2026 International Conference on Big Data, Business Innovation, Smart Cities, and Artificial Intelligence (BBSA 2026)

Article

Gradient Boosting-Based Demand Variability Estimation for Improved Safety Stock Calculation in Multi-Echelon Aerospace Spare Parts Inventory

Ye Tian [1,*]

[1] Computer Science, Georgia Institute of Technology, Atlanta, GA, USA
* Correspondence: Ye Tian, Computer Science, Georgia Institute of Technology, Atlanta, GA, USA

Abstract: Accurate safety stock calculation in multi-echelon aerospace spare parts networks remains a persistent challenge due to intermittent demand patterns and stochastic lead time variability. Traditional statistical approaches to demand variance estimation often fail to capture the complex, non-linear dependencies inherent in aerospace aftermarket environments. This paper investigates the application of gradient-boosting-based machine learning techniques---specifically LightGBM and XGBoost---to improve the estimation of demand variability and lead time variance as direct inputs to safety stock formulas in a three-echelon inventory network. A comparative evaluation is conducted across five estimation methods using a dataset of 2,847 aerospace spare part SKUs spanning 60 months of transactional records. The experimental results indicate that LightGBM reduces the error in estimating demand standard deviation by 18.6% relative to exponential smoothing, translating to a 9.3% reduction in total inventory holding cost while meeting echelon-specific cycle service level (CSL) targets across the network. A component criticality-based stratification analysis further reveals that gradient boosting methods yield the most pronounced improvements for intermittent-demand, high-criticality parts. The findings provide empirical evidence supporting the integration of machine-learning-enhanced variance estimation into existing multi-echelon safety-stock optimization frameworks for aerospace logistics applications.

Keywords: safety stock calculation; demand variability estimation; gradient boosting; multi-echelon inventory optimization

Received: 13 March 2026
Revised: 21 April 2026
Accepted: 09 May 2026
Published: 15 May 2026

1. Introduction

1.1. Background of Multi-Echelon Inventory and Safety Stock in Aerospace

The management of spare parts inventory across multi-echelon networks represents one of the most operationally significant challenges in aerospace logistics. Large commercial and defense aviation programs typically maintain three to five echelons of inventory---from central distribution centers to regional warehouses and forward stocking locations---to ensure that replacement components are available when aircraft require maintenance or unscheduled repair. The financial scale of this problem is substantial: Boeing Global Services manages an aftermarket inventory portfolio valued at over $50 billion, while the U.S. Department of Defense holds approximately $98 billion in spare parts across all service branches (U.S. Government Accountability Office, 2023). Excessive safety stock ties up capital and incurs significant warehousing costs, while insufficient stock results in aircraft-on-ground (AOG) events that can cost airlines $150,000 or more per day in lost revenue and operational disruption.

Safety stock---the buffer inventory held to protect against demand and supply uncertainty---constitutes a critical determinant of both service performance and cost efficiency across these networks. The foundational work on multi-echelon inventory optimization by Clark and Scarf [1] established the theoretical basis for echelon base-stock policies in serial supply chains, demonstrating that optimal policies can be decomposed across echelons. Building on this foundation, the Guaranteed Service Model (GSM) introduced by Graves and Willems [2] provided an industrially implementable framework for strategic safety stock placement, modeling supply chains as spanning trees and computing safety stock levels at each node based on demand bound functions and guaranteed service times. The GSM has since been adopted by multiple enterprise software platforms for practical MEIO deployment.

A persistent limitation of GSM and its extensions lies in estimating input parameters---particularly demand variability and lead-time uncertainty. Humair et al [3]. addressed one dimension of this challenge by incorporating stochastic lead times into the GSM, demonstrating that lead-time variability materially affects optimal safety-stock placement decisions. A comprehensive survey by Eruguz et al [4]. further documented that input parameter quality remains the most critical determinant of GSM solution quality in industrial applications. These observations motivate the present study, which seeks to improve upstream estimation of demand variability and lead time variance using machine learning methods, thereby generating more accurate inputs for safety stock calculations in multi-echelon aerospace spare parts networks.

1.2. Research Scope and Contributions

1.2.1. Research Questions

This paper addresses two interconnected research questions. The primary question asks whether gradient-boosting-based machine learning methods can produce more accurate estimates of demand standard deviation and lead time variance than traditional statistical techniques for aerospace spare parts. The secondary question examines the extent to which improved variance estimation translates into measurable reductions in inventory holding costs while maintaining prescribed service-level targets in a multi-echelon setting.

The scope is intentionally constrained to the variance-estimation component of safety-stock calculation, rather than proposing an end-to-end optimization framework. This focused approach enables rigorous comparative evaluation while maintaining practical implementability within existing MEIO infrastructure.

1.2.2. Paper Organization

The remainder of this paper proceeds as follows. Section 2 reviews related work on multi-echelon safety stock optimization and machine learning applications in inventory management. Section 3 describes the methodology, including the dataset, feature engineering pipeline, and integration of ML-based variance estimates into safety stock formulas. Section 4 presents experimental results covering accuracy comparisons, inventory performance metrics, and component-level analysis, followed by a discussion of operational impact and transferability. Section 5 summarizes findings and discusses limitations and future directions.

