Journal of Highway and Transportation Research and Development (English Edition)
Open Access | Editor-in-Chief: Wanheng Li, Daming Lin | ISSN 2095-6215

J. highw. transp. res. dev.

https://doi.org/10.26599/HTRD.2026.9480121

Article Type: Research Article

Risk Assessment for Urban eVTOL Operations and Air-Ground Cooperative Hazard Avoidance Mechanism

Yuetan Zhang[1,*], Honghai Zhang[1,**], Wei Pan[1], Yulin Yuan[2]

[1] *College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China*

[2] *Nanjing Tiancheng Transportation Research Institute Company Limited, Nanjing, Jiangsu211500, China*

*Email: Yuetan Zhang, zhang23@nuaa.edu.cn;

**Corresponding author: Honghai Zhang, honghaizhang@nuaa.edu.cn

Received: date(20-Oct-2026); Revised:date(27-Oct-2026); Accepted: date(04-Nov-2026)

Just Accepted

Abstract: This study develops an eight-indicator framework for describing low-altitude operational risk and a candidate air-ground response architecture. The evidence is drawn from three distinct supplementary datasets: 1,000 source-labelled sortie records (772 retained after exposure and duration quality control; 740 operational and 32 simulated), a trajectory/alert slice aggregated to 123 occupied 500 m × 500 m by 5-min cells, and 6,079 occupied 5-min windows from 55 days of operational records. These datasets are analysed separately. In the 55-day series, the adjusted negative-binomial model estimated a rate ratio of 3.32 per one-level increase in mean sortie risk score (95% CI 3.13–3.52), while density had an independent rate ratio of 1.11 (95% CI 1.07–1.14). In the 700-cell trajectory pipeline, the score–alert association was not statistically distinguishable from unity (RR 1.0026 per point, 95% CI 0.9957–1.0095; p = 0.464). Near-perfect rolling-window AUCs were attributable to the close constructional dependence between sortie risk scores and event records and are not interpreted as independent predictive validation. Because the observations are low-altitude UAV operations and simulation records rather than commercial passenger eVTOL flights, the findings support a pilot

assessment workflow, not eVTOL accident-risk estimates or validated operational thresholds.

Keywords: eVTOL; urban air mobility; safety governance framework; risk assessment; Security Governance; negative binomial regression

1 Introduction

As urban air mobility (UAM) progresses from proof-of-concept to demonstration operations and commercialization, electric vertical take-off and landing aircraft (eVTOL), as its core carrier, are being gradually integrated into the comprehensive urban transportation system. While eVTOL operations enhance urban transportation efficiency and alleviate ground traffic congestion, the complexity of their operating environment-including high-density traffic flow in low-altitude airspace, occlusion and interference from urban buildings, rapid changes in meteorological conditions, and high exposure of ground populations and critical infrastructure-poses unprecedented systemic safety challenges. The focus of safety governance has expanded from single-dimension aircraft airworthiness and onboard collision-avoidance technology to a cross-domain, cross-level comprehensive governance issue involving dynamic airspace management, Vertiport capacity coordination, ground risk control, and multi-stakeholder emergency response coordination.

Currently, research on eVTOL operational safety has predominantly focused on real-time detection and resolution algorithms for air traffic conflicts[1], single-aircraft reliability analysis[2], or risk simulation based on specific scenarios, lacking a multidimensional and quantifiable risk assessment framework capable of simultaneously reflecting aerial exposure probability, ground consequence severity, and overall system vulnerability and resilience[3]. Existing management practices have treated air traffic management (ATM) and ground operation management (such as Vertiport scheduling) as separate domains, and a set of rules capable of automatically and efficiently translating real-time risk states into cooperative hazard-avoidance actions spanning both airspace and ground operations has yet to be established. The process also lacks sufficient transparency and auditability, making it difficult to meet the stringent requirements for liability determination and safety regulation demanded by future large-scale, commercialized operations[4].

The study makes three bounded contributions. First, it specifies an eight-indicator framework and documents a reproducible reference implementation for trajectory-to-cell aggregation and fuzzy clustering. Second, it translates candidate risk bands into a proposed airspace/vertiport response matrix with explicit authority and release conditions. Third, it tests associations and robustness across separately defined operational and simulation datasets. The empirical evidence concerns recorded UAV operational proxies and does not constitute passenger-eVTOL flight validation.

Although this study draws upon general management experience regarding unmanned aerial vehicles (UAVs) and low-altitude airspace in its literature review and policy analysis, the core research object is strictly limited to electric vertical take-off and landing aircraft (eVTOL) used for urban passenger transport or bulk logistics. Compared with micro/light

Just Accepted

consumer-grade UAVs, manned/load-carrying eVTOLs are subject to far more stringent requirements regarding takeoff weight, airworthiness standards, system redundancy, and airspace access conditions, and their operations exhibit characteristics of "high vulnerability" and "extremely low risk tolerance." Consequently, conventional collision-avoidance rules designed for micro/light UAVs cannot be directly transposed to eVTOLs; a tailor-made, air-ground cooperative comprehensive governance mechanism based on multidimensional, forward-looking risk perception must instead be developed specifically for this category of aircraft.

2 Research Progress on Low-Altitude Safety Governance

As shown in Figure 1, the policy background, academic research, and practical frameworks related to urban eVTOL safety governance are systematically reviewed to establish the theoretical foundation and practical basis for the proposed multidimensional risk assessment and air-ground cooperative rule system. The content covers the policy evolution of low-altitude safety management, the current state of research on urban air traffic, systematic governance models, and the limitations of existing studies, thereby clarifying the positioning and contributions of this research.

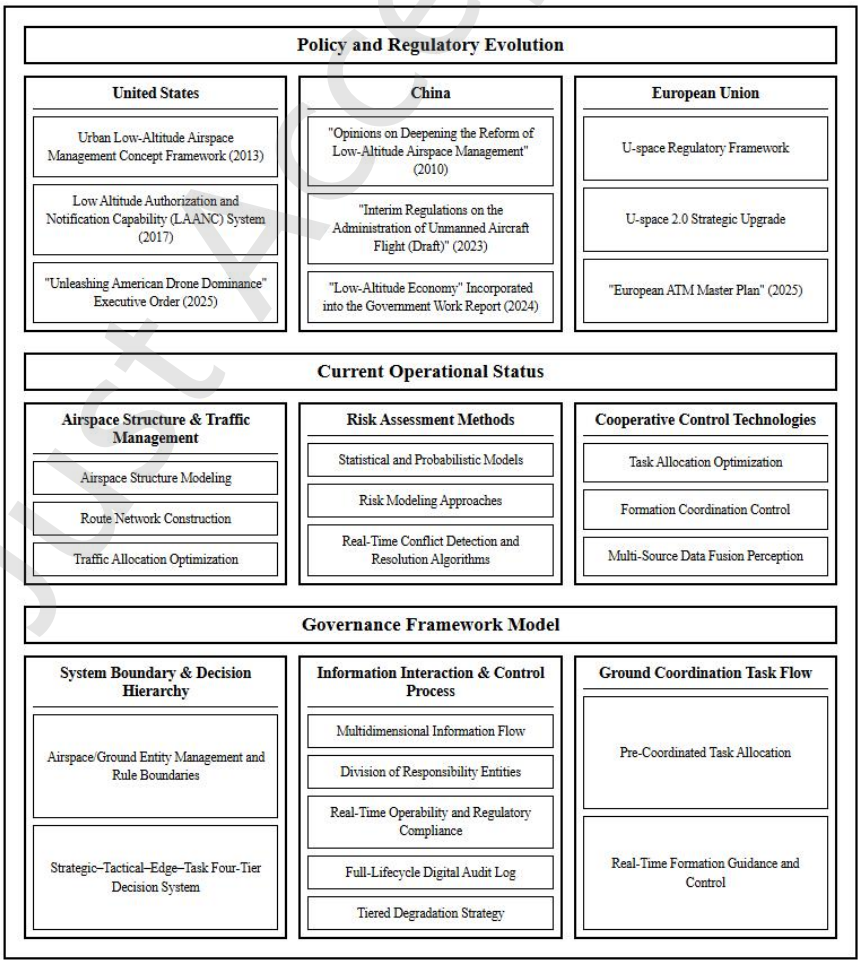


Fig. 1 Research progress on domestic and international low-altitude safety governance.

Just Accepted

2.1 Evolution of Low-Altitude Safety Management Policies

Major aviation economies around the world are actively advancing reforms and innovations in low-altitude airspace management policies to accommodate the development of new low-altitude aircraft such as eVTOLs.

The Low Altitude Authorization and Notification Capability (LAANC) system, implemented in 2017, achieved automated access authorization for controlled airspace. In 2023, the Federal Aviation Administration (FAA) of the United States released the Advanced Air Mobility (AAM) Implementation Plan, further specifying a roadmap for achieving large-scale commercial eVTOL operations by 2028. The European Union, through the U-space framework, has systematically promoted the integration of unmanned aerial vehicles and air traffic management. According to the European ATM Master Plan (2020 Edition) and subsequent regulations (EU 2021/664), the core objective of U-space is to provide services including digital geofencing and dynamic airspace authorization, so as to support the high-density operation of novel aircraft. China's low-altitude safety management is accelerating onto a fast track of legalization and standardization. The Interim Regulations on the Administration of Unmanned Aircraft Flight, promulgated in 2023 by the State Council and the Central Military Commission (effective from January 1, 2024), marked a key advancement at the regulatory level. In 2024, the "low-altitude economy" was incorporated into the Government Work Report for the first time, and the Civil Aviation Administration of China (CAAC) subsequently issued a series of guiding opinions to promote the revitalization of airspace resources and the implementation of airworthiness certification standards for eVTOLs.

2.2 Current Status of Research on Urban Low-Altitude Airspace Operations

The academic community has conducted multi-faceted research on the safety and operational management of urban low-altitude airspace, mainly focusing on airspace design, conflict risk assessment, and traffic management methods.

In terms of airspace structure and traffic management, scholars have proposed various models to reduce management complexity[5,6,7,8,9,10]. For example, Wang et al. proposed an optimization framework for airspace traffic allocation[5]; Zhai et al. constructed a low-altitude public route network based on global subdivision grid theory[7]; and Lee et al. systematically analyzed and compared the advantages and disadvantages of three airspace structure models-Air-Matrix, Air-Network, and AirTube[8]. These studies have provided a theoretical foundation for structured airspace operations.

In terms of safety risk assessment, research methods are becoming increasingly diversified[11,12,13,14,15,16,17]. Some studies are based on statistical and probabilistic models, such as Underhill et al., who used Bayesian networks to evaluate the applicability of uncorrelated airspace encounter models at low altitudes[13]; and Sun et al., who conducted UAV accident risk assessment based on data-driven Bayesian networks[11]. Another category of research focuses on real-time risk fields and conflict resolution, such as Zhao et al., who proposed a flight risk field model for advanced low-altitude traffic systems[3]; and Zhang et al., who studied three-dimensional path planning and real-time conflict resolution algorithms in complex urban low-altitude environments[18]. In addition, some scholars have focused on

Just Accepted

assessment methods under uncertainty[15,16], such as Wang et al., who proposed a low-altitude UAV operational risk assessment method that accounts for Monte Carlo uncertainty[16]. Related trajectory-planning and collision-avoidance methods are also reported in [19,20,21,22].

