A Metric-Driven Framework for Evaluating Prognostic Based Failure Estimation on Spare Part Inventory Management

Tanmay Daga[1], Robert Meissner[1], Ahmad Ali Pohya[1] and Gerko Wende[1]

[1] *German Aerospace Center (DLR), Hamburg, 21129, Germany*
tanmay.daga@dlr.de
robert.meissner@dlr.de
ahmad.pohya@dlr.de
gerko.wende@dlr.de

ABSTRACT

In recent years, the aviation sector has been at the forefront of adopting Industry 4.0 technologies, including artificial intelligence, additive manufacturing, cyber-physical systems, big data analytics, and the Internet of Things (IOT). These technologies have accelerated the development of advanced maintenance strategies, such as Predictive and Prescriptive Maintenance, especially in the field of spare part inventory management. By leveraging insights from Prognostic Health Management (PHM) technologies, logistics and maintenance service providers can optimize inventory levels to reduce costs while maintaining service levels, thereby minimizing aircraft downtimes. Despite these potential advantages, the widespread adoption of PHM strategies in the Maintenance, Repair and Overhaul (MRO) industry remains a challenge, primarily due to their modelling complexity, high cost of adoption, regulatory challenges, data availability, and impact assessment. However, a more targeted allocation of development resources, to address these barriers, can be achieved if economic benefits can clearly be demonstrated for individual stakeholders (such as logistics) and different PHM technology maturity levels. Therefore, the aim of this study is to quantify the benefits of prognostic-based inventory policies in comparison to traditional reliability-based approaches across different demand patterns. Specifically, this study investigates the influence of different prognostic accuracy and prognostic horizon levels on key performance indicators, such as total cost and service level. It also evaluates the robustness of the proposed methodology against noise factors like prediction biases and false alarms. Based on these comparisons, minimum performance requirements for prognostics based policies can be established to ensure tangible benefits. Consequently, this study not only provides the readers with a methodology to quantify the impact of prognostics-based failure prediction on spare part inventory management, but it also proposes a lightweight framework which could act as surrogate for prognostic models to assist in the development of future prescriptive maintenance strategies.

1. INTRODUCTION

The availability of spare parts in aviation has always been of critical importance, as aircraft downtime directly undermines operational reliability, safety performance, and financial outcomes. The provisioning of spare parts is intricately embedded in complex, globally distributed supply networks which are bound to constraints regarding the availability of critical raw materials (Hoff, Sprecher, Pohya, Wende, & Peck, 2025), stringent regulatory requirements (Meissner, Pohya, Weiss, Piotrowski, & Wende, 2025), and extended lead times for part replenishment (Oliver Wyman & International Air Transport Association, 2025). Even minor malfunctions can propagate rapidly, potentially resulting in Aircraft on Ground (AOG) events with extensive operational irregularities and substantial economic losses for airline operators (Alomar & Nikita, 2025). To account for such risks and gain flexibility to react to technical breakdowns, Maintenance, Repair and Overhaul (MRO) logistic providers tend to maintain higher inventory levels of essential spare parts. While this strategy can help to stabilize operations through improved logistics service levels, it increases the associated spare unit holding costs significantly, resulting in additional operational expenses of \$1.3 billion in the year 2025 alone (Oliver Wyman & International Air Transport Association, 2025). Consequently, the development of Prognostic Health Management (PHM) systems that can (a) aid in the accurate prediction of demand, (b) support efficient inventory management solution, and (c) improve resilience against external shocks (e.g., sudden supply chain interruptions) remain of pivotal importance to the industry.

