# Degradation of Automatic Dependent Surveillance-Broadcast Integrity: Temporal Clustering and Anomaly Persistence

Eugene Pik[a,*]

[a]Mevocopter Aerospace, 70 Samuel Oster Ave, Vaughan, L4J 7C8, ON, Canada

## Abstract

As safety systems like the Airborne Collision Avoidance System X (ACAS X) increasingly depend on Automatic Dependent Surveillance-Broadcast (ADS-B), the structural reliability of these data streams is critical. Traditional monitors assume independent, stochastic failures. This study challenges that assumption by building a global baseline of longitudinal degradation from 229.0 billion state vectors. The framework integrates multiscale permutation testing with day-resolution, exposure-adjusted survival analysis. Significant temporal clustering appears in 98.9% of the high-utilization commercial cohort ($\geq$20 flight-hours/month; $N$=30,716) and in 67.2% of the sparser remainder under the uniform null; 81.3% of sparse-cohort detections survive a stricter block-preserving null. Clustering peaks at a 3-second characteristic scale, and the mean sustained duration of 28.83 seconds exceeds the 5-second fault tolerance of decentralized fusion trackers. Across the fleet, 58.4% of aircraft record clustered anomalies on two or more distinct observation days, and 55.5% across two or more calendar months. With follow-up censored at the first >30-day observation gap, 56.0% of aircraft experience a first clustered anomaly, with a median of 9 observed days. A recurrent-event hazard model links recurrence to recent utilization (hazard ratio 1.19 per standard deviation) and, more strongly, to network observability coverage (hazard ratio 1.52). The negative time-to-first exposure coefficient is largely an artifact of onset timing. These results shift integrity monitoring from point-in-time anomaly detection to predictive system health management.

*Corresponding author

*Email address:* eugene.pik@mevocopter.com (Eugene Pik)

*Keywords:* Automatic Dependent Surveillance-Broadcast, Surveillance Integrity, Temporal Clustering, Survival Analysis, Air Traffic Management, System Safety, Anomaly Persistence

---

## 1 Nomenclature

**Acronyms**

| | |
|---|---|
| AAM | Advanced Air Mobility |
| ACAS X | Airborne Collision Avoidance System X |
| ADS-B | Automatic Dependent Surveillance-Broadcast |
| ANSP | Air Navigation Service Provider |
| FDR / FWER | False Discovery Rate / Family-Wise Error Rate |
| GET | Global Envelope Thresholding |
| O/E | Observed/Expected Ratio |
| RSP | Required Surveillance Performance |
| TBO | Trajectory-Based Operations |
| UTM | Unmanned Aircraft Systems Traffic Management |

2

**Symbols**

| | |
|---|---|
| $C_{\mathrm{obs}}$ | Cumulative cluster-window mass (observed clustering) |
| $D_{\max}$ | Worst-case integrity loss (maximum duration) |
| $\bar{D}_{\mathrm{cluster}}$ | Sustained degradation duration (mean) |
| $E[C_{\mathrm{sim}}]$ | Expected clustering baseline |
| $\bar{G}_{\mathrm{gap}}$ | Recovery time between clusters (mean) |
| $T_i$ | Discrete time to first clustered anomaly (days) |
| $X_{\mathrm{exp}}$ | Cumulative surveillance exposure (tracking hours) |
| $X_{\mathrm{cov}}$ | Pre-event observability coverage (share of at-risk days observed) |
| $w^*$ | Characteristic degradation temporal scale (peak clustering) |
| $\mathcal{W}$ | Evaluated temporal window scales $\{3, 5, 7, 10, 15\}$ |
| $\beta$ | Cox Proportional Hazards regression coefficient |

3

# 1. Introduction

## 1.1. Background and Motivation

The transition toward 4D Trajectory-Based Operations (TBO) under global modernization frameworks such as NextGen and SESAR fundamentally relies on the continuous, high-fidelity transmission of cooperative surveillance data [1, 2, 3]. Automatic Dependent Surveillance-Broadcast (ADS-B) serves as the primary data link for these predictive systems. However, as established in the foundational GET-25 global baseline, crowdsourced and terrestrial ADS-B networks are subject to severe, localized kinematic inconsistencies, ranging from geometric teleportation to altitude sensor divergence [4, 5].

While the spatial distribution and macro-temporal volatility of these errors have been quantified at the systemic level, the trajectory-level structure of this degradation remains uncharacterized. ICAO Annex 10 and RTCA DO-260B specify stringent integrity requirements for ADS-B OUT performance. Determining whether kinematic instability manifests as isolated, stochastically distributed measurement noise, or as persistent, temporally clustered degradation is critical for assessing compliance with Required Surveillance Performance (RSP) standards. Modern 4D-TBO relies on extrapolating ADS-B position reports for conflict prediction and metering. Temporally clustered kinematic anomalies introduce systematic rather than random errors into these predictions, potentially degrading separation assurance algorithms such as the Airborne Collision Avoidance System X (ACAS X) below required levels of safety [6, 7].

## 1.2. Analytical Framework and Scope

This study shifts the unit of analysis from the macro-level airspace to the longitudinal trajectory of the individual aircraft to systematically investigate the integrity and availability of cooperative surveillance data over time. Persistent anomaly patterns from specific aircraft or geographic regions may indicate degraded surveillance coverage that compromises controller situational awareness, even when individual error magnitudes periodically return to tolerance bounds.

The analytical framework remains entirely descriptive, non-causal, and non-attributional. The methodology isolates structural degradation from usage-driven confounding by evaluating the longitudinal persistence of these anomalies across the global fleet. Root-cause diagnosis of specific anomalies (e.g., intentional GNSS spoofing or electromagnetic interference [8, 9, 10, 11]) is deferred to the discussion of unobserved environmental variables; the focus remains on the degradation of surveillance integrity as perceived by the ground network.

### *1.3. Study Objectives and Organization*

To address the critical gap in surveillance integrity monitoring, this study is driven by four primary research questions. First, it evaluates how the frequency of temporally clustered kinematic events within individual aircraft trajectories compares to the expectation of isolated, independent measurement noise. Second, it aims to determine the characteristic temporal scale of anomaly clustering and how cluster duration and inter-cluster intervals distribute across the fleet. Third, it assesses what proportion of the global fleet exhibits persistent kinematic inconsistencies over the operational year, and how stable these recurrent patterns are. Finally, it investigates whether recurrence risk remains statistically significant after controlling for cumulative surveillance exposure (in hours) and network observability coverage ($X_{\mathrm{cov}}$).

This study makes four contributions. First, it establishes a global, day-resolution baseline of clustered-anomaly onset and persistence across 233,859 aircraft and a full calendar year. Second, it introduces a dual-null significance framework, pairing a uniform null with a block-preserving null and validating both with an empirical calibration study. Third, it specifies a survival design with explicit, contrasted censoring conventions anchored to observation gaps. Fourth, it separates utilization from network observability in recurrence modeling, showing that observability coverage, not flight hours, dominates the hazard.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature and characterizes the gap in longitudinal anomaly persistence modeling. Section 3 details the data architecture and outlines the sequential phases of the analytical framework, from permutation testing to survival modeling. Section 4 presents the multi-scale temporal clustering results and the exposure-adjusted recurrence risk. Section 5 discusses the operational implications for aviation certification and automated safety systems, alongside methodological limitations. Finally, Section 6 concludes the study and outlines directions for future research.

## 2. Literature Review

The expansion of crowdsourced surveillance networks, such as the OpenSky Network [12, 13], has facilitated extensive research into the vulnerabilities of unencrypted ADS-B protocols. Most existing literature approaches ADS-B kinematic inconsistency as a tactical, point-in-time detection problem. Numerous studies propose anomaly-detection frameworks based on Deep Neural Networks [14], Gaussian Mixture Models [15], and two-step Kalman filters [16] to identify discrete GNSS spoofing events and trajectory deviations.

While these highly localized, tactical models offer superior predictive sensitivity for individual state vector corruption, they frequently rely on heavily pre-filtered training datasets bounded to specific airspace sectors over short temporal windows [17]. Standard preprocessing techniques systematically discard incomplete or malformed packets (e.g., enforcing non-null velocity constraints) prior to analysis [18, 19]. As established in recent global baselines, this introduces a profound survivorship bias, inadvertently masking the true systemic observability of the network and ignoring severe structural anomalies such as "Ghost Records" [4, 5].

Furthermore, a critical methodological gap remains regarding the longitudinal structure of these anomalies. While the spatial distribution and cross-sectional macro-volatility of ADS-B errors have been quantified [20], there is a distinct lack of research tracking the *persistence* of kinematic degradation across individual aircraft trajectories over extended operational horizons. The GET-25 baseline established the anomaly thresholds, the severity taxonomy, and the cross-sectional spatial distribution of these errors. What it could not address is whether individual aircraft repeat: the onset timing, recurrence, and hazard structure of anomalies at the trajectory level. Those longitudinal questions are the subject of this study. In aerospace reliability engineering, survival analysis techniques, such as Kaplan-Meier estimators [21] and Cox Proportional Hazards models [22], are standard paradigms for evaluating component fatigue and maintenance intervals. However, these time-to-event frameworks have not yet been applied to evaluate the continuous integrity of cooperative digital surveillance streams.

This gap in longitudinal characterization has direct operational consequences. Modern decentralized tracking algorithms, including those underpinning ACAS X and AAM/UTM frameworks [23], rely on probabilistic state-transition matrices to extrapolate aircraft positions. If kinematic anomalies manifest as isolated, independent measurement noise, they can often be mitigated through standard mathematical smoothing or coasting algorithms. However, if anomalies exhibit non-random temporal clustering or persist across multiple flights for specific airframes, they introduce systematic errors that can severely degrade collision prediction accuracy. Consequently, establishing a multi-scale, exposure-adjusted baseline of anomaly persistence is a prerequisite for hardening the integrity monitoring frameworks of NextGen and SESAR infrastructures.

For readers prioritizing results: primary clustering findings appear in Section 4 (Tables 2 and 3), longitudinal persistence in Section 4.3 (Fig. 2), and exposure-adjusted risk in Section 4.4 (Fig. 3). Methodological details for each phase are outlined sequentially in Section 3, with supporting sensitivity analyses provided in Appendix A.

## 3. Data Architecture and Methodology

### *3.1. Phases of the Analytical Framework*

The analytical framework is implemented via a modular three-stage computational pipeline designed to process 229.0 billion records while preserving trajectory-level kinematic integrity, as illustrated in Fig. 1.

#### *3.1.1. Phase I: Database Architecture and Sequential Streaming*

*Data Extraction Pipeline.* The gross 2025 volume comprises 246.7 billion ADS-B messages, spanning January 1, 2025, 00:00:00 UTC through December 31, 2025, 23:59:59 UTC (January 1, 2026 is the exclusive upper bound, for SQL partition efficiency). From this volume, an optimized out-of-core sequential streaming architecture isolates 229.0 billion filtered, airborne state vectors. To extract the data, a custom pipeline was engineered to execute initial data sanitation, temporal sorting, and sequential state-pairing via native SQL window functions directly within the OpenSky Network's Trino cluster. The extraction enforces a 60-second maximum gap constraint to preserve kinematic continuity.

*Kinematic Thresholding.* Trajectories were evaluated against the Global Envelope Thresholding (GET-25) P99 baseline constraints established in prior global analysis [4], utilizing the open-access GET-25 replication dataset [5]. To ensure self-contained reproducibility, the kinematic thresholds defining an anomalous state vector are: derived speed $> 6670.3$ m/s, geometric-barometric altitude divergence $> 2323.1$ m, sequential heading delta $> 170.7°$, sequential speed delta $> 84.9$ m/s, and vertical rate delta $> 13.0$ m/s.