Despite the preliminary progress achieved by existing research, urban low-altitude airspace management still faces systemic challenges: the operating environment is complex and highly variable, with dynamic and static disturbances from buildings, weather conditions, and bird flocks; the diversification of carrier vehicles means that the airworthiness and operational models of novel aircraft such as eVTOLs have not yet been validated at scale; the requirements for refined planning are stringent, necessitating multidimensional coordination among the airspace network, takeoff/landing facilities, urban space, and ground traffic; and the regulatory model remains unclear, involving multiple departments such as air traffic control, transportation, and public security, with the cooperative responsibilities and mechanisms for air-ground integration still not well defined. Relevant discussions of low-altitude economy, UAM management, and integrated operations are provided in [23,24,25,26,27].

2.3 Research on Urban Low-Altitude Safety Governance Mechanisms

2.3.1 System Governance Boundaries and the Hierarchical Responsibility-Based Decision-Making System

Effective governance must first clarify the boundaries of system operation and decision-making. The boundaries of the eVTOL operational system encompass not only the physical boundaries of airspace, but also extend to structured routes and rule-based spaces, operational environments constrained by dynamic geofencing and environmental limiting factors, ground risk control boundaries based on the exposure assessment of population and critical infrastructure, and ground infrastructure boundaries constrained by resource capacity[28,24]. The corresponding decision-making process should follow a hierarchical and coordinated logic, spanning the entire process from macro-level strategic planning to micro-level real-time control, and is specifically constructed as a complete system comprising the following four tiers: the macro strategic management layer, responsible for airspace network planning and safety margin setting; the meso tactical management layer, responsible for flight plan approval and dynamic traffic flow control; the regional edge management layer, i.e., local management of takeoff and landing areas; and the micro individual management layer, responsible for real-time conflict resolution. This system ensures comprehensive coverage ranging from long-term planning to millisecond-level response.

Cooperative governance relies on clear information flow and the delineation of responsibilities. Multidimensional information interaction must achieve shared situational awareness, computational task offloading, control instructions and status reporting, as well as the efficient circulation of airspace authorization, conflict alerts, and meteorological intelligence within traffic management services[24,29,27]. On this basis, the boundaries of responsibility for entities at each level must be clearly defined: the national/regional regulatory layer is responsible for rule-making and certification; the service provision layer is responsible for dynamic authorization and surveillance/alerting; the operation and execution layer bears

Just Accepted

primary responsibility for safe operations; and the infrastructure layer must ensure the performance and reliability of services such as communication and navigation.

2.3.2 Research Progress on Trustworthy Governance and Air-Ground Cooperative Control

To achieve trustworthy governance, a system must possess auditability as well as implementability and robustness under non-ideal conditions[30]. Auditability requires full-lifecycle, tamper-proof logging of key decision-making elements, including "data version-rule version-trigger cause-executed action-response time-outcome record." Implementability, in turn, requires that system design anticipate scenarios such as data loss or communication delay, and incorporate a tiered degradation operating mode-for example, automatically expanding safety separation under high communication latency, or triggering a preset emergency return-to-base or forced landing procedure in the event of prolonged data interruption-so as to maintain a minimum level of safety.

The concept of air-ground cooperative control originates from the military domain; in the civilian low-altitude sector, its core refers to the coordinated and consistent actions taken by ground systems and unmanned aerial vehicles to accomplish shared missions. Current research is primarily divided into two stages: the pre-task cooperative allocation stage, in which the core problem is typically modeled as a multi-constraint optimization problem that simultaneously accounts for economic cost, operational efficiency, and safety performance; and the in-task formation cooperative control stage, in which research mainly focuses on key technologies such as consensus control within formations, formation keeping, and dynamic reconfiguration[25,31]. However, existing research has largely concentrated on pre-task planning or intra-formation coordination, and there remains a lack of systematic mechanism design regarding how to dynamically translate real-time risk assessment results into rule-based cooperative hazard-avoidance instructions spanning the "airspace-ground" domains, and how to embed such instructions into an auditable governance process. Air-ground trajectory planning and emergency transport optimization provide further application examples[32,33].

2.4 Research Limitations and Positioning

In summary, when examining urban eVTOL safety governance, existing research still leaves several deep-seated, root-cause pain points that require further elucidation and deeper investigation:

(1) Multidimensional fragmentation of institutional and operational physical boundaries: Traditional low-altitude traffic management has predominantly focused on single-aircraft collision avoidance or airborne conflict resolution, resulting in a severe disconnect-both in physical boundaries and management mechanisms-between "high-frequency aerial conflict perception" and "fine-grained ground Vertiport resource provisioning (takeoff/landing scheduling)." This institutional fragmentation causes risk perception and cooperative control measures to be frequently disjointed, preventing the formation of an adaptive closed-loop governance mechanism.

Just Accepted

(2) Cross-domain multi-source data silos and the absence of empirical calibration: Cross-domain multi-source data-including real-time trajectories, meteorological conditions, alerts, and infrastructure information-lack deep integration under a unified spatiotemporal reference frame. Due to the scarcity of real collision data at the early stage of low-altitude economy development, existing safety models struggle to perform dynamic empirical calibration using safety proxy signals from actual operational systems, leaving the governance process mired in a "black-box" dilemma with insufficient transparency.

(3) Nonlinearity of dynamic risk evolution and bottlenecks in mathematical modeling: Under high-density low-altitude operating conditions, the systemic risks faced by urban eVTOLs exhibit highly nonlinear, continuously evolving, and boundary-blurred characteristics. Traditional risk assessment approaches largely rely on a coarse "hard-boundary" classification mode, which struggles to accurately capture the flexible, gradual transitions of risk states, thereby impeding the adaptive and auditable circulation of the "risk–trigger–action" closed-loop governance process.

In response to the aforementioned deep-seated root-cause bottlenecks, as shown in Figure 2, this study is positioned to construct an integrated urban eVTOL comprehensive safety governance mechanism combining "multidimensional risk assessment–air-ground cooperative rules–auditable implementation." In order to break through the limitations of existing research, the core innovations of this study are closely and individually aligned with the three deep-seated root-cause pain points identified above, with the specific positioning as follows:

①Breaking the fragmentation between aerial traffic conflict perception and ground Vertiport support resources at the physical and jurisdictional boundary level. This study pioneers a cross-domain cooperative closed-loop governance architecture that flexibly links high-frequency aerial situational awareness with ground resource scheduling, achieving "air-ground integrated" cooperative disposition.

② The study addresses the scarcity of accident observations by analysing recorded operational proxy signals, including conflict alerts and abrupt-heading-change indicators. These signals are not treated as accidents, and the reported models estimate proxy-event associations only.

③ The study evaluates fuzzy classification and a reference adaptive-genetic FCM implementation. In the available 700-cell trajectory sample, AGA-FCM and ordinary FCM reached the same reported XB value (0.085257); therefore, the evidence supports repeatability of the implementation, not a demonstrated clustering advantage of AGA over FCM.

The framework is presented as a pilot governance and analysis design. Its parameters and action thresholds require external validation before use in operational eVTOL decision-making.

Just Accepted

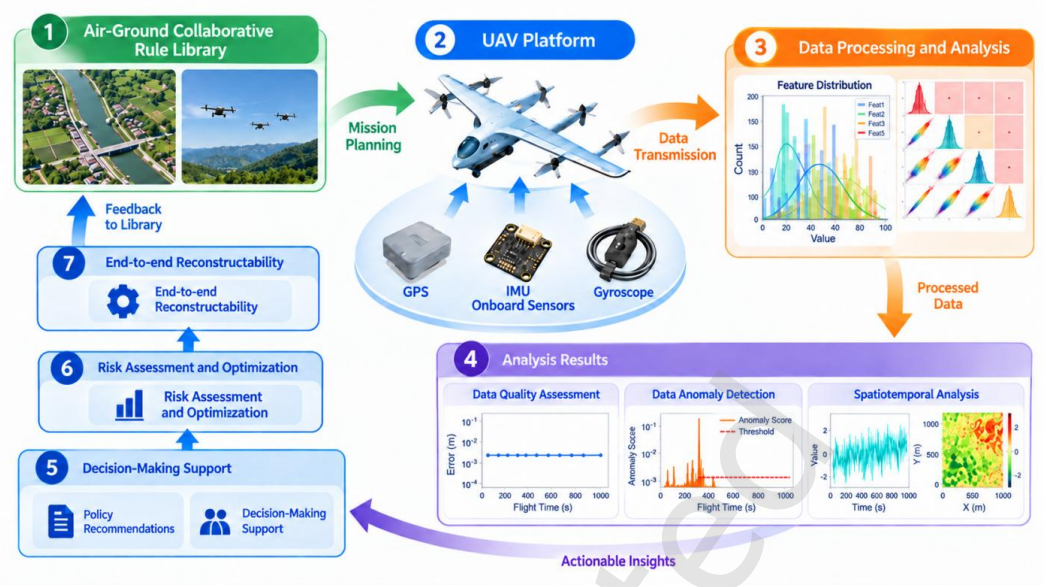


Fig. 2 eVTOL operational risk assessment and collaborative risk mitigation mechanism.

3 Data and Research Scenarios

The supplied supplementary archive adds row-level exports, processing scripts, and analysis outputs. These materials support three distinct, reproducible datasets, which must not be conflated with the manuscript's original 785-sortie/64-unit analysis: (A) 1,000 source-labelled sortie records, of which 772 pass the positive-exposure and ≤24-h duration screen (740 operational and 32 simulated); (B) a capped trajectory/alert slice yielding 123 occupied 500 m x 500 m by 5-min cells across 47 grids and 29 windows; and (C) a separate 55-day series yielding 6,079 occupied 5-min windows. The archive's capped exports do not independently reproduce the original 785/64 cohort. Its exact reconciliation should use the original processing records, while the supplemental results are reported under their own sample definitions.

3.1 Data Sources and Inventory

This study integrates multi-source heterogeneous data covering real-time operations, infrastructure, environment, and population, in order to comprehensively capture the risk exposure, incident consequences, and resilience of the urban air mobility system. Table1 lists the main data categories, their sources, governance relevance, and specific use in this study.

Just Accepted

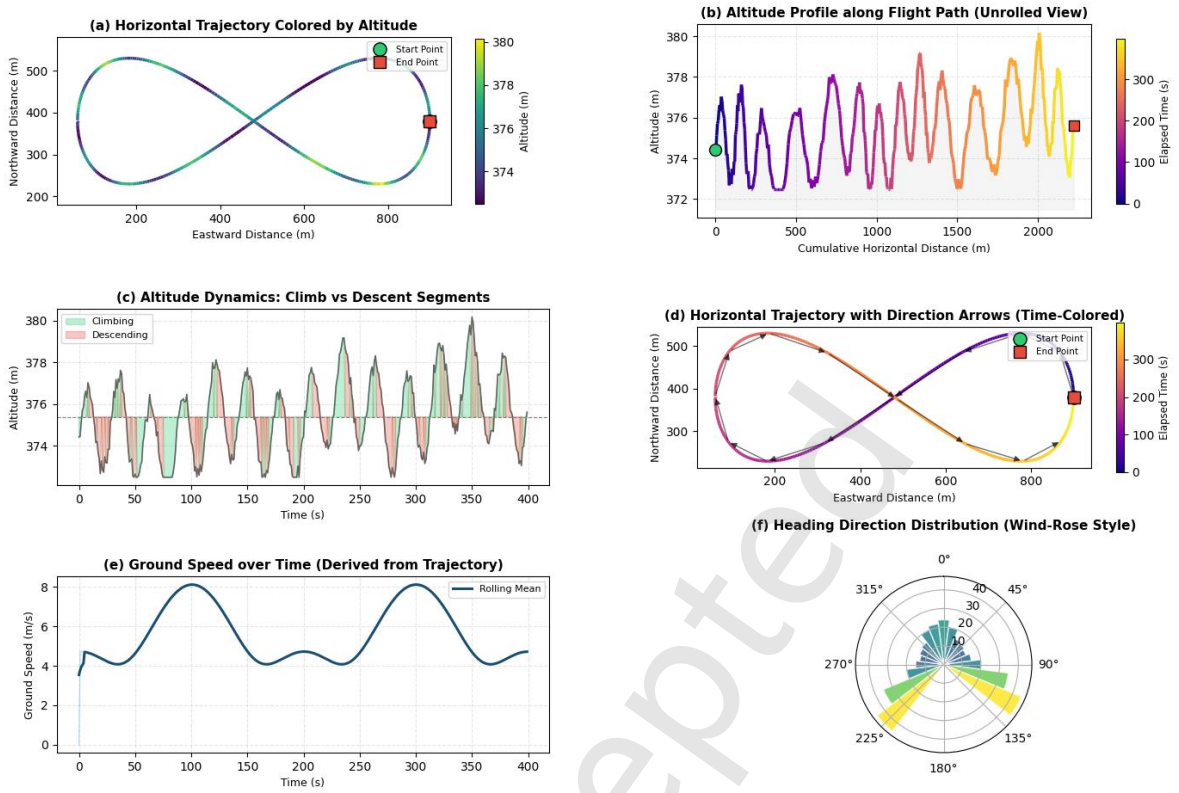


Fig. 3 Flight trajectory diagram.