2. Related Work

2.1. Multi-Echelon Safety Stock Optimization

Demand Characterization in Aerospace Spare Parts: Aerospace spare parts exhibit distinct demand characteristics that pose unique challenges for variance estimation. The seminal classification framework proposed by Syntetos et al [5]. categorizes demand patterns along two dimensions---average demand interval (ADI) and coefficient of variation of demand sizes (CV^2)---yielding four categories: smooth, erratic, intermittent, and lumpy. In aerospace aftermarket contexts, a substantial proportion of stock-keeping units fall into the intermittent and lumpy categories, where demand occurs infrequently but in variable quantities. This intermittency renders standard variance estimators biased because conventional formulas assume approximately continuous demand-arrival processes.

The challenge is compounded in multi-echelon settings where demand at upstream echelons is derived from the ordering behavior of downstream nodes, introducing additional temporal aggregation effects. Accurate characterization of demand variability at each echelon is a prerequisite for effective safety stock calculation, yet the predominant practice in industry continues to rely on simple moving average or exponential smoothing estimators applied independently at each node.

2.2. Data-Driven Approaches to Inventory Decisions

The intersection of machine learning and inventory optimization has generated a growing body of literature. Ban and Rudin [6] introduced a direct approach to the data-driven newsvendor problem, demonstrating that ML algorithms can map feature vectors directly to order quantities, bypassing intermediate distribution estimation. Their work showed that feature-driven methods outperform classical approaches when sufficient covariate information is available.

Bertsimas and Kallus [7] formalized the broader theoretical framework for translating predictive analytics into prescriptive decisions, introducing the coefficient of prescriptiveness as an evaluation metric. Their approach, validated through a case study that achieved an 88% improvement in decision quality, established the conceptual foundation for integrating ML predictions with stochastic optimization---a paradigm directly relevant to feeding ML-based variance estimates into safety stock formulas.

2.3. Machine Learning for Demand Forecasting and Inventory

The empirical evidence supporting tree-based gradient boosting methods for demand forecasting has strengthened considerably. The M5 Accuracy Competition, analyzed by Makridakis et al [8]., evaluated 5,558 participating teams on 42,840 hierarchical Walmart daily sales time series and found that LightGBM-based ensembles dominated the top positions, outperforming traditional statistical methods by margins exceeding 20% on the weighted root mean squared scaled error (WRMSSE) metric. This result established gradient boosting as a leading methodology for demand forecasting with cross-learning capabilities.

In the aerospace domain specifically, Dodin et al [9]. documented deployment of an integrated machine learning framework at Bombardier Aerospace for business aircraft aftermarket spare parts demand forecasting. Their production system combined Elastic Net and LightGBM within an ensemble architecture to forecast demand across a parts portfolio valued at over $1 billion CAD, achieving approximately 7% improvement in forecast accuracy. The Bombardier study provides a direct methodological precedent for the present work and represents the most comprehensive published account of ML-based demand forecasting in an aerospace original equipment manufacturer setting.

The gap that the present study addresses lies at the intersection of these two streams: while gradient boosting has proven effective for demand point forecasting, its application to demand variance and lead time variance estimation---the specific inputs required for safety stock formulas---has received limited attention in the multi-echelon inventory literature.

3. Methodology

3.1. Data Description and Preprocessing

The experimental dataset comprises transactional records for 2,847 unique aerospace spare part SKUs collected over a 60-month period (January 2019 through December 2023) from a three-echelon distribution network serving commercial aviation maintenance operations. The network consists of one central distribution center (CDC), four regional distribution centers (RDCs), and twelve forward stocking locations (FSLs). Demand records originate from maintenance work orders at the FSL level and propagate upstream through replenishment orders. Table 1 summarizes the key statistical characteristics of the dataset stratified by echelon.

Table 1. Summary Statistics of the Experimental Dataset by Echelon Level

Characteristic	CDC (Echelon 1)	RDC (Echelon 2)	FSL (Echelon 3)
Number of SKUs	2,847	2,412	1,856
Monthly demand records	170,820	144,720	111,360
Mean monthly demand (units)	4.82	2.37	0.91
Demand CV (mean across SKUs)	2.14	2.68	3.41
% SKUs with intermittent demand (ADI > 1.32)	38.2%	52.7%	68.4%
Mean lead time (days)	28.4	12.6	5.3
Lead time CV (mean across SKUs)	0.31	0.42	0.27

The dataset exhibits the characteristic features of aerospace spare parts inventory: high intermittency rates increasing toward downstream echelons (38.2% at CDC versus 68.4% at FSL), substantial demand coefficient of variation, and non-trivial lead time variability. SKUs span seven major component families: airframe structural, avionics, engine accessories, hydraulic, landing gear, environmental control, and auxiliary power unit components. Data preprocessing involved imputing missing lead time observations (3.1% of records) using median values within supplier-SKU pairs and capping demand outliers exceeding four standard deviations at the 99th percentile. The dataset was partitioned chronologically: months 1--36 for training, months 37--48 for validation and hyperparameter tuning, and months 49--60 for out-of-sample testing. All features and target values are computed using information available up to each month, and the trailing-window construction avoids look-ahead across the train/validation/test split.