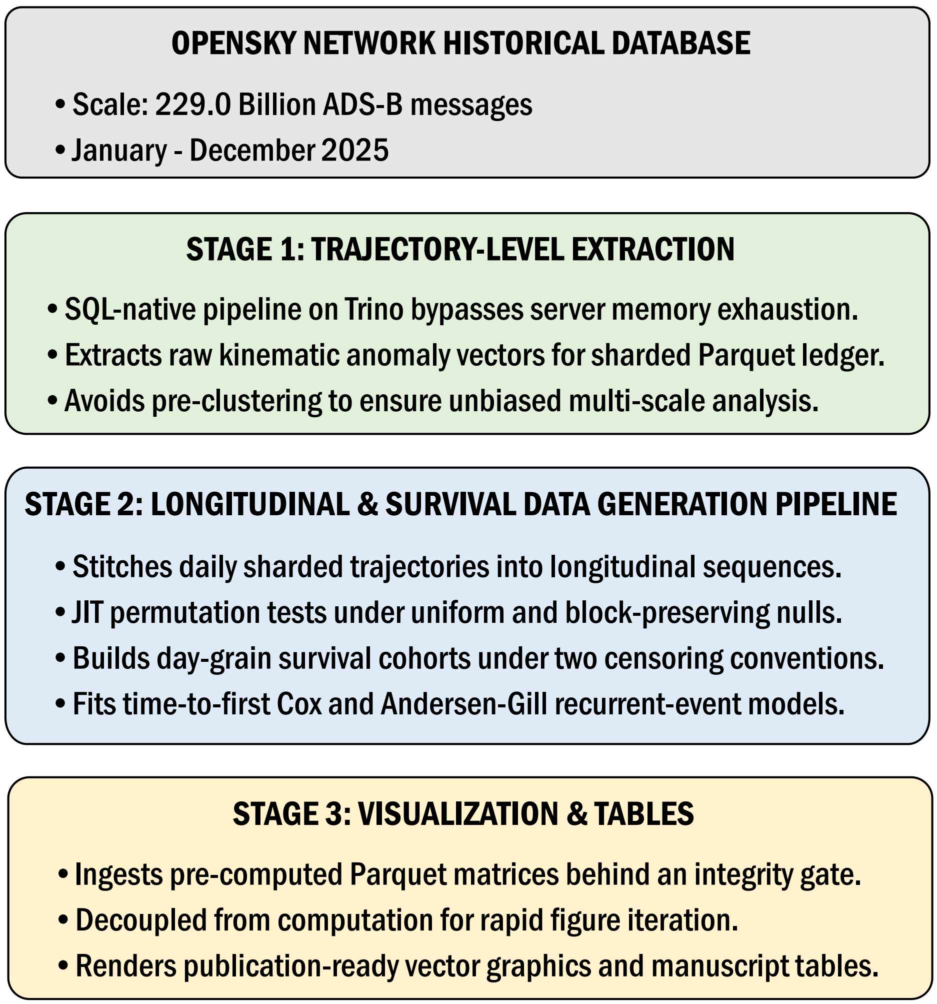


Figure 1: Conceptual architecture of the three-stage computational pipeline, illustrating the transition from distributed SQL extraction to longitudinal survival modeling.

*Data Sanitization Protocols.* To ensure dataset integrity and reproducibility, the pipeline applies POSIX regular expression filtering to validate structural formatting. It incorporates wildcard exclusions (`0000%` and `FFFF%`) alongside specific target exclusions (`000000`, `FFFFFF`, and `123456`) to eliminate known malformed, unallocated, or factory-default transponder hex blocks. Furthermore, to ensure mathematical stability during cluster evaluation, aircraft-day cross-sections (the daily extraction grain) containing fewer than 5 valid sequential message pairs were excluded from the analysis cohort. All deterministic permutation seeds were cryptographically salted to sever any theoretical correlation with sequential ICAO24 manufacturer block allocations. A full-key integrity scan across all 229.0 billion rows found zero duplicate (`icao24`, timestamp) pairs, so no aircraft-second is counted twice. Because every kinematic test compares fields within a single message pair, shared network-timing artifacts cannot manufacture divergence between those fields.

*Spatial Pre-filtering.* To prevent GPS micro-jitter and float-precision errors from spuriously triggering the $v_{\text{derived}}$ teleportation threshold during near-stationary ground movements, an orthogonal bounding-box spatial pre-filter of 0.004° (maintaining an approximate 444-meter longitudinal boundary globally via cosine-scaling) was enforced. This was applied prior to the heavy Haversine kinematic evaluation to computationally reject stationary jitter.

*Memory Management.* To prevent worker-node memory exhaustion during the map-reduce phase of global aggregation, trajectory sequences are chunked into discrete daily bounds, utilizing atomic writes and full-file MD5 checksums to guarantee data integrity and support robust state-recovery. The resulting high-dimensional trajectory matrices are processed locally via the PyArrow engine. To bypass the out-of-memory (OOM) failures typical of monolithic data ingestion, the pipeline dynamically downscales ingestion batch sizes (e.g., from 500,000 rows during database extraction to 50,000 rows during localized evaluation) to safely stream Parquet row groups.

During this trajectory reconstruction, to prevent sliding analytical windows from bridging independent observations across midnight boundaries, a precise buffer equal to the maximum evaluated window scale ($w_{\max} = 15$) is injected between daily arrays. This halts cross-day window bleeding. The buffer slots are counted inside the total trajectory length $L$. Simulated anomalies can land in them; observed anomalies never can. The buffer share of $L$ is small (mean 0.4% of slots; 90th percentile 0.8%), so its dilution of the uniform null is negligible. The injected-independence calibration study confirms this empirically: the unconditional false-positive rate is 0.040 against the nominal 0.05 in every density stratum (Fig. A.5).

#### *3.1.2. Phase II: Scrambled-Temporal Permutation Test*

To test for temporal clustering while controlling for non-stationary flight phases (e.g., altitude-dependent maneuvering variance), observed clustering is evaluated against a distribution-free scrambled-temporal control group.

*Algorithm Optimization.* Shuffling full binary arrays of length $L$ scales at $\mathcal{O}(N \times L)$ and is computationally prohibitive. The methodology therefore exploits the sparsity of kinematic errors. The algorithm isolates the $k$ indices where anomalies occur and redraws these positions as independent uniform draws across the interval $L$, sampling with replacement. Position collisions, possible in principle, render the null marginally *more* clustered and the test correspondingly conservative; a dedicated sensitivity re-test of the 500 densest FDR-significant aircraft using exact without-replacement sampling ($N$=1,000) produced zero significance changes and a maximum $|\Delta p|$ of 0.000. Re-running the cluster classification with duration-standardized windows ($w' = 5\,\mathrm{s}/\overline{\Delta t}$ per aircraft-day) reproduces the fixed-index classification across all 18,243,276 aircraft-days: 60 discordant aircraft-days out of 4,391,404 flagged (Jaccard 0.99999). Index windows therefore correspond to wall-clock seconds for this fleet.

By deploying a highly optimized two-pointer sliding window algorithm across the sorted anomaly arrays, this formulation bounds computational complexity to $\mathcal{O}(N \times k \log k)$. To ensure computational feasibility across the global fleet, this core simulation loop is accelerated via pure C-level Just-In-Time (JIT) compilation using Numba, enabling the execution of $N = 10{,}000$ permutations per aircraft in milliseconds rather than seconds.

*Statistical Thresholding and FDR Correction.* Sustained duration metrics (Section 3.1.3) use strict 1D array convolution to capture total cumulative sliding-window mass, but executing $N = 10{,}000$ such convolutions per aircraft is computationally prohibitive. The permutation engine therefore deploys the two-pointer boundary heuristic to tally distinct cluster occurrences. This tally slightly undercounts true convolution mass, so the significance threshold it supports is conservative while remaining computationally feasible.

For each aircraft, $N = 10{,}000$ permutations are executed without early stopping, giving a minimum resolvable $p$-value of 0.0001 and ample headroom for the Benjamini-Hochberg False Discovery Rate (FDR) correction ($\alpha = 0.05$) [24], so that highly kinematically unstable aircraft are ranked accurately. Furthermore, because testing identical aircraft across expanding window scales ($\mathcal{W} = \{3, 5, 7, 10, 15\}$) inherently tests the same structural temporal hypothesis, applying the FDR correction independently within each scale risks inflating the Family-Wise Error Rate (FWER) [24]. To control the expected proportion of false rejections across the entire study, the Benjamini-Hochberg FDR correction ($\alpha = 0.05$) is applied globally across the concatenated pool of all evaluated window sizes, rather than attempting overly conservative FWER control that would severely degrade statistical power.

*Effect Size Quantification.* The statistical difference between observed clustering ($C_{\text{obs}}$) and expected clustering ($E[C_{\text{sim}}]$) is quantified using the Observed/Expected (O/E) ratio and the standardized $Z$-statistic, defined as $Z = (C_{\text{obs}} - E[C_{\text{sim}}])/\sigma_{\text{sim}}$. To prevent division-by-zero instability when $E[C_{\text{sim}}]$ approaches zero (common among sparse operators with few expected clusters), expected counts are floor-clamped at $\epsilon = 0.1$ prior to O/E ratio calculation. This preserves ordinal rankings while ensuring finite ratios. Under this formulation, a positive $Z$-statistic directionally indicates excess temporal clustering beyond the independent noise baseline.

*Conservatism of the Uniform Null Model Under Non-Stationarity.* A known characteristic of crowdsourced ADS-B surveillance is the presence of strong temporal non-stationarity, driven primarily by diurnal traffic cycles and regional operational constraints (e.g., nighttime cargo flights). The uniform scrambled-temporal null model acts as a highly conservative benchmark for cluster detection under these conditions.

Consider an aircraft experiencing $k$ legitimate kinematic anomalies concentrated entirely within a narrow operational subinterval $I \subset [1, L]$ (such as a 4-hour nighttime flight). Under a block-preserving null model that restricts permutations strictly to interval $I$, the probability of a cluster forming by random chance is governed by the dense binomial coefficient $\binom{|I|}{w}$. Conversely, under the deployed uniform null, the $k$ anomalies are distributed across the vastly larger global timeframe $L$, significantly lowering the baseline probability of dense cluster formation. Because the uniform null disperses the baseline hazard, the observed data must exhibit high clustering density to achieve statistical significance ($p < 0.05$). Consequently, this approach may overestimate the significance threshold for aircraft with non-uniform observation patterns.

The reported proportions of kinematically unstable aircraft should therefore be interpreted as conservative lower bounds on systemic prevalence.

Two structural properties of the index space merit explicit statement. First, the permutation universe comprises the 190,830 aircraft with at least one anomaly, $L \geq 20$, and $k \leq 50{,}000$. Anomaly-free aircraft are untestable by construction and are excluded from all significance denominators. Second, sequential pairs with $\Delta t > 60$ s are removed before indexing. Same-day flights separated by ground time therefore sit adjacent in index space, less one pair, and observed windows can span such gaps. The block-preserving null bounds this exposure by confining permutation to one-hour operational blocks. It retains 83.2% of uniform-null detections at $w^*$=5 (Fig. A.6).

To address the potential for anti-conservatism among low-utilization operators (e.g., business jets with sparse observation patterns), the analytical framework executed a supplementary block-preserving sensitivity analysis. This localized test restricted permutation shuffling within contiguous active flight blocks. To guarantee comparability across heterogeneous transponder equipment, block boundaries were standardized to absolute time rather than raw index counts. The permutation engine dynamically defined the mathematical boundaries by dividing a one-hour temporal separation threshold (3,600 seconds) by the aircraft's specific median update interval ($\Delta t$). Permutation shuffling within these active flight blocks was bounded to a symmetric boundary envelope of $\pm w$ indices applied to the absolute outermost observed anomalies (yielding a total padding of $2w$ indices), strictly clamped to the trajectory limits.

Because this boundary scales dynamically with the evaluated window size, for the maximum evaluated scale of $w = 15$, this symmetric padding represents a total buffer of 30 seconds for commercial aircraft operating at a standard 1 Hz update rate (scaling up to 150 seconds for sparse operators transmitting at 0.2 Hz). Rather than preventing ground-gap bleeding (which is natively handled by the block segmentation itself), this localized margin preserves true intra-block temporal density across heterogeneous equipment profiles by confining simulated anomalies to the immediate operational neighborhood of the actual events. Independent flights are thereby isolated consistently across the fleet, whether an aircraft broadcasts at 1 Hz or 0.2 Hz, and the block-preserving null partitions operations in true wall-clock time across heterogeneous equipment profiles.

While this conservatism is robust for high-utilization commercial airliners, it inverts for highly sparse operators (e.g., business jets logging only 10 hours per month), where the uniform null can become artificially anti-conservative. To neutralize this localized bias, final incidence rates are dual-reported under both the uniform-null and the block-preserving null assumptions. As an interpretation guideline, the uniform-null model is primary for continuous, high-utilization commercial fleets, where its artificial dispersion acts as a conservative lower bound. For sparse, low-utilization operators (e.g., general aviation and business jets), the block-preserving null is treated as primary, protecting against artificial inflation of statistical significance. To keep the two null assumptions comparable, the block-preserving analysis mirrors the primary model by applying the Benjamini-Hochberg FDR correction globally across the concatenated $\mathcal{W}$ pool.

