Commun Transp Res

Editorial

Scenario-based testing platform for urban air mobility traffic management systems

Cong Zhang[1], Shulu Chen[2,✉], Kai Wang[1,2]

Cite this article: Zhang C, Chen S L, K Wang. *Commun Transp Res* 2026, **6**(1): 9640007. https://doi.org/10.26599/COMMTR.2026.9640007

1 Overview of UAM traffic management systems

With population growth and accelerated urbanization, travel demand has increased, and cities have become increasingly congested. The emergence of urban air mobility (UAM) offers a promising solution by extending transportation into the third dimension to mitigate congestion and improve traffic efficiency. Today, three-dimensional urban low-altitude transportation through UAM is gradually becoming a reality. For example, Joby Aviation and Volocopter have recently validated the technical feasibility of UAM operations through electric vertical take-off and landing (eVTOL) flight tests. In addition, Tsinghua University exemplifies air–ground mobility (Lv et al., 2025) through the development of a compound-wing modular flying car. Compared with traditional commercial aviation, UAM is more flexible and supports a wider range of missions (e.g., urban commuting, emergency medical transport, and security inspection). However, unlike traditional commercial aviation, which operates in a relatively simple and structured flight environment, UAM operates in a more congested and heterogeneous environment. This environment involves various aircraft types as well as multiple constraints, such as communication interference, intruders such as birds and unlawful drones, and weather factors. These challenges may increase safety risks and operational uncertainties. Therefore, developing a comprehensive UAM traffic management system to ensure flight safety and operational efficiency is an urgent need.

There are several existing UAM traffic management systems. For example, ANRA Technologies provides an airspace management software platform for advanced air mobility called ANRA UAM (PSU) (ANRA Technologies, 2025), and Eve Air Mobility has completed its urban air management (Urban ATM) prototype to ensure safe, scalable integration of urban air mobility operations. These existing systems generally include strategic conflict management and tactical deconfliction (Chen et al., 2024). On the one hand, strategic conflict management focuses on decision-making processes that balance air traffic demand. It emphasizes long-term planning involving scheduling (e.g., terminal-area scheduling policies, vertiport scheduling algorithms) and flight planning for UAM operations. On the other hand, tactical deconfliction deals with near-term aircraft maneuvers, with representative systems such as the Traffic Alert and Collision Avoidance System (TCAS) and the Airborne Collision Avoidance System (ACAS). The existing algorithms and management systems are diverse and complex; thus, it is necessary to rigorously validate these systems' safety, efficiency, and operational robustness before real-world implementation (Lee et al., 2024).

2 UAM operational challenges and possible solutions

Developing a testing procedure is the first step toward validating the UAM traffic management system. Testing procedures for conventional air traffic management (i.e., commercial and general aviation) have been established through certification standards for many years. However, one limitation of such frameworks is the lack of "intelligence", as they rely heavily on human management, making them vulnerable to human errors and inefficiencies. Notably, low-altitude air traffic has inherited many of general aviation's standards and regulations, but the situation that low-altitude aviation faces is much more complex than that of general aviation. The aforementioned high-density and heterogeneous flying environments impose additional requirements on low-altitude traffic management, demanding more automated validation procedures. To meet these challenges, the testing framework for UAM traffic management systems must evolve toward greater intelligence to build safer and more efficient UAM operations through robust evaluation. However, existing validation procedures remain fragmented and largely unsystematic (Federal Aviation Administration, 2023). For example, most studies addressed only isolated problems such as conflict mitigation or traffic management. As a result, there is still a lack of (1) a unified scenario library for benchmarking, (2) systematic generation of rare safety-critical scenarios, and (3) a closed-loop framework to improve system performance iteratively.

A comprehensive scenario-based testing platform is needed to evaluate UAM traffic management systems. There are two reasons for this need. First, unlike commercial aviation, which has decades of real-world experience, low-altitude operations are still new and

[1] School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China. [2] State Key Laboratory of Intelligent Green Vehicle and Mobility, Tsinghua University, Beijing 100084, China.

