

A BENCHMARK FOR UNCERTAINTY QUANTIFICATION AND OPTIMIZATION UNDER UNCERTAINTY IN CONCEPTUAL AIRCRAFT DESIGN: THE F25-OUU CHALLENGE PROBLEM

Ben D. Phillips[1], Eliot Aretskin-Hariton[2], Antoine DeBlois[3], Jasveer Singh[3], Mohammad Abu-Zurayk[4], John A. Schaefer[5] & Sean Wakayama[6]

[1]NASA Langley Research Center, Hampton, VA. United States
[2]NASA Glenn Research Center, Cleveland, OH. United States
[3]Bombardier Aviation, Montreal, Canada
[4]German Aerospace Center (DLR), Braunschweig, Germany
[5]The Boeing Company, St. Louis, MO. United States
[6]The Boeing Company, Huntington Beach, CA. United States

Abstract

In parallel with advances in multidisciplinary design optimization (MDO), uncertainty quantification (UQ) is a rapidly growing discipline within the aerospace community. Performing UQ can provide decision-makers with valuable information about candidate aircraft designs. However, system-level UQ and optimization under uncertainty have not yet achieved widespread industrial adoption. This paper extends a previously published MDO benchmark problem with a comprehensive UQ and optimization under uncertainty formulation, providing an open, complex test case to bridge the gap between academic development and industrial adoption. The benchmark defines 10 uncertain parameters (5 aleatory, 5 epistemic) spanning aerodynamics, structures, propulsion, and takeoff/landing performance, and it is structured into four progressive tiers ranging from deterministic point optimization to fleet-level multi-mission optimization under uncertainty driven by real-world operational data. Accompanying this paper is an open-source repository with a self-contained conceptual aircraft optimization problem. To our knowledge, the highest tier is the first open conceptual aircraft design benchmark to pose fleet-level optimization under uncertainty with technology-coupled uncertainties and joint trajectory optimization.

Keywords: Aircraft Design, Uncertainty Quantification, Optimization Under Uncertainty, MDO

1. Introduction

The Computational Fluid Dynamics (CFD) Vision 2030 [1] sets ambitious capability goals for uncertainty quantification (UQ) and multidisciplinary design optimization (MDO), including technology milestones such as "Incorporation of UQ for MDO," culminating in a technology demonstration of "UQ-Enabled MDO" upon completion of the roadmap in 2030. These goals build on an earlier NASA roadmap [2] that identified the needs and opportunities for uncertainty-based multidisciplinary design methods in aerospace. A progress update at the end of 2024 [3] notes an increase in UQ-related publications and mentions recent UQ challenge problems [4, 5] in addition to recent publications [6, 7]. The broader literature on uncertainty-based multidisciplinary design optimization for aerospace vehicles has been surveyed by Yao et al. [8]. An earlier survey targeted specifically at aerospace optimization under uncertainty is given by Padula, Gumbert, and Li [9]. A broader cross-industry review of classical mean–variance robust design optimization, addressing designs that remain insensitive to production tolerances, operational drift, and model uncertainties, is given by Beyer and Sendhoff [10]. While recent advancements have both improved the state-of-the-art and increased

industry adoption, significant progress is still required to achieve the goals of the CFD Vision 2030. To advance this effort, a multinational team developed a benchmark test case for MDO [11] based on a next-generation, open configuration — the DLR-F25 [12]. The aim was to establish a baseline that could accommodate progressively more complex optimization tasks in future releases. The initiative, through collaboration between research institutions and industry partners, integrated expertise across single and multidisciplinary optimization, aircraft conceptual design, aerodynamics, loads and structure, propulsion systems, and UQ. In addition to detailing discipline-specific and overall aircraft design variables, constraints, and objectives, a simplified UQ problem was presented. Building on that work, this paper contributes more-complex, system-level UQ and optimization under uncertainty problems for the DLR-F25 configuration, referred to herein as the F25-OUU challenge problem. Prior UQ challenge problems in aerospace have focused on single-discipline propagation, notably fluids [4, 5]. The present challenge complements that work at the system level by posing UQ and optimization under uncertainty on a complete aircraft configuration with coupled disciplines.

The objectives are to encourage the incorporation of uncertainty into MDO problems, specifically early in conceptual aircraft design; to drive future capability development beyond the current state-of-the-art in the field; to include uncertainties associated with technology assumptions; and to consider "as-operated" versus "as-designed" in aircraft design. The challenge problem is divided into four tiers of increasing complexity to encourage broad participation. A participant without prior UQ infrastructure can engage at Tier 0, which is a deterministic optimization with no UQ, or Tier 1, which is UQ propagation at the fixed F25 baseline with no design optimization. Higher tiers add optimization under uncertainty on the single design mission (Tier 2) and on the prescribed multi-mission fleet (Tier 3) — see (§6.5). To our knowledge, Tier 3 is is the first open conceptual aircraft design benchmark to pose fleet-level optimization under uncertainty with technology-coupled uncertainties and joint trajectory optimization.

The paper proceeds as follows. Section 2. describes the DLR-F25 reference vehicle and the operational mission mix that motivates the multi-mission tier of the problem. Section 3. introduces the uncertainty framework (physics-based scatter, direct uncertainty assignment, and technology-based uncertainty) and the UQ vocabulary used throughout. Section 4. presents the four discipline models (aerodynamics, wing mass and structures, propulsion, and takeoff/landing) and the uncertain parameters each contributes. Section 5. details the calibration of the model reproducing published F25 reference values for mass breakdown and design mission fuel burn. Section 6. consolidates the full problem specification (uncertain parameters, design variables, constraints, robust objective, and four tiers). Section 7. lists the deliverables expected from the participants.

2. The DLR-F25 Reference Configuration and Mission

The DLR-F25 concept shown in Fig. 1 is a research baseline aircraft configuration representing a 2035 entry-into-service (EIS) technology scenario for the short-medium range market segment. It was developed by the German Aerospace Center (DLR) to provide a common benchmark for evaluating advanced propulsion architectures and unconventional aircraft concepts.

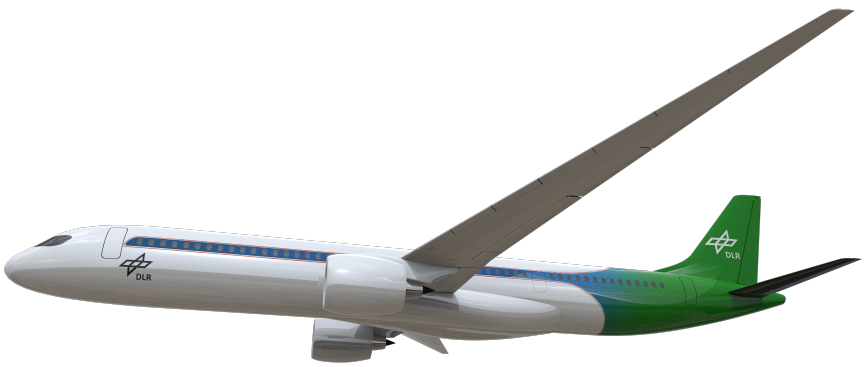
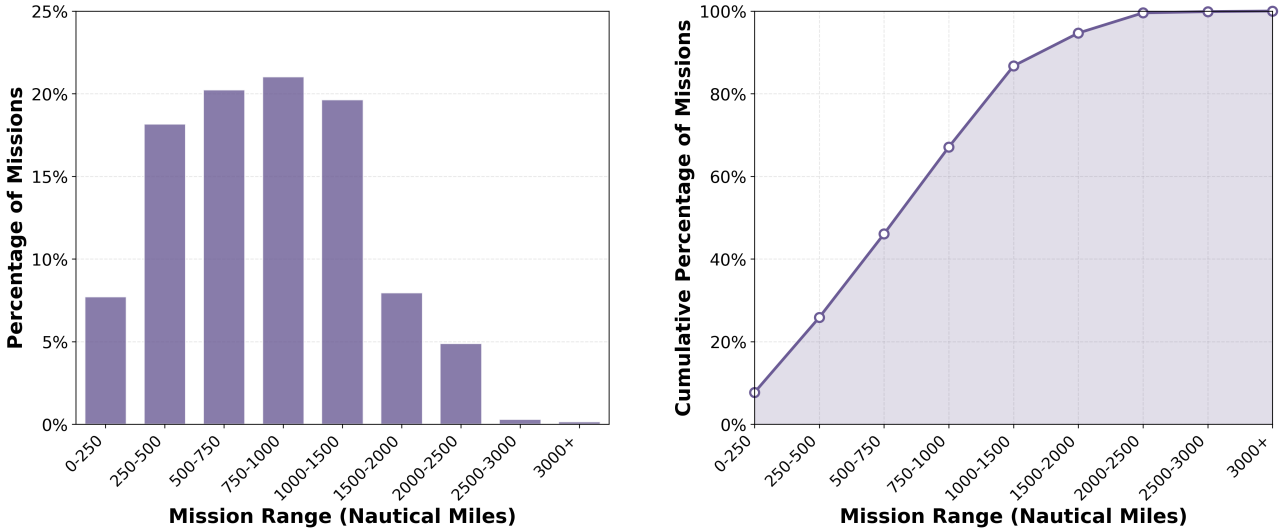


Figure 1 – The DLR-F25 configuration.

The configuration is a kerosene-powered, 239-passenger aircraft with a 2,500 NM design range, featuring evolutionary technologies that include a very high aspect ratio wing (aspect ratio 15.6, 45 m span with folding wingtips), an ultra high bypass ratio geared turbofan (bypass ratio approximately 15), carbon fiber reinforced polymer (CFRP) wing construction with advanced load alleviation and active flutter suppression, and a more-electric systems architecture. The design payload mass of 25,000 kg corresponds to 239 passengers at 95 kg each (including bags, consistent with the EASA standard [13]) plus approximately 2,300 kg of miscellaneous cargo. Relative to a 2020 baseline aircraft, the DLR-F25 achieves approximately 17.7% block fuel reduction on an 800 NM evaluation mission, with wing technologies contributing the majority of the improvement [12].
Although the F25 is sized to the 2,500 NM design mission described above and traditional design optimization assumes the aircraft is flown at that single range, real-world operations rarely match that assumption. To illustrate the disparity, T-100 segment data from the Bureau of Transportation Statistics (BTS) for calendar year 2024 were analyzed [14]. The dataset includes over 5.5 million flights by Boeing 737 and Airbus A320 family variants — aircraft representing the DLR-F25's target short-medium range market. The data indicate about 95% of all missions flown were less than 2,000 NM. Figure 2 shows the range distribution, which is heavily skewed to lower range missions. When performing aircraft optimization, designers should consider aligning the problem formulation as close as possible to "as-operated." For simplicity the challenge problem assumes a constant payload. BTS data indicate a correlation between distance flown and passenger load factors which is not addressed.



(a) Mission range histogram.

(b) Mission range CDF.

Figure 2 – Boeing 737 and Airbus A320 variants mission range departing or arriving at US airports in 2024.

The continuous BTS range distribution was discretized into the five mission ranges and frequencies given in Table 1. The discretized mission mix reflects current operations and is intended to anchor the as-operated framing of the challenge problem rather than to forecast the operating environment of the 2035 EIS DLR-F25. These ranges are flown in the multi-mission formulation (Tier 3) of the challenge problem. This multi-mission, fleet-level framing builds on prior work on conceptual-design optimization over a prescribed mission mix [15, 16, 17], on rapid integrated fleet-level modeling for aircraft system-level studies [18], and on coupled design-and-allocation optimization at higher fidelity [19].

3. Uncertainty Framework

3.1 UQ in Conceptual Aircraft Design

In conceptual aircraft design, margins have historically been applied to sensitive quantities such as weight, fuel reserves, and takeoff field length to absorb the cumulative effect of unmodeled uncer-

Table 1 – Mission range distribution prescribed for the multi-mission (Tier 3) formulation.

Mission Range (NM)	Frequency
500	30%
1000	35%
1500	20%
2000	10%
2500	5%

tainty. UQ formalizes this practice. Rather than selecting margins through rule of thumb, the practitioner prescribes a distribution or interval for each uncertain parameter, propagates samples from the joint set of uncertainties through the analysis chain, and reports the resulting statistics or intervals of every quantity of interest [8, 20]. This approach is more computationally expensive than deterministic analysis, as the design must be re-evaluated across the uncertainty samples. The benefits include the traceability of every margin to a specific physical parameter, the ability to rank uncertainty sources by impact through global sensitivity analysis [21, 22], and the faculty to optimize directly for robustness rather than for nominal performance [10].

Two types of uncertainty are treated separately throughout the paper. Aleatory uncertainty refers to irreducible random scatter, such as build-to-build manufacturing variation, day-to-day operational scatter, and measurement noise; it is represented by a probability distribution. Epistemic uncertainty refers to a (potentially) reducible lack of knowledge, such as model form, future technology assumptions, or extrapolation beyond the calibrated database; it is represented by a bounded interval with no probability density asserted [23]. The distinction matters because epistemic uncertainty can in principle be reduced by collecting more data or refining the model, while aleatory uncertainty cannot.

