Evaluating Hybrid Electric Turboprop Development with Novel Metrics: Failure Impact and Uncertainty Impact

Tavish Pattanayak * and Dimitri Mavris†
Georgia Institute of Technology, Atlanta, Georgia, 30332

This paper introduces two new metrics, Failure Impact (FI) and Uncertainty Impact (UI), to address the limitations of traditional evaluation methods like Technology Readiness Level (TRL) and Integration Readiness Level (IRL) in the development of hybrid-electric turboprop aircraft. These metrics, scored on a 1-9 scale, quantify the severity of component failures and the impact of uncertainties. The study applies these metrics to a 50-passenger hybrid-electric turboprop, focusing on four critical components: battery, motor, inverter, and thermal management system (TMS).

Using a Multi-Attribute Decision-Making (MADM) framework, the paper evaluates these components under four strategic scenarios: Balanced, Safety-First, Maturity-Focused, and Uncertainty-Reduction. The results from various MADM methods, including MAUT, AHP, BWM, and TOPSIS, consistently identify the battery as the highest priority for testing and resource allocation due to its high FI and UI scores. The TMS consistently ranks as the second priority. The rankings of the motor and inverter vary depending on the strategic focus, particularly under the Uncertainty-Reduction scenario. A non-compensatory method, ELECTRE III, provided a contrasting ranking by placing the battery last due to its extreme risk scores triggering vetoes. The study concludes that the Best-Worst Method (BWM) is particularly well-suited for aerospace applications due to its simplicity and consistency.

I. Introduction

The aviation industry faces mounting pressure to achieve net-zero carbon emissions by 2050, driving innovation in sustainable propulsion systems [1]. Hybrid electric turboprops, designed for 30–100 passengers, offer a promising solution for regional flights (250-500 km), combining electric motors with conventional engines to reduce fuel consumption while maintaining reliability [2, 3]. However, developing these systems introduces significant challenges: immature technologies, complex integration, and stringent safety and certification requirements [4]. Established metrics like Technology Readiness Level (TRL) and Integration Readiness Level (IRL) assess technological and integration maturity but fail to capture the severity of component failures or uncertainties in performance and scalability [5, 6].

*Graduate Research Associate, Daniel Guggenheim School of Aerospace Engineering, and AIAA Member
†Distinguished Regents Professor and Director of ASDL, Daniel Guggenheim School of Aerospace Engineering, and AIAA Member

This paper proposes two novel metrics - Failure Impact (FI) and Uncertainty Impact (UI) - scored on a 1-9 scale, to quantify failure consequences and development uncertainties for critical components. We then apply these metrics to a hybrid-electric turboprop aircraft use case. However, these metrics offer a versatile framework for decision-makers across industries, such as electric vehicles and renewable energy, enabling risk-informed design and certification. By evaluating components through four decision-making scenarios (Balanced, Safety-First, Maturity-Focused, Uncertainty-Reduction), this study demonstrates FI and UI's utility in prioritizing research, ensuring safety, and advancing sustainable aviation systems.

II. Background and Motivation

Hybrid electric turboprops are gaining traction as a practical step toward sustainable regional aviation, leveraging electric propulsion for efficiency and turboprops for redundancy [7]. However, components like batteries, electric motors, inverters, and thermal management systems introduce unique risks and uncertainties. For instance, battery failures pose safety hazards (e.g., thermal runaway), while the unproven scalability of thermal systems limits certification confidence [4, 8]. Existing metrics fall short in addressing these challenges. TRL and IRL, defined by NASA and Sauser et al., assess maturity but overlook failure severity and uncertainty [5, 6]. Probabilistic Risk Assessment (PRA) and Failure Mode and Effects Analysis (FMEA) focus on failure likelihood or operational risks, not component-specific impact or development uncertainties [9]. Uncertainty quantification in environmental models addresses output variability, not technology development [10]. This gap necessitates new metrics. Beyond aviation, FI and UI can guide decision-makers in fields like electric vehicles (e.g., assessing battery fire risks) or renewable energy (e.g., evaluating solar panel scalability). By applying FI and UI to a hybrid turboprop, this paper provides a framework to prioritize safety-critical components, reduce uncertainties, and inform certification, offering a systems-level approach for sustainable technology development.

III. Methodology

This study introduces two novel metrics, Failure Impact (FI) and Uncertainty Impact (UI), to evaluate critical components for any complex project, alongside Technology Readiness Level (TRL) and Integration Readiness Level (IRL). Scored on a 1-9 scale, FI quantifies the severity of component failures on safety, mission success, and system redundancy, while UI measures the impact of component uncertainties on system-level outcomes. Unlike TRL and IRL, which assess maturity [5, 6], or probabilistic risk assessments (PRA) and failure mode analyses (FMEA), which focus on likelihood [9], FI and UI address consequence severity and uncertainty impact, critical for aviation certification, and cross-industry applications. This section defines FI and UI, explains their calculation, and outlines their application to a 50-passenger hybrid turboprop use case, evaluated through four decision-making scenarios.

A. Failure Impact (FI)

FI measures the severity of a component's failure on safety, mission success, and system redundancy, scored from 1 (negligible) to 9 (catastrophic), calculated as:

$$\text{FI} = (0.5 \times \text{Safety}) + (0.3 \times \text{Mission Impact}) + (0.2 \times \text{Redundancy Impact}) \tag{1}$$

Weights reflect aviation priorities: safety (0.5) is paramount due to zero-tolerance for catastrophic failures, [4, 9]; mission impact (0.3) addresses operational reliability; and redundancy impact (0.2) accounts for built-in backups for redundancies (e.g., turboprop mitigating electric failures). Aviation safety analyses consistently highlight the disproportionate consequences of safety-related failures [9], justifying the 0.5 weight for a factor that can lead to catastrophic outcomes like thermal runaway. Mission impact (0.3) and redundancy impact (0.2) account for operational reliability and the effectiveness of built-in backups, which are significant but secondary to catastrophic safety events[8]. Alternative weighting schemes, such as equal weights, were considered, but they did not align with aviation's safety-first ethos. The formula normalizes the final score to a 1–9 via rounding, ensuring compatibility with the TRL scale [5]. By integrating failure severity with maturity, FI provides a holistic risk assessment that enables decision-makers to prioritize mitigation efforts as technologies progress. Sub-variables are scored using literature review, industry reports, or expert input, with definitions in Table 1.

B. Uncertainty Impact (UI)

UI quantifies a component's uncertainty contribution to system-level outcomes, scored from 1 (negligible) to 9 (dominant). Unlike traditional uncertainty quantification, which often focuses on environmental variability [11], UI targets the impact of component-level uncertainty, a critical factor for certifying emerging technologies. UI is derived from the total-order Sobol indices through sensitivity analysis [12]. These indices measure the fraction of variance in system outcomes attributable to each component's input uncertainty, including all interaction effects. This approach offers a systems-level perspective on risk that is distinct from maturity-focused metrics like TRL [5] or probability-based methods like PRA [9].

The process involves four steps: (1) modeling system outcomes, (2) defining uncertainty distributions for component inputs, (3) computing total-order Sobol indices via Monte Carlo simulations, and (4) mapping the resulting indices to a 1-9 scale. The thresholds for this mapping are justified by sensitivity analysis literature [13–15], which suggests that total Sobol index greater than 0.3 indicates a dominant uncertainty contributor, while an index below 0.15 reflects minimal impact. These thresholds were calibrated to meet aviation's need for precise uncertainty quantification, ensuring UI prioritizes components with the greatest system-level risk.

The 1–9 UI scale is defined in Table 2.

Table 1 Definition of FI Sub-Variables (1–9 Scale)

Score	Safety	Mission Impact	Redundancy Impact
1	No safety impact	No mission impact	Redundancy fully mitigates failure impact
2	Negligible safety impact (minor discomfort)	Negligible mission impact (<5% performance degradation	Redundancy mostly mitigates failure impact (negligible residual impact)
3	Minor safety impact (slight injury)	Minor mission impact (5-10% degradation)	Redundancy partially mitigates failure impact (minor residual impact)
4	Moderate safety impact (injury requiring medical attention)	Moderate mission impact (10-20% degradation)	Redundancy moderately mitigates failure impact (moderate residual impact)
5	Significant safety impact (serious injury)	Significant mission impact (20-50% degradation)	Redundancy slightly mitigates failure impact (significant residual impact)
6	Major safety impact (severe injury)	Major mission impact (>50% degradation)	Redundancy minimally mitigates failure impact (major residual impact)
7	Severe safety impact (life-threatening injury)	Severe mission impact (mission compromised)	Redundancy barely mitigates failure impact (severe residual impact)
8	Critical safety impact (multiple life-threatening injuries)	Critical mission impact (mission failure with partial objectives met)	Redundancy does not mitigate failure impact (critical impact)
9	Catastrophic safety impact (fatalities)	Catastrophic mission impact (total mission failure)	No redundancy, failure impact is unmitigated (catastrophic/single point failure)

Table 2 Definition of UI Levels (1–9 Scale) Based on Sobol Indices (S_{Ti})

Score	Uncertainty Impact Description
1	Negligible impact ($S_{Ti} < 0.05$): Component uncertainty has minimal effect on outcomes (e.g., motor efficiency on range)
2	Very low impact ($0.05 \leq S_{Ti} < 0.10$): Slight contribution to outcome variance
3	Low impact ($0.10 \leq S_{Ti} < 0.15$): Minor influence, manageable uncertainty (e.g., inverter on efficiency)
4	Low-moderate impact ($0.15 \leq S_{Ti} < 0.20$): Noticeable but not critical contribution
5	Moderate impact ($0.20 \leq S_{Ti} < 0.25$): Significant uncertainty requiring attention
6	Moderate-high impact ($0.25 \leq S_{Ti} < 0.30$): Notable influence on system outcomes (e.g., thermal system on safety)
7	High impact ($0.30 \leq S_{Ti} < 0.35$): Substantial uncertainty affecting performance
8	Very high impact ($0.35 \leq S_{Ti} < 0.40$): Major contributor to system uncertainty
9	Dominant impact ($S_{Ti} \geq 0.40$): Critical uncertainty driving outcome variability (e.g., battery on range)

C. Application Use-Case

The application use-case evaluates the proposed Failure Impact (FI) and Uncertainty Impact (UI) metrics within a broader decision-making framework designed to help decision-makers allocate resources in the development of a 50 pax hybrid electric turboprop aircraft. The use-case prioritizes testing and development of critical components—battery,

electric motor, inverter, and thermal management system (TMS). Hybrid-electric turboprops couple these components to a conventional gas-turbine core. Because resources often preclude simultaneous testing of all four subsystems to flight-test maturity, the resource allocation problem can be framed as a multi-attribute decision-making (MADM) problem whose alternatives are the four components and whose attributes capture the four dominant risk dimensions identified. The question to answer here becomes: "Which component should receive the next tranche of resources, given multiple risk dimensions that matter to regulators, engineers, and program managers?" The attributes are summarized in the table 3.