✉ Corresponding author. E-mail: csl_uam@mail.tsinghua.edu.cn

Received: October 27, 2025; Revised: November 28, 2025; Accepted: December 19, 2025

lack large-scale validation data. Second, as aviation is a safety-critical domain in which even small errors may result in serious outcomes, the testing framework must have strong safety guarantees. Validating such massive scenarios in the real world operations would require long operational experience, which is costly and inefficient. This motivates us to introduce a framework addressing these challenges intelligently, which includes scenario library generation, evaluation of generated scenarios, and iterative improvement of the management system.

The operational challenges and management solutions are illustrated in Fig. 1. These challenges arise from diverse operational scenarios. The scenario space is designed to be comprehensive, covering both normal scenarios and safety-critical scenarios such as conflicts, severe weather, communication/navigation/surveillance (CNS) failures, and final approach and takeoff area (FATO) outages. Each scenario is paired with corresponding management solutions (e.g., conflict management, demand–capacity balancing, and weather avoidance) using the chosen testing scenarios for evaluation. Together, these challenges and solutions form the basis for systematic evaluation. Moreover, the platform systematically measures safety metrics (e.g., the number of loss-of-well-clear (LoWC) and the number of near mid-air collisions (NMACs)) and time metrics (e.g., ground delay, on-time arrival rate, flight time) to evaluate whether the strategies achieve safe and efficient operations. The performance metrics for evaluation under different scenarios are different. Table 1 provides a systematic map linking each scenario to the safety and/or time-based metrics used for assessment.

3 Scenario-based testing platform for UAM traffic management

To validate UAM traffic management systems, a testing platform is required to support large-scale, scenario-based validation, including both normal scenarios and safety-critical scenarios. Fig. 2 highlights how the platform conducts validation through a scenario-based, closed-loop framework. Using tools such as generative artificial intelligence (AI), it builds a cost-effective yet comprehensive scenario library that covers various conditions. To further stress-test the system, sampling methods (e.g., importance sampling (Feng et al., 2021) can be applied to draw scenarios from the generated scenario library, amplifying rare scenarios to accelerate evaluation. In addition, the generated scenarios should be evaluated in terms of their realism, diversity, and causal validity, considering generative AI, rules, and causal inferences. Instead of evaluating individual flight operations in isolation, the focus should be on the whole system. Once the chosen generated scenarios are tested in the UAM schedule management system, the testing platform should provide diagnostic feedback by identifying bottlenecks and failure modes to improve its scheduling strategies. The following subsections outline the key steps in this framework.

3.1 Testing scenario generation

To validate UAM traffic management systems, the scenario generation task should focus on testing at the mission level (e.g., sequencing of UAV, drone, and/or eVTOL missions) rather than the trajectory level. A comprehensive scenario library must consider varying loads and resource availabilities to evaluate and optimize the performance of the management system for conflict prevention, efficiency improvement, and system robustness verification. Meanwhile, generative models and rule-based constraints are used to synthesize safety-critical scenarios. For example, the platform can create scenarios where a CNS outage coincides with an emergency medical mission, stressing the system's conflict resolution and priority management modules. Importance sampling is a common method that can be applied to increase the frequency of these critical scenarios in simulations. Following civil aviation safety objectives (e.g., FAA AC 25.1309