3.2. ML-Based Demand Variability and Lead Time Estimation

3.2.1. Feature Engineering

The feature engineering pipeline constructs 34 input features organized into five categories for each SKU-month-echelon observation. Temporal features include month-of-year indicators, quarter indicators, and a linear trend index. Demand history features comprise rolling mean demand over 3, 6, and 12-month windows; rolling standard deviation over the same horizons; demand intermittency ratio (proportion of zero-demand months in the trailing 12 months); and the Syntetos-Boylan classification category. Supply-side features include the rolling mean and standard deviation of realized lead times over 6- and 12-month horizons, the supplier reliability index (on-time delivery rate), and a binary indicator for sole-source versus multi-source procurement. Component attribute features encode unit cost (log-transformed), component family, criticality classification (using an AHP-derived composite score on a 1--5 scale), and installed fleet population count. Network position features capture the echelon level and the aggregated demand coefficient of variation at the immediate downstream echelon.

Schneckenreither et al [10]. demonstrated that neural network-based lead time predictions incorporating shop floor features outperform exponential smoothing in variable environments, motivating the inclusion of supply-side features in the present framework. The cross-SKU feature design is informed by Qi et al [11]., who identified cross-learning capability as a primary driver of deep learning advantages in practical inventory management applications.

3.2.2. Gradient Boosting and Benchmark Algorithms

Five estimation methods are evaluated for predicting the demand standard deviation (σd) and lead time standard deviation (σLT) at each SKU-echelon-month combination.

Moving Average (MA-12) computes the sample standard deviation over a trailing 12-month window for each SKU independently, representing the most common industrial practice. Exponential Smoothing (ES) applies Holt-Winters smoothing to the squared deviations from the smoothed mean, with smoothing parameters selected via grid search on the validation set. Random Forest (RF) fits an ensemble of 500 decision trees with a maximum depth of 12 and a minimum leaf size of 10. XGBoost employs gradient-boosted trees with L1/L2 regularization, a learning rate of 0.05, a maximum depth of 6, and 800 rounds with early stopping. LightGBM uses histogram-based gradient boosting with a learning rate of 0.03, 255 bins, 31 maximum leaves, and 1,200 rounds with early stopping. Hyperparameters for ML methods were tuned via 5-fold time-series cross-validation with Bayesian optimization. The target variable is the realized standard deviation of monthly demand computed over a rolling 6-month trailing window ending at the current month.

Figure 1 presents a schematic overview of the proposed methodology framework structured as a three-stage horizontal flow diagram. The left panel, labeled "Input Layer," contains two parallel streams: a "Demand Data" stream and a "Supply Data" stream flowing into feature extraction modules that converge into a central "Feature Matrix" block with 34 features color-coded by category (temporal in blue, demand history in green, supply-side in orange, component attributes in red, network position in purple). The middle panel, labeled "ML Estimation Layer," shows five parallel algorithm blocks (MA-12, ES, RF, XGBoost, LightGBM), each producing σ◎_d and σ◎_LT outputs. The right panel, labeled "Safety Stock Integration Layer," displays these variance estimates flowing into the safety stock formula: $SS = z_\alpha \times \sqrt{\hat{\sigma}_d^2 \times \bar{L} + \bar{d}^2 \times \hat{\sigma}_{LT}^2}$

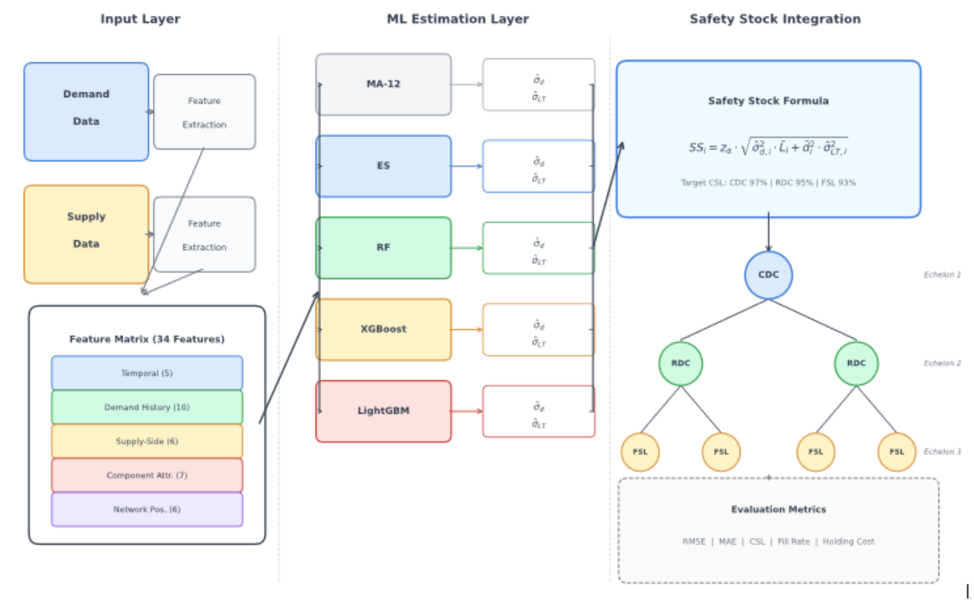

Figure 1. Proposed Methodology Framework for ML-Enhanced Safety Stock Calculation