Attribute	**Description**	**Desired Direction**
Failure Impact (FI)	Severity of the consequences if the component were to fail during operation	Higher –> Higher Priority
Uncertainty Impact (UI)	Severity of potential consequences arising from uncertainties associated with the component's performance	Higher –> Higher Priority
Technology Readiness Level (TRL)	Maturity level of the component technology	Lower –> Higher Priority
Integration Readiness Level (IRL)	Readiness of the component for integration with other elements of the subsystem	Lower –> Higher Priority

Table 3 Attributes identified for the MADM problem

Table 4 compiles FI, UI, TRL, and IRL scores for the components from the most recent literature review.

The battery subsystem received an FI of 9 (on the 1–9 impact scale), reflecting its potential for catastrophic failure. This is derived from the Equation 1 wherein the scores for safety, mission impact, and redundancy impact were all 9 for the battery. An internal cell fault or thermal runaway can lead to fire or explosion, jeopardizing the entire aircraft. If battery fails, then that is considered to be total mission failure, and there is no redundancy as this is a single-point failure. FAA tests have shown that an overheating lithium-ion battery can trigger a chain reaction in which adjacent cells ignite in succession. Such propagation is extremely difficult to extinguish in flight [16, 17]. This highest failure impact rating is warranted given historical incidents (e.g. cargo and 787 battery fires [18]) and the need for rigorous containment to prevent a single-cell failure from escalating into a loss-of-aircraft event. The battery's UI was also rated high due to significant uncertainty in its novel aviation application. Despite extensive ground testing, there remain unknowns in how a high-energy-density battery will behave across all flight conditions (vibration, altitude, thermal cycling) [19]. The X-57 program evidenced this uncertainty: an early battery pack test in 2016 resulted in an unexpected thermal runaway, prompting a complete module redesign for safety [20]. Such lessons learned underline the unpredictability and risk margins associated with advanced batteries, justifying a high uncertainty impact score. In terms of maturity, the battery technology itself is fairly mature as lithium-ion systems have been proven in aerospace use (NASA classifies battery tech at roughly TRL 9 for space applications) [21]. However, the TRL assigned was 4 to reflect the state of the integrated aircraft battery pack, as well as the high specific energy required to fulfill future flights. No crewed electric aircraft has

flown with a battery of this scale; X-57's development had to advance the battery (of a lower specific energy) from a lab prototype to a flight-ready subsystem (achieving on the order of TRL 5–6) through iterative testing [20]. By the time of this study, the battery pack had demonstrated full functionality and safety in a relevant ground test environment, but not yet in an actual flight. The advancements in specific energy also remain uncertain. The battery's IRL was judged to be 4 as well, acknowledging the integration challenges remaining. Safely installing a large Li-ion pack into an airframe requires managing thermal loads, electrical integration, and containment for fault conditions. The extensive redesign and validation of X-57's battery modules, which included propagation tests to ensure a single-cell failure could be isolated [20], illustrate that significant integration effort was needed before the battery could be deemed flightworthy.

Component	**FI**	**UI**	**TRL**	**IRL**	Principal Evidence
Battery	9	9	4	4	FAA SAFO 10017 & TC-13-53 thermal-runaway fire data[16, 17]; NASA Small Spacecraft Tech SOA 2024 lists MW-class packs at TRL 4 [22]
Electric Motor	8	1	6	6	NEAT 350 kW motor HIL ground test met TRL 6/IRL 6 criteria [23, 24]
Power Converter/Inverter	7	3	8	8	X-57 Maxwell traction-power system flight-qualified (Mod II), achieving TRL 8 & IRL 8 [25]
Thermal Management System (TMS)	8	1	5	4	Clean-Aviation TheMa4HERA aims at TRL 5 by 2025; project notes cascading safety effects of TMS failure [26]

Table 4 Components with their assigned attribute values and sources

For the electric motor, the safety score is 8 as it is critical, but not always fatal. The mission impact is 8 as there is a scope for partial recovery even though a motor failure in flight would result in immediate loss of thrust. Similarly, the redundancy impact is 8, as there is some redundancy, but it is still a major issue. From these scores, the FI was calculated from the Equation 1 to be 8, slightly lower than the battery but still very high. In a twin-motor configuration, losing one unit can critically impair climb performance and controllability, so each motor is treated as a flight-critical component. The FI rating reflects this safety impact, albeit recognizing that a motor failure alone is somewhat less dire than a battery fire. The UI for the motors was assessed at 1, because electric motors themselves are a well-understood technology with extensive ground heritage (brushless AC motors have decades of development [27]). We assign the motors a TRL of 6 on the NASA scale. Electric propulsion motors had been demonstrated in relevant environments through dynamometer tests and were nearly flight-ready by the end of X-57's Mod II development. The X-57 team initially sought to use off-the-shelf motors with high TRL to reduce risk [28], and the cruise motors ultimately underwent extensive qualification testing (spin tests, vibration, thermal) to ensure they met aircraft requirements. By 2023, the X-57 motors had run at full power in ground trials, which corresponds to a prototype demonstrated in a relevant environment (TRL 6). Full flight demonstration (TRL 7+) had not yet occurred, keeping the rating at the system/subsystem prototype

level. The IRL for the motor was rated similarly at 6. This accounts for the effort needed to integrate the motors with propellers, mountings, power electronics, and the airframe. The motor and inverter had to function as a tightly coupled unit, and issues like torque oscillations or electrical harmonics needed to be tuned out during system integration testing [29]. X-57's reports indicate that even with high-TRL components, significant challenges arose in integrating lower-TRL hardware into the aircraft [28], underscoring that integration maturity lagged behind component maturity. Successful dynamometer tests of the complete motor–inverter assembly and its control software [29] demonstrated progress in integration; furthermore, dedicated efforts were made to address control stability in the system [30]. These achievements justify an IRL in the mid-range (the motors were proven in a combined hardware-in-loop environment). Still, until the motors operate reliably as part of the full aircraft in flight, their integration readiness cannot be considered to be fully mature.

For the inverter, safety is scored at 7, recognizing that high-voltage inverter faults can cause localized heating, arc faults, or ejection of material, posing physical risk to nearby systems. However, these events are less likely to escalate into full-system failures if containment measures are in place. Mission impact is also rated at 7 as a failed inverter will disable the associated motor, resulting in a major loss of thrust and likely mission compromise. Redundancy impact is rated 7, as inverter failure in a distributed electric propulsion setup may be partially mitigated by other functioning inverters or propulsion nodes, but system-level consequences remain significant. Applying the weights to the Equation 1, produces a final score of 7, indicating a high but somewhat bounded level of failure severity. An inverter failure would automatically render the corresponding motor inoperable. Moreover, failure modes of high-power inverters can be hazardous in their own right; for example, a shorted semiconductor device can cause a direct fault on the high-voltage DC bus. Testing on X-57 showed that MOSFET (Metal-Oxide-Semiconductor Field-Effect Transistor) failures in the inverter could short circuit the DC bus, rapidly heat, and explode/eject material, if not properly contained [31]. Such an event could damage surrounding systems or even ignite the battery. The UI for the inverter was evaluated to be 3 in line with the definition given in Table 2. The inverter is assigned a TRL of 8. This acknowledges that the technology had progressed beyond lab breadboards and was demonstrated in a relevant environment (the X-57 cruise motor controllers had passed ground qualification tests by 2023). They achieved key performance goals such as 98% efficiency at full power, enabling purely air-cooled operation [32], and survived rigorous thermal/vibration trials. Finally, the inverter's IRL was also given to be 8. By the end of ground testing, the inverters had been successfully integrated with the motors and batteries on the X-57 test article, including operation of the complete electric propulsion system. Achieving this required overcoming integration hurdles such as ensuring electromagnetic compatibility and coordinating control algorithms with the rest of the aircraft's systems [30]. The inverter was demonstrably working as part of the integrated ground-test system, warranting a high IRL.

