Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

One standard, many airports: Evaluating a uniform ground-time planning parameter for KC-390 airlift operations

Ramiro de Paiva Fagundes
Instituto Tecnológico de Aeronáutica
Lucas Orbolato Carvalho
orbolatoloc@fab.mil.br

Instituto Tecnológico de Aeronáutica
Daniel Alberto Pamplona
Instituto Tecnológico de Aeronáutica
Níssia Carvalho Rosa Bergiante
Instituto Tecnológico de Aeronáutica

Research Article

Keywords: Turnaround time, Ground time, Military airlift, Schedule buffer, Service level, Aerodrome heterogeneity

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11235475/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

One standard, many airports: Evaluating a uniform ground-time planning parameter for KC-390 airlift operations

Ramiro de Paiva Fagundes[1], Lucas Orbolato Carvalho[1*], Daniel Alberto Pamplona[1], Níssia Carvalho Rosa Bergiante[1]

[1]Aeronautics Institute of Technology (ITA), São José dos Campos, SP, Brazil.

*Corresponding author(s). E-mail(s): orbolatoloc@fab.mil.br;

Abstract

Ground-time allowances link aircraft ground operations to mission feasibility and crew-duty margins, but uniform values can conceal differences between airports. This study evaluates the two-hour allowance used to plan Brazilian Air Force KC-390 airlift missions against 1,679 ground events reconstructed from 2023–2024 mission reports. The standard covers 57% of stops, implying an overrun-to-unused-allowance cost ratio of about 1.3. A two-regime mixture separates routine service from extended ground events and supports a conservative 4.5-h trimming threshold; recorded durations show no anchoring to the planning value. Median and one-sided trimmed-mean models identify three of 14 aerodromes where the standard is systematically too short and only 25–35% of stops meet it. These findings survive Holm correction and empirical Bayes partial pooling. Two further aerodromes warrant monitoring with explicit sample-size rules. Extended-event shares range from 1% to 12%, distinguishing routine shortfalls from tail risks. In mission-level simulations, localized allowances of 2.39–2.47 h reduce the probability of a cumulative overrun above 30 minutes from 66% to 38% on a route through two affected aerodromes. A tiered adjust–monitor–retain policy translates these findings into mission-planning inputs, using an aggregate-data workflow compatible with restricted operational records.

Keywords: Turnaround time, Ground time, Military airlift, Schedule buffer, Service level, Aerodrome heterogeneity

Acknowledgments. The authors thank the Brazilian Air Force for authorizing access to the mission-report data used in this study. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Brazilian Air Force or the Brazilian Ministry of Defense.

1 Introduction

Planning parameters translate uncertain operational processes into the fixed numbers that schedulers, crew planners and commanders use to decide whether a sequence of flights is feasible. In commercial aviation the most studied of these parameters is the buffer embedded in scheduled block and turnaround times, whose size trades aircraft utilization against punctuality and the propagation of delay through the network [1–3]. Much less is known about how such parameters perform when they are fixed by regulation, applied uniformly across heterogeneous airports and used to plan non-scheduled operations. Military and other state air transport operators fall squarely into this category. Their planning manuals prescribe standard ground times by aircraft type [e.g. 4], and missions are then built, crewed and dispatched on that basis. Whether these standards reflect what happens on the ground is seldom assessed: in the United States, for example, the Government Accountability Office found that the Air Mobility Command did not systematically collect the operational data needed to judge its planning factors against observed performance [5].

Ground-time assumptions also connect local aircraft handling to system-level performance. Recent aeronautical research has examined how ground-support equipment affects delays across locations [6] and how exhausted rotation margins transmit delays to subsequent flights [7]. Validating the ground-time input is therefore a prerequisite for credible mission timelines and operational reliability assessments.

This article examines one such standard. The Brazilian Air Force plans KC-390 missions with a fixed allowance of two hours between landing and the next takeoff [8]. The KC-390 is a multimission airlifter employed in logistic transport, airdrop, aerial refueling, firefighting and humanitarian support [9], and it operates from a wide range of aerodromes, from major air bases to remote airfields with limited ground-handling resources. The allowance matters because it enters every mission timeline. An optimistic value consumes the crew's duty margin and can force an unplanned overnight; a pessimistic value leads planners to assign augmented or relief crews, whose off-duty members occupy seats that would otherwise carry passengers and who generate additional per diem costs, and it leaves a scarce aircraft idle. A uniform value is administratively simple, but it assumes that ground processes are homogeneous across aerodromes—an assumption that the turnaround literature gives little reason to accept [10, 11].

Using two years (2023–2024) of mission reports, we address three research questions. *RQ1*: How reliable is the uniform two-hour standard at fleet level, and what service level does it implicitly target? *RQ2*: Are there aerodromes at which the standard is systematically too short, and do such findings survive alternative treatments of extreme observations, control of multiplicity across aerodromes and partial pooling of the aerodrome estimates? *RQ3*: Which planning values and decision rules follow from the evidence, and what are their consequences for multi-stop missions?

Two constraints shape the analysis, and both are common in state aviation and increasingly in commercial settings. First, the mission reports record landing and takeoff times but not the on-block and off-block milestones that define turnaround time in the airline literature, so ground time must be reconstructed from consecutive legs. Second, the leg-level records are restricted and cannot be released. We address

the first constraint by constructing a landing-to-departure measure that matches the regulatory definition of the planning parameter. We address the second with a two-layer design: estimation that requires leg-level records was carried out within the data owner's environment, while a set of extended analyses—reliability profiles, regime separation, simultaneous inference, partial pooling and mission-level simulation—relies only on aggregate statistics reported in this article, so that readers can replicate them without access to the restricted records.

The article makes four contributions. Empirically, it provides, to our knowledge, the first systematic evaluation of a military airlift ground-time standard against operational records, covering 1,679 KC-390 ground events. Conceptually, it reframes a regulatory ground-time allowance as an implicit service level—the probability that a stop is completed within the allowance—which links ground-time standards to the newsvendor logic used to set scheduled block times [3, 12] and exposes the cost trade-off embedded in the current rule. Methodologically, it offers a transparent and disclosure-compatible workflow that separates routine ground service from extended ground events, combines two robust estimators with family-wise error control and empirical Bayes partial pooling, and converts inconclusive evidence into explicit monitoring requirements. Managerially, it proposes a tiered policy—retain the fleet-wide default, adjust it where the evidence is robust and monitor where it is only suggestive—and quantifies the effect of that policy on overrun risk and unused aircraft time in representative mission sequences.

The remainder of the paper is organized as follows. Section 2 reviews the literature on turnaround operations, schedule buffers, military airlift planning and the statistical treatment of skewed operational durations. Section 3 describes the operational context and the data. Section 4 presents the methods, Section 5 the results and Section 6 the implications and limitations. Section 7 concludes.

2 Literature review

2.1 Aircraft turnaround operations and airport heterogeneity

The turnaround comprises the ground processes performed between an aircraft's arrival at the stand (on-block) and its departure (off-block): deboarding and unloading, refueling, cleaning, catering, technical checks, loading and boarding. These processes run partly in parallel and partly in sequence, so turnaround duration is governed by a critical path whose composition depends on the aircraft type, the operator's procedures and the resources available at the airport [11, 13, 14]. Early work showed that turnaround efficiency is a lever on aircraft operating costs and that its variability propagates through the day's rotation [15, 16]. Subsequent studies modeled individual process times stochastically—boarding being the most intensively studied [17]—and found that process-time distributions and delay patterns are specific to airports and operators [10, 18]. Monitoring frameworks were proposed to identify the sub-processes that drive late departures [19], and recent work exploits milestone data from airport collaborative decision-making (A-CDM) systems: Okwir et al. [20] examine how collaborative decision-making shapes turnaround performance, Rott et al. [21] apply process mining to turnaround event logs at Munich Airport, and Tang et al. [22, 23]

predict turnaround milestones with cascaded machine-learning models that account for the remaining buffer time.

A second strand examines turnarounds under resource constraints. Simulation shows that turnaround duration responds non-linearly to shortages of ground-handling staff, so that below a threshold the additional time can no longer be absorbed by schedule buffers [24], and that infrastructure and process design determine feasible turnaround times at a new airport [25]. Optimization models allocate ground-handling vehicles and staff [26, 27], and the effects of automation and post-pandemic hygiene requirements on ground time have also been assessed [28, 29]. Taxi times, which are part of the landing-to-takeoff interval analyzed here, likewise vary systematically across airports and operating conditions [30].

Recent work in *CEAS Aeronautical Journal* strengthens the connection between ground processes, local resources and operational prediction. Tucci and Facchini [6] analyze ground-support equipment and flight delays across locations to support investment decisions. Murad and Ruocco [31] evaluate synthetic operational data for pre-tactical flight-delay and turnaround forecasting, showing both the usefulness of synthetic records and the limits of forecasts based only on scheduled information. These studies concern equipment decisions and flight-specific forecasting, respectively; the present study evaluates a fixed planning allowance against observed durations and derives aerodrome-level adjustment rules.

In complementary work, Gök et al. [32] combine simulation and optimization to coordinate turnaround activities and ground-service teams under uncertainty while limiting the exchange of sensitive information. Their approach optimizes the execution of apron operations; the present analysis estimates and validates the duration allowances that enter mission plans when detailed process-level data are unavailable.

Two features of this literature matter for the present study. First, the evidence comes almost entirely from scheduled commercial operations at airports that record milestone-level data. Second, it consistently shows that turnaround performance is airport-specific. The implications of that heterogeneity have not been examined for operators that apply a single regulatory ground-time allowance across heterogeneous, and often austere, aerodromes and whose records contain only landing and takeoff times.

2.2 Schedule buffers, delay propagation and reliability-based planning

Because ground and flight processes are stochastic, schedules embed buffers. Wu [1] distinguished delays that are inherent in the schedule design from those caused by operational disruptions and showed that the allocation of buffer time determines a schedule's operational reliability. Buffers are costly: slack improves punctuality but lowers aircraft utilization, and its value depends on where in the network it is placed [2, 33–35]. Empirical studies of schedule padding confirm that airlines set buffers strategically and that padding affects punctuality and consumer choice [36–39], and robust scheduling has been associated with measurable reductions in delay [40].

A central insight of this literature is that a planned duration can be read as a quantile of the distribution of actual durations. Deshpande and Arıkan [12] model

the choice of scheduled block time as a newsvendor problem in which the airline balances the costs of under- and over-allocation, so that the optimal schedule sets the probability of on-time completion equal to a critical fractile; Hao and Hansen [3] similarly relate scheduled block times to block-time reliability. The newsvendor model has a long history in operations research [41, 42] and offers a natural language for regulatory planning factors: a fixed allowance implicitly selects a service level and, with it, an implicit ratio between the cost of a shortfall and the cost of idle time. We are not aware of studies that apply this logic to ground-time standards in state or military aviation.

The consequences of insufficient buffers are well documented. Delay propagates through aircraft rotations, crew connections and passenger connections [43–48], and recovery requires the coordinated re-planning of turnarounds, aircraft and crews [49–51]. Crew schedules built on deterministic durations perform worse in operation than their planned cost suggests [52], which motivates robust crewing strategies such as move-up crews [53]. In military airlift the crew-duty constraint is often binding, and because duty limits exist to manage fatigue, the consumption of duty time by unplanned ground delay is a safety concern as well as an efficiency one [54].

Within aeronautical network modeling, Berling et al. [7] jointly address primary and reactionary delay in a dynamic simulation of the European air transport system. Their treatment of delay propagation through aircraft rotations reinforces the need to evaluate ground-time assumptions at both the stop and mission levels. Our mission-level analysis isolates the consequences of changing ground-time allowances under the stated site-model and independence assumptions; it does not represent a full network recovery model.

Finally, statistical studies emphasize that delay distributions are asymmetric with heavy right tails [55, 56] and that the factors driving typical outcomes differ from those driving extreme ones [57]; data-driven delay research more broadly is reviewed by Carvalho et al. [58]. For planning parameters this implies that central tendency and tail behavior should be analyzed separately, a principle we adopt by distinguishing routine ground service from extended ground events.

2.3 Military airlift planning and ground operations

Operations research on military airlift has long treated ground times as inputs to large-scale models. Optimization models of airlift mobility allocate aircraft, crews and cargo across airlift networks subject to airfield and ground-time constraints [59–61]; simulation-based optimization captures the stochastic nature of the airlift problem [62]; and intratheater models support routing in operational settings [63]. At the airfield level, queuing models of ground processes—from fork-join networks for sortie generation to mean-value analysis of airlift operations at an individual airfield—show that ground resources and parking constraints can limit throughput [64–66], a conclusion also reached in work on airfield capacity for airlift operations [67]. Planning manuals operationalize these ideas through standard ground times by aircraft type [4].

What these contributions share is that the ground-time value itself is rarely estimated from operational records, and the concern raised by U.S. Government Accountability Office [5] about planning factors that are not validated against observed

operations applies equally to ground times. Evidence from humanitarian operations reinforces the point: airports become bottlenecks when relief demand surges, and aircraft waiting and ground times depend strongly on airport capacity and coordination [68–71]. This is directly relevant to the KC-390, whose missions include domestic humanitarian support. Recent research has also examined the management of military transport aircraft, including capability-cost replacement modeling and pilots' fuel-saving behavior in logistics missions [72, 73]; evidence-based ground-time planning complements these contributions by addressing the operational productivity of the fleet.

