

Article

Validation Framework for Automated Aircraft Load Control Operations

Ivan Jakovljević ^{1,2,*}, Olja Čokorilo ¹  and Ljubiša Vasov ¹

¹ Department of Air Transport and Traffic, Faculty of Transport and Traffic Engineering, University of Belgrade, 11000 Belgrade, Serbia; lj.vasov@sf.bg.ac.rs (L.V.)

² Ink Innovation, 03001 Alicante, Spain

* Correspondence: ivan.jakovljevic@innovation.ink

Abstract

This paper presents a validation framework for automated aircraft mass and balance operations, combining statistical process control with integrated risk scoring and addressing the critical gap between traditional centralised load control and emerging automated operations. A dual-validation system integrates univariate control limits, a weighted multivariate T^2 -type statistic, and composite scoring with operational risk assessment; risk scoring quantifies centre of gravity proximity and maximum mass margins. The framework was validated using 48 flights operated by Airbus A350-900/1000 aircraft. The framework was evaluated using a two-phase design—flights 1–30 for calibration (Phase I) and flights 31–48 for prospective evaluation (Phase II). Across all 48 flights it yields an 81.25% automation-eligibility rate (39 GREEN), with 4.17% YELLOW (non-blocking supervisor notification within a defined service window) and 14.58% RED (mandatory review). RED cases result from cargo offloads, dangerous goods changes, or significant baggage/ZFW variations, while cargo operations show exceptional stability (the exclusively cargo-loaded Compartment 2 has $\sigma = 9$ kg, versus 219 kg for total cargo). The framework enables transition from centralised load controller roles to exception-based supervision, with risk scoring outputs driving targeted, proportional safety measures. To the best of the authors' knowledge, it provides the first systematic method for determining when automated load control can be trusted without human intervention—a critical requirement as the industry transitions towards full automation.

Keywords: aircraft; load control; statistical process control; multivariate statistical process control; risk scoring; automation



Academic Editor: Antonio Gil Bravo

Received: 6 July 2026

Revised: 23 July 2026

Accepted: 30 July 2026

Published: 3 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Aircraft mass and balance management represents one of the most safety-critical functions in aviation ground operations. Accurate calculation of aircraft mass and centre of gravity position is essential for flight safety, directly affecting aircraft performance, fuel efficiency, and structural integrity. Traditional load control operations rely on human oversight, where load controllers manually verify all calculations before flight release.

Effective safety management in aviation requires the systematic identification, assessment, and mitigation of operational risks [1].

The aviation industry is undergoing rapid digitalisation, with initiatives such as IATA's Digital Loading Verification (DLV) and the X565 Aircraft Semi-Permanent Data Exchange standard (specified via the IATA Airport Handling Manual, AHM 565/X565 [2])

transforming how ground operations data is captured and validated. These developments create opportunities for increased automation in load control operations. The present work concerns internal mass and balance—the distribution of payload, fuel and dangerous goods within the airframe; the distinct problem of external stores and aerodynamic external-load effects is treated elsewhere [3,4] and is outside the present scope.

While Yelken and Yilmaz [5] examined the success factors in ground handling automation, and Chatzi [6] developed safety management system frameworks, the specific application to load control process validation remains unaddressed.

However, to the best of the authors' knowledge and on the basis of the structured review above, no comprehensive framework for determining when automated load control decisions can be trusted without human intervention has been reported. This paper addresses this gap by developing a dual-validation approach combining statistical process control with operational risk assessment.

2. Literature Review

2.1. Mass and Balance Regulation and Academic Literature

EASA CAT.POL.MAB.105 mandates that operators ensure that if a digital tool is used for mass and balance calculations, the output of the tool is verified [7]. ICAO Annex 6 establishes international standards for operational procedures [8], while IATA's Airport Handling Manual (AHM) chapter 503 provides industry standards for load sheet preparation using electronic systems [2].

The academic literature on load control has focused primarily on loading optimisation rather than validation. Mongeau and Bès [9] developed mixed-integer linear programming models for container loading. Verstichel et al. [10] demonstrated improvements over manual planning, reducing average centre of gravity (CG) deviation from 4.75% to 0.04%. Zhao et al. [11] incorporated CG envelope constraints based on aircraft manufacturer specifications, solving practical instances within seconds.

Macalintal and Ubando [12] developed fuzzy linear programming enabling multi-objective optimisation across payload, priority, and cost. Halder [13] documents centralised load control system benefits including a 40% reduction in load planning time and a 78.57% reduction in misload incidents.

2.2. Statistical Process Control

Statistical process control methods have been applied in aviation maintenance [14]. The Hotelling T^2 statistic [15] provides multivariate quality control. Montgomery [16] establishes theoretical foundations, while Mason and Young [17] provide implementation guidance.

2.3. Research Gap Analysis

While substantial work exists in flight operations automation, air traffic management, and predictive maintenance, ground operations—particularly load control—remain largely unexplored from a statistical validation perspective.

Recent bibliometric analyses confirm this gap. Demir et al. [18] reviewed 224 articles on AI in aviation safety (2004–2024), finding research concentrated in air traffic management, pilot behaviour, and maintenance—with no studies addressing automated load control validation. Similarly, Lozano Tafur et al. [19] analysed 120 AI applications in air operations, highlighting trajectory prediction and performance optimisation while noting the absence of ground operations validation frameworks.

In the ground handling domain, Kovynyov and Mikut [20] surveyed digital transformation at airports, identifying automation advances in baggage handling, passenger

processing, and ramp operations—yet load control validation was not addressed, with the focus on cost optimisation rather than safety. Alonso Tabares and Mora-Camino [21] analysed aircraft ground handling automation potential, focusing on vehicle docking and autonomous ground support equipment rather than decision validation processes. Sunaryo and Basaria [22] conducted a systematic review of airport apron automation (2019–2023), confirming that while labour-intensive tasks show automation progress, analytical validation of automated decisions remains unexplored.

The mass and balance domain specifically shows this pattern. Chaves et al. [23] developed onboard weight estimation using flight mechanics and statistics for A320 aircraft—an in-flight estimation approach rather than pre-departure validation. Moshe et al. [24] applied neural networks for in-flight weight identification, again focusing on airborne systems rather than ground operations validation. Both approaches address post-departure estimation rather than pre-departure validation—a fundamentally different operational context where errors cannot be corrected after flight release.

