AIAA SciTech Forum
6-10 January 2025, Orlando, FL
AIAA SCITECH 2025 Forum

10.2514/6.2025-0036

Introduction of Applied Aerodynamics Surrogate Modeling Benchmark Cases

Philipp Bekemeyer*
German Aerospace Center (DLR), Braunschweig, 38108, Germany

Nathan Hariharan†
DoD High Performance Computing Modernization Program, Alexandria, VA 22315

Andrew Wissink‡
US Army DEVCOM AvMC, Moffett Field, CA 94035

Jason Cornelius§
NASA Ames Research Center, Moffett Field, CA 94035

From aircraft design to certification, a significant volume of aerodynamic data is required to ensure optimal performance, meet regulatory standards, and maintain structural integrity. These data must span the entire flight envelope, encompassing pressure and shear stress distributions, global coefficients, and derivatives. Traditionally sourced from flight tests, wind tunnel experiments, or numerical simulations, the data are often of varying fidelity, ranging from handbook methods to high-resolution simulations. In recent years, the demand for efficient use of these data has grown, driven by advancements in artificial intelligence and machine learning, enabling the development of fast-running surrogate models. Unlike traditional high-fidelity simulations or experimental setups, which can be resource-intensive, surrogate models trained on these data sets deliver rapid predictions comparable to database queries. The AIAA Applied Aerodynamics Surrogate Modeling (AASM) group was formed to bring focus to data-driven and AI modeling in aerospace sciences, uniting experts from academia, industry, and government agencies worldwide. The AASM group prioritizes the development, accuracy, and applicability of surrogate modeling for aerospace applications, including design optimization, uncertainty quantification, systems engineering, and mission analysis—all critical to a digital engineering ecosystem. To support evaluation and comparison of methodologies, this paper introduces four benchmark cases: an aerodynamic database of integrated airfoil performance coefficients, a missile case for 6DOF generation, and two data sets focusing on surface pressure distributions. These benchmarks highlight associated surrogate modeling challenges and will be made publicly available through AIAA, offering valuable resources for the aerospace community.

I. Introduction

During aircraft design and certification aerodynamic data plays a fundamental role. It is crucial to accurately provide aerodynamic data, whether they are scalar- or vector-valued quantities in order to analyze performance, comply with certification requirements or ensure structural integrity to name just a few examples. In fact, a correct interpretation of data helps the engineers to both gain insights into complex physical phenomena and reliably investigate new technologies. In fact, aerodynamic data sets represent the common interface to other disciplines such as flight mechanics, loads analysis or overall aircraft design. The quantity and quality of these data sets highly depend on the methodology used to gather them. While flight tests are known to provide data within a real environment but at a substantial cost, numerical analyses are a cheaper alternative at a reduced fidelity level. The efficient use, combination and handling of such data sets have always been a focus of the aerodynamics community. However, in the past decade, this trend has significantly intensified due to increasing computational resources available and the rise in popularity of machine learning (ML)

*Team Leader, Institute for Aerodynamics and Flow-Technology, Lilienthalplatz 7, 38108 Braunschweig, Germany, AIAA Member
†HPCMP CREATE^{TM} Chief Scientist & Technologist, Alexandria, VA 22315, AIAA Associate Fellow
‡CREATE Air Vehicles Project Manager, AIAA Associate Fellow
§Aerospace Engineer, AIAA Member

methods. In fact, big data, machine learning, and deep learning (DL) are regularly seen as driving factors of the aerospace industry in the upcoming years [1–3].

Providing comprehensive aerodynamic data sets for an aircraft configuration is one of the key tasks for aerodynamic engineers. These data sets should contain performance values (lift and drag), handling quantities as well as loads at all flight points throughout the envelope. Parametric variations of the geometry are also of interest for design activities and can be introduced to the data set. High-fidelity analysis tools whether it be numerical simulations or wind tunnel tests are routinely used to generate such data sets. However, they inherently entail a multi-query problem and quickly become prohibitive with respect to computational cost. Also such complex analysis methods do not fulfill rapid turnaround requirements that are becoming more crucial to be competitive. Data-driven models, which are able to draw information from discrete sample points, are a potential way forward to circumvent these restrictions [4].

It should be mentioned that surrogate modeling techniques are not a new discipline and approaches like system identification have a long-standing history in aerodynamics and are a core element in nearly all modern aircraft design programs [5–7]. Over the past decades, various other surrogate models have been proposed and investigated to handle aerodynamic data sets. Notable examples are radial basis functions, polynomial interpolation, and Gaussian Processes [8–14]. More recently, neural networks have gained increased attention due to the availability of advanced hardware on the one hand and powerful libraries that can be used and adapted on the other hand. With the ambition to further improve the performance of surrogate models by transferring knowledge from other fields to aerodynamics, a multitude of investigations are available in the literature for the prediction of scalar quantities, e.g. [15–18]. A similar statement can be made for prediction of distributed quantities such as pressure distributions. Purely data-driven methods are especially well-suited for integration into other workflows, e.g. overall aircraft design or multi-disciplinary optimization. The arguably most common methodology is proper orthogonal decomposition (POD) [4] as a dimensionality reduction technique combined with an interpolation method such as radial basis functions or Gaussian Processes. Application examples within aerodynamics are widespread and can be found in [19–23]. Models are reported as easy to construct but suffer from the underlying POD assumption of linearity that conflicts with the nonlinear nature of fluid dynamics problems, especially at transonic flow conditions. Hence, several alternatives have been proposed in recent years including Isomap [21] or deep neural networks [24–30].

Inspired by the rapid advancements in surrogate modeling and its potential impact, the AIAA Applied Aerodynamics Technical Committee initiated the Applied Aerodynamics Surrogate Modeling (AASM) group. The objective of this group is to unite experts in the field, fostering an environment for collaboration, exchange of ideas, and presentation of cutting-edge research, while ensuring that the core challenges of applied aerodynamics are addressed. At the SciTech 2024 conference, three special sessions were organized to advance these objectives. These sessions featured a total of 10 papers covering a range of topics and application cases, including recent advancements and future directions for software solutions relevant to surrogate modeling [31–40]. Additionally, an open discussion group was convened to identify current limitations and blind spots in the field. A key issue highlighted during this meeting was the absence of common, publicly available benchmark cases specifically tailored to applied aerodynamic challenges. While benchmark cases are a well-established tool for evaluating methods in the general surrogate modeling domain—such as assessing classifier performance—an equivalent resource for applied aerodynamics has yet to be developed.

This paper seeks to address this gap by introducing four distinct applied aerodynamics benchmark cases. These include an aerodynamic database of integrated airfoil performance coefficients, a missile case featuring stability and control values, and two data sets focused on surface pressure distributions. The paper begins with an overview of the general setup and approach underlying these cases, followed by detailed descriptions of each benchmark data set. For each case, the background and objectives are first outlined. This is followed by insights into the data generation process and a summary of the resulting data sets. Finally, example surrogate modeling results are presented to demonstrate potential applications of these data sets and highlight the types of insights they can provide. The paper concludes with a discussion of envisioned community engagement and the steps needed to advance these benchmarks as a resource for the broader aerospace community.

II. Surrogate Modeling Benchmark Cases

To advance the field of applied aerodynamics surrogate modeling, the AIAA Applied Aerodynamics Surrogate Modeling (AASM) group has curated a set of benchmark cases tailored to the unique challenges of this domain. These benchmarks aim to provide a standardized framework for evaluating surrogate modeling methods and foster collaboration across academia, industry, and government. The initial set of cases, which includes diverse applications such as airfoil performance, missile stability, distributed surface quantities, and full aircraft configurations, will be hosted on the AIAA

AASM website. This centralized repository will serve as a living resource, expanding over time as new benchmarks are introduced to address emerging needs in aerospace design and analysis. By offering detailed datasets and clearly defined use cases, these benchmarks empower the community to test methodologies, compare results, and drive innovation in data-driven modeling for aerospace engineering.