Traditionally, spare parts demand forecasting has relied on stochastic methods, such as Croston's two-step approach for intermittent demand (Croston, 1972), its bias-corrected variant Syntetos and Boylan Approximate (SBA) (Syntetos & Boylan, 2005), the Teunter-Syntetos-Babai (TSB) method for declining demand (Teunter, Syntetos, & Babai, 2011), and exponential smoothing (Romeijnders, Teunter, & van Jaarsveld, 2012). These methods, however, require extensive information about historic demand to extract meaningful trends. To counter this, several studies exist which rather focus on individual component's reliability to predict the failure, like, the development of a conditional Weibull approach to predict the failure time of individual Line Replaceable Unit (LRU) for new aircraft (Yongquan, Xi, He, Yingchao, & Quanwu, 2016). More recently, sensor-driven prognostics have evolved, with studies predicting a turbofan engine component's Remaining Useful Life (RUL) through functional principal component analysis and penalized regression to select informative sensors and fuse degradation signals (Fang, Paynabar, & Gebraeel, 2017). However, these approaches are largely constrained by limited data availability and predefined statistical assumptions, restricting their ability to capture complex degradation patterns.

With the recent advancements in the Industry 4.0 concepts and integration of large-scale sensors and data across process and maintenance systems, there has been an increasing emphasis on the benefits of prognosis in failure prediction, allowing improved system reliability maintenance planning (Reis & Gins, 2017). In aerospace applications specifically, PHM-driven RUL estimation and prescriptive maintenance frameworks have demonstrated significant gains in safety, cost efficiency, emission reduction, and delay mitigation (Meissner, Rahn, & Wicke, 2021; Kordestani, Orchard, Khorasani, & Saif, 2023). Studies on complex systems like turbofan engines have demonstrated that machine learning algorithms, such as random forests and neural networks, can effectively predict the RUL of aircraft components for predictive maintenance (Mathew, Toby, Singh, Rao, & Kumar, 2017). Probabilistic machine learning using a multi-channel Convolutional Neural Network (CNN) has also been applied to predict the RUL of aircraft turbofan engines for predictive maintenance (Lee & Mitici, 2023). Studies have also combined Long Short-Term Memory (LSTM) based models RUL prediction with a latent wiener process to model turbofan degradation (Deng, Bucchianico, & Pechenizkiy, 2020).

However, despite its potential, the application of PHM-based maintenance approaches remain limited, in particular within the aerospace sector. Implementation of prognostic capabilities is constrained by a lack of availability and ownership of quality data, high cost of adoption, and the complexity of the models (Teubert, Pohya, & Gorospe, 2023). Furthermore, existing Condition Based Maintenance (CBM) and regulatory frameworks provide limited guidance on translating high-level system objectives into measurable prognostic performance requirements, making it difficult to formally certify prognosis tools for safety-critical aircraft functions (Meissner et al., 2025).

Building on these challenges, there is a clear necessity to develop a framework that can assess the practical value of prognostics in real-world applications, e.g., spare part inventory management. Some ground work has already been laid in this field through prognostics-based just-in-time spare part ordering (Cai et al., 2022; Bousdekis, Papageorgiou, Magoutas, Apostolou, & Mentzas, 2017) and the development of a cost optimization model for prognostics-based spare part ordering and replacement (Z. Wang, Hu, Wang, Kong, & Zhang, 2015). However, these studies are limited in terms of identifying technological maturity requirements (e.g., expressed in terms of Prognostic Accuracy (PA) and Prognostic Horizon (PH)) in order to break-even with traditional statistics-based inventory management approaches.

Through this study, by utilizing the prognostic performance metrics PA and PH as indicators of different technological maturity levels and surrogates for full prognostic models, we propose a methodology that enables a systematic comparison between traditional statistics-based and prognostic-based inventory management strategies. The developed framework allows the evaluation of inventory-management-related costs and service level while explicitly accounting for external disturbances such as prediction skewness and (diagnostic) False Alarm (FA). The obtained results will (a) provide insights into potential benefits and limitations of adopting prognostics-based strategies in spare parts management and (b) support developers to identify the necessary prognostic performance of their developed algorithm to justify deviation from conventional statistics-based approaches.

