Benchmark Problem for Autonomous Urban Air Mobility

Newton H. Campbell Jr.*, Irene M. Gregory†, Michael J. Acheson‡, Hari S. Ilangovan§, Shivakumar Ranganathan¶
NASA Langley Research Center, Hampton, VA, 23681, USA

This paper presents a comprehensive study of methods to evaluate advancement of autonomy and mission risk acceptability in Urban Air Mobility (UAM) systems. We introduce a Community Benchmark Problem (CBP) for Intelligent Contingency Management (ICM) in UAM. The paper underscores the integration of two pivotal frameworks: UL 4600, an American National Standards Institute standard for autonomous product safety evaluation, and the Capability Maturity Model Integration for Development (CMMI-DEV), a process improvement approach.

These methodologies bring a novel approach to UAM, with UL 4600 extending its technology-neutral stance to emphasize safety case construction, risk analysis, and autonomy validation of UAM systems. CMMI-DEV offers a structured, validated process for understanding maturity, thereby enhancing the quality and predictability of new UAM operations. Together, these frameworks allow us to establish a capability-maturity model specifically for ICM in UAM. This approach facilitates dynamic scoring and tracking within the UAM context, addressing the unique challenges and evolving nature of these systems. Illustrating the practical application of these methodologies, the paper applies the proposed measures to our Generic Urban Air Mobility simulation, serving as a model for future research and development in intelligent flight systems.

Nomenclature

AI Artificial Intelligence
CAR Causal Analysis and Resolution
CBP Community Benchmark Problem
CM Configuration Management
CMMI Capability Maturity Model Integration
CMMI-DEV Capability Maturity Model Integration for Development
DAR Decision Analysis and Resolution
eVTOL electric Vertical Takeoff and Landing
FAA Federal Aviation Administration
GUAM Generic Urban Air Mobility (UAM)
HMI Human-Machine Interface
ICM Intelligent Contingency Management
IPM Integrated Project Management
LaRC Langley Research Center
MA Measurement and Analysis
NASA National Aeronautics and Space Administration
PI Product Integration
PMC Project Monitoring and Control
PP Project Planning
PPQA Process and Product Quality Assurance
RD Requirements Development
REQM Requirement Management
RSKM Risk Management
SAM Supplier Agreement Management
SAT Boolean Satisfiability Problem
SG Specific Goal
SoS System of Systems
SP Specific Practice
SPI Safety Performance Indicator
TS Technical Solutions
TTT Transformational Tools and Technologies
UAM Urban Air Mobility
UCAT UAM Coordination Assessment Team
UML UAM Maturity Level

*SME III - Data Scientist, Everything Lives LLC, NASA Enterprise Applications Service Technologies.
†Senior Technologist for Advanced Control Theory and Application, irene.m.gregory@nasa.gov, Fellow AIAA.
‡Senior Research Engineer, Dynamic Systems and Control Branch, michael.j.acheson@nasa.gov, Member AIAA.
§Senior Data Scientist, Science Applications International Corporation (SAIC), NASA Enterprise Applications Service Technologies.
¶Senior Principal Solutions Architect, Science Applications International Corporation (SAIC), NASA Enterprise Applications Service Technologies.

I. Introduction

Urban Air Mobility (UAM) aims to revolutionize transportation by providing on-demand aviation services in urban environments. As UAM garners increased interest, one challenge stands prominent: the absence of standardized benchmarks for evaluating the performance, safety, handling, and flight quality of UAM vehicles and operations. Benchmark problems have played a pivotal role in the broader landscape of autonomy. The Defense Advanced Research Projects Agency (DARPA) Grand Challenge [1], Urban Challenge [2], and Subterranean Challenge [3], for example, have significantly advanced the autonomous land vehicles domain. The National Aeronautics and Space Administration (NASA) Urban Air Mobility Noise Working Group (UNWG) [4] has made strides by proposing noise prediction tools, metrics, and reference vehicle designs. Similarly, the RoboCup Soccer competition [5] set standards in multi-agent systems. Academia, industry, and government consistently refer to these benchmark problems when producing their own research, such that the community has a common, well-known set of constraints and metrics from which to compare.

Establishing UAM benchmarks is a complex task given the diversity in vehicle designs, propulsion systems, and flight patterns intrinsic to UAM. Yet, the importance of establishing benchmarks cannot be understated. UAM involves various vehicle types, propulsion systems, flight modes, mission profiles, operational environments, and user preferences. Also, noise, handling, and flying qualities are critical in determining the effectiveness and safety of operations, influencing factors such as vehicle stability, maneuverability, and passenger comfort. A benchmark for UAM needs to assess these aspects and pave the way for innovation, collaboration, and standardization – all crucial for UAM's success. These do not encompass the gamut of UAM scenarios, neglecting factors like vehicle installation effects, unsteady flight conditions, and varying user expectations. UAM is a dynamic and evolving field that requires constant adaptation and learning. Therefore, benchmark constraints and scoring metrics must be flexible enough to accommodate different situations and objectives, and rigorous enough to provide meaningful and reliable results.

This paper presents the following contributions to benchmarking in the domain of intelligent flight systems:

- Introduction of novel methodologies for assessing mission complexity and mission risk acceptability. This provides a structured approach to evaluating and enhancing Intelligent Contingency Management (ICM) capabilities for a UAM Community Benchmark Problem (CBP).
- Pioneering UL 4600 [6] and Capability Maturity Model Integration for Development (CMMI-DEV) [7] standards integration into UAM frameworks. This contribution enhances the safety and maturity models within intelligent flight systems and offers a versatile benchmark for comparing and improving intelligent flight solutions.
- Showcasing the practical application of these methodologies in real-world scenarios. This includes their use within both commonly used and specific UAM frameworks, providing valuable insights and practical guidance for the broader intelligent flight community.

We describe two methodologies that were developed for measuring the progress of developing mission complexity and mission risk acceptability into the simulation framework developed under the NASA Langley Research Center (LaRC) ICM for UAM sub-projects [8]. The methodologies were developed to measure the sub-project's progress during a yearly benchmark exercise. The framework for scoring is extendible and aligned with Boolean logic rules that enable the community to contribute to future tools. Finally, we describe how the constraints and scoring mechanisms were applied to a common, widely available UAM framework before describing its application to our own framework, the Generic UAM (GUAM) simulator.

II. Background

Autonomy is a crucial enabler for UAM, offering improvements in safety, efficiency, scalability, and affordability. However, it also introduces significant challenges in verification, validation, certification, and regulation [9]. Therefore, the measurement and comparison of autonomy levels in UAM vehicles and systems are essential. NASA's "Enabling Autonomous Flight and Operations in the National Airspace System (NAS) sub-project" is an initiative consolidating stakeholder perspectives to develop a national strategy for advanced autonomous operations [10]. This sub-project, through workshops, has identified core requirements and pathways for operationalizing increasingly autonomous systems, applicable to both UAM and small uncrewed aerial system (sUAS). NASA's Integrated Aviation Systems Program (IASP) is centered on the practical application and flight demonstrations of advanced aviation technologies [11], with its Integration of Automated Systems (IAS-1) flight tests focusing on the automation necessary for scalable UAM operations [12]. The UAM Operational Concept (OpsCon) document provides an in-depth overview of the UAM environment, discussing the vision, goals, operational practices, and performance metrics for UAM [13]. It reflects NASA's UAM Coordination Assessment Team (UCAT) efforts in documenting the development of UAM [14]. The

UAM Maturity Level (UML) scale, developed under UCAT, categorizes the evolutionary phases of UAM systems from their inception to full integration into everyday life [15]. Various stakeholders have informed this scale, focusing primarily on passenger-carrying scenarios, and considering different risk tolerances for other use cases. These initiatives are significant in the UAM domain. They address various challenges associated with autonomy, enhance human-machine interactions, and provide frameworks to assess the readiness of autonomous systems.

The problem of a benchmark evaluation methodology for UAM autonomy is a combination of the respective problems of evaluating System of Systems (SoS), evaluating autonomy algorithms, and evaluating Artificial Intelligence (AI). As of the time of this writing, there is no consensus on a framework for holistically evaluating any of these. The concept of SoS describes a superstructure of operationally independent systems targeting a unified goal [16, 17]. Similarly, evaluating autonomy algorithms and AI involves assessing the ability of uncrewed systems to perform complex tasks with minimal human intervention and adapt to changing environments [18, 19]. This is apparent in the Society of Automotive Engineers (SAE) scale [20] later adopted for UAM by Radovic [21]. However, there is also no universally accepted definition or measure of autonomy for uncrewed systems [22], and different methods have been proposed based on different levels of autonomy [23, 24].

Benchmark definitions for evaluating UAM Frameworks, essential in simulating and executing missions of varying complexities, should align with established operational standards. The specifics of UAM Corridor operations, as described in Federal Aviation Administration (FAA)-regulated guidelines [25], are in a state of evolution. FAA regulations will likely intersect with urban planning, particularly in establishing and managing vertiports and other essential infrastructure. As UAM is a nascent and rapidly evolving field, these concepts, guidelines, and regulations are subject to continuous refinement and development, influenced by ongoing research, technological advancements, and input from various stakeholders.

A. Safety and Maturity Modeling

UL 4600, the American National Standards Institute (ANSI) standard for the safety evaluation of autonomous products, is primarily designed for self-driving cars [6]. Its scope extends to applications in diverse fields like mining, agriculture, and maintenance, and includes lightweight Uncrewed Aerial Vehicles (UAVs), making it a versatile framework potentially applicable to UAM. The standard is founded on a claim-based approach, which does not prescribe specific technological solutions but instead focuses on creating a comprehensive safety case. UL 4600 addresses critical areas, including safety case construction, risk analysis, safety aspects of the design process, testing, tool qualification, autonomy validation, data integrity, and human-machine interaction. Importantly, it includes security as a requirement, does not delve into performance criteria or define pass/fail criteria for safety, and does not set acceptable risk levels or ethical product release requirements.

A UL 4600 Safety Case is a structured argument, supported by evidence, that a system is safe for a given application in a given environment. It typically includes:

1) **Claim:** A statement expressing a characteristic of the autonomous product or system relevant to safety.
2) **Argument:** A logical structure that links evidence to the claim (e.g., Performance targets, Safety Performance Indicators (SPIs), Methodologies).
3) **Evidence:** Data supporting the claim, including test results, analyses, and expert judgments.

The safety case aims to address potential risks, demonstrating adequate safety measures comprehensively. Claims can start out broad and be narrowed into sub-claims with a specified strategy, which also requires arguments and evidence.

CMMI-DEV is a process improvement approach that provides organizations with the essential elements for effective process improvement [7, 26]. Capability Maturity Model Integration (CMMI) has been recognized as one of the most renowned models in the software development industry since its inception in 1987. The model has undergone several adjustments over the years to maintain its relevance and applicability in evolving industry contexts. This continued evolution and adaptability have contributed to its sustained popularity in development. The Software Engineering Institute (SEI) has maintained data on the "time to move up" for organizations that have adopted CMMI. For example, since the release of CMMI, the median time for organizations to move from CMMI Maturity Level 1 to Level 2 is around five months, with a median movement to CMMI Maturity Level 3 taking an additional 21 months [27]. The CMMI-DEV model provides a comprehensive framework for evaluating and enhancing process maturity, focusing on improving the quality and predictability of processes to yield higher-quality products or services. Its structured levels of process maturity range from initial (ad hoc, chaotic processes) to optimized (continuous process improvement).

In the context of UAM Framework verification, validation, and testing, CMMI-DEV offers a robust framework for establishing and improving processes that ensure the reliability and safety of systems. It emphasizes process areas like

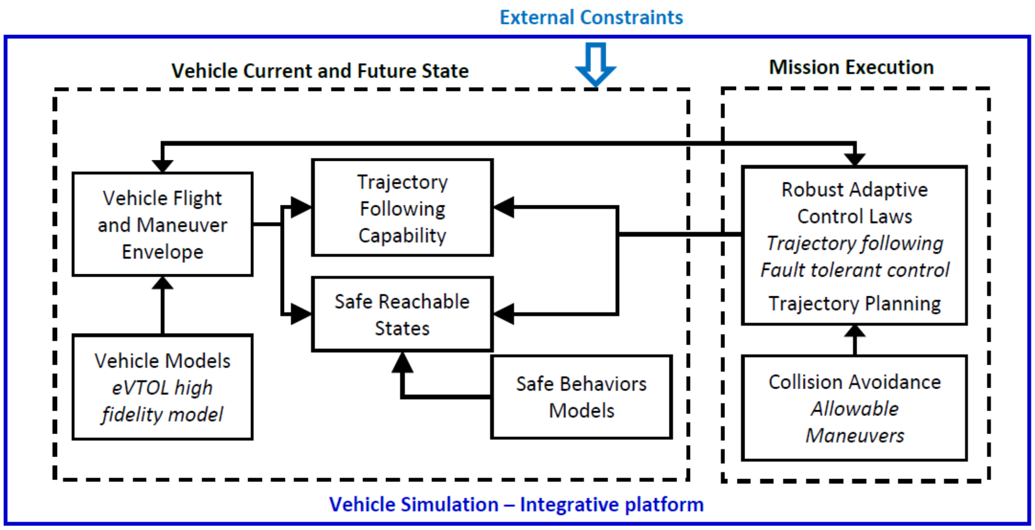


Fig. 1 ICM GUAM Architecture

requirements management, technical solutions, product integration, verification, validation, and decision analysis and resolution. Most organizations deploying any practical level of UAM use the practices and processes of CMMI-DEV, particularly for software development [28], to some extent. The latest CMMI-DEV techniques and methods provide a comprehensive framework for executing the described steps in a safety case evaluation [29, 30]. Introducing new domains such as Data Management, People Management, and Virtual Work allows for a more nuanced classification and evaluation of evidence. These domains, especially Data Management, align well with categorizing evidence like experimental UAM Framework data and analytics. The model's focus on continuous improvement and outcomes-based performance, including updates in Agile and DevSecOps, supports the mapping and evaluation of evidence against CMMI-DEV levels. Section III.B describes relevant aspects of the CMMI-DEV process areas in more detail.