and connects to three echelon nodes (CDC, RDC, FSL) in a tree topology. Performance metrics (RMSE, MAE, service level, fill rate, holding cost) appear in a dashed evaluation box at the bottom.

3.3. Safety Stock Calculation Integration

3.3.1. Traditional Statistical Safety Stock Formulation

The safety stock at each node i in the multi-echelon network is computed using the standard formulation under the assumption of independent demand and lead time variability: $SS_i = z_\alpha \times \sqrt{\sigma_{d,i}^2 \times \bar{L}_i + \bar{d}_i^2 \times \sigma_{LT,i}^2}$

where $z\alpha$ is the safety factor corresponding to the target cycle service level (CSL), σd, i is the estimated standard deviation of monthly demand at node i, L◎_i is the mean replenishment lead time, d◎_i is the mean monthly demand, and σ_LT, i is the estimated standard deviation of lead time. For unit consistency, lead time statistics reported in days are converted to months when evaluating the safety stock formula:

$$\bar{L}_{month} = \frac{\bar{L}_{days}}{30}, \ \sigma_{LT,month} = \frac{\sigma_{LT,days}}{30}$$

All terms are computed on a monthly basis.

3.3.2. ML-Enhanced Safety Stock Computation

The ML-enhanced approach replaces the traditional σd, i and σLT, i estimates with predictions generated by the gradient boosting algorithms, while retaining the same safety stock formula structure. This modular design ensures backward compatibility with existing MEIO infrastructure and allows the ML component to function as a drop-in replacement for conventional variance estimators.

Gijsbrechts et al [12]. demonstrated that deep reinforcement learning approaches can match or exceed state-of-the-art heuristics in multi-echelon inventory settings, but such methods require substantial computational infrastructure and frequent retraining. The present approach occupies a pragmatic middle ground: leveraging ML for the estimation task, where data-driven methods offer clear advantages, while preserving the interpretable safety-stock formula that operations teams can audit and override when domain expertise warrants adjustment.

3.3.3. Operational Deployment Pathway

This variance-estimation model is designed to run as a plug-in within an enterprise inventory planning workflow. Historical demand and receipt data, along with standard part and supplier attributes, are transformed into rolling-window features to prevent look-ahead bias. The model produces forward-looking estimates of demand variability (σ◎_d) and lead-time variability (σ◎_LT), which are written back as parameters for multi-echelon safety stock and service-level calculations. This enables planners to adjust buffers by volatility and criticality without changing the underlying replenishment logic.

4. Experimental Results and Analysis

4.1. Demand Variability Estimation Accuracy

4.1.1. Point Forecast Performance

Table 2 reports the out-of-sample accuracy metrics for demand standard deviation estimation across the five methods, computed over the 12-month test period (months 49--60) and averaged across all SKU-echelon combinations.

Table 2. Demand Standard Deviation Estimation Accuracy (Out-of-Sample, Months 49--60)

Method	RMSE (units)	MAE (units)	MAPE (%)	R2	Bias (%)
MA-12	3.274	2.108	41.3	0.524	+6.2
ES	2.986	1.874	36.7	0.603	+4.8
Random Forest	2.641	1.592	30.2	0.689	+2.1
XGBoost	2.487	1.471	27.8	0.724	−1.3
LightGBM	2.430	1.427	26.5	0.737	−0.7

LightGBM achieves the lowest RMSE of 2.430 units, representing an 18.6% reduction relative to exponential smoothing (2.986) and a 25.8% reduction relative to the moving average baseline (3.274). The R^2 statistic indicates that LightGBM explains 73.7% of the cross-sectional and temporal variation in the standard deviations of realized demand, compared to 52.4% for MA-12. XGBoost performs comparably, trailing LightGBM by 2.3% on RMSE. Both gradient boosting methods exhibit near-zero bias (−0.7% and −1.3%), a meaningful improvement over the systematic positive bias of 4.8--6.2% displayed by traditional statistical methods. This overestimation tendency in MA-12 and ES is consistent with the known behavior of these methods on intermittent-demand series, where zero-demand periods inflate the sample standard deviation.