A. Benchmark Case 1 - PALMO Database

This section introduces the motivation for creating the OVERFLOW Machine Learning Airfoil Performance (PALMO) database and the expected benefits it will provide to the aerospace community. The initial database contains airfoil performance coefficients from 52,480 simulations generated using the OVERFLOW CFD solver second-order accurate in space and fourth-order accurate in time with Spalart-Allmaras turbulence closure. These simulations are a parametrization of the NACA 4-series airfoil family over a wide range of Mach number, Reynolds number, and angle of attack relevant to many aerospace engineering applications. PALMO is well suited to be a benchmark data set for the development and testing of machine learning methods in aerospace engineering, which enables access to OVERFLOW-quality airfoil performance predictions without high-performance computing.

1. Background

Although the PALMO database was conceptualized to meet a need in the rotorcraft community, it is suitable for use in fixed-wing, rotary-wing, and any other aerospace applications requiring efficient yet accurate airfoil performance predictions.

In the aerospace community, airfoil look-up tables are often used in low- and mid-fidelity tools for the prediction of aircraft and rotorcraft performance. In the rotorcraft community, these tables are used extensively in design and analysis tools such as Blade Element Momentum Theory (BEMT) and Comprehensive Analysis (CA). For the fixed-wing community, the same procedure is carried out in applications such as lifting-line tools towards aircraft design, performance analysis, flight simulation, and design optimization. In either application, obtaining accurate results requires a tailored set of look-up tables with the design airfoils at appropriate Reynolds and Mach numbers for the unique wing or rotor design. An example rotor blade airfoil look-up table discretization is shown in Fig. 1. The rotor blade is discretized into 10 stations to capture radial changes in these parameters. Higher discretization is especially important for modeling variable-speed (varying RPM) rotor systems. In design applications with evolving geometry, numerous airfoil look-up tables are required, which becomes laborious and computationally expensive to compute with high-order accurate CFD data.

2. Data set Generation

The first release of the PALMO database includes 52,480 simulations, which is sixteen airfoil base-cubes each having 3,280 data points. For each airfoil, simulations are parametrized across Mach number, Reynolds number, and angle of attack as reported in Tab. 1.

The database is expanded beyond Mach number, Reynolds number, and angle of attack by further parametrizing the airfoil. The NACA 4-series airfoil family is fully defined by two values: thickness and camber. The full database consists of sixteen total base-cubes as shown in Fig. 2. Twelve of the base-cubes (in blue) parametrize a rectangular grid of camber and thickness, while the red stars denote additional test data. Although the data can be used in many ways,

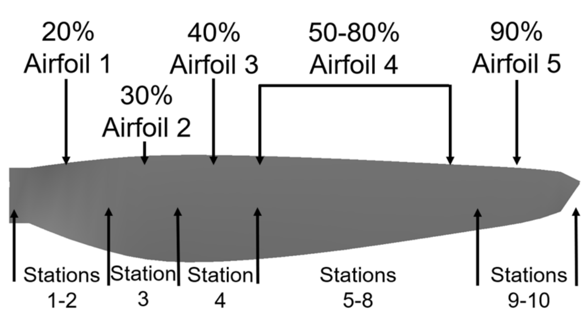


Fig. 1 Example Airfoil Look-up Table Discretization for a Rotor Blade.

Parameter	Discretization
Mach number	0.25, 0.35, 0.45, 0.55, 0.65, 0.70, 0.75, 0.80, 0.85, 0.9
Reynolds number	75k, 125k, 250k, 500k, 1M, 2M, 4M, 8M
Angle of attack	-20 to 20 degrees, 1-degree increments

Table 1 The 3,280 Parametrized Conditions in a PALMO Base-cube.

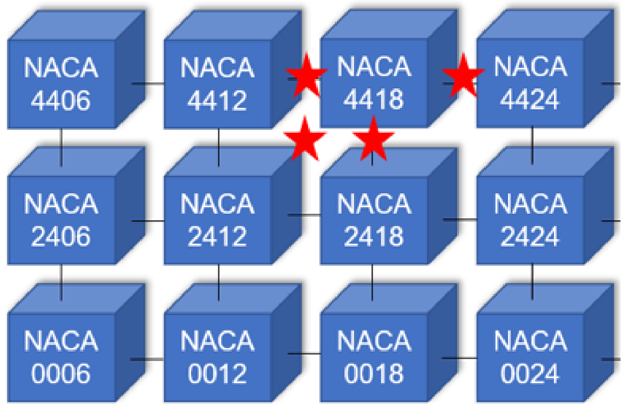


Fig. 2 PALMO Database, NACA 4-series (Red stars are suggested test data).

one suggested implementation for surrogate model development and testing is to use the blue base-cubes for model training and the red star base-cubes for testing.

Additional details about the database generation, including airfoil coordinate file generation, grid convergence studies, and OVERFLOW setup have been documented [41]. This publication also has embedded within it the airfoil coordinates and airfoil performance coefficients for all conditions.

3. Sample Surrogate Model Results

To demonstrate a possible downstream use of the PALMO database, a simple feed-forward neural network was trained on the data set. The neural network has three layers with 200 neurons per layer and was trained on the 12 blue base-cubes shown in Fig. 2. Predictions were made on part of the held-out test data, the NACA 3415. Fig. 3 shows

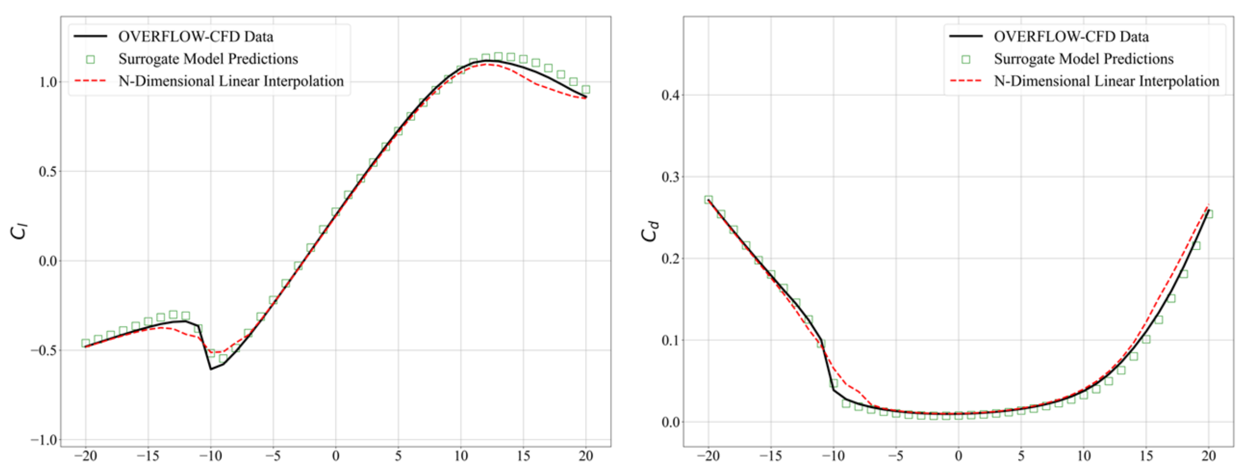


Fig. 3 PALMO Surrogate Model Predictions, NACA 3415, Mach Number 0.25, Reynolds Number 1M.