2.4 Statistical treatment of skewed operational durations

Ground times are non-negative, right-skewed and contaminated by events generated by a different process, such as crew rest, maintenance or cargo unavailability. The arithmetic mean is sensitive to such observations, whereas the median and trimmed means provide robust alternatives with known efficiency trade-offs [74–76]. Trimmed means were popularized in signal processing as a compromise between averaging and median filtering [77] and have been applied wherever extreme values must be removed without discarding structure [78]. For asymmetric distributions, symmetric trimming biases the location estimate, and one-sided or asymmetric truncation is preferable [79].

Parametric modeling of durations typically compares candidate families using information criteria and goodness-of-fit statistics [80–82], with confidence intervals obtained from asymptotic theory or the parametric bootstrap [83]. When data are available only in grouped form, likelihood methods for interval-censored observations remain valid [84, 85]. Heaping—the concentration of recorded values at focal points—indicates that the recording or generating process is influenced by reference values [86], and in operations management the observation that work expands to fill the time available implies that observed service times depend on the time allowed, which biases capacity estimates that ignore the effect [87]. When many units are compared simultaneously, multiplicity must be controlled [88, 89], and estimates for units with small samples can be stabilized by partial pooling toward a common mean [90–92].

2.5 Research gap

Table 1 summarizes the four streams. The turnaround literature establishes that ground times are stochastic and airport-specific but relies on milestone data from scheduled commercial operations. The buffer literature provides a decision-theoretic interpretation of planned durations but has not been applied to regulatory ground-time allowances. Military airlift models use ground times as fixed inputs without validating them against operations. Robust statistics offers the necessary tools, but these have rarely been combined into an auditable procedure for evaluating a planning standard. This study sits at the intersection of the four streams.

Table 1 Literature streams informing this study and the gap addressed

Stream	Representative studies	Typical data and methods	Main insight	Gap addressed here
Turnaround operations	Wu and Caves [13]; Oreschko et al. [10]; Schmidt [11]; Malandri et al. [24]; Rott et al. [21]; Tang et al. [22]	Milestone (A-CDM) data at commercial airports; stochastic simulation; process mining; machine learning	Ground processes are stochastic and airport-specific	Evidence for non-scheduled operators using only landing and takeoff times
Buffers and reliability	Wu [1]; AhmadBeygi et al. [2]; Deshpande and Arıkan [12]; Hao and Hansen [3]; van Schilt et al. [35]	Airline schedules and delay records; newsvendor and optimization models	Planned durations are implicit service levels with a cost trade-off	Service-level reading of a regulatory ground-time allowance
Military airlift	Morton et al. [59]; Baker et al. [61]; Dietz [65]; Brown et al. [63]; Department of the Air Force [4]; U.S. Government Accountability Office [5]	Network optimization, queuing, planning factors	Ground times constrain throughput and crew use	Empirical validation of a ground-time planning factor
Robust statistics	Tukey and McLaughlin [74]; Marazzi and Ruffieux [79]; Delignette-Muller and Dutang [82]; Holm [88]; DerSimonian and Laird [91]	Trimmed means, distribution fitting, multiplicity control, partial pooling	Tools for skewed data and many small units	Combined into an auditable adjust, monitor or retain rule

3 Operational context and data

3.1 Airlift organization and the KC-390 ground-time standard

Logistic air transport moves personnel and materiel in support of national logistics needs, ranging from planned exercises to operational and humanitarian missions [93]. Two commands structure the activity. The Preparation Command (COMPREP) is responsible for training and for the standards that govern it, and the Aerospace Operations Command (COMAE) orders employment missions, that is, missions other than training, which are passed to air bases and then to transport squadrons that assign aircraft and crews. Employment missions are activated on demand and must preserve rapid-response capacity, so their planning combines route design with detailed assumptions about ground support. This study considers only employment missions ordered by COMAE.

COMPREP's standard NOPREP/SGV/01E sets, among other planning factors, the ground time to be assumed between a landing and the next takeoff. For the KC-390 the value is two hours, intended for unloading, refueling, provisioning and reloading [8]. The same standard defines three crew compositions—a simple crew with two pilots, a composite crew with three and a relief crew with at least four—each associated with a different maximum duty period, since off-duty pilots can rest in flight. Ground time therefore enters crew planning directly: the sum of planned flight and ground times along a route determines which crew composition a mission requires. The standard does not differentiate between aerodromes.

The KC-390 can carry up to 26 tonnes of cargo or 80 troops [9, 94]. Its multimission design adds tasks that are absent from airline turnarounds—installing or removing seats, reconfiguring the cargo compartment and adjusting the load for weight and balance—and that may differ between legs of the same mission.

3.2 Data source

The data are mission reports filed by aircraft commanders in the Hercules 2 system for employment missions flown in 2023 and 2024. The extraction contained 36,773 flight-leg records from 25 aircraft fleets, of which 2,778 were KC-390 legs of missions recorded as totally or partially accomplished. The main fields (Appendix Table 11) include landing and takeoff dates and times, departure and arrival aerodromes, passengers, cargo mass and volume, fuel and lubricant consumption, flight time, completion status, mission and order identifiers, and two free-text fields. We initially searched the mission-report free text for the term *delay* (*atraso*), which appeared in 1,001 of the 2,778 records. The free-text fields are, however, replicated across all legs of a mission, so a reported delay cannot be attributed to a specific ground event; they were therefore not used. Aerodromes are identified by anonymized codes (AD-01, AD-02, ...) for confidentiality.

3.3 Ground-time construction

Because the reports contain no on-block or off-block times, ground time was constructed by ordering the legs of each mission and pairing each departure with the preceding landing in the same mission:

$$GT_{m,i} = DT^{\mathrm{dep}}_{m,i} - DT^{\mathrm{land}}_{m,i-1}, \tag{1}$$

where $DT^{\mathrm{dep}}_{m,i}$ is the date and time of takeoff of leg i of mission m and $DT^{\mathrm{land}}_{m,i-1}$ is the date and time of landing of the previous leg. Each ground event is attributed to the departure aerodrome of leg i, which coincides with the arrival aerodrome of leg $i-1$ in all but four of the events retained. First legs, for which no preceding landing exists, were excluded, leaving 2,196 intervals.

Five non-positive intervals, attributable to recording errors, and 512 intervals longer than 10 h were then excluded. The applicable standard prescribes a minimum crew rest of 10 h between landing and the next takeoff [8]. An interval of that length is therefore more plausibly an overnight or crew-rest stop, a rescheduling or a deliberate

hold than a ground-service event, and retaining it would mix the crew duty cycle with the ground process that the two-hour parameter is meant to represent. The resulting analytical sample contains 1,679 ground events, hereafter also referred to as stops. Fig. 1 summarizes the construction of the sample and the structure of the analysis.

The landing-to-takeoff interval is longer than the on-block-to-off-block turnaround because it includes taxi-in and taxi-out, and it may include short waits for clearance, cargo or passengers. It is, however, exactly the interval governed by the planning standard, which makes it the appropriate measure for evaluating that standard.

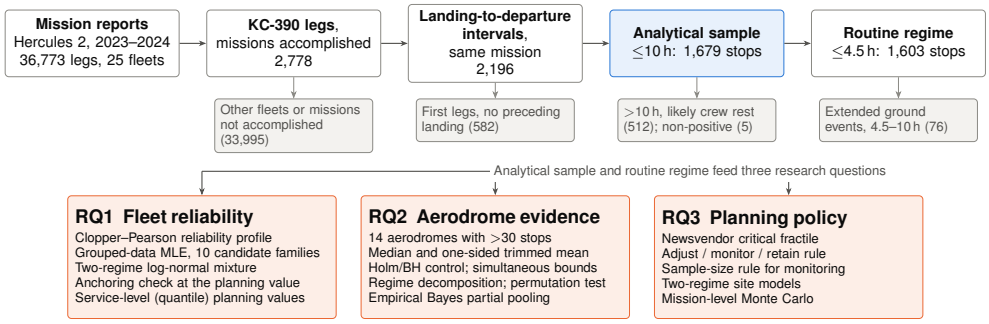


Figure 1 Construction of the analytical sample and structure of the analysis. Counts in gray boxes are records removed at each step

4 Methods

4.1 Analytical design

The analysis proceeds in two layers. The *leg-level layer* requires the restricted records and comprises the construction of ground times, the descriptive statistics by aerodrome and the fitting of probability models to aerodrome-level samples (Section 4.3). The *aggregate layer* uses only statistics reported in this article: the fleet-level frequency distribution in half-hour classes (Table 2), the frequencies of the five-minute values around the planning value (Section 5.1), the aerodrome-level counts, means and medians before and after trimming (Table 5), and the model-based estimates and confidence intervals (Table 6). Given these inputs and the procedures specified below, the aggregate-layer results can be replicated without access to the restricted records. Results that involve resampling or simulation (bootstrap intervals, the permutation test and the mission-level simulation) will agree up to Monte Carlo error, and quantities recomputed from tabulated values may differ in the last reported digit because of rounding. The aggregate layer was implemented in Python 3.12 with NumPy, SciPy and pandas, and the leg-level layer in R with the `fitdistrplus` package [82].

4.2 Fleet-level reliability and regime separation

We define the *reliability* of a planning allowance t as $R(t) = \Pr(GT \leq t)$, the proportion of stops that are completed within the allowance, and refer to $R(2)$ as the

service level of the current standard. Empirical reliabilities are reported at half-hour thresholds with Clopper–Pearson 95% confidence intervals [95].

The fleet-level data consist of counts n_j in 20 half-hour classes $(a_j, b_j]$ covering $(0, 10]$ h. Since events longer than 10 h were excluded, the distribution is right-truncated, and for a candidate cumulative distribution function F_θ the grouped-data log-likelihood is

$$\ell(\theta) = \sum_{j=1}^{20} n_j \log \frac{F_\theta(b_j) - F_\theta(a_j)}{F_\theta(10)}, \tag{2}$$

which is the standard likelihood for interval-censored data [84, 85]. We fitted the nine candidate families specified for the aerodrome-level analysis (exponential, normal, log-normal, gamma, Weibull, log-logistic, inverse Gaussian, Burr XII and generalized gamma) and a two-component log-normal mixture, $F = \pi F_R + (1 - \pi)F_E$, in which F_R represents routine ground service and F_E extended ground events. Models were compared by AIC and BIC, and absolute fit was assessed with Pearson's χ^2 statistic after pooling classes with expected counts below five. For the mixture, the posterior probability that an event of duration t belongs to the extended regime is

$$P_E(t) = \frac{(1 - \pi)f_E(t)}{\pi f_R(t) + (1 - \pi)f_E(t)}, \tag{3}$$

which provides a data-driven benchmark for the trimming threshold.

The one-sided trimming threshold of 4.5 h was set at the first half-hour class at which at least 95% of the analytical sample had accumulated (Section 5.1). We assess this choice in two ways: by locating it on the posterior curve (3), and by computing the fitted conditional mean $E[GT \mid GT \le c]$ for thresholds c between 3 and 10 h.

Planning values for a target service level p are the quantiles $t_p = R^{-1}(p)$. We report empirical quantiles, obtained by linear interpolation within classes [96], with 95% intervals from 5,000 multinomial bootstrap resamples, and mixture-based quantiles with intervals from 300 parametric bootstrap refits [83].

Finally, we checked whether recorded ground times are anchored to the plan, that is, whether departures tend to be recorded at the planning value, or just after it, rather than when ground service is completed. Such anchoring would make observed ground time partly a product of the standard itself [87] and would appear as heaping of recorded values at the planning value [86]. Because times are recorded on a five-minute grid, we compare the frequencies of the values 2 h 00 min and 2 h 05 min in the trimmed sample with the average frequency of the other five-minute values between 1 h 45 min and 2 h 20 min, using a Poisson reference distribution.

4.3 Aerodrome-level robust inference

Aerodromes with more than 30 ground events were analyzed individually; 14 aerodromes, accounting for 1,251 of the 1,679 events, met this criterion. Two robust central measures were used: the median of the analytical sample and the mean of the one-sided trimmed sample ($GT \le 4.5$ h), hereafter the trimmed mean. Aerodromes whose sample median (four aerodromes) or trimmed mean (five aerodromes) exceeded two

hours were modeled. For each, the candidate families were fitted by maximum likelihood with `fitdistrplus`; of the nine families specified, the generalized gamma could not be estimated with the software used and one further family failed to converge in one case, so seven or eight families were compared in each case. The family with the lowest AIC was selected, and its fit was examined graphically (Appendix C). Two 95% confidence intervals were computed for the theoretical median or mean θ of the selected distribution. The asymptotic interval is $\theta \pm z_{0.975}\,\mathrm{se}$, with $\mathrm{se} = \{4nf(\theta)^2\}^{-1/2}$ for the median, where f is the fitted density, and $\mathrm{se} = \{\mathrm{Var}(GT)/n\}^{1/2}$ for the mean, where the variance is that of the fitted distribution. The parametric-bootstrap interval is formed by the 2.5th and 97.5th percentiles of the medians (or means) of 5,000 samples of size n drawn from the selected distribution with its parameters fixed at the maximum-likelihood estimates.

