Mission Ready, If the Forecast Says So:
Improving MRO Demand Forecasting for U.S. Naval Aviation

by

Luis Pujol Sánchez de León

and

Dolev Starr

Submitted on May 8th, 2026, in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management
at the Massachusetts Institute of Technology

ABSTRACT

Demand forecasting for military Maintenance, Repair, and Overhaul (MRO) consumables is uniquely challenging: Parts are expensive, lead times are long, and a stockout can ground aircraft rather than result in a lost sale. This project partners with the Defense Logistics Agency (DLA) to improve forecast accuracy for U.S. Naval Aviation MRO consumables. The proposed framework clusters Stock Keeping Units (SKUs) by volume, volatility, and intermittency using K-means, then evaluates 25 forecasting models for each SKU, ranging from Naive baselines to deep learning architectures, and selects the winners based on Mean Absolute Error (MAE). Accuracy improvements over the Naïve Mean baseline range from 16% to 59%, depending on the cluster. Probabilistic forecasts from the top three models per cluster feed directly into a base-stock inventory policy, enabling planners to set order-up-to levels to meet a target service level. The study demonstrates that the cluster-then-forecast paradigm, previously validated in commercial settings, extends effectively to defense logistics and provides DLA with an operational tool to reduce both stockouts and excess inventory across the naval aviation supply chain.

Capstone Advisor: Dr. Ilya Jackson

Title: Research Scientist at the MIT Center for Transportation & Logistics

Table of Contents

1. Introduction

1.1. Background

The study builds on MIT CTL's research in demand forecasting (Ma et al., 2025) and extends its application to a new context: the national defense industry.

Our partnering organization is the Defense Logistics Agency (DLA), and the scope is limited to the United States (U.S) Naval Air Forces' Maintenance, Repair, and Overhaul (MRO) consumables. The Navy operates approximately 3,700 aircraft worldwide, across dozens of air bases and aircraft carriers (Hill Aerospace Museum, 2021). These units operate under varied pressures, including their role in the U.S. Military's ability to project force worldwide, their high cost of $20B annually (U.S. Naval Institute, 2025), and the extensive maintenance required to ensure aircraft can fly under any conditions. These pressures underscore the importance of accurate demand forecasting for Naval Aviation MRO. A poor forecast may have real-world consequences beyond a simple "lost sale." As shown in Figure 1, the U.S. Government Accountability Office (GAO, 2022) found that the mission-capable supply rate, the percentage of time an aircraft cannot execute its assigned missions due to the unavailability of a required repair part, decreased by an average of 13.5% between 2015 and 2020 for four aircraft sampled.

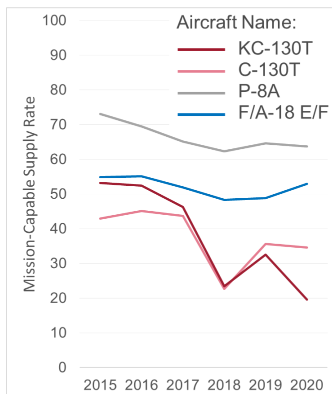

Figure 1: Mission-Capable Supply Rate Decrease Over Years. (GAO, 2022)

1.2. Motivation

Our partnering organization faces a challenge in accurately forecasting demand for MRO consumables. In this context, MRO consumables encompass the entirety of the materials required to keep the Navy's Aviation Fleet at the highest level of readiness.

The DLA navigates unique constraints in this regard, such as the extremely high cost of these MRO materials and the need to supply these parts to aircraft carriers at sea and to other forces globally. While a "standard" civilian organization may have the luxury of offsetting a poor demand forecast with higher inventory levels, the U.S. Government Accountability Office (GAO, 2000) noted that the Navy lacks this luxury due to monetary constraints. This lack of flexibility is compounded by the limited physical space aboard aircraft carriers and by the challenge of resupplying ships deployed at sea.

Our partner organization was interested in determining which forecasting model and methodology would best improve forecasting accuracy, thereby reducing stock-out costs and enhancing the fleet's readiness. It sought a Quantitative Supply Chain (QSC)-based model that could account for the wide range of constraints and variables, including forecasting up to two years ahead, lead-time constraints, and more.

1.3. Problem Statement and Key Questions

To address the challenges mentioned above, this research aimed to answer the following problem statement: "Which framework should the DLA adopt to improve its demand forecasting accuracy and inventory planning for its MRO consumables?"

The hypothesis posited that treating all Stock Keeping Units (SKUs) equally and applying a single forecasting methodology can lead to inaccuracies with severe downstream consequences. By contrast, the planned solution segmented the products into at least four clusters and applied a distinct forecasting methodology to each.

Therefore, the study formulated the following research questions:

1. How should SKUs be clustered to allow for differentiated forecasting? What are the optimal number and types of clusters?
2. Across classical, machine learning, and deep learning models, which forecasting methodology best fits each cluster based on accuracy performance?
3. How should demand uncertainty be quantified and translated into inventory planning recommendations?

1.4. Project Goals and Expected Outcomes

The project had two main goals. From the CTL research perspective, the goal was to demonstrate that the cluster-then-forecast approach to demand forecasting has practical applications in the defense industry. From DLA's perspective, the goal was to improve the accuracy of its forecast for MRO parts demand.

The expected outcomes included:

1. A multidimensional SKU segmentation based on demand volume, volatility, and intermittency.
2. Cluster-specific forecasting model recommendations with supporting accuracy metrics across 25 evaluated models.
3. Actionable inventory level recommendations for each SKU, grounded in probabilistic demand forecasts and a user-defined service level.

This project aligned the goals of CTL and the DLA as they shared the ultimate objective of developing a practical, data-driven framework to improve demand-forecasting accuracy and operational decision-making within the defense supply chain.

2. State of the Practice

Defense logistics is one of the most complex and high-stakes industries in the world. Parts must meet strict technical requirements, lead times may span years, and the consequences of both under- and over-stocking extend far beyond a lost sale. Demand forecasting sits at the center of this challenge, and yet it remains an underdeveloped discipline within this domain.

This section reviews the state of practice across three areas. First, the industry context covers defense logistics and the Naval Aviation maintenance structure. Second, the methodological context covers SKU clustering, traditional and machine learning forecasting models, and probabilistic forecasting. Third, we bring the two together by examining prior forecasting work in the defense industry and identifying the specific gap this study addresses.

2.1. Industry Context - Naval Aviation MRO Consumables

The importance of defense logistics has been unquestionable for millennia. As early as the fourth century B.C.E., a Greek philosopher and military leader noted, "For without provisions, neither general nor private is of any use" (Xenophon, 1921). In 1473, Vegetius (n.d.) echoed this sentiment by observing that shortages, rather than combat, more often destroy armies. Lastly, in the modern era, the U.S. Department of the Army (2024) recognizes sustainment as a mission-critical enabler that supports a commander's freedom of action and endurance.

The U.S. Department of Defense (2010) defines logistics as "Planning and executing the movement and support of forces." However, this simple definition does not capture the true scale of the field. As the Joint Staff (2011) codified, the true scale encompasses a diverse array of functional areas, including supply, maintenance, transportation, civil engineering, and more. The modern art of defense logistics encompasses not only these fields, multiple government agencies, and procurement programs spanning decades, but also the coordination of an industrial base of over 1 million workers across nearly 60,000 companies, with a combined spending of over $400 billion USD annually (Nicastro, 2024).