B. ICM for UAM

The NASA's Transformational Tools and Technologies (TTT) Autonomous Systems (AS) Project has been instrumental in the UAM space. The project focuses on the transition to greater levels of autonomy for new air transportation modes such as UAM. It identifies both the technological gaps and the challenges related to human-autonomy interactions. As a sub-project of TTT, the ICM sub-project [31] seeks to promote principles and experimentation that can address some of these challenges. The sub-project's integration of AI [32] with high-quality autonomy algorithms [33–35] and high-fidelity modeling [36–38] has produced the necessary knowledge to propose measurements that address community concerns. The ICM research team is conducting theoretical and experimental studies on various aspects of ICM, such as data uncertainty quantification, vehicle health management, trajectory optimization, contingency detection and diagnosis, mission replanning and reconfiguration, human-machine interaction, and verification and validation. The team has developed a UAM simulation framework that integrates various tools and models to test and evaluate ICM concepts and algorithms in realistic scenarios. The public version of this simulation framework, GUAM, permits the research community to introduce algorithms at the intersection of UAM, SoS thinking, and AI.

GUAM's architecture, shown in Figure 1, aims to facilitate research into more intelligent, adaptive UAM strategies. It has two primary components: Vehicle Current and Future State, and Mission Execution. The former focuses on monitoring and projecting the vehicle's internal and external states, employing sensor fusion and trajectory prediction to assess vehicle health and situation awareness. The latter deals with planning and executing missions, incorporating adaptive control laws, onboard trajectory planning, and collision avoidance to ensure safety and efficiency, even in the presence of uncertainties and faults [39, 40]. A critical element in GUAM's success is the modeling of vehicles, particularly the Lift-Plus-Cruise (L+C) configuration, a type of electric Vertical Takeoff and Landing (eVTOL) design [41]. This is the primary vehicle for ICM testing, under GUAM, although other UAM-type vehicles are available. The modeling approach integrates computational methods with real-time aerodynamic monitoring, using machine learning

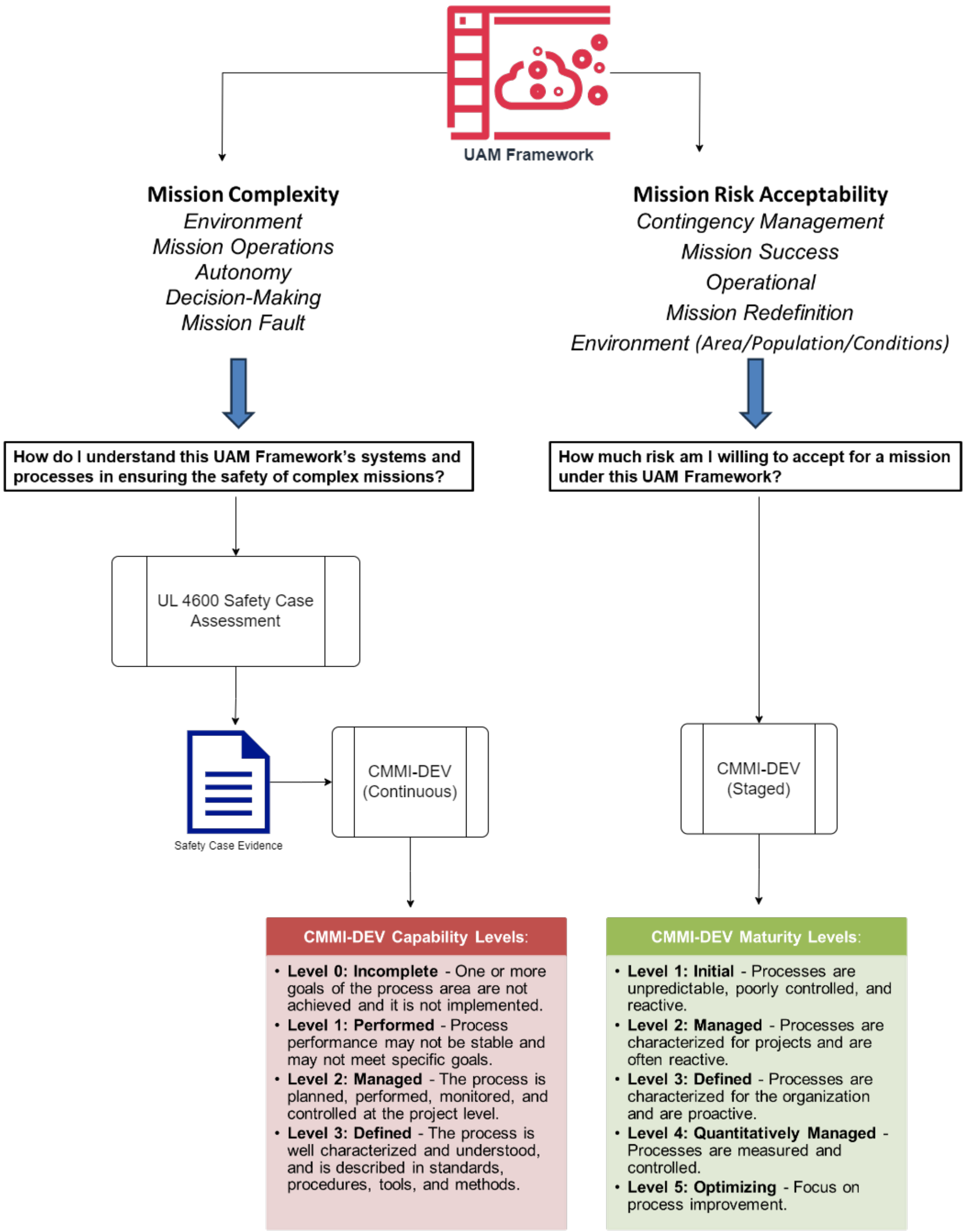


Fig. 2 Methodology for Scoring the UAM for ICM Community Benchmark Problem

for system identification and adjustments to dynamic aerodynamic changes. The progressive development of GUAM is focused on expanding the coverage of contingencies, including unforeseen situations, while reducing reliance on rigid rule-based systems. The architecture permits continuous integration of machine learning techniques, including multi-agent reinforcement learning and neural networks.

III. Methodology

We establish a maturity model for UAM autonomy by leveraging two frameworks: UL 4600 and CMMI-DEV. This model allows us to track both the complexity with which a UAM Framework can apply its underlying autonomy to missions and the level of risk that can be accepted for missions. We denote these concepts as Mission Complexity and

Mission Risk Acceptability, respectively. The purpose of the methodology, illustrated in Figure 2, is to leverage our capability-maturity model to compute and track these concepts over time.

The Mission Complexity and Risk Acceptability of a UAM Framework are evaluated across several domains. UL 4600 Safety Cases are developed for each domain to assess the evidence supporting that domain's complexity. The process includes:

1) Identify Evidence Categories: Classify the evidence within each safety case with labels such as experimental data and analytics.
2) Map Evidence to Capability Levels: Link each evidence piece to an appropriate CMMI-DEV Capability level.
3) Evaluate Maturity in Handling Risk: Assess each category's maturity concerning the organization against the CMMI-DEV Maturity model, focusing on aspects like documentation completeness and process effectiveness.
4) Aggregate Capability Level Scores: Calculate an overall capability level for each safety case by combining individual evidence scores, considering their importance within the UAM framework.
5) Periodic Reassessment: Regularly reassess evidence capability and organizational maturity through benchmark exercises to stay aligned with technological and operational updates in UAM frameworks.

The evidence for each sub-criterion within these domains undergoes a CMMI-DEV continuous capability level assessment. The capability level, denoted as $C_i(t)$, is a function of time t, reflecting the evolving nature of the evidence supporting each domain of Mission Complexity.

Mission Risk Acceptability directly applies CMMI-DEV to aspects of the project, rather than filtering through the lens of UL 4600 safety case evidence. This results in a time-dependent score $R_j(t)$, the ratio of Specific Goals (SGs) met to total defined SGs for each domain j. The methodology emphasizes dynamic scoring and tracking for both Mission Complexity and Mission Risk Acceptability:

$$C_i(t) = \text{CMMI-DEV capability level for UL 4600 evidence in Complexity domain } i \tag{1}$$

$$R_j(t) = \text{\% SGs at CMMI-DEV maturity level for Risk Acceptability domain } j \tag{2}$$

Growing the UAM Framework to the point where it is "safe enough," is integrated with the development of a UL 4600 safety case. We establish evidence, performance targets, and SPIs for the safety cases in alignment primarily with the FAA Concept of Operations for UAM [25] and the NASA UML [15]. With the safety cases established, we then use first-order logic to formalize the methodology and track the evolution of the UAM Framework's complexity and safety capabilities. Using this approach, we craft a maturity model that describes how each maturity level is characterized by mission risk acceptability, and capability level characterizes mission complexity.

A. Measuring Complexity

Mission Complexity in a UAM Framework addresses key aspects of safety and capability maturity across several domains:

- **Complexity of Environment:** Assesses navigational adaptability in varying urban airspaces.
- **Complexity of Mission Operations:** Evaluates the framework's capability in mission planning and execution, including standard and emergency operations.
- **Complexity of Autonomy:** Measures the level of system automation, from minimal to advanced AI, impacting the operator's role
- **Complexity of Decision-Making:** Assesses the framework's capacity for strategic and responsive decision-making at various operational levels.
- **Complexity of Mission Fault:** Explores the framework's approach to identifying and addressing potential mission faults.

We describe the UL 4600 Safety Case, comprised of claims that the UAM Framework can perform and adapt to these levels of complexity in the next subsections*. The relationship between claims and evidence are described by formal Boolean Satisfiability Problem (SAT) descriptions [42], where ϵ_x describes the existence and validation of x, which describes measurements, metrics, or classes of algorithms that need to be implemented, tested, and continuously or iteratively verified to justify a sub-claim.

*NASA will publish the UL 4600 safety cases in full upon review. In this paper, we include brief summaries of each segment.

1. Complexity of Environment

Examples of quantitative measurements x for evidence ϵ_x to justify a claim about Complexity of Environment include Dynamic Density [43], Temporal and Spatial metrics, Airspace Saturation metrics and Weather Model Complexity [44], and Terrain Complexity Index (TCI) [45], with research utilizing Light Detection and Ranging (LIDAR) data to analyze terrain complexity, including factors like elevation range and obstacle density, and developing trajectory planning algorithms for emergency landings. Our study developed eight UL 4600 Safety Cases guided by such measurements to shape evidence, performance targets, and SPIs. The sub-claims are combinations of environmental complexity characteristics: Structured/Unstructured, Known/Unknown, and Static/Dynamic.

- **Structured Environments**: Characterized by well-defined patterns, enhancing predictability.
- **Unstructured Environments**: Lack regular patterns, presenting chaotic conditions.
- **Known Environments**: Well-explored and familiar terrains.
- **Unknown Environments**: Comprising uncharted or unfamiliar elements.
- **Static Environments**: Exhibiting minimal changes over time.
- **Dynamic Environments**: Undergoing rapid and constant changes.

Our methodology involves assigning each safety case to a three-dimensional category (e.g., Structured, Unknown, Dynamic), with each case featuring tailored evidence claims, evidence, and performance targets that reflect the specific environmental complexity. This approach comprehensively assesses the safety of the UAM Framework across diverse operational scenarios. These complexity assignments necessitate distinct safety strategies, ensuring the UAM Framework aligns with the varying operational conditions per UL 4600 standards. Formal SAT descriptions align the UAM Framework with each unique environmental assignment, based on the evidence described for each sub-claim in the UL 4600 Safety Cases:

$$\begin{aligned}&(\textbf{Structured} \wedge \textbf{Known} \wedge \textbf{Static})\\ &\wedge(\epsilon_{\text{VVR}} \wedge \epsilon_{\text{VOC}} \wedge \epsilon_{\text{ACM}})\\ &\wedge(\epsilon_{\text{CAC}} \wedge \epsilon_{\text{TCI}}) \wedge (\epsilon_{\text{TCIc}} \wedge \epsilon_{\text{DD}})\end{aligned} \tag{3}$$

$$\begin{aligned}&(\textbf{Structured} \wedge \textbf{Known} \wedge \textbf{Dynamic})\\ &\wedge(\epsilon_{\text{VVR}} \wedge \epsilon_{\text{VOC}} \wedge \epsilon_{\text{ACM}})\\ &\wedge(\epsilon_{\text{CAC}} \wedge \epsilon_{\text{TCI}}) \wedge (\epsilon_{\text{RA}} \wedge \epsilon_{\text{TPP}})\end{aligned} \tag{4}$$

$$\begin{aligned}&(\textbf{Structured} \wedge \textbf{Unknown} \wedge \textbf{Static})\\ &\wedge(\epsilon_{\text{VVR}} \wedge \epsilon_{\text{VOC}} \wedge \epsilon_{\text{ACM}})\\ &\wedge(\epsilon_{\text{NSA}} \wedge \epsilon_{\text{UCS}}) \wedge (\epsilon_{\text{TCIc}} \wedge \epsilon_{\text{DD}})\end{aligned} \tag{5}$$

$$\begin{aligned}&(\textbf{Structured} \wedge \textbf{Unknown} \wedge \textbf{Dynamic})\\ &\wedge(\epsilon_{\text{VVR}} \wedge \epsilon_{\text{VOC}} \wedge \epsilon_{\text{ACM}})\\ &\wedge(\epsilon_{\text{EMUE}} \wedge \epsilon_{\text{ADNC}}) \wedge (\epsilon_{\text{RTWU}} \wedge \epsilon_{\text{TPP}})\end{aligned} \tag{6}$$

$$\begin{aligned}&(\textbf{Unstructured} \wedge \textbf{Known} \wedge \textbf{Static})\\ &\wedge(\epsilon_{\text{ABM}} \wedge \epsilon_{\text{SMP}})\\ &\wedge(\epsilon_{\text{CAC}} \wedge \epsilon_{\text{TCI}}) \wedge (\epsilon_{\text{TCIc}} \wedge \epsilon_{\text{IR}})\end{aligned} \tag{7}$$

$$\begin{aligned}&(\textbf{Unstructured} \wedge \textbf{Known} \wedge \textbf{Dynamic})\\ &\wedge(\epsilon_{\text{TEMP\_PROB}} \vee \epsilon_{\text{MDP}})\\ &\wedge(\epsilon_{\text{CAC}} \wedge \epsilon_{\text{TCI}}) \wedge (\epsilon_{\text{RTWU}} \wedge \epsilon_{\text{TPP}})\end{aligned} \tag{8}$$

$$\begin{aligned}&(\textbf{Unstructured} \wedge \textbf{Unknown} \wedge \textbf{Static})\\ &\wedge(\epsilon_{\text{EA}} \wedge \epsilon_{\text{NSN}})\\ &\wedge(\epsilon_{\text{PDA}} \wedge \epsilon_{\text{NAT}}) \wedge (\epsilon_{\text{TCIc}} \wedge \epsilon_{\text{DD}})\end{aligned} \tag{9}$$

$$\begin{aligned}&(\textbf{Unstructured} \wedge \textbf{Unknown} \wedge \textbf{Dynamic})\\ &\wedge(\epsilon_{\text{MLA}} \vee \epsilon_{\text{EAA}})\\ &\wedge(\epsilon_{\text{DYN\_IN}} \vee \epsilon_{\text{EMUE}}) \wedge (\epsilon_{\text{RTWU}} \wedge \epsilon_{\text{TPP}})\end{aligned} \tag{10}$$

Where

$\epsilon_{\textbf{ABM}}$ = High Modeling Accuracy for Agent-Based Models.
$\epsilon_{\textbf{ACM}}$ = Sufficient Airspace Traffic Complexity based on chosen metrics(e.g., [46]).
$\epsilon_{\textbf{ADNC}}$ = Adaptation to Novel Conditions Timeframe.
$\epsilon_{\textbf{CAC}}$ = Certification of Aircraft Compliance.
$\epsilon_{\textbf{DD}}$ = Sufficient Dynamic Density (or equivalent metric).
$\epsilon_{\textbf{DYN\_IN}}$ = Application of Inference Models for Dynamic Responses to Unknown Conditions (e.g., [47]).
$\epsilon_{\textbf{EA}}$ = High Exploration Algorithm Success Rate.
$\epsilon_{\textbf{EAA}}$ = High Evolutionary Algorithm Optimization Effectiveness.
$\epsilon_{\textbf{EMUE}}$ = Efficient Management Rate of Unpredictable Elements.
$\epsilon_{\textbf{IR}}$ = Low Incident Rate.