*Empirical Subset Validation.* To empirically validate that reduced permutation counts suffice for identifying extreme outliers, a supplementary study compared classification concordance between $N = 200$ and $N = 10,000$ permutations on a stratified subset of 500 aircraft. On this subset, the reduced $N = 200$ benchmark reproduced the full-run significance classifications with 100.0% concordance at the standard $\alpha = 0.05$ threshold. While $N = 200$ lacks the mathematical resolution to achieve the $p < 0.0001$ margin required for rigorous global FDR headroom, it provided sufficient baseline stability to validate the permutation architecture prior to deploying the full $N = 10{,}000$ global execution.

#### *3.1.3. Phase III: Multi-Scale Rolling Window Analysis*

A custom computational pipeline developed by the author replicates the permutation test across a predefined set of window sizes $\mathcal{W} = \{3, 5, 7, 10, 15\}$. These window scales span operational regimes from rapid burst detection ($w = 3$, approximately 3 seconds at a 1 Hz update rate) to sustained degradation identification ($w = 15$, approximately 15 seconds). Intermediate values capture the characteristic ACAS X coasting threshold ($w = 5$) and typical maneuvering durations ($w = 7$, $w = 10$). Formally, a temporal window of size $w$ is classified as *clustered* if it contains $\tau \geq 2$ anomalous observations, a threshold specifically selected to mirror the fault-tolerance limits of $\alpha - \beta$ fusion trackers in automated collision avoidance systems [7]. For each $w \in \mathcal{W}$, the $C_{\text{obs}}(w)$ and empirical distributions are recalculated to identify the dominant granularity (characteristic scale) of integrity loss.

Crucially, while this clustering analysis strictly evaluates the binary presence of systemic degradation (independent of specific physical failure modes), the subsequent survival modeling framework evaluates the persistence of distinct severity classes to identify distinct hardware failure modes. This dual-resolution approach intentionally separates the detection of systemic integrity loss from the etiological tracking of localized failures.

The methodology further decomposes clustering into four orthogonal temporal metrics: $C_{\text{obs}}$ (cumulative cluster-window mass), $\bar{D}_{\text{cluster}}$ (sustained degradation duration), $D_{\max}$ (worst-case integrity loss), and $\bar{G}_{\text{gap}}$ (recovery time). By defining $C_{\text{obs}}$ as the cumulative sum of valid sliding-window positions rather than a discrete event frequency, the metric accurately captures the total volumetric severity of the clustered anomalies. To provide a standardized operational baseline, the pipeline extracts these sustained duration metrics consistently at the $w = 5$ window scale. This scale was selected *a priori* from operational safety thresholds, not from post-hoc statistical optimization. It mirrors the fault-tolerance limits of decentralized collision avoidance systems such as ACAS X, where an airborne $\alpha$-$\beta$ fusion tracker mandates degraded coasting after two consecutive corrupted target hits within a 5-second horizon [7].

Algorithmically, the sequential rolling evaluation in Python (utilizing zero-padded convolution) mathematically replicates the logic of a backward-looking `ROWS BETWEEN 4 PRECEDING` SQL window function. Executing this rolling evaluation locally in Python, rather than during the initial database extraction, prevents pre-aggregation bias and enables true independent multi-scale testing. Furthermore, to prevent computational memory exhaustion without inducing selection bias against high-frequency commercial airliners, trajectories exceeding 100,000 observation windows were evaluated utilizing a randomly sampled, contiguous 100,000-window slice to accurately estimate these duration metrics.

#### *3.1.4. Phase IV: Time-to-First Anomaly Probability Modeling*

To evaluate the longitudinal persistence of kinematic instability, the analysis shifts from static point-in-time anomaly detection to time-to-event survival analysis. The methodology deploys a prevalent cohort design, evaluated via time-to-first-event modeling from the onset of observation. In this framework, "Time Zero" ($t_0$) is anchored at each aircraft's first observed day of transmission within the dataset. Just as the foundational GET-25 framework resolved kinematic survivorship bias by actively retrieving malformed state vectors [4], this longitudinal design must address observational survival biases. Because the cohort evaluates aircraft already in active operation, the design acknowledges the left-truncation bias inherent to unobservable early equipment failures (true hardware inception dates remain unobservable).

By conditioning survival from first sight, the analysis bypasses the need for delayed-entry risk set modifications (e.g., the Tsai-Wang estimator [25]), measuring the survival probability from the onset of network observability. While the multi-scale permutation test identifies the theoretical structural peak of clustering ($w^*$), the longitudinal survival modeling evaluates time-to-anomaly at a fixed operational baseline of $w = 5$ with a minimum cluster threshold of $\tau \geq 2$. As established in Phase III (Section 3.1.3), these constraints mirror ACAS X fault-tolerance thresholds, anchoring the survival probabilities to a physical operational reality rather than a shifting statistical optimum.

The time-to-event scale $T$ is measured in *days* from each aircraft's first observed day, reconstructed from the daily extraction ledgers whose calendar dates are preserved in the shard naming; monthly aggregation is retained only as a reported sensitivity. Day resolution keeps the tie mass of the discrete-time estimators roughly thirtyfold below monthly aggregation, while the estimators remain actuarial life-table approximations at the daily level.

Let $T_i$ denote the discrete time to the first clustered kinematic anomaly for aircraft $i$, measured in days from $t_0$. Because the observational horizon is bounded to a single 12-month calendar year (2025), chronological year-boundary wrap-around cannot occur, ensuring all computed discrete intervals remain positive. To account for aircraft that exit the observation window before experiencing a clustered anomaly, the survival model employs right-censoring. The censoring indicator $\delta_i$ is defined such that $\delta_i = 1$ if the anomaly is observed, and $\delta_i = 0$ if the observation is right-censored. Censoring occurs either due to the administrative end of the study period (December 31, 2025) or due to severe network dropouts.

To prevent immortal time bias [26], in which an aircraft appears at risk while it is unobservable, dropout censoring is anchored to the first inter-observation gap that exceeds the 30-day absence threshold; follow-up ends there. This convention, termed gap-anchored censoring, is the primary specification. A second convention, event-priority censoring, lets a later clustered anomaly count as an event even when such a gap precedes it; it is reported as a sensitivity (Fig. A.1). The two conventions differ for 33,703 aircraft: gap-anchored censoring yields 131,067 events (56.0%), while event-priority censoring yields 164,770 (70.5%). The observation calendar used for gap detection is the same valid-position, $\geq$5-pair daily universe in which anomalies are defined. This choice is conservative relative to any-message daily spans.

The survival framework non-parametrically estimates the baseline time-to-first-anomaly survival function using the Kaplan-Meier product-limit estimator [21]:

$$S(t) = \prod_{t_j \le t} \left(1 - \frac{d_j}{n_j}\right) \tag{1}$$

where $S(t)$ represents the probability that an aircraft maintains surveillance integrity (i.e., experiences no clustered anomaly) beyond day $t$. The term $d_j$ denotes the number of aircraft experiencing a first clustered anomaly at time $t_j$, and $n_j$ denotes the risk set, representing the number of aircraft still active and under observation just prior to $t_j$. To test for structural differences in degradation trajectories among different physical error modes, the analytical framework stratifies the failed cohort by severity class and statistically compares the onset curves using the non-parametric log-rank test. Because permanent survivors (aircraft with zero lifetime anomalies) inherently lack a localized error mode, this specific log-rank evaluation compares the conditional time-to-onset distributions among the sub-cohort of aircraft that ultimately experience kinematic degradation.

#### *3.1.5. Phase V: Exposure-Adjusted Risk Modeling*

While the Kaplan-Meier estimator provides a robust descriptive baseline, raw time-to-onset profiles fail to account for differential aircraft utilization. An aircraft operating 300 hours per month naturally presents more mathematical opportunities to encounter localized RF interference or coverage gaps than an aircraft operating 20 hours per month.

To isolate intrinsic structural tracking vulnerabilities from usage-driven exposure confounding, this study deploys an Exposure-Adjusted Risk Model utilizing the Cox Proportional Hazards architecture [22]. The hazard function $h(t \mid \mathbf{X})$, which represents the instantaneous risk of a clustered anomaly at time $t$ given a set of covariates $\mathbf{X}$, is formulated as:

$$h(t \mid \mathbf{X}) = h_0(t) \exp\left(\beta_1 X_{\text{exp}} + \beta_2 (X_{\text{exp}} - \tilde{X}_{\text{exp}})^2 + \beta_3 X_{\text{cov}}\right) \tag{2}$$

where $h_0(t)$ is the unspecified baseline hazard function. To assess non-linear threshold effects, the model incorporates a quadratic exposure term. Raw linear and quadratic terms correlate heavily, so the linear exposure covariate is median-centered ($\tilde{X}_{\text{exp}}$) before polynomial expansion ($\beta_2 (X_{\text{exp}} - \tilde{X}_{\text{exp}})^2$).

This orthogonalization bounds the Variance Inflation Factor (VIF) and the condition number of the design matrix. It resolves multicollinearity while preserving the detection of non-linear risk. Predictive discrimination is quantified with the concordance index (C-index), read *a priori* on standard conventions: 0.5 is random chance, 0.5–0.7 poor to moderate, 0.7–0.8 good, and $> 0.8$ excellent. The baseline (time-to-first) model evaluates all covariates strictly on the pre-event history, which prevents contemporaneous look-ahead bias. To capture dynamics and recurrence beyond the first event, an Andersen–Gill counting-process model [27] complements it over the aircraft-month panel. Exposure enters lagged by one month, and coverage enters as a running pre-period share, so the covariates stay strictly predictable while utilization varies through time. Recurrent cluster-months constitute its events.

*Cumulative Surveillance Exposure ($X_{exp}$).* This covariate captures the tracking exposure (in hours) strictly prior to the initial event. To prevent non-linear scaling errors caused by heterogeneous transponder equipment, the daily volumetric exposure is first calculated independently for each active day (total consecutive message pairs multiplied by the median update interval $\Delta t$, divided by 3,600). To eliminate contemporaneous look-ahead bias, the overall exposure covariate is derived as the cumulative sum of daily surveillance hours strictly before the event day for event aircraft, and through the censoring day for censored aircraft. Aircraft failing on their first observed day accumulate 0.0 pre-exposure hours.

Missing observation days (e.g., extended unobservable oceanic routing) contribute zero exposure, which may slightly conflate routing-based unobservability with true non-usage. Because aircraft failing early in their observation history inherently accumulate near-zero pre-exposure hours, a supplementary sensitivity analysis excluded all first-interval failures (events in an aircraft's first observed month) to verify that the Cox coefficients are not distorted by this zero-exposure subpopulation.

Because this extraction explicitly drops inter-message gaps exceeding 60 seconds, this metric rigorously captures cumulative continuous sequence duration rather than gross chronological observation time. It isolates uninterrupted network tracking exposure and is distinct from actual airborne flight-hours (Hobbs time). Furthermore, to protect the Cox model from severe outlier values induced by data artifacts, the preprocessing script winsorizes this covariate at the 99th percentile. While a logarithmic transformation ($\log(X_{\mathrm{exp}})$) is a standard statistical approach for right-skewed utilization data, it could not be applied to this framework due to two fundamental constraints. Mathematically, aircraft failing on their first observed day accumulate exactly 0.0 pre-exposure hours; a $\log(0)$ transformation yields $-\infty$, and shifted alternatives (e.g., $\log(X+1)$) introduce arbitrary scaling distortions for fractional hours. Physically, in unencrypted crowdsourced ADS-B networks, extreme upper-tail outliers frequently do not represent genuine high-utilization fleets (e.g., long-haul cargo). Instead, they represent hex-code duplication or transponder spoofing, in which multiple emitters transmitting the same identifier concurrently accrue physically impossible flight times (e.g., $> 744$ hours in a single 31-day month). Winsorization effectively clamps these physically impossible artifacts to the maximum plausible utilization boundary without permanently excluding the corrupted airframes from the survival cohort.

*Pre-Event Observability Coverage ($X_{cov}$).* This covariate captures how completely the network observed the aircraft during its at-risk interval: the share of at-risk days on which the aircraft was actually observed, computed strictly on the pre-event history. The median update interval itself is uninformative here. After the retention filters, its 10th, 50th, and 90th percentiles all equal 1.00 s across 18.24 million aircraft-days (99th percentile: 1.017 s), so it carries essentially no variance. Coverage is therefore the operative observability dimension. It separates continuously tracked continental operations from intermittently observed routing.