Comparing robust designs across participant submissions requires scalar performance metrics. A probability distribution can be sampled to produce them directly; a bounded interval cannot because it asserts no density to sample. One potential solution is to assign each epistemic interval a uniform density and compute ordinary moments from the result. However, this simplification reclassifies every epistemic parameter as aleatory, erasing the distinction just established. More sophisticated methods avoid this by operating on the bounding distributions directly. Horsetail matching, for example, matches the upper and lower CDFs to a target, yielding a single robust metric without assigning a density to the epistemic parameters [24]. Such methods come at the cost of requiring gradient information and per-design surrogate models. In an attempt to balance computational cost against the proper classification of uncertainties, the challenge problem adopts second-order probability [25]; the epistemic intervals are swept in an outer loop and the aleatory distributions propagated in an inner loop, bounding each quantity of interest by a probability box (p-box) rather than a single cumulative distribution function (CDF). The aggregate performance metrics are then read from this envelope as *intervals of statistics* — a location $\tilde{\mu}$ and spread $\tilde{\sigma}$ defined in Fig. 3 — rather than as moments of an assumed distribution.

The bounding CDFs, F_l and F_u, give the range of cumulative probability attainable at each value of a quantity of interest. The central 95% interval $[L,U]$ is taken across the envelope, with its lower edge, L, drawn from F_l and its upper edge, U, from F_u, and the location and spread are reported as the interval midpoint $\tilde{\mu}$ and half-width $\tilde{\sigma}$ (Figure 3). Drawing the two edges from opposite bounds makes $[L,U]$ the widest 95% interval consistent with the p-box, so $\tilde{\sigma}$ reflects both aleatory scatter and epistemic interval width. When no epistemic parameter is active, the bounds coincide and $\tilde{\mu}$ and $\tilde{\sigma}$ reduce to the midpoint and half-width of a single distribution's central 95% interval.

Propagation of the uncertainties through the analysis chain produces a p-box on every quantity of interest, from which the prescribed deliverable metrics (location, spread, probability of constraint satisfaction, and tail-risk measures) are computed. Numerous propagation methods are available, including direct sampling approaches [26], spectral expansions [27], and surrogate-assisted or sample-reuse approaches [28]. The accompanying repository contains a Python-based, deterministic aircraft optimization problem representing the F25 in an Aviary [29] model, but the challenge problem is agnostic to the UQ methods used. Participants are encouraged to apply their own methods and best

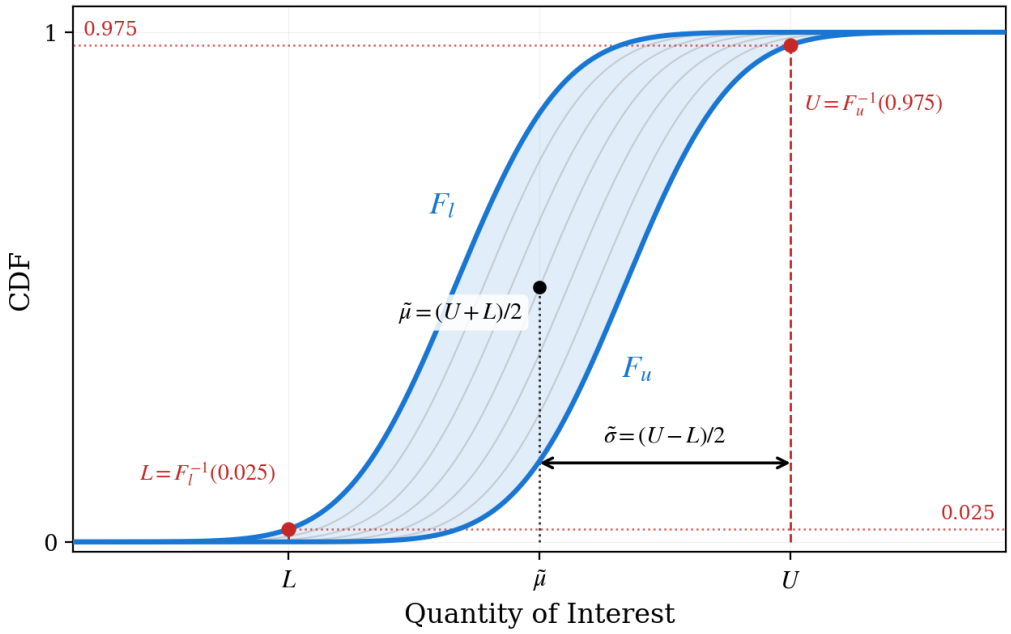


Figure 3 – Probability box (p-box) for a generic quantity of interest, bounded by the lower- and upper-quantile CDFs F_l and F_u. The central 95% interval $[L, U]$ has edges $L = F_l^{-1}(0.025)$ and $U = F_u^{-1}(0.975)$; the location $\tilde{\mu}$ is its midpoint and the spread $\tilde{\sigma}$ its half-width.

practices and to report the chosen propagation scheme and its convergence diagnostics as part of the deliverables (§7.).

A benefit of applying UQ in an MDO problem is the ability to enforce probabilistic constraints of the form $\mathbb{P}[g(x,\xi) \leq \text{limit}] \geq \alpha$, evaluated over the propagated uncertainty in ξ. A robust objective combines central performance and dispersion into a single scalar criterion. Traditionally, for a maximization problem like the present one, this takes the canonical mean–variance form

$$J(x;\lambda) = \mathbb{E}[f(x,\xi)] - \lambda \cdot \sigma[f(x,\xi)], \tag{1}$$

where the weighting parameter $\lambda \geq 0$ tunes the trade-off between expected performance and robustness [10, 30]. Because f is here bounded by a p-box rather than a single distribution, the challenge problem retains this form but substitutes the envelope metrics of Fig. 3 for the moments,

$$J(x;\lambda) = \tilde{\mu}[f(x,\xi)] - \lambda \cdot \tilde{\sigma}[f(x,\xi)], \tag{2}$$

where $\tilde{\mu}$ and $\tilde{\sigma}$ are the central-interval midpoint and half-width defined in Fig. 3. Since specific air range (SAR) is maximized, the minus sign rewards a higher central value and penalizes a wider envelope; at $\lambda = 1$ the objective reduces to $\tilde{\mu} - \tilde{\sigma} = L$, the conservative lower edge of the 95% interval. The minimization-convention literature uses the opposite sign, so that minimizing $\tilde{\mu} - \lambda \cdot \tilde{\sigma}$ similarly penalizes dispersion. Sweeping λ generates a Pareto front in the $(\tilde{\mu}, \tilde{\sigma})$ plane that defines the achievable design trade. A tail-risk measure, the conditional value-at-risk CVaR_q, introduced by Rockafellar and Uryasev [31] and applied to aerospace design optimization by Quagliarella and Iuliano [32] and Chaudhuri et al. [33], is the expectation of f conditioned on the worst $q\%$ of the propagated distribution; it may additionally be reported for designs where downside behavior matters more than dispersion alone. (The symbol q disambiguates this tail level from the constraint satisfaction level α used for probabilistic constraints in §6.3.)

Probabilistic constraints and robust objectives address distinct features of the same propagated distribution. Figure 4 illustrates the distinction. A robust objective concerns the central tendency (location and spread) of the propagated distribution, while a reliability-based probabilistic constraint enforces a bound on the tail probability of exceeding a limit. The F25-OUU challenge problem combines both into a single formulation, referred to in the literature as robust and reliability-based design optimization (RBRDO) [34, 35].

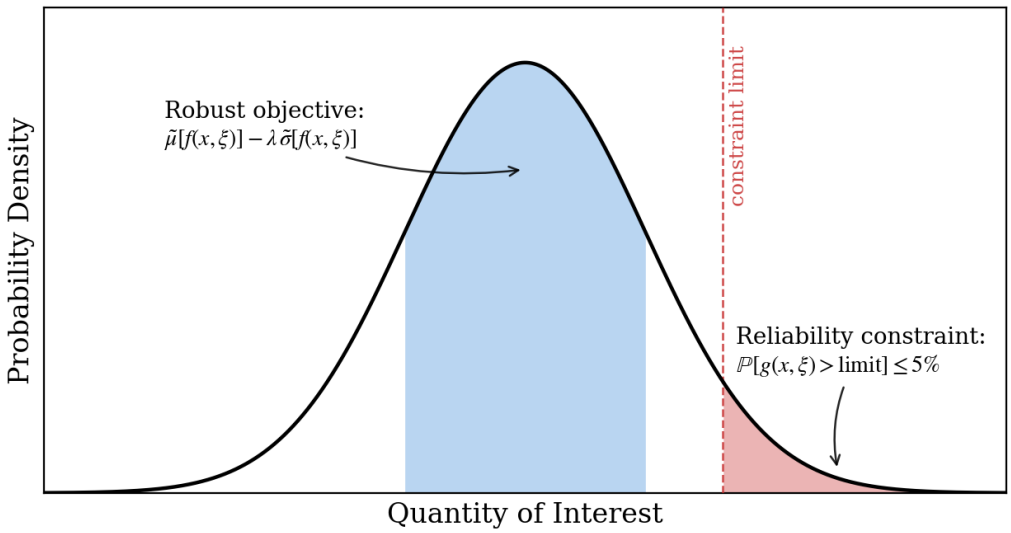


Figure 4 – Probability density of a propagated quantity of interest, with the central region (blue) and the right-tail region beyond the constraint-limit line (red). Adapted from Zang et al. [2].

3.2 Categories of Uncertainty in This Problem

The challenge problem organizes uncertainty into four categories that are instantiated, discipline-by-discipline, in §4.. *Calibrated parameter uncertainty* captures aleatory scatter derived from data fitting, where the calibration itself supplies the distribution shape (regression prediction intervals, build-to-build manufacturing scatter). *Model-form uncertainty* captures epistemic structural uncertainty about the mathematical model itself — the validity of a fitted functional form, the inclusion of a particular correction term, or the extrapolation of a calibration outside its training domain — and is represented as a bounded interval where no probability density can be defended. *Direct uncertainty assignment* is used when the upstream physics that would naturally produce variability in an intermediate quantity are not present in the conceptual design toolchain, so a distribution is assigned at the intermediate quantity directly. *Technology-based uncertainty* captures the additional risk associated with future-EIS technology assumptions and design extrapolation beyond the certified-aircraft database. As the challenge problem is scoped specifically for conceptual design, higher-fidelity specific uncertainties, such as those arising from CFD grid resolution, are not addressed. The focus throughout is on the *process* of performing UQ, on the traceability of each uncertain parameter to its physical origin, and on the (often overlooked) effective communication of study results.

3.2.1 Calibrated Parameter Uncertainty

Calibrated parameter uncertainty arises when a model's parameters are fitted to data, and the fit itself produces a quantified residual that can be propagated as an aleatory distribution. Examples of calibrated parameter uncertainty arise throughout conceptual aircraft design. Wherever a regression-based method is employed to predict a response, an inherent uncertainty exists. These regression-based methods are routinely treated as deterministic without consideration to the quality of fit, and potentially the domain of validation. The underlying data to which the model is calibrated becomes the source of the model-parameter uncertainty. The defining feature is that the dispersion is grounded in data rather than in judgment, and the shape and width of the distribution are dictated by the calibration methodology rather than freely chosen. The challenge problem includes calibrated parameter uncertainties in the aerodynamics and structures disciplines (§4.1,§4.2).

3.2.2 Model-Form Uncertainty

Model-form uncertainty is epistemic uncertainty about the mathematical assumptions of the model. This uncertainty is structural rather than parametric. No probability density on the model-form parameter is asserted; the representation is a bounded interval. In conceptual design, epistemic uncertainty

is typically encountered when trading lower computational costs for physics simplifications, for example, running a panel method for aerodynamic analysis inside an MDO problem instead of a Reynolds-Averaged Navier–Stokes (RANS) simulation. There is uncertainty associated with the panel method's aerodynamic analysis due to the simplification of physics in comparison to the higher-fidelity methods. Model-form uncertainty is not unique to low-fidelity tools; even RANS carries it through the turbulence closure, and methods exist to quantify it as an epistemic interval [36]. This trade-off is a common one, and this paper seeks to make the use of uncertainty modeling with model-form uncertainty common as well. The challenge problem assigns model-form uncertainties in the aerodynamics and structures disciplines (§4.1, §4.2), each as an epistemic interval on a structural parameter of the underlying model.

3.2.3 Direct Uncertainty Assignment

Another uncertainty challenge in conceptual design is that the upstream physics models that would naturally produce variability in an intermediate quantity often do not exist at this stage of design. Maximum lift coefficients $C_{L_{\text{max,TO}}}$ (maximum lift coefficient during takeoff) and $C_{L_{\text{max,LDG}}}$ (maximum lift coefficient during landing) in a high-fidelity environment would be the output of high-lift CFD calibrations or wind tunnel campaigns. In conceptual design, the practitioner does not have that pipeline; what is available is a single tabulated number or an empirical correlation.