The Thermal Management System (TMS) was derived to have an FI of 8. It earns a safety score of 8, since inadequate thermal control can lead to overheating in batteries, inverters, or motors, any of which may then fail in a

dangerous mode. While the TMS itself might not be explosive or flammable, its failure can indirectly cause catastrophic outcomes by disabling critical subsystems. The mission impact is rated at 7, as a degraded or failed TMS can lead to derating of power systems, overheating shutdowns, or emergency procedures, effectively compromising the flight. Redundancy impact is scored at 7, reflecting that while some passive thermal buffering or airflow can provide limited backup, the TMS typically operates as a single integrated loop with little modular redundancy. Using the Equation 1, FI = 0.5×8 + 0.3×7 + 0.2×7 = 7.5, which rounds to 8, the TMS is identified as a component with significant failure consequences, especially through its cascading effects on other systems. While a thermal system failure might not be as immediately dramatic as, say, a battery fire, it has broad and potentially dire implications. Inadequate cooling will lead to overheating of the battery, inverters, and motors; over time, this can trigger performance degradation or cascading failures (for instance, overheated cells can go into thermal runaway, or electronics can fail and short). Thus, the TMS plays a direct role in preventing other high-impact failures. X-57 documentation notes that the inverter enclosure and design had to account for worst-case MOSFET failures and ensure the TMS could "reject the required heat" to prevent catastrophic outcomes [31]. A loss of cooling capacity, especially in an electric aircraft where heat loads are high, could precipitate exactly the kind of battery thermal event we seek to avoid. Therefore, the TMS is regarded as a critical safeguard and earns a high FI score. As with the others, the UI was evaluated in line with the definition given in Table 2. The TMS's technological maturity was assessed to be the lowest among the four components, at TRL 5. The fundamentals of thermal management (heat exchangers, cooling loops, etc.) is known, but the specific implementation for a high-power-density electric aircraft is only at a breadboard or prototype level. In X-57's case, the chosen approach relied heavily on the propulsion components' high efficiency (minimizing waste heat) and forced-air cooling for electronics [32]. This approach was validated in ground tests to the extent possible, with chamber tests showing the controllers stayed within limits with airflow cooling, characteristic of TRL 5 (component/system validation in relevant environment). Yet, because the TMS had not been proven in an actual flight scenario (with real aerodynamic and environmental conditions), it was not credited with a higher TRL. Lastly, the TMS's IRL was rated at 4, reflecting that significant integration work was still needed to fully harmonize the thermal subsystem with the aircraft. The TMS must interface with every other power component (battery modules, inverters, motors) and the airframe (e.g., coolant ducts or heat sinks placed in the structure). Ensuring that all parts of the system stay within allowable temperatures during joint operation is a complex integration task. NASA engineers noted that bringing all of Maxwell's systems together and making them work in concert was one of the biggest challenges of the project, and thermal management is a big part of the challenge [32]. By the end of the development, the TMS had been partially integrated (cooling provisions were built into the X-57's design and tested on the ground), but a fully integrated demonstration under flight conditions had not yet occurred, justifying the IRL 4 assigned. It was proven that the TMS could function in the ground test environment, but its readiness to perform in an operational flight context remains to be conclusively demonstrated.

The subsequent text in this subsection goes into the methodology of single-attribute and multi-attribute decision-

making. However, before that, it is important to note that to preserve a consistent higher-score = higher-priority convention, the two maturity attributes are first transformed. While attributes such as Failure Impact (FI) and Uncertainty Impact (UI) inherently adhere to this directional principle (where a larger score signifies a more critical concern) Technology Readiness Level (TRL) and Integration Readiness Level (IRL) operate on an inverse scale, where higher scores denoting greater maturity and thus less immediate concern. To reconcile this, these maturity metrics are inverted:

$\overline{\text{TRL}}$ = 9 - TRL, $\overline{\text{IRL}}$ = 9 - IRL,

so that larger values represent larger maturity gaps.

1. Single-attribute Decision Making

Before advocating for any integrated prioritisation scheme, the study first establishes a single-attribute baseline: each component is ranked four separate times, once by Failure-Impact (FI), once by Uncertainty-Impact (UI), once by Technology-Readiness-Level (TRL), and once by Integration-Readiness-Level (IRL). This baseline lets readers see how much the preferred test list can swing when decision-makers focus on only one risk dimension at a time. In later sections, performance gains from multi-attribute methods will be expressed as improvements relative to this one-dimensional benchmark.

For each attribute, a self-contained ranked list was generated by ordering the four components strictly according to that single metric. Components are then ranked in descending order of:

- FI (highest failure impact first)
- UI (highest uncertainty impact first)
- $\overline{\text{TRL}}$ (least mature component first)
- $\overline{\text{IRL}}$ (least likely to integrate component first)

Ties, although unlikely with a 1–9 scale, are resolved conservatively by awarding the same rank to both components and leaving subsequent ranks blank. These four deterministic league tables constitute the single-attribute baseline against which multi-attribute rankings will be compared in the next section.

2. Multi-Attribute Decision-Making Framework

Real-world programs often need to balance multiple objectives at the same time. Therefore, a multi-attribute decision making framework is tested in order to enable decision making for resource allocation towards technology maturation in the context of a 50 pax hybrid electric turboprop aircraft. Doing so:

1) combines the four risk attributes into a single priority score for each component
2) allows stakeholders to express different strategic emphases through weight vectors
3) is transparent, reproducible, and statistically auditable

To prevent any single attribute from dominating the analysis due to its numerical range, min-max scaling was applied

Scenario	Weight Recipe (w_{UI},w_{TRL},w_{IRL},w_{IRL}, w_{FI})	Usage
Balanced	(0.25, 0.25, 0.25, 0.25)	Early road-mapping phases when there's an equal desire to understand all potential risks and opportunities without prejudging future challenges
Safety-First	(0.10, 0.10, 0.10, 0.70)	Critical stages like certification audits or public-safety reviews: emphasis on FI ensures that components with even a remote possibility of catastrophic failure are prioritized for rigorous testing
Maturity-Focused	(0.10, 0.40, 0.40, 0.10)	Periods leading up to a significant milestone, this approach pushes immature or poorly integrated components to the forefront of the test queue to mitigate schedule risks
Uncertainty-Reduction	(0.70, 0.10, 0.10, 0.10)	Reduce errors and improve predictive accuracy before committing to physical building or costly fabrication

Table 5 Table showing the four scenarios for decision making to be tested in MADM, along with some suggested use cases for each

to normalize each attribute column to a standardized [0,1] range. This robust transformation ensures that the "worst" score in any column becomes 1 (highest urgency) and the "best" score becomes 0 (lowest urgency), preserving rank order while giving decision-makers an unbiased picture of where each component stands.

Real-world programs must balance multiple objectives, and priorities can shift. Therefore, it was decided to define four distinct, easily interpretable scenarios, each represented by a unique weight vector that reflects a different strategic emphasis. Table 5 summarizes these scenarios, their attribute weights, and their intended use cases. This approach allows us to see how component rankings change when strategic priorities shift from a balanced view to one focused on safety, maturity, or uncertainty reduction.

To ensure our results are not an artifact of a single optimization routine, five conceptually distinct MADM algorithms were applied to each of the four weighting scenarios. This comprehensive approach provides a reliable and thoroughly validated set of recommendations. A summary of each algorithm is found in the Table 6. The detailed mathematical procedure for each of these five methods is provided in Appendix A.

Employing this suite of algorithms yields twenty priority lists (five methods × four weighting scenarios).

IV. Results

The results are presented in five subsections. The first subsection establishes a single-attribute baseline—the four "league tables" obtained when each risk metric is considered in isolation. The subsequent subsections build on this baseline by applying the compensatory (MAUT, AHP, BWM, TOPSIS) and non-compensatory (ELECTRE III) methods described earlier. Unless stated otherwise, all rankings are derived from the attribute scores in Table 4 after direction harmonization and min–max scaling.

Algorithm Name	Core Principle & Calculation	Key Characteristics & Justification	Reference
Multi-Attribute Utility Theory (MAUT) (Additive Form)	Each component's aggregate utility U_i is calculated as a weighted sum of its normalized attribute scores: $U_i = \sum_j w_j v_{ij}$.	This approach is fully compensatory, meaning a low score in one attribute can be offset by high scores in others. Its mathematical transparency makes the decision logic easily traceable.	Keeney & Raiffa (1976) [33]
Analytic Hierarchy Process (AHP)	Stakeholder preferences are systematically elicited through a consistency-checked pairwise comparison matrix. The principal eigenvector of this matrix yields the attribute weights, which are then used in a weighted-sum scoring model to rank alternatives.	AHP provides a structured framework for incorporating subjective judgments, ensuring internal consistency in preference elicitation.	Saaty (1980) [34]
Best-Worst Method (BWM)	Decision-makers identify the single most-important and least-important attributes. A linear programming model then derives attribute weights by minimizing the maximum relative deviations, thereby reducing the cognitive burden on decision-makers while maintaining high consistency.	BWM offers a streamlined approach to weight elicitation compared to AHP, requiring fewer comparisons and often leading to more consistent results.	Rezaei (2015) [35]
Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)	Components are geometrically positioned in a multi-dimensional normalized attribute space. Priority is assigned based on the Euclidean distance from an "ideal point" (representing the best possible score on every attribute) and an "anti-ideal point" (representing the worst score on every attribute).	TOPSIS identifies solutions that are simultaneously close to the ideal and far from the anti-ideal solution, offering a comprehensive relative ranking.	Hwang & Yoon (1981) [36]
ELECTRE III	An outranking procedure establishes pairwise dominance relations between components using pre-defined indifference, preference, and veto thresholds. A subsequent "descending distillation" process then produces a complete ranking.	This method can effectively penalize a single unacceptable score, even if other attributes are favorable. ELECTRE III is particularly useful when dealing with uncertainty and imprecision in preferences, allowing for non-compensatory criteria where a poor performance on one attribute can disqualify an alternative.	Roy (1991) [37]

Table 6 Comparison of Multi-Criteria Decision-Making Algorithms

A. Single Attribute Decision Making

Table 7 presents the single-attribute rankings for the four components based on FI, UI, $\overline{\text{TRL}}$, and $\overline{\text{IRL}}$, establishing the baseline for multi-attribute analysis. These rankings are obtained when only one risk dimension is considered at a time as the decision rule.

- **Failure Impact (FI)** – Test the component whose failure would be most catastrophic.
- **Uncertainty Impact (UI)** – Test the component whose parameter uncertainty drives the largest share of output variance.
- **Technology Maturity Gap** $\left(\overline{\text{TRL}} \equiv 9 - \text{TRL}\right)$ – Test the least-mature technology first.
- **Integration Maturity Gap** $\left(\overline{\text{IRL}} \equiv 9 - \text{IRL}\right)$ – Test the component least ready for integration.

Ranks are assigned in descending order of each metric; ties receive the same ordinal number, and the following rank is skipped.