We formalize the decision problem as a one-sided test of $H_0 : \theta_a \le 2$ h against $H_1 : \theta_a > 2$ h, where θ_a is the median or trimmed mean at aerodrome a. The standard error is recovered from the asymptotic interval as $\widehat{\mathrm{se}}_a = (U_a - L_a)/(2z_{0.975})$, and the one-sided p-value is $1 - \Phi\{(\hat{\theta}_a - 2)/\widehat{\mathrm{se}}_a\}$. Because 14 aerodromes were screened, the relevant family has 14 members for each measure. Aerodromes that were not modeled have point estimates of at most two hours and hence one-sided p-values of at least 0.5; we assign them $p = 1$, which leaves the family size unchanged and makes the adjustment conservative. We report Holm-adjusted p-values [88], Benjamini–Hochberg q-values [89] and simultaneous one-sided 95% lower confidence bounds $\hat{\theta}_a - z_{1-0.05/14}\,\widehat{\mathrm{se}}_a$ with $z_{1-0.05/14} = 2.69$. Each aerodrome is then assigned to one of three tiers: *adjust* if the Holm-adjusted p-value is below 0.05 for both the median and the trimmed mean; *monitor* if at least one point estimate exceeds two hours but the evidence does not meet the adjustment criterion; and *retain* otherwise.

For aerodromes in the monitor tier we compute the minimum detectable excess with 80% power at the one-sided 5% level, $\mathrm{MDE}_a = (z_{0.95} + z_{0.80})\,\widehat{\mathrm{se}}_a$, and the sample size at which the observed excess would be detectable, $n_a^{*} = n_a\{(z_{0.95} + z_{0.80})\widehat{\mathrm{se}}_a/(\hat{\theta}_a - 2)\}^2$. Dividing $n_a^{*} - n_a$ by the annual accrual rate observed in 2023–2024 gives the additional years of data required.

4.4 Routine and extended regimes across aerodromes

For each aerodrome a with n_a events, n_a^{T} of which do not exceed 4.5 h, the share of extended events is $p_a = (n_a - n_a^{T})/n_a$, reported with Clopper–Pearson intervals. Writing $\bar{x}_a$ and $\bar{x}_a^{T}$ for the means before and after trimming, the mean duration of the extended events is $\mu_a^{E} = (n_a \bar{x}_a - n_a^{T} \bar{x}_a^{T})/(n_a - n_a^{T})$, and

$$\bar{x}_a - \bar{x}_a^{T} = p_a\,(\mu_a^{E} - \bar{x}_a^{T}) \tag{4}$$

decomposes the difference between the two means into the frequency and the severity of extended events. Heterogeneity of p_a across the 14 aerodromes and the pooled remaining aerodromes was tested with Pearson's χ^2 statistic, whose p-value was obtained from 20,000 permutations with fixed margins; each aerodrome was also compared with all others by a one-sided Fisher exact test with Holm adjustment.

4.5 Partial pooling across aerodromes

Aerodrome medians based on 32–265 events differ in precision, and extreme estimates from small samples partly reflect noise. We therefore fitted a random-effects model to the log medians, $y_a = \log \hat{m}_a \sim N(\theta_a, s_a^2)$ with $\theta_a \sim N(\mu, \tau^2)$, estimating τ^2 by the DerSimonian–Laird method and summarizing heterogeneity with I^2 [91, 92]. The sampling variance of each median follows from the asymptotic result $\mathrm{Var}(\hat{m}_a) \approx \{4 n_a f_a(m_a)^2\}^{-1}$, with the density f_a taken from the site models described in Section 4.6, and $s_a = \mathrm{se}(\hat{m}_a)/\hat{m}_a$ by the delta method. Empirical Bayes estimates shrink each log median toward the pooled mean, $\tilde{\theta}_a = B_a y_a + (1 - B_a)\hat{\mu}$ with $B_a = \hat{\tau}^2/(\hat{\tau}^2 + s_a^2)$ [90], and yield the posterior probability that the median exceeds two hours and a 95% prediction interval for the median of an aerodrome not in the sample. As sensitivity analyses, we replaced s_a by the model-based standard errors where available and inflated all s_a by 50%.

4.6 Planning policy analysis

Service levels and implied costs

Let c_u denote the cost of each hour by which ground time exceeds the allowance (duty time consumed, downstream delay and the risk of an unplanned overnight stop) and c_o the cost of each hour of allowance that is not used (idle aircraft and crew time). The allowance that minimizes $c_u\, E[(GT - t)^+] + c_o\, E[(t - GT)^+]$ satisfies the newsvendor condition

$$R(t^*) = \frac{c_u}{c_u + c_o}, \tag{5}$$

[12, 41]. Conversely, an allowance with service level $R(t)$ is optimal only if $c_u/c_o = R(t)/\{1 - R(t)\}$, which we call the implied cost ratio of the standard.

Site models

Site-level reliabilities and mission simulations require a distribution for each aerodrome. We represent each aerodrome by a two-regime model calibrated only to published statistics: with probability $1 - p_a$ a routine event drawn from a log-normal distribution truncated to $(0, 4.5]$ h whose median and mean equal the aerodrome's trimmed median and trimmed mean, and with probability p_a an extended event drawn from the grouped distribution of the 76 events between 4.5 and 10 h in Table 2: a half-hour class is selected with probability proportional to its count, and a value is drawn uniformly within that class. The extended events are therefore represented by their grouped fleet-level distribution, not by the individual durations, which are not disclosed. The implied medians of the full distributions, which are not used in calibration, are compared with the reported medians as a validation check (Appendix B).

Planning values

For each aerodrome in the adjust tier we consider two candidate allowances derived from the model-based estimates: a *suggested* value, equal to the larger of the fitted median and the fitted trimmed mean, for routine planning; and a *conservative* value, equal to the largest upper 95% confidence bound across the asymptotic and bootstrap

intervals of the two measures, for missions in which an overrun would be especially costly. Their service levels at each aerodrome are obtained from the site models.

Mission-level simulation

We compare four policies: a uniform allowance of 2.00 h; a localized policy that retains 2.00 h except at aerodromes in the adjust tier, where the suggested value is used; a localized conservative policy that uses the conservative value at those aerodromes; and a service-level policy that sets the allowance at each stop to the aerodrome's 70th percentile. Three stylized mission sequences are considered: two stops at adjust-tier aerodromes (S1); three stops combining an adjust-tier aerodrome with a high-traffic aerodrome and a fast aerodrome (S2); and three stops drawn in proportion to traffic across the 14 aerodromes (S3). For each policy and sequence we simulate 200,000 missions assuming independent ground times across stops and report the probability that cumulative ground time exceeds the cumulative allowance by more than 0, 30 and 60 minutes, the expected overrun and the expected unused allowance.

5 Results

5.1 Fleet-level reliability of the two-hour standard

Table 2 reports the distribution of the analytical sample in half-hour classes and the corresponding reliability profile. Of the 1,679 ground events, 962 (57.3%; 95% CI 54.9–59.7%) were completed within two hours, so the standard was exceeded at 42.7% of stops. The fleet-level median was 1.83 h (95% CI 1.79–1.88 h): the standard lies above the typical stop but corresponds only to its 57th percentile. Under Eq. (5), a service level of 57.3% is optimal only if an hour of overrun costs 1.34 times as much as an hour of unused allowance (95% CI 1.22–1.48). The share of events completed within 4.5 h was 95.5%, and the remaining 76 events were spread thinly between 4.5 and 10 h.

The distribution is strongly right-skewed (Table 3; Fig. 2a). Light-tailed families (exponential, normal, Weibull and gamma) are clearly rejected, and the log-normal, inverse Gaussian and generalized gamma distributions also fail Pearson's test. Only the two-component log-normal mixture ($p = 0.69$) and the Burr XII distribution ($p = 0.10$) are compatible with the data; the mixture has the lowest AIC, whereas BIC, which penalizes its five parameters more heavily, favors the log-logistic distribution, which fails the goodness-of-fit test at the 5% level. The mixture separates a routine component with a median of 1.78 h and a log-scale standard deviation of 0.46 from a diffuse extended component that accounts for 6.4% of events. The posterior probability of the extended regime is 0.02 at two hours, 0.21 at 4.5 h and exceeds 0.5 only beyond 5.7 h (Fig. 2d). The 4.5 h threshold therefore lies in the transition zone and is conservative: it removes the events most likely to belong to the extended regime—the mixture attributes 57% of the events above 4.5 h to that regime—while also trimming some long routine stops. The choice is not consequential for the central estimate: the fitted conditional mean equals 1.86, 1.91, 1.94 and 1.99 h for thresholds of 4.0, 4.5, 5.0 and 6.0 h, a range of eight minutes.

Recorded ground times show no anchoring to the planning value (Fig. 2b). All but one of the 1,603 events in the trimmed sample lie on the five-minute grid. The values

Table 2 Fleet-level distribution and reliability profile of the 1,679 ground events in half-hour classes

		Completed within the upper limit		
Class (h)	Events in class	Events	Share (%)	95% CI (%)
0.0–0.5	12	12	0.7	[0.4, 1.2]
0.5–1.0	168	180	10.7	[9.3, 12.3]
1.0–1.5	411	591	35.2	[32.9, 37.5]
1.5–2.0	371	962	57.3	[54.9, 59.7]
2.0–2.5	285	1,247	74.3	[72.1, 76.3]
2.5–3.0	187	1,434	85.4	[83.6, 87.1]
3.0–3.5	90	1,524	90.8	[89.3, 92.1]
3.5–4.0	54	1,578	94.0	[92.7, 95.1]
4.0–4.5	25	1,603	95.5	[94.4, 96.4]
4.5–5.0	16	1,619	96.4	[95.4, 97.3]
5.0–5.5	15	1,634	97.3	[96.4, 98.0]
5.5–6.0	10	1,644	97.9	[97.1, 98.5]
6.0–6.5	8	1,652	98.4	[97.7, 98.9]
6.5–7.0	6	1,658	98.7	[98.1, 99.2]
7.0–7.5	6	1,664	99.1	[98.5, 99.5]
7.5–8.0	2	1,666	99.2	[98.7, 99.6]
8.0–8.5	3	1,669	99.4	[98.9, 99.7]
8.5–9.0	3	1,672	99.6	[99.1, 99.8]
9.0–9.5	3	1,675	99.8	[99.4, 99.9]
9.5–10.0	4	1,679	100.0	[99.8, 100.0]

Note: The class counts are the grouped data used in Eq. (2); the cumulative share at the upper limit of a class is the reliability of an allowance equal to that limit. Clopper–Pearson intervals.

Table 3 Distributions fitted to the grouped fleet-level data (1,679 events in 20 half-hour classes, right-truncated at 10 h), ordered by AIC

Distribution	k	Log-likelihood	ΔAIC	ΔBIC	χ^2 (df)	p
Log-normal mixture	5	-3461.0	0.0	5.6	8.2 (11)	0.694
Burr XII	3	-3466.6	7.2	1.9	18.6 (12)	0.100
Log-logistic	2	-3469.4	10.7	0.0	22.4 (12)	0.033
Generalized gamma	3	-3478.7	31.4	26.1	46.5 (11)	<0.001
Log-normal	2	-3482.4	36.7	26.0	46.0 (11)	<0.001
Inverse Gaussian	2	-3490.1	52.1	41.4	58.4 (11)	<0.001
Gamma	2	-3548.3	168.5	157.8	176.1 (10)	<0.001
Weibull	2	-3670.7	413.4	402.7	346.5 (10)	<0.001
Normal	2	-3811.4	694.7	684.0	588.9 (10)	<0.001
Exponential	1	-4069.8	1209.6	1193.5	1059.2 (17)	<0.001

Note: k is the number of parameters; the log-normal mixture has two components. χ^2 is Pearson's statistic after pooling classes with expected counts below five.

2 h 00 min and 2 h 05 min were recorded 55 and 51 times, whereas the other five-minute values between 1 h 45 min and 2 h 20 min were recorded 55, 74, 50, 65, 61 and

43 times (1 h 45 to 1 h 55 min and 2 h 10 to 2 h 20 min), an average of 58. Neither value exceeds that average ($p = 0.67$ and $p = 0.84$ against a Poisson reference). There is therefore no sign that departures are systematically recorded at, or just after, the planned allowance, which supports reading the observed intervals as the durations of the ground processes rather than as a reflection of the standard.

Table 4 converts the fleet distribution into planning values for target service levels. Empirical and mixture-based quantiles agree to within 0.04 h at every tabulated service level. A planning value of 2.08 h would have covered 60% of stops, 2.37 h 70%, 2.76 h 80% and 3.43 h 90%; each ten-point increase in the service level therefore requires a progressively larger allowance, reflecting the long right tail.

Table 4 Fleet-level planning values for target service levels and the cost ratio each implies

		Planning value (h) [95% CI]	
Service level (%)	Implied c_u/c_o	Empirical	Mixture
50	1.00	1.83 [1.79, 1.88]	1.81 [1.77, 1.85]
60	1.50	2.08 [2.01, 2.14]	2.04 [1.99, 2.09]
70	2.33	2.37 [2.32, 2.43]	2.33 [2.27, 2.40]
75	3.00	2.53 [2.46, 2.62]	2.51 [2.44, 2.59]
80	4.00	2.76 [2.68, 2.84]	2.73 [2.66, 2.82]
85	5.67	2.98 [2.91, 3.11]	3.02 [2.93, 3.14]
90	9.00	3.43 [3.30, 3.59]	3.45 [3.34, 3.61]
95	19.00	4.34 [3.99, 4.74]	4.33 [4.08, 4.64]

Note: Empirical quantiles by linear interpolation within half-hour classes, with intervals from 5,000 multinomial bootstrap resamples; mixture quantiles with intervals from 300 parametric bootstrap refits. The current standard (2.00 h) corresponds to a service level of 57.3% and an implied cost ratio of 1.34.