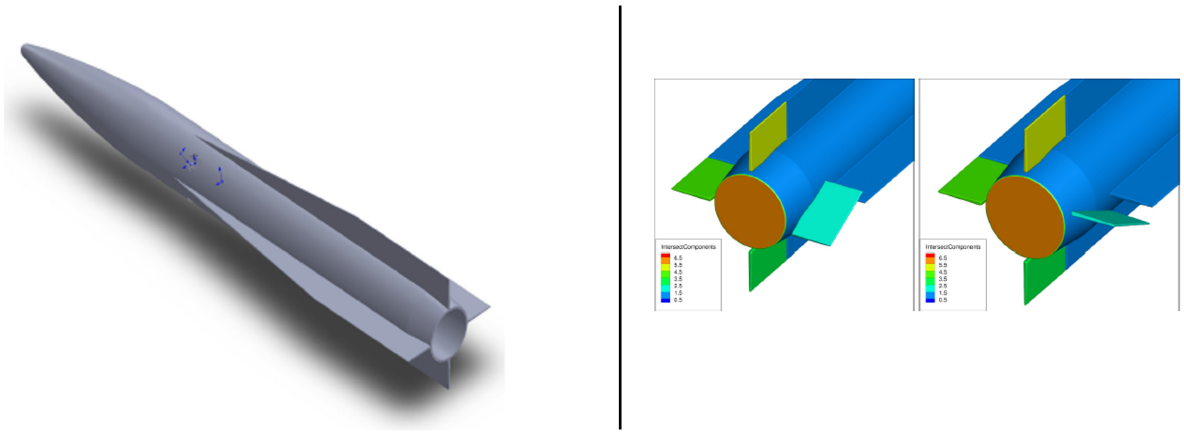


Fig. 4 ARL's Laboratory Technology Vehicle (LTV).

Parameter	Lower Bound	Upper Bound
Mach number M_∞	0.65	4.0
Angle of attack α	0	20
Sideslip angle β	0	20

Table 2 Input parameters for benchmark case 2, the LTV Missile.

the surrogate model prediction versus linear interpolation of the blue base-cubes to predict the actual OVERFLOW CFD data. The surrogate model is observed to be a better fit to the CFD data in almost all conditions. It also correctly captures the corner of the drag bucket, whereas linear interpolation of the training data does not.

This demonstrates the use of the PALMO database to predict airfoil performance with accuracy similar to OVERFLOW without high-performance computing, which is useful in a variety of aerospace applications. OVERFLOW level of accuracy airfoil look-up tables can be generated at any arbitrary combination of airfoil thickness, camber, Mach number, Reynolds number, and angle of attack within the bounds of the database. The PALMO database also provides a good benchmark data set for the development of best practices in aerospace surrogate modeling applications.

B. Benchmark Case 2 - Missile

Surrogate models can be an alternative to traditional aerodynamics databases for six-degree-of-freedom model prediction for a vehicle of interest. Given a set of inputs that include the vehicle velocity (u,v,w) , angle of attack α, sideslip β, Mach number M_∞, atmospheric conditions (ρ, P), wind velocity, and control surface deflections (δ_e,δ_a,δ_r), the surrogate is trained to provide aerodynamic forces (lift, drag, side force) and moments (pitch, roll, yaw) which can be used to calculate the accelerations and rates for guidance and control, trajectory modeling, handling qualities assessment, and more. This particular case is intended to explore techniques for generating surrogate models for a 3DOF or 6DOF model of a generic missile case using both steady and time-dependent data sets.

The US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) introduced the Laboratory Technology Vehicle (LTV) as an engineering testbed in a public Technical Report [42] as well as at the 2021 AIAA SciTech Meeting [43]. The LTV (compare Fig. 4) is composed of a symmetric cylindrical body with four fins and moving control surfaces. For this initial data set steady calculations of the LTV are performed using two different fidelity solvers; low-fidelity semi-empirical calculations using Missile DATCOM [44] and high-fidelity implicit RANS calculations using HPCMP CREATE ™ AV Kestrel [45]. The tradespace covers a range of Mach numbers M_∞, angles of attack α, and sideslip angle β to create a 3D, or three-degree-of-freedom surrogate.

Table 2 shows the input variables for the 3D surrogate of the LTV. M_∞ is varied between 0.65 and 4.0, and both α and β are varied between 0^o and 20^o. Calculations of both Missile DATCOM and Kestrel are performed at $3x3x3$ (27 samples), $6x6x6$ (216 samples), and $9x9x9$ (729 samples) uniform grid sampling points. Also included are 200 and 500 Latin Hypercube Sampling (LHS) points. These different fidelity results at different sampling points are intended to give users the opportunity to experiment with different combinations of sampling, fidelity, and surrogate construction techniques.

In the future this benchmark case will add pitch, roll, and yaw control deflections to create a 6D design space which could then act as a 6DOF model for stability and control, flight dynamics, or other analyses. It will also potentially include time-dependent data which could be used to understand dynamic control. Finally, further variables that may be considered are altitude and temperature. The size of the data set for this case is thus expected to grow with time.

C. Benchmark Case 3 - Airfoil

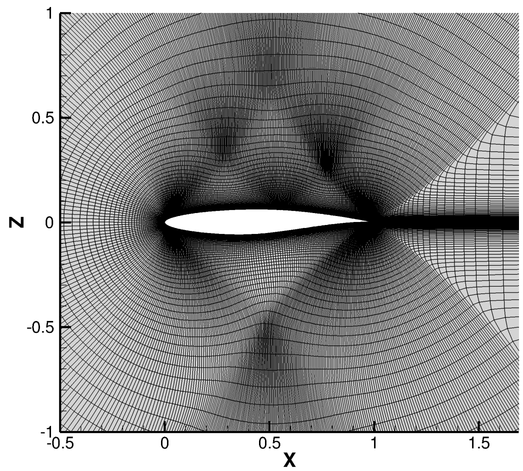
During aircraft design and optimization typically a configuration is varied based on a specific parameterization. Besides integral performance as well as stability and control quantities surface distributions such as the pressure or skin fiction distribution are often also of interest. These might be used to derive loads and perform further analysis. Hence, the third benchmark case focuses on the prediction of the distributed surface quantities, i.e. pressure coefficient and skin friction coefficients. The well-established transonic airfoil RAE2822 is selected to serve as baseline upon which the parameterization is applied and new designs are derived. For the parameterization the class shape function transformations(CST) is employed to model and modify the airfoil's geometry as outlined in [46]. The CST parameterization describes a two-dimensional geometry using a combination of a class function $C(x/c)$, a shape function $S(x/c)$ based on Bernstein binomials, and an additional term for the trailing edge thickness:

$$
\begin{aligned}
&\frac{\mathrm{z}}{\mathrm{c}} = C\left(\frac{\mathrm{x}}{\mathrm{c}}\right) S\left(\frac{\mathrm{x}}{\mathrm{c}}\right) + \frac{\mathrm{x}}{\mathrm{c}}\,\frac{\Delta z_{TE}}{c} \\
&C\left(\frac{\mathrm{x}}{\mathrm{c}}\right) = \left(\frac{\mathrm{x}}{\mathrm{c}}\right)^{N_1}\left(1-\frac{\mathrm{x}}{\mathrm{c}}\right)^{N_2} \quad \text{for} \quad 0 \leq \frac{\mathrm{x}}{\mathrm{c}} \leq 1 \\
&S\left(\frac{\mathrm{x}}{\mathrm{c}}\right) = \sum_{i=0}^{n}\left[X_i K_{i,n}\left(\frac{\mathrm{x}}{\mathrm{c}}\right)^{i}\left(1-\frac{\mathrm{x}}{\mathrm{c}}\right)^{n-i}\right]
\end{aligned}
\tag{1}
$$