This analysis reveals that while optimisation algorithms, operational efficiency metrics, and airborne estimation methods exist independently, no research integrates statistical process control with operational risk assessment for validating automated load control decisions prior to flight release. The proposed framework addresses this gap.

This gap is particularly significant given regulatory requirements. EASA CAT.POL. MAB.105(b) mandates that operators using computerised mass and balance systems verify the integrity of the output data [7], yet no academic framework exists for implementing such verification systematically. The goal of this research is to address this gap by developing the first dual-validation framework combining statistical process control with operational risk scoring for load control automation.

3. Methodology

This research implements a mixed-methods approach combining statistical data analysis with operational validation for framework development. The methodology is presented through five phases:

- Phase 1—Data acquisition: collection of operational data from loading instruction reports (LIRs) and final loadsheets, capturing planned versus actual values for 15 mass and balance parameters.
- Phase 2—Statistical baseline establishment: calculation of statistical distributions (μ , σ) for each parameter from historical data, enabling identification of operational stability patterns and control limit derivation.
- Phase 3—Framework development: design of the four-level hierarchical validation system integrating univariate control (Level 1), a weighted multivariate T^2 -type statistic (Level 2), composite scoring (Level 3), and operational risk assessment (Level 4).
- Phase 4—Threshold calibration: determination of decision boundaries using statistical confidence levels (95% and 99.73%) aligned with aviation safety standards and operational requirements.
- Phase 5—Validation: application of the framework to the complete dataset, with anomaly analysis to verify that flagged flights correspond to genuine operational irregularities.

The following subsections detail each component of the validation framework.

4. Validation Framework

The validation framework operates on the principle of variance analysis: comparing planned (projected) values against actual (final) values for all mass and balance parameters. The fundamental assumption is that under normal operations, variances between planned and actual values follow predictable statistical distributions. Anomalies—whether

from data entry errors, operational irregularities, or equipment malfunctions—manifest as statistical outliers.

The framework implements a four-level hierarchical validation system. Level 1 (univariate control) checks each parameter individually against control limits derived from historical variance distributions. Level 2 (multivariate control) applies a weighted multivariate T^2 -type analysis to detect anomalous patterns across correlated parameters that might not trigger individual alerts. Level 3 (composite scoring) synthesises all statistical measures into a single decision metric with clear operational thresholds. Level 4 (risk assessment) quantifies proximity to operational safety limits independent of statistical behaviour.

Each risk component scores from 0 (no risk) to 1 (maximum risk). The composite risk score combines all elements, with any single component exceeding defined thresholds sufficient to escalate the flight for human review.

4.1. Dual Validation Logic

Both statistical validation (Levels 1–3) and risk assessment (Level 4) must pass for automatic approval. A flight with normal statistical behaviour but high CG proximity risk requires review. Equally, a flight with comfortable safety margins but unusual variance patterns requires review. This dual-gate approach ensures that neither statistical anomalies nor operational risk exposure can bypass human oversight.

4.2. Data Collection

Data was collected from a single-destination flight using Airbus A350-900/1000 aircraft. The dataset comprises 48 flights (November–December 2025) with complete loading instruction reports as well as loadsheets data. Parameters monitored include forward hold delta (FWD), aft hold delta (AFT), total loading delta, compartment 1–4 deltas, bulk hold delta, baggage weight delta, cargo delta, dangerous goods delta, forward ZFM index delta, aft ZFM index delta, total ZFM index delta, and ULD mass discrepancies. This provides 15 monitored parameters for multivariate analysis; as detailed below, these are organised in a parent–child hierarchy and are not statistically independent.

For validation the 48 flights were split chronologically into a Phase I calibration set (flights 1–30, 5 November–8 December 2025), used to estimate baseline means, standard deviations and control limits, and a Phase II set (flights 31–48, $n = 18$, 9–26 December 2025) that was not used to estimate the limits. The limits were then applied to all 48 flights, so flights 31–48 constitute an out-of-sample check. This two-phase design follows standard statistical process control practice and ensures that the data used to develop the model differ from the data used to assess its performance. All 30 calibration flights are retained (no exclusions): the two flights with assignable causes are kept deliberately so that the cargo and dangerous goods parameters retain the non-zero variance required for standardisation. Both are still correctly flagged RED when scored.

The monitored parameters are organised in a parent–child hierarchy and are therefore not statistically independent: $FWD = C1 + C2$; $AFT = C3 + C4 + \text{bulk}$; $\text{total} = FWD + AFT = \text{baggage} + \text{cargo} + \text{dangerous goods}$; and $\text{total ZFM index} = \text{forward ZFM index} + \text{aft ZFM index}$, while ULD discrepancies are independent. Redundant aggregates are retained deliberately: aggregate parameters localise systemic shifts, component parameters identify their source, and the operational data show breaches occurring independently at either level. This exact linear dependence also makes the sample covariance matrix singular, which motivates the diagonal statistic adopted in Section 4.3.2.

4.3. Statistical Validation (Levels 1–3)

The statistical component detects anomalies through variance analysis. Parameters are weighted according to operational stability: cargo and dangerous goods parameters

receive the highest weight ($3\times$) due to exceptional stability, fuel and special loads receive moderate weight ($2\times$), while baggage parameters receive standard weight ($1\times$), reflecting inherent variability. This weighting ensures that stable parameters appropriately influence multivariate analysis without being masked by variable parameters.

Weights are assigned in two steps. Each parameter first receives a base weight from its stability category (Table 1): the most stable parameters (the all-cargo compartment, the stable cargo hold, and dangerous goods) receive weight 3, moderately stable parameters weight 2, and highly variable parameters weight 1. The base weight is then adjusted for mass-and-balance effect and safety criticality: the aft hold delta is raised to weight 2 despite its high variance because it constrains the binding aft centre of gravity limit. The resulting weights (Table 2) therefore reflect operational stability and safety criticality rather than arbitrary assignment, and sum to 30.

Table 1. Parameter weighting scheme based on operational stability analysis.

