

Article

Demand Forecasting and Inventory Optimization of Aviation Rotable Parts: A Markov Queuing Approach

Guannan Chen ¹, Yue Teng ^{2,*} , Yangyang Zhang ³  and Zhenxing Gao ³ ¹ China Eastern Airlines Holding Co., Ltd., Shanghai 201100, China; chenguanna@ceair.com² China Eastern Technology Application Research and Development Center Co., Ltd., Shanghai 201707, China³ College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; zhang_yy@nuaa.edu.cn (Y.Z.); z.x.gao@nuaa.edu.cn (Z.G.)

* Correspondence: tengyue1@ceair.com

Abstract

High-value aviation rotatable parts require accurate demand forecasting because each failure simultaneously creates a replacement demand, a repair workload, and a potential aircraft on-ground (AOG) risk. This study develops a continuous-time Markov chain (CTMC) queuing framework for forecasting demand and optimizing target stock levels under realistic maintenance, repair, and overhaul (MRO) capacity constraints. Historical service-engineering records from B737 mechanical engine control (MEC) units are first benchmarked and filtered to retain inherent random failures during the useful-life phase. A Kolmogorov-Smirnov goodness-of-fit test is then used to verify the exponential time-between-failure assumption required by the Markov model. Based on this validated failure-pattern classification, the study derives steady-state probability models for both an ideal $M/M/\infty$ repair system and a finite-capacity $M/M/c$ repair system. The main contribution is not the isolated use of Weibull or exponential reliability models, Markov chains, or inventory optimization, which are established methods, but their auditable integration into an airline rotatable-parts workflow that links failure-pattern screening, finite repair capacity, service-level constraints, and engineering validation. The empirical results show that repair bottlenecks shift probability mass toward higher numbers of failed units, create a fat-tailed backlog distribution, and double the required target stock from 6 to 12 MEC units under a 95% service-level requirement. Sensitivity experiments further show that repair turnaround time, fleet size, and MRO channel capacity jointly determine the inventory-capacity cost trade-off. The proposed framework provides an interpretable decision tool for airlines to align spare-part procurement with actual repair-system performance.

Keywords: aviation rotatable parts; inventory optimization; Markov queuing theory; maintenance capacity constraints; aircraft on-ground



Academic Editor: Sujin Bureerat

Received: 10 April 2026

Revised: 30 June 2026

Accepted: 6 July 2026

Published: 8 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

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

1. Introduction

In recent years, the global civil aviation industry has been facing a highly competitive market environment. With profit margins continually squeezed by changing operational environments and strict regulatory demands, improving fleet availability while reducing operational costs has become a primary strategic goal for airlines [1,2]. Among various controllable operational costs, the procurement and inventory management of aviation spare parts account for a large proportion [3,4]. The availability of these spare parts is closely related to flight safety and the prevention of severe operational disruptions, especially the costly aircraft on-ground (AOG) events [5,6].

Within the aviation material inventory, components are generally divided into expendable parts and rotatable parts. Rotatable parts, such as engine mechanical control (MEC) units, avionics computers, and landing gear assemblies, are characterized by their high unit prices and repairability. Instead of being discarded upon failure, a rotatable part enters a closed-loop lifecycle of removal, queuing for repair, maintenance, and eventually returning to the available inventory [7,8]. As a result, the capital invested in rotatable parts inventory constitutes a substantial portion of the assets within airline maintenance, repair, and overhaul (MRO) departments. Scientifically determining the target stock level (TSL) and developing accurate demand forecasts for these high-value rotatable parts is therefore of great importance to achieve cost reduction without compromising aviation safety standards [9,10].

Effective management of rotatable parts relies on accurate inventory classification. Traditionally, airlines have largely depended on financial or empirical rules, such as the activity-based classification based on unit cost, or the vital, essential and desirable (VED) classification based on operational criticality. However, as modern aircraft fleets grow in scale and complexity, these single-factor and static methods have shown obvious limitations, often resulting in severe spare parts overstock and capital occupation [9]. To overcome these limitations, recent studies have increasingly adopted multi-criteria decision-making and data-driven classification methods. Researchers have proposed comprehensive portfolio classification methods that combine maintenance history, logistics, and equipment data to connect different departmental information systems [7,11]. Furthermore, recent studies highlight the importance of considering actual operational features and demand lumpiness in classification models [5,12]. Despite these improvements, a major theoretical deficiency persists: existing classification frameworks are predominantly value-driven or consumption-driven, rarely categorizing rotatable parts based on their statistical failure patterns, such as exponential, normal, or Weibull distributions. Here, the term failure-pattern-based does not imply that this paper models microscopic fracture, wear, fatigue, or plastic-deformation physics directly. Rather, it means that inventory classification is linked to the empirically observed reliability behavior generated by those mechanisms, especially the distributional form of time-between-failure records. Consequently, ignoring the component's statistical failure behavior would reduce the accuracy of subsequent forecasting models [13].

After component classification, accurately predicting spare parts demand is a major challenge. For a high-reliability rotatable part, the key reliability parameter is the individual failure rate λ . When failures follow an exponential time-between-failure distribution, the aggregate removals over a short planning interval follow a low-rate Poisson process. This explains why operational demand records contain many zero-demand periods interrupted by occasional failure clusters. Because high-value rotatable parts usually have high reliability and low failure rates, their consumption records show highly fluctuating, intermittent time series with severe data imbalance, often having long periods of zero demand with occasional unexpected peaks [14–16]. To forecast such intermittent and imbalanced demand patterns, researchers have heavily focused on artificial intelligence (AI) and machine learning (ML) techniques [17]. Recent literature shows the successful use of tree-based ensemble models and artificial neural networks (ANNs) in forecasting intermittent spare parts demand in automotive and electric vehicle manufacturing [18]. In military and aviation logistics, advanced methods using stacked generalization, hybrid models for end-of-life products, and new ensemble approaches combined with focal loss and synthetic minority over-sampling technique have been proposed to handle data imbalances [19–21]. In addition, combining installed-base data and digital twin simulations with Adaptive particle swarm optimization (PSO) has shown good potential in improving demand forecasting across multi-stage supply chains [22,23]. Some researchers have also studied hybrid

after-sales strategies that combine traditional and 3D printing manufacturing to deal with demand uncertainty [24].

Nevertheless, these data-driven methods are not adopted simply because of the general popularity of AI. They are useful when nonlinear covariates and demand intermittency dominate the prediction task, but they are less transparent when the objective is to connect failures, repair queues, and inventory availability. However, although these data-driven algorithms show good statistical performance in time-series fitting, they mostly operate as black boxes. They are effective in detecting historical patterns but do not explicitly describe the operational dynamics of the rotatable repair cycle, including operation, failure, removal, queuing, repair, and return to stock [25]. More importantly, the majority of existing ML methods and statistical models assume infinite maintenance capacity and immediate replenishment, which is unrealistic.

In practice, the management of aviation rotatable parts is a complex queuing system constrained by limited repair capacities and long turnaround times [26]. Rotatable parts are typical exchangeable items in operations research, where the repair system must consider tolerable wait times, random lead times, and long-down constraints, such as maximum allowed AOG durations [6,27]. When modeling such dynamic, capacity-constrained repair environments, Markov chains provide strong physical interpretability [28]. By mathematically formulating the transition probabilities between system states, Markov models can dynamically evaluate the performance of spare parts inventory systems. The effectiveness of Markov-based heuristic models has been demonstrated in managing intermittent inventory flows [29]. Furthermore, partially observable Markov decision processes have been successfully employed to integrate spare unit selection, condition-based maintenance, and inventory control into a unified optimization framework [30].