Table 1 Data inventory, governance relevance, and intended use.

Data category	Specific data item	Source	Update frequency	Governance relevance	Research use
Flight operations data	Real-time trajectory (latitude, longitude, altitude, speed, heading)	UTM / operator real-time data stream	Second-level	Airspace exposure calculation, conflict detection	Computing density, complexity, encounter rate
	Planned trajectory and flight intent	Flight planning system	Pre-flight / real-time updates	Risk prediction, decision support	Used for risk prediction and cooperative planning
	Status (takeoff, landing, cancellation)	Vertiport operations system	Event-driven	System status and capacity monitoring	Computing takeoff/landing throughput, delay, cancellation rate
Safety monitoring data	Conflict alerts (TCAS/UTM alerts)	ATM/UTM monitoring logs	Real-time	Direct safety signal	Risk model calibration (Y~u,t~)
	Separation infringement events (lateral, longitudinal, vertical)	Surveillance system records	Event-driven	Safety status indicator	Risk calibration and rule-effectiveness evaluation

Just Accepted

Data category	Specific data item	Source	Update frequency	Governance relevance	Research use
	Route/corridor network structure	Airspace design documents	Static / periodically updated	Airspace cell delineation	Defining spatiotemporal unit u (corridor link)
Infrastructure and airspace data	Vertiport layout, capacity, availability status	Vertiport management database	Near real-time	Ground capacity and vulnerability	Computing Vu,t, triggering ground-side rules
	Restricted zones, no-fly zones, sensitive areas	Geographic information database	Static / dynamically updated	Airspace constraints and consequence weighting	Affecting Cu,t and rule actions
	Weather (wind speed, visibility, precipitation)	Meteorological service API	Hourly / real-time	Risk exposure correction facto	Control variable Zu,t
Environmental and external data	Ground population distribution (gridded population density)	Public security population data	Annual / monthly updates	Ground consequence weighting	Computing Cu,t
	Critical facilities (hospitals, schools, transportation hubs)	Urban GIS platform	Static / periodically updated	Consequence sensitivity weighting	Adjusting Cu,t weighting
Operational logs and metadata	Data collection timestamp, version number	Logs of each system	Real-time	Audit and reproducibility	Supporting data provenance and version management

Just Accepted

3.2 Data Analysis and Fusion Workflow

The processing workflow comprises source screening, trajectory quality control, spatial-temporal assignment, event aggregation, and model-specific analysis[34]. The 1,000-record source-labelled sortie sample is screened by requiring flight_min > 0 and flight_min ≤ 1,440 min; this retains 772 records (740 operational, 32 simulation). The larger 55-day series contains 20,000 exported sortie rows; 119 negative-duration records are excluded, leaving 19,881 sorties and 6,079 occupied 5-min windows for the density-adjusted count models.

For the trajectory slice, invalid geographic points are removed and duplicate timestamps are dropped before calculating kinematic features. The core operating area is bounded by longitude 118.70–118.84°E and latitude 32.02–32.12°N. Coordinates are projected locally to metres; 500-m grid cells are combined with non-overlapping 300-s windows. Alert locations are assigned to the same cells and windows, and repeated alerts are also summarized by distinct mission_id. The exported trajectory/alert tables are capped extracts, so the resulting slice is used for supplementary reconstruction and not represented as a complete fleet census.

The trajectory/alert slice begins at 2026-09-11 14:00 local dataset time. It covers 47 active grid cells and 29 five-minute windows, producing 123 occupied grid-window cells. The median number of sorties per occupied cell is one; 9 cells contain conflict alerts after assignment. Exposure is approximated as the number of trajectory points in the cell multiplied by the median within-mission sampling interval (2 s in the indicator pipeline; the separate cell-formation demonstration uses its own estimated interval). These are export-slice reconstruction counts and do not redefine the manuscript's original 785-sortie/64-unit analysis; the supplied archive is added as supplementary evidence, and reconciliation of the original counts is tied to the corresponding processing records.

For the trajectory-slice analysis, each cell stores the number of missions, trajectory points, exposure seconds, mean sortie risk score, raw conflict-alert count, and alert count deduplicated by mission. The 55-day series uses a different aggregation: events and exposure seconds are summed in fixed 5-min windows based on recorded mission start times; the mean risk score and mean concurrent-aircraft density are calculated for each window. These cell definitions are kept distinct because they have different spatial support and outcome construction.

$$D_{u,t} = \{\mathrm{Expo}_{u,t},\quad Y_{u,t},\quad Z_{u,t}\} \tag{1}$$

The source tag is retained for every sortie. The 1,000-record source sample comprises 967 operational sorties (96.7%) and 33 simulated sorties (3.3%); the screened regression sample comprises 740 operational and 32 simulated sorties. The simulated subset has zero recorded events and no variation in risk score, so simulation-only and risk-by-source interaction estimates are not identifiable. The mixed-sample estimate is therefore reported alongside the operational-only estimate, rather than treated as validation of simulation scenarios.

3.3 Bias Analysis and Mitigation Strategies

Safety outcomes in these analyses are recorded proxies, not actual accidents. The source-labelled sortie dataset contains event_count and risk_score fields from the monitoring export; the trajectory slice uses monitoring-system conflict alerts; abrupt-heading changes are maneuver indicators. None is equated with collision, injury, or accident occurrence. Their interpretation depends on alert coverage, export completeness, and the construction of the risk score.

(1) Interpretation: The selected signals are treated as observable indicators for statistical analysis. Their relationship with accident risk is not assumed to be one-to-one, and this study does not estimate accident probability.

For sortie-level and 55-day window models, Y denotes the recorded event_count aggregated at the relevant unit, and exposure is the sum of flight seconds. For the trajectory slice, conflict alerts are separately counted and deduplicated by mission_id within a grid–window cell. A binary proxy-event indicator is one when the relevant unit contains at least one recorded event and zero otherwise. The event definitions are not combined across datasets.

Just Accepted

(2) Potential Biases and Mitigation:

Reporting bias: monitoring coverage and alert thresholds may vary across missions and times. Source labels, quality-control flags, and the distinct event-construction rules are retained; where those fields are unavailable, results are interpreted as observed-data associations rather than fully adjusted population estimates.

Exposure Dependence: The number of events naturally increases with traffic volume. By introducing an exposure offset term $\log \mathrm{Expo}_{u,t}$ into the calibration model, the modeling target shifts from the absolute number of events to the event occurrence rate, thereby controlling for the effect of traffic volume.

Causal interpretation: all fitted coefficients are associational. In particular, sortie risk_score and event_count are produced by the same operational monitoring environment and may share inputs. Models do not establish causal effects or accident probabilities.

4 Multidimensional Risk Assessment Model

4.1 Risk Dimensions and Indicator Definitions

In order to systematically reflect the operational safety posture of urban low-altitude airspace, a multidimensional risk assessment system comprising 8 indicators is constructed. These indicators capture the state of airspace operational risk from different perspectives, including conflict frequency, degree of proximity, and difficulty of conflict resolution, among others. The specific definitions are as follows:

(1) Airspace Conflict Count

For each unordered aircraft pair, the pair-level indicator is summed as follows:

$$C(u,t) = \sum_{i<j} C_{ij}(u,t) \tag{2}$$

Here, Cij(u,t)=1 if pair (i,j) meets the stated conflict criterion at least once during unit (u,t), and Cij(u,t)=0 otherwise; N(u,t) is the number of observed aircraft. Pairs are unordered, and continuous encounters are deduplicated by the stated persistence rule.

(2) Airspace Conflict Rate

This indicator jointly considers the effects of both the conflict count and the flight volume within the airspace Ω during the statistical period $t_0 - t_1$ .The expression is given as follows: The flight-volume denominator must be positive. This descriptive conflict count per observed aircraft is not a probability and may exceed one.

$$RC^{\Omega}_{t_0-t_1} = \frac{C^{\Omega}_{t_0-t_1}}{N^{\Omega}_{t_0-t_1}} \tag{3}$$

Just Accepted

(3) Degree of Airspace Convergence Proximity

This indicator refers to the degree of proximity among all converging aircraft within the airspace Ω during the statistical period $t_0 - t_1$. The calculation expression for the average proximity rate of all aircraft $AP^{\Omega}_{t_0-t_1}$ within the airspace is given as follows:

$$AP^{\Omega}_{t_0-t_1} = \frac{\sum_l \sum_{i=1}^{n_l} \sum_{j=1, j\neq i}^{n_l} AP_{ij}^{t_0-t_1}}{\sum_l n_l(n_l-1)} \tag{4}$$

where l denotes the altitude layer and n_l is its number of observed aircraft. Self-pairs are excluded. Define $P_{ij} = p_j - p_i$ and the relative closing velocity $v_{ij} = v_i - v_j$. The nonnegative pairwise convergence measure is $AP_{ij}^{t_0-t_1} = max(0, \frac{P_{ij} \cdot v_{ij}}{\|P_{ij}\|^2})$, evaluated for nonzero relative position. This convention makes approaching motion positive and clips receding motion to zero. The aggregate is assigned zero when no eligible pair is present. Coincident positions require a separate conflict flag rather than evaluation of the singular quotient.

(4) Airspace Collision Risk Intensity

This indicator refers to the magnitude of the repulsive force—i.e., the collision risk intensity—among low-altitude aircraft within the airspace Ω during the statistical period $t_0 - t_1$, based on the artificial potential field method. It is expressed as follows:

$$CS^{\Omega}_{t_0-t_1} = \sum_i \frac{\eta}{D_i^2(t)} \left(\frac{1}{D_i(t)} - \frac{1}{R_S} \right) \cdot I\left(D_i(t) \le R_S\right) \tag{5}$$

where η denotes the proportional position gain coefficient， $D_i(t)$ denotes the horizontal separation distance between low-altitude aircraft i and the nearest low-altitude aircraft at time t， R_S denotes the minimum required horizontal safety separation for low-altitude aircraft i， I is an indicator function, which equals 1 when $D_i(t) \le R_S$, and 0 otherwise. The expression is evaluated only for positive separation and a positive reference separation. Coincident positions must be flagged separately rather than substituted into the singular inverse-distance expression.