Rank	FI	UI	$\overline{\text{TRL}}$	$\overline{\text{IRL}}$
1	Battery	Battery	Battery	Battery, TMS
2	Electric Motor, TMS	Inverter	TMS	–
3	–	Electric Motor, TMS	Electric Motor	Electric Motor
4	Inverter	–	Inverter	Inverter

Table 7 Single-Attribute Rankings for Hybrid Electric Turboprop Components. Components are ranked in descending order based on Failure Impact (FI), Uncertainty Impact (UI), and transformed Technology Readiness Level ($\overline{\text{TRL}} = 9 - TRL$) and Integration Readiness Level ($\overline{\text{IRL}} = 9 - IRL$). Ties are assigned the same rank, with subsequent ranks skipped.

The battery subsystem tops every column, confirming that it is simultaneously the most hazardous (FI 9), the most uncertain (UI 9), and the least mature when maturity is expressed as a residual gap. The Thermal-Management System (TMS) ties for first place under the integration-gap metric but drops to second under FI and $\overline{\text{TRL}}$ and disappears from the podium under UI, foreshadowing its weight-dependent behaviour in multi-attribute analyses. The electric motor appears ahead of the inverter when failure consequences dominate, but the inverter overtakes when uncertainty alone is judged, suggesting that the inverter's risk profile is knowledge-driven rather than consequence-driven. Because each one-dimensional ranking emphasizes a different component for testing priority, stakeholders cannot rely on any single metric without risking blind spots, which serves as the motivation for the multi-attribute methods, whose results are presented next.

B. Multi-Attribute Utility Theory (MAUT)

MAUT aggregates the four risk attributes—*Failure Impact* (FI), *Uncertainty Impact* (UI), and the two inverted maturity gaps $\overline{\text{TRL}}$ and $\overline{\text{IRL}}$—into a single utility score *.

*Each attribute vector was first direction-harmonised and min–max normalised III.C.2

Table 8 presents the component rankings derived from the Multi-Attribute Utility Theory (MAUT) algorithm across the four decision-making scenarios, based on normalized attribute scores and scenario-specific weights.

Scenario	Rank 1	Rank 2	Rank 3	Rank 4
Balanced	Battery	TMS	Electric Motor	Inverter
Safety-First	Battery	TMS	Electric Motor	Inverter
Maturity-Focused	Battery	TMS	Electric Motor	Inverter
Uncertainty-Reduction	Battery	TMS	Inverter	Electric Motor

Table 8 Multi-Attribute Utility Theory (MAUT) Rankings for Hybrid Electric Turboprop Components. Components are ranked based on utility scores calculated as weighted sums of normalized attribute scores (FI, UI, $\overline{TRL}$, $\overline{IRL}$) across four decision-making scenarios.

In all four scenarios, the battery subsystem retains first place, underscoring its simultaneous exposure to catastrophic failure, high epistemic uncertainty, and low maturity gaps. Even when Uncertainty Impact receives a 0.70 weight, the battery's already extreme FI score keeps it atop the utility list. The TMS claims second position in every scenario, reflecting a broad middle-ground risk profile: it is less catastrophic than the battery, but more uncertain than the motor or inverter, and its integration gap ties it with the battery under the $\overline{\text{IRL}}$ metric. This robustness signals that funding TMS tests pays dividends across a wide range of management priorities. The ordering of the two electric-power-train components (Motor vs. Inverter) flips only under the *Uncertainty-Reduction* scenario. There, the inverter's relatively larger UI contribution outweighs its higher maturity, nudging it above the motor in overall utility. In the other three scenarios, the motor's higher Failure Impact is decisive.

Although ranks provide an intuitive hierarchy, the raw MAUT scores reveal that the utility gap between Rank 1 and Rank 2 is roughly twice the gap between Rank 2 and Rank 3. Hence, the decision to test the battery first is comparatively "hard," whereas the choice between motor and TMS is "soft". These findings will be compared against the other MADM methods - the judgement-centric AHP and BWM, the geometry-based TOPSIS, and outranking-based ELECTRE III.

C. Analytic Hierarchy Process (AHP)

AHP converts scenario-specific pairwise comparisons into weight vectors and utilities via the eigenvector procedure. The resulting rankings are summarized in Table 9, accompanied by the internal *Consistency Ratio* (CR) for each scenario. [†]

Three of the four scenarios reproduce the MAUT order exactly. Divergence appears only in Uncertainty-Reduction, where AHP places the electric motor above the inverter. Although the Uncertainty-Reduction matrix exceeds Saaty's 0.10 guideline (0.195), sensitivity tests show the battery remains first by a large margin even if the most extreme 7:1 judgements are relaxed to 5:1. The principal effect of the inconsistency is the motor–inverter swap at the tail of the

[†]Balanced, Safety-First, and Maturity-Focused matrices were perfectly consistent (CR = 0) because their pairwise judgements formed exact multiplicative chains. The Uncertainty-Reduction matrix shows CR = 0.195 owing to the strong 7:1 intensities assigned to UI; it was decided to retain this moderate inconsistency and document it rather than dilute the stated emphasis on uncertainty.

Scenario	Rank 1	Rank 2	Rank 3	Rank 4	CR
Balanced	Battery	TMS	Electric Motor	Inverter	0.000
Safety-First	Battery	TMS	Electric Motor	Inverter	0.000
Maturity-Focused	Battery	TMS	Electric Motor	Inverter	0.000
Uncertainty-Reduction	Battery	TMS	Electric Motor	Inverter	0.195

Table 9 AHP rankings and consistency ratios.

ranking, compared to the MAUT rankings.

The next subsection applies the Best–Worst Method (BWM) to the same data set, providing a still leaner elicitation variant and an independent weight-generation check.

D. Best–Worst Method (BWM)

The Best–Worst Method (BWM) uses a leaner judgment process to derive scenario-specific weight vectors, requiring only five pairwise comparisons per scenario. Table 10 presents the weight vectors obtained for each scenario by solving the corresponding linear programs that minimize maximum deviation from the expert-provided "best-to-others" and "others-to-worst" comparisons.

Scenario	w_{FI}	w_{UI}	$w_{\overline{\mathrm{TRL}}}$	$w_{\overline{\mathrm{IRL}}}$
Balanced	0.25	0.25	0.25	0.25
Safety-First	0.70	0.10	0.10	0.10
Maturity-Focused	0.10	0.10	0.40	0.40
Uncertainty-Reduction	0.10	0.70	0.10	0.10

Table 10 Attribute weights derived from BWM across the four decision-making scenarios. All vectors sum to 1.00.

These weight vectors match the MAUT scenario recipes exactly because the pairwise judgments were chosen to reproduce those proportions. All four BWM linear programs returned a minimum deviation value $\xi^* = 0$, yielding a BWM consistency ratio of $\mathrm{CR}_{\mathrm{BWM}} = 0.000$ in every case.

Table 11 shows the component rankings obtained when these weights are applied to the same min-max normalized attribute matrix used for MAUT and AHP.

Scenario	Rank 1	Rank 2	Rank 3	Rank 4
Balanced	Battery	TMS	Electric Motor	Inverter
Safety-First	Battery	TMS	Electric Motor	Inverter
Maturity-Focused	Battery	TMS	Electric Motor	Inverter
Uncertainty-Reduction	Battery	TMS	Inverter	Electric Motor

Table 11 Component rankings under the Best–Worst Method (BWM).

Because the BWM weights were constructed to match the MAUT scenario recipes, the component rankings reproduce

the MAUT results exactly. This validates the judgment inputs and highlights that the MAUT and BWM frameworks are compatible in both formulation and outcome.

The BWM linear model guarantees the lowest possible inconsistency for a given set of best–worst comparisons. The fact that all four cases achieved $\xi^* = 0$ implies that the original judgements were internally coherent and needed no reconciliation. BWM's minimal comparison burden (five values per scenario) and deterministic output make it an attractive choice when stakeholder time is limited or when rapid iteration is needed to reflect evolving programme priorities.

Next, we examine the results for TOPSIS, which shifts from a weighting logic to a geometric lens—ranking components based on their Euclidean proximity to the best and worst possible composite attribute vectors.

E. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

TOPSIS evaluates alternatives by measuring their geometric distance to an ideal (worst-case) and anti-ideal (best-case) composite profile across all attributes. Each component is scored on a 0–1 scale for Failure Impact (FI), Uncertainty Impact (UI), and the inverted maturity gaps ($\overline{\text{TRL}}$, $\overline{\text{IRL}}$). For each decision scenario, these attributes are weighted, and the relative closeness of each component to the worst-case risk vector is computed:

$$C_i^* = \frac{S_i^-}{S_i^- + S_i^+}$$

where S_i^+ and S_i^- are the Euclidean distances from component i to the ideal and anti-ideal points, respectively. A higher C_i^* indicates greater urgency for testing.

Table 12 shows the resulting rankings.

Scenario	Rank 1	Rank 2	Rank 3	Rank 4
Balanced	Battery	TMS	Motor	Inverter
Safety-First	Battery	TMS	Motor	Inverter
Maturity-Focused	Battery	TMS	Motor	Inverter
Uncertainty-Reduction	Battery	TMS	Inverter	Motor

Table 12 TOPSIS rankings for hybrid-electric turboprop components across four decision scenarios. Rankings are based on relative closeness C_i^* to the ideal composite risk vector.

In all four scenarios, the battery receives a perfect closeness score of $C^* = 1.000$, indicating that it lies precisely at the worst-case corner of the weighted attribute space. This result is geometrically expected because the battery holds the maximum score on every attribute (FI = UI = $\overline{\text{TRL}}$ = $\overline{\text{IRL}}$ = 1) before weighting. The Thermal Management System consistently ranks second, regardless of how weights are distributed. Its moderate-to-high risk profile across all four attributes ensures it remains closer to the ideal than either the motor or the inverter, but never surpasses the battery.

In the uncertainty reduction scenario, UI receives a 0.70 weight, magnifying the penalty for epistemic uncertainty.