Based on the underlying statistical failure behavior observed in service records, aircraft components are classified into two primary categories. Category I includes wear-out components governed by normal or Weibull distributions, and their predictable degradation is managed via deterministic hard-time replacements, making stochastic modeling unnecessary. Conversely, category II includes complex rotatable parts, such as MEC assemblies. In reliability engineering, the lifecycle of such complex repairable systems typically follows the classic bathtub curve. After passing the initial burn-in phase, these complex rotatables enter a prolonged useful life phase. During this period, failures are predominantly triggered by random external stressors, such as sudden temperature fluctuations or vibrations, rather than predictable age-related wear-out. Consequently, the component operates in a random failure phase where the failure rate remains strictly constant. Throughout their useful-life phase, these complex systems exhibit random failure patterns, with their times-to-failure strictly governed by an exponential distribution. This filtering step is also the way the proposed model addresses dependent or coupled failure modes: failures showing clear dependence, common-cause behavior, infant-mortality effects, or wear-out trends are excluded from Category II and should be handled by condition-based or multi-state reliability models rather than by the CTMC formulation used here. Isolating these Category II rotatables establishes the fundamental theoretical premise for the stochastic modeling framework proposed in this study. It ensures the degradation process possesses the memoryless property, the absolute prerequisite for deploying the continuous-time Markov chain (CTMC) models, to optimize capacity-constrained repair cycles.

In summary, while current research has made significant progress in AI-based demand forecasting and multi-criteria classification, existing models often separate the intrinsic link between a component's physical failure distribution, its dynamic repair cycle, and the limited capacity of the maintenance facility. Motivated by this theoretical deficiency, this study proposes a comprehensive demand forecasting and inventory optimization framework for

aviation rotatable parts. The novelty of the study lies in the problem-specific integration and validation of established reliability, Markov queuing, and inventory optimization tools under actual airline MRO constraints, rather than in claiming a new standalone mathematical algorithm. By integrating empirical failure curve classification with a Markovian birth-death queuing process, this research establishes a mathematically rigorous and physically interpretable methodology for airline spare parts management. The main contributions of this paper are as follows.

Contribution 1: Formulation of a refined classification strategy based on empirical failure curves. Unlike traditional financial metrics, this study categorizes rotatable parts based on their actual statistical failure density functions. This grounds the inventory model in observed statistical failure behavior rather than only financial value, and it defines a clear applicability boundary for Markov modelling.

Contribution 2: Construction of multi-scenario Markov queuing models for the airline rotatable repair loop. By mathematically modeling the state transitions of rotatable parts, this study derives steady-state probability equations and inventory calculation formulas under two operational scenarios: a base model without repair queuing and a constrained queuing model reflecting limited maintenance workshop capacity.

Contribution 3: Derivation of practical inventory thresholds via empirical validation. Using historical fault data from MEC components of a B737 fleet, the proposed model calculates the minimum discrete inventory level required to ensure a 95% serviceability rate. This achieves a practical balance between reducing AOG risks and improving capital efficiency.

Contribution 4: Engineering validation of transferability and cost trade-offs. Beyond the baseline MEC case, this study evaluates service-level sensitivity, fleet-size pooling, MRO capacity investment, and parameter-transfer stress scenarios, thereby showing how the framework can support auditable engineering decisions rather than only fitting one historical demand series.

This paper is organized as follows. Section 2 describes the methodology integrating empirical failure curve-based classification with a CTMC queuing network, which is employed to model the capacity-constrained repair cycle and optimize the target stock level. Section 3 demonstrates the proposed framework through simulation experiments using actual operational data of MEC components from a B737 fleet. The discussion and conclusions are presented in Sections 4 and 5, respectively.

2. Materials and Methods

In a typical airline maintenance ecosystem, a complex rotatable part, such as a MEC unit, operates on the aircraft until a random failure occurs. Upon failure, the component is removed and sent to the MRO facility, simultaneously triggering a demand for a serviceable spare part from the inventory. If a spare is available, it is immediately installed to prevent an AOG event. If the inventory is depleted, a backorder occurs, resulting in severe operational penalty costs.

To ensure theoretical rigor and logical consistency, the methodology bridges the proposed multi-distribution classification framework with a stochastic queuing model. The classification step serves as a critical filter, while the CTMC developed in this section is exclusively applied to complex, high-value rotatable parts that exhibit random failures. By isolating components that follow an exponential failure distribution, the model adheres to the memoryless property required for Markov processes. The overall research framework is shown in Figure 1.

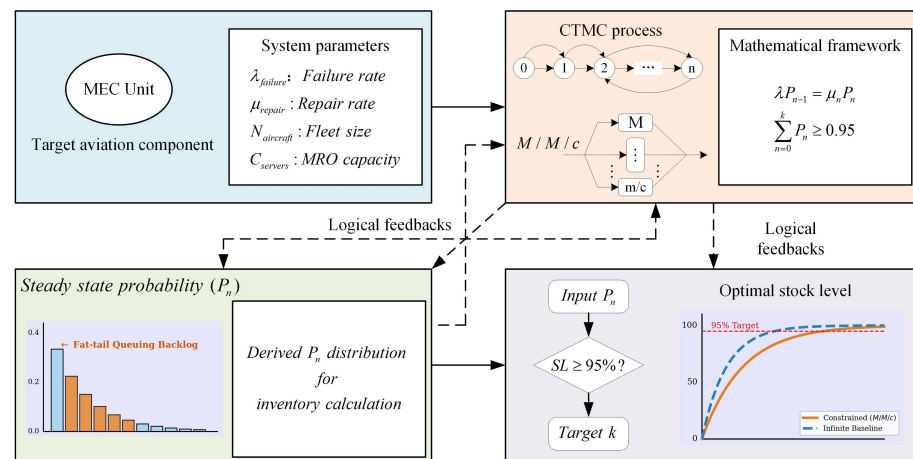


Figure 1. Overall research framework.

2.1. Problem Description and Fundamental Assumptions

To formulate this continuous-time stochastic repair loop into a tractable mathematical model, the following rigorous assumptions are established.

Assumption 1. Following the classification strategy, the targeted high-value rotatable parts exhibit a constant failure rate during their useful life. Thus, the failure process of the installed fleet follows a Poisson process with parameter λ , and the inter-arrival time of failures follows an Exponential distribution.

Assumption 2. The repair time at the MRO facility is an exponentially distributed random variable with a mean repair rate μ . The repair process restores the component to a serviceable condition.

Assumption 3. The total number of operating aircraft in the active fleet is denoted as N . Since each aircraft is equipped with 2 MEC units, the installed base is $2N$. The aggregate demand rate is $\Lambda = 2N\lambda$.

Assumption 4. The inventory system operates on a continuous review and base-stock policy, equivalent to a $(S - 1, S)$ policy. Due to the high unit cost of rotatables, batch ordering is not viable, and every failure instantly triggers a single repair and replenishment request.