Rather than constructing artificial upstream models for the sole purpose of generating variability, the challenge problem assigns distributions directly to the intermediate quantities themselves. The assigned distribution is shaped to envelop the aggregate variability that an upstream pipeline would have produced, and the uncertainty propagates downstream through the supplied analytical performance routines. This practice is standard in aerospace conceptual design. It is referred to here as *direct uncertainty assignment* to distinguish it cleanly from the calibrated and model-form uncertainties.

Three of the ten uncertain parameters in this challenge problem are direct assignment cases: the shared takeoff/landing aerodynamic parameter, t_{HL}, that scales $C_{L_{\text{max,TO}}}$ and $C_{L_{\text{max,LDG}}}$ together (§4.4), the takeoff zero-lift drag coefficient, $C_{D_{0,\text{TO}}}$ (§4.4), and the non-wing operating empty mass scaler, σ_{OEM} (§4.2). Each is assigned a distribution at the intermediate quantity rather than derived from a propagated upstream model.

Direct assignment trades traceability for tractability. A single assigned distribution cannot be decomposed into the upstream parameters whose variability it represents. A sensitivity analysis on t_{HL} cannot directly answer the question of which specific component of uncertainty in the high-lift system is responsible for a takeoff field length constraint violation. A higher-fidelity follow-on study would replace each direct assignment distribution with explicit upstream propagation, recovering that traceability at substantially higher computational cost.

3.2.4 Technology-Based Uncertainty

Concept vehicles with future EIS dates, such as the DLR-F25, include assumptions about technologies where the technology readiness level (TRL) of the subsystem under consideration is considerably lower than that of a flight-ready system. As the design deviates further from current certified vehicles, the general level of uncertainty should increase. For example, assumptions for aircraft designs with an EIS 30 years from now should model considerably more technology-based uncertainty than those for aircraft with EIS in 5 years. The challenge problem models this based on two factors: TRL maturity and deviation from currently certified aircraft. Equation 3 shows that a technology-specific uncertainty, U_{tech}, (which could take the form of a subsystem mass or cost) scales inversely with TRL maturity. Equation 4 shows a certification-based uncertainty, U_{cert}, that scales with the difference between a currently certified design, X_{cert}, and the proposed design, X_{design}. Notionally, the overall uncertainty scales as the candidate design deviates from TRL_{flight} or X_{cert}.

$$U_{tech} \propto (TRL_{flight} - TRL_{tech})\sigma_{tech} \tag{3}$$

$$U_{cert} \propto (X_{cert} - X_{design})\sigma_{cert} \tag{4}$$

The propulsion discipline (§4.3) instantiates both frameworks. The wing mass discipline (§4.2) uses the concept of U_{cert} to widen the thin-wing structural extrapolation interval as AR moves beyond the certified envelope.

4. Discipline Models and Uncertainties

Each discipline contributes a model used in the mission analysis and a set of uncertain parameters from the categories of §3.. The four disciplines are aerodynamics, wing mass and structures, propulsion, and takeoff/landing. The discipline-level uncertain parameters are summarized in table 3 and discussed in detail in §6.1. Throughout, uncertain parameters follow a prefix convention. The prefix z_X denotes an aleatory standard-normal parameter. The prefix t_X denotes an aleatory triangular parameter. The prefix u_X denotes an epistemic bounded-interval parameter. Parameters with a clear physical name and range, such as a Raymer exponent or the leading coefficient C, are named directly.

4.1 Aerodynamics

Aviary's [29] FLOPS-based [37] aero is calibrated within its empirical framework to the extent possible to match CFD data. With this calibration, FLOPS overpredicts F25 cruise drag by 11–15% relative to a CFD database covering the F25 design space, because the empirical correlations did not cover the F25's specific high-aspect-ratio, swept-supercritical geometry, and because conceptual design methods routinely embed excrescence and installation drag in their skin-friction allowances that clean CFD typically omits. A CFD-calibrated reduced-order drag polynomial is supplied with the challenge problem to close that geometry-specific gap. The general approach of building a CFD database over a transport aircraft design space and fitting a reduced-order surrogate for use in optimization is well established [38, 39]. The calibration here replaces the FLOPS compressibility/wave-drag formulation with the classical Lock fourth-power form, propagated in conceptual design MDO by Malone and Mason [40], with critical-Mach calibration informed by Krull's recent fit across 16 transport aircraft [41], plus a polynomial residual fit to the CFD database and geometry-only scaling factors on the friction and induced-drag terms inherited from FLOPS. A constant +25 drag count (dc) full-vehicle anchor offset converts the wing-body-tail (WBT) CFD into a full-airplane drag estimate. The full calibration process is multi-stage, comprising a sparse CFD database spanning the design space and a weighted-least-squares fit whose leverage-augmented prediction interval becomes the prescribed aleatory uncertainty. A complete treatment is beyond the scope of the present paper. A follow-on publication will document the CFD database construction, the regression methodology, the diagnostic checks, and the cross-validated prediction-interval calibration in full. Participants in the challenge problem need only the supplied calibrated polynomial and its prescribed uncertainty (§6.1). Both the deterministic C_D surrogate and the heteroscedastic σ_{aero} field are exposed as queryable functions, so participants are not required to reproduce the calibration to engage with the UQ formulation.

Figure 5 shows the F25 baseline drag polar at cruise Mach. The FLOPS curve (no CFD calibration) sits above the CFD. Adding the +25 dc full-vehicle offset to the raw CFD brings the data up to the calibrated full-airplane reference. The shaded band is the 95% regression prediction interval. Its standard deviation varies across both the flight envelope and the design space and becomes the prescribed aleatory aerodynamic uncertainty, propagated at every flight node as

$$\Delta C_D \;=\; z_{\text{aero}} \cdot \sigma_{\text{aero}} = z_{\text{aero}} \cdot f(M, C_L, AR, \Lambda, S, t/c), \tag{5}$$

where z_{aero} is a standard normal random variable and fixed within each uncertainty realization. The flight condition dependence captures the regression's larger scatter in the transonic drag rise region and at elevated C_L, while the geometry dependence captures the leverage inflation of the prediction interval as a candidate design moves away from the CFD training centroid, so designs with higher aspect ratio than the database carry a larger σ_{aero}, formalizing the certification-gap framework U_{cert} (§ 3.2.4) for the aero discipline. An additional epistemic ± 5 dc bound on the full vehicle offset is included to acknowledge that the anchor offset constant is itself uncertain. At the F25 baseline cruise condition ($M = 0.78$) shown in Fig. 5, the 95% envelope of the propagated C_D (second-order probability p-box approach) has a total width of roughly 72 drag counts (± 36 dc about the nominal of 299 dc, or about

$\pm 12\%$). This magnitude is large in absolute terms but consistent with the conceptual design stage of an advanced configuration. Four factors contribute to the overall aerodynamic uncertainty. First, the analysis uses low-order physics tools rather than RANS-in-the-loop within the optimizer. Second, the calibration is built from a limited CFD database fitted by a reduced-order surrogate. Third, the F25 configuration has a high aspect ratio and span that lie outside the certified-aircraft database. A small design modification to an aircraft already certified, such as a fuselage stretch, would not carry uncertainty of this magnitude. The F25 sits well outside anything operating today, and rigorously propagating heterogeneous conceptual design stage uncertainties is one of the goals of the challenge problem. Fourth, in generating the CFD database, the local airfoil shape was not re-optimized when planform parameters were changed. This leads to higher than expected uncertainty. The band is intentionally retained at this magnitude to expose the design risk of an advanced configuration; reducing it would mask the very uncertainty the challenge problem aims to surface. The presence of significant uncertainty has been cited as reasoning for not performing UQ. An objective of this paper is to challenge that traditional thought process and encourage UQ, even in the presence of significant uncertainty.

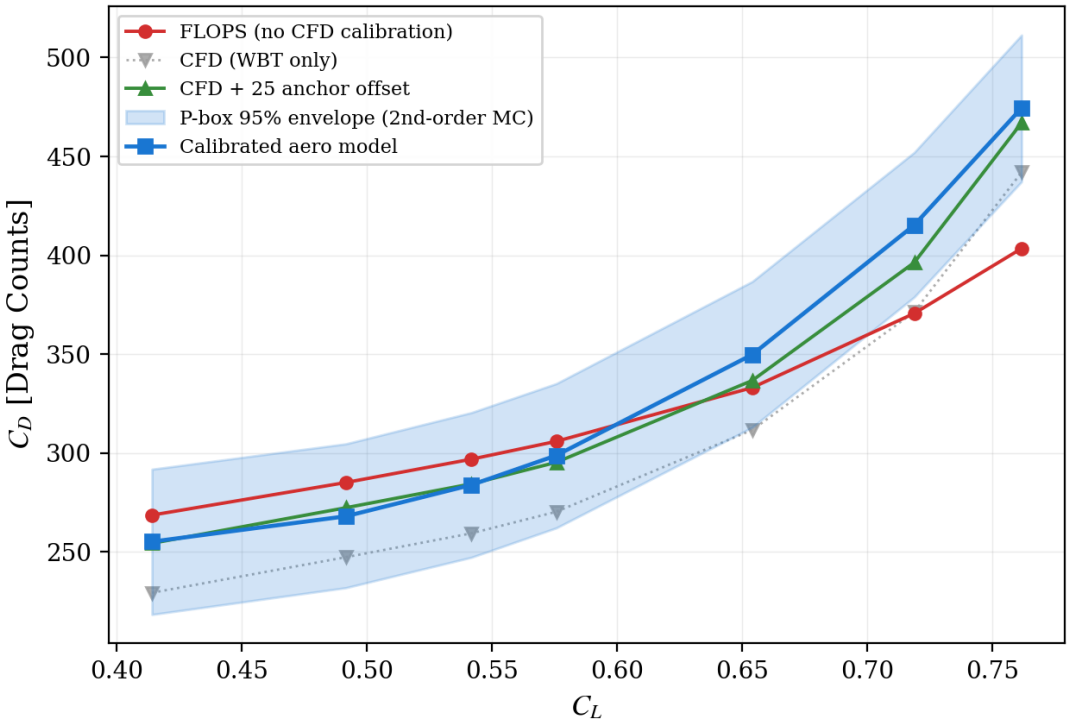


Figure 5 – F25 baseline drag polar at cruise ($M = 0.78$) with the propagated p-box 95% envelope on C_D.

4.2 Wing Mass and Structures

Wing mass is calculated from a recalibrated form of the Raymer transport aircraft equation [42], with a coupled folding-wingtip mass penalty for spans exceeding the 36 m gate-limit constraint and a one-sided thin-wing interaction term that captures structural penalties for designs more aggressive than the current state-of-the-art. The base wing mass is given by Eq. 6.

$$M_{\text{base}} = K_{\text{cal}}\, C\, (M_{\text{dg}} N_z)^{0.557}\, S_w^{0.649}\, AR^{0.5}\, (t/c)^{-0.4}\, (1+\lambda)^{0.1}\, \cos(\Lambda)^{-1.0}\, S_{\text{csw}}^{0.1}\, I^{p_\Gamma} \tag{6}$$

Here M_{dg} is the design gross mass, N_z is the ultimate load factor, S_w is the wing reference area, AR is the aspect ratio, t/c is the average thickness-to-chord ratio, λ is the wing taper ratio, Λ is the quarter-chord sweep angle, and S_{csw} is the wing-mounted control surface area. K_{cal} and C are calibration constants discussed below, I is the thin-wing interaction term defined in Eq. 7, and p_Γ is the design-coupled exponent on the interaction term (Eq. 14), taking the calibrated value 0.50 deterministically

and an AR-dependent epistemic interval under propagation.

The thin-wing interaction term, I, is the ratio of the geometric thin wing index, $AR/((t/c)\cos(\Lambda))$, to the same quantity evaluated at a current state-of-the-art reference geometry ($AR_{\text{ref}} = 11.0$, $(t/c)_{\text{ref}} = 0.14$, $\Lambda_{\text{ref}} = 25°$):

$$I = \frac{AR\,/\,((t/c)\cos(\Lambda))}{AR_{\text{ref}}\,/\,((t/c)_{\text{ref}}\cos(\Lambda_{\text{ref}}))} \tag{7}$$

The interaction is one-sided; it activates only in the thin-wing regime where $I > 1$ (designs more aggressive than the current state-of-the-art reference) and is held at unity below via an activation function. The high-AR/thin-wing/high-sweep extrapolation beyond Raymer's transport aircraft database is captured exclusively by the interaction term I^{p_Γ}, leaving the underlying regression unmodified.

The calibration factor, $K_{\text{cal}} = 0.70$, represents the 30% wing mass technology credit attributed by Wöhler et al. [12] to the combination of CFRP construction, load alleviation, active flutter suppression, and adaptive flaps. This full credit is realized only at the current state-of-the-art reference geometry where the thin-wing interaction term is inactive ($I = 1$). For designs more aggressive than the reference, the interaction term lifts wing mass back up, partially eroding the credit. The leading coefficient, $C = 5.45 \times 10^{-3}$, and the interaction exponent, $p_\Gamma = 0.50$ (its deterministic value), are solved from two anchors. The first anchor, in the inactive regime, enforces a wing mass equal to 70% of the corresponding FLOPS empirical weights prediction [43] (the technology-credit asymptote). The second anchor, at the F25 baseline geometry (where the interaction is active), recovers the published total wing mass of 7,866 kg, giving an effective ratio of roughly 0.86 relative to FLOPS at that point.