MRO in an aviation context encompasses operations such as inspection, parts replacement, sealant application, coating repair, and lubrication (Ceruti et al., 2019). This paper narrows that definition to the military context, adopting the Office of the Under Secretary of Defense for Acquisition and Sustainment (2018) definition, which classifies MRO consumables as supplies normally expended beyond recovery, and consumable repair parts as items used in the maintenance of an end item or weapon system.

2.2. Demand Forecasting

Hyndman and Athanasopoulos (2021) define forecasting as "predicting future outcomes based on the available information." Hyndman and Athanasopoulos (2021); Armstrong (2001); and Stojanović and Regodić (2017) all maintain that forecasting is a critical part of supply chain management, as it directly influences procurement, production, inventory management, budgeting, staffing, replenishment, and long-term capacity decisions.

Ernst & Young (2021) notes that changing customer preferences and geopolitical and macroeconomic volatility have made demand forecasting more challenging. Simultaneously, researchers such as Petropoulos et al. (2022) and Makridakis et al. (2022) have observed that methods to address these challenges, including machine learning and deep learning, have demonstrated superior performance to traditional approaches in handling complex demand patterns. Together, these trends are shaping industry adoption, with a McKinsey & Company survey (Destino et al., 2022) finding that 90% of supply chain executives have plans to modernize their forecasting tools.

The unique nature of MRO consumables makes accurate forecasting more challenging than for other types of inventory. While some consumables, such as oil used per flight, may be relatively easy to forecast, other segments of MRO inventory, primarily spare parts, depend on inherently random events, such as damage and part breakage, which may trigger demand (Bacchetti & Saccani, 2012).

2.2.1. SKU Clustering: Volume, Volatility, and Intermittency

Stojanović and Regodić (2017) trace the evolution of the ABC-XYZ segmentation method back to the 1950s, noting that a General Electric manager developed the ABC component to categorize SKUs based on a single criterion: demand volume. They explain that the XYZ component was added later to account for a second dimension: demand volatility.

In managing both MRO and spare parts inventories, intermittency is a key consideration that significantly complicates forecasting (Van der Auweraer et al., 2017). Unlike high-volume, stable goods, these inventory types are characterized by long stretches of zero-demand periods punctuated by periods of non-zero demand. In the early seventies, Croston (1972) proposed a forecasting method specifically for such demand periods. Nearly 30 years later, Syntetos and Boylan (2001) found that Croston's original formula exhibited an overestimation bias and developed a bias-adjusted modification to Croston's method. In 2005, Syntetos, Boylan, and Croston (Syntetos et al., 2005) compared these two methods and proposed four categories of demand patterns: erratic, lumpy, smooth, and intermittent. These four categories have since become the standard for demand forecasting.

2.2.2. Forecasting Models

To provide a comprehensive evaluation, our research compared 25 models spanning five categories: Baseline and Benchmark, Classical Statistical, Intermittent Specialists, Machine Learning, and Deep Learning. Beyond point forecasts, the study also incorporates probabilistic forecasting to quantify demand uncertainty. The following sections cover at least one representative model from each category to enable a technical comparison of the diverse approaches evaluated.

2.2.2.1. Naïve Mean

Naïve Mean assumes that the best predictor for future demand is simply the average of all past historical data. Naïve Mean assigns equal weight to every historical period, regardless of whether it was the most recent or 100 periods ago. This method creates a model that is stable and resistant to outliers but cannot capture trends or seasonality. In summary, older, potentially irrelevant data anchors the model down.

2.2.2.2. Exponential Smoothing

Hyndman and Athanasopoulos (2021) position exponential smoothing (ETS) as a balanced model between the two extreme methods, naïve and average, in time series analysis. While the naïve method assumes that the only important observation is the latest, the average method assumes that all observations are equally important, including those from years ago. This model assigns weights that decrease exponentially with age.

2.2.2.3. Croston's Method

Traditional methods, such as ETS or Naive Mean forecast, produce misleadingly low numbers when faced with a higher proportion of zero-demand periods. Croston's method (1972) aims to address this by constructing a model that can account for a large proportion of zero-demand periods. Croston's method decomposes a time series into two parts: the demand magnitude (the size of demand) and the demand interval (the number of time periods between demand occurrences).

2.2.2.4. Automated Trigonometric Box-Cox ARMA Trend Seasonal (AutoTBATS)

De Livera et al. (2011) developed the TBATS model (Trigonometric, Box-Cox, ARMA errors, Trend, and Seasonal components) to forecast time series with more complex patterns, such as multiple or frequent seasonality that classical models cannot account for. The major benefit of the Auto version of the model is that it automatically selects the best-performing configurations and weights, eliminating the need for manual tuning. The TBATS model consists of the following five components:

1. Trigonometric: Employs trigonometric pairs to model seasonal patterns
2. Box-Cox: Stabilizes variance to cause the dataset to more closely resemble a normal distribution
3. ARMA Errors: Captures the autocorrelation
4. Trend: Calculates the linear or dampened trend
5. Seasonal: Models the seasonal component

2.2.2.5. Random Forest

Breiman (2001) developed Random Forest, a machine learning model that uses ensemble learning to combine the predictions of many individual decision trees into a single, more robust forecast. Each decision tree is trained on a random subset of the data, and at each split, only a random subset of the features is available to the tree, ensuring each tree is unique. The final prediction is calculated by averaging the predictions of all the trees in the "forest." By averaging many different trees, Random Forest produces a robust model that minimizes the variance inherent in a single tree.

2.2.2.6. Decomposition Linear (DLinear)

Zeng et al. (2023) developed the DLinear (Decomposition Linear) model as a deep learning model specifically for time series forecasting. Decomposition models are a method in which a time series is broken down into typically three to four distinct components, such as trend, seasonality, and noise. DLinear pairs this with an additional single-layer linear step. After breaking down the time series into those components, it passes each one through an additional linear layer. In other words, the model predicts each component separately, allowing it to specialize in forecasting each component rather than forecasting the entire model concurrently.

2.2.2.7. Probabilistic Forecasting

Traditional point forecasts output a single expected value but fail to account for uncertainty, limiting their usefulness when forecasting items with high demand variability. Probabilistic forecasting fills the gap by estimating a range of future values and their associated likelihoods, as illustrated in Figure 2, thereby providing a more comprehensive view of variability risk. (Rodrigo & Ortiz, n.d.) Two factors drive our decision to shift towards the probabilistic forecasting approach. First, many conventional inventory models rely on the assumption that demand follows a normal distribution; however, these models fail to account for inventory that follows a 'lumpy' or 'intermittent' demand pattern. Secondly, classical inventory models aim to forecast a fixed mean, whereas probabilistic models allow us to target a flexible service level.

The selection of a probabilistic forecasting method depends on the underlying model type. Baseline and Classical models use Gaussian intervals derived from historical residuals; Intermittent Specialist models require Occurrence + Size Distributions to handle the high number of zero-demand periods; and machine learning models use Pinball loss Quantile Regression to quantify risk.

First, Gaussian Intervals estimate uncertainty by constructing a range around the point forecast, defined as $\hat{y}_{t+1} \pm k * \sigma_h$, where σh is the forecast standard deviation and k is the coverage factor corresponding to the desired service level. In this study, demand follows a Gamma distribution, and the interval range is calculated accordingly.