2.2. State Definition and Continuous-Time Markov Chain

Let n denote the total number of rotatable parts currently in the repair network at time t , encompassing both the components actively undergoing repair and those waiting in the queue. The state space of the system is discrete, defined as $S = \{0, 1, 2, \dots, n, \dots\}$.

The equations in this section are derived by the authors by specializing the standard birth-and-death CTMC balance principle to the rotatable-parts repair loop. The Markov and queuing foundations follow established repairable-inventory and spare-parts models [27–30], while the specific state definition, service-level constraint, and cost formulation are adapted to the MEC inventory problem studied here. The state transitions follow a classic birth-and-death process. A birth corresponds to a component failure, incrementing the system state from n to $n + 1$ with a constant transition rate Λ , and a death corresponds to a successful repair, decrementing the state from n to $n - 1$. The death transition rate μ_n is state-dependent and governed by the specific capacity constraints of the MRO facility.

Under steady-state conditions ($t \rightarrow \infty$), the probability flux entering state n must equal the flux leaving that state. Based on the infinitesimal generator matrix of the CTMC, the global balance equations are formulated as Equation (1):

$$\begin{cases} \Lambda P_{n-1} + \mu_{n+1} P_{n+1} = (\Lambda + \mu_n) P_n, & \text{for } n \geq 1 \\ \mu_1 P_1 = \Lambda P_0, & \text{for } n = 0 \end{cases} \quad (1)$$

where P_n represents the steady-state probability that exactly n components are in the repair system.

2.3. Steady-State Probabilities Under Multi-Scenarios

Because the repair transition rate μ_n is intrinsically linked to the resource availability (e.g., test benches, engineers) at the MRO workshop, this study derives exact steady-state probabilities under two distinct operational scenarios.

(1) Scenario A: Non-queuing repair system

As a baseline, it is assumed that the MRO facility possesses infinite parallel repair capacity. Consequently, any arriving failed component commences repair immediately without queuing. This corresponds to an $M/M/\infty$ queuing model. The effective repair rate is linearly proportional to the number of failed units in the system: $\mu_n = n\mu$. Substituting this into Equation (1) yields Equation (2):

$$P_n = \frac{1}{n!} \left(\frac{\Lambda}{\mu} \right)^n P_0 \quad (2)$$

Applying the normalization condition $\sum_{n=0}^{\infty} P_n = 1$, the probability of an empty system is $P_0 = e^{-\frac{\Lambda}{\mu}}$. Thus, the steady-state probability distribution follows the Poisson form shown in Equation (3):

$$P_n = \frac{\rho^n}{n!} e^{-\rho} \quad (3)$$

where $\rho = \frac{\Lambda}{\mu}$ represents the traffic intensity, indicating the expected number of components concurrently circulating in the repair pipeline.

(2) Scenario B: Queuing repair system

In realistic operational settings, MRO facilities face stringent capacity constraints. Assume the workshop possesses a limited number of parallel repair channels, denoted as c . If $n \leq c$, components are repaired immediately. However, if $n > c$, the excess components must wait in a buffer queue. This structural bottleneck reflects an $M/M/c$ queuing configuration.

The state-dependent repair rate becomes the piecewise function in Equation (4):

$$\mu_n = \begin{cases} n\mu, & 0 \leq n \leq c \\ c\mu, & n > c \end{cases} \quad (4)$$

By recursively solving Equation (1) with the piecewise rate in Equation (4), the steady-state probability P_n under finite capacity is derived as Equation (5):

$$P_n = \begin{cases} \frac{\rho^n}{n!} P_0, & 0 \leq n \leq c \\ \frac{\rho^c}{c!} \left(\frac{\rho}{c} \right)^{n-c} P_0, & n > c \end{cases} \quad (5)$$

The boundary probability P_0 is evaluated by the normalization condition as Equation (6):

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!(1-\rho/c)} \right]^{-1} \quad (6)$$

For the system to reach a stable equilibrium (preventing the queue from growing infinitely), the ergodicity condition $\rho < c$ (i.e., $\Lambda < c\mu$) must be strictly satisfied.

2.4. Target Stock Optimization and Service Level Evaluation

The ultimate objective of this mathematical modeling is to determine the optimal target stock level k to satisfy the strict service level (SL) α . In aviation spare-parts planning, a 100% or 99.99% fill-rate target is usually not economically attainable because the tail of the demand distribution is unbounded and would require disproportionately large capital investment. Therefore, this study uses 95% as a baseline operational service-level target that balances AOG-risk mitigation and inventory affordability, and then evaluates higher levels through sensitivity analysis.

Under the $(S-1, S)$ policy with an initial inventory of k spares, an out-of-stock backorder is triggered only when the total number of components in the repair network n exceeds the available shelf inventory k . Therefore, the probability of fulfilling a replacement demand immediately from stock is equivalent to the cumulative probability that n is less than or equal to k , as given by Equation (7):

$$P(n \leq k) = \sum_{n=0}^k P_n \quad (7)$$

To minimize the capital allocated to inventory while guaranteeing operational availability, the optimization problem is formulated as the probabilistically constrained integer programming model in Equation (8):

$$\begin{cases} \min_{k \in \mathbb{N}} C(k) = kV \\ \text{Subject to: } \sum_{n=0}^k P_n \geq \alpha \end{cases} \quad (8)$$

where V denotes the unit procurement value of the rotatable part. Because the cumulative probability $\sum P_n$ is a monotonically increasing function of k , this non-linear constraint can be efficiently solved using marginal analysis. By embedding the P_n derived from scenario A or scenario B, airlines can determine the minimum inventory thresholds satisfying the service-level constraint, achieving a robust balance between AOG risk mitigation and financial efficiency.

3. Results

3.1. Parameter Settings

To validate the theoretical queuing framework, an empirical study was conducted utilizing the operational and maintenance records of a major commercial airline. For confidentiality reasons, the specific airline is anonymized. The dataset focuses on the MEC units installed on a mature B737 fleet.

The primary data extraction covered a consecutive observation window of 8 years (from 2018 to 2025). During this period, the airline operated an active fleet of 50 B737 aircraft, resulting in a stable installed base of $N = 100$ MEC units. To ensure statistical reliability, the raw maintenance logs were rigorously preprocessed. Only inherent component failures confirmed by service-engineering diagnosis were retained. Non-inherent records, including

no-fault-found removals, scheduled replacements, documentation errors, handling damage, and clearly human-induced events, were excluded because they do not represent the stochastic reliability behavior of the MEC unit itself. Ultimately, a total of 292 valid, inherent failure events were isolated.

To explicitly validate the memoryless property assumption required for the subsequent Markov modelling, a rigorous goodness-of-fit analysis was conducted on the time-between-failure data. The estimated mean time between failures is approximately 1000 days, yielding a constant individual failure rate of $\lambda = 0.001$ failures/day. The Kolmogorov-Smirnov (K-S) test is applied to compare the empirical cumulative distribution function (ECDF), $F_n(t)$, of the 292 valid failure events against the theoretical exponential cumulative distribution function, $F(t) = 1 - e^{-\lambda t}$. The K-S test statistic D , representing the maximum absolute distance between the ECDF and the theoretical CDF, is defined as:

$$D = \max_{1 \leq i \leq n} |F_n(t_i) - F(t_i)|$$

For a sample size of $n = 292$ and a significance level of $\alpha = 0.05$, the critical value $D_{critical} \approx 0.0796$. The statistical software calculated a test statistic of $D = 0.045$, which is strictly less than the critical value, alongside a high p -value of 0.38 ($p > 0.05$). Consequently, the null hypothesis cannot be rejected, mathematically confirming that the time-between-failure data of the MEC units rigorously follows an exponential distribution. Currently, backend MRO tracking data reveals that the average shop TAT for an unserviceable MEC, including induction, teardown, component replacement, and final test cell calibration, is rigorously tracked at 30 days.