$\epsilon_{\mathbf{MDP}}$ = High Markov Decision Processes Success Rate.
$\epsilon_{\mathbf{MLA}}$ = High Machine Learning Algorithm Success Rate in adapting to unknown scenarios.
$\epsilon_{\mathbf{NAT}}$ = High Neural Network Accuracy for Traffic Prediction.
$\epsilon_{\mathbf{NSA}}$ = Novel Scenario Adaptation.
$\epsilon_{\mathbf{NSN}}$ = Navigation Success in Unknown Static Environments.
$\epsilon_{\mathbf{PDA}}$ = High Performance Data Accuracy.
$\epsilon_{\mathbf{RA}}$ = Validation of Rapid Environmental Adaptation.
$\epsilon_{\mathbf{RTWU}}$ = Validation of Real-Time Weather Updates.
$\epsilon_{\mathbf{SMP}}$ = Stochastic Model Predictions Alignment with real-world data.
$\epsilon_{\mathbf{TCI}}$ = Low Terrain Complexity Index Discrepancy.
$\epsilon_{\mathbf{TCIc}}$ = Low TCI Conformance Deviation.
$\epsilon_{\mathbf{TEMP\_PROB}}$ = High Temporal Probabilistic Model Accuracy for dynamic changes in air traffic (e.g., [48]).
$\epsilon_{\mathbf{TPP}}$ = Low Traffic Pattern Prediction Error Rate.
$\epsilon_{\mathbf{UCS}}$ = High Flight Stability measurements In Unforeseen conditions.
$\epsilon_{\mathbf{VOC}}$ = High Vertiport Operations Compliance.
$\epsilon_{\mathbf{VVR}}$ = Verification and Validation of Records Matching FAA or Experimental Environment Structure.

These descriptions form an accompanying knowledge base for these and other UL 4600 Safety Case claims and sub-claims. The UAM Framework should be assigned to a combination of environmental characteristics by identifying the above satisfiable equations. If the UAM Framework under assessment is not mature enough such that the quantitative score or metric relevant to a SAT clause cannot be reasonably computed, then the corresponding variable for that threshold is regarded as False. Note the emphasis on artificial intelligence and evolutionary algorithms in the last two categories. This is a recognition that we cannot meet these levels of autonomously handling uncertain environments without such capabilities. The remainder of the breakdowns for the UL 4600 Safety Claims will be described more briefly in subsequent sections with definitions in the Appendix.

2. Complexity of Mission Operations

The following categories describe the Complexity categories of the UAM Framework to facilitate Mission Operations:

- **Simple Flight Plan:** Direct or minimally routed trajectories, requiring basic navigation and minimal contingency planning.
- **Complex Flight Plan:** Intricate routing with multiple waypoints, requiring dynamic rerouting as needed.
- **Simple Flight Tasks:** Basic maneuvers like takeoff, cruise, and landing.
- **Complex Flight Tasks:** Sophisticated maneuvers requiring precise control in challenging environments and complex system interactions.
- **Normal Operations:** Routine flights within standard parameters.
- **Abnormal Operations:** Flights under challenging conditions, demanding enhanced skills.
- **Recoverable Failures:** Scenarios where the system maintains mission continuity despite faults.
- **Unrecoverable Failures:** Critical failures necessitating comprehensive emergency response protocols.
- **Fundamental Handling and Flight Quality:** Focuses on ensuring optimal handling and flight quality in basic operational scenarios, emphasizing passenger comfort and cargo security.
- **Advanced Handling and Flight Quality:** Maintains superior handling and flight quality under complex operational scenarios, focusing on enhancing passenger experience and safeguarding cargo integrity.

Unrecoverable faults must accurately reflect the cascading effects and system-wide impacts that would lead to a mission's failure, demanding a deeper integration with the simulation's logic and possibly requiring more advanced algorithms to model these scenarios realistically. While still challenging, recoverable faults are generally less complex, have more localized effects, and require the simulation to model and execute less complex recovery protocols that return the system to normal operations without the mission's failure. Table 4 in the Appendix describes the corresponding SAT sub-claims for Complexity of Mission Operations. This complexity category captures the UAM Framework's ability to manage unanticipated environmental variables and operational anomalies, also testing AI's capability to maintain operational integrity and execute contingency protocols under diverse conditions. In the case of fault injection, the safety cases assess the framework's ability to simulate, identify, and respond to faults, ranging from recoverable to catastrophic scenarios. This assessment provides insights into the AI's diagnostic and prognostic capabilities, response strategies to system failures, and recovery mechanisms. When failure modes can have significant consequences, this evaluation is pivotal in validating the AI's competence in risk assessment, failure impact analysis, and emergency management. Including both levels of Handling and Flight Quality requires the development of Mission Task Elements [49]. This includes considering and testing realistic flight missions, which is increasingly important in the context of passenger-centric UAM services and sensitive cargo operations.

3. Complexity of Autonomy

Using automotive autonomy standards [20] and NASA's UML [15] as guides, we delineate eight distinct levels of Complexity of Autonomy for the CBP. For the CBP, **assisted flight** refers to levels where the human operator retains significant control or oversight of the operations. The following categories fall under this classification:

- **Manual Control (Level 1)**: At this level, all flight operations are manually conducted/commanded by the operator, including navigation and stabilization.
- **Flight Stability (Level 2)**: Introduces basic automated systems, leveraging Inertial Measurement Units (IMUs) (e.g., gyroscopes and accelerometers), to assist the operator in maintaining flight stability.
- **Envelope Protection (Level 3)**: Incorporates advanced safety mechanisms to prevent the aircraft from entering unsafe flight conditions, with the operator overseeing these systems.
- **Navigation and Collision Avoidance (Level 4)**: Automated systems guide the aircraft along predetermined routes and assist in avoiding obstacles while the operator monitors and provides strategic courses of action.

Autonomous flight encompasses levels where the aircraft can perform operations with minimal to no human intervention. These levels include:

- **Conditional Automation (Level 5)**: The vehicle(s) can conduct entire flight operations autonomously under specific conditions, but requires operator intervention when conditions are not met.
- **Conditional Automation with AI (Level 6)**: Enhances conditional automation with AI capabilities, where the operator supervises AI systems and intervenes with alternative courses of action as necessary.
- **High Automation (Level 7)**: Signifies a near-complete shift to autonomy, with the operator's role primarily focused on system monitoring and providing courses of action in exceptional circumstances.
- **Full Automation (Level 8)**: The pinnacle of autonomous UAM Frameworks, where the vehicle operates autonomously in any scenario without the need for human oversight, and requires only high-level requirements from operators.

This framework, described in further detail in Tables 5 and 6 in the Appendix, describes characteristics that flow toward autonomy, necessitating a comprehensive understanding of the various levels of automation and their implications on operational dynamics. The progression from assisted to autonomous flight in UAM Frameworks should reflect a significant technological and operational shift. Implementing automated systems requires rigorous testing and validation to ensure they perform reliably under various conditions. The systems must be sensitive and responsive to environmental changes, yet robust enough to maintain stability without excessive interference in the operator's control. This balance is critical for safety and operational efficiency, representing a step forward in the journey towards higher levels of autonomy.

4. Complexity of Decision-Making

The complexity of decision-making within a UAM Framework is integral in assessing its capability for autonomous flight management. This complexity is defined by the framework's ability to make informed, autonomous decisions across various operational aspects. The sophistication of these decisions ranges from immediate operational responses to anticipatory, strategic planning for long-term mission effectiveness. The following categories detail the complexity levels at which the UAM Framework manages decision-making:

- **Mission-Level Decision-Making**: Involves strategic decisions based on overarching mission objectives and constraints. Advanced levels involve optimizing mission goals across a series of missions.
- **Task-Level Decision-Making**: Focuses on decisions related to the sequencing and prioritization of specific mission tasks. A mature framework dynamically adjusts task sequencing for optimal mission flow and efficiency.
- **Plan-Level Decision-Making**: Pertains to planning decisions for each task with adaptability, considering changing resources and conditions. Higher maturity allows for dynamic planning adjustments in response to evolving scenarios.
- **Maneuver-Level Decision-Making**: Involves decisions on maneuver selection for plan execution. An advanced framework can proactively optimize mission performance through strategic maneuver selection.
- **Control-Level Decision-Making**: Concerns decisions about executing maneuvers through control inputs. At higher maturity levels, the framework can enhance precision and efficiency in control execution.
- **Health-Level Decision-Making**: Centers on decisions based on the health status of the aircraft and subsystems. An evolved framework anticipates and mitigates future risks, extending operational life.
- **Fault-Level Decision-Making**: Involves detecting and preemptively addressing potential faults and failures.
- **Recovery-Level Decision-Making**: Relates to decisions for initiating recovery actions post-fault or failure. Higher

maturity involves comprehensive recovery strategies to minimize mission disruption.

These decision-making areas are quantitatively evaluated within the UAM Framework to ensure alignment with safety and efficiency criteria in the corresponding UL 4600 Safety Cases, described in Table 7 in the Appendix. The framework's capability to handle various levels of decision-making complexity is pivotal, as rapid, accurate, and foresighted decision-making is critical for safety and efficiency. Notably, it is common for a UAM Framework to operate concurrently at multiple, non-successive decision-making categories. This also reflects the framework's evolving autonomy, ensuring competence in managing complexities at each decision-making stage.

5. Complexity of Mission Fault

Managing mission faults is critical to ensuring safety and reliability. We classify mission faults into various categories based on their nature and correctability, drawing from principles outlined in safety-critical systems research [50]. The following categories shape the scores and UL 4600 safety case sub-claims for this category:

- **Expected Faults**: Anticipated based on historical data or predicted scenarios. Scored on predictive analytics and historical data analysis.
- **Unexpected Faults**: Occur without warning. Scored on real-time diagnostics and response agility.
- **External Faults**: Resulting from environmental or external operational factors. Scored on environmental awareness and interaction with external systems like air traffic control.
- **Internal Faults**: Originating from within the vehicle, such as mechanical or software issues. Scored on internal monitoring, self-diagnostics, and fail-safe mechanisms.
- **Correctable Faults**: Resolvable during or before the next mission. Scored on adaptive mission planning and in-mission correction efficiency.
- **Uncorrectable Faults**: Severe, mission-critical faults. Scored on emergency response protocols and minimization of mission disruption and safety compromise.

The UAM Framework employs a three-dimensional scale for evaluating Mission Fault Complexity. This scale reflects the multifaceted nature of operational challenges, drawing parallels with aviation's Threat and Error Management (TEM) practices [51]. Expected faults, predictable through historical data and predictive models, generally include routine maintenance or known software issues. Unexpected faults, however, such as sudden system failures, require a robust response due to their unforeseen nature and are heavily influenced by AI uncertainty [52]. The spectrum from External to Internal faults covers the fault origin, with external faults being environmental challenges like adverse weather, and internal faults being system-intrinsic issues like hardware malfunctions. Finally, the Correctable vs. Uncorrectable fault dichotomy is vital for operational and emergency planning; correctable faults permit in-mission adjustments, while uncorrectable faults often lead to mission aborts or emergency responses. An example of a correctable fault is a temporary GPS signal loss, whereas an uncorrectable fault might be a critical failure in the propulsion system.

B. Measuring Risk Acceptability

In this section, we outline the Mission Risk Acceptability components of the benchmark, crucial for ensuring UAM safety, efficiency, and public acceptance. Determining risk acceptability involves quantitative and qualitative approaches, including numerical risk estimates and value judgments. Five risk acceptability categories are considered: *Contingency Management*, *Mission Success*, *Operational*, *Mission Redefinition*, and *Environment*. Each is assessed using CMMI-DEV Staged assessment processes, with scoring aligned to CMMI-DEV Maturity levels as shown in Table 1. Relevant CMMI-DEV process areas for each category are detailed in Table 2.

Contingency Management concerning Mission Risk Acceptability describes how a system is prepared to handle unexpected events or emergencies, reflecting its readiness and adaptability in crisis situations. This is crucial for responding effectively to unexpected events in the dynamic urban airspace, encompassing challenges like adverse weather, equipment malfunctions, or airspace congestion. UAM systems require robust contingency plans, including alternative routes and emergency landing options, to ensure safety and mission completion. Inadequate management in this area can lead to significant mission delays or accidents. This involves crucial CMMI-DEV processes such as Risk Management (RSKM) for risk mitigation, Configuration Management (CM) for tracking and managing system changes, and Project Monitoring and Control (PMC) for continuous oversight of UAM projects, ensuring adaptability and response to sudden changes in conditions and effective integration of new components.