Because the temporal interval to first clustered anomaly $T$ is evaluated as a discrete, integer-valued daily metric to maintain computational feasibility across 229.0 billion records, heavily tied event times are expected across the fleet. During the maximization of the partial likelihood, these ties induce matrix singularities. To resolve this collinearity across the tied daily bounds, the Cox model implementation relies on standard baseline tie-handling methods (Breslow's approximation [28]) in conjunction with an L2 (Ridge) regression penalizer (penalty coefficient $\lambda = 0.1$). This penalizer is applied uniformly across all evaluated covariates, mitigating variance inflation arising specifically from the correlation between the centered linear and quadratic exposure terms. The penalty value was selected after a sensitivity analysis across $\lambda \in \{0.01, 0.05, 0.1, 0.5, 1.0\}$, with Breslow tie-handling held fixed for comparability. Across this grid (Fig. A.3), coefficient magnitudes shrink monotonically with $\lambda$, as expected under L2 regularization, while signs and rank ordering do not change. All coefficients in the text are reported at

$\lambda = 0.1$.

Additionally, a supplementary analysis confirmed the robustness of the survival estimates to the network dropout censoring threshold, evaluated concurrently at 15, 30, 45, and 60 days (Fig. A.2). To assess whether informative censoring (e.g., oceanic dropouts) masks the true kinematic hazard rate, the comprehensive demographic profiles of the censored cohort versus the event cohort are explicitly tracked and fully reported in Table 1. Visual comparison between the two censoring conventions is provided in Fig. A.1. Table 1 confirms that exposure and operational profiles remain systematically characterized across survival states under both conventions.

Table 1: Demographic Profiles of Censored vs. Event Cohorts

| **Status** | **Aircraft Count** | **Median Time ($T$, days)** | **Median Exposure (hrs)** |
|---|---|---|---|
| *Gap-anchored censoring (primary)* | | | |
| Event Observed | 131,067 | 9 | 4.33 |
| Network Dropout (>30 d) | 97,386 | 2 | 1.28 |
| Administrative Censoring | 5,406 | 18 | 3.79 |
| *Event-priority censoring (sensitivity)* | | | |
| Event Observed | 164,770 | 19 | 3.58 |
| Network Dropout (>30 d) | 63,683 | 1 | 0.91 |
| Administrative Censoring | 5,406 | 18 | 3.79 |

*Note:* Median follow-up and pre-endpoint surveillance exposure by survival state, under both censoring conventions. Dropout-censored aircraft show short follow-up and low exposure, consistent with early departure from network coverage rather than with hidden events.

Furthermore, to provide intuitive visual analytics, the continuous Cox model is decoupled from the categorical visual representation. The fleet is discretized into equal-sized exposure tertiles (light, moderate, heavy). To resolve the boundary overlap inherent to highly zero-inflated cumulative exposure data, the algorithm attempts strict quantile discretization, dynamically falling back to rank-based percentage binning ($X_{\text{exp}} \rightarrow \text{Tertiles}$) where tied bounds trigger matrix singularities. These tertiles are subsequently plotted as purely descriptive, exposure-stratified Kaplan-Meier curves, ensuring that the visual representations remain independent of the continuous parametric constraints.

Regarding the proportional hazards assumption, scaled Schoenfeld residual tests [29] were computed on the day-grain model for all covariates (full outputs in `model_schoenfeld_daygrain.txt`). All three covariates formally reject proportionality ($p < 5 \times 10^{-5}$). At $n$=233,859, even modest time-variation in effects is decisive. A sensitivity fit that excludes first-interval events localizes the source. The exposure coefficient attenuates from $-0.431$ to $-0.272$ per standard deviation, while the coverage coefficient strengthens to $+1.268$. The baseline-model hazard ratios are therefore read as time-averaged associations over the year. The Andersen–Gill companion model absorbs this drift through its time-varying covariates and carries the dynamic interpretation.

### *3.2. Computational Infrastructure and Reproducibility*

Processing 229.0 billion ADS-B surveillance messages required a highly optimized out-of-core architecture to prevent memory exhaustion. Stage 1 data sanitization and sequential state-pairing were executed utilizing native SQL window functions on a distributed Trino cluster. To ensure total scientific and computational reproducibility, the subsequent localized analysis was decoupled into two distinct processing stages. Stage 2 (Data Generation) executes the multi-scale permutation tests and survival modeling, utilizing process-based parallelization to bypass memory constraints during CPU-bound matrix evaluations. The pipeline then saves these intermediate results to disk. Stage 3 (Visualization and Reporting) subsequently ingests these pre-computed Parquet matrices to generate all tables, statistical models, and figures. This architectural decoupling explicitly separates the massive computational loads from the aesthetic iteration of results.

All extraction scripts, configuration ledgers, and synthesized matrices are made openly available via the project's repository. While optimal performance scales with available hardware, the localized pipeline integrates dynamic RAM auto-throttling (via `psutil`), systematically stepping down `loky` worker-process parallelization to prevent crashes on machines with as little as 8 GB of system memory. Under unthrottled conditions, Stage 1 extraction requires approximately 72 hours on the Trino cluster, and Stage 2 requires approximately 48 hours, depending on fleet volume and hardware. All exact library dependencies (including `numba`, `joblib`, `psutil`, `lifelines` [30], `statsmodels`, `pyarrow`, and `trino`) are pinned within the repository's `requirements.txt` environment specification.

## 4. Results

The findings are derived from the evaluation of 233,859 unique aircraft trajectories observable via crowdsourced networks, spanning 229.0 billion filtered state vectors and 63.6 million tracked surveillance hours across the 2025 calendar year. The results characterize both the short-term clustering of kinematic anomalies and their longitudinal persistence.

### 4.1. Temporal Clustering of Kinematic Noise

To evaluate whether kinematic inconsistencies manifest as isolated measurement errors or structured bursts, the scrambled-temporal permutation test was executed across the global fleet. To explicitly prevent the uniform null model from artificially inflating significance rates among sparse, low-utilization operators (e.g., general aviation), the headline clustering prevalence was strictly stratified by utilization intensity. Among the high-utilization commercial cohort ($\geq 20$ flight hours/month; $N$=30,716), 98.9% of aircraft exhibited significant temporal clustering under the uniform independent noise model (FDR-corrected $p < 0.05$). In the sparse operator cohort ($N$=160,114), 67.2% were significant under the same null, and 81.3% of those detections were retained under the stricter block-preserving null.

As detailed in Table 2, which reports the aggregate uniform baseline, the median $Z$-statistic was positive at 45.13: the shift is toward excess clustering, not anti-clustering. The median Observed/Expected (O/E) ratio was 33.69 (IQR: 18.88–51.91). When anomalies occur, they concentrate within discrete temporal windows rather than spreading through the flight path.

As described in Section 3, because the uniform null distributes probability mass across the entire trajectory, the expected baseline ($E[C_{\text{sim}}]$) can approach zero for sparse operators. With expected simulation counts floor-clamped at 0.1 to resolve division instability, the reported median O/E ratio characterizes the full, representative sample of significant aircraft, serving as a descriptive magnitude of deviation, while formal statistical significance remains gated exclusively by the FDR-corrected $p$-values and standardized $Z$-statistics. At this sample size, statistical power is effectively unlimited, and even trivial deviations reach $p < 0.001$. The O/E ratio therefore carries the operational interpretation: it establishes the physical magnitude of clustering that threatens ACAS X fault-tolerance limits, independently of raw sample size. The dual-reported incidence rates are aggregated and compared across all window scales in Table 3.

Table 2: Per-Aircraft Temporal Clustering Statistics Across Multi-Scale Window Analysis. Median O/E ratios exceeding unity across all scales confirm systematic temporal clustering, with peak deviation at $w = 3$ indicating multi-second anomaly burst structure.

| Window Size ($w$) | $w = 3$ | $w = 5$ | $w = 7$ | $w = 10$ | $w = 15$ |
|---|---|---|---|---|---|
| Total Aircraft Evaluated | 190,830 | 190,830 | 190,830 | 190,830 | 190,830 |
| Significant Aircraft ($p < 0.05$, FDR) | 71.1% | 72.3% | 72.8% | 73.1% | 73.2% |
| Median $Z$-statistic | 71.84 | 45.13 | 35.19 | 28.68 | 24.70 |
| Median O/E Ratio$^\dagger$ | 64.13 | 33.69 | 23.84 | 17.82 | 14.03 |
| O/E IQR | 36.45–100.70 | 18.88–51.91 | 13.00–37.96 | 9.33–30.39 | 7.03–25.87 |
| Mean $C_{\text{obs}}$ | 328.88 | 430.97 | 548.99 | 750.82 | 1151.96 |
| Mean $E[C_{\text{sim}}]$ | 9.89 | 25.75 | 48.37 | 94.38 | 201.22 |

$^\dagger$O/E ratios reported only for aircraft with $E[C_{\text{sim}}] \geq 1.0$ to prevent division instability. Note: All statistical tests used $N = 10{,}000$ permutations with global FDR correction at $\alpha = 0.05$.

Table 3: Comparison of Significance Rates (Uniform vs. Block-Preserving Null)

| **Window Size ($w$)** | **Total Tested** | **Significant (Uniform)** | **Significant (Block)** | **Block Retention Rate** |
|---|---|---|---|---|
| $w = 3$ | 135,922 | 133,138 | 117,845 | 88.51% |
| $w = 5$ | 138,311 | 135,591 | 114,106 | 84.15% |
| $w = 7$ | 139,402 | 136,927 | 107,118 | 78.23% |
| $w = 10$ | 140,055 | 138,088 | 98,755 | 71.51% |
| $w = 15$ | 140,484 | 139,771 | 93,213 | 66.69% |

*Note:* Retention of statistical significance when permutations are confined to contiguous active flight blocks, insulating sparse operators from uniform-null assumptions.

Across all windows, the stratified significance rates span 98.8–99.0% for the high-utilization cohort ($N$=30,716) and 65.7–68.3% for the sparser cohort ($N$=160,114). Among uniform-significant aircraft, the block-preserving null retains 89.8% of the high-utilization cohort and 81.3% of the sparse cohort at $w^*$=5 (Fig. A.6). The headline prevalence is therefore robust to the null geometry in the stratum where the uniform null is least conservative.

### *4.2. Multi-Scale Degradation Signatures*

To eliminate pre-aggregation bias, the data extraction pipeline evaluated trajectories for raw, instantaneous kinematic violations. The multi-scale rolling window permutation test was then executed independently across $\mathcal{W} = \{3, 5, 7, 10, 15\}$ to identify the characteristic temporal scale of surveillance degradation without circular dimensionality constraints. The fleet-aggregate O/E ratio peaked at $w^* = 3$, showing that kinematic corruption is most highly structured at the 3-second granularity.

While $w^* = 3$ represents the statistically dominant clustering scale globally, duration metrics are reported at the $w = 5$ operational baseline. At this scale, the sequential scan revealed that the mean sustained degradation duration ($\bar{D}_{\text{cluster}}$) was 28.83 seconds, with the worst-case integrity loss ($D_{\max}$) spanning 4,930 seconds. Because Phase I bounds kinematic evaluation to sequential messages received within 60 seconds of one another, this $D_{\max}$ represents 82 minutes of continuous corrupted transmission without the transponder ever going silent for more than a minute. The operational danger is that these transponders fail active rather than silent: instead of dropping out, which tracking systems can safely coast through, they keep injecting severe geometric and velocity errors into the network. The mean recovery time between clusters ($\bar{G}_{\text{gap}}$) averaged 11,825.07 seconds (approximately 3.28 hours), reflecting the episodic nature of these severe tracking failures.

### *4.3. Longitudinal Persistence and Time-to-Anomaly*

Moving beyond intra-flight clustering, the longitudinal analysis follows each aircraft day by day from its first observed transmission. The Time-to-Onset Survival Function (Kaplan–Meier estimator, Fig. 2) shows the probability that an aircraft remains free of clustered anomalies as observed days accumulate. To prevent conditioning-on-event bias during the visual stratification of specific failure phenotypes, the plot incorporates the global baseline curve. This curve retains the right-censored permanent survival cohort (aircraft maintaining perfect integrity with zero clustered anomalies) to ensure an unbiased representation of systemic onset probabilities.