Consequently, the mean repair rate per active channel is parameterized as $\mu = 0.0333$ units/day. The unit procurement value V for an MEC is set at 500,000, and the safety-critical target service level α is universally enforced at 95% to strictly mitigate AOG risks.

3.2. Experimental Results

3.2.1. Steady-State Distribution and Queuing Effect

Based on the established parameters, the model first computes the steady-state probability P_n of having n failed components in the repair network. Here, n does not denote the number of physical components installed in one engine; it denotes the number of MEC units simultaneously outside service across the fleet and circulating in the repair pipeline. The finite capacity case is evaluated over a truncated numerical range large enough for the remaining tail probability to be negligible. In scenario A with non-queuing, P_n exhibits a standard Poisson distribution, and the probability of an empty system is $P_0 \approx 0.0498$. However, in scenario B with finite capacity, the structural bottleneck of $c = 4$ induces queuing delays when failures surge. This causes the probability distribution curve to shift to the right, reducing P_0 to 0.0377 and significantly inflating the tail probabilities for states $n \geq 5$, which explicitly represent the accumulation of backlogs in the system. As illustrated in Figure 2. The steady-state probability mass function P_n for the number of components in the repair network exhibits a profound divergence depending on the assumed MRO capacity constraints.

For states where the number of failed components is relatively small $n \leq 3$, the infinite capacity model of scenario A, represented by the blue bars, yields higher probabilities than the finite capacity model of scenario B. Mathematically, this represents an idealized condition where failed components are inducted instantaneously without queuing, enabling the system to efficiently clear unserviceable units and maintain a minimal system footprint. Consequently, the probability of having a near-empty system is maximized.

A critical shift occurs as the number of failed parts reaches and exceeds the MRO workshop's physical channel limit $c = 4$. This is why the finite-capacity probability bars become higher than the infinite-capacity bars after approximately five failed units: once all

four repair channels are occupied, additional failed units cannot begin service immediately and must accumulate in the queue. The finite model is therefore not “better” than the infinite model in this region; it is showing the higher probability of backlog caused by the bottleneck. In scenario B, represented by the orange bars, the strict capacity constraint prevents any more than 4 units from being repaired simultaneously, forcing subsequent arrivals into a backlog queue. Visually, this creates a pronounced rightward shift and a fat tail in the probability distribution. The probabilities of having 5, 6, 7, or more components stuck in the MRO pipeline are significantly magnified compared to the rapid exponential decay seen in Scenario A.

This evidence indicates a severe operational risk. If an airline’s inventory planning department relies on the infinite-capacity Poisson distribution, it will underestimate the likelihood of extreme accumulation events (e.g., having 6 or 7 parts out of service at once). This systemic underestimation can translate into premature stock depletion and increased AOG exposure. Therefore, incorporating the $M/M/c$ queuing dynamics is necessary for capturing the degradation of serviceability caused by workshop bottlenecks.

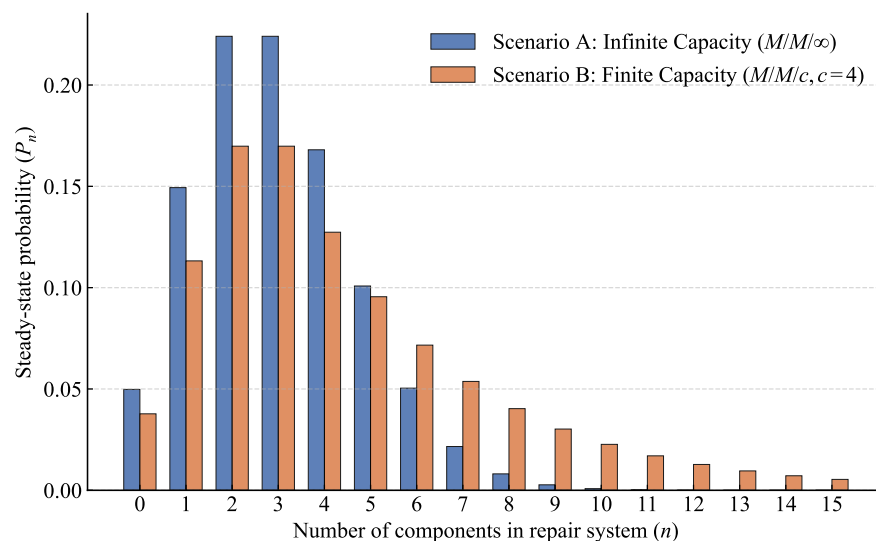


Figure 2. Steady-state probability mass function of scenario A and B.

As illustrated in Figure 3, the cumulative distribution function of the system state determines the expected operational fill rate as a step function of the initial inventory investment k . The red dashed horizontal line indicates the 95% target service level used for mitigating AOG risks.

The blue step curve, representing the infinite-capacity $M/M/\infty$ baseline, ascends rapidly. Because failed components are immediately accepted for processing, without queuing, the probability mass is highly concentrated at lower system states. Consequently, the blue curve crosses the 95% threshold early, mathematically demonstrating that an optimal target stock level of $k_A = 6$ units is theoretically sufficient to sustain fleet operations.

By contrast, the orange step curve representing the realistic finite-capacity $M/M/c$ model ($c = 4$) climbs at a noticeably slower rate. At very low inventory levels, the two curves should be interpreted as cumulative probabilities rather than cost performance curves. The infinite-capacity curve remains above the finite-capacity curve at every stock level because the no-queue assumption keeps fewer units in the repair pipeline. The horizontal lag between the two curves captures the queuing penalty caused by the structural bottleneck of the MRO facility. Because components spend additional time in the repair backlog, the overall pipeline expands. To compensate for this operational friction and push

the fill rate above the red target line, the optimal inventory threshold shifts rightward, requiring $k_B = 12$ units.

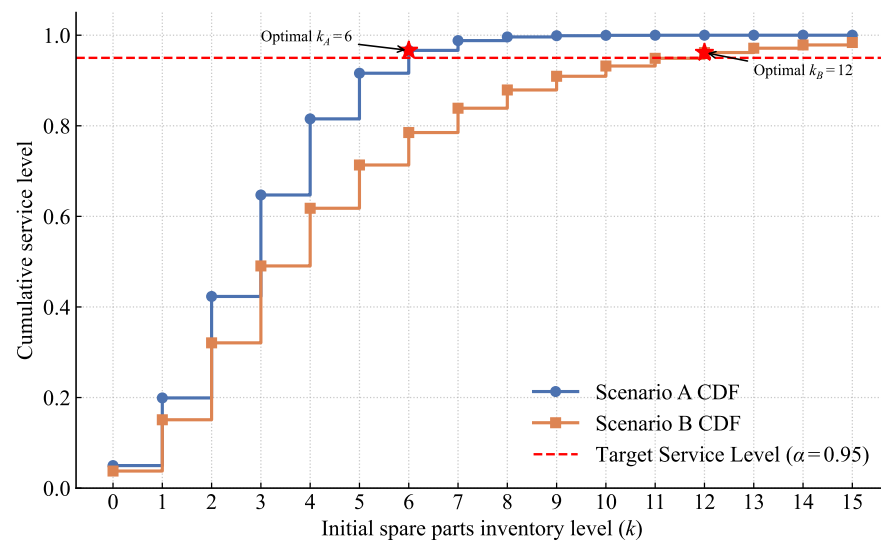


Figure 3. Cumulative distribution function of scenario A and B.