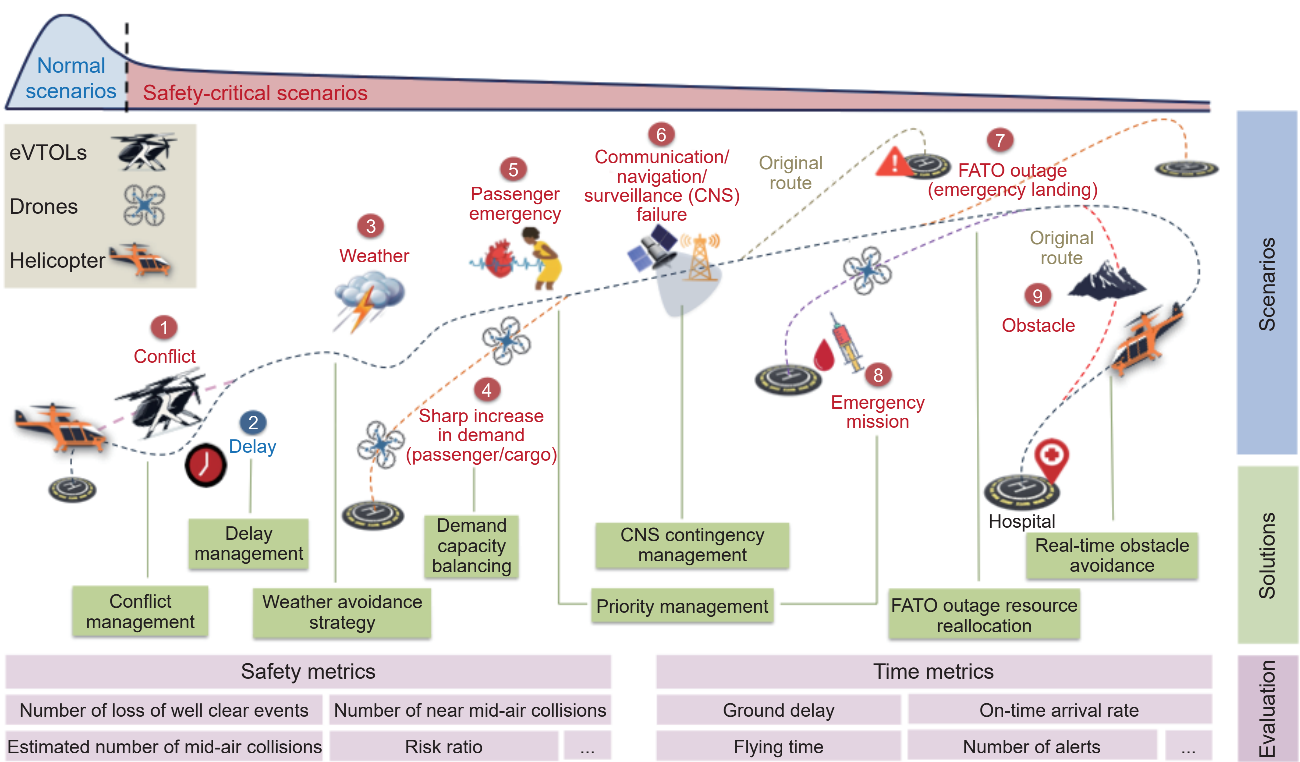

Fig. 1 UAM operational challenges and management solutions.

Table 1 Scenario classification and corresponding performance metrics for UAM testing

Scenario category	Representative scenario	Metric
Conflict	Loss of separation	Number of LoWC events NMAC
	Near-mid-air collision	Estimated number of mid-air Collisions Risk ratio
Delay	Vertiport congestion	Ground delay Flying time
	Route resequencing	On-time arrival rate
Weather	Strong wind/gusts Thunderstorms Low visibility Heavy rain	Ground delay Flying time Number of alerts Mission completion rate
Sharp increase in demand	Passenger demand surge Cargo demand surge	Ground delay On-time arrival rate Number of alerts
Passenger emergency	Passenger medical emergency	Number of alerts Flying time On-time arrival rate
CNS failure	Communication failure Navigation failure Surveillance failure	Number of alerts Flying time
FATO outage (Emergency landing)	Blocked landing pad Emergency diversion Reduced vertiport capacity	On-time arrival rate Flying time Number of alerts
Emergency mission	Time-sensitive emergency task (e.g., medical delivery)	Flying time Mission completion rate
Obstacle	Real-time obstacle appearance Unexpected static obstacle	Number of alerts Minimum obstacle clearance