(5) Average Airspace Conflict Resolution Behavior Ratio

This indicator refers to the ratio of the sum of conflict resolution behavior ratios to time within the airspace Ω during the statistical period $t_0 - t_1$. The expression is given as follows:

$$\overline{CR^{\Omega}_{t_0-t_1}} = \frac{1}{t_1 - t_0} CR^{\Omega}_{t_0-t_1} \tag{6}$$

where $CR^{\Omega}_{t_0-t_1}$ denotes the conflict resolution behavior ratio， defined as the ratio of the number of flight conflict resolution behaviors to the flight volume， a smaller value indicates that conflict resolution is easier to achieve. The expression is given as follows：

Just Accepted

$$CR_{t_0-t_1}^{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=t_0}^{t_1}(heading_{i,t} + altitude_{i,t} + velocity_{i,t}) \tag{7}$$

where $heading_{i,t}$, $altitude_{i,t}$, $velocity_{i,t}$ denote, respectively, the number of heading changes, altitude changes, and speed changes made by low-altitude aircraft i at time t for the purpose of conflict avoidance and resolution. If aircraft i performs a heading change, altitude change, or speed change action, then the corresponding term $heading_{i,t} = 1$, $altitude_{i,t} = 1$, $velocity_{i,t} = 1$; otherwise, it equals 0.

(6) Average Airspace Heading Change

This indicator refers to the average heading change among all low-altitude aircraft that adjust their heading for conflict avoidance within the airspace Ω during the statistical period $t_0 - t_1$, reflecting the difficulty of conflict resolution for low-altitude aircraft. It is expressed as follows:

$$\underline{\Delta\theta_{t_0-t_1}^{\Omega}} = \frac{\sum_{i=1}^{n_{\Delta\theta}} d_{circ}(\theta_i^t, \theta_i^*)}{n_{\Delta\theta}} \tag{8}$$

where $n_{\Delta\theta}$ denotes the number of heading-change actions performed, $\theta_{i,t}$ denotes the instantaneous flight heading of low-altitude aircraft i at time t, θ_i^* denotes the intended (planned) heading of low-altitude aircraft i at time t. Heading differences use the shortest circular distance in degrees: d_circ(a,b) = |((a − b + 180°) mod 360°) − 180°|, with the modulo result in [0°, 360°). Headings of 359° and 1° therefore differ by 2°. Equations (8)–(10) require a nonempty set of eligible maneuver observations; an empty set is flagged as no observed maneuver and is not evaluated by division by zero.

(7) Average Airspace Altitude Change

This indicator refers to the average altitude change among all low-altitude aircraft that adjust their altitude for conflict avoidance within the airspace Ω during the statistical period $t_0 - t_1$, reflecting the difficulty of conflict resolution for low-altitude aircraft. It is expressed as follows:

$$\overline{\Delta H_{t_0-t_1}^{\Omega}} = \frac{\sum_{i=1}^{n_{\Delta h}} \left| h_i^t - h_i^* \right|}{n_{\Delta h}} \tag{9}$$

where $n_{\Delta h}$ denotes the number of altitude-change actions performed, h_i^t denotes the instantaneous flight altitude of low-altitude aircraft i at time t, h_i^* denotes the intended (planned) altitude of low-altitude aircraft i at time t.

Just Accepted

(8) Average Airspace Speed Change

This indicator refers to the average speed change among all speed-adjusting aircraft within the airspace Ω during the statistical period $t_0 - t_1$, used to reflect the difficulty of conflict resolution for aircraft. The calculation expression is given as follows:

$$\overline{\Delta V_{t_0-t_1}^{\Omega}} = \frac{\sum_{i=1}^{n_{\Delta v}} \left| v_i^t - v_i^* \right|}{n_{\Delta v}} \tag{10}$$

where $n_{\Delta v}$ denotes the flight volume of aircraft performing speed-adjustment actions, v_i^t denotes the instantaneous speed of aircraft i at time t, v_i^* denotes the cruising speed of aircraft i.

Table 2 Multidimensional Risk Indicator System and Governance Implications

Indicator name	**Indicator meaning**	**Formula correspondence**	**Governance implication**
Airspace conflict count	The total number of flight conflicts occurring within the airspace during the statistical period, used to characterize the overall level of conflict exposure	Eq.(2)	Reflects the degree of tension in airspace operations; can serve as a basic quantity for conflict monitoring and early-warning triggering
Airspace conflict rate	Characterizes the level of conflict occurrence per unit of operational activity, accounting for flight volume	Eq. (3)	Normalizes conflicts by observed aircraft volume; residual density effects require separate assessment and adjustment
Degree of airspace convergence proximity	Characterizes the relative closeness and strength of convergence tendency among aircraft	Eq. (4)	Reflects potential instability trends; can be used to identify high-density convergence risk zones in advance
Airspace collision risk intensity	Measures the strength of collision risk among aircraft based on the artificial potential field method	Eq. (5)	Can serve as an important basis for dynamic safety separation control and prioritization of conflict resolution

Just Accepted

Indicator name	Indicator meaning	Formula correspondence	Governance implication
Average airspace conflict resolution behavior ratio	Reflects the frequency of maneuvering actions taken to avoid conflicts, per unit of flight volume	Eqs. (6)–(7)	Used to measure the system's regulatory pressure and operational load, assisting in judging whether current operations are approaching capacity limits
Average airspace heading change	Statistically characterizes the average magnitude of heading adjustments during conflict avoidance	Eq. (8)	Can be used to assess the cost of lateral avoidance and the degree of route deviation
Average airspace altitude change	Statistically characterizes the average magnitude of altitude adjustments during conflict avoidance	Eq. (9)	Used to assess the complexity of vertical avoidance and the degree of disturbance to altitude-layer operations
Average airspace speed change	Statistically characterizes the average magnitude of speed adjustments during conflict avoidance	Eq. (10)	Used to identify the burden of speed-control adjustments and their impact on operational efficiency

Table 2 systematically summarizes the eight core indicators adopted in this chapter in order to enhance the comparability among the various risk indicators and to provide structured support for the subsequent construction of the composite risk score. These indicators characterize the operational risk state of urban low-altitude airspace from three complementary dimensions-conflict exposure, proximity intensity, and resolution cost-and jointly constitute the indicator foundation for the subsequent calculation of risk scores and the triggering of governance rules.

Just Accepted

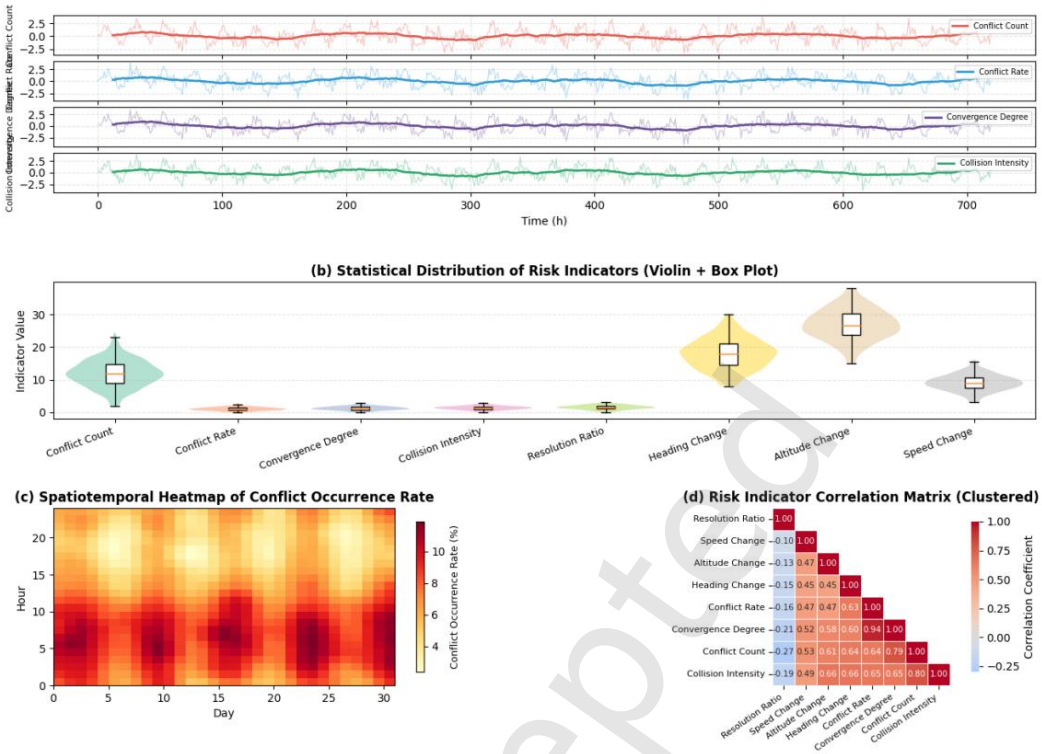


Fig. 4 Comprehensive risk indicator analysis.

To characterize the dynamic features and statistical properties of each risk indicator, a multidimensional risk indicator visualization analysis system is constructed, as shown in Figure 4. This analysis system adopts a four-panel composite view to comprehensively depict the spatiotemporal distribution patterns of the risk indicators: (a) the trend comparison chart presents the temporal evolution patterns of four core indicators-conflict count, conflict rate, convergence degree, and collision risk-with standardization applied to ensure comparability across indicators of different dimensions; (b) the statistical distribution box plot reveals the central tendency, dispersion, and outlier distribution of each risk indicator, with the conflict resolution behavior ratio exhibiting a notably negative-skewed distribution; (c) the spatiotemporal heatmap presents the spatiotemporal heterogeneity of conflict occurrence rates in a gridded format of 24 hours × 30 days, where red regions represent high-risk time periods and airspace zones; (d) the correlation matrix quantifies the strength of association among the 8 risk indicators, with results showing a strong positive correlation between conflict count and conflict rate ($r = 0.87$), and a significant negative correlation between the conflict resolution behavior ratio and collision risk ($r = -0.72$). This visualization system enables multidimensional perception of complex risk postures, providing a data foundation and intuitive basis for the subsequent construction of the risk score.

To further reveal the intrinsic associative structure among the risk indicators, a risk indicator correlation network model is constructed, as shown in Figure 5. In this network model, the 8 risk indicators are abstracted as nodes, with edges constructed based on Pearson correlation coefficients:

Just Accepted

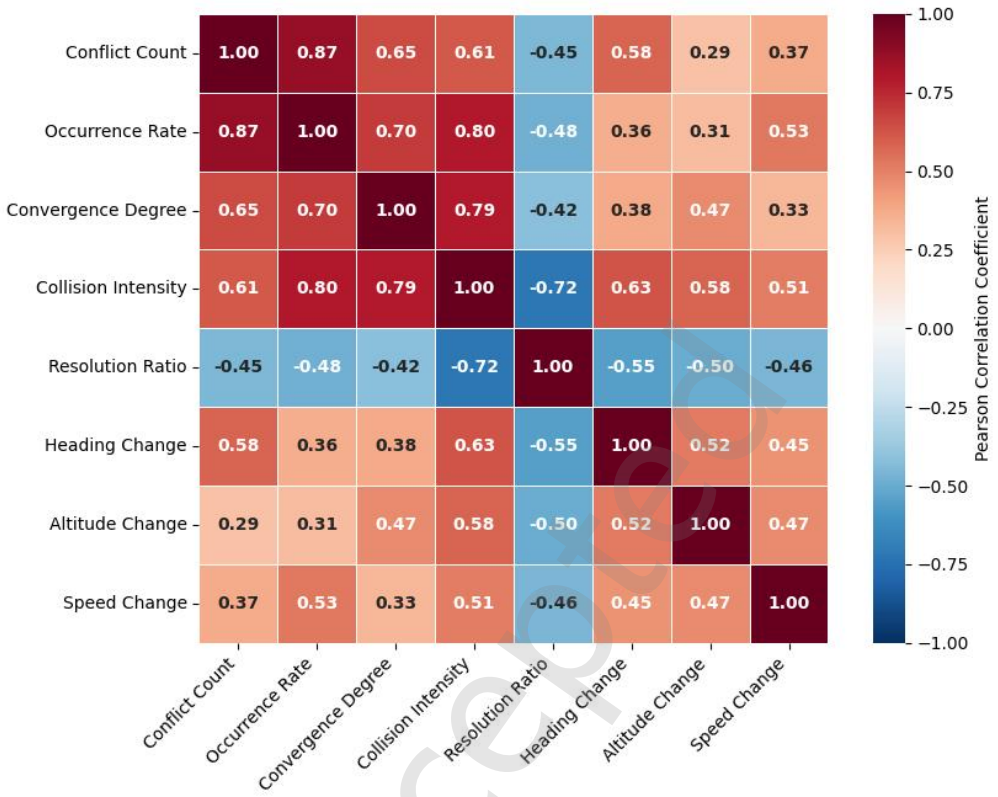


Fig. 5 Correlation analysis of risk indicators.