The gap between these two optimal points $\Delta k = 6$ is the most critical finding of this optimization framework. It quantitatively isolates the hidden inventory requirement induced solely by maintenance capacity constraints. If an airline's material planning department utilizes traditional forecasting models that implicitly assume infinite repair capacity, they will procure only 6 units, inadvertently creating a deterministic shortage. This systemic miscalculation would severely expose the fleet to unmitigated AOG risks whenever repair demands exceed the 4 available test channels. Therefore, integrating the exact capacity parameter c into the inventory calculation is imperative for robust supply chain resilience.

As depicted in Figure 4, the sensitivity of the optimal target stock level k to variations in the average repair TAT is distinctly non-linear, revealing critical insights into the system's queuing dynamics. The figure illustrates the inventory trajectories under three different service level requirements: 90%, 95% (baseline), and 99%.

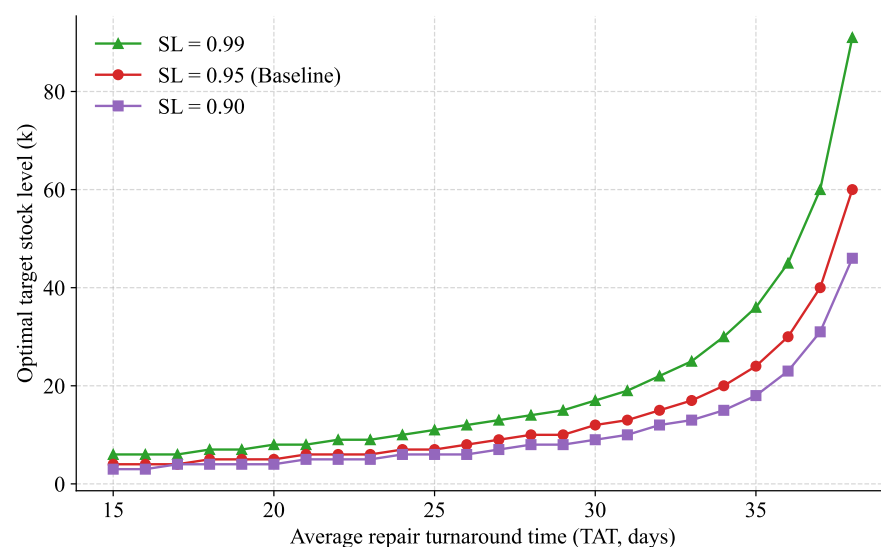


Figure 4. Sensitivity of the optimal target stock level to variations in the average repair TAT.

When the MRO facility operates with high efficiency, such as rapid repair processes, the required inventory levels are low and tightly clustered. For instance, at a TAT of 15 days, securing a 99% service level requires only one more spare part than a 90% service level. This demonstrates that when repair velocity is high, airlines can afford a higher service-level buffer at a negligible marginal inventory cost.

As the TAT increases towards the right side of the x-axis, the inventory curves diverge significantly and exhibit an exponential upward trend. This is fundamentally driven by the $M/M/c$ queuing theory limit. Because the total demand arrival rate is 0.1 and the capacity is 4, the system approaches its absolute saturation point as the TAT approaches 40 days. In this near-bottleneck state (e.g., TAT = 38 days), the queue length explodes. Consequently, maintaining a 95% SL requires a massive inventory injection, exceeding 50 units, and achieving a 99% SL becomes financially prohibitive, exceeding 80 units.

This visual evidence reveals that relying solely on inventory buffering to mitigate supply chain disruptions is unsustainable. As MRO turnaround times worsen, the marginal cost of maintaining a target service level skyrockets. Therefore, capital investments should be dynamically reallocated rather than passively purchasing more 500,000 MEC units to hedge against long TATs. The 25–30 day region in Figure 4 corresponds to the observed service-engineering baseline and therefore serves as the empirical anchor for the sensitivity experiment. It is vastly more cost-effective to invest in MRO process improvements, such as adding test benches, cross-training staff, to compress the TAT and shift the system back to the stable, low-inventory zone.

In aviation strategic planning, as an airline introduces new aircraft into its fleet, the demand for rotatable spare parts does not necessarily increase linearly. To quantify the impact of fleet expansion on inventory requirements, an experiment was conducted varying the fleet size N from 20 to 150 aircraft, adjusting the aggregate demand arrival rate Λ proportionally while keeping the target service level at 95% and the MRO channels constant $c = 6$ to accommodate maximum flow.

As illustrated in Figure 5, the optimal target stock level k exhibits a sub-linear growth trajectory in response to fleet expansion. This non-linear relationship is numerically detailed in Table 1. For a small fleet of 20 aircraft, 3 spare MEC units are required, resulting in a high per-aircraft stock ratio of 0.150. However, when the fleet scales up to 100 aircraft, the required inventory only increases to 10 units, dropping the per-aircraft ratio significantly to 0.100. The real outcome of Figure 5 is that the model converts fleet-expansion plans into discrete spare-part procurement requirements and shows when common pooling can reduce the per-aircraft stock burden. This mathematically demonstrates the risk pooling effect inherent in queuing networks. A larger installed base smooths out the variance of intermittent random failures, allowing airlines to achieve massive economies of scale in inventory holding costs.

A critical managerial dilemma in airline operations is allocating capital between purchasing spare parts and investing in MRO repair capacity, such as adding test benches or hiring more technicians. To evaluate this trade-off, a joint cost optimization experiment was formulated. The total cost is defined as the sum of the inventory holding cost kV and the capacity investment cost cC_{cap} . Assuming the annualized investment equivalent for adding one dedicated MRO channel C_{cap} is 800,000, and the unit cost of an MEC remains 500,000. The fleet size is fixed at $N = 100$, that is, $\Lambda = 0.2$.

Table 2 and Figure 6 present the optimization results across different MRO capacity constraints c . Because $\rho = \Lambda/\mu = 6.0$, the system requires a strict minimum of $c = 7$ channels to maintain stability, that is, $\rho < c$. At the bottleneck boundary $c = 7$, extreme queuing delays force the airline to hold 23 spare units to satisfy the 95% service level, driving the total cost to 17.1 million. However, investing in two additional repair

channels $c = 9$ eliminates the bottleneck, drastically reducing the required inventory to 12 units. The global optimum is achieved at $c = 9$, minimizing the total system cost to 13.2 million. This proves that moderate over-capacity in MRO facilities is economically superior to holding excessive high-value inventory. This experiment also adds an explicit cost calculation to the service-level analysis: the optimal decision is not simply the lowest inventory level, but the minimum total cost that still satisfies the target fill-rate constraint.