Effective *Environmental* risk management in this context requires a comprehensive understanding and modeling of these varied environments to assess their impact on operations. Mitigation strategies must factor in traffic congestion, obstacle avoidance, and communication systems for safe and efficient operations. Key CMMI-DEV Process Areas vital

Table 1 Mission Risk Acceptability Category per CMMI Maturity Level

CMMI Maturity	*Contingency Management*	*Mission Success*	*Operations*	
Level 1: Initial	Inadequate	Vague	Reactive	
Level 2: Managed	Feasible	Post-Analysis	Basic Monitoring	
Level 3: Defined	Well-Planned	Scenario-Based	Systematic	
Level 4: Quantitatively Managed	Human Comparable	Conditional	Predictive	
Level 5: Optimizing	Fully Explainable	Attainable	Adaptive and Innovative	
CMMI Maturity	*Mission Redefinition*	*Environment (Area/Population/Weather Resilience)*		
Level 1: Initial	Rigid	Deserted	None	None
Level 2: Managed	Algorithm Modifiable	Rural	Low	Low
Level 3: Defined	Human Modifiable	Suburban	Low-Medium	Medium
Level 4: Quantitatively Managed	Human+AI Modifiable	Urban	Medium	High
Level 5: Optimizing	AI-Modifiable	Emergency	Extremely High	Extremely High

Table 2 Most relevant CMMI-DEV Process Areas for Mission Risk Acceptability Categories

Mission Risk Acceptability Category	**CMMI-DEV Process Area**	**Maturity Level**	**Significance**
Contingency Management	CM PMC	Level 2	Ensures robust contingency plans and risk mitigation strategies for effective response to unexpected events.
	RSKM	Level 3	
Environment	MA REQM	Level 2	Involves managing the impact of environmental factors on UAM operations.
	PI RD TS	Level 3	
	CAR	Level 5	
Mission Redefinition	DAR IPM RSKM	Level 3	Enables flexible adaptation and redefinition of mission parameters in changing conditions.
Mission Success	PMC PP REQM	Level 2	Focuses on managing mission requirements and project planning to achieve desired outcomes.
	RSKM	Level 3	
Operations	CM MA PPQA SAM	Level 2	Focuses on managing mission operational challenges, maintaining safety and flight quality standards.

for Environment Risk Acceptability include Requirements Development (RD) and Requirement Management (REQM), which ensure systems are designed and operated with these environmental factors in mind. This could entail developing specific operational requirements tailored to different environments, such as noise reduction measures or enhanced navigation systems. Measurement and Analysis (MA) plays a crucial role in continuously assessing the operating environment's impact on UAM operations, involving data analysis on traffic patterns, weather, and airspace constraints. Causal Analysis and Resolution (CAR) focuses on maintaining services effectively across diverse conditions, planning for system redundancy and robust operation protocols. Additionally, Technical Solutions (TS) and Product Integration (PI) enable agility in responding to environmental risks, with TS developing adaptable technical solutions and PI ensuring their seamless integration into the UAM system, like advanced weather prediction tools or adaptable anti-collision systems.

Mission Redefinition addresses the need for flexibility and adaptability in response to changing conditions or goals in urban environments, like unexpected airspace restrictions or emergencies. This requires the systems to be agile and capable of modifying mission objectives, routes, or strategies, often through a combination of algorithmic decision-making and human expertise. Balancing these elements is crucial to optimizing mission outcomes, ensuring safety, and complying with regulations. Key processes in managing this risk include Integrated Project Management (IPM), RSKM, and Decision Analysis and Resolution (DAR). IPM focuses on flexible planning and execution, allowing for rapid mission changes in response to evolving scenarios. RSKM involves preemptive preparation for scenarios that may necessitate quick mission changes, such as analyzing urban traffic patterns or weather conditions. Finally, DAR aids in making informed decisions during mission redefinition, evaluating various factors like fuel efficiency, time delays, and safety, ensuring a balanced approach between algorithmic adaptability and human judgment.

Mission success concerning Mission Risk Acceptability describes how well the system achieves its defined objectives, considering the effectiveness and efficiency of its operations under varying conditions. It hinges on managing risks related to achieving desired outcomes, considering factors like mission complexity, operational efficiency, and plan adherence. This begins with vague risk criteria in early development stages, acknowledging but not fully quantifying risks, and progresses to Post-Analysis criteria based on past mission evaluations. Advanced stages involve Scenario-based criteria using detailed planning and simulations for different operational scenarios, and Conditional criteria based on specific operational conditions. Attainable risk criteria are set, reflecting realistic standards achievable with current technology. This comprehensive approach integrates REQM for defining mission requirements, Project Planning (PP) and PMC for planning and execution, and RSKM for identifying and mitigating potential risks, ensuring a robust framework for mission success in diverse UAM scenarios.

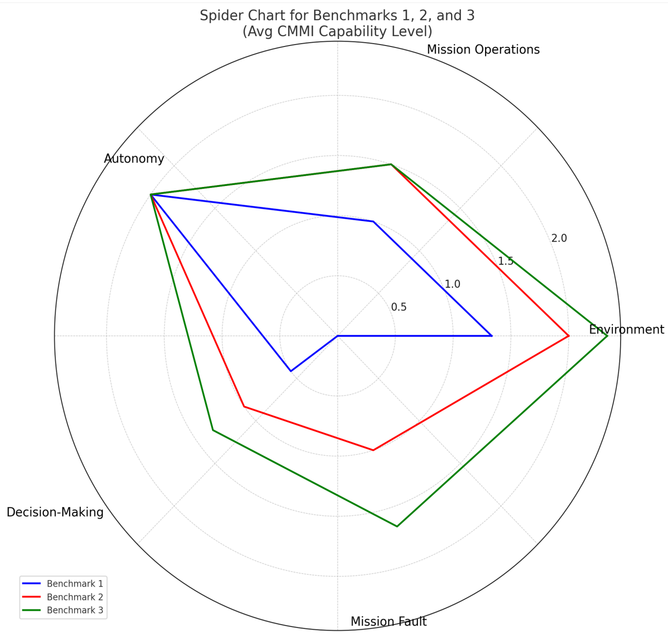
Mission Operations in this context refers to the entire system's capacity to execute and manage flight tasks, encompassing routine and complex operational aspects and responsiveness to changing scenarios. It addresses the challenges and uncertainties inherent in daily tasks such as resource allocation, airspace management, traffic coordination, and maintenance. Effective management of these risks involves establishing efficient operational practices and adapting them based on real-time data analysis for optimal resource utilization and safety. This management is initially reactive to operational needs, eventually becoming systematic and potentially adaptive (with the aid of AI). Crucial to this is MA, which leverages vast amounts of operational data for efficiency enhancements, and Process and Product Quality Assurance (PPQA), ensuring adherence to the highest safety and quality standards through regular checks and compliance with aviation regulations. Additionally, CM plays a key role in the systematic management of operational changes, while Supplier Agreement Management (SAM) focuses on maintaining standards with critical aerospace component suppliers. Poor operational risk management can lead to issues like inefficient resource use, increased costs, and compromised safety measures, underscoring the importance of these processes in maintaining smooth and safe UAM operations.

IV. Use Case: Measuring the Autonomy of GUAM

We applied our methodologies to the NASA TTT ICM sub-project simulator, known as GUAM. The NASA LaRC development team for GUAM successfully completed three benchmark exercises, showcasing the ability to handle vehicle and operational environment contingencies safely and cost-effectively. Benchmark 1 was pivotal, focusing on creating an infrastructure that supports collaborative research within the ICM architecture:

1) GUAM Simulation Framework: Development of a MATLAB™[†]/Simulink environment, adaptable to various vehicle types and research needs, forming the foundation for comprehensive simulation and analysis.
2) Data Integration and Analytics: Establishment of a robust data pipeline linking the GUAM framework with external machine learning and analytics tools, enabling seamless data exchange essential for intelligent contingency management.

[†]The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.



Benchmark 1		**Benchmark 2**		**Benchmark 3**	
Complexity Category	***Sub-Claims***	***Complexity Category***	***Sub-Claims***	***Complexity Category***	***Sub-Claims***
Environment	Structured, Known, Static	Environment	Structured, Known, Dynamic	Environment	Structured, Known, Dynamic
Mission Operations	Simple Flight Plan, Simple Flight Tasks, Normal Operations, Recoverable Faults, Fundamental Handling and Flight Quality	Mission Operations	Simple Flight Plan, Simple Flight Tasks, Normal Operations, Recoverable Faults, Fundamental Handling and Flight Quality	Mission Operations	Simple Flight Plan, Complex Flight Tasks, Normal Operations, Recoverable Faults, Fundamental Handling and Flight Quality
Autonomy	Flight Stability	Autonomy	Navigation and Collision Avoidance	Autonomy	Navigation and Collision Avoidance
Decision-Making	Control-Level, Task-Level	Decision-Making	Task-Level, Fault-Level, Control-Level	Decision-Making	Task-Level, Fault-Level, Control-Level
Mission Fault	Expected, Internal, Correctable	Mission Fault	Expected, External, Correctable	Mission Fault	Expected, External, Correctable

Fig. 3 UL 4600 Safety Case Claims and CMMI-DEV Capability Levels for GUAM Benchmarks

3) Data Storage and Sharing: Implementation of efficient data storage and sharing system using databases, enhancing data accessibility and collaborative research.
4) AI and Analytics Tools: Utilization of NASA's computing resources for in-depth analysis and decision-making processes.
5) Data Visualization: Demonstrating proficiency in data visualization with tools like Unreal Engine™and Microsoft (MS) AirSim, augmented with resources from NASA supercomputing.

Benchmark 2 for GUAM marked a significant step in evaluating vehicle control and replanning capabilities. This benchmark involved flying a standard trajectory with various atmospheric challenges like winds and turbulence, navigating around severe weather, and dealing with uncooperative intruders and gradual propulsive performance degradation. The primary goals were to assess and adapt the vehicle's performance using generalized control metrics, and to monitor these adaptations. There were notable enhancements to the infrastructure supporting collaborative research in the ICM architecture. GUAM evolved to allow for more versatile modeling, offering different levels of detail. A key development was the integration of machine learning, which significantly improved the generation and utilization of training data for machine learning models, notably in fault detection and adapting to vehicle dynamics changes. Additionally, the introduction of Parametric Differential Dynamic Programming (PDDP) represented a major advancement in trajectory planning, allowing for the simultaneous optimization of trajectories and time-invariant parameters. The safety and control capabilities of GUAM were further enhanced by integrating collision avoidance algorithms, like Optimal Reciprocal Collision Avoidance (ORCA), and using Bézier curves [53] for vehicle dynamics-aware planning. These updates collectively marked a substantial progression in GUAM's capabilities and its application in UAM scenarios.

A. Mission Complexity Analysis

Figure 3 showcases the evolution of CMMI-DEV Capability Levels and UL 4600 Sub-Claims across the yearly Benchmark Exercises for GUAM and its counterpart. Benchmark 1 addressed various contingencies and challenges such as propulsive failures, control authority degradation, traffic management issues, atmospheric disturbances, and complex decision-making scenarios. It utilized variable fidelity models to understand the intricate aerodynamics of eVTOL vehicles and baseline machine learning techniques to assess vehicle capabilities and state changes. This phase also initiated a significant overhaul in infrastructure to improve data analysis, visualization, and introduced new machine-learning models for analyzing failures and loss-of-control scenarios.

By Benchmark 2, the decision-making of GUAM was focused solely on creating the Mission Manager system,

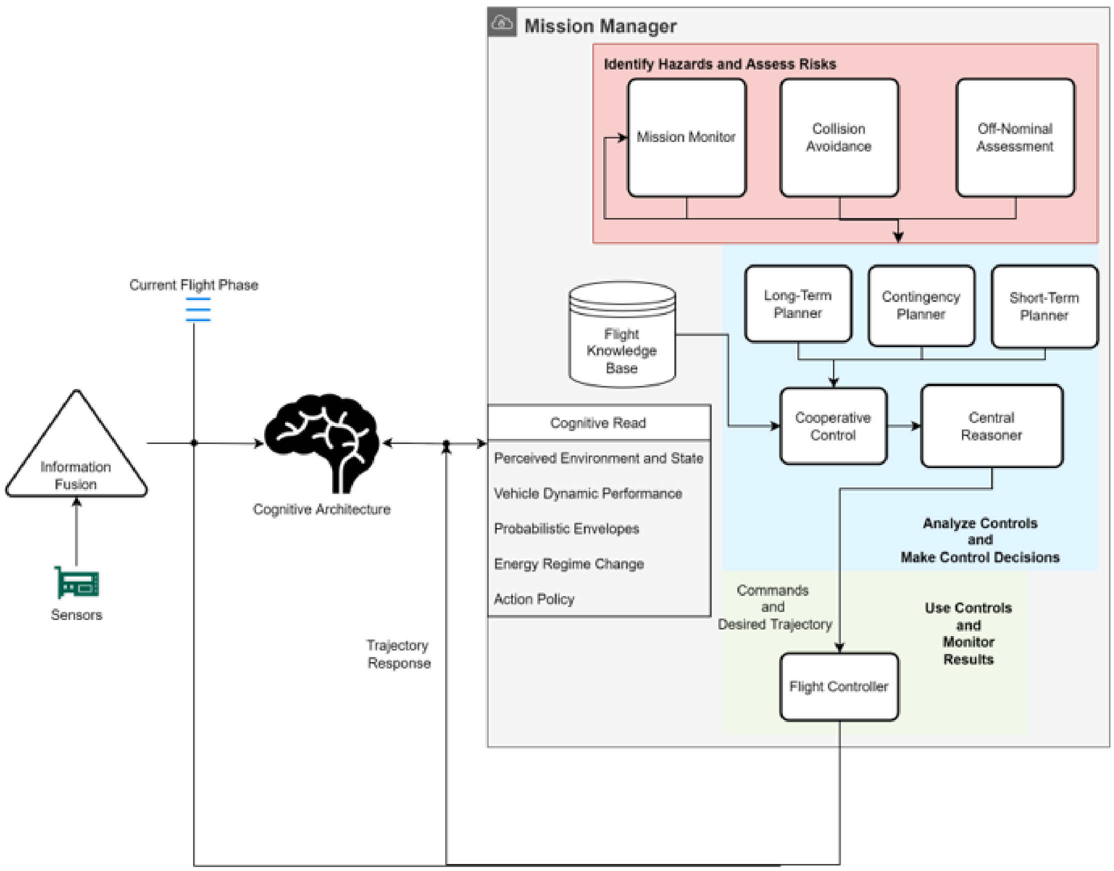


Fig. 4 Mission Manager for the GUAM Framework

illustrated in Figure 4. The Mission Manager, a critical decision-making component in GUAM, encompasses various advanced flight mission management modules. The Cognitive Read provides mission-critical data, supplied by the Cognitive Architecture, which provides advanced vehicle analyses to guide decision-making. Collision Avoidance focuses on detecting and resolving conflicts using physics and perception. Placeholders were made for future support of the other components: The Mission Monitor maintains mission progress and situational awareness, classifying objectives into Short-Term Goals (STGs) and Long-Term Goals (LTGs), while the Off-Nominal Assessment module identifies potential failures. The Short-Term Planner handles trajectory replanning, and the Long-Term Planner ensures overall mission success. The Contingency Planner prepares for emergencies, crafting contingency plans for rapid activation. Cooperative Control evaluates plans, and the Central Reasoner makes final decisions based on comprehensive contextual reasoning.