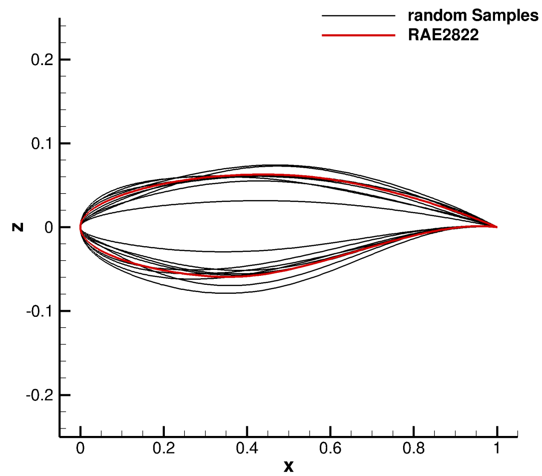
In this equation, $K_{i,n} = \frac{n!}{i!(n-i)!}$. The exponents N_1 and N_2 are chosen to reflect the desired geometry type. For an airfoil, typically $N_1 = 1/2$ and $N_2 = 1$ are used, as $\sqrt{x/c}$ produces rounded leading edges and $(1 - x/c)$ leads to sharp trailing edges. The weight factors X_i represent the design variables. Note that, one equation with a set of coefficients is used for the upper surface $X_{u,i}$ and another for the lower surface $X_{l,i}$. CST parameterization ensures C^2 continuity of the surfaces and effectively captures a range of smooth airfoil shapes. Ten design parameters (five for the upper surface and five for the lower surface) define this parameterization. However, in order to ensure C^2 continuity, the first CST parameter on the lower surface $X_{l,1}$ is chosen equivalent to the first parameter on the upper surface $X_{u,1}$, effectively resulting in nine design parameters labeled CST_i. Even though the theoretical range for CST parameters is $[-1, 1]$ to ensure somehow feasible airfoils the CST parameters are bound relative to the baseline RAE2822 transonic airfoil design $\boldsymbol{x}_0$ allowing for a maximum deviation of 50%. The resulting upper and lower bounds for all 9 CST design parameters are given in Tab. 3. Besides the CST parameters also the global flow parameters Mach number, angle of attack and Reynolds number are varied. This enables to investigate airfoil performance and occurring loads throughout the theoretical flight envelope. Bounds are also given in Tab. 3 and are supposed to be representative of an operating envelope a civil aircraft configuration encounters.

The CFD solver DLR TAU is used to solve the Reynolds-Averaged Navier-Stokes (RANS) equations in conjunction with the Spalart-Allmaras turbulence model to construct the data set. Once a specific set of CST parameters is defined, a mesh deformation routine is employed to translate the new surface shape to the volume mesh reusing an initial mesh. Fig 5 illustrates the initial volume mesh in close proximity to the airfoil as well as sample airfoil profiles together with the RAE2822 baseline. A central scheme with scalar artificial dissipation for the inviscid fluxes discretization and the Green-Gauss approach for the computation of the exact gradient of viscous and source terms are used. A four level multigrid scheme is employed to accelerate convergence. The simulations are regarded as sufficiently converged once the density residual is below $1 \cdot 10^{-8}$. This has been achieved for the majority of the simulations but not for all due to challenging flow conditions, e.g. high Mach number and angles of attack, combined with unconventional, aerodynamically not beneficial airfoil shapes. These simulations are terminated once the maximum number of 100,000 iterations has been reached and a sufficient convergence of the integral coefficients has been ensured.

Note that, the ambition of this data set is not be the most physical accurate representation of the flow around the geometry - which would potentially call for a more sophisticated turbulence model and/or stricter numerical settings - but instead to provide a feasible data set that is representative for the underlying applied aerodynamic surrogate modeling challenges.



(a) Volume mesh around the airfoil



(b) Sample airfoil shapes within the benchmark case 3

Fig. 5 Volume mesh around the airfoil surface together with several airfoil shape samples which are part of the database.

A Design of Experiment (DoE) based on a Sobol [47] sequence is used to determine in total 600 combinations of the parameters, CST plus operating parameters, within their bounds. Employing a quasi-random low-discrepancy sequence has the benefit of trying to optimally invest the available computational budget, hence, spreading simulations equally throughout the design space. Moreover, it enables adding further samples at a later point in time since the Sobol sequence is deterministic. Out of the 600 simulations in total 3 simulations did not yield a result because the combination of CST parameters did result in an invalid mesh. Hence, in total 597 surface results are available within benchmark case 3. These are split in a training set of 497 and a fixed test set of 100 solutions. Note that, if the model of interest requires a validation set, like most deep learning methods do, these validation samples should be taken from the 498 solutions labeled training data set. The data set is accessible through the website of the applied aerodynamics surrogate modeling challenge and consists of 597 surface solution files in ASCII format. Each file comprises first the

Parameter	Lower Bound	Upper Bound
Mach number	0.2	0.7
Angle of attack	-3.0	5.0
Reynolds number	$1e^6$	$6.5e^6$
$X_{u,1}$ = CST_1	0.0644	0.1932
$X_{u,2}$ = CST_2	0.0688	0.2064
$X_{u,3}$ = CST_3	0.0961	0.2883
$X_{u,4}$ = CST_4	0.0961	0.2882
$X_{u,5}$ = CST_5	0.1010	0.3030
$X_{l,2}$ = CST_6	0.0680	0.2039
$X_{l,3}$ = CST_7	0.1126	0.3377
$X_{l,4}$ = CST_8	0.0381	0.1143
$X_{l,5}$ = CST_9	-0.0586	-0.0195

Table 3 Input parameters with their lower and upper bound for the benchmark case 3. Note that values for the CST parameters are rounded to four digits after the decimal dot for better visualization.

parameters used for this specific simulation together with the resulting global aerodynamic coefficient followed by a data block containing the surface coordinates, normals, pressure coefficients as well as skin friction coefficients.

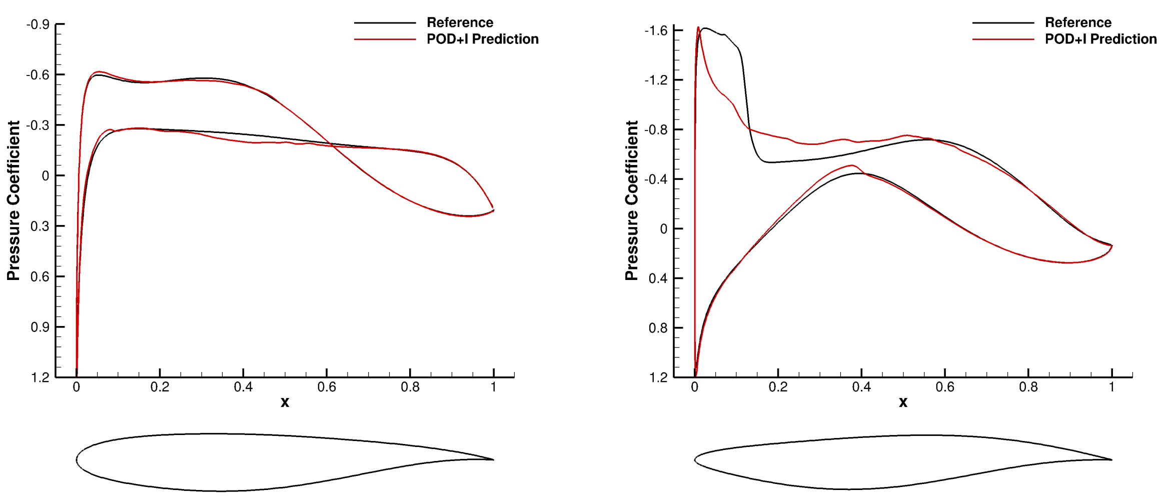
1. Sample Surrogate Modeling Results

Next some exemplary surrogate modeling results are presented for the benchmark case 3 using the 497 training solutions to build a model and evaluating its prediction accuracy by comparing results to the 100 test solutions. The intent is to showcase how the usage of the data set could look like and what type of information could in general be presented. This is by any means not meant to be an exhaustive set of results but should rather be seen as an initial guidance when approaching this case.