Second, Occurrence + Size Distribution models are a two-step probabilistic approach tailored to intermittent demand. The Occurrence component applies a Bernoulli Distribution to determine whether demand occurs in a given period, and the Size component estimates its magnitude using either a Poisson distribution for low-volume, discrete demand or a Gamma distribution for larger, continuous, non-negative patterns.

Third, quantile regression estimates each quantile independently by minimizing the Pinball Loss Function, which asymmetrically penalizes over- and under-predictions (Vermorel, 2012). For example, at the 90th percentile, under-predictions carry a heavier penalty, pushing the model to place approximately 90% of demand values below the forecast, as illustrated in Figure 3.

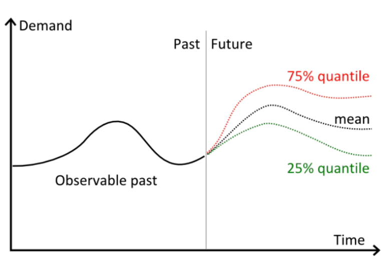

Figure 2: Illustration of Quantile Regression (Lokad, 2012)

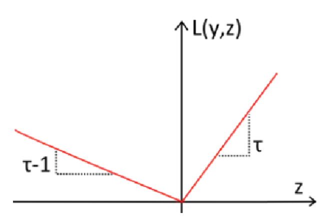

Figure 3: Pinball Graph and Formula (Lokad, 2012)

2.3. Demand Forecasting Within Naval Aviation MRO Consumables

A synthesis of issues was identified during our thorough review of reports primarily produced by the U.S. Government Accountability Office and the Department of Defense Inspector General. Table 1 illustrates the root causes and organizational culture that may heavily contribute to an inaccurate forecast.

Source Title	Summary of Relevant Findings
Evaluation of the Spare Parts Onboard U.S. Navy Ships in the Indo-Pacific Region	Onboard nine sampled Navy ships, spare parts inventory accuracy ranged from 83-94.9%, not meeting the stated goal of 98% in every case. (Inspector General of the U.S. Department of Defense, 2025).
Management Actions Needed to Improve the Cost Efficiency of the Navy's Spare Parts Inventory	From 2004 to 2007, the Navy had an average of $18.7B in secondary (spare parts) inventory, with $3.75 billion per year considered excess (U.S. Government Accountability Office, 2008).
F-35 Aircraft: DOD Should Assess and Update Its Engine Sustainment Strategy to Support Desired Outcomes	The D.O.D. and the F-35 engine's prime contractor did not agree on a model for forecasting spare parts (U.S. Government Accountability Office, 2022).
Navy Inventory: Parts Shortages Are Impacting Operations and Maintenance Effectiveness	"Accurately forecasting the demand for parts is difficult because of the large number of variables that affect demand, including flying hour frequency and environment" (U.S. General Accounting Office, 2001).
Audit of the Reliability of Army Spare Parts Forecasts Submitted to the Defense Logistics Agency	The Army had only a 20% forecast accuracy rate for spare parts in FY 2011. This report also found that inaccurate forecasts caused unnecessary excess inventory costs and stockouts for other parts simultaneously. Lastly, they found multiple instances in which personnel were unable to recreate submitted forecasts (Inspector General of the U.S. Department of Defense, 2023).
Forecasting Critical Aircraft Launch & Recovery Equipment (ALRE) Components Demand	The research attempted to optimize the forecasting model for a critical part used aboard aircraft carriers, which exhibits highly erratic demand patterns. It concluded that grouping by demand criteria and then identifying the most accurate model for each group is a possible solution (Laryea et al., 2018).
Military Readiness: Actions Needed to Further Implement Predictive Maintenance on Weapon Systems	"Proactive and accurate planning (of predictive maintenance) is necessary to ensure the timely availability of spare parts for the maintenance process, especially since the acquisition lead-time for spare parts can range from days to years" (U.S. Government Accountability Office, 2022). Thus, one possible strategy to reduce variability in spare parts demand is to ensure that predictive maintenance is fully implemented.
Defense Logistics Agency Aviation Retained Excessive V-22 Osprey Spare-Parts Inventory	*"For 22 of 53 spare parts that we nonstatistically selected for review, DLA Aviation retained spare-parts that was excessive."* In another case, DLA ordered 21 door assemblies, with an average annual demand of 1, thereby creating 21 years of inventory (Inspector General of the U.S. Department of Defense, 2015).
F-35 Program: DOD Needs Better Accountability for Global Spare Parts and Reporting of Losses Worth Millions	"The F-35 Joint Program Office was unable to provide the cost, total quantity, and locations of spare parts in the global spares pool, and continues to rely on the prime contractors' records for this information" (U.S. Government Accountability Office, 2023).

Table 1: Summary Table of Relevant Government Documents.

2.3.1. Literature Gap: Defense and Aviation MRO Consumables Specific Forecasting Work

As Choi and Suh (2020) argue, defense and aviation MRO forecasting present unique forecasting challenges that are underserved by conventional commercial forecasting methods. Both civil and military air fleets rely on spare parts with infrequent, highly variable demand patterns (Heijenrath & Verhagen, 2023), and the U.S. Government Accountability Office (2007) has highlighted that spare-part acquisition is subject to significant lead-time variability. Inaccurate demand forecasting, compounded by the exceptionally high costs of spare parts for Naval Aviation aircraft (U.S. Government Accountability Office, 2000), has created a situation in which meeting the strict readiness requirements is possible only through infusions of billions of dollars in additional funding for spare parts. (U.S. Government Accountability Office, 2003).

Zsidisin et al. (2020) identified several systemic causes of the research gap in defense logistics. They note that most research is done at civilian institutions, which naturally focus on profit-driven, commercial supply chains. The lack of literature integration may be due to a bias towards commercial supply chains, poor readership interest, and a perceived lack of citation potential. Furthermore, they theorize that the gap may stem from perceptions that the defense and commercial fields are incomparable, the interdisciplinary nature of defense logistics, and the view that defense logistics is a domain of practice rather than theory.

3. Methodology

Our study followed a six-step methodology, as illustrated in the setup flow in Figure 4. First, to address the lengthy approval timelines for sensitive data at the DLA, we generated a synthetic dataset to validate our proposed framework. Second, we categorized the SKUs using K-means clustering based on their volume, volatility, and intermittency characteristics. Third, we applied 25 different forecasting models to all SKUs. Fourth, we compared the performance of these models and created a ranking for each cluster. Fifth, we conducted probabilistic forecasting to account for uncertainty. Sixth, in the inventory planning phase, we used the forecast results to recommend a base stock level.

We detailed the daily operations flow in the results section, but the bottom row of Figure 4 provides context for our methodology choices. Once the setup is complete, which we recommend doing quarterly, semiannually, or even annually, the model will already know the clusters and top-performing models. Therefore, updating the forecasts and inventory plans requires minimal computational effort.