The Pearson correlation matrix shown in Figure 5 is presented as a color-coded heatmap, in which cell color intensity-ranging from deep blue (strong negative correlation) through white (near-zero correlation) to deep red (strong positive correlation)-visually encodes the direction and strength of pairwise associations among the eight risk indicators, with exact coefficient values annotated within each cell for precise quantitative reference.

The analysis reveals that the risk indicators exhibit a distinct modular structure: conflict-related indicators-including conflict count, occurrence rate, convergence degree, and collision intensity-form a tightly positively correlated cluster, with coefficients ranging from 0.61 to 0.87, appearing as a dense block of deep red cells in the upper-left region of the matrix. In contrast, resolution-related indicators-including heading change, altitude change, and speed change-form another moderately positively correlated cluster (coefficients of 0.45–0.63), visible as a secondary red-toned block in the lower-right region. Notably, resolution ratio exhibits consistently negative correlations with all other indicators (ranging from −0.42 to −0.72), displayed as a distinct blue-toned row/column that visually separates the two positive clusters. Collision intensity, acting as the core hub indicator, exhibits moderate-to-strong correlations with multiple indicators across both clusters (e.g., 0.80 with occurrence rate, 0.79 with convergence degree, 0.63 with heading change, and −0.72 with resolution ratio—the strongest negative correlation in the entire matrix). This correlation structure uncovers the intrinsic associative characteristics of the risk system, providing a theoretical basis for constructing a concise and effective risk early-warning indicator system.

Just Accepted

The matrix in Figure 4(d) is the correlation panel within the composite indicator figure; the separate correlation matrix is Figure 5. Coefficients describe pairwise association and do not establish causal pathways.

As the horizontal or vertical separation between aircraft nonlinearly approaches the critical safety threshold, the repulsive force field in the artificial potential field model erupts in a sharply nonlinear manner, causing "collision risk intensity" to surge within an extremely short time window, thereby pushing the entire spatiotemporal unit into a high-risk state. This systemic risk escalation generates a strong reactive effect at the operational control level, in turn compelling the UTM system and onboard collision-avoidance equipment to execute high-frequency conflict-resolution maneuvers (frequently triggering abrupt changes in heading, altitude, and speed). This mechanism perfectly explains the transmission pathway among exposure indicators, internal situational indicators, and operational cost indicators, ensuring that the multidimensional risk assessment charts are no longer presented in isolation but are instead deeply embedded within the dynamical logic of systemic situational awareness.

4.2 Risk Score Construction and Uncertainty

A reproducible reference implementation uses K=4 clusters, fuzzifier m=2.0, objective-change tolerance 10^{-6}, maximum 500 FCM iterations, and 30 repeated runs with seed 20260911. The adaptive genetic search uses a population of 60, at most 150 generations, early stopping after 25 stagnant generations, two elite solutions, inner FCM limit 15, and Srinivas adaptive crossover[35] $Pc \in [0.6,0.9]$ and mutation $Pm \in [0.02,0.15]$. These values describe the supplementary reference implementation used here. Across 30 runs, AGA-FCM and FCM both yielded XB=0.085257 and silhouette=0.8620; K-means yielded XB=0.1516–0.1736 and silhouette=0.8568. The results do not establish an AGA-FCM advantage over FCM.

(1) Multi-source Indicator Data Preprocessing

Each of the eight indicators is first checked for directionality and distribution. Indicators with absolute skewness above 1 are transformed using log1p; all indicators are then robustly standardized as $(x-\text{median})/(1.4826\times\text{MAD})$, with standard deviation used when MAD is degenerate. The transformed indicators are oriented so that larger values correspond to greater operational concern. The supplementary pipeline uses 700 occupied cells from the trajectory slice for this clustering analysis.

(2) AGA-FCM-based Risk Situation Clustering Logic

The reference risk-clustering procedure combines adaptive genetic initialization with FCM refinement. Its core mechanisms include:

The reference adaptive genetic algorithm uses floating-point cluster-centre chromosomes and the reciprocal FCM objective as fitness. The Srinivas operators adapt crossover and mutation probabilities within the ranges stated above. Each candidate is evaluated with up to 15 FCM iterations; the full search stops at 150 generations or after 25 generations without improvement, retaining two elite candidates. Thirty runs use the fixed random seed for reproducibility. Since the resulting XB equals the FCM result, genetic initialization did not improve the attained objective in this sample.

Just Accepted

FCM memberships are iterated until the objective change is below 10^{-6} or 500 iterations are reached. The number of clusters is fixed at four for the reference analysis; labels are ordered by the cluster-level score. XB is used as a compactness/separation diagnostic, not as evidence of operational safety or external validity.

(3) Multidimensional Risk Score System Evaluation

Figure 6 illustrates the intended risk-scoring and four-level classification workflow schematically. Its displayed distributions, class proportions, decision zones, and time series are illustrative and are not empirical outputs of the 700-cell reference pipeline. The empirical sample counts and exploratory boundaries are reported separately below.

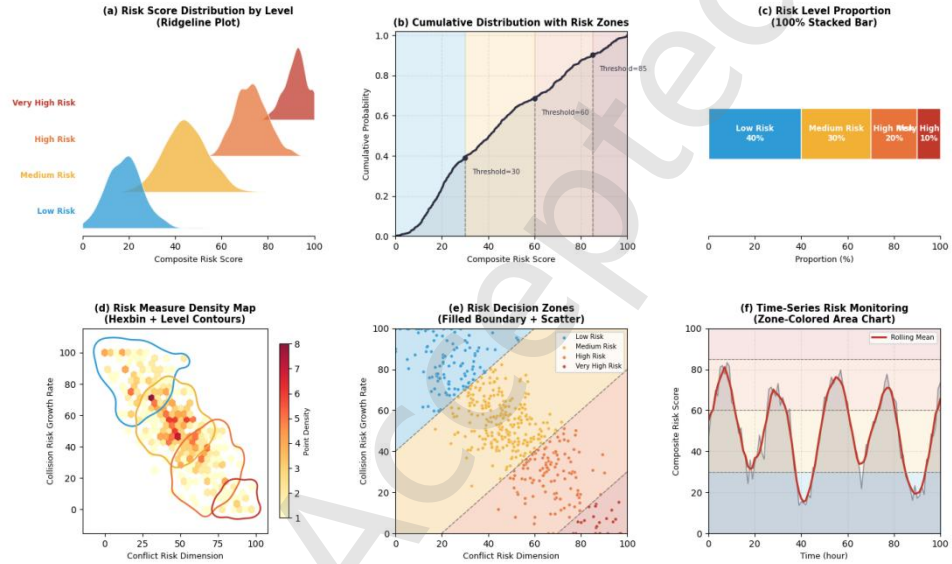


Fig. 6 Schematic illustration of multidimensional risk scoring and classification (illustrative distributions and proportions).

(a–c) The panels illustrate risk-level distributions, cumulative risk zones, and a generic composition of four classes. Their displayed proportions must not be used as the observed proportions of the 700-cell sample.

The four-level score is defined as $R_i=100\cdot\Sigma_k\ \mu_{ik}\ s_k$, with ordinal scores $s=\{0,1/3,2/3,1\}$. Candidate boundaries are B01=34.98, B12=45.43, and B23=64.54, obtained from adjacent-cluster membership crossover candidates; B23 uses a midpoint rule because fewer than three qualifying crossover candidates were available. These boundaries are exploratory and sample-specific, not validated operational thresholds.

The observed class counts in the 700-cell slice are 567, 88, 33, and 12 for levels 0–3, respectively. Level 3 is small (n=12); its estimated contributions and boundary behaviour are correspondingly uncertain. The class proportions are not interpreted as Heinrich-triangle evidence.

(d–e) The scatter and decision-surface panels schematically illustrate a four-level partition; they are not fitted or independently validated decision boundaries for the reported sample.

Just Accepted

(f) The time-series panel schematically illustrates monitoring across risk zones; it is not an observed or prospective real-time warning evaluation.

(4) Threshold Sensitivity and Robustness Analysis

A 27-scenario sensitivity analysis reran affected indicator, standardization, clustering, score, boundary, and negative-binomial steps. Across the tested settings, the most influential changes were the maneuver-heading threshold (30° to 45°: ΔRR=+0.0034; B23 shifted +14.2 points; 5.0% level migration), FCM fuzzifier (m=1.5: ΔRR=+0.0019; B23 shifted +29.4 points; 5.9% migration), and ordinal score mapping (maximum ΔRR=+0.0011). Coupled R_s/H_warn changes of ±50% changed RR by at most 0.00015. All RR confidence intervals included 1 and overlapped the baseline interval. Interpretation is limited by sparsity: only 24 of 700 cells contained an aircraft pair and only 12 cells were in level 3.

Clustering quality was compared using the Xie–Beni index and silhouette coefficient across 30 repeated runs of the reference implementation. The same scaling, K=4, and feature set were used for AGA-FCM and FCM; K-means used a hard partition. These diagnostics assess within-sample partition geometry only.

AGA-FCM and FCM both achieved XB=0.085257 and silhouette=0.8620; K-means XB ranged from 0.1516 to 0.1736 and its silhouette was 0.8568. Thus, K-means had a less compact partition under this implementation, while AGA-FCM did not outperform FCM. The comparison is based on a sparse 700-cell export slice and is not evidence of eVTOL-specific classification performance.

4.3 Risk Model Calibration and Validation

Calibration and validation are conducted separately for each available evidence stream. The source-labelled sortie sample supports operational-only versus mixed-source regression; the 55-day series supports density-adjusted temporal analyses; and the 700-cell trajectory slice supports indicator construction, clustering, parameter sensitivity, and supplementary event analysis. These samples have different units, outcomes, and time coverage, so estimates are not pooled.