A surrogate model combining proper orthogonal decomposition (POD) combined with an interpolation method acting on the latent space is the surrogate model of choice for showcasing some exemplary results for predicting surface pressure distributions. Such a model is oftentimes also labeled as a POD+I model. More specifically a radial basis function interpolation method based on a thin-plate spline is used as interpolation model. Surrogate Models are constructed using the Surrogate Modeling for AeRo-data Toolbox in python (SMARTy) developed at DLR [48]. Model hyperparameters are selected based on previous experience and are a retained relative information content of 0.9999, a linear trend function for the interpolation model. Moreover, the mean is subtracted from all snapshots as a pre-processing step and input parameters are scaled to unit-hypercube. Since, strictly speaking, no data is needed for model validation during the training process, as hyperparameters are fixed beforehand, all 497 training samples are used for model training. A straightforward extension would be to reserve some of the 498 solutions for model validation and employ a model selection algorithm to automatically determine an optimal set of hyperparameters for this specific data set. Once the model has been trained it can be used to rapidly predict solutions at all conditions within the test set. Such solutions can afterwards be compared to the ground truth values and global error metrics such as the mean absolute error, the root mean squared error and the coefficient of determination R^2 can be computed. For the POD+I model described above global errors are reported in Tab. 4. Besides global errors also individual solutions can be compared to the reference values in more detail. Figure 6 shows comparisons of the surface pressure distributions for two selected samples within the test set. While the model yields accurate results at benign flow conditions (compare Fig. 6a) the

mean absolute error	the root mean squared error	coefficient of determination R^2
0.0313698	0.07793855	0.97371587

Table 4 Global error values of the exemplary POD+I model on the entire test set for benchmark case 3



(a) Test sample number 4 **(b) Test sample number 81**

Fig. 6 Sample POD+I surrogate modeling results for benchmark case 3

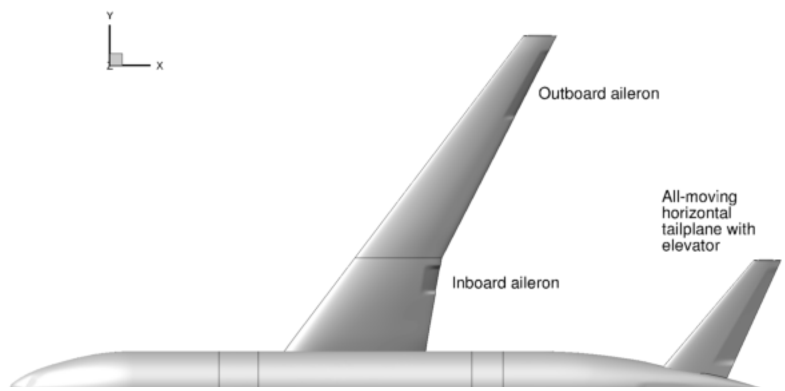


Fig. 7 Control surface deflections of the NASA CRM. Only the half-configuration is shown for better visualization.

prediction accuracy deteriorates once non-linear flow phenomena such as shocks are present (compare Fig. 6b). This is well-known for POD+I models and should not be discussed further here. Instead it clearly emphasizes that more advanced modeling techniques are needed to achieve accurate surrogate model predictions for this benchmark case. It could also be of interest to predict the skin friction coefficients alongside the pressure coefficients. Solutions including local error estimates as well as other error metrics would further enable insight into employed modeling approaches. Moreover, users are encourage to go beyond the aforementioned points in all aspects.

D. Benchmark Case 4 - NASA Common Research Model

Even though benchmark case 3 allows to investigate surrogate model performance for the prediction of distributed surface quantities, the underlying geometrical complexity and, hence, also the size of each solution is still rather small and far away from industrial-relevant configurations. In fact, the scaling of surrogate modeling techniques towards large-scale 3D cases is often overlooked. Therefore, benchmark case 4 focuses on a civil aircraft configuration, namely the NASA Common Research Model (CRM) that is well-established within the AIAA community as an open-source configuration. For this case in total 6 parameters are varied. This includes the Mach number, angle of attack, inboard aileron deflection angle, outboard aileron deflection angle, elevator deflection angle and horizontal tailplane (HTP) deflection angle. Hence, this set-up tries to emulate the challenge of providing consistent aerodynamic data sets for loads purposes throughout the envelope for a geometrically fixed configuration. Bounds for all 6 parameters are given in Tab. 5 and the geometric layout of the control surfaces are shown in Figure 7. The altitude was set to $37,000$ ft and kept fix for all investigations. This configuration and data set has been previously used to investigate surrogate model performance in [28, 30].

The CFD solver DLR TAU is used determine solutions at different parameter combinations. For all of these solutions the Reynolds-Averaged Navier-Stokes (RANS) equations together with the Spalart-Allmaras turbulence model are solved. The computational grid comprises approximately 43 million points and the corresponding surface grid, shown in Figure 8, consists of $454,404$ surface points. The grid generation approach is based on experience at DLR gained during the AIAA Drag Prediction Workshop [49, 50]. A central scheme with scalar artificial dissipation for the inviscid fluxes

Parameter	Lower Bound	Upper Bound
Mach number	0.5	0.88
Angle of attack	-2.5°	7.5°
Outboard aileron angle	-20.0°	10.0°
Inboard aileron angle	-20.0°	20.0°
Elevator angle	-10.0°	10.0°
HTP angle	-2.0°	2.0°

Table 5 Input parameters with their lower and upper bound for the benchmark case 4.