Although the motor performs better on maturity gaps, its UI = 0 allows no contribution to closeness when UI dominates the composite risk space. In contrast, the inverter, despite being worse on maturity, has a nonzero UI score, bringing it geometrically closer to the ideal. This reverses the ranking seen in MAUT and AHP, where maturity was more strongly emphasized. The score gap between TMS and the motor in the Balanced case is roughly $0.540 - 0.396 = 0.144$, indicating a moderately firm preference. By contrast, the difference between motor and inverter in the Uncertainty-Reduction case is less than 0.03, underscoring how close those options are under UI-heavy weighting.

In summary, the TOPSIS method confirms the dominant role of the battery subsystem while revealing subtle geometric trade-offs among the remaining components, especially under asymmetric weight distributions.

F. Elimination and Choice Expressing Reality (ELECTRE III)

ELECTRE III offers a fundamentally different approach to ranking: it is a non-compensatory, outranking-based method that incorporates real-world features such as imprecise preferences and veto thresholds. Rather than aggregating scores into a single utility value, ELECTRE compares components pairwise to determine whether one outranks another, based on thresholds of indifference q, preference p, and veto v. In this study, all attributes were assigned thresholds $q = 0.05$, $p = 0.20$, and $v = 0.40$ to reflect typical test-planning tolerances.

For each decision scenario, a credibility matrix was constructed, and each component's total number of strong outrankings (credibility score > 0.65 with no veto triggered) was tallied. Table 13 shows the resulting rankings based on descending outrank count.

Scenario	**Rank 1**	**Rank 2**	**Rank 3**	**Rank 4**
Balanced	Inverter	Motor	TMS	Battery
Safety-First	TMS	Inverter	Motor	Battery
Maturity-Focused	TMS	Inverter	Motor	Battery
Uncertainty-Reduction	Inverter	TMS	Motor	Battery

Table 13 ELECTRE III rankings based on outranking counts. A component ranks higher if it strongly outranks more alternatives (credibility > 0.65) without violating any veto threshold.

In all four scenarios, the battery is vetoed by every other component, due to having the maximum possible risk score on every attribute. Its outranking count is zero across the board, which sharply contrasts with the additive methods (e.g., MAUT, BWM, AHP, TOPSIS), where the battery consistently ranks first. The thermal-management system (TMS) outranks all others in both the Safety-First and Maturity-Focused scenarios, while the inverter does the same for the other two scenarios.

ELECTRE III thus highlights failure modes that other methods overlook and provides an important counterpoint to additive ranking strategies.

V. Discussions

This study introduced and applied a systematic Multi-Attribute Decision-Making (MADM) approach for prioritizing component testing in the development of a 50-passenger hybrid-electric turboprop aircraft. The results derived from the various MADM methods (MAUT, AHP, BWM, TOPSIS, and ELECTRE III) reveal critical insights regarding the prioritization of components based on different strategic decision-making scenarios: Balanced, Safety-First, Maturity-Focused, and Uncertainty-Reduction.

Across all additive methods (MAUT, AHP, BWM, and TOPSIS), the battery subsystem consistently ranked as the highest priority. This outcome is primarily attributed to the battery's uniquely severe combination of high failure impact (FI = 9), substantial uncertainty impact (UI = 9), and the lowest technology and integration readiness (TRL = 4, IRL = 4). This consistency underscores the significant risk that the battery poses to safety, mission success, and system reliability, necessitating prioritized testing regardless of strategic perspective. This consistency also indicates robustness in the MADM methods applied, particularly in scenarios where component risks span multiple dimensions.

The Thermal Management System (TMS) consistently secured the second rank across most scenarios and methods, highlighting its critical intermediate risk profile characterized by moderate-to-high failure impacts and integration challenges. The robustness of TMS's ranking across varied decision-making frameworks suggests a persistent requirement for testing to mitigate system-level risks, particularly those associated with integration complexities that could cascade into broader system performance and safety issues.

A notable divergence emerged in the ranking of the electric motor and the inverter. In the scenario focusing on uncertainty reduction, the inverter typically outranked the motor due to its relatively higher uncertainty. In contrast, scenarios that prioritize failure impacts and maturity metrics favored the electric motor. This outcome highlights the sensitivity of ranking decisions to attribute weight assignments, emphasizing the need to carefully align decision-making strategies with clearly articulated stakeholder objectives.

The non-compensatory ELECTRE III approach provided a markedly different prioritization pattern, placing the battery at the lowest rank across all scenarios due to its extreme attribute scores triggering veto thresholds. This unique outcome emphasizes the importance of using complementary decision-making methods to uncover critical vulnerabilities that compensatory models may inadvertently obscure.

Comparing the single-attribute decision-making outcomes with those from the MADM approaches further justifies the use of a multi-attribute framework. Single-attribute decisions showed significant swings in prioritization based on isolated metrics, potentially leading to oversight of critical risks in multidimensional scenarios. MADM approaches, in contrast, provide balanced and nuanced insights that reflect the complexities inherent in the maturation of aerospace technology, where multiple risk factors must simultaneously guide resource allocation.

Considering the comparative effectiveness of the various MADM methods tested, the Best–Worst Method (BWM) emerges as particularly suitable for aerospace application contexts. Its minimal cognitive burden (requiring fewer

comparisons), inherent consistency checks, and straightforward interpretation offer substantial advantages in high-stakes, rapidly evolving project environments like hybrid-electric aviation development. Furthermore, BWM's robustness to varied scenario weightings makes it particularly valuable in adaptive resource allocation contexts.

VI. Conclusion

This study introduced two novel metrics, Failure Impact (FI) and Uncertainty Impact (UI), specifically designed to quantify critical risks and uncertainties inherent in developing hybrid-electric propulsion systems. The practical application of these metrics within a comprehensive Multi-Attribute Decision-Making (MADM) framework demonstrated their utility and effectiveness for prioritizing testing and resource allocation among components such as the battery, Thermal Management System (TMS), electric motor, and power converter.

The robust and consistent prioritization of the battery in various MADM methods and strategic scenarios emphasizes the importance of integrating these new metrics into risk assessments to achieve a holistic view of the vulnerabilities of the components. The consistently high ranking of the Thermal Management System further illustrates the need to consider multiple attributes simultaneously, given its significant yet nuanced risk profile.

In addition, this study validates the use of MADM methodologies in single-attribute decision-making by revealing significant oversights that can arise when only isolated metrics are considered. Among the MADM methods evaluated, the Best–Worst Method (BWM) emerged as particularly advantageous due to its simplicity, interpretability, consistency, and robustness across varying strategic emphases.

Future research should further explore the generalization of the FI and UI metrics across other aerospace and engineering domains, and focus on integrating real-time data analytics into these MADM frameworks. Doing so could lead to adaptive, risk-informed systems capable of dynamically allocating resources in complex and rapidly evolving development environments.

Acknowledgments

The authors would like to thank Dr. Raphael Gautier for his valuable insights that helped further improve and polish this paper.

References

[1] International Civil Aviation Organization, "Environmental Report 2022: Aviation and Environment," Tech. rep., ICAO, 2022. URL https://www.icao.int/environmental-protection/Documents/ICAO_Environmental_Report_2022.pdf, accessed: 2025-06-05.

[2] Pattanayak, T., Uzodinma, J., Gautier, R. H., Walter, M., Zaidi, T., and Mavris, D., "Uncertainty analysis of turbofan and

turboprop architectures for hybrid electric propulsion in nasa epfd vehicles," *AIAA AVIATION FORUM AND ASCEND 2024*, 2024, p. 3826.

[3] Pattanayak, T., Uzodinma, J., Gautier, R., Zaidi, T., and Mavris, D., "Uncertainty Analysis for Hybrid Electric Propulsion in NASA EPFD Vehicles," *2025 IEEE/AIAA Transportation Electrification Conference and Electric Aircraft Technologies Symposium (ITEC+ EATS)*, IEEE, 2025, pp. 1–6.

[4] European Union Aviation Safety Agency, "Special Condition for Electric/Hybrid Propulsion Aircraft (SC-EPHA)," Tech. rep., EASA, 2021. URL `https://www.easa.europa.eu/downloads/128058/en`, accessed: 2025-06-05.

[5] National Aeronautics and Space Administration, "Technology Readiness Level Definitions," Tech. rep., NASA, 2017. URL `https://www.nasa.gov/directorates/heo/scan/engineering/technology/technology_readiness_level`, accessed: 2025-06-05.

[6] Sauser, B. J., Ramirez-Marquez, J. E., Magnaye, R., and Tan, W., "A Systems Approach to Expanding the Technology Readiness Level within Defense Acquisition," *International Journal of Defense Acquisition Management*, Vol. 1, 2010, pp. 39–58. https://doi.org/10.1109/SYSOSE.2008.4724149.

[7] Cardone, L., Petrone, G., De Rosa, S., Franco, F., and Greco, C., "Review of the recent developments about the hybrid propelled aircraft," *Aerotecnica Missili & Spazio*, Vol. 103, No. 1, 2024, pp. 17–37.

[8] Hepperle, M., "Electric Flight – Potential and Limitations," *NATO Science and Technology Organization*, 2012, pp. 1–24. URL https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/RTO-MP-AVT-209/MP-AVT-209-09.pdf, sTO-MP-AVT-209.

[9] SAE International, "Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment," Tech. Rep. ARP4761, SAE International, 1996. URL https://www.sae.org/standards/content/arp4761/.

[10] Wilbanks, T. J., Zimmerman, R., Julius, S., Kirshen, P., Smith, J. B., Moss, R., Solecki, W., Ruth, M., Conrad, S., Fernandez, S. J., et al., "Toward indicators of the performance of US infrastructures under climate change risks," *Climatic change*, Vol. 163, 2020, pp. 1795–1813.

[11] Sahai, T., and Pasini, J. M., "Uncertainty quantification in hybrid dynamical systems," *Journal of Computational Physics*, Vol. 237, 2013, pp. 411–427.