5.2 Aerodrome heterogeneity: routine service and extended events

Table 5 and Fig. 3 summarize the 14 aerodromes with more than 30 events. Medians range from 1.33 h at AD-06 and AD-07 to 2.50 h at AD-02. Twelve aerodromes have mean ground times above two hours, but only four have medians and five have trimmed means above that value, which illustrates how strongly the arithmetic mean is driven by a minority of long events. The decomposition in Eq. (4) makes this explicit. The difference between the mean and the trimmed mean ranges from 0.05 h (AD-07) to 0.59 h (AD-14) and is driven by both the frequency and the duration of extended events; at AD-02, for example, 17 extended events with a mean duration of 6.3 h raise the mean by 0.47 h.

The share of extended events varies from 1.0% at AD-06 to 11.7% at AD-02 (Fig. 3b). The hypothesis of a common share across the 14 aerodromes and the pooled

Table 5 Aerodromes with more than 30 ground events: central tendency before and after one-sided trimming, and decomposition of extended ground events (above 4.5 h)

	Analytical sample (≤10 h)			Trimmed sample (≤4.5 h)			Extended events (>4.5 h)			
Aerodrome	n	Mean (h)	Median (h)	n	Mean (h)	Median (h)	n	Share (%) [95% CI]	Mean (h)	Mean inflation (h)
AD-01	265	2.14	1.92	251	1.91	1.83	14	5.3 [2.9, 8.7]	6.3	0.23
AD-02	145	2.80	2.50	128	2.33	2.25	17	11.7 [7.0, 18.1][a]	6.3	0.47
AD-03	135	2.46	2.33	132	2.39	2.29	3	2.2 [0.5, 6.4]	5.5	0.07
AD-04	119	2.17	1.83	114	2.00	1.83	5	4.2 [1.4, 9.5]	6.0	0.17
AD-05	112	2.06	1.83	108	1.92	1.83	4	3.6 [1.0, 8.9]	5.9	0.14
AD-06	100	1.59	1.33	99	1.53	1.33	1	1.0 [0.0, 5.4]	7.4	0.06
AD-07	66	1.61	1.33	65	1.56	1.33	1	1.5 [0.0, 8.2]	5.0	0.05
AD-08	61	2.10	1.67	58	1.79	1.63	3	4.9 [1.0, 13.7]	8.2	0.31
AD-09	56	2.07	1.79	55	2.00	1.75	1	1.8 [0.0, 9.6]	6.1	0.07
AD-10	47	2.31	2.17	44	2.14	2.08	3	6.4 [1.3, 17.5]	4.8	0.17
AD-11	41	2.72	2.42	37	2.40	2.33	4	9.8 [2.7, 23.1]	5.7	0.32
AD-12	40	2.46	1.88	39	2.28	1.83	1	2.5 [0.1, 13.2]	9.4	0.18
AD-13	32	2.03	1.67	31	1.78	1.67	1	3.1 [0.1, 16.2]	9.8	0.25
AD-14	32	2.24	1.67	29	1.65	1.42	3	9.4 [2.0, 25.0]	7.9	0.59

Notes: Mean inflation is the difference between the mean and the trimmed mean (Eq. 4). Shares with Clopper–Pearson intervals. Mean durations of extended events are computed from the rounded trimmed means and are accurate to within ±0.04–0.5 h; the bound is widest at aerodromes with a single extended event. [a] Share significantly higher than at all other aerodromes (one-sided Fisher exact test, Holm-adjusted $p < 0.05$).

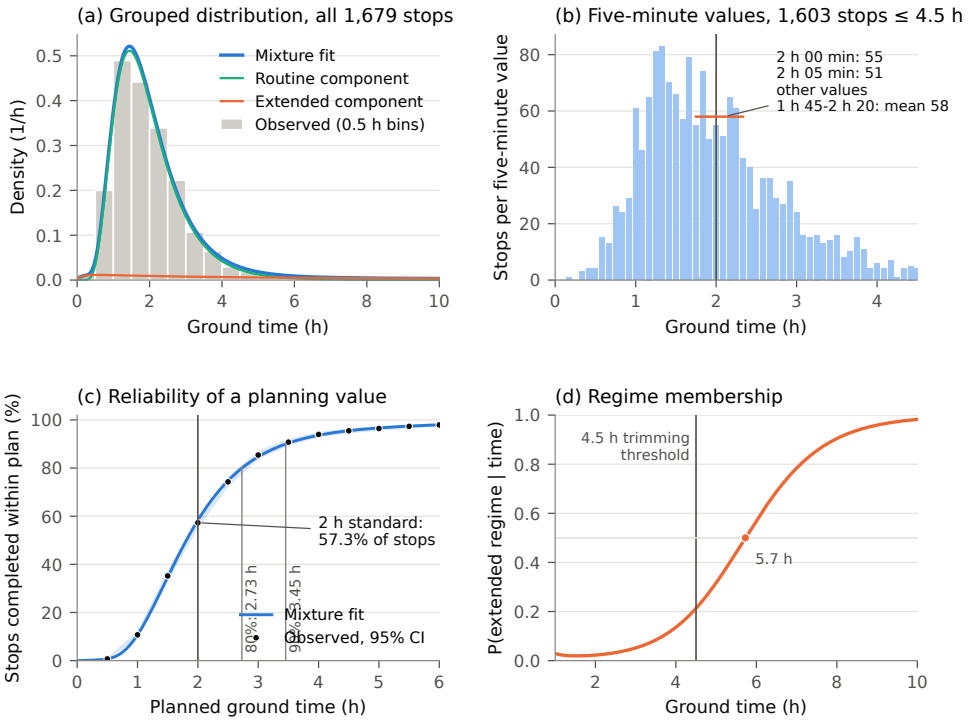


Figure 2 Fleet-level distribution of KC-390 ground time. (a) Grouped distribution of the analytical sample with the fitted two-component log-normal mixture. (b) Frequencies of the five-minute values in the trimmed sample; the horizontal segment is the average frequency of the values between 1 h 45 min and 2 h 20 min other than 2 h 00 and 2 h 05 min. (c) Reliability of a planning allowance with Clopper–Pearson intervals; the vertical lines mark the allowances required for 80% and 90% service levels under the mixture fit (Table 4). (d) Posterior probability that an event of a given duration belongs to the extended regime

remaining aerodromes is rejected ($\chi^2_{14} = 31.2$, permutation $p = 0.007$). AD-02 stands out: its share is 2.6 times the fleet average, and it is the only aerodrome whose share is significantly higher than elsewhere after Holm adjustment. Heterogeneity is therefore present both in routine service and in the frequency of extended events, and the two do not coincide: AD-03 combines a long routine service with a low share of extended events, whereas AD-14 combines a short routine service with a high share.

5.3 Aerodrome-level inference and robustness

Table 6 and Fig. 4 report the model-based estimates for the aerodromes whose median or trimmed mean exceeded two hours. AD-02, AD-03 and AD-11 exceed the standard on both measures. Their one-sided p-values remain below 0.05 after Holm adjustment over the 14 screened aerodromes (the largest adjusted value is 0.028, for the AD-11 median), and all six simultaneous one-sided lower bounds exceed two hours, ranging from 2.02 h (AD-11, median) to 2.20 h (AD-03, trimmed mean). Benjamini–Hochberg q-values lead to the same classification: at most 0.011 for these aerodromes and at

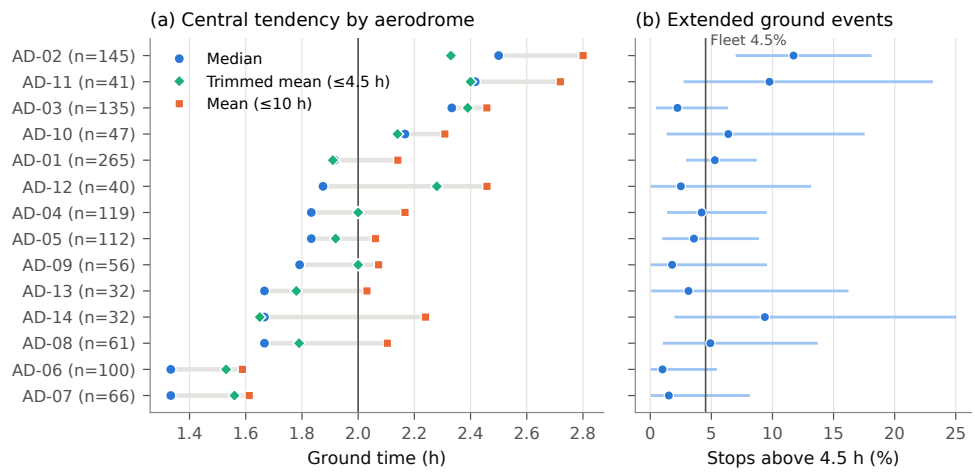


Figure 3 Aerodromes with more than 30 ground events. (a) Median, trimmed mean and mean; the gray segment spans the median and the mean. (b) Share of events above 4.5 h with Clopper–Pearson 95% intervals; the vertical line is the fleet-wide share

least 0.21 for the others. The asymptotic and parametric-bootstrap intervals differ by at most 0.03 h, so the conclusions do not depend on the interval method. These three aerodromes are assigned to the adjust tier.

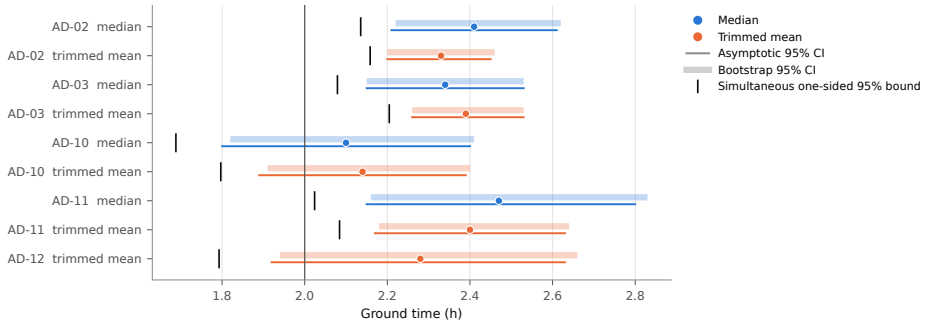


Figure 4 Model-based estimates for aerodromes with a median or trimmed mean above two hours: asymptotic and parametric-bootstrap 95% confidence intervals and simultaneous one-sided 95% lower bounds (Bonferroni, 14 aerodromes). The vertical line marks the current standard

AD-10 and AD-12 have point estimates above two hours but do not meet the adjustment criterion and are assigned to the monitor tier. Table 7 shows why the evidence is inconclusive and what would resolve it. At AD-12 the trimmed mean exceeds the standard by 0.28 h, but with 39 events the minimum detectable excess is 0.45 h; about 101 events—roughly three further years at the 2023–2024 rate—would be needed to detect an excess of the observed size. At AD-10 the observed excess is small (0.10–0.14 h, or six to eight minutes) and would require 226 to 681 events, that

Table 6 Model-based inference for aerodromes with a median or trimmed mean above two hours, with multiplicity control over the 14 screened aerodromes

Aerodrome	Measure	Distribution	n	Estimate (h)	Asymptotic 95% CI	Bootstrap 95% CI	z	Holm p	Simultaneous bound (h)	Tier
AD-02	Median	Burr	145	2.41	2.21–2.61	2.22–2.62	4.02	<0.001	2.14	Adjust
AD-02	Trimmed mean	Normal	128	2.33	2.20–2.45	2.20–2.46	5.17	<0.001	2.16	Adjust
AD-03	Median	Gamma	135	2.34	2.15–2.53	2.15–2.53	3.51	0.003	2.08	Adjust
AD-03	Trimmed mean	Weibull	132	2.39	2.26–2.53	2.26–2.53	5.66	<0.001	2.20	Adjust
AD-10	Median	Log-logistic	47	2.10	1.80–2.40	1.82–2.41	0.65	1.000	1.69	Monitor
AD-10	Trimmed mean	Gamma	44	2.14	1.89–2.39	1.91–2.40	1.10	1.000	1.80	Monitor
AD-11	Median	Log-logistic	41	2.47	2.15–2.80	2.16–2.83	2.83	0.028	2.02	Adjust
AD-11	Trimmed mean	Gamma	37	2.40	2.17–2.63	2.18–2.64	3.41	0.004	2.08	Adjust
AD-12	Trimmed mean	Inverse Gaussian	39	2.28	1.92–2.63	1.94–2.66	1.55	0.672	1.79	Monitor

Notes: Estimates are the theoretical median or mean of the selected distribution (diagnostics in Appendix C). $z = (\hat{\theta} - 2)/\widehat{\mathrm{se}}$ with the standard error recovered from the asymptotic interval. Holm-adjusted one-sided p-values over the 14 screened aerodromes. Simultaneous bound: one-sided 95% lower confidence bound with Bonferroni correction for 14 aerodromes. AD-12 appears only for the trimmed mean because its sample median did not exceed two hours.

is, 8 to 27 further years, to be detected; an excess of that size is too small to justify a change, and monitoring serves to detect any drift. The remaining nine aerodromes, whose medians and trimmed means do not exceed two hours, form the retain tier.