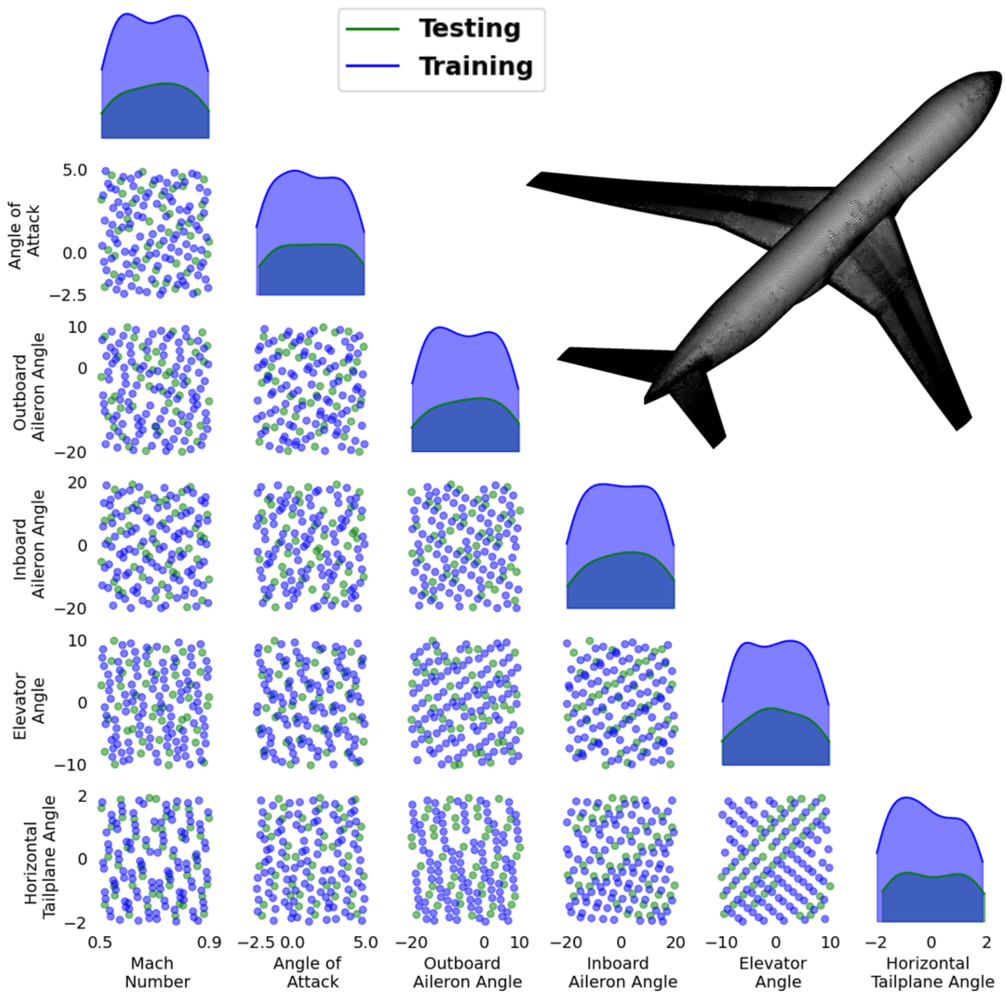


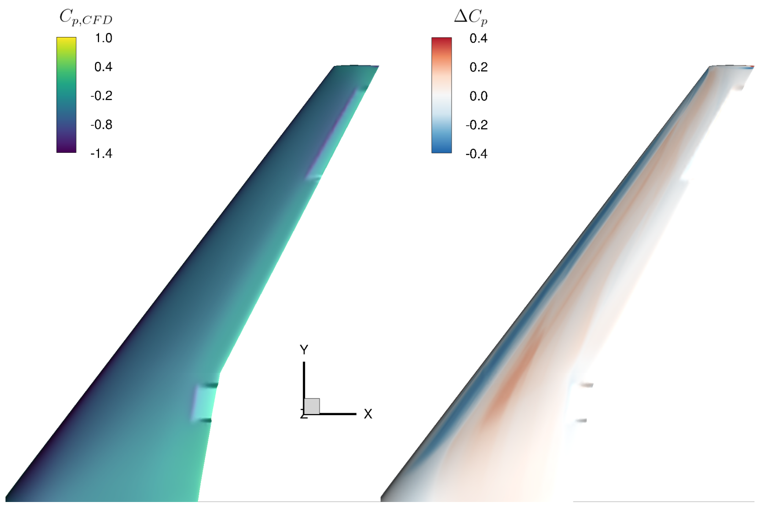
Fig. 8 Pairplot of sampling strategy and computational mesh for benchmark case 4.

discretization and the Green-Gauss approach for the computation of the exact gradient of viscous and source terms are used. Further, a 5v multi-grid scheme was employed during all simulations to accelerate convergence. Simulations were ran until a residual below e−8 was reached. Just as the grid generation approach these settings are based in experience at DLR and closely correlation to settings used during the AIAA Drag Prediction Workshops. Hence, solutions provided as part of this data set are expected to be comparable in accuracy. Even though the surface shape is assumed fixed a realistic 1g flight load deformation is obtained using an initial fluid–structure simulation. Results of this initial deformation simulation are compared to results from a European transonic wind-tunnel test campaign in [51] with good agreement. For further details, such as mesh convergence studies, turbulence model influence and comparison to experimental results as well as other codes, the interested reader is referred to the aforementioned literature. As for case 3 the ambition of the data set as part of the applied aerodynamic benchmark cases is not to be as physically accurate as possible but instead being representative for challenges faces when investigating surrogate modeling techniques.

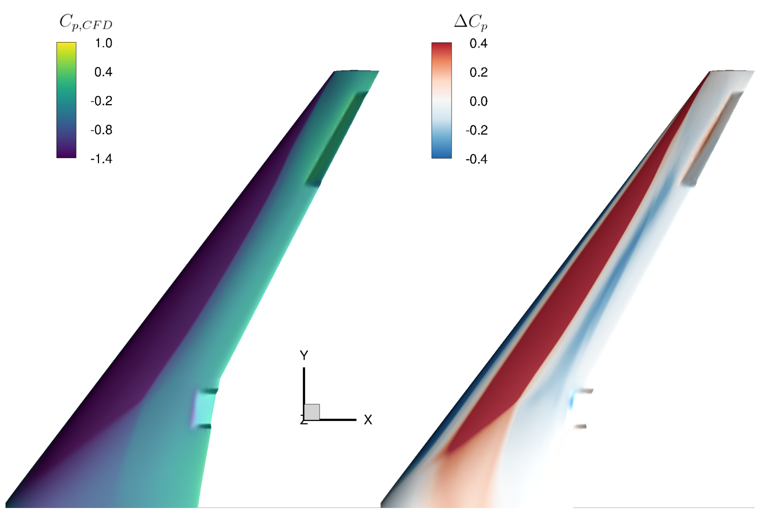
A Halton sequence was used to produce a set of in total 157 six-dimensional operational conditions, of which 149 were retained after running the CFD simulations. Out of the 149 total samples, the first 105 should be used for training, while the remaining 44 samples are exclusively reserved for testing. If for a specific model type a set of validation samples is needed, these are to be taken from the 105 training samples. In [30] the last 30 samples from the training set are used for validation during model training resulting in a 75-30-44 split. The data set is accessible through the website of the applied aerodynamics surrogate modeling challenge and consists of 149 surface solution files. Each file comprises first the parameters used for this specific simulation together with the resulting global aerodynamic coefficient followed by a data block containing the surface coordinates, normals, pressure coefficients as well as skin friction coefficients.

mean absolute error	the root mean squared error	coefficient of determination R^2
0.04173218	0.08256017	0.9326680

Table 6 Global error values of the POD+I model on the entire test set for benchmark case 4



(a) Upper wing surface for test sample number 12



(b) Upper wing surface for test sample number 13

Fig. 9 Sample POD+I surrogate modeling results showing pressure coefficient predictions for benchmark case 4

1. Sample Surrogate Modeling Results

Next some sample surrogate modeling results are presented for the benchmark case 4 using the 105 training solutions to build a model and evaluating its prediction accuracy by comparing results to the 44 test solutions. The ambition is to showcase how the usage of the data set could look like and what type of information could in general be of interest. This is by any means not meant to be an exhaustive set of results but should rather be seen as an initial guidance when approaching this case.

As for the surrogate modeling results for benchmark case 3, proper orthogonal decomposition (POD) combined with an interpolation method is selected. Again, a radial basis function interpolation method based on a thin-plate spline is used as interpolation model and the SMARTy toolbox is used [48]. Model hyperparameters are selected based on previous experience and are a retained relative information content of 0.99, a linear trend function for the interpolation model. Moreover, the mean is subtracted from all snapshots as a pre-processing step and input parameters are scaled to unit-hypercube. No data is needed for model validation during the training process of a POD+I model and, hence, all 105 training samples are used for model training. Once the model has been trained solutions at all conditions within the test set are computed and compared to the ground truth values. Global error metrics are computed and reported in Tab. 6. Individual solutions at specific parameter combinations can also be predicted by the model and then compared to the reference values to get a deeper insight into the predictive performance. Figure 9 presents such a comparison for surface pressure distributions for two selected samples within the test set. Just like before, the model yields more accurate results at more benign flow conditions (compare Fig. 9a) while the prediction accuracy deteriorates significantly once non-linear flow phenomena such as shocks are present (compare Fig. 9b). Similar to the benchmark case 3 the investigations here could focus on various aspects beyond the few sample results presented herein and the community is encourage to do so.