Figure 6 shows the resulting wing mass response across aspect ratio with Maximum Takeoff Weight (MTOW) closed iteratively at each AR via a fixed-point operating empty mass loop. The band shown is the 95% Conservative Symmetric Probability Interval (CSPI) [44] of a propagated probability box (p-box), generated by a second-order Monte Carlo that samples the epistemic parameters in the outer loop and the aleatory parameters in the inner loop, with the MTOW closure re-converged for every joint sample. The iso-percentage reference curves indicate wing mass as a percentage of the closed MTOW at each AR.

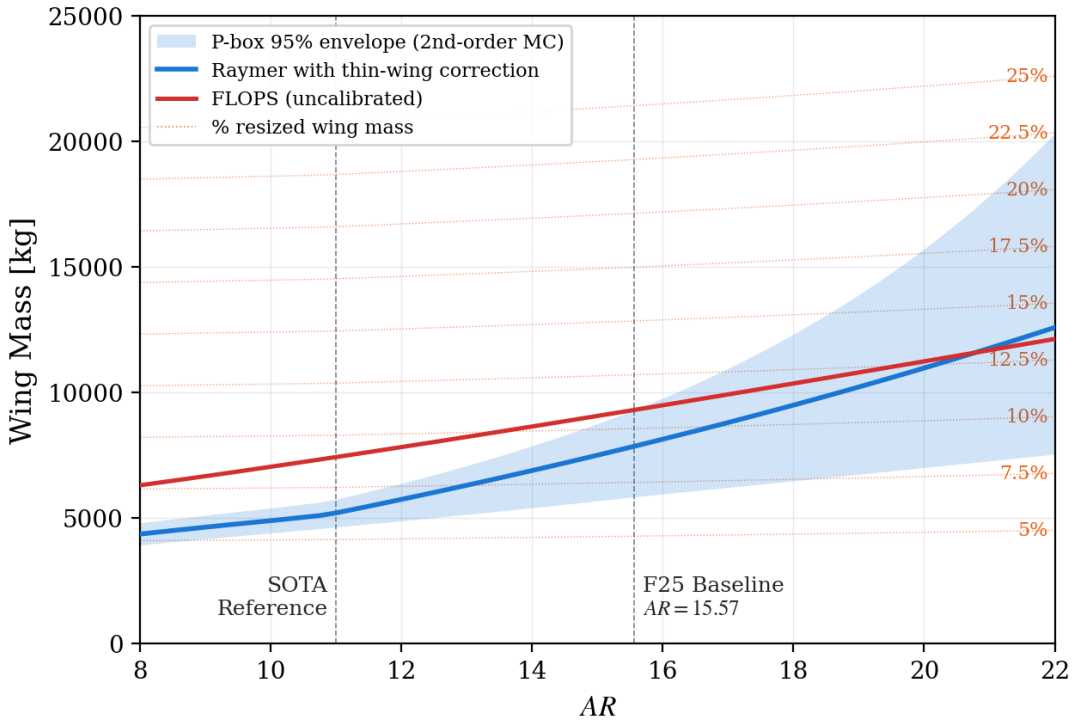


Figure 6 – Wing mass vs aspect ratio: recalibrated Raymer (blue) and FLOPS native (red), with the propagated p-box 95% envelope on the Raymer response.

The folding wingtip introduces a coupled mass penalty; the hinge mass, M_{fold}, scales with the *total* wing mass, M_{wing}, which itself includes the fold mass. This formulation is given in Eq. 8, where M_{base}

is the mass of the wing without the hinge mass and the fold mass is a fraction g of the total wing mass, $M_{\text{fold}} = g\,M_{\text{wing}}$, with g developed below.

$$M_{\text{wing}} = M_{\text{base}} + M_{\text{fold}} \tag{8}$$

As wing span is a design variable in the challenge problem, the folding wing mechanism must be resized for different wing span values. A longer span outboard from the wing hinge requires a larger hinge mechanism. A literature survey did not uncover any sizing relationships for commercial transonic high aspect ratio folding wings. Scholz [45] investigated the folding hinges of carrier-based fighter aircraft, whose highly loaded hinges and short bending-moment arms differ fundamentally from a high aspect ratio transport wingtip. In commercial transonic transports, Gu et al. [46] treats the hinge as a constant lump mass ($\sim$200 kg per hinge) and explicitly does not model how it scales. In this problem, therefore, a simple sizing model is adopted where the hinge mass is modeled as the sum of two contributions, given in Eq. 9. The first contribution is derived from a *fixed* mechanism mass taken as independent of outboard span. The second contribution is derived from a *variable* structural mass which scales with the wing bending moment of the portion of the span outboard of the hinge. To complete the sizing relationship, the contributing percentage of total hinge mass for both the fixed and variable components must be determined. That is, how much of the total hinge mass is variable based on wing span and how much of the total hinge mass is invariant to wing span changes. In the absence of any ratio supported by literature, an even split of overall mass at the baseline F25 design is assumed, $\phi = 0.5$. The split fraction ϕ is defined as the variable share of the baseline hinge mass, Eq. 10, with the fixed mechanism taking the complementary share $(1-\phi)$.

$$M_{\text{fold}} = M_{\text{fold}_{fixed}} + M_{\text{fold}_{bending}} \tag{9}$$

$$M_{\text{fold}_{bending}}(b_{\text{ref}}) = \phi\, M_{\text{fold}}(b_{\text{ref}}), \qquad M_{\text{fold}_{fixed}} = (1-\phi)\, M_{\text{fold}}(b_{\text{ref}}) \tag{10}$$

A higher ϕ therefore increases the variable contribution relative to the fixed one. Their ratio is $M_{\text{fold}_{bending}}/M_{\text{fold}_{fixed}} = \phi/(1-\phi)$, the form used in the closed expression below.
To size the variable bending contribution as a function of span, an elliptical spanwise lift distribution is integrated to find the bending moment that the outboard panel exerts about the hinge station. This moment is then taken relative to its baseline-span value. Doing so removes the load magnitude and leaves the purely geometric span ratio $\bar{M}(b)$ of Eq. 11. The geometric factor $G(\eta_h)$ and the full integration are given in Appendix A.

$$\bar{M}(b) = \frac{M(b)}{M(b_{\text{ref}})} = \frac{b\,G(\eta_h)}{b_{\text{ref}}\,G(\eta_{h,\text{ref}})}, \qquad \eta_h = \frac{b_{\text{gate}}}{b}, \quad \eta_{h,\text{ref}} = \frac{b_{\text{gate}}}{b_{\text{ref}}} \tag{11}$$

Throughout, $b_{\text{gate}} = 36$ m is the ICAO Aerodrome Code C wingspan limit and $b_{\text{ref}} = 45$ m is the F25 baseline span [12].
By construction $\bar{M}(b_{\text{ref}}) = 1$, and $\bar{M} \to 0$ as $b \to b_{\text{gate}}$, where the outboard panel and its moment vanish. The variable hinge mass scales directly with this ratio, Eq. 12, so the bending contribution equals its baseline value at $b = b_{\text{ref}}$ and grows or shrinks in proportion to the outboard bending moment away from it. The fixed mechanism mass is independent of span by construction.

$$M_{\text{fold}_{bending}}(b) = M_{\text{fold}_{bending}}(b_{\text{ref}})\,\bar{M}(b) \tag{12}$$

Below the gate-limit span, no wingtip fold is fitted and $M_{\text{fold}} = 0$; immediately above b_{gate}, the variable term vanishes ($\bar{M} \to 0$) and only the fixed mechanism mass remains, producing a discontinuous step that reflects the discrete presence of the hinge mechanism. Because this surviving contribution is independent of ϕ, the epistemic dimension of the uncertainty band closes to a point at the gate limit; only the aleatory scatter on C remains. Carrying the moment in closed form keeps the span dependence exact, so ϕ is the only quantity in the model that is not fixed by geometry or physics.
Because M_{fold} depends on M_{wing}, which contains it, Eqs. 8 and 9 are solved together. Combining the split (Eq. 10) with the span scaling (Eq. 12) expresses the hinge mass as a fraction $g(b)$ of total wing mass; substituting $M_{\text{fold}} = g(b)\,M_{\text{wing}}$ into Eq. 8 gives the closed forms of Eq. 13.

$$M_{\text{wing}} = \frac{M_{\text{base}}}{1-g(b)}, \qquad M_{\text{fold}} = \frac{g(b)}{1-g(b)} M_{\text{base}}, \qquad g(b) = c_{\text{fix}} \left[1 + \frac{\phi}{1-\phi} \bar{M}(b) \right]. \tag{13}$$

Here c_{fix} is calibrated so that at the baseline span with $\phi = 0.5$ and $\bar{M} = 1$, the total fold mass fraction recovers the 5% baseline anchor (393 kg) consistent with the Gu et al. [46] magnitude estimate: $g = c_{\text{fix}}[1+1] = 2\,c_{\text{fix}} = 0.05$, hence $c_{\text{fix}} = 0.025$. It is thus exactly half the baseline fold-mass fraction by construction and is the value $g(b)$ takes for spans immediately above the gate limit, where $\bar{M} = 0$ leaves only the fixed term.

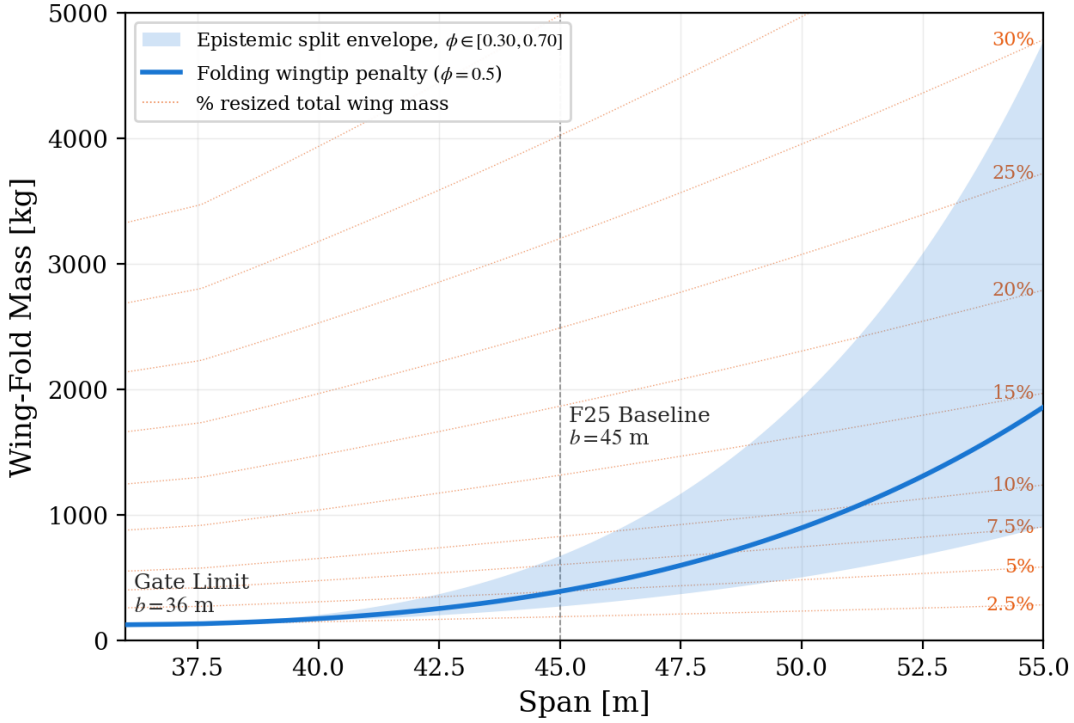


Figure 7 – Folding-wingtip mass penalty vs span, with the epistemic split-fraction envelope ($\phi \in [0.30, 0.70]$) on the fold mass.

Figure 7 shows the fold mass versus span with all other parameters fixed. At the 45 m baseline span and the nominal equal split ($\phi = 0.5$), the fold mass is 393 kg, 5% of total wing mass, rising to roughly 14% at $b = 55$ m. Unlike Fig. 6, which propagates the full p-box, Fig. 7 isolates the split-fraction contribution by holding all other parameters at nominal, so the band shown is purely epistemic in ϕ. The shaded band is the epistemic envelope over the split fraction $\phi \in [0.30, 0.70]$. Because the fixed mechanism mass is pinned and ϕ scales the variable contribution, the split impacts the fold mass at every span where an outboard panel exists. The band is open at the baseline and widens toward higher span as $\bar{M}$ grows. It closes to a point only at the gate limit, where $\bar{M} = 0$ leaves the ϕ-independent fixed mechanism mass and, below the gate, M_{fold} drops to zero in the step noted above. The envelope is asymmetric and wider above the nominal than below because the variable-to-fixed ratio $\phi/(1-\phi)$ is convex in ϕ and the fixed piece clamps the lower edge. The asymmetry is a property of this ϕ-only sweep and would be partly masked under full p-box propagation. The dotted iso-percentage curves give the fold mass as a percentage of the resized total wing mass, which itself grows with $AR = b^2/S$ at fixed area.