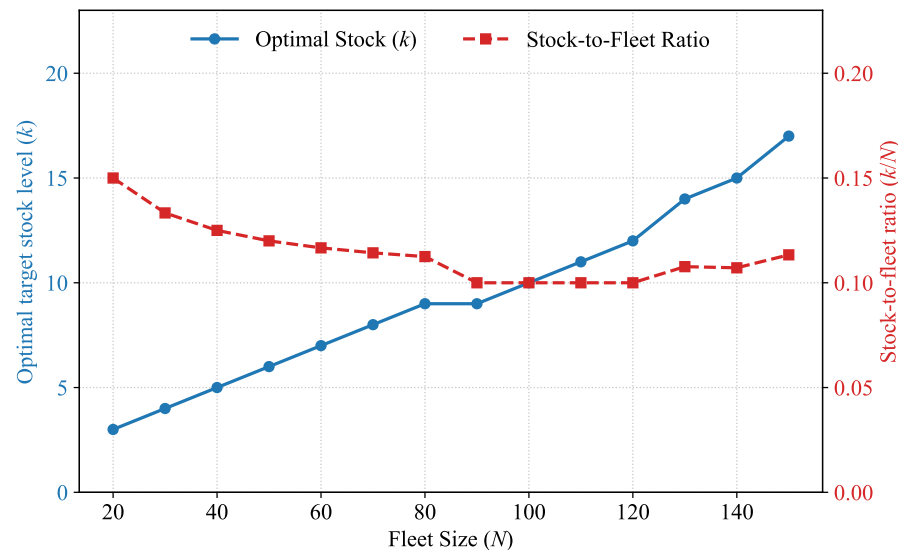


Figure 5. Variation in optimal target stock level and stock-to-fleet ratio under varying fleet sizes.

Table 1. Optimal inventory configurations and economies of scale metrics across different fleet sizes.

Fleet Size (N)	Installed Base	Demand Rate (Λ)	Optimal Stock (k)	Stock-to-Fleet Ratio (k/N)	Total Inventory Cost (Million \$)
20	40	0.04	3	0.150	1.5
50	100	0.10	6	0.120	3.0
80	160	0.16	9	0.113	4.5
100	200	0.20	10	0.100	5.0
150	300	0.30	17	0.113	8.5

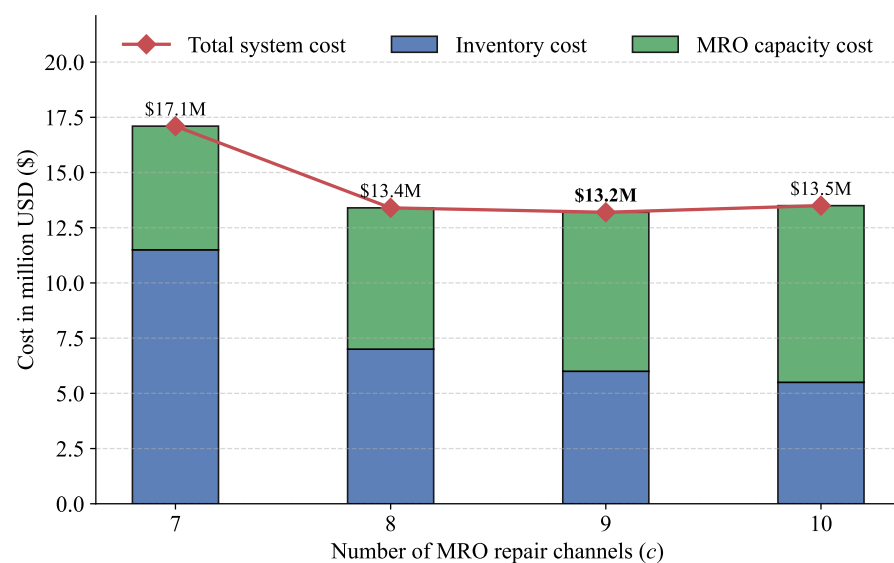


Figure 6. Joint optimization of MRO capacity and target stock level.

Table 2. System cost breakdown and performance metrics under varying MRO channel capacities.

MRO Channels (<i>c</i>)	Traffic Intensity (ρ)	Optimal Stock (<i>k</i>)	Actual Fill Rate (%)	Inventory Cost (Million \$)	Capacity Cost (Million \$)	Total Cost
7	6.0	23	95.82	11.5	5.6	17.1
8	6.0	14	96.14	7.0	6.4	13.4
9	6.0	12	96.03	6.0	7.2	13.2
10	6.0	11	97.45	5.5	8.0	13.5

3.2.2. Additional Validation: Parameter-Transfer Stress Test

To further address the concern that the proposed framework may be limited to a single MEC case, an additional parameter-transfer stress test was conducted. This test is not intended to claim independent empirical validation for every aircraft component class. Instead, it examines whether the same CTMC inventory logic produces stable and interpretable target-stock decisions when the input parameters are shifted to representative ranges for other Category II rotatable parts. The tested component classes include the B737 MEC baseline, an avionics-computer class with a lower aggregate removal rate and shorter TAT, a hydraulic-actuator class with a moderate aggregate removal rate and longer TAT, and a high-stress sensitivity scenario representing degraded reliability and slower repair performance.

As shown in Table 3, the model remains numerically stable whenever the ergodicity condition $\rho < c$ is satisfied, and the resulting optimal stock levels vary in the expected direction. Lower aggregate demand and faster repair reduce the required target stock, while longer TAT or higher failure intensity increases the stock requirement. This additional validation supports two scientific points. First, the model is not merely a curve-fitting calculation for one MEC data set; it is a parameterized repairable-inventory framework. Second, the model's applicability is conditional rather than universal: before applying it to any other rotatable type, the analyst must verify the random-failure assumption and update Λ , μ , c , and V from the corresponding service-engineering records.

Table 3. Parameter-transfer stress test for different Category II rotatable classes.

Scenario	Installed Base	Λ	TAT (Days)	<i>c</i>	ρ	<i>k</i> ₉₅
B737 MEC baseline	100	0.1000	30	4	3.000	12
Avionics computer	50	0.0375	25	3	0.938	3
Hydraulic actuator	100	0.0550	35	4	1.925	5
High-stress sensitivity	100	0.1200	35	6	4.200	11

4. Discussion

This study proposes an aviation rotatable parts inventory optimization framework that integrates empirical failure curve-based classification with a CTMC queuing network. The validation strategy contains four layers: service-engineering filtering of actual maintenance records, statistical goodness-of-fit testing of the exponential failure assumption, operational benchmarking of the resulting inventory thresholds against observed TAT and capacity conditions, and a parameter-transfer stress test across representative Category II rotatable classes. By simulating the actual operational data of MEC components from a B737 fleet,

this section evaluates the theoretical contributions, managerial implications, and future research directions of the proposed model.

Existing aviation spare parts demand forecasting methods frequently rely on data-driven machine learning models. While these models perform well in time-series fitting, they generally treat the repair system as a black box with infinite capacity, overlooking the underlying physical mechanisms of component flow. The theoretical contribution of this study lies in bridging empirically observed statistical failure distributions with queuing theory. Through a pre-classification mechanism, the model isolates components that exhibit random failures, which strictly aligns with the memoryless property assumption of the Markov birth-and-death process. Furthermore, the derivation of steady-state probabilities under finite repair capacity $M/M/c$ mathematically demonstrates the fat-tail effect of queuing backlogs, addressing the theoretical blind spots caused by traditional infinite-capacity $M/M/\infty$ assumptions.