III. Conclusions and Outlook

This paper introduced four benchmark cases designed to address the need for publicly available data sets in applied aerodynamics surrogate modeling. These benchmarks cover a range of challenges, including aerodynamic airfoil performance coefficients, stability and control evaluation for missiles, and surface pressure distribution analysis. By providing detailed descriptions of the data generation process, resulting data sets, and example surrogate modeling results, this work aims to establish a foundation for consistent evaluation and comparison of surrogate modeling methods in applied aerodynamics. The benchmark cases are intended to serve as a shared resource for the aerospace community, facilitating advancements in design optimization, uncertainty quantification, and systems-level analysis.

Looking ahead, the Applied Aerodynamics Surrogate Modeling (AASM) group envisions an active community of researchers and practitioners engaging with these benchmarks to refine methodologies and develop innovative solutions. To support this engagement, a dedicated website has been created to host all benchmark case information, data sets, and associated documentation. This resource will serve as a central hub for the community, enabling users to access data sets, share results, and contribute feedback.

As surrogate modeling methods and processes evolve, the AASM group plans to add additional benchmarks and standardized evaluation procedures to accommodate emerging needs and ensure continued relevance. Future updates to the benchmark cases will incorporate additional complexities, such as multi-disciplinary data sets and unsteady aerodynamic phenomena, to address new challenges in aerospace design.

Further, the group aims to expand the benchmarks to include a broader range of configurations and operational conditions, such as transonic regimes, high angle of attack scenarios, and interactional aerodynamics. The integration of multi-fidelity data sets, combining low- and high-fidelity simulations, will also be explored to extend the applicability of surrogate models. By fostering collaboration among academia, industry, and government, the AASM group seeks to establish these benchmarks as a cornerstone of data-driven aerospace engineering, advancing surrogate modeling capabilities and promoting innovation across the field.

Acknowledgment

The Applied Surrogate Modeling (ASM) steering group would like to acknowledge the support of the AIAA Applied Aerodynamics TC in enabling this effort.

The first author gratefully acknowledges the scientific support and HPC resources provided by the German Aerospace Center (DLR). The HPC system CARA is partially funded by "Saxon State Ministry for Economic Affairs, Labour and Transport" and "Federal Ministry for Economic Affairs and Climate Action".

The second and third authors acknowledge support by the CREATE Applied Surrogates Institute sponsored by the U.S. Department of Defense HPC Modernization Program Office along with colleagues from the Sage software development team.

The fourth author acknowledges colleagues in the NASA Ames Research Center's Aeromechanics Office for their support of the PALMO Database development and use of the NASA High-End Compute Capability.

References

[1] European Comission, "Flightpath 2050 Europe's Vision for Aviation Maintaining Global Leadership and Serving Society's Needs," *Report of the High Level Group on Aviation Research*, 2011.

[2] Cary, A. W., Chawner, J., Duque, E. P., Gropp, W., Kleb, W. L., Kolonay, R. M., Nielsen, E., and Smith, B., "CFD Vision 2030 Road Map: Progress and Perspectives," *AIAA SciTech 2021*, 2021. AIAA 2021-2726.

[3] Liu, X., Furrer, D., Kosters, J., and Holmes, J., "Vision 2040: A Roadmap for Integrated, Multiscale Modeling and Simulation of Materials and Systems," *NASA Report*, 2018. NASA/CR—2018-219771.

[4] Lucia, D. J., Beran, P. S., and Silva, W. A., "Reduced-order modeling: new approaches for computational physics," *Progress in Aerospace Sciences*, Vol. 40, No. 1-2, 2004, pp. 51–117.

[5] Remple, R. K., and Tischler, M. B., *Aircraft and Rotorcraft System Identification*, American Institute of Aeronautics and Astronautics, Inc., 2006. https://doi.org/10.2514/4.861352.

[6] Klein, V., and Morelli, E., *Aircraft System Identification: Theory and Practice*, AIAA education series, American Institute of Aeronautics and Astronautics, 2006.

[7] Morelli, E. A., and Klein, V., "Application of System Identification to Aircraft at NASA Langley Research Center," *Journal of Aircraft*, Vol. 42, No. 1, 2005, pp. 12–25. https://doi.org/10.2514/1.3648.

[8] Powell, M. J. D., "Radial Basis Functions for Multivariable Interpolation: A Review," *Algorithms for Approximation*, Clarendon Press, USA, 1987, pp. 143–167.

[9] Simpson, T. W., Poplinski, J. D., Koch, P. N., and Allen, J. K., "Metamodels for Computer-based Engineering Design: Survey and recommendations," *Engineering with Computers*, Vol. 17, No. 2, 2001, pp. 129–150. https://doi.org/10.1007/PL00007198.

[10] Glaz, B., Liu, L., and Friedmann, P. P., "Reduced-Order Nonlinear Unsteady Aerodynamic Modeling Using a Surrogate-Based Recurrence Framework," *AIAA Journal*, Vol. 48, No. 10, 2010, pp. 2418–2429. https://doi.org/10.2514/1.J050471.

[11] Da Ronch, A., Ghoreyshi, M., and Badcock, K. J., "On the generation of flight dynamics aerodynamic tables by computational fluid dynamics," *Progress in Aerospace Sciences*, Vol. 47, No. 8, 2011, pp. 597–620. https://doi.org/10.1016/j.paerosci.2011.09.001.

[12] Han, Z.-H., Görtz, S., and Zimmermann, R., "Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function," *Aerospace Science and Technology*, Vol. 25, No. 1, 2013, pp. 177–189. https://doi.org/10.1016/j.ast.2012.01.006.

[13] Ghoreyshi, M., Jirásek, A., and Cummings, R. M., "Reduced order unsteady aerodynamic modeling for stability and control analysis using computational fluid dynamics," *Progress in Aerospace Sciences*, Vol. 71, 2014, pp. 167–217. https://doi.org/10.1016/j.paerosci.2014.09.001.

[14] Kou, J., and Zhang, W., "Data-driven modeling for unsteady aerodynamics and aeroelasticity," *Progress in Aerospace Sciences*, Vol. 125, 2021, p. 100725. https://doi.org/10.1016/j.paerosci.2021.100725.

[15] Santos, M., Mattos, B., and Girardi, R., "Aerodynamic Coefficient Prediction of Airfoils Using Neural Networks," *46th AIAA Aerospace Sciences Meeting and Exhibit*, American Institute of Aeronautics and Astronautics, 2008. https://doi.org/10.2514/6.2008-887.

[16] Kou, J., and Zhang, W., "Layered reduced-order models for nonlinear aerodynamics and aeroelasticity," *Journal of Fluids and Structures*, Vol. 68, 2017, pp. 174–193. https://doi.org/10.1016/j.jfluidstructs.2016.10.011.

[17] Wu, H., Liu, X., An, W., Chen, S., and Lyu, H., "A deep learning approach for efficiently and accurately evaluating the flow field of supercritical airfoils," *Computers & Fluids*, Vol. 198, 2020, p. 104393. https://doi.org/10.1016/j.compfluid.2019.104393.

[18] Rozov, V., and Breitsamter, C., "Data-driven prediction of unsteady pressure distributions based on deep learning," *Journal of Fluids and Structures*, Vol. 104, 2021, p. 103316. https://doi.org/10.1016/j.jfluidstructs.2021.103316.

[19] Bui-Thanh, T., Damodaran, M., and Willcox, K., "Proper Orthogonal Decomposition Extensions for Parametric Applications in Compressible Aerodynamics," *21st AIAA Applied Aerodynamics Conference*, American Institute of Aeronautics and Astronautics, Reston, Virigina, 2003. https://doi.org/10.2514/6.2003-4213.

[20] Iuliano, E., and Quagliarella, D., "Proper Orthogonal Decomposition, surrogate modelling and evolutionary optimization in aerodynamic design," *Computers & Fluids*, Vol. 84, 2013, pp. 327–350. https://doi.org/10.1016/j.compfluid.2013.06.007.