Four of the ten uncertain parameters act on the structures and weights discipline: two epistemic and two aleatory. The first epistemic parameter is the fold-mass split fraction, $\phi \in [0.30, 0.70]$, with nominal value $\phi = 0.5$. Because no published data resolve the split for a commercial transport wingtip, the interval is deliberately wide. $\phi = 0.30$ represents a mechanism-dominated joint and $\phi = 0.70$ a structure-dominated one, bracketing the equal-split best estimate. Its influence vanishes only at the gate, where $\bar{M} = 0$. The second epistemic parameter is the interaction exponent, p_Γ, of Eq. 6,

promoted from its deterministic value of 0.50 to an AR-dependent epistemic interval for propagation: its lower bound is fixed at zero (no penalty) and its upper bound widens linearly with AR beyond the reference geometry, formalizing the certification-gap framework (U_{cert}, §3.2.4) of Eq. 14.

$$p_\Gamma \in [0,\ 0.5 + 0.04\,(AR - 11)], \quad AR > 11. \tag{14}$$

At $AR = 11$ the interval is $[0, 0.5]$, with the calibrated value $p_\Gamma = 0.50$ at the upper edge. At $AR = 18$, the upper end of the F25 design space, it widens to $[0, 0.78]$ and the nominal sits well inside. The extension down to zero admits the possibility that advanced load alleviation or structural tailoring removes the penalty entirely. The first aleatory parameter is the leading Raymer coefficient, $C \sim \mathrm{N}(C_{\text{nom}}, 0.05\,C_{\text{nom}})$, with $C_{\text{nom}} = 5.45 \times 10^{-3}$ and a 5% coefficient of variation. Wing mass scales linearly with C, so it sets an aleatory floor on the wing mass uncertainty band across the design space. The second aleatory parameter, $\sigma_{\text{OEM}} \sim \mathrm{N}(1.0, 0.015^2)$, is a multiplicative scaler on the non-wing Operating Empty Mass (OEM) contribution in the mass-balance closure. Its 1.5% coefficient of variation gives a $\pm 2\sigma$ band of about $\pm 3\%$, consistent with typical conceptual design mass scatter. It is classified as direct assignment (§3.2.3) because the supplied toolchain does not propagate it from a component-level mass model.

4.3 Propulsion

The propulsion model is supplied as a tabulated engine deck that returns thrust and Specific Fuel Concumption (SFC) as functions of Mach number, altitude, and throttle setting, with a reference engine sized for the F25 baseline. Two discipline-specific design variables couple to the deck. The engine scale factor, $SF \in [0.7, 1.3]$, scales the reference deck's maximum thrust and engine mass to size the engine to the candidate aircraft; deviations from $SF = 1$ correspond to extrapolation away from the baseline engine. The engine performance technology factor, $\eta_{\text{tech}} \in [0, 1]$, interpolates between current technology performance ($\eta_{\text{tech}} = 0$, representative of a currently certified engine) and the advanced technology performance assumed for the F25 at 2035 EIS ($\eta_{\text{tech}} = 1$). The two design variables are independent. The propulsion discipline instantiates the technology-based framework U_{tech} of §3.2.4, coupling both engine uncertainties to η_{tech}; the SFC and engine-mass bands are evaluated at the baseline scale factor $SF = 1$.

The deterministic SFC scaler interpolates linearly between 1.000 at advanced technology ($\eta_{\text{tech}} = 1$) and 1.036 at current technology ($\eta_{\text{tech}} = 0$). The current-technology endpoint ($\eta_{\text{tech}} = 0$) is anchored to the published reference engine of the DLR D239 concept, from which the SFC and mass values are taken [12]. An asymmetric one-sided 95% prediction interval on u_{SFC} permits only a fixed, -1% scatter on the favorable (lower-SFC) side, while the adverse (higher-SFC) edge is held at the current technology worst case. A bad outcome, therefore, never underperforms proven current technology, and because the nominal improves toward advanced technology, the band widens on the adverse side, reflecting the TRL risk that the advanced technology underdelivers.

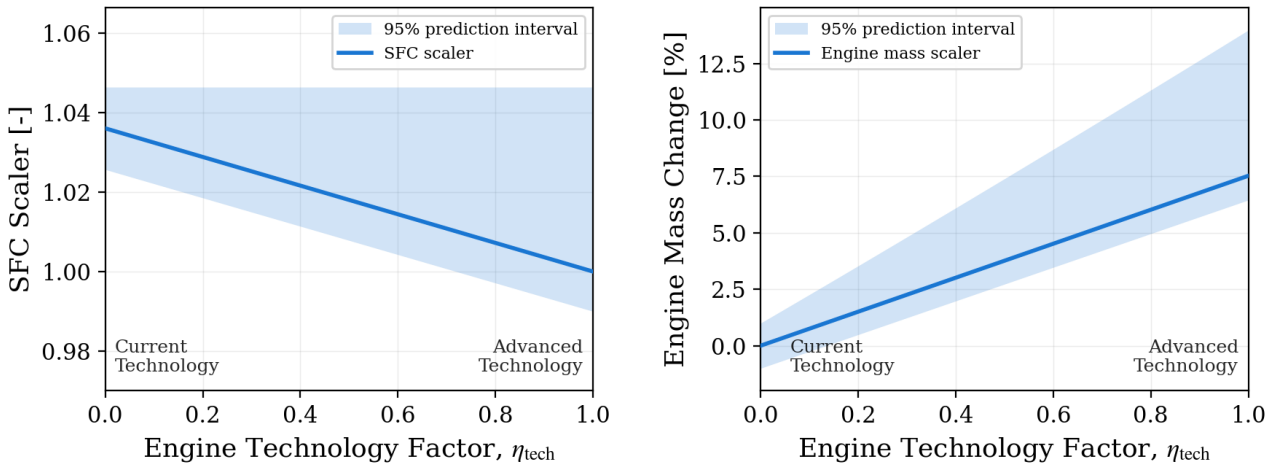
In the present formulation, u_{SFC} enters only as a fuel burn perturbation. The same asymmetric η_{tech}-coupled band can also serve as a proxy for broader operational risk in a system-level model where direct operating cost (DOC) is the objective. Advanced engine architectures that fall short of their entry-into-service projections typically do so not only through missed SFC targets, but also through reduced on-wing time, accelerated module deterioration, higher overall maintenance costs, and lost revenue from unplanned groundings. All of these effects increase DOC.

The engine mass uncertainty mirrors the SFC band on the opposite side. The nominal engine mass rises toward advanced technology, and an asymmetric one-sided 95% prediction interval on u_{eng} permits only a fixed, -1% scatter on the favorable (lighter) side, while the adverse (heavier) edge grows from $+1\%$ at current technology to $+6\%$ at advanced. This reflects the risk that the advanced technology engine comes out heavier than predicted. Engine mass itself follows the FLOPS power law

$$m_{\text{eng}} = m_{\text{eng,ref}} \cdot SF^{1.15}. \tag{15}$$

Figure 8 shows the two η_{tech}-coupled bands.

The two propulsion uncertain parameters (u_{eng}, and u_{SFC}) are each epistemic, bounded-interval parameters on $[-1, +1]$. u_{eng} enters as a perturbation on the engine mass scaler and propagates into OEM and the mass-balance closure. u_{SFC} scales the cruise SFC at every node.



(a) SFC scaler vs η_{tech}.

(b) Engine mass change vs η_{tech}.

Figure 8 – Propulsion uncertainty model: (a) SFC scaler and (b) engine mass change versus engine technology factor η_{tech}, each with its asymmetric 95% band.

4.4 Takeoff and Landing

Three certification-relevant takeoff and landing constraints are enforced: TOFL at sea level, International Standard Atmosphere (ISA), and MTOW; the second-segment one-engine-inoperative climb gradient, γ_{OEI}, defined as the steady climb angle in the critical configuration with one engine failed; and the approach reference speed at maximum landing mass, V_{OBS}, defined as $1.3V_{\text{stall}}$ at Maximum Landing Weight (MLW). The supplied Aviary package includes analytical takeoff and landing performance routines (standard conceptual design formulations following Raymer [42]) that compute these three quantities on every evaluation. The constraints are therefore propagated per uncertainty realization through the supplied physics. The driving aerodynamic inputs ($C_{L_{\text{max,TO}}}$, $C_{L_{\text{max,LDG}}}$, and $C_{D_0,\text{TO}}$) are not derived from an upstream high-fidelity aerodynamics analysis in the supplied toolchain; their uncertainty is instead set directly at the intermediate quantity level following the direct-assignment framework of §3.2.3.

Two primary parameters carry the uncertainty. The shared aleatory parameter $t_{HL} \sim \text{Tri}(-1,0,+1)$ scales both maximum-lift coefficients via the linear maps

$$C_{L_{\text{max,TO}}} = 2.30 + 0.20\,t_{HL}, \tag{16}$$

$$C_{L_{\text{max,LDG}}} = 2.85 + 0.25\,t_{HL}, \tag{17}$$

so that a single random realization affects both takeoff and landing constraints coherently. Figure 9 shows the propagation. The triangular density of t_{HL} on the x-axis (lower band) maps through the linear function (blue line) onto matching triangular marginals on $C_{L_{\text{max,TO}}}$ (left axis) and $C_{L_{\text{max,LDG}}}$ (right axis), with modes at the nominals 2.30 and 2.85 and support endpoints at 2.10/2.60 and 2.50/3.10. The takeoff zero-lift drag coefficient is treated directly as an independent aleatory parameter, $C_{D_0,\text{TO}} \sim \text{N}(0.058, 0.005^2)$.

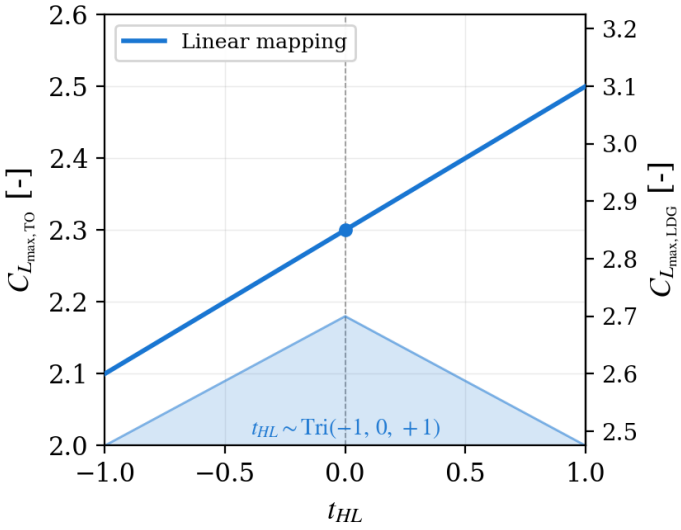


Figure 9 – $C_{L_{\mathrm{max,TO}}}$ (left axis) and $C_{L_{\mathrm{max,LDG}}}$ (right axis) plotted against t_{HL}, with the standard triangular density of t_{HL} shown as the shaded band along the x-axis.

5. Reference Model Validation

The supplied Aviary model is calibrated against the F25 reference values published in the baseline reference [12], ensuring that participants begin from a model that reproduces the documented baseline aircraft. Two sets of anchors are used: the mass breakdown and the design mission fuel burn.

5.1 Mass Breakdown

Most mass items are set directly in the F25 Aviary input file to the values published (fuselage, empennage, gear, pylons, individual systems and equipment line items, furnishings, and operating items). Two mass categories are model computed. The wing mass is produced by the calibrated Raymer model (§4.2). The engine mass is anchored to the deck reference of 3,538 kg per engine and scaled deterministically by η_{tech} (mass factor 0.93 at $\eta_{\mathrm{tech}} = 0$, 1.00 at $\eta_{\mathrm{tech}} = 1$). FLOPS' empirical mass equations supply the items aggregated into the reference totals but not itemized: the engine auxiliaries within the power units aggregate (oil, trapped fuel, mounts) and the crew mass within the operating items aggregate. Table 2 compares the resulting model breakdown to the published reference.

The model reproduces every line item of the reference to within 0.5%. The wing $+5$ kg residual is the Raymer two-anchor calibration accuracy. The ± 130 kg shift between Systems and Furnishings is a deliberate categorization workaround. The reference Systems line includes a 130 kg Fire Protection item that has no corresponding Aviary input variable, so it is bundled into Furnishings in the F25 input (which therefore is 3,730 kg vs the reference 3,600 kg). The Power Units aggregate matches exactly. FLOPS-computed engine auxiliaries (oil, trapped fuel, mounts; 515 kg) fill the gap between the engine deck reference (7,076 kg for two engines) and the published 7,591 kg. The Operating Items aggregate matches within 1 kg. FLOPS-computed crew mass (1,042 kg for 239 passengers) closes the residual after the explicit passenger service and cargo container overrides.