2. SIMULATION ARCHITECTURE

This section presents the methodology used to compare different strategies for demand prediction in spare part inventory management. For this study, we use a rolling horizon methodology, which is well documented for maintenance-driven environments (Nielsen, Kroon, & Maróti, 2012; Sahin, Narayanan, & Robinson, 2013), as it continuously updates decisions based on newly revealed system states and demand information to reduce uncertainty. Within the framework, two distinct demand prediction approaches are evaluated: (a) reliability-based statistical methods and (b) prognostic-based methods. This distinction allows us to assess their respective efficacy in reducing total inventory costs while maintaining high customer service levels.

The rolling horizon framework is encapsulated within a Discrete Event Simulation (DES) architecture (Meissner et al., 2021), which allows a periodic simulation of demand forecasting, inventory replenishment, component failures and replacement events, as shown in Figure 1. At each period, forecasts and inventory planning decisions are generated over a finite horizon, while only the immediate actions are executed. As the simulation advances, the horizon rolls forward, incorporating updated information on the component's condition as well as inventory status, to better prepare the spare parts inventory for the uncertainties associated with stochastic demands. The individual phases are going to be explained in more detail in the following sections, providing readers with information about the analytical methods, theoretical foundations, and the assumptions undertaken at each stage of the architecture.

At the conclusion of the simulation, the prediction methods are evaluated based on the following two performance criteria.

Total cost: - Represents the cumulative cost incurred over the entire simulation. It consists of the components like ordering cost (C_{Ordering}), holding cost (C_{Holding}), and stock-out cost ($C_{\text{Stock-out}}$), and can be expressed as

$$C_{\text{total}} = C_{\text{Ordering}} + C_{\text{Holding}} + C_{\text{Stock-out}} \tag{1a}$$

with

$$C_{\text{Ordering}} = \sum_{t=1}^{T} O_t \tag{1b}$$

$$C_{\text{Holding}} = \sum_{t=1}^{T} h\, I_t \tag{1c}$$

$$C_{\text{Stock-out}} = \sum_{t=1}^{T} p\, S_t \tag{1d}$$

where O_t denotes the fixed cost incurred when an order is placed in period t, h is the unit holding cost per period, I_t represents the on-hand inventory level at the end of period t, p is the penalty cost per unit short, and S_t symbolizes the unfulfilled demand in period t.

Service Level: Measures the ability of the inventory system to satisfy demand without shortages. It is calculated as

$$SL = (1 - \frac{\sum_{t=1}^{T} S_t}{\sum_{t=1}^{T} D_t}) \cdot 100 \tag{2}$$

where D_t denotes the demand in period t and S_t represents the number of stock-outs. A higher service level indicates better demand fulfillment performance and improved operational reliability.

2.1. System Analysis

At the beginning of each epoch, a fleet wide check is performed to identify the active/working components, their current age, and their true/actual failure date (often referred to as "ground truth"). This analysis is essential for two purposes.

First, it allows the simulation to generate the actual failure date for a component upon introduction, by drawing samples from a desired distribution. This study makes use of a Weibull