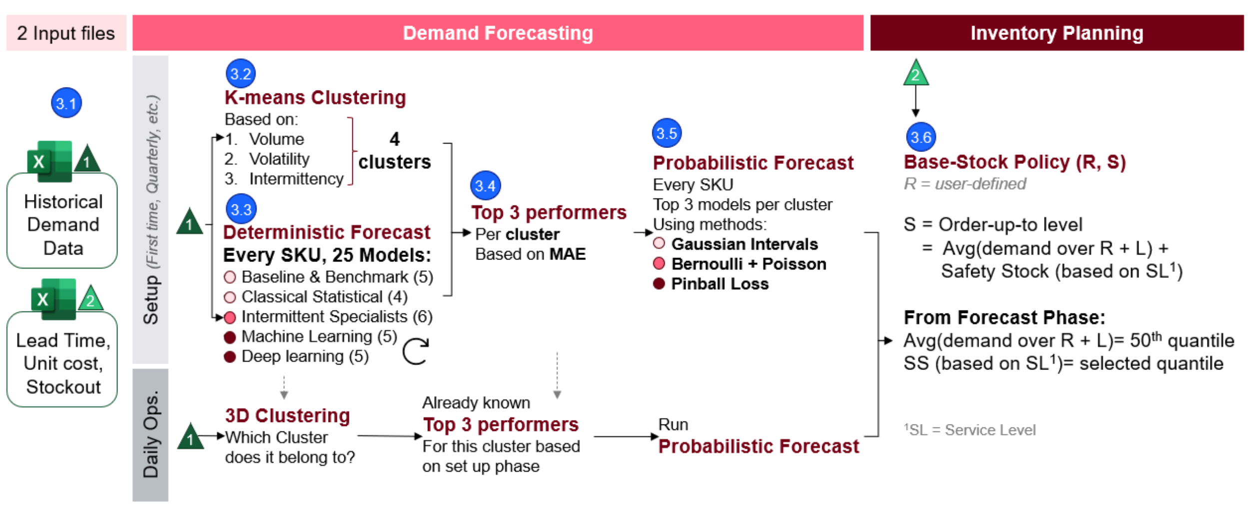

Figure 4: Methodology Summary Flow.

3.1. Synthetic Data Generation

The synthetic data generation involves three steps. First, we created six SKU behavioral types: smooth-stable, high-seasonal, intermittent-lumpy, trend-up, trend-down, and noisy, assigning varying frequencies to each type to reflect realistic scenarios. For instance, smooth-stable patterns account for 30% of SKUs, while intermittent-lumpy patterns account for only 5%. Each profile includes controlled parameters for volume, volatility, trend, seasonality, and intermittency; for example, the “noisy” profile is defined by its high volatility coefficients.

Second, a generator function receives these parameters to synthesize weekly SKU demand. As illustrated in Figure 5, it involves three sub-steps. A deterministic demand curve (blue) is constructed with the base volume, trend, and seasonality. Randomness (red) is added to simulate real-world behavior. An intermittency component (black) is added, assigning zero demand to selected weeks, before rounding to whole units. Third, the model iterates to produce a seven-year-long weekly demand history for 500 SKUs. A fixed random seed is employed to ensure the dataset is reproducible for future validation.

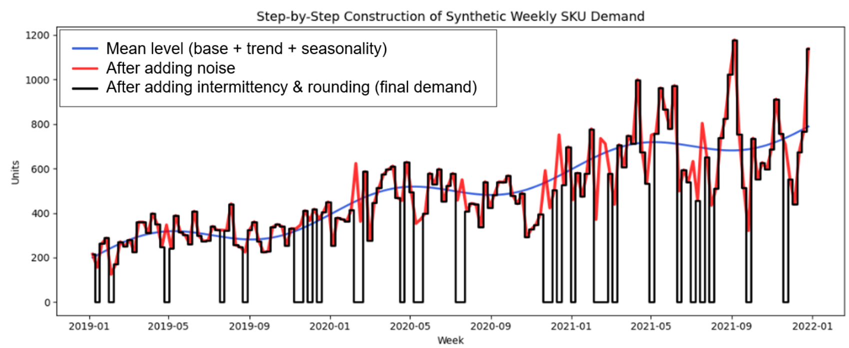

Figure 5: Example Construction of Synthetic Weekly Demand for One SKU with a Shortened Timeline (Three Years).

3.2. Clustering Rationale and Process

Drawing on the State of the Practice section, this study developed a multidimensional clustering method based on volume, volatility (CV), and intermittency. This approach integrated the ABC-XYZ segmentation with the four intermittency categories as originally proposed by Syntetos, Boylan, and Croston. This 3x3x4 matrix theoretically yields 36 distinct SKU categories, which is operationally impractical. To preserve these three dimensions while ensuring operational feasibility, we used K-means clustering to generate four cohorts based on demand behavior.

We selected a cluster count of k=4 as it ensured an even distribution of SKUs across groups, as illustrated in Figure 6. This configuration enabled specialized forecasting without adding excessive planning complexity. If a cluster count of k=3 were used, one group would contain approximately 75% of SKUs, thereby limiting the benefits of specialized forecasting.

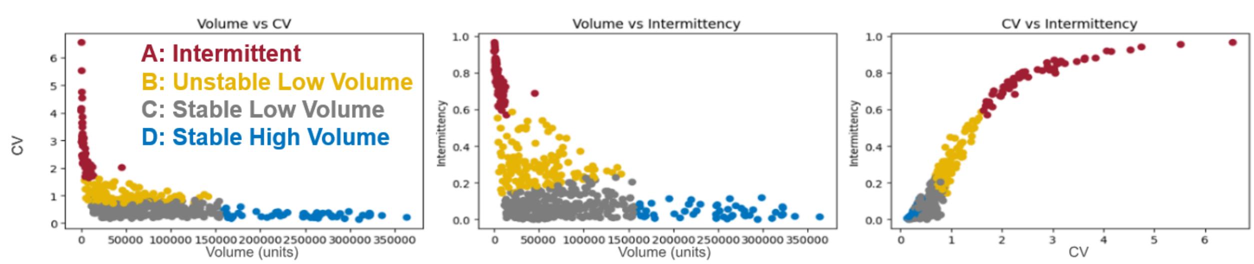

Figure 6: Four Clusters Comparing Volume, Volatility (CV), and Intermittency.

3.3. Deterministic Forecasting

Our study evaluated 25 forecasting models, organized into five categories by model type, as illustrated in Figure 7. Each of these models had different computing power and data history requirements, and some proved more suitable for specific demand patterns than others. This structure ensured broad methodological coverage, spanning from Naive Moving Average, which debuted in the early 1900s, to NLinear, which was introduced at Zhejiang University in China less than three years ago, ensuring that only the best model won the forecasting competition.

Each of the 25 models was executed for each of the 500 SKUs, using a 70-30% training-test split. We were able to allocate 30% to testing, rather than the more conventional 20-25%, because we generated enough synthetic data. This process produced forecasts at weekly intervals, average computational times per model under specific computer specifications, and SKU-model performance metrics, which are elaborated upon in the following subsection.