Stability Category	Variance	Weight (w_i)	Parameters
Very Stable	0.24–1%	3	Cargo, DG, Comp2, Comp3
Moderately Stable	5.72–14.24%	2	Cargo, FWD, TTL, Comp1, Bulk, Index parameters
Variable	19.7–94.47%	1	AFT, Comp4, Baggage, ULD

Table 2. Statistical baseline parameters.

Parameter (Delta)	Mean (kg)	Std Dev (kg)	Weight
Forward	−38	229	2
Aft	−655	327	2
TTL	−693	368	2
Compartment 1	−40	229	2
Compartment 2	2	9	3
Compartment 3	5	19	3
Compartment 4	−561	298	1
Bulk	−99	32	2
Baggage	−655	337	1
Cargo	−40	219	2
Dangerous Goods	2	9	3
ULD	−132	64	1
ZFMi FWD	0.1	0.72	2
ZFMi AFT	−1.12	1.11	2
ZFMi TTL	−1.02	1.36	2

4.3.1. Level 1: Univariate Control Limits

For each parameter i , the delta (variance) is calculated as the difference between actual and planned mass:

$$\Delta_i = X_{i(actual)} - X_{i(planned)} \quad (1)$$

Control limits are established using the standard 3-sigma method [25]:

$$UCL_i = \mu_i + 3\sigma_i \quad (2)$$

$$LCL_i = \mu_i - 3\sigma_i \quad (3)$$

where μ_i is the mean delta and σ_i is the standard deviation for parameter i , calculated from the historical dataset. The 3-sigma limits provide 99.73% confidence bounds, meaning only 0.27% of observations from a stable process would exceed these limits by chance alone.

Due to the parent–child relationship between the parameters, a single child-parameter univariate limit can be breached before a mandatory review is needed.

Formally, let the parent set $P = \{\text{FWD, AFT, total, total ZFM index}\}$ and the child set $C = \{\text{C1–C4, bulk, baggage, cargo, forward/aft ZFM index}\}$, and let B_P and B_C denote the number of parent- and child-level 3σ breaches. The univariate compliance indicator is $C_{uni} = 1$ when $B_P = 0$ and $B_C \leq 1$, and 0 otherwise (Equation (8)). A single isolated child-level breach with all parent aggregates in control reflects localised redistribution and is tolerated at Level 1, but it suppresses GREEN so that the status is capped at YELLOW. Dangerous goods is treated at parent level: any dangerous goods breach forces $C_{uni} = 0$. This rule reproduces every classification in Table 5, including the 19 December case (a single CPT3 UCL breach classified YELLOW).

Several baseline means differ substantially from zero (for example, the aft hold and baggage deltas average approximately -655 kg), reflecting systematic characteristics of the planning process (standard versus scale masses, and full allocation versus actual bookings) rather than individual errors. Following standard SPC practice, control limits are centred on the observed process mean, so the statistical gate detects departures from established behaviour rather than repeatedly flagging a known, stable offset. These persistent offsets are reported separately to load planning as a bias indicator, and a sustained mean shift triggers re-baselining. The implication of accepting such offsets as the baseline is noted in Section 6.3.

4.3.2. Level 2: Weighted Multivariate T^2 -Type Statistic

While univariate control is computationally simple, it fails to account for correlations between parameters. The Hotelling T^2 statistic addresses this limitation by considering the multivariate distribution of all parameters simultaneously.

Classical Hotelling's T^2 requires the inverse of the sample covariance matrix. Here, that matrix is singular by construction, because several monitored parameters are exact linear combinations of others (Section 4.2), and with only about 30 Phase I flights it could not be estimated reliably in any case. The statistic used is therefore a diagonal, variance-scaled T^2 -type statistic: it weights squared standardised deviations without inverting a covariance matrix and does not itself model cross-correlations, which are instead handled by the hierarchical univariate rule and reflected in the calibrated control limits. Because it is not a full Hotelling T^2 , it is referred to throughout as a weighted multivariate T^2 -type statistic; this diagonal construction follows high-dimensional practice for the small-sample, singular-covariance regime [26].

First, each delta is standardised to a Z-score:

$$Z_i = \frac{(\Delta_i - \mu_i)}{\sigma_i} \quad (4)$$

A key innovation of this framework is the introduction of parameter weighting based on operational stability. Parameters exhibiting high stability (low variance) receive higher weights, as deviations in these parameters are more operationally significant. The weighting scheme is shown in Table 1.

The weighted multivariate T^2 -type statistic is then calculated as:

$$T_{weighted}^2 = \sum_{i=1}^n (w_i \cdot Z_i^2) \quad (5)$$

The control limit for the weighted T^2 follows a scaled chi-square distribution. Using the Satterthwaite [27] approximation for weighted sums:

$$UCL_{T^2} = c \cdot \chi_{(df_{eff}, \alpha)}^2 \quad (6)$$

where c is a scaling constant and df_{eff} is the effective degrees of freedom. Because the weighted sum of squared standardised deviations is not exactly chi-square distributed, the control limit uses a Satterthwaite scaled-chi-square approximation. For the implemented weights (Table 2; $\Sigma w_i = 30$, $\Sigma w_i^2 = 66$), the scaling constant is $c = \Sigma w_i^2 / \Sigma w_i = 2.20$ and the effective (Satterthwaite) degrees of freedom are $\nu = (\Sigma w_i)^2 / \Sigma w_i^2 = 13.64$; the chi-square quantile is evaluated at the integer value 13, a slightly conservative choice that tightens the limit for an automated-release screen. Two confidence levels define the boundaries: 95% ($UCL_{T^2} = 49.20$) for the GREEN/YELLOW boundary and 99.73% ($UCL_{T^2} = 69.65$) for the YELLOW/RED boundary. Because the parent–child aggregates induce positive dependence, the nominal significance level is treated as approximate and the limits function as calibrated screening boundaries (Section 5.6).