Just Accepted

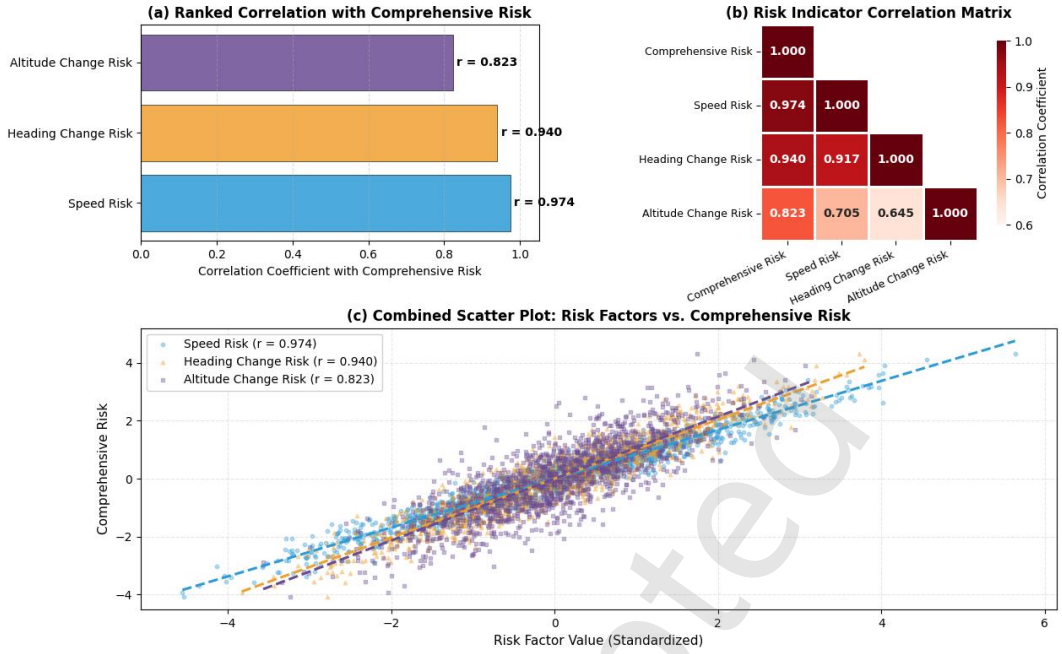


Fig. 7 Risk indicators correlation analysis.

In the 700-cell trajectory slice, the overall negative-binomial association between the composite score and deduplicated conflict-alert count was RR=1.0026 per score point (95% CI 0.9957–1.0095; p=0.464; α=14.53). The null-compatible estimate does not support the earlier claim of a statistically significant 4.9% increase per point. The source-labelled sortie regression gives a separate result: among 772 retained records, RR=10.51 per one-level increase (95% CI 9.59–11.51), with nearly identical operational-only estimate (n=740; RR=10.51, 95% CI 9.59–11.51). The simulation-only model and source interaction are not identifiable because all 32 retained simulated records have zero event counts and no score variation. These sortie-level results are subject to shared construction of risk_score and event_count.

A separate operational series covering 55 days was aggregated to 6,079 occupied 5-min windows after exclusion of 119 records with negative duration. In an NB2 model with log flight-seconds as offset, time-of-day Fourier terms, weekend status, mean risk level, and mean concurrent-aircraft density, the risk-level rate ratio was 3.317 (95% CI 3.128–3.516; n=6,079); density RR was 1.106 (95% CI 1.074–1.139). The risk coefficient changed by about 1% when density was added to the controlled risk model, indicating little density-related change in the risk coefficient in this dataset. This is a within-dataset association, not independent event prediction.

(a) Normal operation scenario: Conflict count registers the highest value among all indicators (0.75), while all remaining indicators-including occurrence rate, convergence degree, collision intensity, and resolution ratio-remain at consistently low levels (0.06–0.10). This sharply contrasting pattern (one dominant indicator against a uniformly low background) exhibits typical low-risk characteristics, reflecting baseline traffic exposure without elevated systemic risk.

Just Accepted

(b) Submarine cable fault scenario: All eight risk indicators rise to a moderate and relatively uniform level (0.35–0.42), without any single indicator exhibiting extreme values. This flat, moderately elevated profile suggests that infrastructure-related disruptions exert a broadly distributed influence across conflict, convergence, and resolution-related indicators simultaneously, rather than concentrating risk in any specific dimension.c) Severe weather scenario: Convergence degree (0.65) and heading change (0.60) rise most prominently, followed by occurrence rate (0.55), while collision intensity, resolution ratio, altitude change, and speed change remain at moderate-to-low levels (0.30–0.45). This pattern reflects the direct impact of meteorological conditions on aircraft trajectory convergence and heading-related maneuvering, consistent with weather-induced routing adjustments and spatial clustering of traffic.

(d) Equipment fault scenario: Collision intensity reaches the highest value across the entire heatmap (0.80), accompanied by a notably elevated occurrence rate (0.70), while the resolution-capability-related indicators-heading change, altitude change, and speed change-remain comparatively low (0.30–0.35). This divergence indicates that system degradation substantially impairs the aircraft's conflict-resolution capacity, resulting in an accumulation of unresolved collision risk rather than active maneuvering response, which explains the markedly elevated collision intensity.

(e) ATC error scenario: All eight indicators exhibit a comprehensive and relatively uniform increase (0.50–0.65), representing the highest overall risk baseline among all five scenarios. This broad-based elevation-without any single indicator standing out disproportionately-indicates that human factors introduce systemic risk that propagates simultaneously across conflict generation, convergence, and resolution-related dimensions.

(f) Composite comparison heatmap: This integrated heatmap clearly illustrates the differences in risk feature patterns across the various scenarios through direct visual and numerical comparison: the concentrated single-peak pattern in normal operation, the uniformly elevated profile in submarine cable fault and ATC error scenarios, the convergence-dominant pattern in severe weather, and the collision-intensity-dominant pattern in equipment fault. These distinct risk "signatures" provide decision support for formulating scenario-specific risk management and control strategies.

Just Accepted

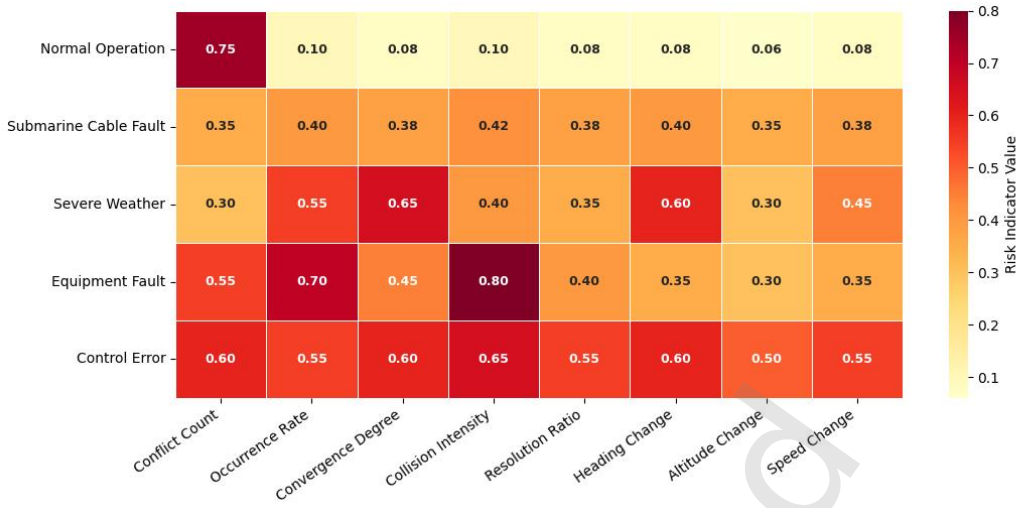


Fig. 8 Risk indicator profiles across operational scenarios.

The 55-day temporal analysis shows stable classification metrics but weak evidence of independent predictive value because risk_score and event_count are closely coupled in the source records. Therefore the rolling AUC is reported as a diagnostic of source-label dependence, not as validation of future eVTOL warning capability.

4.3.1 Calibration Data and Safety Signal Definition

The source-labelled file contains 1,000 sorties, of which 967 (96.7%) are marked operational and 33 (3.3%) simulated. The prespecified exposure screen (flight_min>0 and <=1,440 min) retains 772 records: 740 operational and 32 simulated. Three operational records exceeding 24 h and 225 zero-exposure records are excluded. In the retained sample, simulated sorties have zero recorded events and zero risk-score variation, making simulation-only regression and the risk-by-source interaction non-identifiable. Separately, the trajectory pipeline uses 700 occupied 500 m x 500 m x 5-min cells from a capped 20,000-row trajectory-event export; the core-area reconstruction yields 123 occupied cells (47 grids x 29 windows). The supplied supplementary archive is incorporated to document these export-derived analyses. Because its source tables are capped, they do not independently reproduce the original 785-sortie/64-unit cohort; exact reconciliation of those original counts requires the processing records used to generate that cohort.

Conflict-alert proxy: conflict alerts are read from monitoring records. In the trajectory pipeline, pairwise proximity is operationalized as horizontal separation below H_{warn}=1,096 m and vertical separation below V_{warn}=15 m for at least three consecutive samples; H_{warn} is the median of 3,092 deduplicated alert-pair observations. V_{warn}=15 m and η=2 are reference settings and are varied in sensitivity analyses. This computational criterion is not an accident definition.

Maneuver proxy: an absolute heading change greater than 30° between adjacent trajectory samples no more than 4 s apart is treated as a maneuver. This indicator may reflect avoidance, route correction, or measurement noise and is not labelled an emergency without corroborating evidence.

Just Accepted

For the 55-day series, Y is the recorded sortie event_count aggregated to the 5-min window. For the trajectory pipeline, the primary outcome is the monitoring-system conflict-alert count, with a separate binary indicator for any alert. These outcomes are analysed as safety proxies only and are not treated as accidents or collisions.

4.3.2 Score Calibration Based on Generalized Linear Models

To establish a quantitative association between the composite risk score $R_{u,t}$ and the observable safety event $Y_{u,t}$, this study constructs a statistical calibration framework based on the Generalized Linear Model (GLM). Since the safety event count $Y_{u,t}$ is a non-negative integer and typically exhibits over-dispersion, the standard assumptions underlying traditional Poisson regression do not hold. Negative Binomial Regression is therefore adopted as the base model, as it accommodates the condition in which the variance exceeds the mean through the introduction of a dispersion parameter. The model specification is as follows:

$$Y_{u,t} \sim \mathrm{NegBin}(\lambda_{u,t}, \phi) \tag{11}$$

where $Y_{u,t}$ is the dependent variable (i.e., the count of safety events occurring within spatiotemporal unit u,t， $\lambda_{u,t}$ is the expected value of the event count, $\phi > 0$ is the dispersion parameter. The expected value is linked to the explanatory variables through a log link function to establish a linear relationship, expressed as: The NB2 parameterization has conditional variance $\lambda + \phi\lambda^2$.

$$\log(\lambda_{u,t}) = \beta_0 + \beta_1 R_{u,t} + \sum_{k=1}^{8} \gamma k X_{u,t,k} + \delta Z_{u,t} + \log(Expo_{u,t}) \tag{12}$$

The meaning of each term in the equation is as follows:$R_{u,t}$ is the composite risk score (the core independent variable); $X_{u,t}^{(k)}$ represents the standardized multidimensional risk indicators; $Z_{\mathrm{u},t}$ is the vector of control variables (including meteorological conditions, peak-period dummy variables, etc., introduced to control for external confounding factors); $\mathrm{Expo}_{u,t}$ is the exposure (cumulative aircraft flight seconds) within the spatiotemporal unit, which is introduced into the model as an offset term to eliminate the inherent influence of traffic flow scale on the event count, thereby enabling the modeling of the event occurrence rate rather than its absolute count;β_0 is the intercept term; β_1 is the coefficient of the risk score; and γ_k are the coefficients of the respective risk indicators. Equation (12) is a general candidate specification. The reported score-based regressions set all γk to zero rather than entering the composite score and its eight constituent indicators simultaneously. The 55-day model uses mean recorded ordinal risk level with density and temporal controls; the sortie model uses its recorded risk level, whereas the 700-cell analysis uses the continuous composite score. Rate ratios per risk level and per score point are distinct estimands. An exposure offset does not by itself remove density effects or confounding. Only positive-exposure observations enter models with a logarithmic offset.