By Benchmark 3, the development of the GUAM framework had significantly evolved past the Cognitive Architecture and mission management. This phase focused on enhancing the framework's ability to handle complex operational scenarios, integrating machine learning for fault detection, advanced collision avoidance algorithms, and sophisticated mission planning tools. The emphasis was on improving vehicle performance under various conditions and refining decision-making processes during missions. The framework includes advanced analytics for scalable experimentation and simulation, crucial for predicting vehicle behavior and ensuring robustness and adaptability to new autonomy levels. The expanded decision-making hierarchy in the GUAM framework, from immediate to long-term planning, is exemplified in the evolved Mission Manager system, which incorporates monitoring, assessment, and contingency planning modules, central to mission success and informed decision-making.

Table 3 Results of Mission Risk Acceptability Analysis on GUAM

Mission Risk Acceptability Category	CMMI-DEV Process Area	SGs Met			SPs Met		
		Benchmark 1	Benchmark 2	Benchmark 3	Benchmark 1	Benchmark 2	Benchmark 3
Contingency Management	CM	2/3	2/3	2/3	5/7	6/7	6/7
	PMC	1/2	1/2	1/2	3/5	7/10	7/10
Environment	MA	1	1	1	1	1	1
	REQM	0	0	0	2/5	4/5	4/5
Mission Success	PMC	1/2	1/2	1/2	3/5	7/10	7/10
	PP	1/3	2/3	2/3	1/2	11/14	6/7
	REQM	0	0	0	2/5	4/5	4/5
Operations	CM	2/3	2/3	2/3	5/7	6/7	6/7
	MA	1	1	1	1	1	1
	PPQA	1/2	1/2	1/2	1/2	3/4	3/4
	SAM	1	1	1	1	1	1

B. Mission Risk Acceptability Analysis

In our analysis of Mission Risk Acceptability, we utilized the CMMI-DEV model, aiming to advance our Project team to Maturity Level 2 – Managed. Achieving this level requires meeting all specific and generic goals of the Maturity Level 2 process areas. We concentrated on several Process Areas at this Level relevant to the ICM for UAM sub-project operations, including CM (Contingency Management and Operations Risk), MA (Environment and Operational Risk), PMC (Contingency Management and Mission Success Risk), PP (Mission Success Risk), PPQA (Operational Risk), REQM (Environment Risk), and SAM (Operational Risk).

Results are described in Table 3. We assessed the maturity of our CM processes across the project by evaluating the number of Specific Goals (SGs) and Specific Practices (Specific Practices (SPs)) met, focusing on Mission Operations and Contingency Management. Standardized CMMI checklists were used to gauge our current compliance level in each Mission Risk Acceptability category. Our evaluation showed alignment of our processes with the SGs and SPs of each Process Area. By Benchmark 2, we achieved all SGs for CM. Our project's benchmarks revealed that while the SGs remained largely consistent, there were notable changes in SPs across several areas, particularly within CM, PMC, REQM, PP, and PPQA, indicating progress and adaptations in our approach to managing various risks.

In our analysis of Mission Risk Acceptability, we observed a consistent trend in Contingency Management, Environment, Mission Success, and Operations Risk categories. While the SGs in these categories have remained constant, indicating stable high-level objectives, there has been a notable evolution in the SPs. For Contingency Management, CM and PMC maintained consistent SGs across benchmarks, but SPs showed improvement, especially for PMC, which increased its SPs coverage. This suggests a refinement in how we implement Contingency Management. In the Environment category, despite REQM's SGs remaining at zero, there was a positive shift in SPs, highlighting better execution of requirements management practices. Similarly, for Mission Success, PMC and PP demonstrated enhanced execution of practices, with PP showing significant growth from Benchmark 1 to 3. Operations risk acceptability exhibited stability in SGs for CM and PPQA, with a slight increase in SPs, indicating improved quality assurance practices. SAM remained consistent, reflecting stable supplier management.

V. Implications of the CBP Framework

Applying CMMI-DEV to the development of evidence and measurement tools for UL 4600 Safety Cases, specifically in the context of UAM Frameworks, offered several key benefits:

- Process Maturity Assessment: CMMI-DEV provides a structured approach to assess the maturity level of the processes used to develop evidence and performance targets. It can determine how systematically and effectively these processes are being managed, which is critical for ensuring the reliability of the Evidence-gathering capabilities described in safety cases.

- Capability Enhancement: By adhering to CMMI-DEV while following a safety case about intelligent UAM, organizations can enhance their capability to develop sophisticated tools and algorithms, such as Reinforcement Learning, Genetic Algorithms, and Adaptive Neuro-Fuzzy Inference Systems. This enhancement is especially crucial for dealing with the complexities of dynamic and unstructured UAM scenarios.
- Quality Assurance: CMMI-DEV's emphasis on quality assurance ensures that the tools and methods used to gather evidence and set performance targets are robust, accurate, and reliable. This is especially important for simulations and algorithms that must demonstrate an ability to adapt to rapidly changing conditions in unknown environments.
- Continuous Improvement: The model encourages continuous improvement, which is vital in the context of UAM, where technologies and operational environments are rapidly evolving. The maturity of evidence and measurement tools can be progressively enhanced to meet the evolving safety requirements of UAM systems.
- Risk Management: CMMI-DEV's focus on risk management aids in identifying and mitigating potential errors or shortcomings in the evidence and measurement tools. This aspect is critical for ensuring that the safety cases remain valid and effective even as the operational environment becomes more complex and less predictable.

With CMMI-DEV, the tools and processes stated in developing UL 4600 Safety Cases will mature within our GUAM Framework, addressing the unique challenges posed by the ever-changing UAM ecosystem.

Concerning Mission Risk Acceptability, the static SGs coupled with the dynamic evolution of SPs suggest that our core project goals have stayed the same, but our methods and processes to achieve these goals have become more sophisticated and effective. This progression in SPs, particularly in risk management and quality assurance, shows an enhancement in our ability to manage mission risks. Our practices have become more adept, leading to more effective contingency management, better handling of environmental factors, increased mission success, and more robust operational procedures. And analysis based on the Mission Risk Acceptability, using SPs, supports this increase in effectiveness. This advancement in mission risk acceptability indicates a maturation in our risk management approach, although the future focus on improving our SGs to align with these enhancements may be beneficial.

VI. Conclusion

This research presents a significant advancement in the field of UAM, specifically addressing the critical aspect of ICM. Our comprehensive study has led to developing a CBP that leverages validated capability-maturity models and safety case frameworks to pave the way for future innovations in UAM. Throughout our investigation, we have demonstrated the efficacy of our proposed solutions in tracking the maturity of UAM Frameworks in managing contingencies within urban airspaces. As shown in our results, UL 4600 can be combined with CMMI-DEV to assess growth in capability-maturity in a yearly Benchmark Exercise. Our findings underscore the importance of adaptive and robust control mechanisms that can dynamically respond to the unpredictable nature of urban airspaces.

Furthermore, as we evolve UAM, particularly with the use of AI, it is imperative to embed ethical considerations into the core of our methodologies and findings. The methodology in this paper makes it straightforward to integrate an organization's ethical principles into UL 4600 Safety Cases such that the understanding of how the capability, process, and organization align with these principles is sufficiently tracked. NASA's AI Ethics Framework [54] captures principles that could integrate into the evidence, SPIs, and Methodologies of the UL 4600 Safety Case Claims from this study. Future research will explore adding principles from both NASA's AI Ethics Framework and climate-related goals to our UL 4600 Safety Cases to ensure a systematic method for tracking alignment.

This study adds to the academic literature and offers practical insights for policymakers and industry stakeholders, emphasizing the need for collaborative efforts to advance UAM technologies. Looking ahead, the CBP established in this research opens numerous avenues for future exploration. It invites the academic community to refine further and expand the supplementary material (UL 4600 Safety Cases and Knowledge Base), fostering a collaborative environment for innovation. NASA will internally review and publish an updated version of the supplementary material. Future research could further focus on integrating advanced machine learning algorithms and artificial intelligence to enhance the decision-making processes in UAM Framework development.

Appendix

The appendix summarizes the UL 4600 safety cases for our Mission Complexity measurements[‡].

[‡]The authors plan to publish the full UL 4600 Safety Cases after internal review within a year.

Category	Qualitative Description of Evidence	Boolean SAT Formula
Simple Flight Plan (SFP)	Basic waypoint navigation and smooth trajectory planning, adheres to UAM Corridor and FAA guidelines.	$SFP \wedge \epsilon_{\text{waypoint}} \wedge \epsilon_{\text{smoothTraj}} \wedge \epsilon_{\text{sysAdapt}} \wedge \epsilon_{\text{UAMAlign}}$
Complex Flight Plan (CFP)	Intricate, multi-waypoint routes with dynamic routing and adaptive control systems.	$CFP \wedge \epsilon_{\text{complexRoute}} \wedge \epsilon_{\text{dynamicReroute}} \wedge \epsilon_{\text{mpcAdapt}} \wedge \epsilon_{\text{rtDataProc}}$
Simple Flight Tasks (SFT)	Basic altitude control and maneuvers, with resilience in operational deviations and introductory compliance with UAM Corridor and FAA.	$SFT \wedge \epsilon_{\text{altControl}} \wedge \epsilon_{\text{maneuverExec}} \wedge \epsilon_{\text{opDevResil}}$
Complex Flight Tasks (CFT)	Advanced maneuvers that integrate system analytics, and adapts to environmental changes.	$CFT \wedge \epsilon_{\text{advManeuver}} \wedge \epsilon_{\text{sysInteract}} \wedge \epsilon_{\text{adaptLearn}}$
Normal Operations (NO)	Standard operations within FAA parameters, ensures procedural compliance and operational integrity.	$NO \wedge \epsilon_{\text{normalOps}} \wedge \epsilon_{\text{opIntegrity}}$
Abnormal Operations (AO)	Challenging conditions with advanced navigation and specialized emergency protocols.	$AO \wedge \epsilon_{\text{challCond}} \wedge \epsilon_{\text{highNavAcuity}}$
Recoverable Failures (RF)	Mission continuity despite faults with robust fault tolerance and mitigation strategies.	$RF \wedge \epsilon_{\text{recovFaults}} \wedge \epsilon_{\text{mitigStrat}}$
Unrecoverable Failures (UF)	Critical system failures with advanced fault injection methods and emergency responses.	$UF \wedge \epsilon_{\text{critSysFail}} \wedge \epsilon_{\text{emergResp}} \wedge \epsilon_{\text{advFaultInj}}$
Fundamental Handling and Flight Quality ($FHFQ$)	Smooth passenger experience or secure cargo transit during flights with minimal maneuvering or environmental challenges.	$FHFQ \wedge \epsilon_{\text{smoothFlight}} \wedge \epsilon_{\text{passComfReport}} \wedge \epsilon_{\text{stdCompliance}} \wedge \epsilon_{\text{cargoStab}} \wedge \epsilon_{\text{envImpactAssess}} \wedge \epsilon_{\text{feedbackAnalysis}}$
Advanced Handling and Flight Quality ($AHFQ$)	High standards of passenger comfort and cargo integrity during complex flights with intricate maneuvers or challenging conditions.	$AHFQ \wedge \epsilon_{\text{advCtrlStab}} \wedge \epsilon_{\text{complexOperResponse}} \wedge \epsilon_{\text{realTimeMonit}} \wedge \epsilon_{\text{adaptiveCtrlSys}} \wedge \epsilon_{\text{stressTestSim}}$

Table 4 Complexity of Mission Operations: Categories and Their Corresponding SAT Formulas

1. Complexity of Mission Operations

In Table 4, we summarize the UL 4600 Safety Cases for **Complexity of Mission Operations**:, where

$\epsilon_{\textbf{adaptiveCtrlSys}}$ = Implementation of control systems adapting to changing conditions.

$\epsilon_{\textbf{adaptLearn}}$ = Use of adaptive learning algorithms that accomodate environmental changes.

$\epsilon_{\textbf{advCtrlStab}}$ = Advanced control and stabilization tests under complex flight scenarios.

$\epsilon_{\textbf{advFaultInj}}$ = Use of advanced Fault Injection mechanisms, encompassing methods like Monte Carlo simulations or Chaos Engineering principles.

$\epsilon_{\textbf{advManeuver}}$ = Execution of advanced operational maneuvers.

$\epsilon_{\textbf{altControl}}$ = Execution of fundamental flight tasks like altitude control.

$\epsilon_{\textbf{cargoStab}}$ = Use of sensors and algorithms to ensure cargo stability and security.

$\epsilon_{\textbf{challCond}}$ = Flights under challenging conditions, such as adverse weather or complex urban landscapes.

$\epsilon_{\textbf{complexRoute}}$ = Management of intricate routing with multiple waypoints.

$\epsilon_{\textbf{complexOperResponse}}$ = System response records to high turbulence or maneuver-induced stresses.

$\epsilon_{\textbf{critSysFail}}$ = Critical system failures leading to mission failure.

$\epsilon_{\textbf{dynamicReroute}}$ = Capability for dynamic rerouting in response to evolving circumstances.

$\epsilon_{\textbf{emergResp}}$ = Emergency response systems for critical system failures leading to mission failure.

$\epsilon_{\textbf{envImpactAssess}}$ = Analysis of external environmental factors impacting flight stability.

$\epsilon_{\textbf{feedbackAnalysis}}$ = Collection and analysis of passenger and cargo handler feedback post-flight.

$\epsilon_{\textbf{highNavAcuity}}$ = Requirement for high navigational acuity in abnormal operations.

$\epsilon_{\textbf{maneuverExec}}$ = Execution of basic maneuvers for flight vehicles.

$\epsilon_{\textbf{mitigStrat}}$ = Mitigation strategies employed by the system to manage and mitigate impacts of recoverable faults.