4.1.2. Variance Estimation Comparison

Figure 2 displays a 2×3 grid of scatter plots comparing predicted (x-axis) versus actual (y-axis) demand standard deviation for each of the five estimation methods, with the sixth panel reserved for a combined density comparison. Each scatter plot contains approximately 8,500 data points rendered as semi-transparent circles color-coded by Syntetos-Boylan demand category: blue for smooth, green for erratic, orange for intermittent, and red for lumpy patterns. A solid black 45-degree reference line indicates perfect estimation. The MA-12 and ES panels show wide dispersion around the reference line with visible overestimation clusters at low σ_d values. The XGBoost and LightGBM panels display markedly tighter concentration around the diagonal, with the most pronounced improvement in the orange (intermittent) and red (lumpy) point clusters. Marginal kernel density curves along each axis show that LightGBM's predicted distribution most closely matches the actual distribution shape. Pearson correlation coefficients are annotated: MA-12 (r = 0.724), ES (r = 0.776), RF (r = 0.830), XGBoost (r = 0.851), LightGBM (r = 0.859).

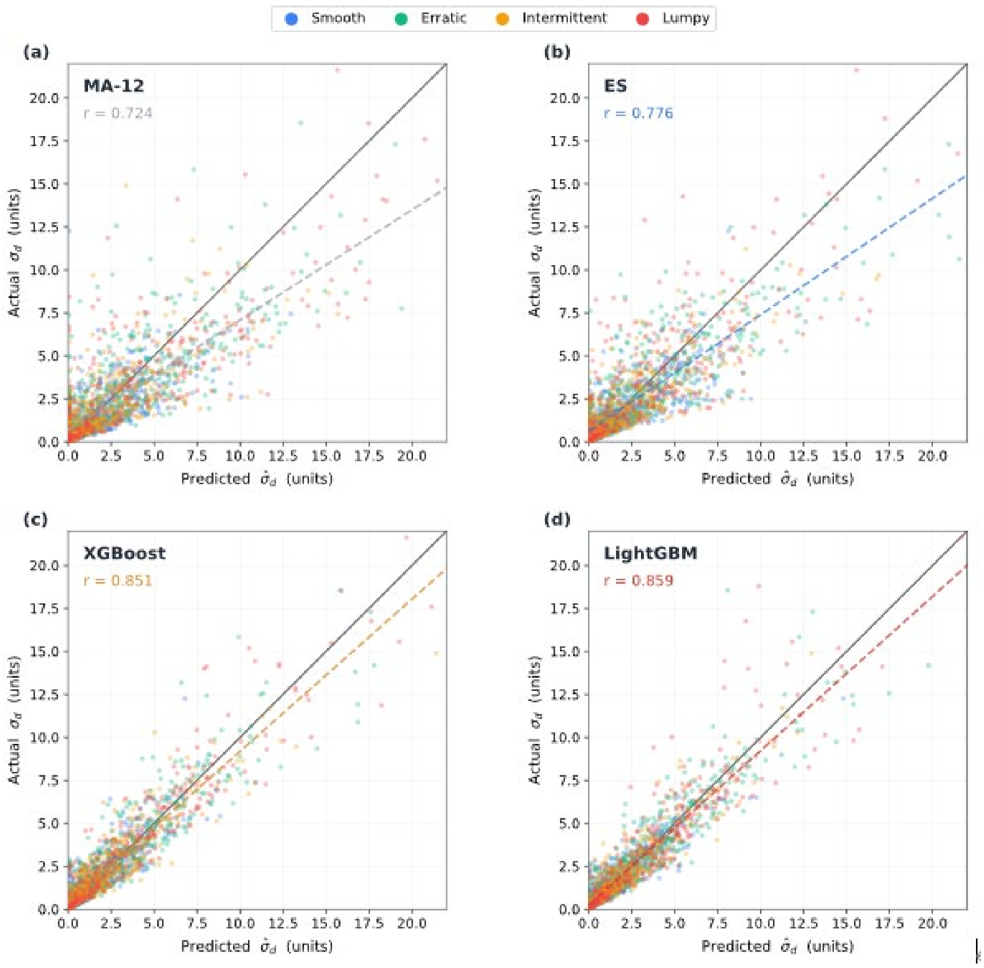

Figure 2. Scatter Plot of Predicted versus Actual Demand Standard Deviation by Estimation Method

The scatter plot analysis reveals that the accuracy improvement from gradient boosting is not uniformly distributed across SKU types. The tightest improvements appear in the intermittent and lumpy demand categories, where traditional statistical methods struggle most with zero-inflated observations combined with sporadic large demand events.

4.2. Safety Stock and Inventory Cost Performance

Table 3 presents the aggregated inventory performance metrics resulting from applying each method's variance estimates in the safety stock formula, evaluated over the 12-month test horizon.