4.3.3. Level 3: Composite Score

The composite statistical score C_{score} integrates univariate and multivariate assessments into a single metric:

$$C_{score} = C_{uni} \cdot C_{T^2} \quad (7)$$

where

$$C_{uni} = \{1 \text{ if } B_P = 0 \text{ and } B_C \leq 1; 0 \text{ otherwise}\} \quad (8)$$

$$C_{T^2} = 1 - \frac{T^2}{UCL_{T^2}} \quad (9)$$

The decision function S maps the composite score to operational status:

$$S_{(C)} = \{GREEN \text{ if } C > 0.294; YELLOW \text{ if } 0 < C \leq 0.294; RED \text{ if } C \leq 0\} \quad (10)$$

GREEN status indicates automatic approval with no human intervention required. YELLOW status triggers the load supervisor notification for review within a defined SLA (a non-mandatory review). RED status requires mandatory review before flight release.

The GREEN/YELLOW threshold on the composite score is not a free parameter: it equals $1 - UCL_{95}/UCL_{99.73}$ for the T^2 -type statistic (≈ 0.294 for these limits), so the composite boundary is fixed by the two control limits rather than tuned. Consistent with Equation (8), a single child-level breach caps the status at YELLOW even when the composite score would otherwise fall in the GREEN band, and any dangerous goods breach forces RED.

4.4. Level 4: Integrated Risk Scoring

Beyond statistical analysis, operational safety requires monitoring proximity to certified limits. The framework introduces integrated risk scoring combining safety compliance, business rules, and operational margins. The risk component quantifies proximity to operational limits through four independent scoring elements.

4.4.1. Compliance Indicators

R_{safe} evaluates compliance with aircraft structural limits, including maximum zero fuel mass, maximum landing mass, and CG envelope boundaries. Any parameter at or beyond certified limits triggers a mandatory review regardless of statistical behaviour.

$R_{business}$ assesses compliance with operator-defined business rules such as load factor targets (loading priorities) and operational preferences (positioning of bags close to the door, etc.) that may be more restrictive than regulatory limits.

$R_{safe} = 1$ if within MZFM, MTOM, MLM, and CG envelope; $R_{business} = 1$ if all business rules are satisfied. These are binary pass/fail criteria.

4.4.2. Risk Indicators

The risk assessment component evaluates proximity to operational limits independent of statistical behaviour. These indicators incorporate operational buffers and assign traffic light status (GREEN/YELLOW/RED) as well as quantitative risk scores normalised to a 0–1 scale. Two primary risk dimensions are assessed: centre of gravity proximity and maximum mass proximity.

4.4.3. Centre of Gravity Proximity Risk Score (R_{CG})

The CG proximity risk quantifies remaining operational margin after accounting for last-minute change (LMC) buffers. This metric ensures that sufficient tolerance exists for post-finalisation adjustments such as passenger movements, late baggage, or cargo changes.

The index and mean aerodynamic chord relationships in Equations (11)–(13) are the standard mass-and-balance computations of the operational loadsheet (trim sheet and balance chart), derived in accordance with the aircraft Weight and Balance Manual and the IATA Airport Handling Manual; their coefficients are read directly from the loadsheet rather than assigned by the authors.

For widebody aircraft, the LMC allowance is ± 1000 kg applied at the most extreme loading positions to calculate maximum index influence. This represents a more restrictive approach than standard LMC application, as it assumes worst-case positioning. The index-to-MAC conversion follows:

$$\Delta\%MAC = \frac{I_{kg} \cdot \Delta Mass}{C_{MAC} \cdot 100} \quad (11)$$

where I_{kg} is the moment index change per kilogram at the extreme position, $\Delta Mass$ is the LMC allowance (1000 kg), and C_{MAC} is the mean aerodynamic chord length. For the Airbus A350, this is calculated as:

$$LMCFWD = 2.22\% MAC \quad (12)$$

$$LMCAFT = 2.50\% MAC \quad (13)$$

The constants in Equations (11)–(14) are aircraft-type-specific mass and balance values taken from the carrier's Weight and Balance Manual (WBM) and Aircraft Handling Manual and are derived from aircraft weighing: I_{kg} is the index-per-kilogram unit that varies based on the loading position (compartment), C_{MAC} is the mean aerodynamic chord length [9.006 m], and $\Delta Mass$ is in kilograms, so $\Delta\%MAC$ is dimensionless (per cent MAC). The last-minute change values used are $LMCFWD = 2.22\% MAC$ and $LMCAFT = 2.50\% MAC$.

These buffers define a more restrictive operational envelope within the certified CG limits. Proximity to this restricted envelope determines status:

- GREEN: CG proximity to restricted limits $> 10\%$;
- YELLOW: CG proximity to restricted limits between 5% and 10% ;
- RED: CG proximity to restricted limits $\leq 5\%$.

The CG risk score is calculated as:

$$R_{CG} = \frac{0.5 - P_{CG}}{0.5} \quad (14)$$

where P_{CG} is the proximity of the actual CG to the nearest limit (forward or aft). The value 0.5 represents the midpoint between forward and aft limits—the position of maximum safety margin. A score of 0 indicates that the CG is centred with maximum available margin; a score approaching 1 indicates that the CG position is near the envelope boundary.

4.4.4. Maximum Mass Risk Score (R_{max})

The maximum mass score incorporates fuel variation statistics and LMC. From operational data, σ_{fuel} represents the standard deviation of final fuel change (planned vs. actual uplift). This fuel change is always positive, as the pilot in command can take more total fuel than issued in the operational flight plan (OFP). Risk thresholds:

$$GREEN : Underload \geq (\delta_{fuel} + LMC) \quad (15)$$

$$YELLOW : MAX(\delta_{fuel}, LMC) \leq Underload < (\delta_{fuel} + LMC) \quad (16)$$

$$RED : Underload < MAX(\delta_{fuel}, LMC) \quad (17)$$

The maximum mass risk score is 0 (safest value) when $\delta_{fuel} = 0$ and $LMC = 0$, i.e., when there are no variations between planned and actual fuel and LMC values, while if the variation equals the remaining underload, the risk score is 1 (the difference between planned and actual will exceed the remaining underload).