$\epsilon_{\textbf{mpcAdapt}}$ = Implementation of Model Predictive Control adaptive algorithms for handling variable speeds

Complexity of Autonomy	Details
Manual Control (MC)	**Evidence Claim:** The UAM Framework is designed to facilitate flight operations under full manual control. **Evidence:** Operator Control Algorithms, Human-Machine Interface (HMI) Effectiveness, Operator Training and Simulation. **SPIs:** Operator Response Accuracy, HMI Usability and Effectiveness, Training Effectiveness. **SAT:** $MC \wedge (\epsilon_{\text{opCtrlAlg}} \wedge \epsilon_{\text{HMIeff}} \wedge \epsilon_{\text{opTrain}}) \wedge (\epsilon_{\text{respAcc}}) \wedge (\epsilon_{\text{situAware}}) \wedge (\epsilon_{\text{manOverride}})$
Flight Stability (FS)	**Evidence Claim:** The UAM Framework ensures enhanced flight stability through automated stabilization systems. **Evidence:** Stabilization Algorithms, Sensor Fusion and Accuracy, Operator-Automation Interface. **SPIs:** Stabilization System Reliability, Deviation from Set Points, Response to Perturbations. **SAT:** $FS \wedge (\epsilon_{\text{stabAlg}} \wedge \epsilon_{\text{sensorFusion}} \wedge \epsilon_{\text{opAutoIntf}}) \wedge (\epsilon_{\text{flightStab}}) \wedge (\epsilon_{\text{sensorAcc}}) \wedge (\epsilon_{\text{opSysInteg}})$
Envelope Protection (EP)	**Evidence Claim:** The UAM Framework ensures safe flight operations with advanced envelope protection systems. **Evidence:** Advanced Envelope Protection Algorithms, Sensor Fusion for Flight Dynamics, Contingency Management Tools. **SPIs:** Effectiveness of Envelope Protection, Responsiveness of Contingency Management, Reliability of Sensor Fusion. **SAT:** $EP \wedge (\epsilon_{\text{envProtAlg}} \wedge \epsilon_{\text{dynMon}} \wedge \epsilon_{\text{contMgmt}}) \wedge (\epsilon_{\text{envEff}}) \wedge (\epsilon_{\text{respTime}}) \wedge (\epsilon_{\text{contAcc}})$
Navigation and Collision Avoidance (NCA)	**Evidence Claim:** The UAM Framework autonomously navigates and avoids collisions, ensuring safety and efficiency. **Evidence:** Collision Avoidance Algorithms, Adaptive Control Systems, Sensor Fusion and Real-Time Monitoring. **SPIs:** Collision Avoidance Effectiveness, Navigation Precision, Sensor Data Reliability. **SAT:** $NCA \wedge (\epsilon_{\text{colAvoidAlg}} \wedge \epsilon_{\text{adaptCtrlSys}} \wedge \epsilon_{\text{realTimeMon}}) \wedge (\epsilon_{\text{collAvoidRate}}) \wedge (\epsilon_{\text{navAcc}}) \wedge (\epsilon_{\text{realTimeResp}})$

Table 5 Summary of UL 4600 Safety Cases for Assisted Flight Levels of Autonomy Complexity

and altitudes.

$\epsilon_{\textbf{normalOps}}$ = Management of standard operational conditions within expected parameters.

$\epsilon_{\textbf{opDevResil}}$ = Resilience in handling operational deviations.

$\epsilon_{\textbf{opIntegrity}}$ = Integrity of operational procedures ensuring compliance with standard flight operations and regulations.

$\epsilon_{\textbf{passComfReport}}$ = Reports collected on passenger comfort during normal flight operations.

$\epsilon_{\textbf{realTimeMonit}}$ = Real-time monitoring systems for passenger comfort and cargo security.

$\epsilon_{\textbf{recovFaults}}$ = Ability to continue the mission despite encountering faults.

$\epsilon_{\textbf{rtDataProc}}$ = Use of real-time data processing algorithms for dynamic rerouting in response to unexpected environmental factors.

$\epsilon_{\textbf{smoothFlight}}$ = Stability and control system validation tests demonstrating smooth flight conditions.

$\epsilon_{\textbf{smoothTraj}}$ = Implementation of smooth trajectory planning, like efficient Bezier curve calculation.

$\epsilon_{\textbf{stdCompliance}}$ = Historical flight data showing consistent adherence to flight quality standards.

$\epsilon_{\textbf{stressTestSim}}$ = Simulated stress testing to evaluate flight quality under complex conditions.

$\epsilon_{\textbf{sysAdapt}}$ = System adaptability to simple operator directives and environmental changes.

$\epsilon_{\textbf{sysInteract}}$ = Managing complex interactions with automated systems.

$\epsilon_{\textbf{UAMAlign}}$ = Initial alignment with UAM Corridor concepts and FAA guidelines.

$\epsilon_{\textbf{waypoint}}$ = Use of waypoint navigation algorithms for straightforward route following.

2. Complexity of Autonomy

Tables 5 and 6 describe the UL 4600 safety cases developed for **Complexity of Autonomy**, where

Complexity of Autonomy	Details
Conditional Autonomy (CA)	**Evidence Claim:** The UAM Framework operates autonomously under specific conditions with human override capability. **Evidence:** Robust Autonomous Flight Systems, HMI and Override Functionality, Environmental Adaptation Systems. **SPIs:** Autonomy Reliability, Human-System Interaction Efficacy, Operational Safety under Varying Conditions. **SAT:** $CA \wedge (\epsilon_{\text{autoFlightSys}} \wedge \epsilon_{\text{HMIoverFunc}} \wedge \epsilon_{\text{envAdaptSys}}) \wedge (\epsilon_{\text{autoOpSuccess}}) \wedge (\epsilon_{\text{manualTrans}}) \wedge (\epsilon_{\text{opSafetyEff}})$
Conditional Autonomy with AI (CAAI)	**Evidence Claim:** AI integration enhances decision-making in conditional autonomy with operator supervision. **Evidence:** AI-Enhanced Decision-Making Systems, Dynamic Environmental Adaptation with AI, AI Supervisory Systems. **SPIs:** Reliability of AI Decision-Making, Effectiveness of Human-AI Interaction, Adaptability of AI Systems. **SAT:** $CAAI \wedge (\epsilon_{\text{AIdecSys}} \wedge \epsilon_{\text{envAdaptAI}} \wedge \epsilon_{\text{AIsupSys}}) \wedge (\epsilon_{\text{AIDecAcc}}) \wedge (\epsilon_{\text{opIntervTime}}) \wedge (\epsilon_{\text{humanIntervMin}})$
High Automation (HA)	**Evidence Claim:** High automation enables autonomous operation with minimal human intervention, except in rare cases. **Evidence:** Advanced Autonomous Operation Systems, Sophisticated Monitoring and Alert Systems, Fail-Safe Mechanisms. **SPIs:** Autonomy Reliability, Monitoring and Alert System Effectiveness, Fail-Safe System Performance. **SAT:** $HA \wedge (\epsilon_{\text{advAutoOpSys}} \wedge \epsilon_{\text{monAlertSys}} \wedge \epsilon_{\text{failSafeMech}}) \wedge (\epsilon_{\text{autoOpRate}}) \wedge \epsilon_{\text{anomDetectAcc}} \wedge \epsilon_{\text{failSafeEngage}}$
Full Automation (FA)	**Evidence Claim:** Full Automation allows for autonomous operation with nominal human oversight in all scenarios. **Evidence:** Fully Autonomous Operational Systems, Robust Hazard Detection and Avoidance, System Redundancy. **SPIs:** Operational Autonomy, Hazard Detection Efficiency, System Redundancy and Reliability. **SAT:** $FA \wedge (\epsilon_{\text{fullAutoSys}} \wedge \epsilon_{\text{hazDetAvoid}} \wedge \epsilon_{\text{sysRedundancy}}) \wedge \epsilon_{\text{autoOpSuccessAll}} \wedge \epsilon_{\text{redundFailZero}} \wedge (\epsilon_{\text{certComplete}})$

Table 6 Summary of UL 4600 Safety Cases for Autonomous Flight Levels of Autonomy Complexity

$\epsilon_{\textbf{advAutoOpSys}}$ = Advanced Autonomous Operation Systems.

$\epsilon_{\textbf{AIDecAcc}}$ = AI Decision-Making Accuracy in Conditional Autonomy with AI.

$\epsilon_{\textbf{AIdecSys}}$ = AI-Enhanced Decision-Making Systems.

$\epsilon_{\textbf{AIsupSys}}$ = AI Supervisory Systems.

$\epsilon_{\textbf{anomDetectAcc}}$ = Anomaly Detection Accuracy in High Automation.

$\epsilon_{\textbf{autoFlightSys}}$ = Robust Autonomous Flight Systems.

$\epsilon_{\textbf{autoOpRate}}$ = Autonomous Operation Success Rate in High Automation.

$\epsilon_{\textbf{autoOpSuccess}}$ = Autonomous operation success rate under specific conditions.

$\epsilon_{\textbf{autoOpSuccessAll}}$ = Autonomous Operation Success Rate in Full Automation.

$\epsilon_{\textbf{colAvoidAlg}}$ = Collision Avoidance Algorithms.

$\epsilon_{\textbf{certComplete}}$ = Completion of Certification for Autonomous Systems in Full Automation.

$\epsilon_{\textbf{collAvoidRate}}$ = Collision Avoidance Success Rate in Navigation and Collision Avoidance.

$\epsilon_{\textbf{contAcc}}$ = Accuracy of Contingency Management.

$\epsilon_{\textbf{contMgmt}}$ = Contingency Management Tools.

$\epsilon_{\textbf{dynMon}}$ = Dynamic Monitoring for Flight Dynamics.

$\epsilon_{\textbf{envAdaptAI}}$ = Dynamic Environmental Adaptation with AI in Conditional Autonomy with AI.

$\epsilon_{\textbf{envAdaptSys}}$ = Environmental Adaptation Systems in Conditional Autonomy.

$\epsilon_{\textbf{envEff}}$ = Effectiveness of Envelope Protection.

$\epsilon_{\textbf{envProtAlg}}$ = Advanced Envelope Protection Algorithms.

$\epsilon_{\textbf{failSafeEngage}}$ = Fail-Safe System Engagement Rate in High Automation.

$\epsilon_{\textbf{failSafeMech}}$ = Robust Fail-Safe Mechanisms.

$\epsilon_{\textbf{flightStab}}$ = Flight Stability Precision.

$\epsilon_{\textbf{fullAutoSys}}$ = Fully Autonomous Operational Systems.

$\epsilon_{\textbf{hazDetAvoid}}$ = Robust Hazard Detection and Avoidance Systems.

$\epsilon_{\textbf{HMIeff}}$ = Effectiveness of the HMI in operation.

$\epsilon_{\textbf{HMIoverFunc}}$ = Human-Machine Interface and Override Functionality.

$\epsilon_{\textbf{humanIntervMin}}$ = Frequency of Minimal Unscheduled Human Interventions in Conditional Autonomy with

AI.

$\epsilon_{\mathbf{manOverride}}$ = Effectiveness of Manual Override Systems.

$\epsilon_{\mathbf{manualTrans}}$ = Rapid Transition to Manual Control in Conditional Autonomy.

$\epsilon_{\mathbf{monAlertSys}}$ = Sophisticated Monitoring and Alert Systems.

$\epsilon_{\mathbf{navAcc}}$ = Navigation Accuracy within predefined routes.

$\epsilon_{\mathbf{opAutoIntf}}$ = Operator-Automation Interface.

$\epsilon_{\mathbf{opCtrlAlg}}$ = Implementation of Operator Control Algorithms for manual control.

$\epsilon_{\mathbf{opIntervTime}}$ = Operator Intervention Time in AI Supervision.

$\epsilon_{\mathbf{opSafetyEff}}$ = Operational Safety and Efficiency in Conditional Autonomy.

$\epsilon_{\mathbf{opSysInteg}}$ = Operator-System Integration.

$\epsilon_{\mathbf{opTrain}}$ = Comprehensive Operator Training and Simulation programs.

$\epsilon_{\mathbf{realTimeMon}}$ = Real-Time Monitoring for Navigation and Collision Avoidance.

$\epsilon_{\mathbf{realTimeResp}}$ = Real-Time Response to Dynamic Obstacles and Changes.

$\epsilon_{\mathbf{redundFailZero}}$ = Redundancy Switch-over Success Rate in Full Automation.

$\epsilon_{\mathbf{respAcc}}$ = Operator response accuracy in control inputs.

$\epsilon_{\mathbf{respTime}}$ = Response Time of Envelope Protection Systems.

$\epsilon_{\mathbf{scenario\_anal}}$ = Scenario Analysis Tools for Impact Assessment.

$\epsilon_{\mathbf{sensorAcc}}$ = Sensor Accuracy in Flight Stability.

$\epsilon_{\mathbf{sensorFusion}}$ = Sensor Fusion and Accuracy in Flight Stability.

$\epsilon_{\mathbf{stabAlg}}$ = Stabilization Algorithms in use.

$\epsilon_{\mathbf{situAware}}$ = Operator's situational awareness level.

$\epsilon_{\mathbf{sysRedundancy}}$ = System Redundancy in Full Automation.