From the perspective of airline asset management, the sensitivity analysis in this study provides a quantitative tool for decision-making. The experimental results reveal that relying on traditional idealized forecasting models may lead to a severe inventory shortfall, thereby escalating the risk of AOG events. Moreover, the model confirms that the optimal target stock level k is highly sensitive to the average TAT. As the repair cycle approaches the capacity bottleneck, the inventory capital required to maintain a 95% service level surges significantly. This finding indicates to management that when facing supply chain constraints for high-value rotables, simply injecting capital to purchase more spare parts is inefficient. Instead, improving the repair rate of the MRO facility through process optimization and technician cross-training presents a more effective pathway to accelerate capital turnover and mitigate supply chain risks.

It is important to state the generalizability of the proposed framework precisely. While the empirical validation in this study focuses on B737 MEC components, the mathematical core of the $M/M/c$ queuing model is not tied to the physical structure of the MEC unit itself. The model can be transferred to other critical Category II components, such as avionics computers or hydraulic actuators, provided that their useful-life failure records satisfy the memoryless assumption and that the input parameters are re-estimated from component-specific data. Variations in failure rates (λ), repair turnaround times (μ), MRO channels (c), and unit procurement costs (V) do not invalidate the structural logic of the system. Instead, they enter the probabilistically constrained integer programming model as auditable engineering inputs. This conditional transferability is a central practical contribution of the study.

Although this study successfully quantifies steady-state inventory thresholds using a queuing model, the complexity of real-world aviation maintenance ecosystems presents significant opportunities for future research.

First, the current model primarily allocates inventory based on reactive failure arrival rates. Future research could transition toward proactive condition-based maintenance by integrating backend inventory control with frontend aviation engine health management systems. Specifically, to address complex coupled faults frequently observed during engine operation, such as the simultaneous occurrence of gas path performance degradation and sensor drift, explainable reinforcement learning or meta-learning architectures could be introduced for high-fidelity data reconstruction and early fault diagnosis. Upgrading from reactive queuing inventory prediction to a dynamic, state-driven supply chain allocation system could further push the boundaries of fleet operational safety and economic efficiency.

Second, while this study utilizes a mature B737 fleet as a baseline case, future empirical research should extend this optimization framework to the commercial operation of

new-generation aircraft. For example, by considering the introductory operational environments of carriers, researchers could explore the unique maintenance queuing patterns of domestic aircraft models. This analysis could assist in developing a holistic operational readiness framework that aligns spare parts provisioning with advanced personnel training paradigms, including evidence-based training and competency-based training and assessment. Additionally, expanding the CTMC model to encompass multi-echelon inventory networks serves as another critical direction for expanding the model's practical applicability.

5. Conclusions

With the global civil aviation industry's dual pursuit of cost reduction and high dispatch reliability, the scientific management of high-value rotatable parts inventory remains a core operational challenge. To address the limitations of traditional models that overlook underlying statistical failure behavior and maintenance capacity constraints, this paper presents a quantitative evaluation and optimization methodology. The main conclusions of this study are as follows:

- (1) Development of a failure-pattern-based classification framework. The study overcomes the limitations of traditional financial-based classification by explicitly isolating complex rotatable parts that exhibit random failures. This isolation provides the necessary theoretical foundation, specifically the memoryless property, for applying CTMC queuing models.
- (2) Quantification of the amplification effect of maintenance capacity bottlenecks. By constructing a closed-loop $M/M/c$ queuing network with finite capacity constraints, the research mathematically demonstrates that maintenance backlogs generate a noticeable rightward shift and fat-tail phenomenon in the distribution of failed parts. Compared to the infinite-capacity baseline, this model successfully captures the AOG risk exposure hidden within queuing delays.
- (3) Formulation of a dynamic inventory-efficiency trade-off strategy. Simulation results based on actual fleet MEC components indicate that to sustain a 95% service level, the target inventory must be dynamically calibrated against the MRO workshop's actual TAT. The sensitivity analysis mathematically confirms the primary role of optimizing MRO repair efficiency in controlling marginal inventory costs.

In summary, the Markov queuing inventory model developed in this paper provides an analytical framework with statistical reliability interpretability for aviation maintenance resource allocation. Its contribution is the construction of a verifiable engineering bridge from failure-pattern screening to finite-capacity repair modelling and service-level inventory decisions, rather than the isolated reuse of mature mathematical tools. The findings contribute to intermittent aviation spare parts demand forecasting and offer decision support for airlines and MRO enterprises navigating complex capacity constraints.

Author Contributions: G.C.: conceptualization, investigation, project administration, visualization, and writing—original draft; Y.T.: data curation, resources, supervision, writing—review and editing, and correspondence; Y.Z.: methodology, validation, software, and writing—original draft; Z.G.: funding acquisition, supervision, validation, and writing—review and editing; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2023YFB4302400).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from China Eastern Airlines and are available from Yue Teng with the permission of China Eastern Airlines.

Acknowledgments: The authors acknowledge the data provided by China Eastern Airlines.

Conflicts of Interest: Guannan Chen was employed by the China Eastern Airlines Holding Co., Ltd., and Yue Teng was employed by the China Eastern Technology Application Research and Development Center Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AOG	Aircraft on-ground
CTMC	Continuous-time Markov chain
MEC	Mechanical engine control
MRO	maintenance, repair, and overhaul
TAT	Turnaround time
AI	Artificial intelligence
ML	Machine learning
ANNs	Artificial neural networks
PSO	Particle swarm optimization
SL	Service level