Table 7 Monitoring requirements for aerodromes in the monitor tier

Aerodrome	Measure	n	Observed excess (h)	MDE (h)	Required n	Additional years
AD-10	Median	47	0.10	0.38	681	27.0
AD-10	Trimmed mean	44	0.14	0.32	226	8.3
AD-12	Trimmed mean	39	0.28	0.45	101	3.2

Notes: MDE: minimum excess over two hours detectable with 80% power in a one-sided test at the 5% level. Required n: sample size at which the observed excess would be detectable with 80% power. Additional years at the 2023–2024 accrual rate.

5.4 Partial pooling across aerodromes

Aerodrome medians are strongly heterogeneous ($Q = 138.0$, 13 df, $p < 0.001$; $I^2 = 0.91$; $\hat{\tau} = 0.16$ on the log scale), and the heterogeneity persists when model-based standard errors are used ($I^2 = 0.89$) or when all standard errors are inflated by 50% ($I^2 = 0.79$). The pooled median is 1.88 h (95% CI 1.71–2.07 h), and the 95% prediction interval for the median of an aerodrome outside the sample is 1.30–2.73 h. Shrinkage toward the pooled median is small for the busiest aerodromes and substantial for the smallest ones, reaching 49% of the distance for AD-14 (Tables 5 and 8; Fig. 5).

After shrinkage, the medians of AD-02 (2.47 h; 95% interval 2.33–2.63 h), AD-03 (2.30 h; 2.13–2.49 h) and AD-11 (2.35 h; 2.12–2.61 h) remain above two hours. Their posterior probabilities of exceeding two hours are at least 0.999 in the baseline specification, at least 0.991 with model-based standard errors and at least 0.96 with inflated standard errors, so the adjust tier is robust to partial pooling. AD-10 also has a high posterior probability of exceeding two hours (0.94 in the baseline and 0.81–0.82 in the sensitivity analyses), but its shrunken median of 2.14 h exceeds the standard by only eight minutes, which is consistent with its placement in the monitor tier; AD-12 has a posterior probability of 0.28. At the other end of the distribution, the shrunken medians of AD-06 (1.42 h) and AD-07 (1.48 h) have upper interval bounds of 1.63 and 1.78 h, so the standard is systematically generous at these aerodromes.

5.5 Planning values and mission-level consequences

The site models translate these differences into reliabilities (Table 8; Fig. 6). The share of stops completed within two hours ranges from 25% at AD-11 and 30% at AD-02 to 75% at AD-06, around the fleet-wide 57%. At the three adjust-tier aerodromes the standard is met at only 25–35% of stops. The suggested allowances of 2.39–2.47 h raise this share to 53%, close to the level that the current standard delivers fleet-wide,

Table 8 Aerodrome-level reliability of the two-hour standard, service-level planning values from the site models and partially pooled medians

Aerodrome	Within 2 h (%)	Service-level value (h)		Pooled median (h) [95% interval]	P(median >2 h)
		70%	80%		
AD-01	58	2.21	2.47	1.92 [1.83, 2.01]	0.03
AD-02	30	2.83	3.25	2.47 [2.33, 2.63]	1.00
AD-03	35	2.83	3.19	2.30 [2.13, 2.49]	1.00
AD-04	54	2.58	3.07	1.84 [1.62, 2.10]	0.11
AD-05	59	2.21	2.48	1.84 [1.70, 1.98]	0.01
AD-06	75	1.84	2.22	1.42 [1.24, 1.63]	0.00
AD-07	72	1.93	2.38	1.48 [1.24, 1.78]	0.00
AD-08	63	2.21	2.63	1.71 [1.48, 1.98]	0.02
AD-09	56	2.58	3.10	1.82 [1.49, 2.23]	0.18
AD-10	41	2.44	2.68	2.14 [1.96, 2.34]	0.94
AD-11	25	2.83	3.17	2.35 [2.12, 2.61]	1.00
AD-12	54	2.64	3.15	1.88 [1.52, 2.32]	0.28
AD-13	67	2.08	2.36	1.71 [1.47, 1.99]	0.02
AD-14	63	2.34	3.01	1.77 [1.41, 2.22]	0.15

Notes: Shares within two hours and service-level values are derived from the two-regime site models (Appendix B). Pooled medians are empirical Bayes estimates from the random-effects model with standard errors derived from the site models.

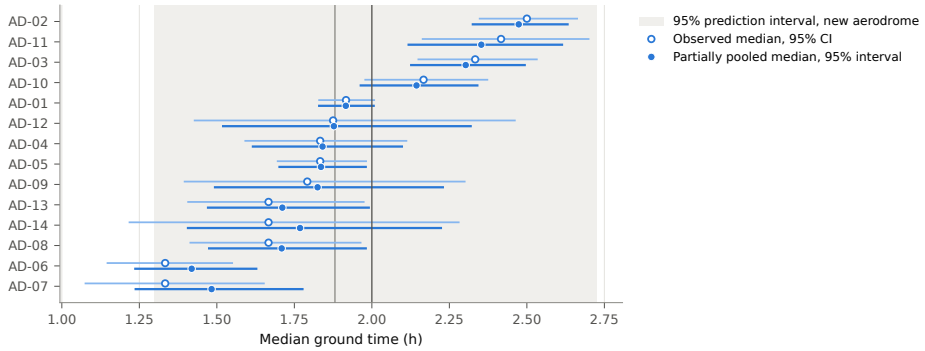


Figure 5 Observed and partially pooled (empirical Bayes) aerodrome medians with 95% intervals. The shaded band is the 95% prediction interval for the median of an aerodrome not in the sample, and the thin vertical line is the pooled median

and the conservative allowances of 2.53–2.83 h raise it to 59–70% (Table 9). A 70% service level at each of them would require about 2.83 h.

Table 10 and Fig. 7 report the mission-level consequences. For a sequence with stops at AD-02 and AD-11 (S1), the uniform standard leads to a cumulative overrun of more than 30 minutes in 66% of simulated missions and of more than one hour in 49%. Localized allowances reduce these probabilities to 38% and 29%, at the cost

Table 9 Planning values at the adjust-tier aerodromes and the service levels they deliver

	Standard	Suggested		Conservative		Service-level value (h)	
Aerodrome	Within 2 h (%)	Value (h)	Within (%)	Value (h)	Within (%)	70%	80%
AD-02	30	2.41	53	2.62	62	2.83	3.25
AD-03	35	2.39	53	2.53	59	2.83	3.19
AD-11	25	2.47	53	2.83	70	2.83	3.17

Notes: Suggested and conservative values as defined in Section 4.6, computed from the estimates in Table 6. Service levels from the site models (Appendix B).

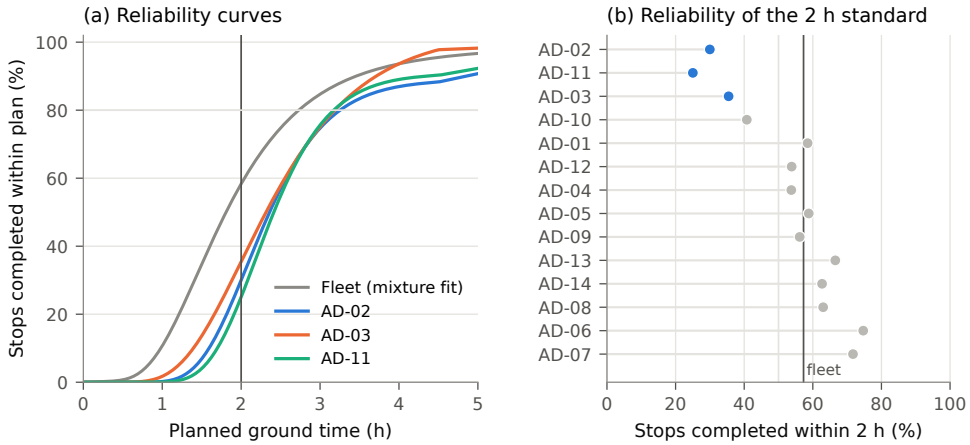


Figure 6 Reliability of planning allowances by aerodrome. (a) Share of stops completed within a given allowance at the adjust-tier aerodromes and for the fleet. (b) Share of stops completed within the two-hour standard at each aerodrome; adjust-tier aerodromes are highlighted and the vertical line marks the fleet-wide share

of 0.88 h of additional planned ground time and an increase in expected unused allowance from 0.07 to 0.35 h; the conservative values reduce them further to 28% and 23%. In the mixed sequence (S2), in which only AD-03 is adjusted, the probability of an overrun above 30 minutes falls from 35% to 27%, and in three-stop missions drawn by traffic share (S3) from 42% to 36%. Setting every allowance at the local 70th percentile yields the lowest overrun risk in all sequences but roughly doubles the expected unused allowance in S2 and S3 (1.14–1.25 h, against 0.56–0.57 h under the uniform standard). Because extended events are included in the simulation, a residual risk remains under every policy: even with the service-level policy, 14–21% of missions overrun by more than one hour. Extended events must therefore be managed through contingency arrangements rather than planned buffers.

Table 10 Mission-level consequences of alternative planning policies (200,000 simulated missions per cell)

Sequence	Policy	Planned ground time (h)	Overrun probability (%) >0	>30 min	>60 min	Expected overrun (h)	Expected unused (h)
S1	Uniform 2.00 h	4.00	83	66	49	1.64	0.07
	Localized, suggested	4.88	52	38	29	1.04	0.35
	Localized, conservative	5.45	36	28	23	0.79	0.67
	70th percentile at each stop	5.66	32	25	21	0.72	0.81
S2	Uniform 2.00 h	6.00	47	35	25	0.77	0.57
	Localized, suggested	6.39	37	27	19	0.60	0.80
	Localized, conservative	6.53	34	25	18	0.55	0.89
	70th percentile at each stop	6.88	27	19	14	0.44	1.14
S3	Uniform 2.00 h	6.00	53	42	33	1.13	0.56
	Localized, suggested	6.32	45	36	28	0.96	0.70
	Localized, conservative	6.47	42	33	26	0.89	0.79
	70th percentile at each stop	7.20	29	22	17	0.63	1.25

Notes: S1: stops at AD-02 and AD-11; S2: stops at AD-01, AD-03 and AD-06; S3: three stops drawn in proportion to traffic across the 14 aerodromes. Overrun: cumulative ground time minus cumulative allowance across the stops of a mission. Localized policies retain 2.00 h except at AD-02, AD-03 and AD-11, where the suggested or conservative values (Table 9) are used; the last policy sets each allowance at the aerodrome's 70th percentile. Ground times are simulated from the site models, including extended events, and are assumed independent across stops.

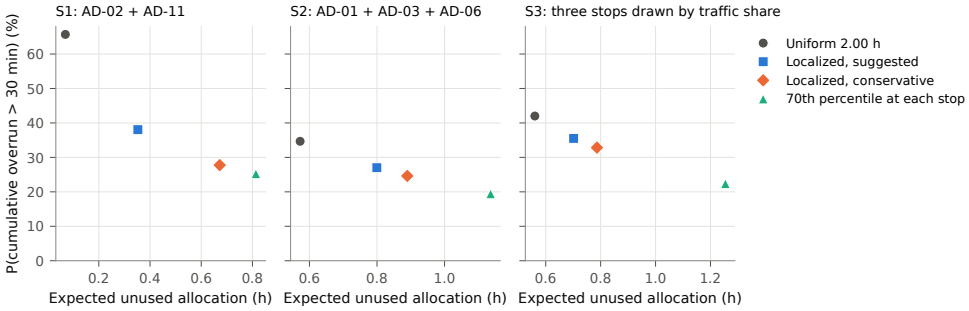


Figure 7 Trade-off between the probability of a cumulative overrun above 30 minutes and the expected unused allowance for four planning policies in three mission sequences

6 Discussion

6.1 A planning standard is an implicit service level

The two-hour standard is not wrong on average: the fleet median of 1.83 h lies below it. Read as a service level, however, it covers only 57% of stops, which implies a cost ratio c_u/c_o of about 1.3 (Section 5.1). In airlift operations that ratio appears low. An overrun consumes crew duty time, can cascade into an unplanned overnight stop that delays cargo and passengers by many hours, and at the limit compromises the fatigue margins that duty limits are designed to protect [54]; an unused hour mostly idles an aircraft that is already committed to the mission. If planners judged the ratio to be 4:1, the corresponding fleet-wide allowance would be about 2.75 h (Table 4). We do not claim to know the correct ratio. The point is that the current standard embeds a cost judgement implicitly, and that making it explicit—as airlines do when they set scheduled block times with reference to on-time targets [3, 12]—would allow the choice of allowance to be debated on operational grounds rather than inherited.

The same framing clarifies why a uniform increase of the standard would be a poor remedy. At the fastest aerodromes the two-hour value already covers about three quarters of stops, so raising it everywhere would add idle time where none is needed. The problem is not the level of the standard but its uniformity.

6.2 Heterogeneity in routine service and in extended events

The aerodrome-level evidence identifies two distinct sources of heterogeneity. The first concerns routine service. At AD-02, AD-03 and AD-11 both robust measures exceed two hours, the finding survives family-wise error control and partial pooling, and the site models suggest that only 25–35% of stops are completed within the standard. An excess of 20–30 minutes per stop may appear modest, but it accumulates along multi-stop sequences and is precisely the kind of recurrent, predictable shortfall that a planning buffer should absorb [1]. The second concerns extended events, whose frequency ranges from 1% to 12% across aerodromes and is significantly heterogeneous. AD-02 is exposed on both dimensions, whereas AD-03 has a longer routine service but few extended events. These profiles call for different responses: routine excess is addressed by a larger allowance in the plan, while extended events are better managed by contingency arrangements—the availability of a relief crew, pre-coordinated overnight options or a review of the causes of long ground events—than by padding every stop [57].