[12] Sobol, I. M., "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and computers in simulation*, Vol. 55, No. 1-3, 2001, pp. 271–280.

[13] Zhang, X.-Y., Trame, M. N., Lesko, L. J., and Schmidt, S., "Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models," *CPT: pharmacometrics & systems pharmacology*, Vol. 4, No. 2, 2015, pp. 69–79.

[14] Wainwright, H. M., Finsterle, S., Jung, Y., Zhou, Q., and Birkholzer, J. T., "Making sense of global sensitivity analyses," *Computers & Geosciences*, Vol. 65, 2014, pp. 84–94.

[15] Xu, R., Pang, Y., and Hu, Z., "Sensitivity analysis of external conditions based on the MARS-Sobol method: case study of Tai Lake, China," *Water Supply*, Vol. 21, No. 2, 2021, pp. 723–735.

[16] of Transportation, U. D., and Service, F. A. A. F. S., *Safety alert for operators*, 2010. URL https://www.faa.gov/sites/faa.gov/files/airports/airport_safety/aircraft_rescue_fire_fighting/SAFO10017.pdf.

[17] Maloney, T., Technology, and Management International, L., *EXTINGUISHMENT OF LITHIUM-ION AND LITHIUM-METAL BATTERY FIRES*, 2014. URL https://www.fire.tc.faa.gov/pdf/TC-13-53.pdf.

[18] Kolly, J. M., Panagiotou, J., and Czech, B., "The investigation of a lithium-ion battery fire onboard a Boeing 787 by the US National Transportation safety board," *Safety Research Corporation of America: Dothan, AL, USA*, 2013, pp. 1–18.

[19] Pattanayak, T., and Mavris, D., "Battery technology for sustainable aviation: a review of current trends and future prospects," *Applied Energy*, Vol. 397, 2025, p. 126356.

[20] Sigler, D., and Sigler, D., "NASA freely shares X57 lessons › Sustainable skies," , Jul. 2018. URL https://sustainableskies.org/x57-lessons-shared-freely/.

[21] Jakupca, I., Apr 2019. URL https://www.hydrogen.energy.gov/docs/hydrogenprogramlibraries/pdfs/review19/ia011_jakupca_2019_o.pdf.

[22] Weston, S. V., Burkhard, C. D., Stupl, J. M., Ticknor, R. L., Yost, B. D., Austin, R. A., Galchenko, P., Newman, L. K., and Soto, L. S., "State-of-the-art small spacecraft technology," 2025.

[23] Dyson, R., and Center, N. G. R., *Hybrid Electric MC-12 Ground Testing Plan Chapter*, ???? URL https://ntrs.nasa.gov/api/citations/20205000528/downloads/20205000528%20Book%20Chapter%20Final.pdf.

[24] Sachs-Wetstone, J., Bianco, S., Kratz, J. L., Horning, M., Amthor, A. E., and Connolly, J. W., "Hybrid-Electric Aero-Propulsion Controls Testbed Results," *AIAA AVIATION 2023 Forum*, 2023, p. 4233.

[25] Clarke, S., and Terry, J., "X-57 Cruise Motor and Inverter Failure Modes," *Committee Meeting for F44 on General Aviation*, 2024.

[26] Cordis, C., "Thermal management for the hybrid electric regional aircraft," , Dec. 2022. URL https://cordis.europa.eu/project/id/101102008?

[27] Pillay, P., and Freere, P., "Literature survey of permanent magnet AC motors and drives," *Conference Record of the IEEE Industry Applications Society Annual Meeting,*, 1989, pp. 74–84 vol.1. https://doi.org/10.1109/IAS.1989.96634.

[28] Reynolds, J., "X-57 Flight Controls Lessons Learned," *Aerospace Control and Guidance Systems Meeting 133*, 2024.

[29] Green, C., Conn, B., and Shemenski, M., "Test Campaign of X-57 Maxwell's 80kW Electric Propulsion System," *AIAA Aviation 2023 Forum*, 2023, p. 4272.

[30] Maliska, H., and Clarke, S., "X-57 Maxwell Aircraft," *AIAA Science and Technology (SciTech) Forum and Exposition 2023*, 2023.

[31] Clarke, S., Terry, J., Harris, K., and Avanesian, D., "X-57 Traction power and command systems development," *34th Congress of the International Council of the Aeronautical Sciences*, 2024.

[32] Marquand, Y. L., Feb. 2023. URL https://www.revolution.aero/news/2023/02/13/nasa-completes-thermal-motor-testing-of-its-x-57-maxwell/.

[33] Keeney, R. L., and Raiffa, H., *Decisions with multiple objectives: preferences and value trade-offs*, Cambridge university press, 1993.

[34] Saaty, R. W., "The analytic hierarchy process—what it is and how it is used," *Mathematical modelling*, Vol. 9, No. 3-5, 1987, pp. 161–176.

[35] Rezaei, J., "Best-worst multi-criteria decision-making method," *Omega*, Vol. 53, 2015, pp. 49–57.

[36] Hwang, C.-L., Yoon, K., Hwang, C.-L., and Yoon, K., "Methods for multiple attribute decision making," *Multiple attribute decision making: methods and applications a state-of-the-art survey*, 1981, pp. 58–191.

[37] Roy, B., "The outranking approach and the foundations of ELECTRE methods," *Theory and decision*, Vol. 31, 1991, pp. 49–73.

[38] Malone, P. K., "Applying System Readiness Levels to Cost Estimates–A Case Study Part 2," *2020 IEEE Aerospace Conference*, IEEE, 2020, pp. 1–20.

[39] Olechowski, A., Eppinger, S. D., and Joglekar, N., "Technology readiness levels at 40: A study of state-of-the-art use, challenges, and opportunities," *2015 Portland international conference on management of engineering and technology (PICMET)*, IEEE, 2015, pp. 2084–2094.

[40] Sauser, B., Ramirez-Marquez, J. E., Magnaye, R., and Tan, W., "A system approach to expanding the technology readiness level within defense acquisition," Tech. rep., Acquisition Research Program, 2008.

[41] Garrett, M. J., Kowalewski, S. R., and Granger, M. G., "X-57 Maxwell High Lift Motor Controller Electromagnetic Interference Test Results," 2024.

[42] Mohan, V., Jeyaraj, A. K., and Liscouët-Hanke, S., "Systems integration framework for hybrid-electric commuter and regional aircraft," *Aerospace*, Vol. 10, No. 6, 2023, p. 533.

[43] Nerone, T., "NASA Sustainable Flight National Partnership Panel," *AIAA AVIATION Forum*, 2024.

VII. Appendix: Detailed MADM Methodologies

Multi-Attribute Utility Theory (MAUT) - Multi-attribute utility theory is used as the baseline compensatory framework to synthesize four heterogeneous risk attributes, Failure Impact (FI), Uncertainty Impact (UI), Technology Readiness Level (TRL), and Integration Readiness Level (IRL), into a single numerical priority score for each component. A compensatory model permits trade-offs among attributes: a component that is highly immature (low TRL) but presents only moderate failure consequences may still outrank a safer yet very mature component if the decision context emphasizes technological advancement.

The calculations were done for the four scenarios identified in Table 5. Each scenario is encoded in a weight vector $\mathbf{w}^{(\sigma)}$:

$$\mathbf{w}^{(\sigma)} = \left[w_{\mathrm{FI}}^{(\sigma)}, w_{\mathrm{UI}}^{(\sigma)}, w_{\mathrm{TRL}}^{(\sigma)}, w_{\mathrm{IRL}}^{(\sigma)} \right], \quad \sum_{j=1}^{4} w_j^{(\sigma)} = 1, \tag{2}$$

where σ indexes the scenario. Using explicit numerical weights rather than implicit preference statements ensures that subsequent sensitivity analysis can trace ranking changes directly to stakeholder emphasis shifts.

For each component i and scenario σ, the scenario-specific utility is computed as

$$U_i^{(\sigma)} = \sum_{j=1}^{4} w_j^{(\sigma)} \tilde{v}_{ij}. \tag{3}$$

where $0 \leq \tilde{v}_{ij}$ and $0 < w_j^{(\sigma)} \leq 1$, utilities are bounded in [0,1]. A component attains the maximal utility of 1 only if it is simultaneously the most critical on every attribute, given the scenario's priorities.

Ranking: Components are ordered by descending $U_i^{(\sigma)}$; ties are rare but, if present, indicate that the risk-priority under that scenario's weighting logic is indistinguishable.

Compensatory behavior: The linear form means that a shortfall in one attribute can be offset by excellence in another.

Scenario analysis: Comparing rankings across σ reveals how priority lists evolve when attention shifts.

Analytic Hierarchy Process (AHP) - The Analytic Hierarchy Process offers a second, more judgment-centred route for transforming the four risk attributes—Failure Impact, Uncertainty Impact, $\overline{\mathrm{TRL}}$, and $\overline{\mathrm{IRL}}$—into a single priority score. Whereas MAUT assumes the decision-maker can state attribute weights explicitly, AHP derives those weights from a small set of intuitive pairwise comparisons, and then checks their internal consistency.

1) Pairwise comparison matrix construction - For four criteria, a 4 x 4 reciprocal matrix is created

$$A = [a_{ij}], a_{ij} = \begin{cases} 1 & i = j, \\ \text{weights from scenarios} & i \neq j, \\ 1/a_{ji} & \text{reciprocal} \end{cases} \tag{4}$$

Each scenario defined earlier in Table 5 is translated into a coherent pairwise matrix by following the weights assigned to each metric in each scenario.

2) Deriving the priority vector - Once A is populated, weights are obtained as the principal-eigenvector of the matrix:

$$\mathbf{Aw} = \lambda_{\max}\,\mathbf{w}, \sum_{j=1}^{4} w_j \quad = 1. \tag{5}$$

Essentially, this means that each column is first normalized by its sum, then each row is averaged, and again the resulting vector is renormalized to unit sum.