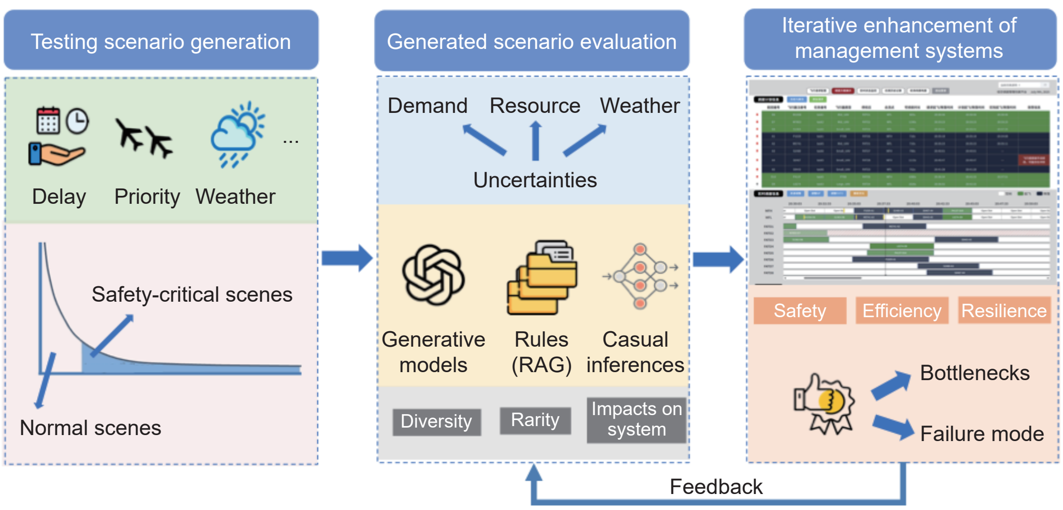

Fig. 2 Scenario-based testing framework.

(Federal Aviation Administration, U.S. Department of Transportatio, 2024) and SAE ARP4761), catastrophic failure conditions should occur with a probability of no greater than 10^{-9} per flight hour. Conducting sufficient validation in real-world operations to meet this level of safety would be prohibitively costly and inefficient. UAM research typically adopts this target level of safety (TLS) as a long-term goal for high-density low-altitude operations. Therefore, importance sampling and other rare-event amplification techniques are essential to achieve statistically significant evaluations within a feasible timeframe.

3.2 Generated scenario evaluation

Before using the generated scenarios to test the UAM traffic management system, the generated scenarios should be comprehensively evaluated to ensure that they are representative, diverse, and operationally meaningful. First, a unified scenario classification (e.g., nominal and safety-critical) and consistent evaluation criteria need to be established. Second, operational uncertainties should be assessed systematically, including weather conditions (e.g., wind fluctuations, thunderstorms), resource availability (e.g., FATO availability reductions, CNS failures), and pop-up demand (e.g., passenger/cargo sharp increase). Data-driven and generative models can be used to generate large-scale data at a low cost, injecting disruptions such as severe weather, resource failures, and heavy traffic to cover rare but challenging scenarios. Beyond evaluating the generated scenarios, the evaluation framework must also ensure that these scenarios support comprehensive comparisons across different management systems. To achieve this, each scenario class should be associated with a baseline performance profile derived from a large number of simulations over its parameter space, which characterizes the inherent difficulty of that scenario class. Scenario sampling is then designed to ensure accurate evaluation relative to these baselines by combining representative real-world conditions with amplified safety-critical cases through importance sampling. To further distinguish the performance differences between management systems, the sampling mechanism should cover the conditions that are most likely to lead to divergent behaviors and/or the conditions that have the largest performance variance. Together, these methods make evaluation more comprehensive and explainable, reduce manual efforts, and support large-scale testing and rapid iterative improvement.