By construction, R_{max} is defined on the operating domain $\delta_{fuel} + LMC \leq \text{underload}$, over which it is bounded to $[0, 1]$; the boundary $R_{max} = 1$ corresponds to the combined variation equalling the remaining underload. A combined variation exceeding the underload denotes an overweight condition and a structural limit breach, for which $R_{safe} = 0$ and the status is RED. Across all 48 flights, $\delta_{fuel} + LMC \leq \text{underload}$ held, so $R_{max} \in [0, 1]$ throughout the dataset.

$$R_{max} = 1 - \left(\frac{Underload - \delta_{fuel} - LMC}{Underload} \right) \quad (18)$$

4.4.5. Overall Risk Score (R_{risk})

The overall risk score takes a conservative approach:

$$R_{risk} = \max(R_{CGscore}, R_{maxscore}) \quad (19)$$

This ensures that proximity to either limit type triggers attention.

4.4.6. Final Status Determination

Final status represents an integrated score between statistical status, risk status, and compliance status:

$$Final\ Status = (Statistical\ Status, Risk\ Status) \text{ when } R_{safe} = 1 \text{ and } R_{business} = 1 \quad (20)$$

If either compliance indicator fails, status is automatically RED. Likewise, if any of the buffers trigger RED or YELLOW status, the overall status becomes RED/YELLOW; otherwise, the status is GREEN.

A YELLOW classification generates a non-blocking notification to the load control supervisor; real-time review is discretionary, undertaken when workload permits and no RED case is pending. All YELLOW cases are, however, subject to a mandatory periodic (for example, weekly) retrospective audit intended to confirm ongoing model calibration, monitor the magnitude and direction of the observed discrepancies, and trigger re-baselining if

systematic drift is detected. This closes the monitoring loop: real-time screening supports dispatch while periodic review preserves calibration integrity.

The final status of Equation (20) is defined explicitly as RED whenever $R_{safe} = 0$ or $R_{business} = 0$, and otherwise the highest severity among the statistical and risk statuses (RED > YELLOW > GREEN). R_{safe} and $R_{business}$ are binary compliance flags (1 = compliant) that must both equal 1 for automated release; $R_{business}$ is set by the operator's commercial-compliance policy.

5. Results

5.1. Statistical Baseline

Analysis of 48 flights reveals distinct operational patterns. Table 2 presents the key parameters.

Cargo operations demonstrate exceptional stability (Compartment 2: $\sigma = 9$ kg, 0.24% variance). Baggage and aft compartments drive variability, accounting for 94.5% of total variance.

5.2. Framework Performance

With control limits estimated on the first 30 flights and applied to all 48, the framework attains an 81.25% automation-eligibility rate (39 GREEN; 95% Wilson CI 68.1–89.8%). Table 3 shows the distribution.

Table 3. Classification distribution.

Status	Count	Percentage	Action
GREEN	39	81.25%	Automatic approval
YELLOW	2	4.17%	Supervisor notification
RED	7	14.58%	Mandatory review

5.3. Anomaly Analysis

The 7 RED classifications correspond to genuine operational irregularities:

- Cargo offloads (three flights): CPT1/CPT3 LCL breaches, FWD LCL, cargo LCL violations;
- Dangerous goods changes (two flights): CPT2 UCL and DG UCL breaches;
- Large baggage/ZFW variations (two flights): significant planned-versus-actual baggage differences flagged by the multivariate T^2 -type statistic; on one flight, individual parameters remained within limits, while the other also showed univariate LCL breaches.

YELLOW classifications (two flights) show intermediate C_{score} values (0.18–0.28) with T^2 threshold exceedance. One flight (13 November) is flagged by the multivariate statistic alone; the other (19 December) has a single child-level UCL breach (CPT3) which, under the hierarchical rule, caps the status at YELLOW rather than promoting it to RED.

5.4. Risk Scoring Results

Table 4 presents the risk score ranges from the dataset.

R_{CG} consistently exceeds R_{max} , indicating that CG proximity is the binding constraint. The MAX function ensures that higher risk drives attention. All flights maintained $R_{safe} = 1$ and $R_{business} = 1$.

Table 4. Risk score summary.

Score	Range	Interpretation
R_{max}	0.107–0.281	Distance from max mass limits
R_{CG}	0.536–0.676	CG envelope margin
R_{risk} (overall)	0.536–0.676	$\max(R_{CG}, R_{max})$

5.5. Final Results

The final results for the 48 analysed flights are presented in Table 5.

Table 5. Complete validation results for 48 flights (November–December 2025).

Date	Aircraft	C _{score}	C _{status}	R _{max}	R _{CG}	R _{risk}	Risk Status	Overall	Reason
5 November 2025	A350-1000	0.75	GREEN	0.13	0.59	0.59	GREEN	GREEN	
6 November 2025	A350-1000	0.67	GREEN	0.16	0.59	0.59	GREEN	GREEN	
8 November 2025	A350-1000	0.64	GREEN	0.11	0.55	0.55	GREEN	GREEN	
9 November 2025	A350-1000	0.65	GREEN	0.24	0.64	0.64	GREEN	GREEN	
10 November 2025	A350-1000	0.95	GREEN	0.15	0.67	0.67	GREEN	GREEN	
11 November 2025	A350-1000	0.76	GREEN	0.28	0.60	0.60	GREEN	GREEN	
12 November 2025	A350-1000	0.73	GREEN	0.11	0.59	0.59	GREEN	GREEN	
13 November 2025	A350-1000	0.28	YELLOW	0.14	0.58	0.58	GREEN	YELLOW	Significant difference between planned and actual baggage (1351 kg): multivariate (T^2) score (CPT3, TTL, bag, bulk and ZFW delta close to UCL)
14 November 2025	A350-1000	0.98	GREEN	0.17	0.58	0.58	GREEN	GREEN	
17 November 2025	A350-900	0.83	GREEN	0.12	0.60	0.60	GREEN	GREEN	
17 November 2025	A350-900	0.00	RED	0.21	0.58	0.58	GREEN	RED	Change in dangerous goods loading; more DG than expected: CPT2 UCL, DGR UCL, T^2 due to >2 UCL breach
18 November 2025	A350-1000	0.86	GREEN	0.15	0.60	0.60	GREEN	GREEN	
19 November 2025	A350-900	0.85	GREEN	0.26	0.54	0.54	GREEN	GREEN	
20 November 2025	A350-1000	0.56	GREEN	0.11	0.54	0.54	GREEN	GREEN	
23 November 2025	A350-900	0.00	RED	0.14	0.57	0.57	GREEN	RED	Cargo offloaded from the forward hold: FWD LCL, CPT1 LCL, cargo LCL, FWD/TTL ZFW UCL
24 November 2025	A350-1000	0.93	GREEN	0.18	0.59	0.59	GREEN	GREEN	

Table 5. Cont.