3) Consistency Ratio (CR) - Because pair-wise judgements are seldom perfectly self-consistent, AHP provides an internal logic check called the Consistency Ratio. To calculate the CR, first, the largest eigenvalue of the comparison matrix is computed. If every judgement was logically exact, then this would equal to the matrix order n; any excess indicates contradiction. This excess is converted into the Consistency Index (CI)

$$CI = \frac{\lambda_{max} - n}{n - 1}. \tag{6}$$

Next, the observed CI is compared with Saaty's Random Index (RI) - a benchmark obtained from thousands of random reciprocal matrices of the same size [34]. For a 4 x 4 matrix RI(4) = 0.90. Finally the CR is calculated as

$$CR = \frac{CI}{RI(n)}. \tag{7}$$

CR, therefore, represents the percentage of inconsistency relative to what one would expect from an arbitrary set of judgments. A widely used rule of thumb accepts CR < 0.10 as sufficiently coherent. Importantly, a CR > 0.10 does not invalidate the weights outright; it merely flags potential contradictions that merit extra attention.

Best-Worst Method (BWM) - The Best–Worst Method (BWM) provides a third, judgment-based route for eliciting attribute weights. Introduced by Rezaei [35], BWM requires only ((2*n)-3) comparisons for n criteria instead of ((n*(n-1))/2) while still producing a full weight vector. In BWM, the step that converts comparisons to weights guarantees the lowest possible inconsistency given those comparisons; therefore, no post-hoc editing of ratings is needed.

1) To implement BWM, first, the best (most important) criterion B and the worst (least important) criterion W are identified. To keep the experimental design coherent with earlier sections, each scenario is aligned with a Best

criterion that embodies its strategic emphasis and a Worst criterion that represents the least-emphasized risk dimension. Below, for each of the four scenarios, the best and worst criterion are defined.

Balanced: B = $\overline{\text{TRL}}$ | W = UI

$\overline{\text{TRL}}$ serves as an appropriate Best anchor in Best-Worst Method analysis even when all weights are nominally equal due to its unique combination of universality, calibration ease, and foundational importance. TRL is defined on a well-established nine-level scale that most subject-matter experts can rank with high confidence, and anchoring on a criterion that everyone can score crisply minimizes inconsistency in the BWM optimization step, producing stable solutions even when other criteria are equally weighted. Beyond its scoring reliability, technology maturity functions as a gateway criterion since it represents a prerequisite for translating any prospective benefit, whether reduced uncertainty, low failure impact, or easy integration, into actual practice. Therefore, using TRL as the "most desirable" pole grounds the comparison in a necessary condition for success rather than favoring an outcome that sits further downstream in the development process. This approach also provides significant leverage on schedule and cost considerations, as it formalizes the common trade study instinct of first asking whether technology is ready enough to merit deeper scrutiny, ensuring that immature options receive appropriately lower composite scores once pairwise comparisons propagate through the model. UI serves as the least useful (Worst) anchor in a balanced scenario due to its high epistemic noise and redundancy with other metrics. UI is inherently subjective, as the same datapoint can appear to have either low or high impact depending on the underlying modeling assumptions used in the analysis. If UI were chosen as the best anchor, its noisier pairwise ratios would drive the entire optimization fit and could magnify expert disagreement, which is precisely what researchers want to suppress when every criterion is supposed to matter equally in the decision framework. Additionally, UI exhibits significant redundancy with other metrics commonly used in aerospace test-planning problems, where large portions of UI variance are already reflected indirectly through TRL (since immature technology is usually more uncertain) and Failure Impact (as high-stakes systems often have poorly characterized hazards). By treating UI as the Worst anchor, analysts can keep it in the model to maintain comprehensiveness without allowing it to double-count risks that other criteria already capture, thereby preventing distortion of the overall assessment while preserving the balanced weighting structure intended for the study.

Safety-First: B = FI | W = UI

FI is the best criterion is self-evident as it is paramount in safety-critical contexts, as the consequences of failure dominate decision-making.

UI is chosen as the worst anchor due to the fundamental distinction between direct and indirect risk mitigation approaches in safety-critical systems. Failure Impact quantifies the consequence side of risk, representing tangible harm that could occur, while uncertainty merely captures lack of knowledge about likelihood or

magnitude without addressing the underlying hazards themselves. In a safety-first posture, organizations mitigate consequences up front through proactive measures such as designing out catastrophic failure modes and adding redundancy, rather than spending equivalent resources on refining probability estimates that leave the fundamental risks unchanged. This prioritization becomes even more pronounced under time-critical decision cadences, where safety reviews often function as go/no-go gates operating on compressed schedules. Resolving epistemic uncertainty through additional testing might delay system fielding without proportionally reducing worst-case outcomes, and when deadlines loom, resources naturally flow toward hazard elimination rather than uncertainty characterization. Perhaps most importantly, UI might suffer from inverse leverage in safety-critical focus, where even large reductions in uncertainty rarely change the maximum credible loss for safety-critical hardware, since regulations typically demand planning for worst-case scenarios regardless of their perceived probability. This regulatory reality means that UI is logically the weakest anchor in a safety-first scenario assessment framework.

Maturity-Focused: B = $\overline{\text{TRL}}$ | W = UI

For a maturity-focused scenario, the best anchor candidates were $\overline{\text{TRL}}$ and $\overline{\text{IRL}}$. $\overline{\text{TRL}}$ is explicitly chosen over $\overline{\text{IRL}}$ due to three central reasons that establish TRL's foundational primacy in system development. First, TRL addresses foundational importance by assessing the maturity of the fundamental technology itself, evaluating basic feasibility and demonstration of functional performance, while IRL evaluates how well that technology integrates with the rest of the system, something that is inherently downstream from basic technology maturation. Without sufficient technology maturation as measured by TRL, integration readiness becomes irrelevant or premature, since organizations cannot meaningfully integrate a technology that isn't yet proven or stable at its core functional level. Second, TRL provides leverage, particularly in the stages targeted by maturity-focused scenarios approaching critical milestones such as design reviews or initial flight tests, where the risks are typically tied directly to raw technology immaturity rather than integration challenges. Resolving technology immaturity through TRL advancement usually removes the largest schedule roadblocks and provides more immediate and measurable benefits than addressing integration maturity, since fundamental technology gaps create cascading delays throughout the entire development timeline. Finally, empirical precedent from historical data and real-world experience in aerospace development programs consistently demonstrates that unresolved basic technology issues manifesting as low TRL cause delays more frequently and significantly than integration challenges reflected in low IRL [38–40]. Technology-level issues typically require fundamental rework, prototype re-design, or extensive research that results in substantial delays, while integration challenges, though significant, generally arise only after basic technology hurdles have been overcome, making TRL the naturally preferred priority when aiming to stabilize the critical path and ensure program success.

Integration path dependency becomes the dominant factor once a component reaches TRL $\geq$ 6, where remaining program risk shifts overwhelmingly toward interfacing details captured by IRL rather than raw uncertainty

about underlying physics or fundamental performance characteristics. At these advanced maturity stages, UI's marginal information value shrinks relative to IRL's practical importance. Budget trade-offs at late development stages further reinforce this prioritization, as money spent on additional uncertainty quantification yields minimal schedule relief compared with investments in integration tooling, harness design, or qualification testing that directly address the remaining barriers to system readiness. This approach finds strong support in empirical precedent from historical data gathered from hybrid-electric demonstrators such as NASA's X-57 and EPS-Alpha programs, which consistently show that once TRL climbs to higher levels, certification bottlenecks are seldom uncertainty-driven but rather emerge from interface and system-engineering challenges that require practical integration solutions [41–43]. The worst criterion position for UI, therefore, represents an evidence-based modeling choice that reflects real-world program dynamics and resource allocation priorities for a maturity-focused scenario.

Uncertainty-Reduction: B = UI | W = $\overline{\text{IRL}}$

In a scenario dedicated to uncertainty reduction, UI naturally becomes the most critical criterion as addressing uncertainty explicitly is the fundamental goal.

$\overline{\text{IRL}}$ is chosen as the worst criterion due to three main reasons. IRL's effect on posterior variance is fundamentally different from other criteria because integrability problems generally manifest as deterministic re-engineering tasks once identified, such as designing new brackets or rerouting harnesses, rather than widening epistemic-uncertainty bounds on system performance to the same extent that fundamental aerodynamic or electrochemical unknowns create. This distinction means that while integration challenges require engineering effort to resolve, they don't contribute to the fundamental uncertainty about whether a system will meet its performance specifications. The decoupling from modeling fidelity further supports IRL's relegation, as the dominant tools for uncertainty propagation treat integration assumptions as deterministic boundary conditions rather than stochastic variables. Finally, resource triage considerations in uncertainty-driven campaigns demonstrate that test-design money is preferentially spent on experiments that collapse probability distributions. Integration experiments typically occur later in the development timeline, when variance bounds have already been established through more fundamental testing, making the worst criterion designation for IRL a reflection of this established funding logic and development sequencing.

Scenario	Best criterion B	Worst criterion W
Balanced	$\overline{TRL}$	UI
Safety-First	FI	UI
Maturity-Focused	$\overline{TRL}$	UI
Uncertainty-Reduction	UI	$\overline{\text{IRL}}$

Table 14 Summary of Best and Worst Criterion Selections

Numerical value	Verbal meaning
1	equal importance
3	moderate importance
5	strong importance
7	very strong importance
9	extreme importance

Table 15 Saaty's 1-9 scale mapped to verbal meaning

2) Next, two short lists of pairwise ratings on Saaty's 1-9 verbal-numerical scale are expressed. Verbal ratings mapped to Saaty's 1-9 scale are given in the Table 15 below:

The two short lists expressed are:

- Best-to-Others vector $A_{B\rightarrow} = (a_{BFI}, a_{BUI}, a_{B\overline{TRL}}, a_{B\overline{IRL}})$, where a_{Bj} is "how many times more important is B than criterion j?"
- Others-to-Worst vector $A_{\rightarrow W} = ((a_{FIW}, a_{UIW}, a_{\overline{TRL}W}, a_{\overline{IRL}W})$, where a_{jW} is "how many times more important is criterion j than W?"