3.3 Iterative enhancement of management systems

Once the scenario library has been generated and evaluated, it can be applied to the UAM traffic management system as a test library to improve its scheduling strategies (e.g., conflict avoidance). Using the aforementioned safety and time metrics (Fig. 1), evaluation results can be obtained and compared across different UAM management systems. Next, by analyzing system performance with the generated scenarios, diagnostic analysis can identify bottlenecks and failure modes. The management system is then iteratively improved using these insights, creating a feedback loop between system improvement and human evaluation. The improved system can then be progressively integrated into real-world operations following thorough simulation-based validation within the management platform.

4 Conclusions

In this work, we outlined a comprehensive scenario-based testing platform for evaluating UAM traffic systems under diverse conditions. Unlike existing ad hoc simulations, the proposed platform integrates scenario generation, scenario evaluation, and iterative enhancement of the management system. By amplifying safety-critical scenarios and applying importance sampling, the platform enables statistically significant and efficient validation of system safety and robustness prior to large-scale deployment. Furthermore, the platform combines scenario generation, policy testing, and human-in-the-loop evaluation to improve safety and efficiency and provides a benchmark for performance comparisons. In conclusion, the suggested testing platform is an essential first step toward developing an open and transparent approach to UAM system evaluation. It could accelerate the safety assessment of urban air mobility for the upcoming generation of transportation systems by bridging the gap between research, operations, and regulation. Future work will expand the scenario library, add economic and network-level metrics, and support multicity comparisons to improve the platform's applicability.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

ANRA Technologies, 2025. An airspace management software platform for operating UAM aircraft in an advanced mobility environment. https://www.anratechnologies.com/home/anra-uam-psu

Chen, S., Evans, A. D., Brittain, M., Wei, P., 2024. Integrated conflict management for UAM with strategic demand capacity balancing and learning-based tactical deconfliction. *IEEE Trans Intell Transp Syst*, **25**, 10049–10061.

Federal Aviation Administration., 2023. Urban Air Mobility (UAM) Concept of Operations, Version 2.0. https://www.faa.gov/sites/faa.gov/files/Urban-Air-Mobility-Concept-of-Operations-2.0.pdf

Federal Aviation Administration, U.S. Department of Transportatio, 2024. 14 C.F.R. § 25.1309—154 Equipment, systems, and installations. Electronic Code of Federal Regulations (eCFR). https://www.ecfr.gov/current/title-14/part-25/section-25.1309

Feng, S., Feng, Y., Yu, C., Zhang, Y., Liu, H. X., 2021. Testing Scenario Library Generation for Connected and Automated Vehicles, Part I: Methodology. *IEEE Trans Intell Transp Syst*, **22**, 1573–1582.

Lee, Y., Gwak, C., Bae, S., Park, K. R., Lee, H. J., Lee, J. W., 2024. Functional analysis and validation of a UAM integrated flight management (UIFM) system. *Int J Aeronaut Space Sci*, **25**, 749–758.

Lv, D., Wang, K., Chen, S., Qu, X., 2025. Flying cars and urban air mobility: Redefining cities in three dimensions. *Commun Transp Res*, **5**, 100213.

Cong Zhang is a Postdoctoral Research Fellow of the Emerging Transportation Solution Lab at the School of Vehicle and Mobility in Tsinghua University. She received the Ph.D. degree from Purdue University in 2025. Her research interests include testing platforms in air traffic management and large language models in transportation.

Shulu Chen is an Assistant Research Fellow at the State Key Laboratory of Intelligent Green Vehicles and Mobility, Tsinghua University. He earned the Ph.D. degree from George Washington University in 2025. His research focuses on reinforcement learning and optimization with applications in air traffic management.

Kai Wang is an Assistant Professor in smart transportation at the School of Vehicle and Mobility in Tsinghua University. He received the Ph.D. degree from The Hong Kong Polytechnic University (PolyU) in 2019. His research focuses on optimization under uncertainty, data-driven decision-making, and large-scale optimization. In particular, he is interested in relevant applications in smart transportation, logistics management, and emerging mobility systems.