Date	Aircraft	C _{score}	C _{status}	R _{max}	R _{CG}	R _{risk}	Risk Status	Overall	Reason
25 November 2025	A350-1000	0.92	GREEN	0.12	0.58	0.58	GREEN	GREEN	
26 November 2025	A350-900	0.81	GREEN	0.14	0.57	0.57	GREEN	GREEN	
27 November 2025	A350-900	0.75	GREEN	0.17	0.57	0.57	GREEN	GREEN	
28 November 2025	A350-1000	0.87	GREEN	0.12	0.60	0.60	GREEN	GREEN	
29 November 2025	A350-1000	0.74	GREEN	0.24	0.58	0.58	GREEN	GREEN	
30 November 2025	A350-900	0.81	GREEN	0.15	0.58	0.58	GREEN	GREEN	
1 December 2025	A350-900	0.60	GREEN	0.27	0.59	0.59	GREEN	GREEN	
2 December 2025	A350-1000	0.89	GREEN	0.11	0.60	0.60	GREEN	GREEN	
3 December 2025	A350-1000	0.76	GREEN	0.13	0.56	0.56	GREEN	GREEN	
4 December 2025	A350-900	0.81	GREEN	0.18	0.58	0.58	GREEN	GREEN	
5 December 2025	A350-900	0.81	GREEN	0.12	0.60	0.60	GREEN	GREEN	
6 December 2025	A350-1000	0.37	GREEN	0.21	0.58	0.58	GREEN	GREEN	
7 December 2025	A350-1000	0.95	GREEN	0.15	0.57	0.57	GREEN	GREEN	
8 December 2025	A350-900	0.94	GREEN	0.26	0.59	0.59	GREEN	GREEN	
9 December 2025	A350-900	0.64	GREEN	0.11	0.56	0.56	GREEN	GREEN	
10 December 2025	A350-1000	0.85	GREEN	0.16	0.68	0.68	GREEN	GREEN	
11 December 2025	A350-1000	0.91	GREEN	0.12	0.59	0.59	GREEN	GREEN	
12 December 2025	A350-900	0.00	RED	0.23	0.59	0.59	GREEN	RED	Offloading of the dangerous goods: CPT2 UCL and DG UCL breached
13 December 2025	A350-900	0.00	RED	0.15	0.59	0.59	GREEN	RED	Offloading of cargo pallet: CPT1 LCL, FWD LCL, cargo LCL, FWD ZFW UCL
14 December 2025	A350-1000	0.00	RED	0.14	0.59	0.59	GREEN	RED	Offloading of cargo pallet: CPT1/CPT3 LCL, FWD LCL, cargo LCL, FWD/TTL ZFW UCL
15 December 2025	A350-1000	0.85	GREEN	0.26	0.57	0.57	GREEN	GREEN	
16 December 2025	A350-900	0.92	GREEN	0.11	0.60	0.60	GREEN	GREEN	
17 December 2025	A350-900	0.89	GREEN	0.13	0.56	0.56	GREEN	GREEN	
18 December 2025	A350-1000	0.92	GREEN	0.17	0.58	0.58	GREEN	GREEN	

Table 5. *Cont.*

Date	Aircraft	C _{score}	C _{status}	R _{max}	R _{CG}	R _{risk}	Risk Status	Overall	Reason
19 December 2025	A350-1000	0.18	YELLOW	0.12	0.59	0.59	GREEN	YELLOW	Change in loading in CPT3: UCL, YELLOW T ² score
20 December 2025	A350-900	0.90	GREEN	0.13	0.57	0.57	GREEN	GREEN	
21 December 2025	A350-900	0.88	GREEN	0.15	0.56	0.56	GREEN	GREEN	
22 December 2025	A350-1000	0.00	RED	0.12	0.57	0.57	GREEN	RED	Significant change between planned and actual baggage count (1685 kg): LCL aft, LCL aft ZFW delta, LCL bag weight
23 December 2025	A350-1000	0.00	RED	0.25	0.55	0.55	GREEN	RED	Significant change between planned and actual baggage count (1685 kg): multivariate T ² breach, no UCL/LCL breach
24 December 2025	A350-900	0.84	GREEN	0.15	0.58	0.58	GREEN	GREEN	
25 December 2025	A350-900	0.74	GREEN	0.28	0.58	0.58	GREEN	GREEN	
26 December 2025	A350-1000	0.85	GREEN	0.11	0.56	0.56	GREEN	GREEN	

5.6. Assumption Verification

The 3σ and scaled-chi-square limits assume approximate normality and process stability, verified as follows on the 30-flight baseline. The continuously varying aggregate parameters are consistent with normality (Shapiro–Wilk p : aft 0.17, total 0.24, baggage 0.17, ULD 0.58), whereas parameters that are zero under normal operations (forward, cargo, dangerous goods) act as event indicators rather than continuous variables. Lag-1 autocorrelations are weak for all parameters ($|r| < 0.16$), consistent with independence. The A350-900 ($n = 20$) and A350-1000 ($n = 28$) subsets do not differ significantly in mean (Welch p : aft 0.55, total 0.80, baggage 0.40) or variance (Levene p : aft 0.22, total 0.30, Baggage 0.16), so they are pooled into a single baseline. Two robustness checks confirm that the classifications are not an artefact of the settings: perturbing any weight by ± 1 category shifts the composite GREEN/YELLOW threshold by at most 0.002 and reclassifies none of the 48 flights; and because the control limit is fixed by the weights, it is unchanged (49.20 and 69.65) regardless of whether the baseline statistics are computed on the first 30 flights or on all 48, with the 30-flight window preceding the December events being retained so that those events do not inflate the limits and mask the irregularities the framework is designed to detect. Because the parent–child aggregates induce positive dependence, the nominal significance level is treated as approximate and the limits act as calibrated screening boundaries; validation on larger multi-route data with the full covariance structure is identified as an avenue for future work.