References

1. Alomar, I.; Nikita, D. Managing operational efficiency and reducing aircraft downtime by optimization of aircraft on-ground (AOG) processes for air operator. *Appl. Sci.* **2025**, *15*, 5129. <https://doi.org/10.3390/app15095129>.
2. Zhang, S.; Huang, K.; Yuan, Y. Spare parts inventory management: A literature review. *Sustainability* **2021**, *13*, 2460. <https://doi.org/10.3390/su13052460>.
3. You, L.; Wang, L.; Lv, X.; Xiang, H.; Wang, Z. Evaluation of spare parts support capacity of civil aircrafts based on type-2 hesitant Pythagorean fuzzy sets and improved TOPSIS. *Appl. Sci.* **2024**, *14*, 7475. <https://doi.org/10.3390/app14177475>.
4. Ekin, E. An application on demand forecasting and stock control for aircraft components. *J. Aviat. Space Sci.* **2022**, *3*, 1–40. <https://doi.org/10.52995/jass.1122940>.
5. Shafi, I.; Sohail, A.; Ahmad, J.; Espinosa, J.C.M.; López, L.A.D.; Thompson, E.B.; Ashraf, I. Spare parts forecasting and lumpiness classification using neural network model and its impact on aviation safety. *Appl. Sci.* **2023**, *13*, 5475. <https://doi.org/10.3390/app13095475>.
6. Lamghari-Idrissi, D.; Basten, R.; Van Houtum, G.J. Spare parts inventory control under a fixed-term contract with a long-down constraint. *Int. J. Prod. Econ.* **2020**, *219*, 123–137. <https://doi.org/10.1016/j.ijpe.2019.05.023>.
7. Didriksen, S.K.; Sigsgaard, K.W.; Mortensen, N.H.; Jespersen, C.B. An empirical data model for spare parts management: Linking maintenance, logistics, inventory, and equipment data to bridge information silos. *Appl. Sci.* **2026**, *16*, 94. <https://doi.org/10.3390/app16010094>.
8. Wang, H.; Liu, H.; Shao, S.; Zhang, Z. Methodology of shipboard spare parts requirements based on whole part repair strategy. *Mathematics* **2024**, *12*, 3053. <https://doi.org/10.3390/math12193053>.
9. Ramírez, X.S.; Solorio, X.G.; Paredes, M.P.; Sifuentes, C.T. Proposal for an improvement model to reduce aircraft spare parts overstock based on ABC methodology and neural network forecasting. In Proceedings of the 2024 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI), Bogotá, Colombia, 2–4 October 2024. <https://doi.org/10.1109/CONIITI64189.2024.10854866>.
10. Vaez-Alaei, M.; Atashbar, N.Z.; Baboli, A. Target stock level and fill rate optimization for worldwide spare parts inventory management: A case study in business aircraft industry. In Proceedings of the 2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), Marrakech, Morocco, 21–23 November 2018. <https://doi.org/10.1109/ITMC.2018.8691215>.
11. Didriksen, S.K.; Sigsgaard, K.W.; Mortensen, N.H.; Jespersen, C.B. Assigning spare parts management decision-making strategies: A holistic portfolio classification methodology. *Appl. Sci.* **2026**, *16*, 1961. <https://doi.org/10.3390/app16041961>.

12. Choi, B.; Suh, J.H. Forecasting spare parts demand of military aircraft: Comparisons of data mining techniques and managerial features from the case of South Korea. *Sustainability* **2020**, *12*, 6045. <https://doi.org/10.3390/su12156045>.
13. Heijenrath, L.M.; Verhagen, W.J. The influence of repair quality on aircraft spare part demand variability. *Aerospace* **2023**, *10*, 731. <https://doi.org/10.3390/aerospace10080731>.
14. Petropoulos, F.; Apiletti, D.; Assimakopoulos, V.; Babai, M.Z.; Barrow, D.K.; Taieb, S.B.; Bergmeir, C.; Bessa, R.J.; Bijak, J.; Boylan, J.E.; et al. Forecasting: Theory and practice. *Int. J. Forecast.* **2022**, *38*, 705–871. <https://doi.org/10.1016/j.ijforecast.2021.11.001>.
15. Ranabhatt, N.; Barreto, S.; Pimpão, M.; Prates, P. Demand forecasting in the automotive industry: A systematic literature review. *Forecasting* **2025**, *7*, 73. <https://doi.org/10.3390/forecast7040073>.
16. Sfiris, D.S.; Koulouriotis, D.E. A new approach to forecast intermittent demand and stock-keeping-unit level optimization for spare parts management. *Appl. Sci.* **2025**, *15*, 12030. <https://doi.org/10.3390/app152212030>.
17. Ekin, E. Quantitative demand forecasting of spare parts in the aviation industry: A comparative analysis. *J. Aviat. Res.* **2026**, *8*, 1–32. <https://doi.org/10.51785/jar.1677452>.
18. Fu, W.; Bian, H.; Chen, J.; Jing, S. Tree-based machine learning intermittent demand forecasting for spare parts in electric vehicle manufacturing. *World Electr. Veh. J.* **2026**, *17*, 127. <https://doi.org/10.3390/wevj17030127>.
19. Kenaka, S.P.; Chakravastia, A.; Ma'ruf, A.; Cahyono, R.T. Enhancing intermittent spare part demand forecasting: A novel ensemble approach with focal loss and SMOTE. *Logistics* **2025**, *9*, 25. <https://doi.org/10.3390/logistics9010025>.
20. Kim, J.D.; Kim, T.H.; Han, S.W. Demand forecasting of spare parts using artificial intelligence: A case study of K-X tanks. *Mathematics* **2023**, *11*, 501. <https://doi.org/10.3390/math11030501>.
21. B Nathan, S.; Reddy, B.V.S.; Sujana, V.S.S.; Sastry, C.C.; Krishnaiah, J.; Jitpichitchai, S. A machine learning framework for long-term forecasting of spare part demand in end-of-life product scenarios. *Sci. Rep.* **2025**, *16*, 1394. <https://doi.org/10.1038/s41598-025-31171-2>.
22. Perera, A.A.; Senanayake, U.; Perera, H.N.; Rosienkiewicz, M. Intermittent demand forecasting of aircraft components through installed base forecasting: A case study. *Eksploat. Niezawodn. Maint. Reliab.* **2026**, *28*, 218545. <https://doi.org/10.17531/ein/218545>.
23. Zhao, J.; Wang, R.; Jing, Y.; Wang, Y.; Pan, C.; Tong, Y. Multi-phase demand modeling and simulation of mission-oriented supply chains using digital twin and adaptive PSO. *Mathematics* **2026**, *14*, 468. <https://doi.org/10.3390/pr14030468>.
24. Sun, M.; Ng, C.T.; Yang, L.; Zhang, T. Optimal after-sales service offering strategy: Additive manufacturing, traditional manufacturing, or hybrid? *Int. J. Prod. Econ.* **2024**, *268*, 109116. <https://doi.org/10.1016/j.ijpe.2023.109116>.
25. Gerrits, B.; Topan, E.; Van der Heijden, M.C. Operational planning in service control towers—heuristics and case study. *Eur. J. Oper. Res.* **2022**, *302*, 983–998. <https://doi.org/10.1016/j.ejor.2022.01.025>.
26. Sachs, F.E.; Helber, S.; Kiesmüller, G.P. Evaluation of unreliable flow lines with limited buffer capacities and spare part provisioning. *Eur. J. Oper. Res.* **2022**, *302*, 544–559. <https://doi.org/10.1016/j.ejor.2022.01.005>.
27. Dreyfuss, M.; Giat, Y. Optimal spares allocation to an exchangeable-item repair system with tolerable wait. *Eur. J. Oper. Res.* **2017**, *261*, 584–594. <https://doi.org/10.1016/j.ejor.2017.02.031>.
28. Hekimoğlu, M.; Van der Laan, E.; Dekker, R. Markov-modulated analysis of a spare parts system with random lead times and disruption risks. *Eur. J. Oper. Res.* **2018**, *269*, 909–922. <https://doi.org/10.1016/j.ejor.2018.02.040>.
29. Pacheco-Velázquez, E.A.; Robles-Cárdenas, M.; Juárez Ordóñez, S.; Dany Solís, A.E.; Cárdenas-Barrón, L.E. A heuristic model for spare parts stocking based on Markov chains. *Mathematics* **2023**, *11*, 3550. <https://doi.org/10.3390/math11163550>.
30. Ogura, N.; Kasuya, M.; Jin, L. Joint optimal policy for maintenance, spare unit selection and inventory control under a partially observable Markov decision process. *Mathematics* **2025**, *13*, 406. <https://doi.org/10.3390/math13030406>.

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