The partial-pooling results reinforce the case for aerodrome-specific planning. Between 79% and 91% of the variation in aerodrome medians, depending on the specification, reflects genuine differences rather than sampling noise, and a new aerodrome could plausibly have a median anywhere between 1.3 and 2.7 h. The fleet-wide value is therefore an uncertain guide for any particular aerodrome, and particularly so for aerodromes with little operational history.

6.3 What recorded ground times measure

Where a planning allowance governs departure times, observed durations can partly reflect the allowance itself: if aircraft that are ready early wait for their planned departure, work appears to expand to fill the time available and the time actually required is overstated [87]. The KC-390 records show no such anchoring. Recorded ground times do not concentrate at, or just after, the two-hour value, so the observed intervals can be read as the durations of the ground processes, including taxiing and short waits, rather than as an echo of the standard. This matters for interpretation: the excess over two hours at AD-02, AD-03 and AD-11 reflects how long stops at those aerodromes actually take, and the shorter ground times elsewhere are not compressed toward the standard. If localized allowances are adopted, the same check should be repeated, because allowances could begin to shape recorded departure times; recording ready-for-departure times would separate process time from planned waiting directly.

6.4 Implications for aircraft operations and mission planning

The results support a tiered policy that retains the administrative simplicity of a default. The two-hour value is retained as the fleet-wide default. At AD-02, AD-03 and AD-11 it is replaced by the suggested values of 2.39–2.47 h for routine planning, with the conservative values of 2.53–2.83 h reserved for high-consequence missions, such as sequences with tight downstream tasking, limited recovery options or no relief crew available. AD-10 and AD-12 are placed under monitoring: their excess is either small (6–8 minutes at AD-10) or uncertain, and the monitoring rule states in advance how much additional evidence would justify a change, which protects the process from reacting to noise. The mission-level simulation indicates what the policy buys: localized allowances reduce the probability of a cumulative overrun above 30 minutes by 28 percentage points on a route through two adjust-tier aerodromes and by 6–8 points on mixed routes, at the cost of less than one hour of additional planned ground time per mission. Localized allowances also restore parity: at the adjust-tier aerodromes they deliver a service level of about 53%, close to the 57% that the standard delivers fleet-wide.

The distinction between routine shortfalls and extended events also helps interpret ground-support interventions. The location-dependent equipment analysis of Tucci and Facchini [6] motivates collecting resource and readiness information at the slower aerodromes; the present duration data alone cannot identify which equipment investment would reduce ground time. Likewise, robust apron scheduling [32] could complement revised allowances where the handling tasks and resource constraints can be observed.

Three further implications follow. First, aerodrome-specific allowances belong in the planning tool used to assemble missions, so that route feasibility and crew composition are assessed against a timeline that reflects local conditions; where the localized timeline approaches a duty limit, the decision to assign an augmented crew rests on evidence rather than on undifferentiated uncertainty, and where it does not, leaner crews preserve payload capacity. Second, the allowances double as a performance indicator: persistent reliance on conservative values at an aerodrome signals a need to

review ground-support equipment, refueling arrangements or cargo readiness, in the spirit of turnaround monitoring frameworks [19]. Third, the workflow is transferable. Regional, cargo and humanitarian operators that apply uniform minimum turnaround times across stations with very different handling resources face the same problem, and the method requires only landing and takeoff times, or their on- and off-block equivalents. Because the extended analyses rely only on aggregate statistics, the workflow also suits operators whose leg-level data are commercially sensitive.

6.5 Limitations and future research

Several limitations qualify these conclusions. The landing-to-takeoff interval includes taxi time and is recorded on a five-minute grid; it is the interval governed by the standard, but it is not a pure measure of ground-service time. No explanatory variables were modeled: cargo mass and configuration, passenger numbers, fuel uplift, mission type, reconfiguration requirements, time of day, weather and ground-support availability are likely drivers of ground time, and their omission limits the explanation of why AD-02, AD-03 and AD-11 are slower. Aerodromes are anonymized, which protects operational information but prevents readers from relating the results to infrastructure characteristics. The aggregate layer relies on approximations: standard errors recovered from Wald intervals, a mixture fitted to half-hour classes, site models calibrated to two moments and, in the simulation, independent ground times across the stops of a mission. The evaluation covers only aerodromes with more than 30 events, so rarely used aerodromes—where the default is least informed—are not assessed, and the two-year window may not capture changes in infrastructure, procedures or mission mix over longer periods. Finally, the observed distribution was generated under the current standard; although recorded times show no anchoring to it, the behavior of crews and ground staff under different allowances cannot be observed in these data.

Synthetic operational data, as investigated by Murad and Ruocco [31], offer a possible extension when a data owner can validate generated records against protected observations. Such validation would be needed before substituting synthetic data for leg-level inference; the present study instead makes its aggregate inputs explicit.

Future research should model the drivers of ground time with leg-level covariates, using quantile regression to target the upper quantiles that matter for buffers [97] and hierarchical models to borrow strength across aerodromes; elicit the cost ratio c_u/c_o from planners so that service-level targets can be set explicitly; and embed localized allowances in a discrete-event simulation of mission networks with crew duty rules to quantify effects on crew allocation, payload and mission completion. The collection of on-block, off-block and ready-for-departure times would allow the separation of process time from planned waiting and would open the way to data-driven turnaround prediction of the kind now used at commercial hubs [22].

7 Conclusions

This study tested the KC-390 two-hour ground-time standard against 1,679 ground events reconstructed from two years of Brazilian Air Force mission reports. Three conclusions emerge. First, the standard is best understood as an implicit service level:

it is met by 57% of stops, which presumes that an hour of overrun costs little more than an hour of idle aircraft time. Second, a uniform value conceals marked heterogeneity: at three aerodromes the standard is systematically too short on two robust measures, a finding that survives multiplicity control and partial pooling, while the frequency of extended ground events varies more than tenfold across aerodromes. Recorded ground times show no anchoring to the standard, so these differences reflect the ground processes themselves. Third, a tiered policy that retains the default, adjusts it where the evidence is robust and monitors where it is suggestive reduces overrun risk in the affected mission sequences at a modest cost in planned ground time. More broadly, the study shows how an operator can audit a planning standard with routinely collected, coarse and restricted operational data, and how the language of service levels can turn an inherited rule into an explicit, evidence-based choice.

Statements and Declarations

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

Author contributions. **Ramiro de Paiva Fagundes:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **Lucas Orbolato Carvalho:** Methodology, Formal analysis, Software, Visualization, Writing – review & editing. **Daniel Alberto Pamplona:** Conceptualization, Supervision, Validation. **Níssia Carvalho Rosa Bergiante:** Conceptualization, Supervision, Validation.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Ramiro de Paiva Fagundes, Lucas Orbolato Carvalho and Daniel Alberto Pamplona are officers of the Brazilian Air Force, which issues the planning standard evaluated in this study.

Data availability. The leg-level mission-report records are restricted operational data of the Brazilian Air Force and cannot be shared. All aggregate statistics used as inputs to the extended analyses are reported in the article.

A Mission-report variables

Table 11 Main variables of the mission reports and their use in this study

Variable	Description	Used
Landing date and time	Date and time of landing of the leg	Yes (ground-time construction)
Takeoff date and time	Date and time of takeoff of the leg	Yes (ground-time construction)
Departure aerodrome	ICAO code of the departure aerodrome (anonymized)	Yes (attribution to aerodrome)
Arrival aerodrome	ICAO code of the arrival aerodrome (anonymized)	Consistency check only
Mission-report number	Identifier used to group the legs of a mission	Yes (pairing of legs)
Model	Aircraft type	Yes (KC-390 filter)
Completion	Mission accomplished totally, partially or not at all	Yes (total or partial only)
Status	Status of the report	No
Issuer	Air unit issuing the report	No
Application	Mission operation or application	No
Registration	Aircraft registration	No
Passengers	Passengers transported on the leg	No
Cargo mass; cargo volume	Cargo transported on the leg (kg; m^3)	No
Fuel; lubricant	Fuel (kg) and lubricant (L) consumed on the leg	No
Flight time	Flight time of the leg (h)	No
Order number	Operational order referenced by the report	No
Commander's report; mission report	Free text, replicated across all legs of a mission	Exploratory search only (not attributable to a leg)

B Site models: calibration and validation

The site models are defined in Section 4.6. Because the mean of a truncated log-normal distribution is not monotone in σ for a given median, the calibration equations can have two solutions, which occurred at AD-02, AD-03, AD-04 and AD-11. We retain the solution with the smaller σ, which at these four aerodromes places at most 10.5% of the routine component above the trimming threshold, whereas the alternative solutions place 13–41% of it above 4.5 h, which is inconsistent with the regime separation.

Independent evidence supports this choice. For the five aerodromes with model-based intervals for the trimmed mean, the standard deviation implied by those intervals ($\widehat{\mathrm{se}}\sqrt{n}$) can be compared with that of the calibrated routine component. For AD-02, AD-03 and AD-11 the retained solutions imply 0.65, 0.82 and 0.61 h, against 0.72, 0.79 and 0.71 h from the model-based intervals, whereas the rejected solutions imply 1.10, 0.93 and 1.07 h. For AD-12 the site model implies 1.08 h against 1.13 h; for

AD-10 it implies 0.52 h against 0.85 h, so the reliability curve of AD-10 is steeper than the data suggest. At AD-09 and AD-12 no exact solution exists—the trimmed mean exceeds the trimmed median by more than any truncated log-normal distribution allows, which at AD-12 is consistent with the bimodal pattern visible in Fig. 9(e)—and the closest solution was used, with residual discrepancies of 0.06 and 0.27 h, respectively; these solutions also place more of the routine component above 4.5 h (26% and 21%).

As an out-of-sample check, the medians of the full distributions implied by the site models, which are not used in calibration, reproduce the reported medians with a mean absolute error of 0.04 h (maximum 0.14 h, at AD-02). Standard errors of the median derived from the site densities are within 35% of the parametric ones for the four aerodromes where both are available (AD-02: 0.08 against 0.10 h; AD-03: 0.10 against 0.10 h; AD-10: 0.10 against 0.15 h; AD-11: 0.14 against 0.17 h), which motivates the sensitivity analyses reported in Section 5.4. Table 12 lists the parameters of all site models together with these checks.

Table 12 Parameters and validation of the two-regime site models

	Routine component						Median (h)	
Aerodrome	μ	σ	SD (h)	Mass above 4.5 h (%)	Extended share (%)	Calibration residual (h)	Reported	Site model
AD-01	0.605	0.298	0.57	0.1	5.3	0.00	1.92	1.87
AD-02	0.814	0.287	0.65	0.8	11.7	0.00	2.50	2.36
AD-03	0.855	0.402	0.82	5.3	2.2	0.00	2.33	2.32
AD-04	0.690	0.649	0.99	10.5	4.2	0.00	1.83	1.89
AD-05	0.605	0.319	0.61	0.2	3.6	0.00	1.83	1.86
AD-06	0.306	0.618	0.86	2.6	1.0	0.00	1.33	1.34
AD-07	0.335	0.728	0.95	5.4	1.5	0.00	1.33	1.35
AD-08	0.503	0.502	0.83	2.3	4.9	0.00	1.67	1.68
AD-09	0.876	0.964	1.12	25.7	1.8	0.06	1.79	1.78
AD-10	0.733	0.241	0.52	0.1	6.4	0.00	2.17	2.12
AD-11	0.848	0.260	0.61	0.6	9.8	0.00	2.42	2.41
AD-12	0.833	0.844	1.08	21.3	2.5	0.27	1.88	1.87
AD-13	0.515	0.372	0.66	0.4	3.1	0.00	1.67	1.70
AD-14	0.425	0.760	0.99	7.8	9.4	0.00	1.67	1.56

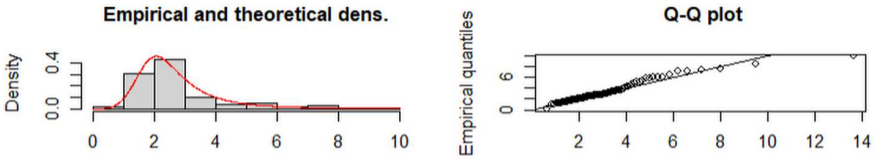
Notes: μ and σ are the log-scale parameters of the routine log-normal component before truncation at 4.5 h; SD is the standard deviation of the truncated component. Calibration residual: largest absolute difference between the fitted and reported trimmed median and trimmed mean. The site-model median is not used in calibration.