Just Accepted

For the 55-day operational series, adjusted NB2 regression yielded RR=3.317 for each one-level increase in mean risk score (95% CI 3.128–3.516; p<0.001), controlling for density and temporal factors. In the source-labelled sortie sample, mixed and operational-only estimates were RR=10.511 (95% CI 9.594–11.514; n=772) and RR=10.506 (95% CI 9.590–11.510; n=740), respectively. In contrast, the 700-cell trajectory analysis yielded RR=1.0026 per score point (95% CI 0.9957–1.0095; p=0.464). Because outcome and risk-score fields can share monitoring-system inputs, the strong associations are interpreted as within-source association and may include criterion contamination; the trajectory-slice result does not show a statistically significant score–alert association.

4.3.3 Model Validation and Robustness Analysis

To ensure the credibility of the calibrated model and its generalization capability under future operational scenarios, a multi-level validation strategy combined with systematic perturbation testing is employed to comprehensively evaluate the model's predictive performance and robustness. To fully assess model performance, this study adopts the following three categories of validation methods:

(1) Rolling-origin temporal analysis: the 55-day series is evaluated with four expanding-window folds (initial 28-day training window, 7-day test windows, 7-day step). AUC ranged from 0.9994 to 1.0000 (mean 0.9997, SD 0.0003), PR-AUC averaged 0.9995, count RMSE averaged 286.1 (SD 87.9), and 20% tolerance coverage averaged 0.096. Risk scores of at least 2 were associated with an event in 100% of sorties, demonstrating near-direct outcome construction dependence; near-perfect AUC is therefore not interpreted as independent predictive performance. Count error also increased in later folds.

(2) Density-adjusted model: the 55-day operational dataset contains 6,079 occupied windows. Density is calculated as the window mean of the second-by-second number of airborne sorties. NB2 models use log summed flight-seconds as offset; controls are two-harmonic hour-of-day terms and a weekend indicator. In the full model, density RR=1.106 (95% CI 1.074–1.139), and risk-level RR=3.317 (95% CI 3.128–3.516). Adding density changed the adjusted risk coefficient from 1.1869 to 1.1989 (about 1.0%).

(3) Perturbation and missingness analyses: 27 parameter scenarios and five repetitions each under MCAR (10% random point loss), MAR (20% loss for above-median speed and 5% otherwise), and MNAR (40% loss in segments within 1,096 m of a nearest-neighbour aircraft) were run on the 700-cell pipeline. Mean RR shifts were −0.0003, −0.0018, and +0.0018, respectively; level migration was at most 1.7%. MNAR reduced mean conflict counts and shifted RR upward, so risk-dependent data loss can bias the association away from the null.

Performance is reported by dataset and outcome definition. The near-perfect AUC from the 55-day series reflects dependence between recorded risk score and event labels; the independent trajectory-slice model has RR=1.0026 (95% CI 0.9957–1.0095; p=0.464). No pooled AUC or causal effect is claimed.

① The rolling AUC criterion is not used as proof of predictive validity. The observed AUC values and their outcome dependence are reported in Section 4.3.3.

Just Accepted

② The 55-day time validation tests temporal stability within the same operational export. It does not constitute geographic external validation; the supplementary 123-cell core-area slice is descriptive and does not support a robust spatial holdout test.

③ Sensitivity results are reported with the actual parameter ranges, boundary shifts, rate-ratio changes, and cell-level sparsity. They apply to the 700-cell reference pipeline.

④ Uncertainty is summarized using the model-based confidence intervals and the 1,000-replicate bootstrap intervals reported for the supplementary contribution analysis; small high-risk cell counts are stated alongside interval-based interpretation.

Together, these analyses characterize the available exports and reference pipeline. They support reproducible description and sensitivity assessment, while the score's independent predictive validity and transfer to commercial eVTOL operations remain unestablished.

Table 3 Design of risk model validation and robustness analysis.

Validation module	Specific method	Data partitioning	Core evaluation metrics	Analytical purpose
Temporal analysis	Four rolling-origin folds	55 days; 6,079 occupied windows; initial 28-day training period, 7-day tests and steps	AUC, PR-AUC, count RMSE and 20% tolerance coverage	Reported within-source temporal analysis; label dependence limits predictive interpretation
Density adjustment	NB2 regression with a log flight-seconds offset	The same 6,079 windows; concurrent-aircraft density and temporal controls	Risk-level RR and density RR with 95% confidence intervals	Reported association analysis; not a causal intervention
Missingness perturbation	MCAR, MAR and MNAR point-loss scenarios	700-cell trajectory pipeline; five repetitions per missingness setting	RR shift and risk-level migration	Reported sensitivity to random and risk-dependent data loss
Parameter sensitivity	27 parameter scenarios	700-cell pipeline; indicator, scaling, clustering, score and model steps rerun	RR change, boundary shift and class migration	Reported within-sample sensitivity; not geographic external validation
Spatial and delay validation	Independent spatial holdout and quantified communication-delay tests	Not established by the reported supplementary analyses	No validated spatial-holdout or 5-s delay performance estimate	Further evaluation required; these tests are not claimed as completed

Just Accepted

Validation module	Specific method	Data partitioning	Core evaluation metrics	Analytical purpose
Uncertainty reporting	Model-based intervals and reported contribution bootstrap	Bootstrap contribution intervals are described for 1,000 resamples	95% intervals; counts of high-risk cells	Intervals must be interpreted with sparse-cell counts and the supplied processing records

Table 3 summarizes the supplementary validation designs. The principal evidence consists of a four-fold temporal analysis of 6,079 operational windows, a density-adjusted NB2 analysis of the same series, and parameter/missingness perturbations of the 700-cell trajectory pipeline. Near-perfect classification metrics are interpreted in light of source-label dependence; count RMSE and the separate null-compatible trajectory-slice estimate are also reported.

4.3.4 Stress Test Analysis of the Air-Ground Cooperative Risk-Avoidance Mechanism under Multi-Scenario, Multi-Intensity Perturbations

The multi-scenario stress test is a simulation-based comparison. The supplementary packet adds operational-data analyses and scenario diagnostics, but it does not turn the stress test into a field intervention. Strategy comparisons are conditional on simulated inputs and are not interpreted as measured reductions in actual accident risk.

Only reported scenario outputs are interpreted. The simulation evidence is kept separate from the 55-day operational window models and the 700-cell trajectory slice; the three sources have distinct event definitions and sampling frames.

The strategy ordering and percentage changes in the stress-test plots are treated as conditional on the chosen simulated scenarios. Operational benefit would require prospective, controlled evaluation with independent outcomes.

The combined evidence supports an auditable candidate governance architecture, but not causal effectiveness or operational readiness. Local procedures and prospective evaluation remain necessary before deployment.

Just Accepted

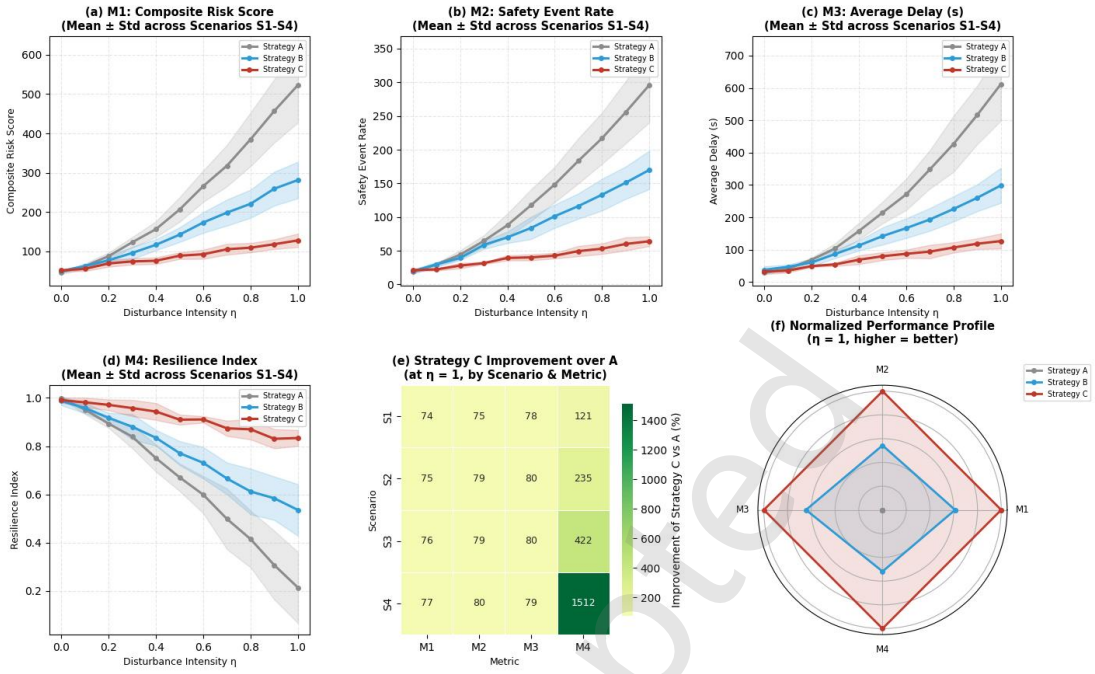


Fig. 9 Comparative stress-test analysis of the air-ground collaborative hazard avoidance mechanism under multi-scenario and multi-intensity disturbances.

5 Conclusions

This study presents a low-altitude operational risk assessment and air-ground response framework and evaluates it with separately analysed operational, simulation, and trajectory-export datasets. The supplied supplementary archive adds reproducible export-derived sample definitions and reports both positive and null-compatible results. The original 785-sortie/64-unit cohort remains a distinct analysis whose final count reconciliation must follow its source processing records; the capped supplementary exports are not substituted for that cohort.

(1) The supplied archive documents a 1,000-record source-labelled export, with 772 records meeting the exposure/duration screen (740 operational and 32 simulated); a separate trajectory/alert reconstruction produces 123 occupied grid-window cells across 47 grids and 29 five-minute windows; and the indicator/clustering pipeline contains 700 occupied cells. These are distinct processing products, not replacements for the original 785-sortie/64-unit cohort. Supplementary row-level records, window parameters, scripts, and per-cell event counts are included; final reconciliation of the original cohort follows its source processing records.

(2) In the 55-day operational series (6,079 occupied five-minute windows), adjusted risk-level RR was 3.317 (95% CI 3.128–3.516), while density RR was 1.106 (95% CI 1.074–1.139). The 700-cell trajectory-slice risk-score association was not statistically significant (RR 1.0026 per point, 95% CI 0.9957–1.0095; p=0.464). Rolling AUC averaged 0.9997, but is not treated as independent predictive validation because the score and event label

Just Accepted

are closely coupled. All findings concern low-altitude operational proxies, not actual eVTOL accidents.

Table 4 Key research findings and management implications

Dimension	Key finding	Value/Characteristic	Management implication and recommendation
Risk assessment system	Distribution of the composite risk score	Exploratory boundaries: 34.98, 45.43, and 64.54 (700-cell reference pipeline)	Use only for exploratory stratification; validate thresholds with independent eVTOL data before operational use
Risk–safety association	Source-specific risk-level and risk-score associations	55-day series: adjusted RR=3.317 per risk level (95% CI 3.128–3.516); separate 700-cell slice: RR=1.0026 per score point (95% CI 0.9957–1.0095; p=0.464)	Report datasets separately; estimates are associational and do not establish an operational trigger
Key risk indicator	Composite-score association with recorded alerts	The 700-cell score–alert association was not statistically distinguishable from unity (p=0.464)	This result does not support prioritizing the score as a validated real-time indicator
Model validation performance	Discriminative power in temporal extrapolation prediction	Rolling AUC mean=0.9997 in the 55-day source series; score and event labels are closely coupled	Treat AUC as a source-dependence diagnostic, not independent predictive validation
System robustness	Performance retention under 10% data loss	Across tested missingness scenarios, mean RR shifts were −0.0003 (MCAR), −0.0018 (MAR), and +0.0018 (MNAR)	Sensitivity findings apply to the sparse 700-cell export pipeline and do not establish field robustness
Latency sensitivity	Impact of a 5-second communication delay	A 5-s communication-latency perturbation was not evaluated in the supplied empirical supplement	Specify and validate latency limits through separate system-level testing before operational deployment

The Risk-Trigger-Action architecture remains a candidate design. Low level: routine monitoring and ground-capacity checks. Moderate level: heightened surveillance, operator notification, and review of departure/landing sequence. High level: the authorized airspace service provider considers increased separation or flow restriction while the vertiport operator meters departures and suspends new clearances if capacity or safety margins are exceeded. Extreme level: the competent authority may close the affected corridor or suspend operations, while the vertiport holds movements except approved emergency activity. Escalation follows a validated threshold crossing or independent safety trigger; de-escalation requires the score to remain below the lower boundary for a locally defined persistence period, hazard resolution, and responsible-authority approval. The supplementary cut points 34.98, 45.43, and 64.54 are exploratory and are not universal operational thresholds.