[21] Franz, T., Zimmermann, R., Görtz, S., and Karcher, N., "Interpolation-based reduced-order modelling for steady transonic flows via manifold learning," *International Journal of Computational Fluid Dynamics*, Vol. 28, No. 3-4, 2014, pp. 106–121. https://doi.org/10.1080/10618562.2014.918695.

[22] Fossati, M., "Evaluation of Aerodynamic Loads via Reduced-Order Methodology," *AIAA Journal*, Vol. 53, No. 8, 2015, pp. 2389–2405.

[23] Ripepi, M., Verveld, M., Karcher, N., Franz, T., Abu-Zurayk, M., Görtz, S., and Kier, T., "Reduced Order Models for Aerodynamic Applications, Loads and MDO," *CEAS Aeronautical Journal*, Vol. 9, No. 1, 2018, pp. 171–193.

[24] Yilmaz, E., and German, B., "A Convolutional Neural Network Approach to Training Predictors for Airfoil Performance," *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, American Institute of Aeronautics and Astronautics, 2017. https://doi.org/10.2514/6.2017-3660.

[25] Jin, X., Cheng, P., Chen, W.-L., and Li, H., "Prediction model of velocity field around circular cylinder over various Reynolds numbers by fusion convolutional neural networks based on pressure on the cylinder," *Physics of Fluids*, Vol. 30, No. 4, 2018, p. 047105. https://doi.org/10.1063/1.5024595.

[26] Bhatnagar, S., Afshar, Y., Pan, S., Duraisamy, K., and Kaushik, S., "Prediction of aerodynamic flow fields using convolutional neural networks," *Computational Mechanics*, Vol. 64, 2019, p. 525–545. https://doi.org/10.1007/s00466-019-01740-0.

[27] Thuerey, N., Weißenow, K., Prantl, L., and Hu, X., "Deep Learning Methods for Reynolds-Averaged Navier–Stokes Simulations of Airfoil Flows," *AIAA Journal*, Vol. 58, No. 1, 2020, pp. 25–36. https://doi.org/10.2514/1.J058291.

[28] Sabater, C., Stürmer, P., and Bekemeyer, P., "Fast Predictions of Aircraft Aerodynamics Using Deep-Learning Techniques," *AIAA Journal*, Vol. 60, No. 9, 2022, pp. 5249–5261. https://doi.org/10.2514/1.J061234.

[29] Sekar, V., Jiang, Q., Shu, C., and Khoo, B. C., "Fast flow field prediction over airfoils using deep learning approach," *Physics of Fluids*, Vol. 31, No. 5, 2019, p. 057103. https://doi.org/10.1063/1.5094943.

[30] Hines, D., and Bekemeyer, P., "Graph neural networks for the prediction of aircraft surface pressure distributions," *Aerospace Science and Technology*, Vol. 137, 2023, p. 108268. https://doi.org/10.1016/j.ast.2023.108268.

[31] Bekemeyer, P., Barklage, A., Hines Chaves, D. A., Stradtner, M., and Görtz, S., "Recent Advances in Data-Driven Modeling for Aerodynamic Applications using DLR's SMARTy Toolbox," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0010.

[32] Wissink, A., and Hariharan, N., "An Overview of the CREATE™ Applied Surrogates Institute," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0011.

[33] Walizer, L., Haehnel, R. B., and Allen, L., "Application of Multi-fidelity Methods to Rotorcraft Performance Assessment," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0012.

[34] Shen, Y., Patel, H. C., Xu, Z., and Alonso, J. J., "Application of Multi-Fidelity Transfer Learning with Autoencoders for Efficient Construction of Surrogate Models," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0013.

[35] Owens, K. L., Bapty, T., Eisele, F., and Scott, J., "Computational Fluid Dynamics Surrogate Based on Convolutional Neural Networks for Design Space Exploration," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0252.

[36] Anhichem, M., Timme, S., Castagna, J., Peace, A., and Maina, M., "Bayesian Approaches for Efficient and Uncertainty-Aware Prediction of Pressure Distributions," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0253.

[37] House, A. M., Kaminsky, A., Wang, Y., Alam, F., Pant, K., Tison, N., and Korivi, V., "A POD-based Input-Output Reduced-Order Modeling Framework for Acceleration of Coupled-Solver Workflows," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0254.

[38] Anand, A., Marepally, K., Safdar, M. M., Lee, B., and Baeder, J. D., "Generalizable Deep Learning Module for Rotorcraft Inverse Design Applications," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0460.

[39] Rataczak, J. A., Boyd, I. D., and McMahon, J. W., "Surrogate Models for Hypersonic Aerothermodynamics and Aerodynamics using Gaussian Process Regression," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0461.

[40] zur Nedden, P. M., Nissen, T., Reumschüssel, J. M., Orchini, A., and Paschereit, C. O., "Surrogate-based modeling for investigating and controlling the response of an axisymmetric jet to harmonic forcing," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2024. https://doi.org/10.2514/6.2024-0462.

[41] Cornelius, J., "PALMO: An OVERFLOW Machine Learning Airfoil Performance Database Version 1.0 NACA 4-Series," *NASA/TM-20240014546*, 2024.

[42] Vasile, J., Bryson, J. T., Sahu, J., Paul, J. L., and Gruenwald, B. C., "Aerodynamic Dataset Generation of a Long-Range Projectile," *CCDC Army Research Lab*, 2020.

[43] Bryson, J., Vasile, J. D., Gruenwald, B. C., Saju, J., and Fresconi, F., "Modeling and Flight Dynamics of a Projectile with Nonlinear, Roll-Dependent Aerodynamics," *AIAA SciTech 2021 Forum*, 2021.

[44] Rosema, C., Doyle, J., Auman, L., Underwood, M., and Blake, W. B., "MISSILE DATCOM User's Manual - 2011 Revision," *AFRL-RB-WP-TR-2011-3071*, 2011.

[45] Morton, S., Tillman, B., McDaniel, D. R., Sears, D. R., and Tuckey, T. R., "Kestrel - A Fixed Wing Virtual Aircraft Product of the CREATE Program," *AIAA-2009-0338*, 2009.

[46] Kulfan, B. M., "Universal Parametric Geometry Representation Method," *Journal of Aircraft*, Vol. 45, 2012, pp. 142–158. https://doi.org/10.2514/1.29958.

[47] Sobol, I. M., "On the Systematic Search in a Hypercube," *SIAM Journal on Numerical Analysis*, Vol. 16, No. 5, 1979, pp. 790–793. https://doi.org/10.1137/0716058.

[48] Bekemeyer, P., Bertram, A., Chaves, D. A. H., Ribeiro, M. D., Garbo, A., Kiener, A., Campomanes, C. S., Stradtner, M., Wassing, S., Widhalm, M., Görtz, S., Florian Jäckel, R. H., and Hoffmann, N., "Data-Driven Aerodynamic Modeling Using the DLR Smarty Toolbox," *AIAA Aviation 2022 Forum*, 2022.

[49] Keye, S., and Mavriplis, D. J., "Summary of Data from the Sixth AIAA CFD Drag Prediction Workshop: Case 5 (Coupled Aero-Structural Simulation)," *55th AIAA Aerospace Sciences Meeting*, American Institute of Aeronautics and Astronautics, 2017. https://doi.org/10.2514/6.2017-1207.

[50] Keye, S., Togiti, V. K., and Brodersen, O. P., "DLR Results of the Sixth AIAA Computational Fluid Dynamics Drag Prediction Workshop," *35th AIAA Applied Aerodynamics Conference*, American Institute of Aeronautics and Astronautics, 2017. https://doi.org/10.2514/6.2017-4232.

[51] Keye, S., and Rudnik, R., "Validation and Assessment of Turbulence Model Impact for Fluid-Structure Coupled Computations of the NASA CRM," *5th CEAS Air & Space Comference*, 2015.