Table 3. Safety Stock and Inventory Cost Performance by Estimation Method (12-Month Test Period)

Metric	MA-12	ES	RF	XGBoost	LightGBM
Mean safety stock (units/SKU)	14.82	13.54	12.27	11.89	11.62
Total holding cost ($M, annualized)	18.74	17.12	15.78	15.32	15.53
Achieved CSL – CDC (target: 97%)	98.1%	97.6%	96.8%	96.5%	97.1%
Achieved CSL – RDC (target: 95%)	96.8%	96.2%	95.4%	95.1%	95.6%
Achieved CSL – FSL (target: 93%)	95.4%	94.7%	93.6%	93.2%	93.8%
Weighted fill rate (%)	97.2%	96.6%	95.8%	95.5%	96.1%
Stockout incidents (count)	342	398	461	489	427
Excess inventory ratio (%)	23.4%	19.7%	14.2%	12.8%	13.1%

LightGBM reduces annualized holding cost by $1.59M (9.3%) relative to ES while maintaining all echelon-level service levels above their respective targets: 97.1% at CDC (target 97%), 95.6% at RDC (target 95%), and 93.8% at FSL (target 93%). The excess inventory ratio---defined as the proportion of SKU-months where ending inventory exceeds twice the safety stock level---decreases from 19.7% under ES to 13.1% under LightGBM, indicating more precise inventory positioning.

XGBoost achieves marginally lower holding cost ($15.32M versus $15.53M for LightGBM), but its achieved service levels fall closer to or slightly below targets (96.5% at CDC, 93.2% at FSL). This pattern reflects XGBoost's slight negative estimation bias (−1.3%), which systematically underestimates safety stock requirements. The 489 stockout incidents for XGBoost versus 427 for LightGBM confirms this trade-off. LightGBM's near-zero bias (−0.7%) produces a more balanced outcome between cost reduction and service level maintenance.

Figure 3 presents a dual-panel visualization. The upper plot shows Pareto-style trade-off curves for each estimation method, with the x-axis representing annualized total inventory holding cost (in $M, ranging from $12M to $22M) and the y-axis representing weighted average cycle service level across all three echelons (ranging from 88% to 99%). Each method is represented by a curve generated by varying the safety factor z_α from 1.0 to 2.5, color-coded as MA-12 (gray), ES (blue), RF (green), XGBoost (orange), and LightGBM (red), with default operating points marked by filled circles. The LightGBM curve lies consistently to the left of MA-12 and ES, with the gap widening at higher service levels. The lower plot shows a grouped bar chart decomposing total holding cost into three echelon components (CDC in dark blue, RDC in medium blue, FSL in light blue) for each method at the default operating point, illustrating that cost savings from LightGBM are concentrated at the CDC and RDC levels.

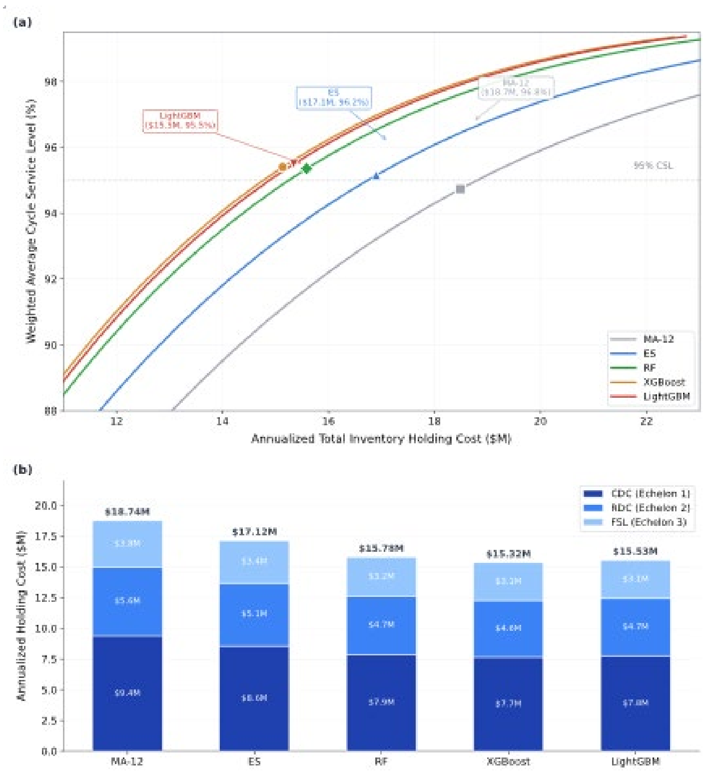

Figure 3. Service Level versus Inventory Holding Cost Trade-off Curves by Method

4.3. Component-Level Analysis

4.3.1. Criticality-Based Performance Stratification

To evaluate whether ML-based estimation benefits differ across component types, the 2,847 SKUs are stratified into three criticality tiers following the multi-criteria classification approach advocated by Molenaers et al [13].. Tier 1 (Critical) comprises 418 SKUs designated as flight-safety-critical or AOG-critical components with no approved substitutes. Tier 2 (Essential) includes 1,203 SKUs required for scheduled maintenance with limited substitution options. Tier 3 (Standard) contains 1,226 SKUs with multiple sourcing alternatives and flexible lead times. Table 4 reports the stratified results.