5.2 Mission Fuel Burn

The 2,500 NM design mission is simulated using the F25 reference trajectory (which itself derives from the reference design mission). The mission profile follows the reference trajectory: takeoff at sea level, climb to a top-of-climb at FL330, step-climb to FL370 at mid-cruise, and descent to landing, with $M = 0.78$ in cruise. Flown at the reference MTOW of 85,688 kg, the model produces a block fuel burn of 12,192 kg, $+53$ kg $(+0.4\%)$ above the reference of 12,139 kg, and an en route fuel burn $+0.5\%$ above the reference. The reference cruise quantities ($C_L \approx 0.59$, $L/D \approx 19.5$) are reproduced by the combination of the calibrated aerodynamics polynomial, the engine deck, and the matched mass breakdown. Figure 10 overlays the simulated trajectory on the reference for the 2,500 NM mission.

Table 2 – F25 baseline mass breakdown. Model values are from a baseline check at the F25 baseline geometry; the Reference column is from the published baseline. Block fuel is referenced against the design mission trajectory of Fig. 10.

Component	Model [kg]	Reference [kg]	Δ	Source
Wing	7,871	7,866	+5 kg	Calibrated Wing Mass Model
Fuselage	11,261	11,261	0	Direct Input Override
Horizontal Tail Plane	480	480	0	Direct Input Override
Vertical Tail Plane	498	498	0	Direct Input Override
Pylons	920	920	0	Direct Input Override
Power Units	7,591	7,591	0	Engine Deck + FLOPS Auxiliaries
Main Gear	1,842	1,842	0	Direct Input Override
Nose Gear	370	370	0	Direct Input Override
Systems	4,780	4,910	−130 kg	Direct Input Overrides (Fire Protection → Furnishings)
Furnishings	3,730	3,600	+130 kg	Direct Input Override (Includes 130 kg Fire Protection)
Manufacturer's Empty Mass (MEM)	39,343	39,337	+6 kg	Sum
Operating Items	6,933	6,934	−1 kg	Overrides + FLOPS Crew
Operating Empty Mass (OEM)	46,276	46,271	+5 kg	MEM + Operating Items
Maximum Payload (239 PAX + Misc Cargo)	25,000	25,000	0	Direct Input Override
Block Fuel (2,500 NM Design)	12,192	12,139	+53 kg	Mission Simulation (§5.2)

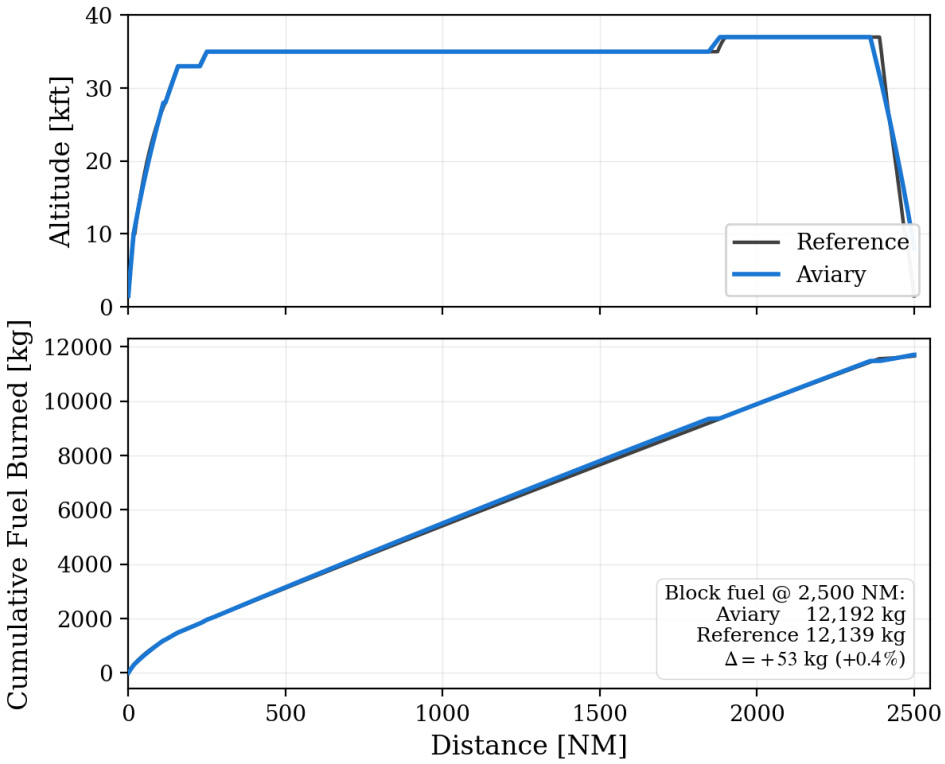


Figure 10 – 2,500 NM design mission validation: Aviary simulation vs reference.

6. Challenge Problem Specification

To encourage broad participation, the challenge problem is divided into multiple tiers consisting of increasingly complex formulations. This section consolidates the challenge problem into a self-

Table 3 – Summary of uncertain parameters in the DLR-F25 challenge problem.

#	Symbol	Discipline	Type	Distribution / Interval	Section
1	z_{aero}	Aero (In-Flight)	Aleatory	$\mathrm{N}(0, 1)$	§4.1
2	u_{wbt}	Aero (In-Flight)	Epistemic	$[-5, +5]$ dc	§4.1
3	t_{HL}	T/O & Landing Aero	Aleatory	$\mathrm{Tri}(-1, 0, +1)$	§4.4
4	$C_{D_0,\text{TO}}$	T/O & Landing Aero	Aleatory	$\mathrm{N}(0.058, 0.005^2)$	§4.4
5	u_{eng}	Propulsion	Epistemic	$[-1, +1]$	§4.3
6	u_{SFC}	Propulsion	Epistemic	$[-1, +1]$	§4.3
7	σ_{OEM}	Structures	Aleatory	$\mathrm{N}(1.0, 0.015^2)$	§4.2
8	p_Γ	Structures	Epistemic	$f(AR)$ interval, see Eq. 14	§4.2
9	ϕ	Structures	Epistemic	$[0.30, 0.70]$	§4.2
10	C	Structures	Aleatory	$\mathrm{N}(C_{\text{nom}}, 0.05\,C_{\text{nom}})$	§4.2

contained specification: the 10 uncertain parameters, the 10 design variables, the 8 constraints, the robust objective, and the 4 problem tiers. Constraints are evaluated and enforced from a probabilistic perspective. For example, the 95% probability of achieving an approach speed of less than or equal to 136 KCAS, given all discipline-specific uncertainties, is enforced as a constraint. After optimization, evaluating constraint sensitivities to uncertain parameters can identify which uncertainties most significantly affect active constraints, helping guide future investments to reduce risk.

6.1 Uncertain Parameters

The 10 uncertain parameters prescribed by the challenge problem (5 aleatory, 5 epistemic) are given in Table 3. Each parameter was introduced in the discipline section listed in the right-most column. Aleatory parameters are sampled from a probability distribution; epistemic parameters are characterized by a bounded interval (no probability density is asserted). The naming follows the role a parameter plays. Some parameters are dimensionless noise variables that scale or perturb a separate physical quantity. These use a prefix convention. The prefix z_X denotes an aleatory standard-normal variable, t_X an aleatory triangular variable, and u_X an epistemic bounded-interval variable. For example, z_{aero} is a standard-normal multiplier on the physical drag scatter σ_{aero}, and u_{eng} is a unit-interval perturbation on the engine-mass band. Other parameters are themselves the physical quantity that carries the distribution, and these are written with their own symbol. Examples are the Raymer leading coefficient C, the interaction exponent p_Γ, the fold-mass split fraction ϕ, and the non-wing OEM scaler σ_{OEM}. The scaler σ_{OEM} is centered on 1.0 and is the physical multiplier itself, so it takes its own symbol rather than a z prefix.

All 10 uncertain parameters in Table 3 are assumed mutually independent. Derived intermediate quantities may share parameters by construction (e.g., t_{HL} drives both $C_{L_{\text{max,TO}}}$ and $C_{L_{\text{max,LDG}}}$ coherently via Eqs. 16–17); this is a deterministic coupling at the model level, not a correlation between Table 3 parameters.

6.2 Design Variables

The design variables for the challenge problem are given in Table 4, grouped into aircraft design variables (six geometric and propulsion variables, plus the design closure $MTOW$) and mission profile variables (the cruise altitude profile and the climb/descent Mach profiles). The first five aircraft design variables are carried over from the original F25 MDO benchmark [11]. The sixth, an engine performance technology factor, η_{tech}, was added in this work to expose the technology readiness gap between current production engines and the advanced engine technology assumed for the F25 as an explicit design axis (§4.3).

$MTOW$ is treated as a free variable. The optimizer must satisfy two conditions on it. The first is the mass-balance closure, OEM + payload + total fuel $=$ MTOW. The second is that the total fuel carried is enough to fly the full design mission, including climb, cruise, descent, and reserves. Together these conditions fix MTOW for any given design, so it is determined by the mass balance rather than driven to its upper bound.

Table 4 – Design variables for the DLR-F25 challenge problem.

Variable	Symbol	Bounds
Aircraft Design		
Wing Area	S	74 – 186 m^2
Wing Span	b	30 – 55 m
Thickness-to-Chord	t/c	0.12 – 0.15
Wing Sweep (c/4)	Λ	$20^\circ - 35^\circ$
Engine Scale Factor	SF	0.7 – 1.3
Engine Performance Technology Factor	η_{tech}	0 – 1
Maximum Takeoff Mass	$MTOW$	Design closure
Mission Profile		
Cruise Altitude	h_{cr}	FL330 – FL370
Climb Mach Profile	M_{climb}	0.39 – 0.78
Descent Mach Profile	M_{descent}	0.36 – 0.78

The wing taper ratio is held fixed at $\lambda = 0.12$, the F25 baseline value [12], throughout. The engine scale factor, SF, and the engine performance technology factor, η_{tech}, are independent design variables. The optimizer may trade engine size against assumed engine technology level.
The trajectory endpoints are prescribed and are not design variables. The climb begins at the takeoff initial altitude of 1,500 ft at $M = 0.39$ (corresponding to a 250 KCAS climb schedule); the descent terminates at 500 ft at $M = 0.36$ (corresponding to a 250 KCAS approach schedule); and cruise is flown at a fixed $M = 0.78$. The cruise altitude is optimized as a cruise-climb profile between FL330 (lower bound and 30-minute climb-time constraint) and FL370 (upper bound). The Mach profiles during climb and descent are optimized between the fixed endpoints.

6.3 Constraints

The challenge problem constraints are given in Table 5. Probabilistic constraints are enforced at $\mathbb{P}[\text{satisfied}] \geq 95\%$ during optimization in Tiers 2 and 3, in Tier 1 they are evaluated at the fixed F25 baseline and $\mathbb{P}[\text{satisfied}]$ is reported as a deliverable, and in Tier 0 they reduce to deterministic inequalities at the nominal design point. All probabilistic constraints are propagated per realization through the supplied Aviary analytical performance routines. The driving aerodynamic inputs ($C_{L_{\text{max,TO}}}$, $C_{L_{\text{max,LDG}}}$, and $C_{D_{0,\text{TO}}}$) enter via the direct-assignment framework of §3.2.3. In the driving parameters column, "Structures" refers collectively to the four structures and weights uncertain parameters of § 4.2: ϕ, p_Γ, C, and σ_{OEM}. All probabilistic constraints are enforced on a one-sided interval, $\alpha = 95\%$, on every mission per uncertainty realization. The single deterministic constraint (rear spar height) is purely geometric; it depends only on the wing design variables and is not driven by any of the prescribed uncertain parameters.

Table 5 – Constraints for the DLR-F25 challenge problem.

Constraint	Limit	Type	Driving Uncertain Parameters
TOFL @ SL, ISA, MTOW	≤ 2200 m	Probabilistic	t_{HL}, $C_{D_{0,\text{TO}}}$, Structures
γ_{OEI} (2nd-Segment)	$\geq 2.4\%$	Probabilistic	t_{HL}, $C_{D_{0,\text{TO}}}$, Structures
V_{OBS} @ MLW (Approach)	≤ 136 KCAS	Probabilistic	t_{HL}, Structures (via MLW)
Climb Time to Cruise Altitude	≤ 30 min	Probabilistic	z_{aero}, Structures (via MTOW)
Descent Time	≤ 40 min	Probabilistic	z_{aero}, Structures (via landing mass)
Required Climb Rate at Cruise	≥ 1.5 m/s	Probabilistic	z_{aero}, Structures (via MTOW)
Wing Fuel Volume	$V_{\text{avail}} \geq V_{\text{req}}$	Probabilistic	z_{aero}, u_{SFC}, Structures (via MTOW)
Rear Spar Height @ $\eta = 0.76$	≥ 0.15 m	Deterministic	—

6.4 Objectives and Robustness Metrics

The performance metric is specific air range (SAR), defined as the range flown in a mission divided by the fuel burned on that mission. Across the prescribed multi-mission fleet (§2.), the fleet-aggregated

SAR is the weighted sum

$$\mathrm{SAR}_{\mathrm{fleet}}(x,\xi) \;=\; \sum_{m=1}^{5} w_m \cdot \frac{\mathrm{range}_m}{\mathrm{fuel}_m(x,\xi)}, \tag{18}$$

where x is the design vector, ξ is the random vector of the 10 uncertain parameters, and w_m are the prescribed mission frequencies of Table 1.