5.7. Ground-Truth Concordance

All 48 flights were dispatched and operated by licenced load controllers and completed without any loading-related incident, so an operational ground truth exists. The reported 81.25% therefore denotes an automation-eligibility rate—the proportion of flights the framework would release automatically—rather than an end-to-end accuracy or safety

metric. Retrospective concordance analysis shows that every GREEN flight matched the accepted final loadsheet with no post-dispatch load discrepancy (no false negatives), while each RED flight corresponds to a verified assignable event (cargo offload, dangerous goods change, or a significant baggage/ZFW difference). Independent assessment by load controllers blinded to the framework output is required to establish accuracy and false negative rates formally and is identified as an avenue for future work.

Although the control limits were estimated on the first 30 flights and applied to all 48, the 18 flights not used for calibration (9–26 December 2025) form an out-of-sample set. On these, the framework released 12 of 18 automatically (66.7%; 95% Wilson interval 43.7–83.7%), with one YELLOW and five RED. Every out-of-sample RED corresponds to a verified assignable event—two cargo-pallet offloads (13 and 14 December), one dangerous-goods change (12 December), and two large baggage/ZFW differences (22 and 23 December). The lower out-of-sample release rate therefore reflects a genuine December cluster of loading irregularities that the framework correctly escalated, not a degradation in performance; contextually, the model produced correct classifications throughout, within the single-route limitations noted in Section 6.3.

6. Discussion

6.1. Contributions

This research establishes the first framework, to the best of the authors' knowledge, combining statistical process control with integrated risk scoring for load control validation. The weighted multivariate T^2 -type approach incorporates operational domain knowledge. The position-specific LMC index calculations combined with fuel-variation-adjusted mass monitoring provide operationally meaningful thresholds.

6.2. Practical Implications

The 81.25% automation rate enables transition from centralised load controller roles to exception-based supervision. RED cases represent genuine operational events (cargo offloads, DG changes, significant variations) that warrant human attention. The risk assessment results enable targeted safety measures: flights with elevated R_{CG} values indicate the need for enhanced last-minute change protocols, while high R_{max} alerts warrant fuel uplift verification procedures. The traffic light classification (GREEN/YELLOW/RED) provides clear escalation paths—GREEN flights proceed autonomously; YELLOW flights trigger proactive supervisor notification, allowing for a non-blocking review; and RED flights enforce mandatory human verification before release. This graduated response ensures that safety interventions are proportional to identified risk levels rather than applied uniformly across all operations. The framework aligns with IATA's Digital Loading Verification and X565 standards, with the ability to apply autonomous ramp loading management until load control supervisor approval/review is needed, increasing automation in a safe and auditable way.

6.3. Limitations

Validation is limited to a single route and operator. The 48-flight sample provides baseline estimation but limits tail probability confidence. Future research should expand across diverse operations.

6.4. Practical Implementation Considerations

Operational deployment requires a route- and aircraft-type-specific calibration phase. Consistent with standard Phase I practice, at least 25–30 flights per route and aircraft family should be collected before control limits are fixed. Control limits are inherently route-

and season-specific: schedule changes, seasonal demand shifts, or a change in aircraft type should trigger re-baselining, monitored through mean-shift detection or a periodic rolling-window update. Aircraft variants may be pooled only after a homogeneity check (Section 5.6); if means or variances differ significantly, each variant requires its own baseline. A data-governance process—recording all fifteen parameters, maintaining assignable-cause investigations, and periodically re-validating baselines—is essential for safe operation. These requirements bound the transferability of the present single-route results and frame the multi-route validation identified as future work.

7. Conclusions

This paper presents a dual-validation framework for automated mass and balance operations, combining statistical process control with integrated risk scoring. The approach achieves 81.25% full automation while ensuring that RED classifications capture genuine operational irregularities requiring human review.

Key contributions include weighted multivariate T^2 -type analysis incorporating operational stability, position-specific LMC buffer calculations for CG proximity, fuel-variation-adjusted maximum mass monitoring, and conservative risk aggregation using $\max(R_{CG}, R_{max})$. The framework supports the industry transition toward automated ground operations.

Author Contributions: Conceptualisation, I.J. and O.Č.; methodology, I.J.; software, I.J.; validation, I.J., O.Č. and L.V.; formal analysis, I.J.; investigation, I.J.; resources, I.J.; data curation, I.J.; writing—original draft preparation, I.J.; writing—review and editing, O.Č. and L.V.; visualisation, I.J.; supervision, O.Č. and L.V.; project administration, O.Č. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to commercial confidentiality.

Conflicts of Interest: Author Ivan Jakovljević is employed by Ink Innovation. The remaining authors declare no conflicts of interest. The company had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations and symbols are used in this manuscript:

AFT	Aft cargo hold
AHM	Airport handling manual
CG	Centre of gravity
CLC	Centralised load control
CPT	Compartment
DG	Dangerous goods
DLV	Digital loading verification
EASA	European Union Aviation Safety Agency
FWD	Forward cargo hold
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
LCL	Lower control limit

LMC	Last minute change
MAC	Mean aerodynamic chord
MLM	Maximum landing mass
MTOM	Maximum take-off mass
MZFM	Maximum zero fuel mass
OFP	Operational flight plan
SLA	Service level agreement
SMS	Safety management system
SPC	Statistical process control
TTL	Total
ULD	Unit load device
X565	Aircraft semi-permanent data exchange standard
ZFM	Zero fuel mass
ZFMi	Zero fuel mass index
Δ_i	Delta (variance) for parameter i
μ_i	Mean delta for parameter i
σ_i	Standard deviation for parameter i
Z_i	Standardised Z-score for parameter i
w_i	Weight assigned to parameter i
T^2	weighted multivariate T^2 -type statistic
Bp	Parent univariate breach
Bc	Child univariate breach
C_{score}	Composite statistical score
Cuni	Univariate compliance indicator
CT^2	Multivariate compliance score
CMAC	Mean aerodynamic chord length
I_{kg}	Moment index change per kilogram
LMCFWD	LMC mass MAC influence at the most extreme forward position
LMCAFT	LMC mass MAC influence at the most extreme aft position
S(C)	Decision function mapping composite score to status
R_{safe}	Safety compliance indicator
$R_{business}$	Business rules compliance indicator
R_{CG}	Centre of gravity proximity risk score
R_{max}	Maximum mass proximity risk score
R_{risk}	Overall risk score
P_{CG}	Proximity of actual CG to nearest limit
δ_{fuel}	Standard deviation of fuel variation
UCL	Upper control limit
df_{eff}	Effective degrees of freedom
χ^2	Chi-square distribution
α	Significance level
c	Scaling constant
%MAC	Percentage of mean aerodynamic chord