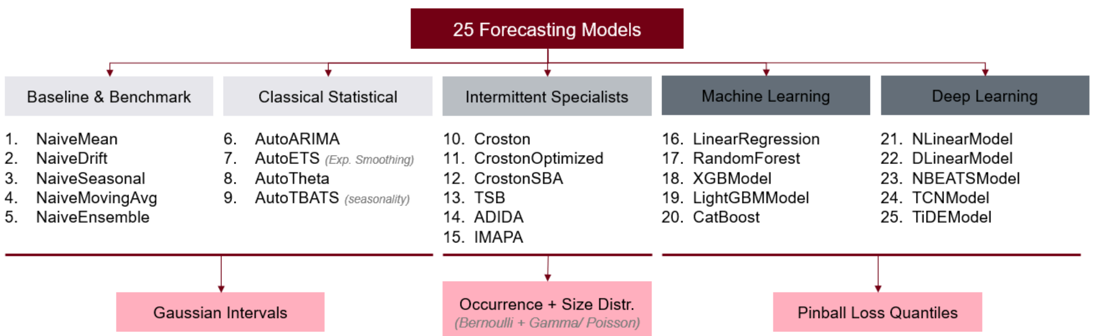

Figure 7: Overview of 25 Forecasting Models by Category and Method Used for Probabilistic Forecasting.

3.4. Model Accuracy Competition

The model accuracy competition integrates the previous phases of clustering and deterministic forecasting to assess the top three performing models within each cluster. Leveraging both the per-SKU accuracy metrics from the forecasting phase and the cluster assignments from the segmentation phase, model performance was aggregated by averaging across SKUs within each cluster to identify the best-fit model per cluster.

Table 2 presents the four accuracy metrics produced by the framework. Given that model rankings varied by metric, we adopted Mean Absolute Error (MAE) as the primary decision criterion for two reasons. First, compared to Mean Squared Error (MSE) and Root Mean Square Error (RMSE), MAE is easier to interpret because it retains the same units and penalizes errors equally. Second, unlike Mean Absolute Percentage Error (MAPE), MAE does not divide by the actual value, which resolves the MAPE-related concerns for intermittent data with frequent zero actual demand. The remaining metrics were calculated and displayed to provide a complete diagnosis.

MSE (Mean Squared Error)	$\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$	RMSE (Root Mean Square Error)	$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}$
MAE (Mean Absolute Error)	$\frac{1}{T}\sum_{t=1}^{T}\left\|y_t - \hat{y}_t\right\|$	MAPE (Mean Absolute Percentage Error)	$100\cdot\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{y_t - \hat{y}_t}{y_t}\right\|$

Table 2: Accuracy Metrics and Their Formulas.

3.5. Probabilistic Forecasting

Having identified the top three-performing models within each cluster, we generated probabilistic forecasts for two reasons. First, point forecasts alone fail to capture demand uncertainty, which is especially important in defense logistics, where over- or understocking has more serious consequences than lost revenue. Second, probabilistic outputs are a prerequisite for the framework's final phase: translating forecasts into inventory planning recommendations.

As Figure 7 illustrates, each category of forecasting model was assigned one of three probabilistic forecasting methods: Gaussian Intervals; Occurrence and Size Distribution Methods; and Pinball Loss Quantiles. To manage computational cost and output speed, probabilistic forecasting was applied only to the top three-performing models in each cluster.

Across all forecasting model families and probabilistic forecasting methods, the 5th-95th percentile quantiles were used by default, as illustrated in Figure 8, which shows both the point and probabilistic forecasts for one SKU. Different SKUs require different inventory levels; thus, we implemented a flexible code that allows the end user to easily adjust the quantile bounds to make them narrower or wider.

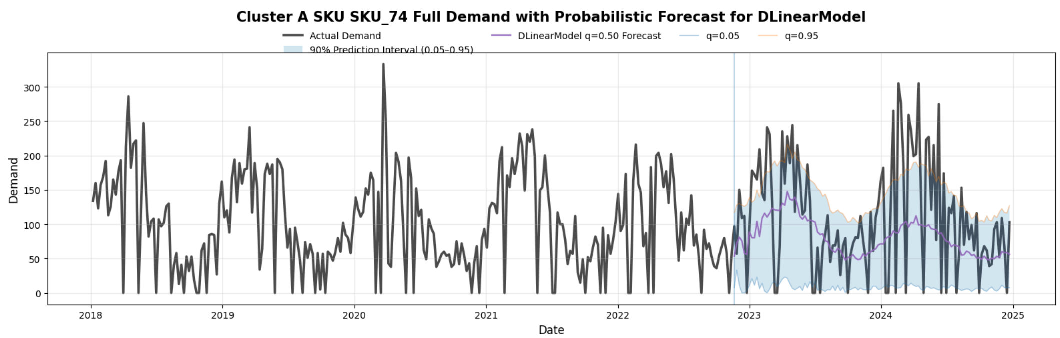

Figure 8: Illustrative Example of SKU-Level Demand with Probabilistic Forecast Interval.

3.6. Inventory Planning: Base Stock Levels

The intent of this section is two-fold. First, forecasts should not be used as the sole input for decisions. Decisions, from tactical replenishment orders to strategic procurement contracts, should include an inventory plan. By extending the framework to this final phase, this study delivers an actionable output rather than stopping at the forecasting stage, where most prior work does. (Goltsos et al., 2022)

Second, it gives our DLA contact the operational grounding he needs to advocate for probabilistic forecasting internally, and demonstrating its operational benefit is more persuasive than citing methodology alone. As Section 2.2.2.7 establishes, the normality assumption that conventional inventory planning defaults to does not hold for intermittent demand. By showing how probabilistic outputs lead to better-informed stock-level decisions for MRO consumables, this phase gives him a concrete, decision-level argument to present to leadership.

We implemented a base-stock policy using an order-up-to-level (R, S) model. Using the following formula, which incorporates probabilistic forecasts as inputs, we calculate the base stock level (S) for each SKU. After each review period (R), the planner places orders to reach this base stock level. Figure 9 presents a sawtooth graph of inventory, illustrating how inventory position and on-hand inventory fluctuate in response to demand and orders.

$$S = Cycle\ Stock + Safety\ Stock$$

$$Cycle\ Stock = Average(forecasted\ demand\ over\ R + L) = 50th\ percentile\ forecast$$

$$Safety\ Stock = 95th - 50th\ percentile\ forecast\ (or\ desired\ service\ level)$$

$$where;\ R = Review\ Period\ (user\ defined),\ L = Leadtime$$

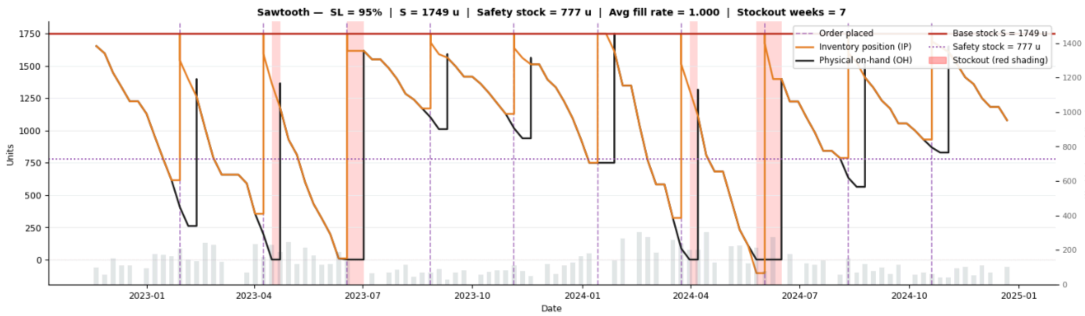

Figure 9: Inventory Sawtooth Graph Example.