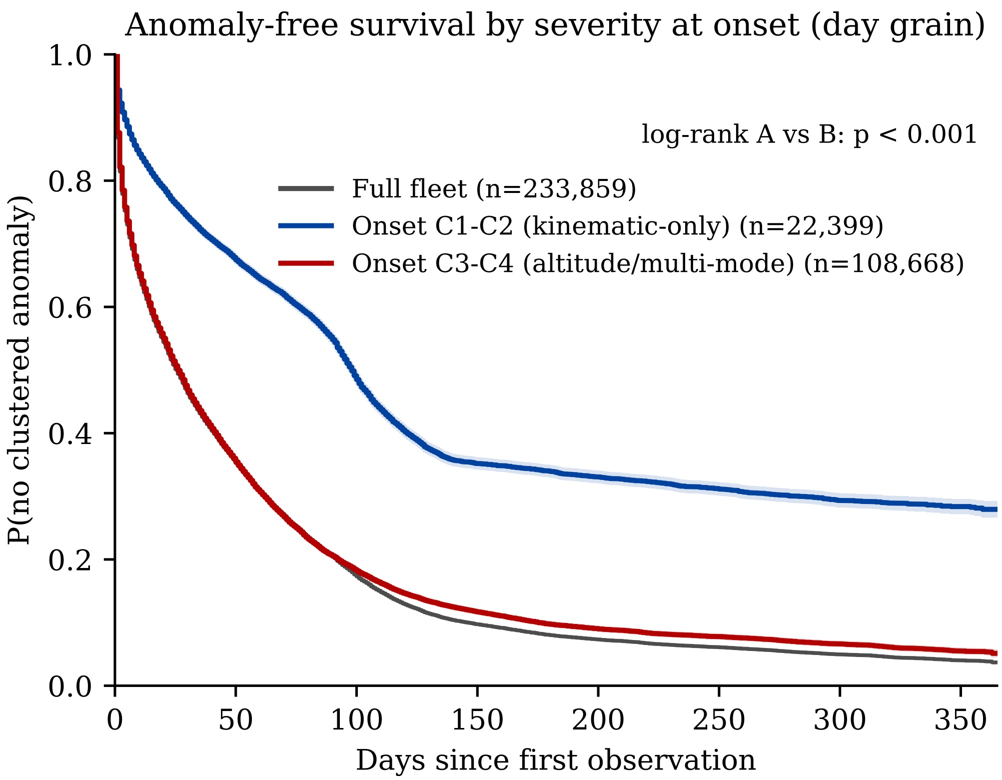


Figure 2: Time-to-First Clustered Anomaly by Severity at Onset (day-grain Kaplan–Meier; gap-anchored censoring). Onset class is the maximum severity on each aircraft's first cluster-day; all 131,067 events are classified.

*Note:* The global baseline curve retains the right-censored permanent survival cohort to ensure unbiased representation of systemic onset probabilities. Shaded regions denote 95% confidence intervals.

To prevent arbitrary calendar demarcations from masking high-frequency degradation, longitudinal persistence was evaluated using two co-primary endpoints. Under the strict definition, 58.4% of the fleet ($N$=136,584) fell into the recurrent cohort, experiencing clustered anomalies on two or more distinct *observation days*; evaluated across two or more distinct *calendar months*, the persistent share is 55.5% ($N$=129,885). However, evaluating the relaxed co-primary endpoint ($\geq$ 2 total clusters regardless of calendar boundaries) expanded the persistent cohort by an additional 27,046 aircraft

$(N = 163,630)$. This confirms that true anomaly persistence, inclusive of rapid same-day sequential degradation, is structurally higher than distinct-day counting implies. Under gap-anchored censoring, the probability of remaining free of clustered anomalies is 0.463 at 30 observed days, 0.204 at 90 days, and 0.037 at 365 days. Under the event-priority convention, the 365-day value is 0.004.

The log-rank test between onset cohorts A (Classes 1–2) and B (Classes 3–4) yields $\chi^2$=634.5, $p$=$5.3 \times 10^{-140}$ (global four-class test $p < 10^{-300}$); median time-to-first is 9 observed days for Cohort B versus 11 for Cohort A. Aircraft whose first cluster-day involved severe geometric divergence or multi-vector corruption (Severity Classes 3 and 4) degraded much faster than those starting with isolated kinematic jumps (Classes 1 and 2). At 90 observed days, the severe-onset cohort retains a 20.5% probability of remaining anomaly-free, against 54.8% for the milder cohort.

Crucially, these internal severity classifications directly map to the structural anomalies identified in the foundational GET-25 baseline. To guarantee standalone interpretability, the severity classes are defined by their physical kinematic failure modes, using the quantitative thresholds of Section 3. **Severity Class 1** (COVERAGE_LATENCY) covers isolated derived-speed exceedances (*Teleportation*). **Severity Class 2** (AVIONICS_VELOCITY) covers extreme, physically impossible jumps in transmitted ground speed (*Avionics Velocity Jump*) or vertical rate (*Vertical Jolt*). **Severity Class 3** (AVIONICS_SENSOR) covers severe kinematic divergence, including unphysical heading reversals (*Directional Jitter*) and 3D geometric decoupling such as large barometric-geometric altitude mismatches (*Sensor Divergence*). **Severity Class 4** (MULTIMODE) covers simultaneous multi-vector corruption, for example concurrent teleportation and altitude failure. Within this taxonomy, Severity Classes 3 and 4 encapsulate the "Ghost Record" phenomenon [4]. A Ghost Record is severe geometric teleportation in which the broadcast position decouples from the actual aerodynamic trajectory, often with zeroed or frozen avionics telemetry. It is distinct from null-packet omissions, which the extraction architecture bypasses structurally. To remain aligned with the non-attributional analytical framework of this study, these cohorts utilize purely neutral physical severity descriptors, avoiding the attribution of these kinematic discrepancies to specific root-cause hardware failures (e.g., RF jamming versus altimeter degradation).

The survival analysis maintains an effective sample size of 131,067 events under gap-anchored censoring. Across the three continuous predictors of the multivariable Cox model (linear exposure, quadratic exposure, and observability coverage), this yields an Events-Per-Variable (EPV) ratio of 43,689. This exceeds the recommended minimum threshold of 10, ensuring robust coefficient estimation and preventing statistical overfitting of the polynomial hazard term [31].

### *4.4. Exposure-Adjusted Risk*

To isolate structural hardware and routing vulnerabilities from simple usage-driven confounding, an Exposure-Adjusted Cox Proportional Hazards Model was fitted to the survival data (Fig. 3). The model controlled for cumulative pre-event surveillance exposure ($X_{\text{exp}}$, linear and centered-quadratic) and pre-event observability coverage ($X_{\text{cov}}$), with robust (sandwich) standard errors.

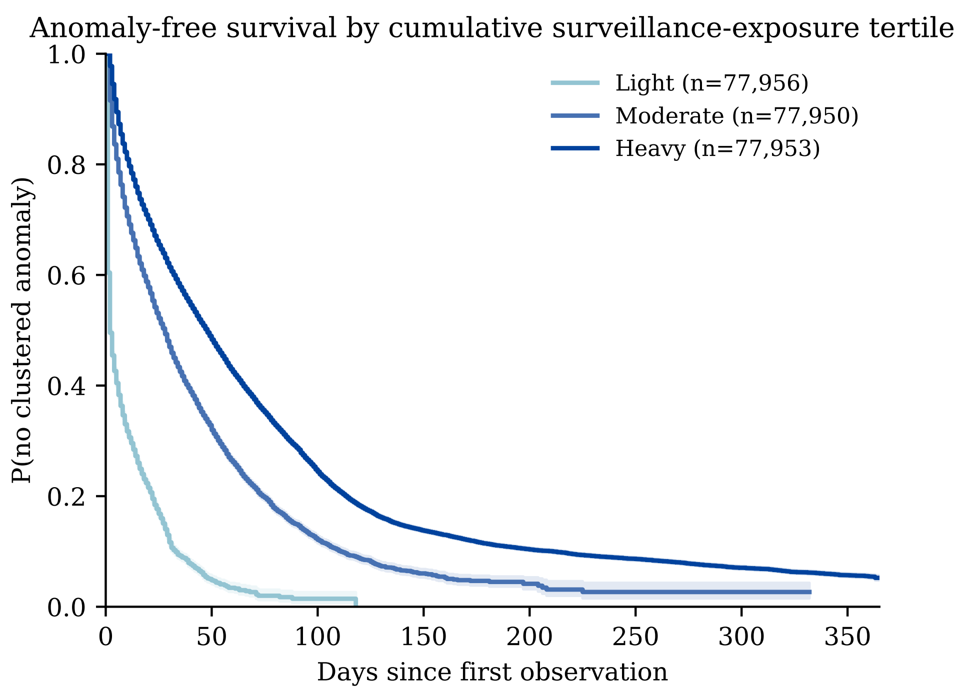


Figure 3: Exposure-Stratified Time-to-First Clustered Anomaly (cumulative surveillance-hour tertiles; day-grain, gap-anchored censoring). Tertiles are descriptive strata, decoupled from the continuous model.

*Note:* Curves are descriptive visualizations of anomaly-free probability by observed exposure tertile. They are distinct from the continuous exposure-adjusted Cox proportional hazards model.

In the day-grain time-to-first model, the standardized exposure coefficient is $\beta = -0.431$ per standard deviation (HR 0.650; robust $p < 10^{-3}$). The centered quadratic term is +0.158. Observability coverage carries +0.264 per standard deviation (HR 1.302). Concordance is $C$=0.627, moderate under the a-priori scale of Section 3.1.5. All coefficients are reported at $\lambda$=0.1 (Fig. A.3).

Three companion analyses fix the interpretation of the negative exposure sign. First, excluding first-interval events attenuates it by 37%, to −0.272, while coverage strengthens to +1.268. Aircraft that fail early necessarily accrue little pre-event exposure, so the timing of onset drives much of the sign. Second, a placebo suite permutes the covariate vectors across aircraft (100 refits). The null exposure coefficient is 0.0001 (SD 0.0027) and null concordance is 0.501. The observed coefficient lies about 160 null standard deviations away, which excludes fitting artifact.

Third, the Andersen–Gill recurrent-event model spans 1,146,596 aircraft-months with 655,588 recurrent cluster-months. Once exposure varies through time, the sign reverses. Lagged monthly utilization carries HR 1.19 per standard deviation, and running observability coverage carries HR 1.52, the dominant term. Recurrence therefore tracks how recently and how completely the network observes an aircraft. It does not reflect a protective effect of flying more.

## 5. Discussion

### 5.1. *Aerospace Operational Implications of Characteristic Scaling*

The multi-scale analysis (Section 3.1.3) fundamentally recharacterizes how kinematic instability manifests within the global airspace. If surveillance degradation were simply a product of isolated, stochastic measurement noise (e.g., random transponder bit-flips), the Observed/Expected (O/E) ratio would remain relatively flat across all evaluated window sizes ($\mathcal{W}$). However, the identification of a distinct peak at the $w^*$ characteristic scale indicates that kinematic corruption is highly structured.

From an operational perspective, this structural clustering has direct implications for the design of ground-station receiver algorithms and decentralized collision avoidance systems. The empirical identification of a characteristic scale peaking at $w^* = 3$, coupled with sustained duration metrics exceeding the $w = 5$ fault-tolerance limits, suggests that short-term degradation is primarily driven by transient environmental factors. These likely include localized RF packet collisions in high-density terminal maneuvering areas or momentary GPS multipath interference, rather than extended, systemic infrastructure failures.

Current tactical tracking algorithms typically rely on simple moving-average smoothing or coasting logic designed to reject single-point outliers. However, the empirical presence of dense, multi-second anomaly bursts indicates these foundational filters are insufficient. To maintain Required Surveillance Performance (RSP) compliance, ANSPs and developers of next-generation ACAS X and UTM frameworks should deploy multi-scale, rolling-window integrity monitors. Such monitors need to quarantine corrupted bursts spanning the identified $w^*$ duration before those errors propagate into 4D trajectory predictions.

Furthermore, current ADS-B integrity requirements specified in RTCA DO-260B and DO-260C [32, 33] mandate that uncontained position errors must be alerted within strict Time-to-Alert (TTA) bounds, typically 2 to 10 seconds depending on the phase of flight. The finding that mean sustained degradation duration reaches 28.83 seconds, with worst-case events spanning 4,930 seconds (approximately 82 minutes), suggests that a subset of the global fleet may operate outside DO-260B and DO-260C compliance parameters for extended intervals. However, because the kinematic thresholds detect anomalies exceeding GET-25 P99 bounds, substantially more stringent than baseline RTCA geometric limits, direct equivalence to certification non-compliance requires additional calibration.

### *5.2. Aviation Certification and the Nature of Persistence*

While transient environmental interference explains short-term clustering, the longitudinal survival analysis (Section 4.3) reveals a more concerning pattern: the persistence of kinematic instability across multi-month horizons. The Kaplan-Meier onset curves show that a distinct sub-cohort of the global fleet fails to maintain surveillance integrity over time.