These two lists contain ((2*n)-3) which is 5 for n = 4 ratings (the entries 1 for B $\rightarrow$ B and W $\rightarrow$ W are fixed).

3) The next step is to derive numerical weights from the provided qualitative judgments. Once we have the two lists, we need a mathematical approach to translate them into a coherent and consistent set of attribute weights. To achieve this, an optimization problem is defined. The objective is straightforward: find attribute weights that closely align with what experts originally opined as per Saaty's scale in the lists. The optimization problem can be defined as:

$$\text{minimise } \xi$$

$$\text{subject to the constraints: } \left|\frac{w_B}{w_j} - a_{Bj}\right| \leq \xi, \quad \text{and} \quad \left|\frac{w_j}{w_W} - a_{jW}\right| \leq \xi, \quad \text{for each criterion } j. \tag{8}$$

Also, the sum of all the weights must equal 1, so as to easily interpret them as percentages

$$w_{FI} + w_{UI} + w_{\overline{TRL}} + w_{\overline{IRL}} = 1, \tag{9}$$

And each weight must be non-negative, as no criterion can have negative importance:

$$w_j \geq 0 \text{ for all criteria j.} \tag{10}$$

The variable ξ is the maximum absolute deviation between the weight ratios and the expert ratios.

4) The optimum ξ obtained from the previous step can be normalized into a Consistency Ratio (CR_{BWM}):

$$CR_{BWM} = \frac{\xi}{\xi_{max}(n)}, \tag{11}$$

where $\xi_{max}(n)$ is the largest deviation expected from a randomly generated judgement set of order n. Empirical benchmarks suggest $CR_{BWM} \leq 0.20$ indicates acceptable reliability, a lighter threshold than AHP as BWM already minimizes inconsistency at the optimization stage.

5) The final step is utility aggregation once the weight vector $w^{(\sigma)}$ is obtained for scenario σ. The utility model is given by:

$$U_i^{(\sigma)} = \sum_{j=1}^{4} w_j^{(\sigma)} \tilde{v}_{ij}. \tag{12}$$

Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) - The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multi-attribute decision-making approach that ranks alternatives by how closely they resemble an ideal solution, and conversely, how far they lie from a worst-case (anti-ideal) solution. TOPSIS explicitly leverages geometry to evaluate each alternative in a straightforward yet robust way. TOPSIS explicitly evaluates proximity both to the ideal (perfect) alternative and the anti-ideal (worst possible) alternative, making it sensitive to both strengths and weaknesses of each candidate. While still using attribute weights, TOPSIS outcomes are often more stable against small changes in weights, since it evaluates actual distances rather than just weighted averages.

The underlying logic of TOPSIS in brief is summarized before moving on to a step-by-step procedure of TOPSIS. Each alternative is described by multiple attributes (FI, UI, $\overline{TRL}$, $\overline{\text{IRL}}$). An ideal solution is an imaginary alternative defined by selecting the best (highest-priority) possible value of each attribute from among all considered alternatives. An anti-ideal solution is similarly defined, but using the worst (lowest-priority) values. Alternatives closest to the ideal and furthest from the anti-ideal are considered most preferable.

The step-by-step procedure of TOPSIS is given below: -

1) Construct the Decision Matrix: - We start by creating a decision matrix D with each row representing a different alternative (here, the aircraft components: Battery, Motor, Inverter, Thermal Management System), and each column representing an attribute (FI, UI, $\overline{\text{TRL}}$, $\overline{\text{IRL}}$). Each entry d_{ij} is the attribute score of alternative i on criterion j.

2) Normalize the Decision Matrix: - Since attributes have different measurement units and scales, we first normalise the decision matrix to standardise scales and ensure fair comparisons:

$$r_{ij} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^{m} (d_{ij})^2}}, \quad \text{for } i = 1, \ldots, m,\ j = 1, \ldots, n \tag{13}$$

This formula normalises values such that the squared sum of values in each column equals 1, making attributes

dimensionless and comparable.

3) Apply Attribute Weights: - Attribute weights are then applied to the normalized matrix. If attribute j has weigth w_j, then the weighted normalised value v_{ij} is calculated as:

$$v_{ij} = w_j.r_{ij}, \text{ for all i,j} \tag{14}$$

Applying weights reflects the relative importance that decision-makers assign to each attribute.

4) Identify the Ideal and Anti-Ideal Solutions: - We now identify the ideal solution (A^+) and the anti-ideal solution (A^-). The ideal solution (A^+) takes the best value of each attribute across all alternatives (highest value for attributes indicating urgency like FI, UI, $\overline{\text{TRL}}$, $\overline{\text{IRL}}$). The anti-ideal solution (A^-) takes the worst value of each attribute.

Formally,

$$A^+ = \{v_1^+, v_2^+, \ldots, v_n^+\}, \quad v_j^+ = \max_i(v_{ij}) \tag{15}$$

$$A^- = \{v_1^-, v_2^-, \ldots, v_n^-\}, \quad v_j^- = \min_i(v_{ij}) \tag{16}$$

5) Compute distances from Ideal and Anti-Ideal: - We calculate the Euclidean (geometric) distance of each alternative from the ideal solution (A^+) and the anti-ideal solution (A^-).

Distance to the ideal solution:

$$S_i^+ = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^+)^2}, \quad \text{for alternative } i \tag{17}$$

Distance to the anti-ideal solution:

$$S_i^- = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^-)^2}, \quad \text{for alternative } i \tag{18}$$

Each component can be seen as a point in the multidimensional space defined by attributes, and these distances show how close or far each component is from ideal and anti-ideal scenarios.

6) Calculate relative closeness to the ideal solution: - Finally, the relative closeness score (C_i^*) of each alternative to the ideal solution is calculated by:

$$C_i^* = \frac{S_i^-}{S_i^+ + S_i^-}, \quad 0 \le C_i^* \le 1 \tag{19}$$

A high relative closeness (close to 1) indicates the alternative is closer to the ideal and further from the anti-ideal,

thus more preferable. A low value (close to 0) means the opposite.

The final TOPSIS score (C_i^*) communicates which alternatives offer the best combination of desirable attributes, balancing both strengths and weaknesses. Alternatives with higher relative closeness to the ideal solution naturally become higher priority for testing and development resources. Because of the explicit geometric reasoning, stakeholders can intuitively visualise why one alternative is ranked higher: it's closer to an ideal state and further from a problematic state in the attribute space.

Elimination and Choice Expressing Reality (ELECTRE) III - ELECTRE III is the final multi-attribute decision-making method employed in this study. ELECTRE III is different from the rest in that it is non-compensatory: an extremely poor performance on any one attribute can block, or veto, an alternative, no matter how strong it is elsewhere. This property can be said to mirror some real-world certification logic, where a single show-stopper can invalidate an otherwise attractive component.

This method has some conceptual building blocks - indifference threshold, preference threshold, and veto threshold. Indifference threshold is a small difference on attribute j that decision makers agree is negligible. Preference threshold is a larger gap on attribute j that clearly favors one alternative. Veto threshold is an extreme gap on attribute j that can single-handedly block an alternative from outranking another. These three thresholds are defined once per attribute.

Now we see the step-by-step procedure for this method: -

1) Indifference threshold, preference threshold, and veto threshold are assigned to every attribute j.

Threshold	Symbol	Default fraction of column range*
Indifference	q_j	$q_j = 0.05$
Preference	p_j	$p_j = 0.15$
Veto	v_j	$v_j = 0.40$

Table 16 The values for the three thresholds assigned to each attribute

*With all attributes scaled 0-1, these fractions translate directly into absolute numbers (e.g., $q_j = 0.05$ means a difference within ±0.05 we consider scores indistinguishable).

2) For every ordered pair of alternatives (A,B) the following indices are computed.

- Concordance on each attribute

$$c_{AB}^{(j)} = \begin{cases} 1, & \tilde{v}_{Aj} \geq \tilde{v}_{Bj} - q_j \\ 0, & \tilde{v}_{Aj} \leq \tilde{v}_{Bj} - p_j \\ \frac{\tilde{v}_{Aj} - \tilde{v}_{Bj} + p_j}{p_j - q_j}, & \text{otherwise.} \end{cases} \tag{20}$$

This means that attribute j "votes" for component A if A is at least as bad as B within the indifference band; votes 0 if clearly better by the preference band and interpolates linearly in between.

- Global concordance

$$C_{AB} = \sum_{j=1}^{n} w_j c_{AB}^{(j)} (0 \le C_{AB} \le 1). \tag{21}$$

Higher C_{AB} means stronger overall evidence that A should be ranked ahead of B.

- Discordance (veto) tests for each attribute,

$$d_{AB}^{(j)} = \begin{cases} 0, & \tilde{v}_{Aj} \ge \tilde{v}_{Bj} - p_j \\ 1, & \tilde{v}_{Aj} \le \tilde{v}_{Bj} - v_j \\ \frac{\tilde{v}_{Bj} - \tilde{v}_{Aj} - p_j}{v_j - p_j}, & \text{otherwise.} \end{cases} \tag{22}$$

If any $d_{AB}^{(j)} = 1$, attribute j triggers a veto: A cannot outrank B regardless of the concordance tally.

- Credibility of outranking

$$\sigma_{AB} = \begin{cases} C_{AB}, & \text{no veto activated,} \\ C_{AB} \times \prod_{d_{AB}^{(j)} < C_{AB}} \left(\frac{1 - d_{AB}^{(j)}}{1 - C_{AB}} \right), & \text{otherwise.} \end{cases} \tag{23}$$

The penalised concordance σ_{AB}(0-1) is the final belief that A outranks B.

3) Ranking by descending distillation - After all the σ have been computed, an arrow is drawn from A to B if the σ $\ge$ chosen cut-off. Nodes with no incoming arrows at a given cut-off form the first stratum or highest priority. They are removed, and the cut-off is lowered. This is continued until all components are placed into strata.