3. *Complexity of Decision-Making*

Table 7 Summary of UL 4600 Sub-Claims for Decision-Making Complexity

Decision-Making Sub-Claim	**Summary of Evidence, Performance Targets, and SPIs**	**SAT**
Mission-level	Focuses on strategic mission planning and real-time adaptation with high accuracy and reliability.	$mission\_level \wedge \epsilon_{\text{strategic\_planning}} \wedge \epsilon_{\text{adaptation\_mechanisms}} \wedge \epsilon_{\text{objective\_optimization}} \wedge \epsilon_{\text{validation}}$
Task-level	High-efficiency task decomposition and multi-option reasoning in task sequencing and accuracy in task prioritization.	$task\_level \wedge \epsilon_{\text{task\_decomposition}} \wedge \epsilon_{\text{multi\_option\_reasoning}} \wedge \epsilon_{\text{task\_prioritization}} \wedge \epsilon_{\text{task\_level\_validation}}$
Plan-level	Feasible plan generation and adaptability in planning, high effectiveness in plan generation and environmental adaptation.	$plan\_level \wedge \epsilon_{\text{plan\_generation}} \wedge \epsilon_{\text{environmental\_adaptation}} \wedge \epsilon_{\text{uncertainty\_management}} \wedge \epsilon_{\text{plan\_level\_validation}}$
Maneuver-level	Precision maneuver selection and aerodynamic optimization, aiming for high accuracy in maneuver selection and efficiency.	$maneuver\_level \wedge \epsilon_{\text{maneuver\_selection}} \wedge \epsilon_{\text{aero\_optimization}} \wedge \epsilon_{\text{real\_time\_data}} \wedge \epsilon_{\text{maneuver\_validation}} \wedge \epsilon_{\text{dynamic\_envelope}} \wedge \epsilon_{\text{offline\_envelope}} \wedge \epsilon_{\text{adaptive\_maneuver}}$
Control-level	Autonomous control input adjustment and adaptive control algorithms, with goals for precision in control adjustments.	$control\_level \wedge \epsilon_{\text{control\_input}} \wedge \epsilon_{\text{control\_precision}} \wedge \epsilon_{\text{adaptive\_control}} \wedge \epsilon_{\text{control\_validation}}$
Health-level	Intelligent health monitoring and active health assessment, focusing on reliability in health monitoring.	$health\_level \wedge \epsilon_{\text{health\_monitoring}} \wedge \epsilon_{\text{active\_assessment}} \wedge \epsilon_{\text{health\_based\_decision}} \wedge \epsilon_{\text{health\_validation}}$
Fault-level	Fault detection and response algorithms, with targets of accuracy in fault detection and effective fault analysis.	$fault\_level \wedge \epsilon_{\text{fault\_detection}} \wedge \epsilon_{\text{fault\_analysis}} \wedge \epsilon_{\text{fault\_response}} \wedge \epsilon_{\text{fault\_validation}}$
Recovery-level	Recovery strategy formulation and mission re-planning tools, aiming for efficiency in recovery strategy execution.	$recovery\_level \wedge \epsilon_{\text{recovery\_strategy}} \wedge \epsilon_{\text{mission\_replanning}} \wedge \epsilon_{\text{post\_recovery\_optimization}} \wedge \epsilon_{\text{fail\_safe\_design}}$

Table 7 summarizes the UL 4600 Safety Case for the **Complexity of Decision-Making** claim, where:

$\epsilon_{\textbf{strategic\_planning}}$ = Implementation of algorithms capable of long-term mission planning.
$\epsilon_{\textbf{adaptation\_mechanisms}}$ = Integration of systems to adapt mission objectives in real-time.
$\epsilon_{\textbf{objective\_optimization}}$ = Utilizing optimization techniques for balancing multiple mission objectives.
$\epsilon_{\textbf{validation}}$ = Comprehensive testing and validation of mission planning algorithms.
$\epsilon_{\textbf{task\_decomposition}}$ = Implementation of algorithms for breaking down complex missions into discrete tasks.
$\epsilon_{\textbf{multi\_option\_reasoning}}$ = The system's ability to evaluate and reason about various optional tasks.
$\epsilon_{\textbf{task\_prioritization}}$ = Dynamic prioritization mechanisms adjusting task priorities in real-time.
$\epsilon_{\textbf{task\_level\_validation}}$ = Extensive validation and testing of task-level decision-making systems.
$\epsilon_{\textbf{plan\_generation}}$ = Integration of algorithms for generating practical and executable plans for each task.
$\epsilon_{\textbf{environmental\_adaptation}}$ = Capability to adjust plans in response to changing environmental conditions.
$\epsilon_{\textbf{uncertainty\_management}}$ = Systems designed to anticipate and accommodate uncertainties in planning.
$\epsilon_{\textbf{plan\_level\_validation}}$ = Thorough testing and validation of plan-level decision-making capabilities.
$\epsilon_{\textbf{maneuver\_selection}}$ = The incorporation of algorithms that accurately select maneuvers.
$\epsilon_{\textbf{aero\_optimization}}$ = Advanced algorithms for optimizing maneuvers considering aerodynamic efficiency.
$\epsilon_{\textbf{real\_time\_data}}$ = Ability to process and integrate real-time data for informed maneuver decisions.
$\epsilon_{\textbf{maneuver\_validation}}$ = Rigorous testing of maneuver selection and execution algorithms.
$\epsilon_{\textbf{dynamic\_envelope}}$ = Capability to compute a real-time dynamic flight envelope.
$\epsilon_{\textbf{offline\_envelope}}$ = Implementing algorithms for offline computation of the flight envelope.
$\epsilon_{\textbf{adaptive\_maneuver}}$ = Using advanced algorithms for selecting maneuvers based on a computed flight envelope.
$\epsilon_{\textbf{control\_input}}$ = System's capability to autonomously determine and adjust control inputs.
$\epsilon_{\textbf{control\_precision}}$ = Implementation of algorithms that accurately set control parameters.
$\epsilon_{\textbf{adaptive\_control}}$ = Use of adaptive control systems that adjust to varying flight conditions.
$\epsilon_{\textbf{control\_validation}}$ = Comprehensive testing and validation of control-level decision systems.
$\epsilon_{\textbf{health\_monitoring}}$ = Implementation of systems for continuous health status monitoring of the aircraft.
$\epsilon_{\textbf{active\_assessment}}$ = Capability for active probing of the aircraft's operational abilities.
$\epsilon_{\textbf{health\_based\_decision}}$ = Integration of decision-making algorithms that respond to health assessments.
$\epsilon_{\textbf{health\_validation}}$ = Comprehensive testing and validation of the health monitoring and decision-making systems.
$\epsilon_{\textbf{fault\_detection}}$ = Implementation of systems for accurate fault detection and diagnosis.
$\epsilon_{\textbf{fault\_analysis}}$ = Use of advanced modeling and simulation methods for fault analysis.
$\epsilon_{\textbf{fault\_response}}$ = Integration of decision-making algorithms for responding to detected faults.
$\epsilon_{\textbf{fault\_validation}}$ = Comprehensive testing and validation of fault-level decision systems.
$\epsilon_{\textbf{recovery\_strategy}}$ = Formulation and implementation of sophisticated recovery strategies.
$\epsilon_{\textbf{mission\_replanning}}$ = Implementation of tools for dynamic mission re-planning post-recovery.
$\epsilon_{\textbf{post\_recovery\_optimization}}$ = Utilization of multi-criteria optimization models for post-recovery operations.
$\epsilon_{\textbf{fail\_safe\_design}}$ = Application of fail-safe system design principles and real-time monitoring.

4. *Complexity of Mission Fault*

Table 8 summarizes the UL 4600 Safety Case for the **Complexity of Mission Fault** claim, where:

- $\epsilon_{\text{acc\_int\_plan}}$ = High accuracy of planning for expected internal uncorrectable faults.
- $\epsilon_{\text{acc\_int\_pred}}$ = High accuracy of fault prediction for expected internal correctable faults.
- $\epsilon_{\text{acc\_plan}}$ = High accuracy of planning for expected external uncorrectable faults.
- $\epsilon_{\text{auto\_detect}}$ = Implementation of automated fault detection systems for internal faults.
- $\epsilon_{\text{bayes\_net}}$ = Use of Bayesian networks or similar probabilistic models for anticipating uncorrectable external faults.
- $\epsilon_{\text{crisis\_mgmt}}$ = Use of crisis management systems for managing uncorrectable faults.
- $\epsilon_{\text{diag\_tools}}$ = Integration of advanced diagnostic tools for detecting internal faults.
- $\epsilon_{\text{emergency\_proto}}$ = Deployment of emergency management protocols for uncorrectable faults.
- $\epsilon_{\text{env\_sense}}$ = Advanced environmental sensing technologies for detecting external changes.
- $\epsilon_{\text{ext\_detect\_rate}}$ = High success rate in detecting unexpected external correctable faults.
- $\epsilon_{\text{ext\_manage\_rate}}$ = High effectiveness in managing un-

Table 8 Summary of UL 4600 Sub-Claims for Mission Fault Complexity

Mission Fault Sub-Claim	Summary of Evidence, Performance Targets, and SPIs	SAT
Expected, External, Correctable	Utilizes machine learning, Monte Carlo simulations, and predictive models for fault management, with high accuracy in predicting external faults and reducing mission impact.	**(Expected $\wedge$ External $\wedge$ Correctable)** $\wedge (\epsilon_{\text{ml\_alg}} \vee \epsilon_{\text{mc\_sim}} \vee \epsilon_{\text{pred\_mod}}) \wedge \epsilon_{\text{env\_sense}} \wedge \epsilon_{\text{real\_time\_data}} \wedge \epsilon_{\text{ext\_detect\_rate}} \wedge \epsilon_{\text{mis\_disrupt\_ext}}$
Expected, External, Uncorrectable	Employs probabilistic models and scenario analysis tools for minimizing mission impact and ensuring safety, targeting high planning accuracy and reduced mission disruption.	**(Expected $\wedge$ External $\wedge$ Uncorrectable)** $\wedge \epsilon_{\text{bayes\_net}} \wedge \epsilon_{\text{scenario\_anal}} \wedge \epsilon_{\text{acc\_plan}} \wedge \epsilon_{\text{min\_disrupt\_ext}}$
Expected, Internal, Correctable	Uses fault prediction algorithms and diagnostics to ensure operational efficiency, with high accuracy in internal fault prediction and reduced mission impact.	**(Expected $\wedge$ Internal $\wedge$ Correctable)** $\wedge \epsilon_{\text{reg\_model}} \wedge \epsilon_{\text{sys\_diag}} \wedge \epsilon_{\text{acc\_int\_pred}}$
Expected, Internal, Uncorrectable	Focuses on minimizing mission impact using statistical and fault tree analysis, aiming for high accuracy in fault prediction and minimizing disruption.	**(Expected $\wedge$ Internal $\wedge$ Uncorrectable)** $\wedge \epsilon_{\text{stat\_anal}} \wedge \epsilon_{\text{fault\_tree}} \wedge \epsilon_{\text{acc\_int\_plan}} \wedge \epsilon_{\text{min\_int\_disrupt}}$
Unexpected, External, Correctable	Ensures operational flexibility with environmental sensing technologies and adaptive control systems, targeting a high success rate in fault detection and limiting mission disruption.	**(Unexpected $\wedge$ External $\wedge$ Correctable)** $\wedge \epsilon_{\text{env\_sense}} \wedge \epsilon_{\text{real\_time\_data}} \wedge \epsilon_{\text{ml\_alg}} \wedge \epsilon_{\text{ext\_detect\_rate}} \wedge \epsilon_{\text{mis\_disrupt\_ext}}$
Unexpected, External, Uncorrectable	Deploys advanced sensors and crisis management systems for managing faults effectively, ensuring safety and mission-critical objectives.	**(Unexpected $\wedge$ External $\wedge$ Uncorrectable)** $\wedge \epsilon_{\text{sensor\_array}} \wedge \epsilon_{\text{crisis\_mgmt}} \wedge \epsilon_{\text{ext\_manage\_rate}} \wedge \epsilon_{\text{safety\_ext}}$
Unexpected, Internal, Correctable	Maintains high operational efficiency using advanced diagnostics, automated detection, and AI systems, with a high success rate in fault resolution and limited mission impact.	**(Unexpected $\wedge$ Internal $\wedge$ Correctable)** $\wedge \epsilon_{\text{diag\_tools}} \wedge \epsilon_{\text{auto\_detect}} \wedge \epsilon_{\text{rapid\_response}} \wedge (\epsilon_{\text{int\_detect\_rate}} \vee \epsilon_{\text{mis\_disrupt\_int}})$
Unexpected, Internal, Uncorrectable	Focuses on safety and operational continuity using system monitoring, decision support, and emergency protocols, aiming for effective fault management and rapid emergency response.	**(Unexpected $\wedge$ Internal $\wedge$ Uncorrectable)** $\wedge \epsilon_{\text{sys\_mon}} \wedge \epsilon_{\text{crisis\_mgmt}} \wedge \epsilon_{\text{emergency\_proto}} \wedge \epsilon_{\text{fault\_manage\_eff}} \wedge \epsilon_{\text{response\_time}}$

expected external uncorrectable faults.
- $\epsilon_{\text{fault\_manage\_eff}}$ = High effectiveness in managing uncorrectable internal faults.
- $\epsilon_{\text{fault\_tree}}$ = Incorporation of fault tree analysis for internal fault pathways.
- $\epsilon_{\text{int\_detect\_rate}}$ = High success rate in detecting unexpected internal correctable faults.
- $\epsilon_{\text{mc\_sim}}$ = Application of Monte Carlo (or similar) methods for fault impact assessment.
- $\epsilon_{\text{min\_disrupt\_ext}}$ = Low level of mission disruption due to expected external uncorrectable faults.
- $\epsilon_{\text{min\_int\_disrupt}}$ = Low level of mission disruption due to expected internal uncorrectable faults.
- $\epsilon_{\text{mis\_disrupt\_ext}}$ = Low level of mission disruption due to unexpected external correctable faults.
- $\epsilon_{\text{mis\_disrupt\_int}}$ = Low level of mission disruption due to unexpected internal correctable faults.
- $\epsilon_{\text{ml\_alg}}$ = Use of machine learning algorithms for predictive analytics.
- $\epsilon_{\text{pred\_mod}}$ = Use of predictive models for environmental and air traffic conditions.
- $\epsilon_{\text{rapid\_response}}$ = Automated systems for rapid fault correction.
- $\epsilon_{\text{real\_time\_data}}$ = The use of real-time data processing systems for external fault management.
- $\epsilon_{\text{reg\_model}}$ = Use of regression models for internal fault prediction.
- $\epsilon_{\text{response\_time}}$ = Low response time for implementing emergency protocols in unexpected internal uncorrectable fault scenarios.
- $\epsilon_{\text{safety\_ext}}$ = High level of safety in scenarios with unexpected external uncorrectable faults.
- $\epsilon_{\text{sensor\_array}}$ = Deployment of sensor arrays for environmental monitoring.
- $\epsilon_{\text{sys\_diag}}$ = Routine system diagnostics for internal fault prediction.
- $\epsilon_{\text{sys\_mon}}$ = Comprehensive system monitoring for internal fault detection.

Acknowledgments

This research was supported by the NASA Aeronautics Research Mission Directorate (ARMD), TTT project, under the Autonomous Systems / ICM. Additional thanks to the NASA LaRC MidRange Computer User's Group for supporting experimental processing and analysis on the K-cluster.

References

[1] Behringer, R., Sundareswaran, S., Gregory, B., Elsley, R., Addison, B., Guthmiller, W., Daily, R., and Bevly, D., "The DARPA Grand Challenge - Development of an Autonomous Vehicle," *IEEE Intelligent Vehicles Symposium, 2004*, 2004, pp. 226–231. URL `https://ieeexplore.ieee.org/document/1336386`.