The exposure-adjusted models isolate the nature of this persistence. Were recurrence usage-driven, utilization would carry a positive hazard: aircraft that fly more would simply meet more environmental hazard. The evidence is more specific. In the time-to-first frame the exposure coefficient is negative, but the companion diagnostics show this is largely a consequence of onset timing rather than a protective effect. Excluding first-interval events removes a third of the coefficient. A covariate-permutation placebo shows the fitted machinery contributes essentially nothing.

Once exposure varies through time, the Andersen–Gill model reverses the sign. Recent monthly utilization raises the recurrence hazard (HR 1.19 per standard deviation). Pre-period observability coverage dominates both frames (HR 1.52 recurrent; 1.30 time-to-first). The picture is one of network-perceived integrity. Clustered-anomaly persistence concentrates where and when the surveillance network observes an aircraft most completely and most recently. Observability structure and routing exposure matter more than airframe wear in these data. For monitoring practice, fleet-health indicators built on these data must therefore be conditioned on coverage; otherwise, apparent degradation trends will partly mirror the network's own visibility field.

Based on these findings, the following measures are recommended for ANSPs and tracking system developers:

1. Implement rolling-window cluster detection ($w \in \{3, 5\}$) in ground station processing chains;
2. Establish fleet-level persistence tracking using exposure-adjusted survival frameworks;
3. Prioritize inspection of aircraft exceeding the 95th percentile O/E ratio for three or more consecutive months.

### *5.3. Methodological Limitations*

To ensure rigorous interpretation of the findings, the following methodological constraints must be acknowledged.

Three statistical limitations bound the survival inferences. First, proportional hazards is formally rejected for all covariates at day grain. The reported baseline-model ratios are time-averaged associations, and the time-varying recurrent-event model carries the dynamic interpretation. Second, day resolution reduces tie mass by roughly thirtyfold, but $T$ remains discrete at the daily level. Third, for sparse aircraft the significance of any single cluster is weak evidence on its own. Under injected independence, a sparse aircraft forms a cluster only 12.8% of the time, so conditioning on an observed cluster already conditions on a rare outcome (Fig. A.5). The persistence endpoints and the pooled FDR therefore carry the sparse-cohort inference. The unconditional false-positive rate of the deployed test is 0.040 against the nominal 0.05.

#### 5.3.1. *Exclusion of Ground-Based Kinematics*

To prevent stationary GPS micro-jitter and terminal multipath interference from inducing artificial derived speeds, the analysis excluded all ground-based position reports (`onground = true`). The transponder-derived `onground` flag can misreport; older transponders sometimes stay in airborne status while taxiing. A supplementary altitude filter, such as $> 1000$ ft AGL, was rejected because it would blind the analysis to initial climbs and final approaches, where low-altitude multipath is prevalent. Instead, the pipeline requires a minimum spatial displacement of 0.004° (about 444 meters) between sequential coordinates before kinematic evaluation is triggered. This safeguard ensures that stationary taxiway jitter cannot be misclassified as a high-speed airborne anomaly. This study therefore quantifies airborne surveillance integrity.

#### 5.3.2. *High-Utilization Aircraft Sampling and Seeding*

To prevent fatal worker-node JVM memory exhaustion within the SQL extraction cluster, the export of raw anomaly arrays was capped via a `SLICE` limit of 250,000 indices per aircraft, per month. For scale, 250,000 corrupted events at a standard 1 Hz update rate equate to approximately 69.4 hours of continuous kinematic failure. Aircraft exceeding this ceiling are categorized at the maximum tier of degradation. Pipeline execution logs confirm that zero airframes (0.00%) within the global fleet triggered this truncation threshold. Because survival modeling (Section 4.3) evaluates the time to the *first* clustered anomaly, capping subsequent anomalies beyond 250,000 events does not alter the aircraft's entry into the "Persistent" cohort.

However, to prevent worker-node memory crashes within the `loky` multiprocessing environment during Phase II, an explicit computational safety limit was implemented. Aircraft exceeding 50,000 cumulative anomalies are completely skipped during the multi-scale permutation test. Consequently, while the SQL truncation preserves initial entry into the persistent survival cohort, the most severely degraded trajectories (>50,000 anomalies) are entirely excluded from the temporal clustering metrics. The reported permutation metrics ($Z$-statistic and O/E ratio) therefore represent a conservative lower bound of systemic network instability.

Furthermore, to prevent computational memory exhaustion during the localized execution of duration metrics (Section 3.1.3), trajectories exceeding 100,000 contiguous observation windows ($L > 100,000$) were evaluated using a randomly sampled 100,000-window slice. For context, this threshold represents approximately 27.8 hours of continuous tracking at a standard 1 Hz update rate, or nearly 139 hours for operators transmitting at 0.2 Hz. While this ensures that the most safety-relevant, high-utilization commercial airliners are retained in the characteristic timescale ($w^*$) analysis, it introduces a localized sampling constraint. To mitigate phase-dependent sampling bias (e.g., anomalies clustering predominantly during terminal approach), the algorithm evaluates aircraft triggering this memory guard across up to three randomly sampled contiguous slices across the trajectory. However, because the computational algorithm inherently bounds the maximum random starting index at $L - 100,000$ to ensure a full slice is extracted, this methodology systematically under-samples the absolute final observation windows for these extended trajectories. The aggregated worst-case maximum reported for $D_{\max}$ must therefore be interpreted as a conservative lower bound that may inadvertently suppress terminal-phase dropouts. Approximately 57% of the evaluated fleet triggered this memory guard threshold. To preserve statistical independence across the fleet, the random slice boundaries were generated from a deterministic SHA-256 seed combining each aircraft's `icao24` identifier with the global configuration salt (`salt_aerospace_42`), preventing correlated slice positions across sequentially allocated ICAO24 blocks.

#### *5.3.3. Two-Sided Sensitivity of Block-Preserving Permutations*

As detailed in Section 3.1.2, a block-preserving null model was deployed to validate the robustness of the uniform permutation test. To evaluate both Type I and Type II error rates among low-utilization operators, this sensitivity analysis covered every aircraft that reached statistical significance, plus a stratified random sample of non-significant aircraft ($N = 5{,}000$ per window scale; $N = 25{,}000$ globally). This two-sided design tests whether uniform-null anti-conservatism suppressed the significance of sparsely operating fleets, so that true clustering is not masked by the baseline assumptions.

#### *5.3.4. Discrete Survival Evaluation and Sub-Daily Persistence*

Time-to-event is evaluated on discrete calendar days reconstructed from the daily extraction ledgers. Recurrence within a single day is therefore not separable from one cluster-day, and a day boundary can split a rapid sequence: clusters late on one day and early on the next count as two cluster-days, while an equally rapid pair inside a single day counts as one.

The two co-primary persistence endpoints defined in Section 4 bracket the operational recurrence rate and prevent calendar demarcations from masking short-term sequential degradation. Day-level resolution captures rapid recurrence; sub-daily dynamics remain a target for localized continuous-time follow-up studies.

### 5.3.5. Informative Censoring and Oceanic Dropouts

A fundamental challenge in longitudinal surveillance analysis is distinguishing between an aircraft ceasing anomalous behavior and an aircraft simply becoming unobservable. In the survival framework, aircraft were right-censored if they experienced a network dropout exceeding 30 consecutive days. While this prevents "missing data" from artificially inflating apparent compliance, it introduces the potential for informative censoring. Aircraft operating predominantly in oceanic or remote airspace (e.g., the NAT corridor) are more likely to experience these 30-day dropouts due to the line-of-sight limitations of terrestrial crowdsourced receivers. If these remote regions inherently cause higher rates of geometric teleportation anomalies, censoring these aircraft could mathematically suppress the true systemic hazard rate.

The observability-coverage covariate ($X_{\text{cov}}$) mitigates this bias by adjusting the hazard for how completely the network observed each aircraft before its endpoint. However, to test whether reported results are robust to oceanic censoring patterns, a supplementary "continental-only" sensitivity analysis was executed. This analysis restricted the cohort strictly to aircraft that never experienced an inter-observation gap exceeding 7 days, thereby completely isolating the model from trans-oceanic line-of-sight dropouts. Results confirmed that the longitudinal hazard trajectories remained structurally consistent when oceanic aircraft were excluded, indicating that informative censoring does not materially distort the primary conclusions (Fig. A.4).

Because the dataset is bounded to the 2025 calendar year, mid-flight gaps crossing the December 31 boundary are truncated. Year-end dropouts are therefore treated as administrative right-censoring rather than network dropout censoring; the aggregate impact on the longitudinal survival estimates is negligible. Additionally, because the discrete risk-set evaluates inter-observation gaps on daily bounding timestamps, a large intra-day dropout (an aircraft observed briefly at 00:01 UTC and again at 23:59 UTC) is recorded as a continuous tracking day. This intra-day blindness marginally overestimates true hourly network coverage.

### 5.3.6. Proportional Hazards Assumption Viability

As detailed in Section 3.1.5, survival times are discrete at the daily level, and tied event times are handled with Breslow's approximation during partial likelihood maximization. Day resolution keeps the tie mass roughly thirtyfold below monthly aggregation, which allowed the scaled Schoenfeld residuals to be evaluated for every covariate. Those tests reject strict proportionality. The continuous Cox coefficients are therefore presented as time-averaged hazard effects across the longitudinal horizon. Should the true baseline hazard fluctuate with seasonal flight volumes, these coefficients represent the weighted mean of that varying risk rather than a fixed constant.

## 6. Conclusion

As global airspace modernization increasingly relies on unencrypted, decentralized cooperative surveillance, the structural integrity of the Automatic Dependent Surveillance-Broadcast (ADS-B) network is paramount. This study extends the planetary-scale spatial baseline of GET-25 into the temporal domain, providing a trajectory-level characterization of kinematic anomaly persistence.

By analyzing 229.0 billion state vectors through an optimized out-of-core computational pipeline, this study shows that kinematic inconsistencies are not merely isolated, stochastic measurement errors. Instead, they manifest as highly structured, temporally clustered bursts, with a characteristic degradation scale of 3 seconds. Furthermore, the application of exposure-adjusted survival analysis revealed that 58.4% of the global fleet records clustered anomalies on two or more distinct observation days (55.5% across two or more calendar months), a persistence that cannot be explained by flight-hour exposure alone.

These findings have direct operational implications for the deployment of 4D Trajectory-Based Operations (TBO) and automated collision avoidance frameworks such as ACAS X. Tactical tracking algorithms need to evolve beyond simple single-point outlier rejection to incorporate multi-scale, rolling-window integrity monitors capable of quarantining sustained data corruption. Additionally, the baseline recurrence probabilities established herein provide aviation regulatory bodies and Air Navigation Service Providers (ANSPs) with an empirical, traffic-normalized framework to identify specific airframes or transponder configurations that routinely fail to meet ICAO and RTCA DO-260B performance standards. Future research can build on these longitudinal baselines to train multi-parameter machine learning classifiers capable of attributing these persistent degradation signatures to their root physical or electromagnetic causes.

From a certification perspective, current RTCA DO-260B and DO-260C compliance testing evaluates ADS-B OUT performance under static, controlled conditions and may not capture the dynamic, longitudinal degradation patterns observed here. Aviation regulatory bodies should consider supplementing laboratory-based type certification with continuous monitoring requirements derived from operational surveillance baselines.