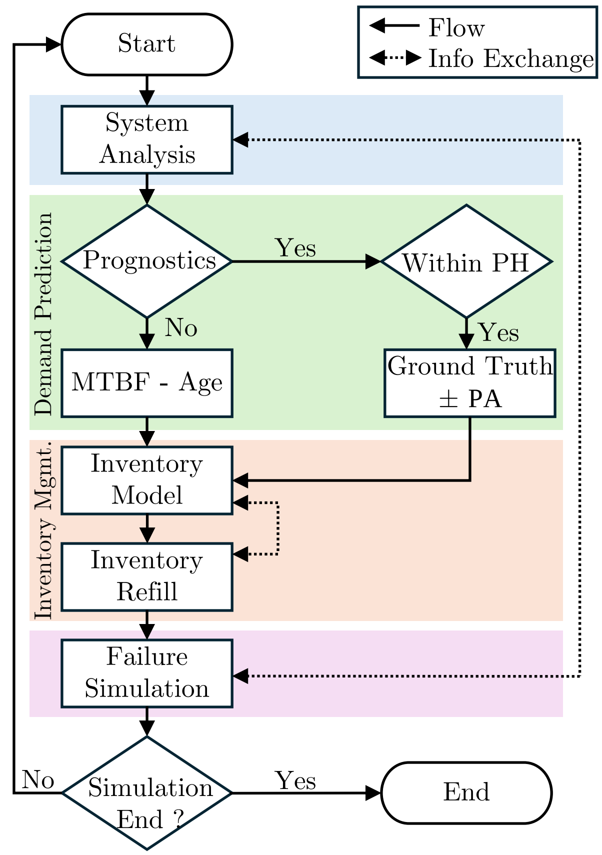


Figure 1. Simulation Architecture

distribution, for it's ability to represent different hazard properties as well as age-dependent failure behavior for a component (Ross, 2014; J. Wang & Yin, 2019).

The Weibull distribution consists of the following two parameters that determine the failure rate evolution and overall reliability characteristics of a component.

Shape Parameter (β). The parameter β dictates how the failure rate behaves over time. Based on the value of β, the failure rate can represent one of the following three conditions:

- $\beta < 1$: Early failures are prominent, with reliability improving with age.
- $\beta = 1$: The distribution shows a constant failure rate, meaning failures occur randomly. In this case, the Weibull distribution reduces to an exponential distribution.
- $\beta > 1$: The failure rate increases over time, with aging components failing more frequently.

A visual representation of these conditions is shown in Figure 2

Scale parameter (η). The parameter η on the other hand determines the spread of the distribution along the time axis. If the Mean Time Between Failure (MTBF) of a component is

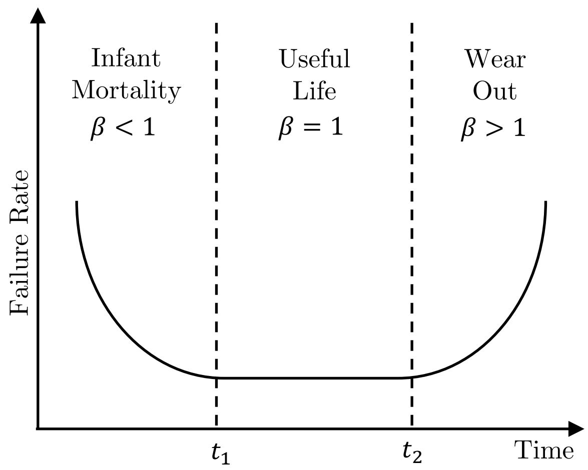


Figure 2. Bathtub Curve

known, the scale parameter can be calculated using

$$\eta = \frac{\text{MTBF}}{\Gamma(1 + 1/\beta)} \tag{3}$$

where Γ represents the Gamma function.

For a given pair of β and η parameters, the time to failure for a component can then be calculated, by drawing samples from the inverse Cumulative Distribution Function (CDF) of the Weibull distribution through

$$T = \eta[-\ln(1-U)]^{\frac{1}{\beta}} \tag{4}$$

where, $U \sim \mathcal{U}(0,1)$ denotes the uniform random variate used to generate Weibull-distributed samples.

The second purpose of this analysis, is to allow the simulation to assimilate up-to-date information on the operational status of the component (active or inactive) along with its accumulated age. This information forms the basis for determining the component's true RUL – a prerequisite for prognosis. If a component is active, its RUL is calculated by

$$\text{RUL} = T_{\text{Fail}} - T_{\text{Current}} \tag{5}$$

with T_{fail} as actual (or predicted) failure time and T_{current} as current age of the component. Once a component reaches its failure time, its state changes to inactive. When a component is inactive, the associated asset is assumed to be grounded. In this situation, the component ceases ageing and, unless replaced, incurs stock-out penalties.