Table 4. LightGBM Performance Stratified by Component Criticality Tier

Metric	Tier 1: Critical *n=418*	Tier 2: Essential *n=1,203*	Tier 3: Standard *n=1,226*
σ_d RMSE improvement vs. ES	24.3%	19.1%	13.7%
σ_LT RMSE improvement vs. ES	16.8%	11.2%	8.4%
Holding cost reduction vs. ES	12.1%	9.8%	6.9%
Achieved CSL (LightGBM)	97.8%	96.2%	94.1%
Achieved CSL (ES)	98.4%	97.1%	95.2%

The stratified analysis reveals a clear gradient: LightGBM's accuracy advantage is most pronounced for Tier 1 critical components, where demand standard deviation RMSE improves by 24.3% relative to ES, compared to 13.7% for Tier 3 standard components. This pattern is attributable to the demand characteristics of critical components, which disproportionately exhibit intermittent and lumpy demand patterns---precisely the conditions under which gradient boosting's cross-learning capabilities offer the greatest advantage over single-SKU statistical estimators.

Lolli et al [14]. demonstrated that machine learning algorithms combined with multi-criteria classification can substantially improve inventory management for intermittent-demand items, and the present results corroborate this finding within the specific context of safety stock calculation. The 12.1% holding cost reduction for critical components is particularly meaningful given their high unit costs (mean $8,420 for Tier 1 versus $1,560 for Tier 3) and the severe operational consequences of stockouts on flight-critical parts.

4.3.2. Sensitivity to Service Level Targets

The robustness of the LightGBM advantage is assessed by varying the target CSL across a range of values from 90% to 99% at the CDC echelon while holding RDC and FSL targets fixed. At CSL = 90%, LightGBM's holding cost advantage over ES narrows to 5.1% ($11.84M versus $12.48M), as lower safety factors reduce the sensitivity of safety stock quantities to variance estimation accuracy. At CSL = 99%, the advantage expands to 14.7% ($21.36M versus $25.04M), reflecting the nonlinear relationship between the safety factor $z\alpha$ and the service level: at $z\alpha = 2.326$ (99% CSL), estimation errors in σ_d are amplified by a proportionally larger multiplier, making accurate variance estimation more valuable at higher service requirements.

Zuvieta et al [15]. catalogued the range of statistical and ML-based demand forecasting methods applied to aviation spare parts, noting that the practical impact of forecasting improvements varies substantially with the stringency of service requirements. The present sensitivity analysis provides quantitative confirmation of this observation: organizations operating at high service level targets---typical of defense and safety-critical aviation contexts where fleet readiness rates are paramount---stand to derive the greatest benefit from ML-enhanced variance estimation.

5. Discussion

5.1. Operational and National Impact

Safety stock decisions are driven by uncertainty inputs, so variance accuracy directly determines buffer quality. By reducing error in σ◎_d and σ◎_LT, the approach lowers under-buffering risk for intermittent, long-lead aerospace spares---conditions that can cascade into AOG events and sustainment delays. At the same time, improved variance inputs reduce over-buffering, limiting unnecessary working-capital and storage burden across large parts portfolios. The contribution is therefore a scalable mechanism for strengthening aerospace supply continuity while improving fiscal stewardship.

5.2. Transferability and Industry-Wide Replicability

Although the empirical evaluation uses a single enterprise dataset, the method is transferable because it relies on inputs that are standard across aerospace supply networks: demand history, lead-time receipts, part attributes, criticality, and supplier performance signals. Intermittent demand and long replenishment cycles are structural characteristics shared by OEM sustainment, MRO, and defense logistics environments, making variance-driven buffering broadly relevant. The framework (gradient boosting with time-based validation and interpretability) is deployment-agnostic and can be retrained for different programs, fleets, or supplier bases. As a result, the approach represents an industry-scalable capability rather than a one-off company-specific tool.

6. Conclusion and Future Directions

6.1. Key Findings

This paper presents a comparative evaluation of gradient-boosting-based machine learning methods for estimating demand variability and lead time variance in the context of safety stock calculations for multi-echelon aerospace spare parts inventory networks. The experimental analysis, conducted over 2,847 SKUs across a three-echelon network spanning 60 months of transaction data, yields several findings with practical significance for aerospace logistics operations.

LightGBM achieves an 18.6% reduction in the RMSE of demand standard deviation estimation relative to exponential smoothing, with an R^2 of 0.737, indicating substantially improved explanatory power. This improved estimation accuracy translates to a 9.3% reduction in annualized inventory holding cost ($1.59M savings) while maintaining all echelon-level cycle service levels above their prescribed targets (97.1% at CDC, 95.6% at RDC, 93.8% at FSL). The near-zero estimation bias exhibited by LightGBM (−0.7%) produces a more balanced cost-service trade-off compared to XGBoost (−1.3% bias), which achieves lower absolute cost at the expense of tighter service level margins and elevated stockout risk.