## Appendix A. Supplementary Figures

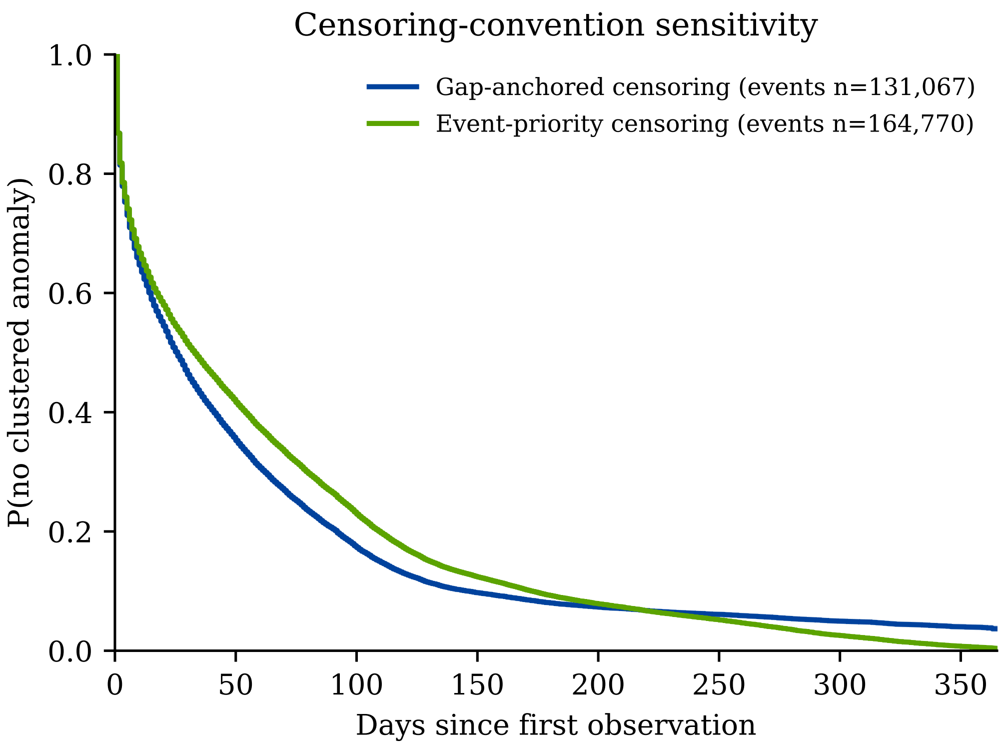


Figure A.1: Sensitivity Analysis: Censoring Convention. Baseline Kaplan–Meier curves under gap-anchored censoring (131,067 events) and event-priority censoring (164,770 events).

*Note:* The conventions differ only in whether a cluster occurring after a qualifying >30-day gap counts as an event.

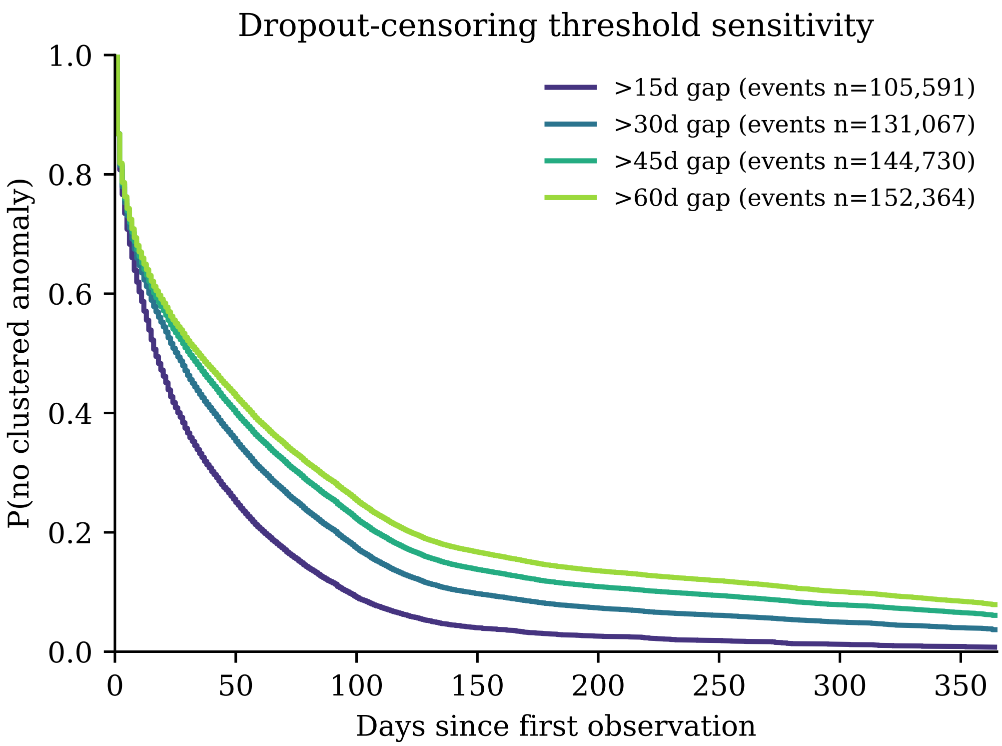


Figure A.2: Sensitivity Analysis: Network Dropout Censoring Thresholds (15, 30, 45, and 60 days; gap-anchored censoring).

*Note:* Event counts are 105,591, 131,067, 144,730, and 152,364. The event-priority count is 164,770 at every threshold. The threshold moves only the censor, never the clusters, and the survival profile keeps its shape.

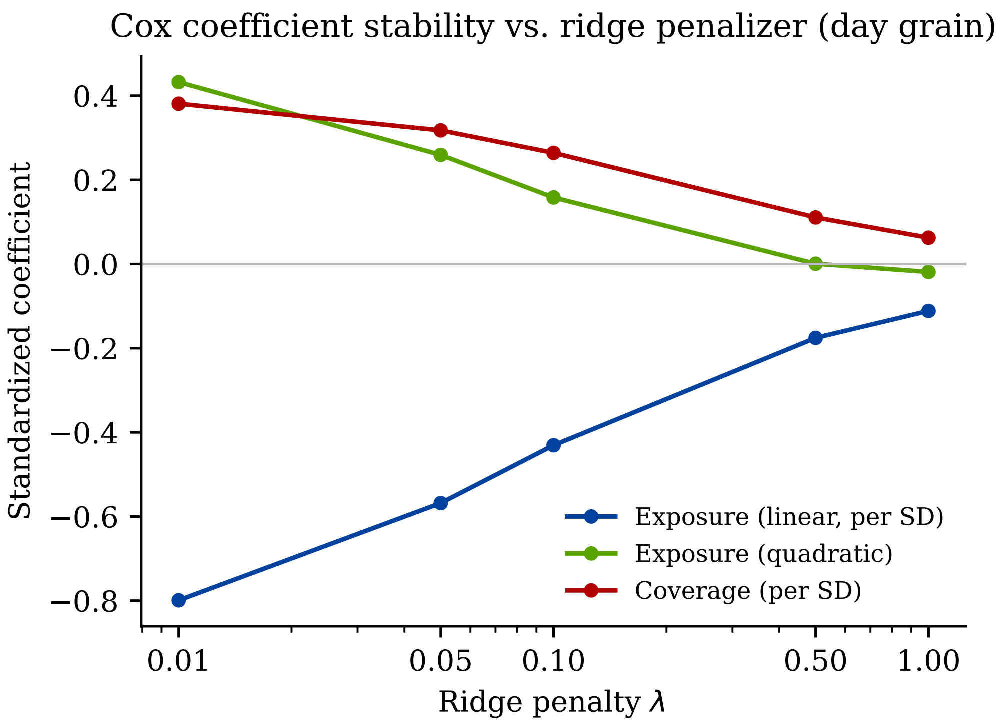


Figure A.3: Cox Coefficient Behavior vs. Ridge Penalizer ($\lambda$), day-grain model.

*Note:* Coefficient magnitudes shrink monotonically with $\lambda$, as expected under L2 regularization. Signs and covariate ordering do not change anywhere on the grid. All coefficients in the text are reported at $\lambda$=0.1.

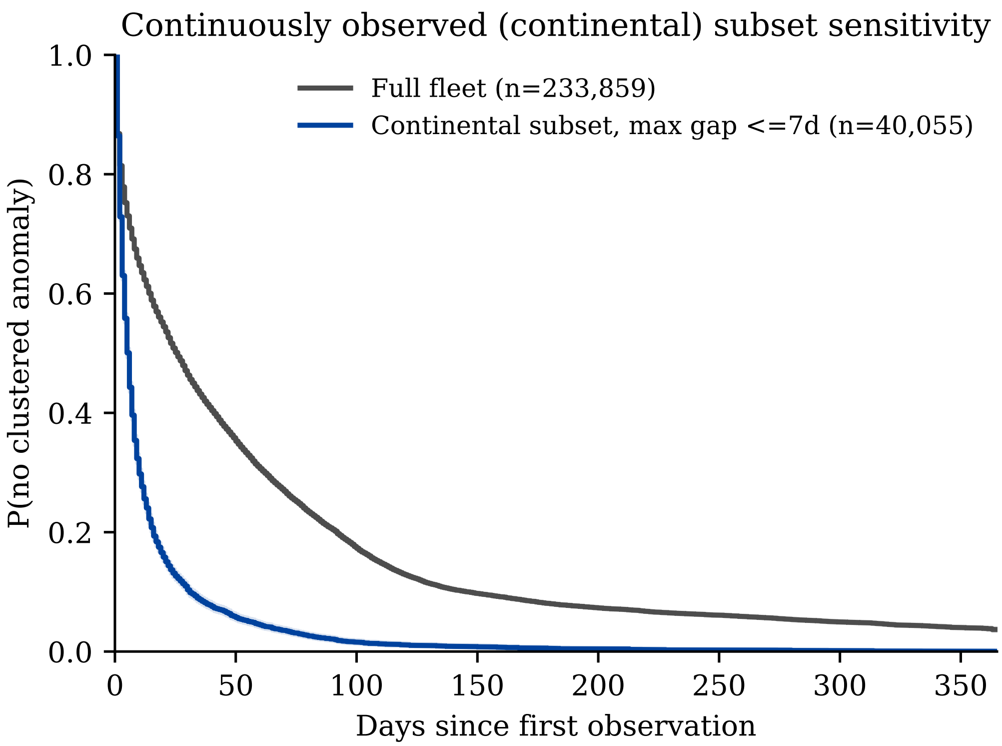


Figure A.4: Sensitivity Analysis: Continuously Observed (Continental) Subset.

*Note:* The subset holds aircraft whose maximum internal observation gap is at most 7 days ($n$=40,055; 14,782 events). Its survival profile tracks the full fleet. This rules out oceanic line-of-sight dropout as the driver of the headline curves.

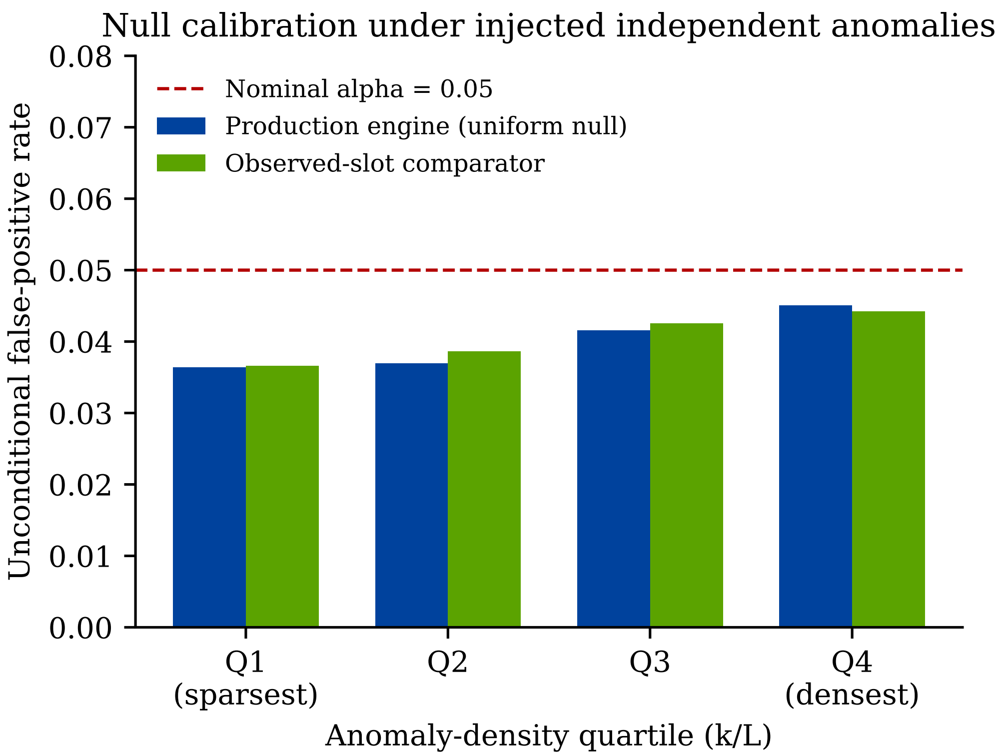


Figure A.5: Null Calibration Under Injected Independent Anomalies.

*Note:* 1,250 simulated aircraft; 50 injections each; $N$=1,000 permutations. The production engine and the observed-slot comparator sit below the nominal $\alpha$=0.05 in every density quartile and agree to about 0.001. The day-boundary buffer slots are immaterial, and the deployed test is slightly conservative.