2.2. Demand Prediction

The next step in the DES constitutes the main focus of this study. Using the information about the component's state derived from the previous step, two distinct methodologies for failure predictions are employed.

2.2.1. Reliability-Based Prediction

The statistics- or reliability-based approach makes use of conditional probabilities to determine the proportion of components which are expected to survive until a given time. Since the component's actual failure date is modeled through a Weibull distribution, the reliability function, representing the probability that a component survives at least until time t, can be expressed as (Nelson, 1982)

$$R(t) = \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \tag{6}$$

From the reliability function, the corresponding CDF, which gives the probability of failure by time t can be defined with

$$F(t) = 1 - R(t) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \tag{7}$$

Traditionally, reliability-based maintenance planning relies on MTBF information as a representative lifetime measure. However, as MTBF reflects only the mean value of the distribution, it should be noted that it does not translate to the median. For example, in case β=1, once the MTBF is reached, 63.2% of the components will have failed already (Wolf, Meissner, Pohya, & Wende, 2026). In consequence, strictly using MTBF data for inventory replenishment planning may lead to distorted failure time estimates, depending on the specific distribution shape.

Therefore, instead of relying on raw MTBF information, we invert the cumulative failure function to determine the time at which a specified fraction p of components is expected to have failed with

$$t_p = \eta\left[-\ln(1-p)\right]^{1/\beta}, \tag{8}$$

where, p denotes the proportion of components expected to have failed by time t_p. Lower values of p correspond to conservative early replacement decisions, whereas higher values lead to more aggressive strategies. For cost and service-level considerations, this study utilizes the median lifetime between successive failures, i.e., p=0.5, as it provides a more balanced approach.

2.2.2. Prognostics-Based Predictions

The other approach for predicting component failures is to employ prognostics, where the core idea is to use prognostic performance metrics as surrogates for a physics-based or data-driven prognostic model. Instead of directly modeling component degradation, the algorithm relies on these performance indicators to generate failure predictions, which emulate the functioning of an actual prognostic model.

This substitution allows the influence of prognostic quality on maintenance and logistics decisions to be studied in isolation. In particular, the results enable the identification of minimum performance thresholds that must be achieved before the use of prognostics-based approaches provide any tangible benefits over purely reliability-based ones.

This study makes use of a simplified version of the α-λ performance metrics (Saxena, Celaya, Saha, Saha, & Goebel, 2010) to simulate failure prognosis as a proxy for a real prognostic model. Here, α or PA represents the acceptable prediction accuracy band around the true RUL, while t_λ defines the earliest time after which all the predicted values adhere to the bounds set by PA, as shown in Figure 3. The difference between the End of Life (EOL) t_{EOL} and t_λ defines the PH.

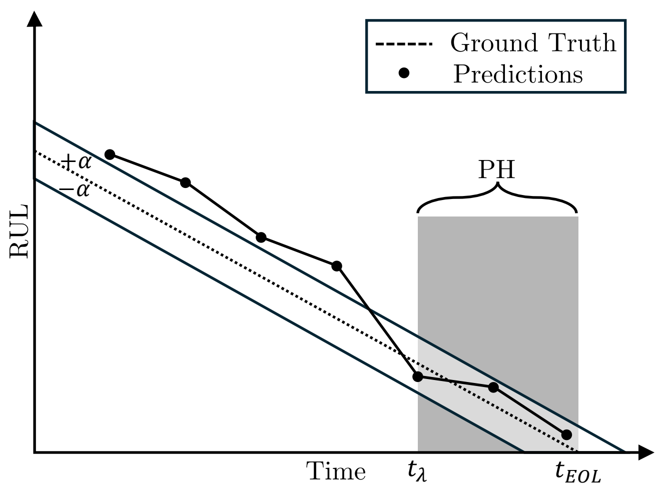


Figure 3. α – λ Performance Metric

In this study, for each component that enters the DES, a t_λ is established by calculating the difference between the component's actual time to failure (t_{EOL}) and the user specified PH. Once, the components accumulated age exceeds that of t_λ, predicted failure times are generated such that their deviation from the true RUL remains within the user specified PA bounds. This approach enables a controlled experimentation on the impact of PA and PH without introducing the additional complexity and uncertainty associated with implementing a full prognostic model.