C Goodness-of-fit diagnostics of the aerodrome-level models

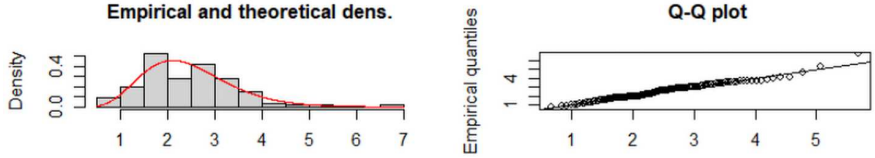
Figs. 8 and 9 show, for each aerodrome and measure in Table 6, the empirical and fitted densities and the Q–Q plot of the selected distribution. Ground times are in hours; in the Q–Q plots, empirical quantiles are plotted against the quantiles of the

fitted distribution. The fitted distributions reproduce the body of the data closely at all aerodromes. The largest departures occur in the upper tail of the full samples used for the median analysis, where the extended events lie (Fig. 8), and at AD-12, whose trimmed sample has a secondary concentration of events between 4 and 4.5 h that no unimodal distribution captures (Fig. 9e).

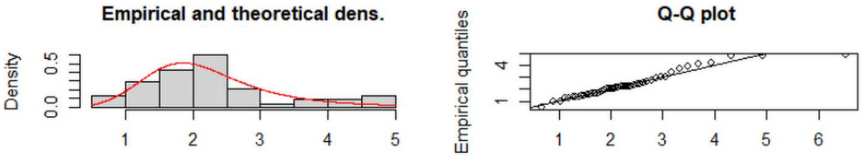
(a) AD-02: Burr distribution, full sample (n = 145)



(b) AD-03: gamma distribution, full sample (n = 135)



(c) AD-10: log-logistic distribution, full sample (n = 47)



(d) AD-11: log-logistic distribution, full sample (n = 41)

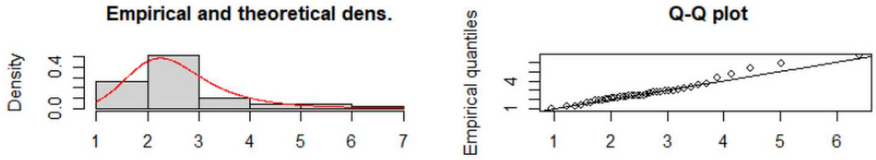
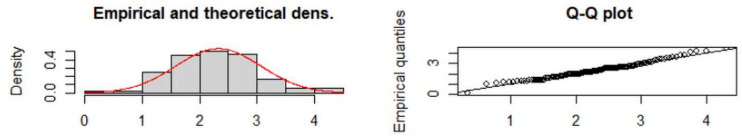
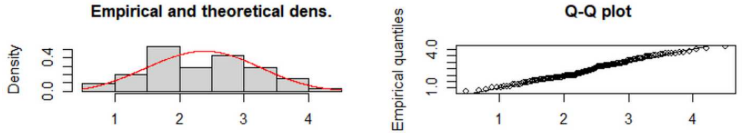


Figure 8 Goodness-of-fit diagnostics of the distributions selected for the median analysis (full samples, ≤10 h): empirical and fitted densities (left) and Q–Q plots (right)

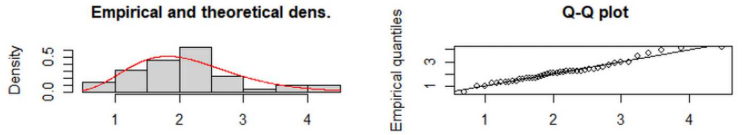
(a) AD-02: normal distribution, trimmed sample (n = 128)



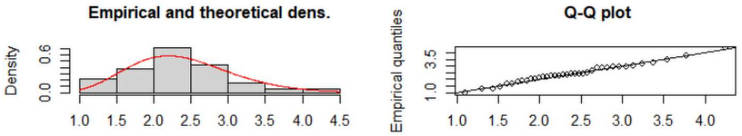
(b) AD-03: Weibull distribution, trimmed sample (n = 132)



(c) AD-10: gamma distribution, trimmed sample (n = 44)



(d) AD-11: gamma distribution, trimmed sample (n = 37)



(e) AD-12: inverse Gaussian distribution, trimmed sample (n = 39)

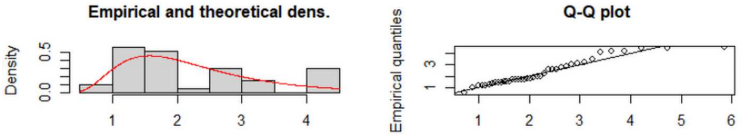


Figure 9 Goodness-of-fit diagnostics of the distributions selected for the trimmed-mean analysis (trimmed samples, $\leq$4.5 h): empirical and fitted densities (left) and Q–Q plots (right)

References

[1] Wu, C.-L.: Inherent delays and operational reliability of airline schedules. Journal of Air Transport Management **11**(4), 273–282 (2005) https://doi.org/10.1016/j.jairtraman.2005.01.005

[2] AhmadBeygi, S., Cohn, A., Guan, Y., Belobaba, P.: Analysis of the potential for delay propagation in passenger airline networks. Journal of Air Transport Management **14**(5), 221–236 (2008) https://doi.org/10.1016/j.jairtraman.2008.04.010

[3] Hao, L., Hansen, M.: Block time reliability and scheduled block time setting. Transportation Research Part B: Methodological **69**, 98–111 (2014) https://doi.org/10.1016/j.trb.2014.08.008

[4] Department of the Air Force: Air force pamphlet 10-1403: Air mobility planning factors. Technical report, United States Air Force, Washington, DC (2018). 24 October 2018

[5] U.S. Government Accountability Office: Defense transportation: Air Mobility Command needs to collect and analyze better data to assess aircraft utilization. Technical Report GAO-05-819, U.S. Government Accountability Office, Washington, DC (2005)

[6] Tucci, H.N.P., Facchini, F.: Ground support equipment and flight delays: a multivariate analysis in different locations. CEAS Aeronautical Journal **17**, 869–887 (2026) https://doi.org/10.1007/s13272-025-00891-6

[7] Berling, J., Lau, A., Gollnick, V.: Constraint reconciliation optimization to minimize primary and reactionary delay in a research network strategic tool. CEAS Aeronautical Journal **15**, 1015–1030 (2024) https://doi.org/10.1007/s13272-023-00708-4

[8] Brazilian Air Force: NOPREP/SGV/01E: Coletânea de normas do Comando de Preparo sobre segurança de voo e do trabalho [Compilation of Preparation Command standards on flight and occupational safety]. Technical report, Comando de Preparo, Comando da Aeronáutica, Brasília (2023). Reissue approved by Portaria COMPREP No. 2.340/SPOG-50, 5 November 2024. In Portuguese

[9] Embraer: Embraer delivers the sixth C-390 Millennium aircraft to the Brazilian Air Force. Press release, 16 June 2023. https://www.embraer.com/media-center/en/?mediatype=NEWS&detail=14139. Accessed 30 September 2026 (2023)

[10] Oreschko, B., Kunze, T., Schultz, M., Fricke, H., Kumar, V., Sherry, L.: Turnaround prediction with stochastic process times and airport specific delay pattern. In: Proceedings of the 5th International Conference on Research in Air Transportation (ICRAT), Berkeley, CA (2012)

[11] Schmidt, M.: A review of aircraft turnaround operations and simulations. Progress in Aerospace Sciences **92**, 25–38 (2017) https://doi.org/10.1016/j.paerosci.2017.05.002

[12] Deshpande, V., Arıkan, M.: The impact of airline flight schedules on flight delays. Manufacturing & Service Operations Management **14**(3), 423–440 (2012) https://doi.org/10.1287/msom.1120.0379

[13] Wu, C.-L., Caves, R.E.: Modelling and optimization of aircraft turnaround time at an airport. Transportation Planning and Technology **27**(1), 47–66 (2004) https://doi.org/10.1080/0308106042000184454

[14] Norin, A., Granberg, T.A., Yuan, D., Värbrand, P.: Airport logistics – a case study of the turn-around process. Journal of Air Transport Management **20**, 31–34 (2012) https://doi.org/10.1016/j.jairtraman.2011.10.008

[15] Wu, C.-L., Caves, R.E.: Aircraft operational costs and turnaround efficiency at airports. Journal of Air Transport Management **6**(4), 201–208 (2000) https://doi.org/10.1016/S0969-6997(00)00014-4

[16] Wu, C.-L., Caves, R.E.: Modelling of aircraft rotation in a multiple airport environment. Transportation Research Part E: Logistics and Transportation Review **38**(3–4), 265–277 (2002) https://doi.org/10.1016/S1366-5545(02)00010-8

[17] Schultz, M.: Implementation and application of a stochastic aircraft boarding model. Transportation Research Part C: Emerging Technologies **90**, 334–349 (2018) https://doi.org/10.1016/j.trc.2018.03.016

[18] Fricke, H., Schultz, M.: Delay impacts onto turnaround performance: Optimal time buffering for minimizing delay propagation. In: Proceedings of the 8th USA/Europe Air Traffic Management Research and Development Seminar (ATM2009), Napa, CA (2009)

[19] Wu, C.-L.: Monitoring aircraft turnaround operations – framework development, application and implications for airline operations. Transportation Planning and Technology **31**(2), 215–228 (2008) https://doi.org/10.1080/03081060801948233

[20] Okwir, S., Ulfvengren, P., Angelis, J., Ruiz, F., Núñez Guerrero, Y.M.: Managing turnaround performance through collaborative decision making. Journal of Air Transport Management **58**, 183–196 (2017) https://doi.org/10.1016/j.jairtraman.2016.10.008

[21] Rott, J., König, F., Häfke, H., Schmidt, M., Böhm, M., Kratsch, W., Krcmar, H.: Process mining for resilient airport operations: A case study of Munich Airport's turnaround process. Journal of Air Transport Management **112**, 102451 (2023) https://doi.org/10.1016/j.jairtraman.2023.102451

[22] Tang, X., Wu, J., Wu, C.-L., Ding, Y., Zhang, S.: Dynamic prediction of aircraft turnaround milestone times using a cascaded gradient boosting model for improved airport collaborative decision-making. Journal of Air Transport Management **128**, 102842 (2025) https://doi.org/10.1016/j.jairtraman.2025.102842

[23] Tang, X., Zhu, W., Wu, C.-L., Zhang, S., Witlox, F.: A method for predicting flight turnaround key milestone events oriented to collaborative optimisation: Integrating effective remaining buffer time and a parallel-cascaded mechanism. Journal of Air Transport Management **135**, 103022 (2026) https://doi.org/10.1016/j.jairtraman.2026.103022

[24] Malandri, C., Mantecchini, L., Reis, V.: Aircraft turnaround and industrial actions: How ground handlers' strikes affect airport airside operational efficiency. Journal of Air Transport Management **78**, 23–32 (2019) https://doi.org/10.1016/j.jairtraman.2019.04.007

[25] Mota, M.M., Boosten, G., De Bock, N., Jimenez, E., Sousa, J.P.: Simulation-based turnaround evaluation for Lelystad Airport. Journal of Air Transport Management **64**, 21–32 (2017) https://doi.org/10.1016/j.jairtraman.2017.06.021

[26] Padrón, S., Guimarans, D., Ramos, J.J., Fitouri-Trabelsi, S.: A bi-objective approach for scheduling ground-handling vehicles in airports. Computers & Operations Research **71**, 34–53 (2016) https://doi.org/10.1016/j.cor.2015.12.010

[27] Guimarans, D., Padrón, S.: A stochastic approach for planning airport ground support resources. International Transactions in Operational Research **29**(6), 3316–3345 (2022) https://doi.org/10.1111/itor.13104

[28] Alonso Tabares, D., Mora-Camino, F., Drouin, A.: A multi-time scale management structure for airport ground handling automation. Journal of Air Transport Management **90**, 101959 (2021) https://doi.org/10.1016/j.jairtraman.2020.101959

[29] Schultz, M., Evler, J., Asadi, E., Preis, H., Fricke, H., Wu, C.-L.: Future aircraft turnaround operations considering post-pandemic requirements. Journal of Air Transport Management **89**, 101886 (2020) https://doi.org/10.1016/j.jairtraman.2020.101886

[30] Park, D.K., Kim, J.K.: Influential factors to aircraft taxi time in airport. Journal of Air Transport Management **106**, 102321 (2023) https://doi.org/10.1016/j.jairtraman.2022.102321

[31] Murad, A., Ruocco, M.: Pre-tactical flight-delay and turnaround forecasting with synthetic aviation data. CEAS Aeronautical Journal (2026) https://doi.org/10.1007/s13272-026-00941-7 . Published online 3 February 2026

[32] Gök, Y.S., Padrón, S., Tomasella, M., Guimarans, D., Ozturk, C.: Constraint-based robust planning and scheduling of airport apron operations through simheuristics. Annals of Operations Research **320**, 795–830 (2023) https://doi.org/10.1007/s10479-022-04547-0

[33] Lan, S., Clarke, J.-P., Barnhart, C.: Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions. Transportation Science **40**(1), 15–28 (2006) https://doi.org/10.1287/trsc.1050.0134

[34] Dunbar, M., Froyland, G., Wu, C.-L.: Robust airline schedule planning: Minimizing propagated delay in an integrated routing and crewing framework. Transportation Science **46**(2), 204–216 (2012) https://doi.org/10.1287/trsc.1110.0395

[35] Schilt, I.M., Kalker, J., Lefter, I., Kwakkel, J.H., Verbraeck, A.: Buffer scheduling for improving on-time performance and connectivity with a multi-objective simulation–optimization model: A proof of concept for the airline industry. Journal of Air Transport Management **115**, 102547 (2024) https://doi.org/10.1016/j.jairtraman.2024.102547

