*, 2020, p. 0677.

[66] D. E. Bryson, "A unified, multifidelity quasi-newton optimization method with application to aero-structural design," Ph.D. dissertation, University of Dayton, 2017.

[67] 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.

[68] S. Volpi, M. Diez, N. J. Gaul, *et al.*, "Development and validation of a dynamic metamodel based on stochastic radial basis functions and uncertainty quantification," *Structural and Multidisciplinary Optimization*, vol. 51, no. 2, pp. 347–368, 2015.

[69] R. Pellegrini, A. Serani, R. Broglia, M. Diez, and S. Harries, "Resistance and payload optimization of a sea vehicle by adaptive multi-fidelity metamodeling," in *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2018, p. 1904.

[70] A. Serani, R. Pellegrini, J. Wackers, *et al.*, "Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels," *International Journal of Computational Fluid Dynamics*, vol. 33, no. 6-7, pp. 237–255, 2019.

[71] J. Wackers, M. Visonneau, A. Serani, R. Pellegrini, R. Broglia, and M. Diez, "Multi-fidelity machine learning from adaptive- and multi-grid rans simulations," in *Proceedings of 33rd Symposium on Naval hydrodynamics, Osaka, Japan, October 18-23*, 2020.

[72] F. Di Fiore, P. Maggiore, and L. Mainini, "Multifidelity domain-aware learning for the design of re-entry vehicles," *Structural and Multidisciplinary Optimization*, vol. 64, no. 5, pp. 3017–3035, 2021.

[73] F. Di Fiore and L. Mainini, "Non-myopic multifidelity bayesian optimization," *arXiv preprint arXiv:2207.06325*, 2022.

[74] F. Di Fiore and L. Mainini, "Non-myopic multifidelity method for multi-regime constrained aerodynamic optimization," in *AIAA AVIATION 2022 Forum*, 2022, p. 3716.

[75] S. Ficini, U. Iemma, R. Pellegrini, A. Serani, and M. Diez, "Assessing the performance of an adaptive multi-fidelity Gaussian process with noisy training data: A statistical analysis," in *AIAA AVIATION 2021 FORUM*, 2021, p. 3098.

[76] 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 *ECCOMAS Congress 2022 8th European Congress on Computational Methods in Applied Sciences and Engineering, 5-9 June 2022, Oslo*, 2010.

[77] T. Scholcz and J. Klinkenberg, "Efficient hull-form optimisation using multi-fidelity techniques," in *AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[78] D. Quagliarella, "Airfoil design using navier-stokes equations and an asymmetric multi-objective genetic algorithm," *Evolutionary Methods for Design, Optimization and Control: Applications to Industrial and Societal Problems, Eurogen*, 2003.

[79] E. Minisci and M. Vasile, "Robust design of a reentry unmanned space vehicle by multifidelity evolution control," *AIAA journal*, vol. 51, no. 6, pp. 1284–1295, 2013.

[80] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii," in *International conference on parallel problem solving from nature*, Springer, 2000, pp. 849–858.

[81] 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.

[82] 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.

[83] R. Pellegrini, A. Serani, M. Diez, J. Wackers, and M. Visonneau, "Adaptive multi-fidelity metamodelling for high-quality shape optimization," in *AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[84] R. Pellegrini, A. Serani, S. Ficini, *et al.*, "Adapt, adapt, adapt: Recent trends in multi-fidelity digital modelling for marine engineering," in *Conference on Computer Applications and Information Technology in the Maritime Industries, COMPIT 2020*, 2020.

[85] R. Pellegrini, A. Serani, M. Diez, M. Visonneau, and J. Wackers, "Towards automatic parameter selection for multifidelity surrogate-based optimization," in *The 9th Conference on Computational Methods in Marine Engineering (Marine 2021)*, 2021.

[86] G. Liuzzi, S. Lucidi, F. Rinaldi, R. Pellegrini, A. Serani, and M. Diez, "Derivative-free line-search algorithm for variable-fidelity optimization," in *AIAA AVIATION Forum*, 2020.

[87] R. Pellegrini, A. Serani, G. Liuzzi, F. Rinaldi, S. Lucidi, and M. Diez, "A derivative-free line-search algorithm for simulation-driven design optimization using multi-fidelity computations," *Mathematics*, vol. 10, no. 3, p. 481, 2022.

[88] H. E. Tekaslan, S. Yildiz, Y. Demiroglu, and M. Nikbay, "Implementation of multidisciplinary multi-fidelity uncertainty quantification methods in sonic boom prediction," in *AIAA AVIATION 2021 FORUM*, 2021, p. 3100.

[89] B. Pirlepeli, M. Nikbay, and K. Kontis, "A multifidelity aerodynamic surrogate model implementation for aeroelastic optimization of a nonplanar lifting surface," 2021.

[90] F. Di Fiore and L. Mainini, "Multifidelity domain-aware scheme for cross-regime airfoil shape optimization," in *The 11th International Conference on Engineering Computational Technology 23-25 August, 2022, Montpellier, France*, to appear, 2022.