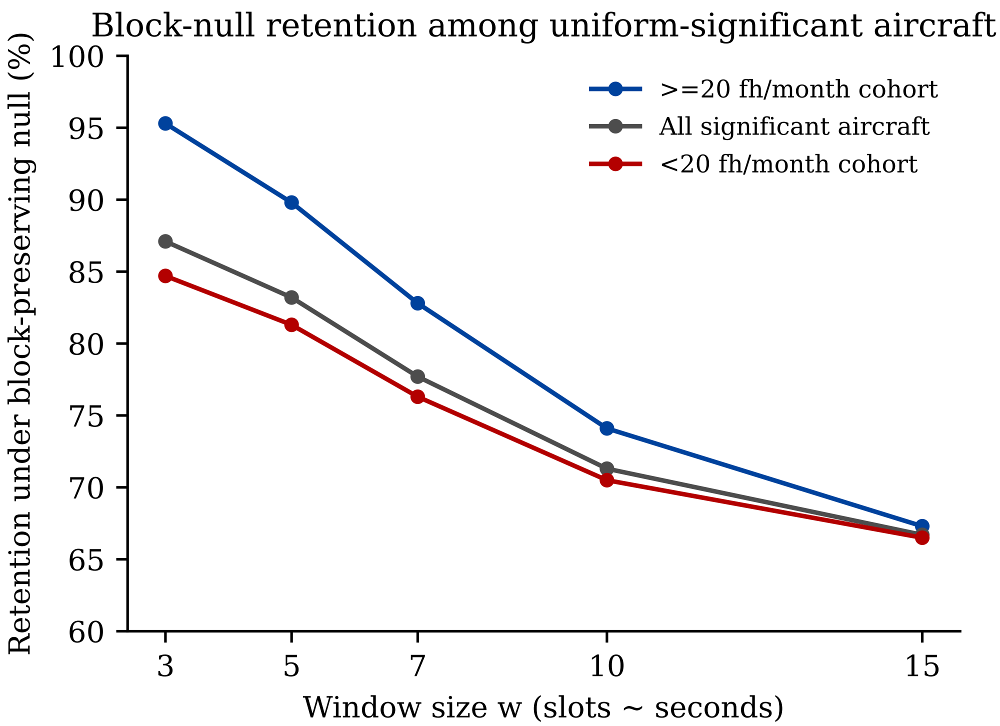


Figure A.6: Block-Preserving Retention Among Uniform-Significant Aircraft.

*Note:* Retention by window size and utilization stratum. At $w^{*}=5$: 89.8% for the $\geq 20$ flight-hours/month cohort, 81.3% below it, and 83.2% overall.

## Data Availability Statement

The raw ADS-B state vectors utilized in this longitudinal study were provisioned by the OpenSky Network. To ensure complete scientific and computational reproducibility, all extraction scripts, configuration ledgers, and synthesized survival matrices have been open-sourced alongside the foundational GET-25 benchmark. The repository includes the cryptographic salts used for deterministic permutation seeding (e.g., `salt_aerospace_42`). It also documents the parallelization architecture (Joblib's `loky` backend) and the memory-safe garbage collection protocols required for Python 3.10+ execution. The complete replication dataset, including the open-science data pipeline and statistical modeling architecture, is available under a CC BY 4.0 license via the Zenodo repository at https://doi.org/10.5281/zenodo.19698274 [34]. Version 2 of the record adds the day-grain survival pipeline (Stages 2b–2d and 3b), the threshold-sweep and figure drivers, configuration ledger V1.1, the day-grain survival cohort, the per-aircraft block-sensitivity matrix (Table S3), and all figures, enabling end-to-end reproduction of every reported value. Independent re-execution reproduced the block-sensitivity aggregates and the validation concordance bit-identically.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author used Claude (Anthropic) to improve the language and readability of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the published article.

## Acknowledgements

The author would like to explicitly thank the OpenSky Network, its founders, and the global network of receiver operators. The crowd-sourced collection and maintenance of this terrestrial surveillance database made this analysis possible.

## CRediT authorship contribution statement

**Eugene Pik:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review and editing.

## Vitae

**Eugene Pik** is a Data Science PhD student at the University of Essex. He holds an M.Sc. in Aviation and Aerospace Sustainability (Aviation Cybersecurity) from Embry-Riddle Aeronautical University, with prior studies in Robotics and Software Engineering. An experienced entrepreneur, he has launched ventures spanning software, healthcare, and philanthropy. In the aerospace sector, Eugene previously led requirements for a hybrid-electric eVTOL aircraft for emergency medical services at Uniqopter. Currently, he is the founder of Mevocopter Aerospace, providing consultancy in aviation cybersecurity.

## References

[1] National Academies of Sciences, Engineering, and Medicine, A Review of the Next Generation Air Transportation System, National Academies Press, Washington, D.C., 2015. doi:10.17226/21721.

[2] T. Bolić, P. Ravenhill, SESAR: The past, present, and future of european air traffic management research, Engineering 7 (4) (2021) 448–51. doi:10.1016/j.eng.2020.08.023.

[3] P. Brooker, SESAR and NextGen: Investing in new paradigms, Journal of Navigation 61 (2) (2008) 195–208. doi:10.1017/S0373463307004596.

[4] E. Pik, J. S. D. Garcia, T. A. Smith, The invisible infrastructure: A global baseline for ADS-B kinematic integrity and the ghost record phenomenon, SSRN Preprint (April 2026). doi:10.2139/ssrn.6666578. URL https://dx.doi.org/10.2139/ssrn.6666578

[5] E. Pik, J. S. D. Garcia, T. A. Smith, GET-25: The global spatiotemporal ADS-B kinematic consistency benchmark, https://doi.org/10.5281/zenodo.18904146, [Dataset] (2026).

[6] M. Strohmeier, I. Martinovic, V. Lenders, Understanding realistic attacks on airborne collision avoidance systems, Journal of Transportation Security 15 (1-2) (2022) 87–118. doi:10.1007/s12198-021-00238-2.

[7] J. E. Holland, M. J. Kochenderfer, W. A. Olson, Optimizing the next generation collision avoidance system for safe, suitable, and acceptable operational performance, Air Traffic Control Quarterly 21 (3) (2013) 275–97. doi:10.2514/atcq.21.3.275.

[8] T. Westbrook, A taxonomy of radiofrequency jamming and spoofing strategies and criminal motives, Journal of Strategic Security 16 (2) (2023) 68–80. doi:10.5038/1944-0472.16.2.2081.

[9] M. Felux, P. Fol, B. Figuet, M. Waltert, X. Olive, Impacts of global navigation satellite system jamming on aviation, Navigation 71 (3) (2024). doi:10.33012/navi.657.

[10] E. Pik, M. Berra, J. Yearwood, J. S. D. Garcia, Detecting GPS anomalies in aviation using ADS-B: Correlating coordinate gaps and GPS deviations with NOTAM warnings, in: AIAA AVIATION FORUM AND ASCEND 2024, 2024, pp. AIAA 2024–4640. doi:10.2514/6.2024-4640.

[11] E. Pik, M. Berra, J. Yearwood, J. S. D. Garcia, GPS anomalies in aviation: Preliminary insights from automatic dependent surveillance–broadcast data, Journal of Aerospace Information Systems 22 (7) (2025) 570–82. doi:10.2514/1.I011527.

[12] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic, M. Wilhelm, Bringing up OpenSky: A large-scale ADS-B sensor network for research, in: Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (IPSN), 2014, pp. 83–94. doi:10.1109/IPSN.2014.6846743.

[13] M. Strohmeier, X. Olive, J. Lübbe, M. Schäfer, V. Lenders, Crowdsourced air traffic data from the OpenSky Network 2019–2020, Earth System Science Data 13 (2) (2021) 357–66. doi:10.5194/essd-13-357-2021.

[14] X. Ying, J. Mazer, G. Bernieri, M. Conti, L. Bushnell, R. Poovendran, Detecting ADS-B spoofing attacks using deep neural networks, in: 2019 IEEE Conference on Communications and Network Security (CNS), 2019, pp. 187–95. doi:10.1109/CNS.2019.8802732.

[15] L. Li, R. J. Hansman, R. Palacios, R. Welsch, Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring, Transportation Research Part C: Emerging Technologies 64 (2016) 45–7. doi:10.1016/j.trc.2016.01.007.

[16] M. Leonardi, G. Sirbu, ADS-B crowd-sensor network and two-step Kalman filter for GNSS and ADS-B cyber-attack detection, Sensors 21 (15) (2021) 4992. doi:10.3390/s21154992.

[17] N. Çevik, S. Akleylek, Comparison of machine learning based anomaly detection methods for ADS-B system, in: G. Mammadova, T. Aliev, K. Aida-zade (Eds.), Information Technologies and Their Applications. ITA 2024, Vol. 1092 of Lecture Notes in Networks and Systems, Springer, Cham, 2025, pp. 275–86. doi:10.1007/978-3-031-73420-5_23.

[18] A. Tabassum, W. Semke, UAT ADS-B data anomalies and the effect of flight parameters on dropout occurrences, Data 3 (2) (2018) 19. doi:10.3390/data3020019.

[19] X. Olive, J. Sun, B. Figuet, R. Alligier, Filtering techniques for ADS-B trajectory preprocessing, Journal of Open Aviation Science 2 (2) (2025). doi:10.59490/joas.2024.7882.

[20] E. Pik, J. S. D. Garcia, T. A. Smith, I. Kocaman, M. Berra, Predicting geopolitical instability through GNSS anomalies and air traffic data, Journal of Transportation Security 18 (1) (2025) 29. doi:10.1007/s12198-025-00323-w.

[21] E. L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, Journal of the American Statistical Association 53 (282) (1958) 457–81. doi:10.1080/01621459.1958.10501452.

[22] D. R. Cox, Regression models and life-tables, Journal of the Royal Statistical Society: Series B (Methodological) 34 (2) (1972) 187–20. doi:10.1111/j.2517-6161.1972.tb00899.x.

[23] T. Prevot, J. Rios, P. Kopardekar, I. Robinson, John E., M. Johnson, J. Jung, UAS traffic management (UTM) concept of operations to safely enable low altitude flight operations, in: 16th AIAA Aviation Technology, Integration, and Operations Conference, AIAA, Washington, D.C., 2016, pp. AIAA 2016–3292. doi:10.2514/6.2016-3292.

[24] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, Journal of the Royal Statistical Society: Series B (Methodological) 57 (1) (1995) 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

[25] W.-Y. Tsai, N. P. Jewell, M.-C. Wang, A note on the product-limit estimator under right censoring and left truncation, Biometrika 74 (4) (1987) 883–886. doi:10.1093/biomet/74.4.883.

[26] S. Suissa, Immortal time bias in pharmacoepidemiology, American Journal of Epidemiology 167 (4) (2008) 492–499. doi:10.1093/aje/kwm324.

[27] P. K. Andersen, R. D. Gill, Cox's regression model for counting processes: A large sample study, The Annals of Statistics 10 (4) (1982) 1100–1120. doi:10.1214/aos/1176345976.

[28] N. Breslow, Covariance analysis of censored survival data, Biometrics 30 (1) (1974) 89–99. doi:10.2307/2529620.

[29] P. M. Grambsch, T. M. Therneau, Proportional hazards tests and diagnostics based on weighted residuals, Biometrika 81 (3) (1994) 515–526. doi:10.1093/biomet/81.3.515.

[30] C. Davidson-Pilon, lifelines: Survival analysis in Python, Journal of Open Source Software 4 (40) (2019) 1317. doi:10.21105/joss.01317.

[31] E. Vittinghoff, D. V. Glidden, S. C. Shiboski, C. E. McCulloch, Regression Methods in Biostatistics: Linear, Logistic, Survival, and Mixed Models, 2nd Edition, Springer New York, 2012. doi:10.1007/978-1-4614-1353-0.

[32] Radio Technical Commission for Aeronautics (RTCA), Minimum operational performance standards for 1090 mhz extended squitter automatic dependent surveillance–broadcast (ads-b) and traffic information services–broadcast (tis-b), Tech. Rep. DO-260B, RTCA, Inc., Washington, DC, USA (Dec. 2009).
URL https://www.rtca.org/products/do-260b-hard-copy/

[33] Radio Technical Commission for Aeronautics (RTCA), Minimum operational performance standards for 1090 mhz extended squitter automatic dependent surveillance–broadcast (ads-b) and traffic information services–broadcast (tis-b), Tech. Rep. DO-260C, RTCA, Inc., Washington, DC, USA (Dec. 2020).
URL https://www.rtca.org/products/do-260c-electronic/

[34] E. Pik, Degradation of ads-b surveillance continuity: Temporal clustering and anomaly persistence, Dataset (Apr. 2026). doi:10.5281/zenodo.19698274.
URL https://doi.org/10.5281/zenodo.19698274