References

1. ICAO. *Safety Management Manual (Doc 9859)*, 4th ed.; International Civil Aviation Organization: Montreal, QC, Canada, 2018.
2. IATA. *Airport Handling Manual*, 44th ed.; International Air Transport Association: Montreal, QC, Canada, 2024.
3. Zhu, R.; Zhang, S.; Zhu, Q.; Liu, H.; Wang, X.; Fei, Q. Improved Regularization-Based Subspace Method for Harmonic Load Identification. *AIAA J.* **2026**, 1–12. [[CrossRef](#)]
4. Liu, Y.; Ng, B.F. Two-Step Correction and Decoupled Sequential Framework for Uncertainty-Oriented Dynamic Force Reconstruction. *AIAA J.* **2025**, 63, 5550–5569. [[CrossRef](#)]
5. Yelken, E.; Kucuk Yilmaz, A. Lean transformation in ground handling: Success factors with cognitive mapping. *Aircr. Eng. Aerosp. Technol.* **2025**, 97, 1354–1364. [[CrossRef](#)]

6. Chatzi, A.V. Safety management systems: An opportunity and a challenge for military aviation organisations. *Aircr. Eng. Aerosp. Technol.* **2019**, *91*, 190–196.
7. EASA. *Commission Regulation (EU) No 965/2012—Annex IV (Part-CAT)*; European Union Aviation Safety Agency: Cologne, Germany, 2012.
8. ICAO. *Annex 6 to the Convention on International Civil Aviation—Operation of Aircraft*, 11th ed.; International Civil Aviation Organization: Montreal, QC, Canada, 2022.
9. Mongeau, M.; Bès, C. Optimization of aircraft container loading. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 140–150. [[CrossRef](#)]
10. Verstichel, J.; Vancroonenburg, W.; Souffriau, W.; Vanden Berghe, G. Automatic air cargo selection and weight balancing: A mixed integer programming approach. *Transp. Res. Part E Logist. Transp. Rev.* **2014**, *65*, 70–83. [[CrossRef](#)]
11. Zhao, X.; Yuan, Y.; Dong, Y.; Zhao, R. Optimization approach to the aircraft weight and balance problem with the centre of gravity envelope constraints. *IET Intell. Transp. Syst.* **2021**, *15*, 1396–1411. [[CrossRef](#)]
12. Macalintal, J.M.V.; Ubando, A.T. Optimal aircraft payload weight and balance using fuzzy linear programming model. *Chem. Eng. Trans.* **2023**, *103*, 613–618.
13. Haldar, P. Centralized load control in aviation. *Int. J. Res. Publ. Rev.* **2025**, *6*, 8690–8694. [[CrossRef](#)]
14. Krishna, G.V.; Thakur, K.; Sundaram, A. Trends in aviation maintenance, safety, and regulation: A probabilistic analysis of global research (2015–2024). *Aircr. Eng. Aerosp. Technol.* **2026**, *in press*. [[CrossRef](#)]
15. Hotelling, H. Multivariate quality control. In *Techniques of Statistical Analysis*; Eisenhart, C., Hastay, M.W., Wallis, W.A., Eds.; McGraw-Hill: New York, NY, USA, 1947; pp. 111–184.
16. Montgomery, D.C. *Statistical Quality Control: A Modern Introduction*, 6th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009.
17. Mason, R.L.; Young, J.C. *Multivariate Statistical Process Control with Industrial Applications*; SIAM: Philadelphia, PA, USA, 2002.
18. Demir, G.; Moslem, S.; Duleba, S. Artificial intelligence in aviation safety: Systematic review and bibliometric analysis. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 279. [[CrossRef](#)]
19. Lozano Tafur, C.; Camero, R.G.; Aldana Rodríguez, D.; Daza Rincón, J.C.; Rativa Saenz, E. Applications of artificial intelligence in air operations: A systematic review. *Results Eng.* **2025**, *25*, 103742. [[CrossRef](#)]
20. Kovynyov, I.; Mikut, R. Digital transformation in airport ground operations. *arXiv* **2018**, arXiv:1805.09142.
21. Alonso Tabares, D.; Mora-Camino, F. Aircraft ground operations: Steps towards automation. *CEAS Aeronaut. J.* **2019**, *10*, 965–974. [[CrossRef](#)]
22. Sunaryo, E.R.; Basaria, F.T. Trends of automation in airport apron area: A systematic literature review. In *Proceedings of the 6th International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 20–21 November 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 364–369.
23. Chaves, F.A.V.; Silvestre, M.A.R.; Gamboa, P.V. Preliminary development of an onboard weight and balance estimator for commercial aircraft. *Aerosp. Sci. Technol.* **2018**, *72*, 316–326. [[CrossRef](#)]
24. Moshe, I.; Iosilevskii, G.; Nazarov, S. In-flight weight and balance identification using neural networks. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit*, Austin, TX, USA, 11–14 August 2003.
25. Shewhart, W.A. *Economic Control of Quality of Manufactured Product*; 50th Anniversary Commemorative Reissue; ASQ Quality Press: Milwaukee, WI, USA, 1980.
26. Srivastava, M.S.; Du, M. A test for the mean vector with fewer observations than the dimension. *J. Multivar. Anal.* **2009**, *100*, 518–532. [[CrossRef](#)]
27. Satterthwaite, F.E. An approximate distribution of estimates of variance components. *Biometr. Bull.* **1946**, *2*, 110–114. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.