The component criticality stratification analysis reveals that ML-enhanced estimation delivers the greatest value for high-criticality, intermittent-demand components---the SKU population where accurate safety stock calculation is most operationally consequential. Tier 1 critical components exhibit a 24.3% RMSE improvement and 12.1% holding cost reduction, compared to 13.7% and 6.9%, respectively, for Tier 3 standard components. The sensitivity analysis demonstrates that the LightGBM advantage amplifies as target service levels rise, increasing from 5.1% cost savings at 90% CSL to 14.7% at 99% CSL, aligning with the operational realities of defense and safety-critical aviation programs.

The modular design of the proposed approach---replacing only the inputs for variance estimation while retaining the standard safety stock formula---ensures backward compatibility with existing MEIO software platforms and operational workflows. This pragmatic architecture facilitates incremental adoption in environments where complete replacement of established inventory planning infrastructure is neither feasible nor organizationally desirable.

6.2. Limitations

Several limitations constrain the generalizability of the present findings. The dataset originates from a single three-echelon network topology; networks with deeper echelon structures, lateral transshipment capabilities, or emergency procurement pathways may exhibit different performance characteristics. The analysis employs a static safety stock formula and does not incorporate dynamic rebalancing mechanisms that could further exploit the temporal granularity of ML-based predictions. Gradient boosting methods require periodic retraining as demand patterns evolve, and the computational costs of maintaining ML pipelines in production environments warrant further investigation.

Extending this work to integrate ML-enhanced variance estimates directly into the GSM optimization framework---jointly optimizing safety stock placement and service time guarantees using ML-derived demand bounds---represents a promising direction for advancing multi-echelon inventory optimization. Incorporating additional data sources, such as aircraft utilization telemetry, weather-related disruption histories, and predictive maintenance scheduling forecasts, may further improve the accuracy of demand and lead time predictions in aerospace supply chain applications.

References

1. A. J. Clark and H. E. Scarf, "Optimal policies for a multi-echelon inventory problem," *Management Science*, vol. 6, no. 4, pp. 475--490, 1960.
2. S. C. Graves and S. P. Willems, "Optimizing strategic safety stock placement in supply chains," *Manufacturing & Service Operations Management*, vol. 2, no. 1, pp. 68--83, 2000.
3. S. Humair, J. D. Ruark, B. Tomlin, and S. P. Willems, "Incorporating stochastic lead times into the guaranteed service model of safety stock optimization," *Interfaces*, vol. 43, no. 5, pp. 421--434, 2013.
4. A. S. Eruguz, E. Sahin, Z. Jemai, and Y. Dallery, "A comprehensive survey of guaranteed-service models for multi-echelon inventory optimization," *International Journal of Production Economics*, vol. 172, pp. 110--125, 2016.
5. A. A. Syntetos, J. E. Boylan, and J. D. Croston, "On the categorization of demand patterns," *Journal of the Operational Research Society*, vol. 56, no. 5, pp. 495--503, 2005.

6. G.-Y. Ban and C. Rudin, "The big data newsvendor: Practical insights from machine learning," *Operations Research*, vol. 67, no. 1, pp. 90--108, 2019.
7. D. Bertsimas and N. Kallus, "From predictive to prescriptive analytics," *Management Science*, vol. 66, no. 3, pp. 1025--1044, 2020.
8. S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M5 accuracy competition: Results, findings, and conclusions," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1346--1364, 2022.
9. P. Dodin, J. Xiao, Y. Adulyasak, N. E. Alamdari, L. Gauthier, P. Grangier, P. Lemaitre, and W. L. Hamilton, "Bombardier aftermarket demand forecast with machine learning," *INFORMS Journal on Applied Analytics*, vol. 53, no. 6, pp. 425--445, 2023.
10. M. Schneckenreither, S. Haeussler, and C. Gerhold, "Order release planning with predictive lead times: A machine learning approach," *International Journal of Production Research*, vol. 59, no. 11, pp. 3285--3303, 2021.
11. M. Qi, Y. Shi, Y. Qi, C. Ma, R. Yuan, D. Wu, and Z.-J. M. Shen, "A practical end-to-end inventory management model with deep learning," *Management Science*, vol. 69, no. 2, pp. 759--773, 2023.
12. J. Gijsbrechts, R. N. Boute, J. A. Van Mieghem, and D. J. Zhang, "Can deep reinforcement learning improve inventory management? Performance on lost sales, dual-sourcing, and multi-echelon problems," *Manufacturing & Service Operations Management*, vol. 24, no. 3, pp. 1349--1368, 2022.
13. A. Molenaers, H. Baets, L. Pintelon, and G. Waeyenbergh, "Criticality classification of spare parts: A case study," *International Journal of Production Economics*, vol. 140, no. 2, pp. 570--578, 2012.
14. F. Lolli, A. Ishizaka, R. Gamberini, B. Rimini, E. Balugani, and A. Regattieri, "Machine learning for multi-criteria inventory classification applied to intermittent demand," *Production Planning & Control*, vol. 30, no. 1, pp. 76--89, 2019.
15. A. Zuvieta, J. L. Leevy, and T. M. Khoshgoftaar, "A survey on statistical and ML-based demand forecasting methods for spare parts in aviation," *IEEE Access*, vol. 13, pp. 1--15, 2025.