[2] Buehler, M., Iagnemma, K., and Singh, S., *The DARPA urban challenge: autonomous vehicles in city traffic*, Vol. 56, springer, 2009. URL `https://link.springer.com/book/10.1007/978-3-642-03991-1`.

[3] Ackerman, E., "Robots Conquer the Underground: What Darpa's Subterranean Challenge Means for the Future of Autonomous Robots," *IEEE Spectrum*, Vol. 59, No. 5, 2022, pp. 30–37.

[4] Doty, M. J., "NASA Urban Air Mobility Noise Working Group Activities," *Transportation Research Board (TRB) 99th Annual Meeting*, 2020. URL `https://ntrs.nasa.gov/citations/20220008095`.

[5] Guangying, Y., "Research of Strategy for RoboCup Soccer Robots Competition," *Procedia Engineering*, Vol. 15, 2011, pp. 649–654. URL `https://www.sciencedirect.com/science/article/pii/S1877705811016225`.

[6] Underwriters Laboratories, "UL 4600: Standard for Safety for the Evaluation of Autonomous Products," `https://users.ece.cmu.edu/~koopman/ul4600/index.html`, 2020.

[7] CMMI Product Team, T., "CMMI for Development, Version 1.3," Tech. Rep. CMU/SEI-2010-TR-033, Nov 2010. URL `https://doi.org/10.1184/R1/6572342.v1`.

[8] Gregory, I. M., Neogi, N. A., Grauer, J. A., Campbell, N. H., Holbrook, J., Bacon, B. J., Murphy, P. C., Moerder, D. D., Simmons, B. M., Acheson, M. J., Britton, T. C., and Cook, J., "Intelligent Contingency Management for Urban Air Mobility," *AIAA Scitech 2021 Forum*, 2021. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2021-1000`.

[9] Jacklin, S., Schumann, J., Gupta, P., Lowry, M., Bosworth, J., Zavala, E., Hayhurst, K., Belcastro, C., and Belcastro, C., *Verification, Validation, and Certification Challenges for Adaptive Flight-Critical Control System Software*, 2004. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2004-5258`.

[10] Kopardekar, P., "Enabling autonomous flight and operations in the national airspace," Tech. Rep. ARC-E-DAA-TN72470, NASA, 2019.

[11] Sakahara, R. D., "Technology Development Project Plan Phase 2: Unmanned Aircraft Systems (UAS) Integration in the National Airspace System (NAS) Integrated Aviation Systems Program (IASP)," Tech. rep., 2017.

[12] Baccheschi, N., and Rorie, C., "Integration of Automation Systems Flight Test Overview," *Radio Technical Commission for Aeronautics (RTCA) SC-147 Face to Face Meeting*, NASA, 2023.

[13] Price, G., Helton, D., Jenkins, K., Kvicala, M., Parker, S., Wolfe, R., Miranda, F. A., Goodrich, K. H., Xue, M., Cate, K. T., et al., "Urban air mobility operational concept (OpsCon) passenger-carrying operations," Tech. Rep. NASA/CR—2020–5001587, NASA, 2020.

[14] Johnson, W. C., "UAM Coordination and Assessment Team (UCAT) NASA UAM for ENRI Technical Interchange Meeting," Tech. rep., 2019.

[15] Goodrich, K. H., and Theodore, C. R., "Description of the NASA Urban Air Mobility Maturity Level (UML) Scale," *AIAA Scitech 2021 Forum*, 2021, p. 1627. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2021-1627`.

[16] Kotov, V., *Systems of Systems as Communicating Structures*, Horwood Publishing, Ltd., USA, 1999, p. 141–154.

[17] Konur, D., Farhangi, H., and Dagli, C. H., "On the Flexibility of Systems in System of Systems Architecting," *Procedia Computer Science*, Vol. 36, 2014, pp. 65–71. URL `https://www.sciencedirect.com/science/article/pii/S1877050914012873`, complex Adaptive Systems Philadelphia, PA November 3-5, 2014.

[18] Wang, Y., and Liu, J., "Evaluation methods for the autonomy of unmanned systems," *Chinese science bulletin*, Vol. 57, 2012, pp. 3409–3418. URL `https://link.springer.com/article/10.1007/s11434-012-5183-2`.

[19] Bauranov, A., and Rakas, J., "Designing airspace for urban air mobility: A review of concepts and approaches," *Progress in Aerospace Sciences*, Vol. 125, 2021, p. 100726. URL `https://www.sciencedirect.com/science/article/pii/S0376042121000312`.

[20] *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, On-Road Automated Driving (ORAD) Committee, Apr 2021. URL `https://doi.org/10.4271/J3016_202104`.

[21] Radovic, M., "Untangling The 5 Levels of Drone Autonomy," , 2019.

[22] Ilachinski, A., "Artificial Intelligence and Autonomy: Opportunities and Challenges," *Center for Naval Analysis*, 2017, pp. 10–2017. URL `https://apps.dtic.mil/sti/citations/AD1041749`.

[23] Huang, H.-M., Albus, J. S., Messina, E. R., Wade, R. L., and English, R. W., "Specifying autonomy levels for unmanned systems: interim report," *Unmanned Ground Vehicle Technology VI*, Vol. 5422, edited by G. R. Gerhart, C. M. Shoemaker, and D. W. Gage, International Society for Optics and Photonics, SPIE, 2004, pp. 386 – 397. URL `https://doi.org/10.1117/12.552074`.

[24] Huang, H.-M., "Autonomy Levels for Unmanned Systems (ALFUS) Framework: Safety and Application Issues," *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems*, Association for Computing Machinery, New York, NY, USA, 2007, p. 48–53. URL `https://doi.org/10.1145/1660877.1660883`.

[25] "Urban Air Mobility (UAM) Concept of Operations," Tech. rep., Federal Aviation Administration, 2023. URL `https://www.faa.gov/air-taxis/uam_blueprint`.

[26] Chrissis, M. B., Konrad, M., and Shrum, S., *CMMI for Development: Guidelines for Process Integration and Product Improvement*, 3rd ed., Addison-Wesley Professional, 2011.

[27] Niazi, M., Wilson, D., and Zowghi, D., "Organisational Readiness and Software Process Improvement," *Product-Focused Software Process Improvement*, edited by J. Münch and P. Abrahamsson, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 96–107. URL `https://link.springer.com/chapter/10.1007/978-3-540-73460-4_11`.

[28] Ayyagari, M. R., and Atoum, I., "CMMI-DEV Implementation Simplified," *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 4, 2019. URL `http://dx.doi.org/10.14569/IJACSA.2019.0100455`.

[29] Kaliyaperumal, L. N., "The Evolution of Security Operations and Strategies for Building an Effective SOC," *ISACA Journal*, Vol. 5, 2021, pp. 1–7.

[30] ISACA, "ISACA Updates CMMI Model with Three New Domains That Help Organizations Improve Quality and Performance," *ISACA Journal*, Vol. 5, 2023, pp. 1–7.

[31] Gregory, I. M., Campbell, N. H., Neogi, N. A., Holbrook, J. B., Grauer, J. A., Bacon, B. J., Murphy, P. C., Moerder, D. D., Simmons, B. M., Acheson, M. J., et al., "Intelligent contingency management for urban air mobility," *Dynamic Data Driven Applications Systems: Third International Conference, DDDAS 2020, Boston, MA, USA, October 2-4, 2020, Proceedings 3*, Springer, 2020, pp. 22–26. URL `https://link.springer.com/chapter/10.1007/978-3-030-61725-7_5`.

[32] Campbell, N. H., Ilangovan, H. S., Gregory, I. M., and Mikaelian, S. S., "Data Augmentation for Intelligent Contingency Management Using Generative Adversarial Networks," *AIAA SCITECH 2022 Forum*, 2022, p. 0622. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2022-0622`.

[33] Gregory, I. M., Acheson, M. J., Patterson, A. P., Houghton, M. D., Oshin, A., Ackerman, K. A., and Cook, J., "Candidate Performance Metrics for Generalized Control for Autonomous Flight," *AIAA SCITECH 2023 Forum*, 2514. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2023-2650`.

[34] Patterson, A., Ackerman, K. A., Cook, J., Acheson, M. J., and Gregory, I. M., "An L1 Adaptive Control Augmentation for a Lift-Plus-Cruise Vehicle," *AIAA SCITECH 2023 Forum*, 2514. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2023-2541`.

[35] Acheson, M. J., Gregory, I. M., and Cook, J., "Examination of Unified Control Incorporating Generalized Control Allocation," *AIAA Scitech 2021 Forum*, 2021. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2021-0999`.

[36] Sarojini, D., Ruh, M. L., Joshy, A. J., Yan, J., Ivanov, A. K., Scotzniovsky, L., Fletcher, A. H., Orndorff, N. C., Sperry, M., Gandarillas, V. E., Asher, I., Chambers, J. T., Gill, H., Lee, S., Cheng, Z., Rodriguez, G., Zhao, S., Mi, C., Nascenzi, T., Cuatt, T., Winter, T. F., Guibert, A. T., Cronk, A., Kim, H. A., Meng, S., and Hwang, J. T., "Large-Scale Multidisciplinary Design Optimization of an eVTOL Aircraft using Comprehensive Analysis," *AIAA SCITECH 2023 Forum*, 2514. URL `https://arc.aiaa.org/doi/abs/10.2514/6.2023-0146`.

[37] Bhandari, R., Mishra, A. A., and Chakraborty, I., "Genetic Algorithm Optimization of Lift-Plus-Cruise VTOL Aircraft with Electrified Propulsion," *AIAA SCITECH 2023 Forum*, 2514. URL https://arc.aiaa.org/doi/abs/10.2514/6.2023-0398.

[38] Comer, A. M., and Chakraborty, I., "Flight Control System Architecture for Urban Air Mobility Simplified Vehicle Operations," *AIAA SCITECH 2023 Forum*, 2514. URL https://arc.aiaa.org/doi/abs/10.2514/6.2023-0399.

[39] Acheson, M. J., Gregory, I. M., and Cook, J., "Examination of unified control incorporating generalized control allocation," *AIAA Scitech 2021 Forum*, 2021, p. 0999.

[40] Houghton, M., Acheson, M., Patterson, A., Oshin, A., and Gregory, I., "Combined Bernstein Polynomial Optimal Reciprocal Collision Avoidance Differential Dynamic Programming for Trajectory Replanning and Collision Avoidance for UAM Vehicles," 2023. https://doi.org/10.2514/6.2023-2544.

[41] Simmons, B. M., and Murphy, P. C., "Wind Tunnel-Based Aerodynamic Model Identification for a Tilt-Wing, Distributed Electric Propulsion Aircraft," *AIAA Scitech 2021 Forum*, 2514. URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-1298.

[42] Cook, S. A., "The complexity of theorem-proving procedures," *Proceedings of the third annual ACM symposium on Theory of computing*, 1971.

[43] Spirkovska, L., Kulkarni, C. S., Watkins, J., and Martin, L., "Urban Air Mobility Airspace Dynamic Density," *AIAA AVIATION 2022 Forum*, 2022, p. 3403. URL https://arc.aiaa.org/doi/abs/10.2514/6.2022-3403.

[44] Andrés, E., Garcıa-Heras, J., González, D., Soler, M., Valenzuela, A., Franco, A., Nunez-Portillo, J., Rivas, D., Radiši T., and Andraši, P., "Probabilistic Analysis of Air Traffic in Adverse Weather Scenarios," *10th International Conference on Research in Air Transportation (ICRAT)*, 2022, pp. 1–8.

[45] Huaxing, L., "Modelling terrain complexity," *Advances in digital terrain analysis*, 2008, pp. 159–176. URL https://link.springer.com/chapter/10.1007/978-3-540-77800-4_9.

[46] Delahaye, D., García, A., Lavandier, J., Chaimatanan, S., and Soler, M., "Air Traffic Complexity Map Based on Linear Dynamical Systems," *Aerospace*, Vol. 9, No. 5, 2022. URL https://www.mdpi.com/2226-4310/9/5/230.

[47] Ragab, M., Ashary, E. B., Aljedaibi, W. H., Alzahrani, I. R., Kumar, A., Gupta, D., and Mansour, R. F., "A novel metaheuristics with adaptive neuro-fuzzy inference system for decision making on autonomous unmanned aerial vehicle systems," *ISA transactions*, Vol. 132, 2023, pp. 16–23. URL https://www.sciencedirect.com/science/article/pii/S0019057822001598.

[48] Kulkarni, C. S., Corbetta, M., and Robinson, E. I., "Systems Health Monitoring: Integrating FMEA into Bayesian Networks," *2021 IEEE Aerospace Conference (50100)*, 2021, pp. 1–11. https://doi.org/10.1109/AERO50100.2021.9438219.

[49] Klyde, D. H., Schulze, P. C., Mitchell, D. G., Sizoo, D., Schaller, R., and McGuire, R., "Mission task element development process: An approach to FAA handling qualities certification," *AIAA AVIATION 2020 FORUM*, 2020, p. 3285. URL https://arc.aiaa.org/doi/10.2514/6.2020-3285.

[50] Belcastro, C. M., and Neogi, N. A., "Verification and Validation of Safety-Critical Aircraft Systems Operating under Off-Nominal, Contingency, and Emergency Conditions," *Briefing is Part of a Virtual Seminar Series on Advanced V&V Methods for Complex Systems*, 2020. URL https://ntrs.nasa.gov/citations/20205009041.

[51] "Threat and error management (TEM)," Seminar, Canadian Aviation Safety, 2005. URL https://skybrary.aero/articles/threat-and-error-management-tem.

[52] Lehman, S. M., Neogi, N. A., and Slagel, J. T., "Tell Me Why: Challenging Surprise Adequacy as a Metric for Safety-Critical Neural Network Testing," *IEEE/ACM International Conference on Software Engineering-" Software Engineering in Practice" Track*, 2023. URL https://ntrs.nasa.gov/citations/20220014352.

[53] Patterson, A., MacLin, G. L., Acheson, M. J., Tabasso, C., Cichella, V., and Gregory, I. M., "On Hermite Interpolation using Bernstein Polynomials for Trajectory Generation," Tech. Rep. TM-2023-0013467, National Aeronautics and Space Administration, Hampton, VA, USA, 2023.

[54] Mclarney, E., Gawdiak, Y., Oza, N., Mattmann, C., Garcia, M., Maskey, M., Tashakkor, S., Meza, D., Sprague, J., Hestnes, P., et al., "NASA Framework for the Ethical Use of Artificial Intelligence (AI)," Tech. rep., NASA, 2021. URL https://ntrs.nasa.gov/citations/20210012886.