[36] Kang, L., Hansen, M.: Behavioral analysis of airline scheduled block time adjustment. Transportation Research Part E: Logistics and Transportation Review **103**, 56–68 (2017) https://doi.org/10.1016/j.tre.2017.04.004

[37] Yimga, J., Gorjidooz, J.: Airline schedule padding and consumer choice behavior. Journal of Air Transport Management **78**, 71–79 (2019) https://doi.org/10.1016/j.jairtraman.2019.05.001

[38] Wang, Y., Zhou, Y., Hansen, M., Chin, C.: Scheduled block time setting and on-time performance of U.S. and Chinese airlines – a comparative analysis. Transportation Research Part A: Policy and Practice **130**, 825–843 (2019) https://doi.org/10.1016/j.tra.2019.09.043

[39] Yimga, J.: Flight schedules under partnership: The effects of capacity purchase agreements on airline schedule buffers. Journal of Air Transport Management **129**, 102876 (2025) https://doi.org/10.1016/j.jairtraman.2025.102876

[40] Lonzius, M.C., Lange, A.: Robust scheduling: An empirical study of its impact on air traffic delays. Transportation Research Part E: Logistics and Transportation Review **100**, 98–114 (2017) https://doi.org/10.1016/j.tre.2016.12.012

[41] Arrow, K.J., Harris, T., Marschak, J.: Optimal inventory policy. Econometrica **19**(3), 250–272 (1951) https://doi.org/10.2307/1906813

[42] Qin, Y., Wang, R., Vakharia, A.J., Chen, Y., Seref, M.M.H.: The newsvendor

problem: Review and directions for future research. European Journal of Operational Research **213**(2), 361–374 (2011) https://doi.org/10.1016/j.ejor.2010.11.024

[43] Pyrgiotis, N., Malone, K.M., Odoni, A.: Modelling delay propagation within an airport network. Transportation Research Part C: Emerging Technologies **27**, 60–75 (2013) https://doi.org/10.1016/j.trc.2011.05.017

[44] Fleurquin, P., Ramasco, J.J., Eguíluz, V.M.: Systemic delay propagation in the US airport network. Scientific Reports **3**, 1159 (2013) https://doi.org/10.1038/srep01159

[45] Kafle, N., Zou, B.: Modeling flight delay propagation: A new analytical-econometric approach. Transportation Research Part B: Methodological **93**, 520–542 (2016) https://doi.org/10.1016/j.trb.2016.08.012

[46] Wu, C.-L., Law, K.: Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model. Transportation Research Part E: Logistics and Transportation Review **122**, 62–77 (2019) https://doi.org/10.1016/j.tre.2018.11.004

[47] Kim, M., Park, S.: Airport and route classification by modelling flight delay propagation. Journal of Air Transport Management **93**, 102045 (2021) https://doi.org/10.1016/j.jairtraman.2021.102045

[48] Erdem, F., Bilgiç, T.: Airline delay propagation: Estimation and modeling in daily operations. Journal of Air Transport Management **115**, 102548 (2024) https://doi.org/10.1016/j.jairtraman.2024.102548

[49] Kohl, N., Larsen, A., Larsen, J., Ross, A., Tiourine, S.: Airline disruption management – perspectives, experiences and outlook. Journal of Air Transport Management **13**(3), 149–162 (2007) https://doi.org/10.1016/j.jairtraman.2007.01.001

[50] Evler, J., Asadi, E., Preis, H., Fricke, H.: Airline ground operations: Optimal schedule recovery with uncertain arrival times. Journal of Air Transport Management **92**, 102021 (2021) https://doi.org/10.1016/j.jairtraman.2021.102021

[51] Evler, J., Lindner, M., Fricke, H., Schultz, M.: Integration of turnaround and aircraft recovery to mitigate delay propagation in airline networks. Computers & Operations Research **138**, 105602 (2022) https://doi.org/10.1016/j.cor.2021.105602

[52] Schaefer, A.J., Johnson, E.L., Kleywegt, A.J., Nemhauser, G.L.: Airline crew scheduling under uncertainty. Transportation Science **39**(3), 340–348 (2005) https://doi.org/10.1287/trsc.1040.0091

[53] Shebalov, S., Klabjan, D.: Robust airline crew pairing: Move-up crews. Transportation Science **40**(3), 300–312 (2006) https://doi.org/10.1287/trsc.1050.0131

[54] Caldwell, J.A.: Fatigue in aviation. Travel Medicine and Infectious Disease **3**(2), 85–96 (2005) https://doi.org/10.1016/j.tmaid.2004.07.008

[55] Tu, Y., Ball, M.O., Jank, W.S.: Estimating flight departure delay distributions – a statistical approach with long-term trend and short-term pattern. Journal of the American Statistical Association **103**(481), 112–125 (2008) https://doi.org/10.1198/016214507000000257

[56] Pérez-Rodríguez, J.V., Pérez-Sánchez, J.M., Gómez-Déniz, E.: Modelling the asymmetric probabilistic delay of aircraft arrival. Journal of Air Transport Management **62**, 90–98 (2017) https://doi.org/10.1016/j.jairtraman.2017.03.001

[57] Zhao, Y.: Decomposing flight delay distributions: Tail, location, and center effects. Journal of Air Transport Management **138**, 103112 (2027) https://doi.org/10.1016/j.jairtraman.2026.103112

[58] Carvalho, L., Sternberg, A., Maia Gonçalves, L., Beatriz Cruz, A., Soares, J.A., Brandão, D., Carvalho, D., Ogasawara, E.: On the relevance of data science for flight delay research: A systematic review. Transport Reviews **41**(4), 499–528 (2021) https://doi.org/10.1080/01441647.2020.1861123

[59] Morton, D.P., Rosenthal, R.E., Weng, L.T.: Optimization modeling for airlift mobility. Military Operations Research **1**(4), 49–67 (1996) https://doi.org/10.5711/morj.1.4.49

[60] Rosenthal, R.E., Baker, S.F., Weng, L.T., Fuller, D.F., Goggins, D., Toy, A.O., Turker, Y., Horton, D.: Application and extension of the Thruput II optimization model for airlift mobility. Military Operations Research **3**(2), 55–74 (1997) https://doi.org/10.5711/morj.3.2.55

[61] Baker, S.F., Morton, D.P., Rosenthal, R.E., Williams, L.M.: Optimizing military airlift. Operations Research **50**(4), 582–602 (2002) https://doi.org/10.1287/opre.50.4.582.2864

[62] Wu, T.T., Powell, W.B., Whisman, A.: The optimizing-simulator: An illustration using the military airlift problem. ACM Transactions on Modeling and Computer Simulation **19**(3), 1–31 (2009) https://doi.org/10.1145/1540530.1540535

[63] Brown, G.G., Carlyle, W.M., Dell, R.F.: Optimizing intratheater military airlift in Iraq and Afghanistan. Military Operations Research **18**(3), 35–52 (2013) https://doi.org/10.5711/1082598318335

[64] Dietz, D.C., Jenkins, R.C.: Analysis of aircraft sortie generation with

the use of a fork-join queueing network model. Naval Research Logistics **44**(2), 153–164 (1997) https://doi.org/10.1002/(SICI)1520-6750(199703)44:2<153::AID-NAV1>3.0.CO;2-8

[65] Dietz, D.C.: Mean value analysis of military airlift operations at an individual airfield. Journal of Aircraft **36**(5), 750–755 (1999) https://doi.org/10.2514/2.2530

[66] Willits, C.J., Dietz, D.C.: Nested fork-join queueing network model for analysis of airfield operations. Journal of Aircraft **38**(5), 848–855 (2001) https://doi.org/10.2514/2.2870

[67] Stucker, J.P., Berg, R.T.: Understanding airfield capacity for airlift operations. Technical Report MR-700-AF/OSD, RAND Corporation, Santa Monica, CA (1998)

[68] Van Wassenhove, L.N.: Humanitarian aid logistics: Supply chain management in high gear. Journal of the Operational Research Society **57**(5), 475–489 (2006) https://doi.org/10.1057/palgrave.jors.2602125

[69] Choi, S., Hanaoka, S.: Estimating the mean waiting time in airports through cooperative disaster response operations. Journal of Air Transport Management **65**, 11–17 (2017) https://doi.org/10.1016/j.jairtraman.2017.05.010

[70] Choi, S., Hanaoka, S.: Prediction of aircraft waiting time at airport during immediate response to disaster. Aerospace **6**(4), 40 (2019) https://doi.org/10.3390/aerospace6040040

[71] Polater, A.: Airports' role as logistics centers in humanitarian supply chains: A surge capacity management perspective. Journal of Air Transport Management **83**, 101765 (2020) https://doi.org/10.1016/j.jairtraman.2020.101765

[72] Hale, M.A., Reiman, A.D., Maywald, J.D., Glassburner, A.V.: A capability-cost replacement model for military transport aircraft. Journal of Air Transport Management **137**, 103069 (2026) https://doi.org/10.1016/j.jairtraman.2026.103069

[73] Gaitan, I., Yun, G., Joo, H., Hwang, S.: Understanding military pilots' fuel-saving intentions for supporting logistics missions. Journal of Air Transport Management **121**, 102689 (2024) https://doi.org/10.1016/j.jairtraman.2024.102689

[74] Tukey, J.W., McLaughlin, D.H.: Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. Sankhyā: The Indian Journal of Statistics, Series A **25**(3), 331–352 (1963)

[75] Huber, P.J., Ronchetti, E.M.: Robust Statistics, 2nd edn. Wiley, Hoboken, NJ (2009). https://doi.org/10.1002/9780470434697

[76] Wilcox, R.R.: Introduction to Robust Estimation and Hypothesis Testing, 4th edn. Academic Press, London (2017)

[77] Bednar, J.B., Watt, T.L.: Alpha-trimmed means and their relationship to median filters. IEEE Transactions on Acoustics, Speech, and Signal Processing **32**(1), 145–153 (1984) https://doi.org/10.1109/TASSP.1984.1164279

[78] Gersztenkorn, A., Scales, J.A.: Smoothing seismic tomograms with alpha-trimmed means. Geophysical Journal International **92**(1), 67–72 (1988) https://doi.org/10.1111/j.1365-246X.1988.tb01121.x

[79] Marazzi, A., Ruffieux, C.: The truncated mean of an asymmetric distribution. Computational Statistics & Data Analysis **32**(1), 79–100 (1999) https://doi.org/10.1016/S0167-9473(99)00018-3

[80] Stephens, M.A.: EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association **69**(347), 730–737 (1974) https://doi.org/10.1080/01621459.1974.10480196

[81] Burnham, K.P., Anderson, D.R.: Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods & Research **33**(2), 261–304 (2004) https://doi.org/10.1177/0049124104268644

[82] Delignette-Muller, M.L., Dutang, C.: fitdistrplus: An R package for fitting distributions. Journal of Statistical Software **64**(4), 1–34 (2015) https://doi.org/10.18637/jss.v064.i04

[83] Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall, New York (1993). https://doi.org/10.1007/978-1-4899-4541-9

[84] Turnbull, B.W.: The empirical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society: Series B (Methodological) **38**(3), 290–295 (1976) https://doi.org/10.1111/j.2517-6161.1976.tb01597.x

[85] Heitjan, D.F.: Inference from grouped continuous data: A review. Statistical Science **4**(2), 164–179 (1989) https://doi.org/10.1214/ss/1177012601

[86] Heitjan, D.F., Rubin, D.B.: Inference from coarse data via multiple imputation with application to age heaping. Journal of the American Statistical Association **85**(410), 304–314 (1990) https://doi.org/10.1080/01621459.1990.10476202

[87] Hasija, S., Pinker, E., Shumsky, R.A.: Work expands to fill the time available: Capacity estimation and staffing under Parkinson's law. Manufacturing & Service Operations Management **12**(1), 1–18 (2010) https://doi.org/10.1287/msom.1080.0250

[88] Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics **6**(2), 65–70 (1979)

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

[90] Efron, B., Morris, C.: Data analysis using Stein's estimator and its generalizations. Journal of the American Statistical Association **70**(350), 311–319 (1975) https://doi.org/10.1080/01621459.1975.10479864

[91] DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. Controlled Clinical Trials **7**(3), 177–188 (1986) https://doi.org/10.1016/0197-2456(86)90046-2

[92] Higgins, J.P.T., Thompson, S.G.: Quantifying heterogeneity in a meta-analysis. Statistics in Medicine **21**(11), 1539–1558 (2002) https://doi.org/10.1002/sim.1186

[93] Brazilian Air Force: DCA 1-1: Doutrina básica da Força Aérea Brasileira [Basic doctrine of the Brazilian Air Force]. Technical report, Ministério da Defesa, Comando da Aeronáutica, Brasília (2024). Approved by Portaria GABAER/GC3 No. 1.563, 23 December 2024. In Portuguese

[94] Embraer: KC-390 Millennium. https://embraer.com/defense-security-air/kc-390-millennium/en/. Accessed 30 September 2026 (2026)

[95] Clopper, C.J., Pearson, E.S.: The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika **26**(4), 404–413 (1934) https://doi.org/10.1093/biomet/26.4.404

[96] Hyndman, R.J., Fan, Y.: Sample quantiles in statistical packages. The American Statistician **50**(4), 361–365 (1996) https://doi.org/10.1080/00031305.1996.10473566

[97] Koenker, R., Bassett, G.: Regression quantiles. Econometrica **46**(1), 33–50 (1978) https://doi.org/10.2307/1913643