4. Results

Our results have validated the core hypothesis: Treating all SKUs as a homogeneous group when forecasting yields poorer results than our cluster-then-forecast approach. As shown in Figure 10, by segmenting SKUs into four clusters, we identified the forecasting model that best captured the unique characteristics of each demand pattern.

Cluster	Model	% Accuracy Improvement vs. NaiveMean
A: Intermittent	AutoTBATS	58.7%
B: Unstable Low Volume	D-Linear	16.3%
C: Stable Low Volume	D-Linear	23.8%
D: Stable High Volume	RandomForest	21.3%

Figure 10: Optimal Forecasting Model and Accuracy Improvement by Demand Cluster.

4.1. Key Findings and Model Selection

Our data confirmed that different demand patterns require different models to forecast accurately. Most notably, for Cluster A (Intermittent Demand), a demand pattern characterized by random, intermittent spikes of high demand, the Classical Statistical model AutoTBATS delivered a nearly 60% improvement over the baseline model. Intermittent demand tends to follow cycles, so the model's strong seasonal components can capture the seasonal patterns arising from those cycles. Additionally, this finding proves that newer or more advanced models, which also require more resources, are not necessarily better than more classical models.

Interestingly, D-Linear, a relatively simple linear model, punched above its weight class and emerged as the winner for both Cluster B (Unstable Low Volume) and Cluster C (Stable Low Volume). This result may justify consolidating those clusters into a single "Low Volume" cluster; however, we maintained the original four-cluster framework for consistency and alignment with existing operational requirements. Retaining this framework allowed for a more granular view of the differences between the two clusters, even though the underlying mathematical model is the same.

In the final cluster, Cluster D (Stable High Volume), the Random Forest Model emerged as the winner, with a 21.3% improvement over the Naive Mean Baseline Model. This result is driven by the fact that high-volume SKUs provide sufficient historical observations for an ensemble model to learn complex, non-linear demand patterns that simpler models cannot capture. This also explains why Random Forest did not win in the lower-volume clusters, where data scarcity limits the advantage of more data-hungry approaches.

4.2. Mitigating Inventory Risk Through Probabilistic Forecasting

To fully understand the risk posed by poor or inaccurate forecasting, our research moved beyond static point forecasts to provide probabilistic forecasts, i.e., the statistical range of likely demand. In the high-stakes world of spare parts inventory management, knowing the forecasted demand is not enough; understanding the uncertainty helps prevent catastrophic failure.

Probabilistic demand forecasting provides decision-makers in the field with a range of tools that, combined with their on-the-ground experience, equip them to make better decisions.

By integrating probabilistic forecasting, we equip decision-makers with the tools to:

- Prevent stockouts: Safety Stock levels can be set using different demand percentiles, based on SKU criticality, rather than a simple average.
- Prioritize by criticality: Inventory buffers can be set based on factors such as SKU importance, holding costs, and more. This flexibility is important when financial resources are constrained.
- Data-Driven resilience: Probabilistic forecasting enables planners to consider dozens of scenarios, further strengthening their decision-making capabilities, especially in the volatile world of Naval Aviation.

4.3. Implementation Guidance

Based on the results of our research, we recommend adopting our cluster-then-forecast framework and have developed an interactive tool to ease this adoption. Our tool transforms our model from a theoretical framework into an operational asset, enabling easy, dynamic, and flexible SKU clustering, forecasting, and inventory planning.

Our tool is designed to give any user the full computational power and flexibility of a wide selection of forecasting modes. The workflow has three entry points, and a video demo is available at this link. This demonstration aligns with the methodology previously illustrated in Figure 4.

- **New Run/Setup:** This is the most comprehensive and computationally intensive function and should therefore be performed only periodically. Our recommended frequency is quarterly, semiannually, or annually.
  - Step 1: Data Input. Enter historical demand data at the SKU level or desired granularity.
  - Step 2: Configuration. To provide flexibility, the user can select the data interval (weekly or monthly), the number of clusters (four by default), the accuracy metric (MAE by default), the desired service levels, and whether to run the inventory phase (which requires an additional input file).
  - Step 3 (optional): Enter inventory parameters. This is the second and final input file required; it should include the review period, lead time, Minimum Order Quantity (MOQ), unit cost, holding cost, and stockout penalty at the SKU level.
  - Step 4: Execute the model. The user clicks "run" to start the code execution. The running times can vary, but for the dataset we used, which contains 500 SKUs and spans seven years of weekly demand, the process required 21 hours to complete.
  - Step 5: Results. The outputs are discussed after the two subsequent entry points, as they are common to both.
- **Update Existing Run**: This step is performed more frequently, typically weekly or monthly. Since the initial setup has already been completed, the model is aware of each SKU's cluster and its top-performing models. As a result, it forecasts only the top three performers in each cluster.

This approach ensures the most up-to-date forecast and inventory recommendations while saving the computational power and time required for the model competition.

- **LLM:** Our tool features a Chat-type interface powered by a Large Language Model (LLM). It is a proof of concept that understands natural-language questions, finds the relevant information in the 10 downloadable output files that the tool creates, and responds in natural language.
    - Users can ask questions in natural language, such as, "What is the 90th percentile forecast for SKU_1?" "How much should I order for SKU_1?", or "How much will I save if I have a service level of 90% instead of 95%?"
    - The chatbot connects to ChatGPT 4.0 mini via an Application Programming Interface (API), allowing it to answer any questions that are directly present in or easily derived from the result files. Its guardrails ensure that it queries only those files and always states its assumptions, calculations, and the source file.
    - Our LLM features demonstrate a strong proof of concept; transitioning to a production environment, particularly one integrated into any defense application, will necessitate rigorous security protocols.

Regardless of the original entry point, each run automatically generates a multi-layered data visualization of the results.

- **Clustering Diagnostics**: Cluster scatter plots display Volume vs. CV, Volume vs. Intermittency, and CV vs. Intermittency, as shown previously in Figure 6.
- **Graphs and Tables**
    - For the forecasting section, the tool displays the top three models for each SKU, including probabilistic ranges (as shown previously in Figure 8), along with downloadable tables containing all relevant information from the forecasting competition.
    - In the inventory section, the tool presents the sawtooth inventory graph, as previously shown in Figure 9, for each SKU, the top three models, and various service level combinations. Downloadable tables are also part of the output.
- **Financial Impact:** Our tool can calculate the dollar value saved compared to the baseline model and provide a service-level agreement trade-off analysis, making it easy to compare the cost of risk minimization.

Lastly, a key takeaway is the trade-off between performance and computational costs. Therefore, we recommend excluding the N-BEATS and AutoARIMA models from the forecasting competition. Neither of these models achieved a performance ranking higher than third place, yet they required significantly more runtime, taking 1300% and 623% longer than the third most computationally demanding model (AutoTBATS), respectively. Table 3 presents the average running time per model for each SKU, using the following computational specifications in Google Colab: CPU, Python 3 runtime, and 12.7GB of RAM.

N-BEATS: 75.8 secs	AutoARIMA: 39.2 secs	AutoTBATS: 5.42 secs
TCN: 4.98 secs	TiDE: 2.94 secs	CatBoost: 2.25 secs
DLinear: 1.6 secs	NLinear: 1.6 secs	XGBoost: 1.5 secs
All Other Models: Below 1 sec		

Table 3: Average Running Time per Model per SKU.