Under uncertainty, the participants will maximizes the location metric $\tilde{\mu}[\mathrm{SAR}]$ of §3.1, read from the propagated p-box. Location alone is not enough; a design with high $\tilde{\mu}[\mathrm{SAR}]$ but a wide envelope across the uncertain parameters is fragile. The spread $\tilde{\sigma}[\mathrm{SAR}]$, the half-width of the central 95% interval of the SAR p-box (Eq. 2, Fig. 3), is reported alongside $\tilde{\mu}[\mathrm{SAR}]$ as a paired robustness metric. Both are read from the second-order propagation of the 10 uncertain parameters in Table 3 using the participant's chosen propagation method (§3.1), and convergence diagnostics for the chosen method must be reported. As guidance, the appropriate diagnostics depend on the chosen method. Direct sampling approaches should report sample size and confidence intervals on the interval edges L and U. Polynomial chaos should report expansion order and validation error, and surrogate-based methods their validation residuals and training sample budget. Any analogous quantities appropriate to the chosen approach are acceptable, provided they demonstrate convergence of the reported metrics to within a participant-specified tolerance.

As noted in §3.1, the metrics $\tilde{\mu}[\mathrm{SAR}]$, $\tilde{\sigma}[\mathrm{SAR}]$, and $\mathrm{CVaR}_{5\%}$ are read from a second-order propagation that sweeps the epistemic intervals in an outer loop and propagates the aleatory distributions in an inner loop, so the epistemic parameters are never assigned a density. This treatment keeps the aleatory and epistemic contributions separable all the way through the analysis to the reported metrics, consistent with the aleatory/epistemic sensitivity partition required in §7.. When a parameter's interval or distribution is design coupled, the propagation uses the interval or distribution evaluated at the current design point. Tier 2 and Tier 3 toolchains, therefore, re-propagate at each outer-loop iteration. Tier 1 propagates once at the F25 baseline values of the design-coupled distributions. The aerodynamic case is illustrative of the convention. At each design x, the heteroscedastic standard deviation, $\sigma_{\mathrm{aero}}(M,C_L,x)$, is a deterministic function of the current design and flight condition, and the scalar z_{aero} is fixed within each uncertainty realization and applied as $\Delta C_D = z_{\mathrm{aero}} \cdot \sigma_{\mathrm{aero}}(M,C_L,x)$.

The $(\tilde{\mu},\tilde{\sigma})$ pair is not exhaustive; two designs with identical location and spread can differ in tail behavior. The left panel of Fig. 11 shows three notional SAR distributions with a common $\tilde{\mu}[\mathrm{SAR}]$. Designs I and II differ only in $\tilde{\sigma}$, while Designs II and III have the same $\tilde{\sigma}$ but Design III has a heavier left tail. The inset confirms that $\tilde{\sigma}$ alone cannot distinguish Designs II and III, while $\mathrm{CVaR}_{5\%}$ does. The right panel of Fig. 11 illustrates the form of a representative submission: a set of non-dominated designs spanning the achievable trade-off between location and spread.

Participants may additionally report $\mathrm{CVaR}_{5\%}[\mathrm{SAR}]$ [31], defined as the expectation of SAR conditioned on the worst 5% of the propagated distribution and capturing performance when the uncertain parameters resolve unfavorably. For SAR maximization, higher $\mathrm{CVaR}_{5\%}$ is preferred. CVaR is reported as an optional metric. The prescribed Pareto front uses only the $(\tilde{\mu},\tilde{\sigma})$ pair because the central interval edges are robustly estimable by most propagation methods, while tail-risk measures such as CVaR generally require greater computational effort or specialized estimators to resolve the tail with comparable accuracy.

Submissions take the form of a Pareto front of designs in the $(\tilde{\mu}[\mathrm{SAR}],\tilde{\sigma}[\mathrm{SAR}])$ plane. This is the canonical mean-variance robust optimization framing surveyed by Beyer and Sendhoff [10] and applied to aircraft conceptual design (under the name Robust Design Simulation) by DeLaurentis and Mavris [30]. At least seven points spanning the achievable trade-off are required per submission and per tier (i.e., Tier 2 submissions report seven points; Tier 3 submissions report seven points for the fleet-aggregated objective), generated, for example, by sweeping a robustness-emphasis parameter (maximize $\tilde{\mu}[\mathrm{SAR}] - \lambda \cdot \tilde{\sigma}[\mathrm{SAR}]$ for several λ values).

6.5 Problem Tiers

The challenge problem is divided into the four tiers given in Table 6. Tier 0 is the deterministic reference. Tier 1 propagates the 10 uncertain parameters through the fixed F25 baseline design with no optimization. Tier 2 performs optimization under uncertainty (OUU) on the single design mission.

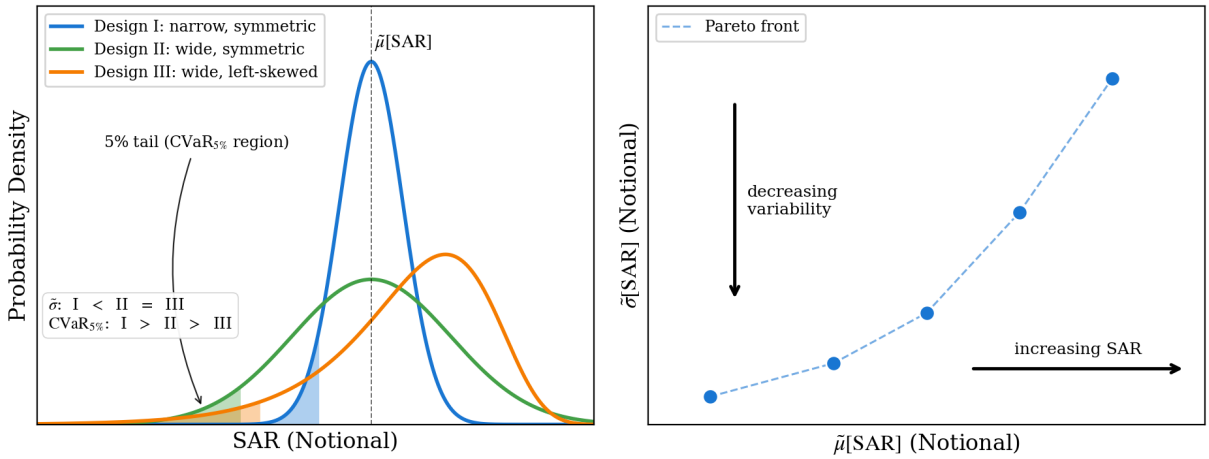


Figure 11 – Schematic illustration of the robust deliverable. Left: three notional SAR distributions with $\tilde{\mu}[\text{SAR}]$ aligned at the dashed line. Right: schematic Pareto front of designs in the $(\tilde{\mu}[\text{SAR}], \tilde{\sigma}[\text{SAR}])$ plane. Both panels are illustrative and do not represent F25 results.

Tier 3 extends OUU to the prescribed multi-mission fleet. All four tiers use the same uncertain parameters (Table 3) and the same constraint set (Table 5). Participants should complete all tiers that their toolchain can support.

Table 6 – Problem tiers for the DLR-F25 challenge problem.

Tier	Approach	Mission Scope	Objective	UQ?
0	Deterministic Optimization	Single, 2,500 NM	max SAR	No
1	UQ Propagation at Fixed F25 Baseline	Single, 2,500 NM	Report $\tilde{\mu}, \tilde{\sigma}, \mathbb{P}[g]$	Yes
2	Single-Mission OUU	Single, 2,500 NM	max $J[\text{SAR}_{\text{design}}]$	Yes
3	Multi-Mission OUU	Fleet, Weighted Mix	max $J[\text{SAR}_{\text{fleet}}]$	Yes

The Tier 2 and Tier 3 objective is the robust scalar $J(x;\lambda)$ of Eq. 2. Tier 2 and Tier 3 submissions additionally report at least seven Pareto-front points spanning the $(\tilde{\mu}, \tilde{\sigma})$ trade-off per the deliverable specification of §6.4. The location-maximizing endpoint is one of those points.

6.5.1 Computational scope by tier.

The tiers are arranged in order of increasing complexity. Tier 0 is a single deterministic optimization at the F25 design point, achievable in under five minutes on a workstation. Participants intending a Tier 3 submission may find it useful to first run Tier 0 over the full weighted mission mix — a deterministic multi-mission optimization with all uncertain parameters held at nominal. This exercises the fleet closure and per-mission profile re-optimization of the Tier 3 convention below without the cost of uncertainty propagation, providing a verified deterministic fleet design as a starting point for the Tier 3 OUU. Tier 1 fixes the design at the F25 baseline and performs a single second-order propagation through the 10 uncertain parameters to compute the deliverable metrics. Because there is no outer design-optimization loop, Tier 1 is the entry point for participants new to optimization under uncertainty, though it still requires the nested aleatory/epistemic propagation of §3.1. Tier 2 nests one UQ propagation inside an outer design optimizer at each iteration, and Tier 3 extends Tier 2 to five missions. The per-design propagation cost depends on the chosen propagation method. The total Tier 2 cost scales as the per-propagation cost times the optimizer iteration count, and the total Tier 3 cost is expected to scale as an order of magnitude higher than Tier 2. Participants are encouraged to engage at the highest tier their toolchain can sustain. A Tier 1 submission alone is a valuable contribution.

6.5.2 Tier 3 mission-closure convention.

A single physical aircraft flies all five Tier 3 missions. The six aircraft design variables in Table 4, including the engine scale factor, SF, and the engine performance technology factor, η_{tech}, are therefore shared across all five missions and do not change between them. The MTOW closure is performed only on the 2,500 NM design mission. That mission sets the design fuel load, drives the MTOW, and sizes the wing fuel volume through the design volume constraint. The four off-design missions are flown using the resulting closed aircraft with a reduced fuel load corresponding to their respective ranges, and the mission profile variables (cruise altitude and climb/descent Mach profiles) are re-optimized for each mission. For mission analysis simplicity, a fixed reserve fuel amount is assumed on every mission; no reserve mission is modeled in the optimization.

6.5.3 Optimizer choice.

The challenge problem is optimizer agnostic. While the supplied Aviary problem is formulated to run a gradient-based optimization, gradient-free, surrogate-assisted, and Bayesian-optimization approaches are all acceptable, as are reliability-index or chance-constrained reformulations of the probabilistic constraints. Participants are encouraged to report optimizer choice and any reformulations of the probabilistic constraints as part of the methodology summary (§7.).

7. Expected Deliverables

Participants are expected to deliver the items below, scaled to the tier of their submission. For Tier 1 submissions, the deliverables reduce to $\tilde{\mu}[\text{SAR}]$, $\tilde{\sigma}[\text{SAR}]$, and $\mathbb{P}[\text{satisfied}]$ for each probabilistic constraint, all evaluated at the fixed F25 baseline design; no design variables are reported and no Pareto front is generated. Tiers 2 and 3 require the full deliverable list:

- Optimized aircraft configurations for Tier 0 (deterministic) and Tiers 2–3 (OUU), reporting the six design variables, design mission MTOW, and fleet-weighted (Tier 3 only) specific air range location metric, $\tilde{\mu}[\text{SAR}]$.
- Probabilistic constraint compliance at the converged design, reported as $\mathbb{P}[\text{satisfied}]$ for each of the probabilistic constraints in Table 5.
- A robustness Pareto front trading $\tilde{\mu}[\text{SAR}]$ against a robustness metric. The percentile-based spread, $\tilde{\sigma}[\text{SAR}]$ (Eq. 2), is the prescribed metric; a tail-risk measure such as $\text{CVaR}_{5\%}$ may be reported additionally.
- Sensitivity rankings of the 10 uncertain parameters (e.g., variance-based Sobol indices [21, 22]) for both $\tilde{\mu}[\text{SAR}]$ and the active probabilistic constraint margins. Participants must partition these sensitivities into aleatory and epistemic contributions to demonstrate which active constraints are driven by irreducible physical scatter versus reducible technology assumptions.
- A methodology summary covering the UQ propagation method, sample size, optimizer settings, and any deviations from the prescribed parameter set.