Just Accepted

Funding

This research was supported by the Major Project of the National Social Science Fund of China (No. 22&ZD169) and the Key Program of the Civil Aviation Joint Fund of the National Natural Science Foundation of China (Nos. U2133207 and U2333214).

Declaration of competing interest

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

Author contribution statement

Yuetan Zhang: Conceptualization, methodology, model development, formal analysis, investigation, data curation, visualization, writing-original draft, project administration. Honghai Zhang: Validation, resources, supervision, writing-review and editing, and funding acquisition. Wei Pan: Investigation, data curation, and writing-review and editing. Yulin Yuan: Resources, validation, and writing-review and editing. All authors have read and approved the final manuscript.

Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Use of AI statement

No AI tools were utilized for this manuscript. All work was performed independently by the authors, who are responsible for the content.

References

[1] Alvarez LE, Jessen I, Owen MP, Silbermann J, Wood P. ACAS sXu: Robust decentralized detect and avoid for small unmanned aircraft systems. In: 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC), 2019: 1-9. https://doi.org/10.1109/DASC43569.2019.9081631.

[2] Ma Z, Bai J, Yan L, Chen Y, Sun S. Safety assessment for airborne intelligent avoidance system based on Bayesian optimization. Acta Aeronautica et Astronautica Sinica, 2026, 47(1): 331973. https://doi.org/10.7527/S1000-6893.2025.31973.

[3] Zhao ZY, Zhang LF, Zhou RD, et al. A flight risk field model for advanced low-altitude transportation system using field theory. Transportation Research Part A: Policy and Practice, 2024, 190: 104268. https://doi.org/10.1016/j.tra.2024.104268.

[4] Haddad J, Mirkin B, Assor K. Traffic flow modeling and feedback control for future Low-Altitude Air city Transport: An MFD-based approach. Transportation Research Part C: Emerging Technologies, 2021, 133: 103380. https://doi.org/10.1016/j.trc.2021.103380.

Just Accepted

[5] Wang Z, Delahaye D, Farges J-L, Alam S. Complexity optimal air traffic assignment in multi-layer transport network for Urban Air Mobility operations. Transportation Research Part C: Emerging Technologies, 2022, 142: 103776. https://doi.org/10.1016/j.trc.2022.103776.

[6] Xie Y, Gardi A, Liang M, Sabatini R. Hybrid AI-based 4D trajectory management system for dense low altitude operations and Urban Air Mobility. Aerospace Science and Technology, 2024, 153: 109422. https://doi.org/10.1016/j.ast.2024.109422.

[7] Zhai W, Han B, Li D, Duan J, Cheng C. A low-altitude public air route network for UAV management constructed by global subdivision grids. PLoS ONE, 2021, 16(4): e0249680. https://doi.org/10.1371/journal.pone.0249680.

[8] Lee U-J, Ahn S-J, Choi D-Y, Chin S-M, Jang D-S. Airspace designs and operations for UAS traffic management at low altitude. Aerospace, 2023, 10(9): 737. https://doi.org/10.3390/aerospace10090737.

[9] Henderson TC, Marston V, Sacharny D. Multi-modal air trajectory traffic management. In: Intelligent Autonomous Systems 18. Springer, 2024: 247-255. https://doi.org/10.1007/978-3-031-44981-9_22.

[10] Ditta CC, Postorino MN. Three-dimensional urban air networks for future urban air transport systems. Sustainability, 2023, 15(18): 13551. https://doi.org/10.3390/su151813551.

[11] Sun X, Hu Y, Qin Y, Zhang Y. Risk assessment of unmanned aerial vehicle accidents based on data-driven Bayesian networks. Reliability Engineering & System Safety, 2024, 248: 110185. https://doi.org/10.1016/j.ress.2024.110185.

[12] Weinert A, Campbell S, Vela A, Schuldt D, Kurucar J. Well-clear recommendation for small unmanned aircraft systems based on unmitigated collision risk. Journal of Air Transportation, 2018, 26(3): 113-122. https://doi.org/10.2514/1.D0091.

[13] Underhill N, Weinert A. Applicability and surrogacy of uncorrelated airspace encounter models at low altitudes. arXiv preprint arXiv:2103.04753, 2021. https://doi.org/10.48550/arXiv.2103.04753.

[14] Yang J-F, Li X-B, Cai M-X, Wang Y-H, Chen K-Q. Flight safety risk assessment method for low altitude multi-rotor unmanned aircraft system. In: 2019 6th International Conference on Information Science and Control Engineering (ICISCE), 2019: 747-752. https://doi.org/10.1109/ICISCE48695.2019.00153.

[15] Han P, Zhou B, Zhang E. Air collision risk of manned drones in multiple scenarios in the terminal area. Journal of Xihua University (Natural Science Edition), 2022, 41(2): 8-11. https://doi.org/10.12198/j.issn.1673-159X.4153.

[16] Wang W, Gan X, Wu Y, Ren J, Wang C. Low-altitude UAV operation risk assessment method considering the uncertainty. Modern Defense Technology, 2022, 50(5): 14-21.

[17] Shang RR, Hu MH, Yang L, et al. Modeling and assessment method of airspace macro traffic risk under integrated manned and unmanned operation. Journal of Transport Information and Safety, 2025, 43(4): 24-36, 45.

[18] Zhang N, Zhang M, Low KH. 3D path planning and real-time collision resolution of multirotor drone operations in complex urban low-altitude airspace. Transportation Research Part C: Emerging Technologies, 2021, 129: 103123. https://doi.org/10.1016/j.trc.2021.103123.

[19] Dong C, Zhang Y, Jia Z, Liao Y, Zhang L, Wu Q. Three-Dimension Collision-Free Trajectory Planning of UAVs Based on ADS-B Information in Low-Altitude Urban Airspace. arXiv preprint arXiv:2404.18436, 2024. https://doi.org/10.48550/arXiv.2404.18436.

Just Accepted

[20] Guo H, Guo X. Local path planning algorithm for UAV based on improved velocity obstacle method. Acta Aeronautica et Astronautica Sinica, 2023, 44(11): 327586. https://doi.org/10.7527/S1000-6893.2022.27586.

[21] Wang F, Zhu X, Zhou Z, Tang Y. Deep-reinforcement-learning-based UAV autonomous navigation and collision avoidance in unknown environments. Chinese Journal of Aeronautics, 2024, 37(3): 237-257. https://doi.org/10.1016/j.cja.2023.09.033.

[22] Zhang Y, Ding M, Yuan Y, et al. Large-scale UAV swarm path planning based on mean-field reinforcement learning. Chinese Journal of Aeronautics, 2025, 38(9): 103484. https://doi.org/10.1016/j.cja.2025.103484.

[23] Guo CY, Ao WZ, Lyu YH. Analysis on the relationship among low-altitude economy, general aviation, unmanned aerial vehicles and urban air mobility. Money China, 2023, (28): 30-32.

[24] Pongsakornsathien N, El Din Safwat N, Xie Y, Gardi A, Sabatini R. Advances in low-altitude airspace management for uncrewed aircraft and advanced air mobility. Progress in Aerospace Sciences, 2025, 154: 101085. https://doi.org/10.1016/j.paerosci.2025.101085.

[25] Jia Y. A scheme for unmanned aerial system traffic management in low-altitude airspace. Acta Aeronautica et Astronautica Sinica, 2025, 46(11): 531399. https://doi.org/10.7527/S1000-6893.2025.31399.

[26] Zhang H, Zou Y, Zhang Q, Liu H. Future urban air mobility management: Review. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 024638. https://doi.org/10.7527/S1000-6893.2020.24638.

[27] Guan X, Shi H, Xu D, et al. The exploration and practice of low-altitude airspace flight service and traffic management in China. Green Energy and Intelligent Transportation, 2024, 3(2): 100149. https://doi.org/10.1016/j.geits.2024.100149.

[28] Zou YY. Research on calibration method of UAV safety separation in smart city environment. Master's thesis. Nanjing: Nanjing University of Aeronautics and Astronautics, 2021.

[29] Lappas V, Zoumponos G, Kostopoulos V, et al. EuroDRONE, a European unmanned traffic management testbed for U-Space. Drones, 2022, 6(2): 53. https://doi.org/10.3390/drones6020053.

[30] Shen B, Ma Q, Zhang Z, Yang G, Jia W. Intelligent evaluation of robustness of unmanned swarms in cooperative sensing scenario. Acta Aeronautica et Astronautica Sinica, 2026, 47(1): 332118. https://doi.org/10.7527/S1000-6893.2025.32118.

[31] Zhang Y, Zhang H, Tang W, et al. Research on collision avoidance methods for logistics unmanned aerial vehicle based on dynamic controlled interactive collaborative fusion. Journal of Field Robotics, 2026, 43(6): 3845-3867. https://doi.org/10.1002/rob.70219.

[32] Huang X, Cao K, Wang J. A trajectory planning method based on air-ground cooperation. Journal of Highway and Transportation Research and Development, 2016, 33(11): 134-139, 145. https://doi.org/10.3969/j.issn.1002-0268.2016.11.020.

[33] Wang C, Chen X. Robust optimization on disaster emergency blood supply chain based on air-ground transport. Journal of Highway and Transportation Research and Development, 2025, 42(6): 212-222. https://doi.org/10.3969/j.issn.1002-0268.2025.06.022.

[34] Weinert A, Underhill N, Gill B, Wicks A. Processing of crowdsourced observations of aircraft in a high performance computing environment. In: 2020 IEEE High Performance Extreme Computing Conference (HPEC), 2020: 1-6. https://doi.org/10.1109/HPEC43674.2020.9286229.

Just Accepted

[35] Srinivas M, Patnaik LM. Adaptive probabilities of crossover and mutation in genetic algorithms. IEEE Transactions on Systems, Man, and Cybernetics, 1994, 24(4): 656-667. https://doi.org/10.1109/21.286385.

Just Accepted

Author biography



Yuetan Zhang is a Ph.D. candidate at Nanjing University of Aeronautics and Astronautics. His research focuses on risk avoidance in autonomous air-ground collaborative operations of low-altitude aircraft, safety governance of low-altitude traffic systems, and intelligent collaborative control mechanisms in low-altitude airspace. He has edited two textbooks, received four awards at provincial/municipal levels and above, published eight journal papers, and filed 10 invention patents.



Honghai Zhang is Vice Dean, Professor, and Ph.D. Supervisor, and Deputy Director of the State Key Laboratory of Air Traffic Management System. He leads a CAAC key innovation team. He has undertaken six national and 48 provincial/ministerial research projects and received 10 ministerial-level science and technology awards.

Just Accepted