5. Recommendations

The results of this study have paved the way for the DLA and its partners. Our cluster-then-forecast approach, validated on synthetic data, delivers accuracy improvements across all demand profiles and yields actionable inventory planning recommendations. This section consolidates the findings of our study into a set of recommendations for the DLA.

5.1. Adopt the Cluster-Then-Forecast Methodology

The primary recommendation of this study is that the DLA replace its current forecasting method with the cluster-then-forecast framework that we established and validated. Our findings confirmed that segmenting SKUs through clustering and tailoring forecasting models to each cluster afterward is both viable and effective for the defense sector.

As we established in the results section, applying a single forecasting model to all SKUs yields substantially lower accuracy than the cluster-then-forecast approach, which assigns an optimal model to each cluster. We recorded accuracy improvements of 16.3%, 21.3%, 23.8%, and 58.7% over the Naïve Mean baseline model. These accuracy improvements represent not only a number, but a fundamental shift in readiness. A more accurate forecast leads to fewer shortages, fewer overstocks, millions of dollars in taxpayer money saved, and, most importantly, more aircraft ready to fly.

5.2. Prioritize Probabilistic Forecasts

Our secondary recommendation is that the DLA implement our tool to shift from point-based to probabilistic forecasting. As discussed in section 2.2.2.7, probabilistic forecasts enable us to incorporate uncertainty, a critical feature when forecasting in austere environments, such as the U.S. Navy. In the commercial world, uncertainty can be offset by inventory buffers, expediting orders, or backorders, but none of those options are available for naval aircraft at sea, which are constrained by a stricter budget, the limited space on board a naval ship, or the lead times for parts that may span months to years.

Probabilistic forecasting closes this gap. A probabilistic model quantifies the range of demand outcomes and their likelihoods, enabling demand planners to set inventory levels that align with a target service or readiness level. For more critical SKUs, a higher quantile (service level) can be designated to minimize stockout risk. For less critical SKUs, a lower quantile can be designated in order to reduce inventory holding costs. This level of flexibility, provided by probabilistic forecasting, gives the DLA a policy lever to maximize readiness at minimum cost.

5.3. Pilot the Framework on Real Demand Data

Our study validated the cluster-then-forecast framework using synthetically generated demand data, a necessary decision given the sensitivity of DLA demand data. While this dataset was carefully constructed to reflect the characteristics of real demand data (including volume, volatility, and intermittency) it still remains synthetic. Therefore, we recommend that the DLA conduct a pilot using a sample of real-world, historical demand data.

This study should span several years, ideally with as much demand as possible, to ensure each model type has a fair chance of capturing as many features as possible when forecasting. Particular attention should be paid to the SKUs in Cluster A (Intermittent), as this family of SKUs demonstrated the most complex demand pattern.

5.4. Full Integration With Systems and Processes in the DLA

While our interactive tool is designed to operate independently, it is a proof of concept, and the greatest strategic value will be achieved if it is integrated into the DLA's existing demand-forecasting and inventory-planning workflows. The DLA must ensure that data input requirements, such as historical demand and inventory parameters, are automatically fed into the tool from DLA systems, while maintaining a single source of truth for all data inflows. This eliminates manual labor and ensures that updating the demand forecast weekly or monthly is easy, hassle-free, and sustainable.

To avoid the internal resistance that often plagues change management, ease of use must be a core requirement. Therefore, the DLA must address this by ensuring the output files generated by our tool are compatible with the DLA's existing systems. This compatibility will build the bridge that ensures our tool is both utilized and easily integrated into daily operations without disruption.

5.5. Computation Power, Time, and Cost Management

The full model competition, consisting of 25 models run 500 times for each SKU, required 21 hours of runtime on our current hardware configuration (CPU, Python 3 runtime, and 12.7GB of RAM). This level of computational demand is appropriate for a quarterly, semiannual, or annual setup phase but is prohibitive for routine use. Therefore, as documented in Table 3, we recommend excluding the N-BEATS and AutoARIMA models from the competition, as they only achieved third place by marginal differences in any cluster and had runtimes substantially longer than those of the next most computationally demanding model.

References

Armstrong, J.S. (Ed.). (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic.

Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, *40*(6), 722-737. https://doi.org/10.1016/j.omega.2011.06.008

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. https://link.springer.com/article/10.1023/a:1010933404324

Ceruti, A., Marzocca, P., Liverani, A., & Bil, C. (2019). Maintenance in aeronautics in an Industry 4.0 context: The role of Augmented Reality and Additive Manufacturing. *Journal of Computational Design and Engineering*, *6*(4), 516–526. https://doi.org/10.1016/j.jcde.2019.02.001

Choi, B., & Suh, J. H. (2020). Forecasting spare parts demand of military aircraft: Comparisons of data mining techniques and managerial features from the case of South Korea. *Sustainability*, *12*(15), 6045. https://doi.org/10.3390/su12156045

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, *23*(3), 289-303. https://doi.org/10.1057/jors.1972.50

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513-1527. https://doi.org/10.1198/jasa.2011.tm09771

Department of Defense. (2010, November 8). *Department of Defense Dictionary of Military and Associated Terms* [JP 1-02]. Joint Chiefs of Staff. https://edocs.nps.edu/2014/December/jp1_02.pdf

Department of the Army. (2024, August). *Sustainment Operations (FM 4-0)*. Central Army Registry (CAR). https://rdl.train.army.mil/catalog-ws/view/100.ATSC/0D7107B6-A6B7-445A-B598-253BB85CC28A-1308676390333/FM4_0.pdf

Destino, M., Fischer, J., Müllerklein, D., & Trautwein, V. (2022, February 9). *To improve your supply chain, modernize your supply-chain IT*. McKinsey & Company. https://www.mckinsey.com/capabilities/operations/our-insights/to-improve-your-supply-chain-modernize-your-supply-chain-it

Ernst & Young. (2021, February 3). *Why companies need to step up their forecasting game*. Ernst & Young. https://www.ey.com/en_gl/alliances/why-companies-need-to-step-up-their-forecasting-game

Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. (2022). Inventory – forecasting: Mind the gap. *European Journal of Operational Research*, *299(2), 397–419*. https://doi.org/10.1016/j.ejor.2021.07.040

Heijenrath, L. M., & Verhagen, W. J.C. (2023). The influence of repair quality on aircraft spare part demand variability. *Aerospace*, *10*(8), 731. https://doi.org/10.3390/aerospace10080731

Hill Aerospace Museum. (2021, October 13). *Air Force vs Navy: Aircraft Comparisons*. Hill Aerospace Museum. Retrieved February 19, 2026, from https://www.aerospaceutah.org/air-force-vs-navy-aircraft-comparisons/

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice* (3rd ed.). OTexts. https://books.google.com/books?id=_bBhDwAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Inspector General of the U.S. Department of Defense. (2015). *Defense Logistics Agency Aviation Retained Excessive V‑22 Osprey Spare‑Parts Inventory (DODIG-2015-136)*. U.S. Department of Defense. https://media.defense.gov/2015/Jun/24/2001713513/-1/-1/1/DODIG-2015-136.pdf