8. Code and Data Availability

All challenge problem artifacts will be released as an open-source, self-contained Aviary repository concurrent with paper publication. The repository will include the F25 input package, the calibrated C_D surrogate and σ_{aero} function (exposed as queryable Python callables), the engine deck, the analytical takeoff and landing performance routines for TOFL, γ_{OEI}, and V_{OBS}, the five reference trajectories for the design and off-design fleet missions, and a standardized submission template (CSV for tabular outputs and JSON for the methodology summary and sensitivity indices). The repository includes a single Python script that reproduces a Tier 0 deterministic optimization in under five minutes on a workstation, providing a verified entry point from which participants can build their UQ propagation. Participants are encouraged to use the standardized submission template to facilitate cross-team comparison; alternative formats are accepted provided the same information content is delivered.

9. Contact Author Email Address

For further information, please contact the authors at benjamin.d.phillips@nasa.gov

10. Acknowledgement

The authors would like to recognize all previous supporters of the DLR-F25 efforts. The development of the DLR-F25 configuration was funded by the German Federal Ministry for Economic Affairs and Climate Action as part of the LuFo project VirEnfREI. NASA support was provided by the Transformational Tools and Technologies Project.

11. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

Appendix A Folding-Wingtip Bending-Moment Derivation

This appendix derives the geometric span ratio $\bar{M}(b)$ of Eq. 11 from an assumed elliptical loading. For a semi-span $s = b/2$ and a spanwise coordinate y measured from the root, the elliptical lift-per-unit-span distribution is given in Eq. 19, where ℓ_0 is the peak (root) loading.

$$\ell(y) = \ell_0 \sqrt{1 - \left(\frac{y}{s}\right)^2} \tag{19}$$

Requiring the distribution to carry the lift per semi-span L (one-half of the aircraft weight, held fixed as span varies) fixes ℓ_0, where the substitution $\eta = y/s$ reduces the integral to the area of a unit quarter circle, $\pi/4$.

$$L = \int_0^s \ell(y)\, dy = \ell_0 s \int_0^1 \sqrt{1-\eta^2}\, d\eta = \frac{\pi}{4}\ell_0 s \quad \Longrightarrow \quad \ell_0 = \frac{4L}{\pi s} \tag{20}$$

The hinge is fixed at the gate station $y_h = b_{\text{gate}}/2$, so the outboard structural mass is sized by the bending moment of the panel outboard of the hinge about that station, Eq. 21.

$$M(b) = \int_{y_h}^{s} (y - y_h)\, \ell(y)\, dy \tag{21}$$

Substituting Eq. 19, applying $\eta = y/s$ ($dy = s\, d\eta$), and writing the hinge station as $\eta_h = y_h/s = b_{\text{gate}}/b$ collects the load magnitude into the prefactor $\ell_0 s^2 = 2Lb/\pi$ (from Eq. 20) and leaves the geometric integral $G(\eta_h)$, Eq. 22.

$$M(b) = \ell_0 s^2 \int_{\eta_h}^{1} (\eta - \eta_h)\sqrt{1-\eta^2}\, d\eta = \frac{2Lb}{\pi} G(\eta_h) \tag{22}$$

Splitting the integrand into its two terms and evaluating each gives Eq. 23.

$$G(\eta_h) = \underbrace{\int_{\eta_h}^{1} \eta\sqrt{1-\eta^2}\, d\eta}_{\frac{1}{3}\left(1-\eta_h^2\right)^{3/2}} - \eta_h \underbrace{\int_{\eta_h}^{1} \sqrt{1-\eta^2}\, d\eta}_{\frac{1}{2}\left[\arccos\eta_h - \eta_h\sqrt{1-\eta_h^2}\right]} \tag{23}$$

which simplifies to the closed form of Eq. 24.

$$G(\eta_h) = \frac{1}{3}\left(1-\eta_h^2\right)^{3/2} - \frac{\eta_h}{2}\left[\arccos\eta_h - \eta_h\sqrt{1-\eta_h^2}\right] \tag{24}$$

This $G(\eta_h)$ is the geometric integral that enters the span ratio $\bar{M}(b)$ of Eq. 11.

References

[1] J. P. Slotnick, A. Khodadoust, J. Alonso, D. Darmofal, W. Gropp, E. Lurie, and D. J. Mavriplis. Cfd vision 2030 study: A path to revolutionary computational aerosciences. Technical Report NASA/CR-2014-218178, NASA, March 2014.

[2] T. A. Zang, M. J. Hemsch, M. W. Hilburger, S. P. Kenny, J. M. Luckring, P. Maghami, S. L. Padula, and W. J. Stroud. Needs and opportunities for uncertainty-based multidisciplinary design methods for aerospace vehicles. Technical Report NASA/TM-2002-211462, NASA Langley Research Center, July 2002.

[3] CFD Vision 2030 Integration Committee Roadmap Subcommittee. Cfd vision 2030 roadmap progress and technology highlights for 2024. Online report, 2024.

[4] A. W. Cary, J. Schaefer, E. P. Duque, M. Khurana, and E. C. DeCarlo. Overview of challenges in performing uncertainty quantification for fluids engineering problems. In *AIAA SciTech Forum*, January 2022. AIAA Paper 2022-2357.

[5] A. W. Cary, J. A. Schaefer, E. C. DeCarlo, E. P. Duque, and M. Khurana. Overview of fluid dynamics uncertainty quantification challenge problem and results. In *AIAA SciTech Forum*, January 2024. AIAA Paper 2024-0705.

[6] L. G. Crespo, S. P. Kenny, and D. P. Giesy. The NASA Langley multidisciplinary uncertainty quantification challenge. In *16th AIAA Non-Deterministic Approaches Conference, AIAA SciTech Forum*, January 2014. AIAA Paper 2014-1347.

[7] L. G. Crespo and S. P. Kenny. The NASA Langley challenge on optimization under uncertainty. *Mechanical Systems and Signal Processing*, 152:107405, 2021.

[8] W. Yao, X. Chen, W. Luo, M. van Tooren, and J. Guo. Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Progress in Aerospace Sciences*, 47(6):450–479, 2011.

[9] S. L. Padula, C. R. Gumbert, and W. Li. Aerospace applications of optimization under uncertainty. *Optimization and Engineering*, 7(3):317–328, 2006.

[10] H.-G. Beyer and B. Sendhoff. Robust optimization – a comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33–34):3190–3218, 2007.

[11] M. Abu-Zurayk, Č. Ilić, P. Wegener, J. Himisch, M. Schulze, S. Wöhler, J. Häßy, S. Dähne, S. Geisbauer, S. Görtz, Ö. Petersson, J. Brezillon, R. Llamas, C. Conlan-Smith, A. DeBlois, J. Singh, B. D. Phillips, E. Aretskin-Hariton, J. Chapman, C. Blondeau, and M. Méheut. Establishing a joint research-industry mdo benchmark based on the dlr-f25 aircraft configuration. In *AIAA SciTech Forum*, January 2026.

[12] S. Wöhler, J. Häßy, and V. Kriewall. Establishing the dlr-f25 as a research baseline aircraft for the short-medium range market in 2035. In *Proceedings of the 34th Congress of the International Council of the Aeronautical Sciences (ICAS)*, Florence, Italy, 2024. ICAS Paper 2024-0398.

[13] European Union Aviation Safety Agency (EASA). Final report: Review of standard passenger weights. Technical report, EASA, 2022.

[14] Bureau of Transportation Statistics. T-100 segment data. U.S. Department of Transportation, 2024.

[15] R. A. McDonald. Establishing mission requirements based on consideration of aircraft operations. *AIAA Journal of Aircraft*, 50(3), 2013.

[16] Y. Cai, D. Rajaram, and D. N. Mavris. Multi-mission multi-objective optimization in commercial aircraft conceptual design. In *AIAA Aviation Forum*, June 2019. AIAA Paper 2019-3577.

[17] Y. Cai, D. Rajaram, J. C. Gladin, and D. N. Mavris. Design and fleet assessment of commercial aircraft with range and speed optimization. *Journal of Aircraft*, 2026. Online ahead of print.

[18] J. E. Bernardo, T. Zaidi, M. LeVine, H. Jimenez, and D. N. Mavris. Rapid integrated interdependent fleet-level environmental model. *Journal of Aircraft*, 54(3):939–954, 2017.

[19] J. T. Hwang, J. P. Jasa, and J. R. R. A. Martins. High-fidelity design-allocation optimization of a commercial aircraft maximizing airline profit. *Journal of Aircraft*, 56(3):1164–1178, 2019.

[20] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM, Philadelphia, PA, 2014.

[21] I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414, 1993.

[22] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis: The Primer*. Wiley, 2008.

[23] W. L. Oberkampf, J. C. Helton, C. A. Joslyn, S. F. Wojtkiewicz, and S. Ferson. Challenge problems: Uncertainty in system response given uncertain parameters. *Reliability Engineering & System Safety*, 85(1–3):11–19, 2004.

[24] Laurence W. Cook, Jerome P. Jarrett, and Karen E. Willcox. Extending horsetail matching for optimization under probabilistic, interval, and mixed uncertainties. *AIAA Journal*, 2017.

[25] M. S. Eldred, L. P. Swiler, and G. Tang. Mixed aleatory-epistemic uncertainty quantification with stochastic expansions and optimization-based interval estimation. *Reliability Engineering & System Safety*, 96(9):1092–1113, 2011. Quantification of Margins and Uncertainties.

[26] J. C. Helton and F. J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69, 2003.

[27] D. Xiu and G. E. Karniadakis. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

[28] L. W.-T. Ng and K. E. Willcox. Monte carlo information-reuse approach to aircraft conceptual design optimization under uncertainty. *Journal of Aircraft*, 53(2):427–438, 2016.

[29] J. Gratz, J. Kirk, C. Recine, J. Jasa, E. Aretskin-Hariton, K. Moore, and K. Marfatia. Aviary: An open-source multidisciplinary design, analysis, and optimization tool for modeling aircraft with analytic gradients. In *AIAA Aviation Forum and ASCEND 2024*, July 2024. AIAA Paper 2024-4219.

[30] D. A. DeLaurentis and D. N. Mavris. Robust design simulation: A probabilistic approach to multidisciplinary design. *Journal of Aircraft*, 36(1):298–307, 1999.

[31] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

[32] D. Quagliarella and E. Iuliano. Robust design of a supersonic natural laminar flow wing-body. *IEEE Computational Intelligence Magazine*, 12(4):14–27, 2017.

[33] A. Chaudhuri, B. Kramer, M. Norton, J. O. Royset, and K. Willcox. Certifiable risk-based engineering design optimization. *AIAA Journal*, 60(2):551–565, 2022.

[34] R. M. Paiva, C. Crawford, and A. Suleman. Robust and reliability-based design optimization framework for wing design. *AIAA Journal*, 52(4):711–724, 2014.

[35] X. Du and W. Chen. Sequential optimization and reliability assessment method for efficient probabilistic design. *Journal of Mechanical Design*, 126(2):225–233, 2004.

[36] G. Iaccarino, A. A. Mishra, and S. Ghili. Eigenspace perturbations for uncertainty estimation of single-point turbulence closures. *Physical Review Fluids*, 2(2):024605, 2017.

[37] L. A. McCullers. Aircraft configuration optimization including optimized flight profiles. In *NASA Langley Research Center Recent Experiences in Multidisciplinary Analysis and Optimization, Part 1*, Hampton, VA, 1984.

[38] R. P. Liem, C. A. Mader, and J. R. R. A. Martins. Surrogate models and mixtures of experts in aerodynamic performance prediction for aircraft mission analysis. *Aerospace Science and Technology*, 43:126–151, 2015.

[39] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. K. Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28, 2005.

[40] B. Malone and W. H. Mason. Multidisciplinary optimization in aircraft design using analytic technology models. *Journal of Aircraft*, 32(2):431–438, 1995.

[41] M. Krull. Identifying wave drag for the generic drag polar equation – unveiling polars of 16 passenger aircraft. Project report, Aircraft Design and Systems Group (AERO), Hamburg University of Applied Sciences, March 2025. URN: urn:nbn:de:gbv:18302-aero2025-03-07.011.

[42] D. P. Raymer. *Aircraft Design: A Conceptual Approach*. AIAA, Reston, VA, 3rd edition, 1999.

[43] D. P. Wells, B. L. Horvath, and L. A. McCullers. The flight optimization system weights estimation method. Technical Report NASA/TM-2017-219627, Volume I, NASA Langley Research Center, 2017.

[44] John A. Schaefer, Andrew W. Cary, and Manas Khurana. Surrogate Model and Discretization Error Impacts on the Fluid Dynamics Uncertainty Quantification Challenge Problem. In *AIAA SCITECH 2024 Forum*, Orlando, FL, January 2024. American Institute of Aeronautics and Astronautics. AIAA Paper 2024-0706.

[45] Dieter Scholz and Vladimir Zhuravlev. Mass Estimation of Folding Wings. In *83rd International Conference on Mass Properties Engineering*, Melbourne Beach, FL, USA, May 2024. Society of Allied Weight Engineers (SAWE).

[46] H. Gu, F. J. Healy, D. Rezgui, and J. E. Cooper. Sizing of high-aspect-ratio wings with folding wingtips. *Journal of Aircraft*, 60(2):461–475, 2023.