Inspector General of the U.S. Department of Defense. (2023). *Audit of the Reliability of Army Spare Parts Forecasts Submitted to the Defense Logistics Agency (DODIG‑2023‑123)*. U.S. Department of Defense. https://media.defense.gov/2023/Sep/21/2003305897/-1/-1/1/DODIG-2023-123.PDF

Inspector General of the U.S. Department of Defense. (2025). *Evaluation of the Spare Parts Onboard U.S. Navy Ships in the Indo-Pacific Region (Report No. DODIG-2025-100)*. U.S. Department of Defense. https://media.defense.gov/2025/May/15/2003715601/-1/-1/1/DODIG-2025-100_REDACTED_FINAL_SECURE.PDF

The Joint Staff. (2011). *Joint Logistics Lexicon*. Joint Chiefs of Staff. https://www.jcs.mil/portals/36/Documents/052311152615_Joint_Staff_J4_Joint_Logistics_Lexicon.pdf

Laryea, M., Coleman, D., & Grimes, J. (2018). *Forecasting Critical Aircraft Launch & Recovery Equipment (ALRE) Components Demand (NPS-LM-19-017)*. Naval Postgraduate School. https://dair.nps.edu/bitstream/123456789/2319/1/NPS-LM-19-017.pdf

Ma, B. J., Jackson, I., Huang, M., Villegas, S., & Macías-Aguayoa, J. (2025). A data-driven and context-aware approach for demand forecasting in the beverage industry. *International Journal of Logistics Research and Applications*, 1-28. https://doi.org/10.1080/13675567.2025.2566806

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International journal of forecasting*, *38*(4), 1346-1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Nicastro, L. A. (2024). *The U.S. Defense Industrial Base: Background and Issues for Congress*. Congressional Research Service. https://www.congress.gov/crs-product/R47751

Nikolopoulos, K. (2021). We need to talk about intermittent demand forecasting. *European Journal of Operational Research*, *291*(2), 549-559. https://doi.org/10.1016/j.ejor.2019.12.046

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2018). *DoD Supply Chain Materiel Management Procedures: Demand and Supply Planning*. Washington Headquarters Services (WHS). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/414001m/414001m_vol02.pdf

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv*. https://doi.org/10.48550/arXiv.1905.10437

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., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Oliveira, F. L. C., Baets, S. D., Dokumentov, A., … Fiszeder, P. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3), 705-871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Robus, S., Kőműves, Z. S., & Walter, V. (2022). An overview of performance measurement for demand forecasting based on artificial neural networks. *Regional and Business Studies*, *14*(2), 5-20. https://doi.org/10.33568/rbs.4434

Rodrigo, J. A., & Ortiz, J. E. (n.d.). *Probabilistic Forecasting: Quantile Regression*. Skforecast Documentation. https://skforecast.org/0.17.0/user_guides/probabilistic-forecasting-quantile-regression

Stojanović, M., & Regodić, D. (2017). The significance of the integrated multicriteria ABC-XYZ method for the inventory management process. *Acta Polytechnica Hungarica*, *14*(5), 29-48. https://acta.uni-obuda.hu/Stojanovic_Regodic_76.pdf

Syntetos, A. A., & Boylan, J. E. (2001). On the bias of intermittent demand estimates. *International journal of production economics*, *71*(1-3), 457-466. https://doi.org/10.1016/S0925-5273(00)00143-2

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495-503. https://doi.org/10.1057/palgrave.jors.2601841

U.S. Department of the Navy. (2010). *Naval Operations Concept*. Federation of American Scientists. https://irp.fas.org/doddir/navy/noc2010.pdf

U.S. General Accounting Office. (2001). *Navy Inventory: Parts Shortages Are Impacting Operations and Maintenance Effectiveness (GAO-01-771)*. U.S. General Accounting Office. https://www.gao.gov/assets/gao-01-771.pdf

U.S. Government Accountability Office. (2000). *Prices of Navy Aviation Spare Parts Have Increased*. https://www.gao.gov/assets/gao-01-23.pdf

U.S. Government Accountability Office. (2003). *Defense inventory: Navy logistics strategy and initiatives need to address spare parts shortages (GAO-03-708)*. U.S. Government Accountability Office. https://www.gao.gov/assets/gao-03-708.pdf

U.S. Government Accountability Office. (2007). *Defense inventory: Opportunities exist to improve the management of DOD's acquisition lead times for spare parts (GAO-07-281)*. U.S. Government Accountability Office. https://www.gao.gov/assets/gao-07-281.pdf

U.S. Government Accountability Office. (2008). *Management Actions Needed to Improve the Cost Efficiency of the Navy's Spare Parts Inventory (GAO-09-103)*. U.S. Government Accountability Office. https://www.gao.gov/assets/gao-09-103.pdf

U.S. Government Accountability Office. (2022). *Actions Needed to Address Persistent Sustainment Risks*. https://www.gao.gov/assets/gao-22-104533.pdf

U.S. Government Accountability Office. (2022). *F-35 aircraft: DOD should assess and update its engine sustainment strategy to support desired outcomes (GAO-22-104678)*. U.S. Government Accountability Office. https://www.gao.gov/assets/gao-22-104678.pdf

U.S. Government Accountability Office. (2022). *Military readiness: Actions needed to further implement predictive maintenance on weapon systems (GAO-23-105556)*. U.S. Government Accountability Office. https://www.gao.gov/assets/gao-23-105556.pdf

U.S. Government Accountability Office. (2023). *F-35 program: DOD needs better accountability for global spare parts and reporting of losses worth millions (GAO-23-106098)*. U.S. Government Accountability Office. https://www.gao.gov/assets/gao-23-106098.pdf

U.S. Naval Institute. (2025, June 26). *Department of the Navy Fiscal Year 2026 Budget Highlights - USNI News*. USNI News. Retrieved February 19, 2026, from http://news.usni.org/2025/06/26/department-of-the-navy-fiscal-year-2026-budget-highlights

Van der Auweraer, S., Boute, R., & Syntetos, A. (2017). *Forecasting spare part demand with Installed Base information: a review*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3104452

Vegetius. (n.d.). *De Re Militari*. University of Richmond. https://facultystaff.richmond.edu/~wstevens/history331texts/Vegetius.html

Vermorel, J. (2012, February). *Pinball Loss Function Definition*. Lokad. Retrieved February 28, 2026, from https://www.lokad.com/pinball-loss-function-definition/

Vermorel, J. (2012, February). *Quantile Regression*. Lokad. Retrieved February 28, 2026, from https://www.lokad.com/quantile-regression-time-series-definition/

Wallström, P., & Segerstedt, A. (2010). Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics*, *128*(2), 625-636. https://doi.org/10.1016/j.ijpe.2010.07.013

Xenophon. (1921). *Hellenica, books VI & VII. Anabasis, books I-III*. Heinemann. https://books.google.com/books/about/Hellenica.html?id=dNMNAAAAIAAJ

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? *Proceedings of the AAAI conference on artificial intelligence*, *37*(9), 11121-11128. https://doi.org/10.1609/aaai.v37i9.26317

Zsidisin, G. A., Bresler, A., Hazen, B., Snider, K. F., & Wilkerson, T. H. (2020). Research in defense logistics: where are we and where are we going? *Journal of Defense Analytics and Logistics*, *4*(1), 3-17. https://doi.org/10.1108/JDAL-07-2019-0012