For a user-specified error bound PA > 0, the prediction error ε is modeled as a bounded continuous random variable defined over the interval [-PA, PA]. For the calculation, the following two alternative distributional assumptions are considered.

Random (uniform) error distribution. In a standard scenario, the prediction error is assumed to be uniformly distributed over the limits established by PA, implying that all deviations within the bounds are equally likely. The corresponding Probability Density Function (PDF) is then given by

$$f_\varepsilon(x) = \begin{cases} \dfrac{1}{2\text{PA}} & \text{if} -\text{PA} \le x \le \text{PA}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Skewed (scaled beta) error distribution. To incorporate prediction bias (Lee & Mitici, 2023), the error term is alternatively modeled using a scaled Beta distribution defined over the same bounded support. Let $Z \sim \text{Beta}(a, b)$ with shape parameters $a, b > 0$, the error variable is obtained via the linear transformation

$$\varepsilon = 2\alpha Z - \alpha \tag{10}$$

Then, the results in the following PDF can be calculated as

$$f_\varepsilon(x) = \begin{cases} \dfrac{1}{2\text{PA}\, B(a,b)} \left(\dfrac{x+\text{PA}}{2\text{PA}}\right)^{a-1} \\ \quad \times \left(1 - \dfrac{x+\text{PA}}{2\text{PA}}\right)^{b-1} & \text{if } -\text{PA} \le x \le \text{PA}, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

where $B(a, b)$ denotes the Beta function.

Here, a left-skewed error structure arises when $a > b$, resulting in a higher probability mass near -PA (early failures). Conversely, a right-skewed distribution is obtained when $a < b$, indicating a greater likelihood of prediction near +PA (late failures).

Upper/lower bounds of PH and PA. The value of PH is varied from 10% to 100% in relation to the MTBF, to allow for a comprehensive analysis of its effects on the resulting cost and service levels. The value of PA, however, is subject to constraints imposed by another prognostic metric, i.e., relative accuracy, which measures how close the predicted RUL (RUL_{Pred}) is to the true RUL (RUL_{True}), so that

$$\text{RA} = 1 - \frac{|\text{RUL}_{\text{Pred}} - \text{RUL}_{\text{True}}|}{\text{RUL}_{\text{True}}} \tag{12}$$

Relative accuracy (RA) is bounded in the interval $[0, 1]$, where a value of 1 corresponds to a perfect prediction and 0 represents the maximum allowable deviation (Saxena et al., 2010). From Eq. (12), a perfect prediction ($RA = 1$) occurs when

$$\text{RUL}_{\text{Pred}} = \text{RUL}_{\text{True}} \tag{13}$$

In contrast, the worst-case scenario ($RA = 0$) arises when the prediction error equals the true remaining useful life,

$$|\text{RUL}_{\text{Pred}} - \text{RUL}_{\text{True}}| = \text{RUL}_{\text{True}}, \tag{14}$$

which leads to the bounds

$$\text{RUL}_{\text{Pred}} = 0 \quad \text{or} \quad \text{RUL}_{\text{Pred}} = 2\,\text{RUL}_{\text{True}} \tag{15}$$

While RA captures the magnitude of the prediction error, it does not distinguish between overestimation and underestimation. To incorporate this directionality, the PA is defined as the signed error between the predicted and true RUL,

$$\mathrm{PA} = \mathrm{RUL}_{\text{Pred}} - \mathrm{RUL}_{\text{True}} \tag{16}$$

To evaluate these bounds in the proposed prognostics-driven methodology, consider the onset of the prediction horizon, where predictions are first issued. At this point, $\mathrm{RUL}_{\text{True}} = PH$ by definition, as illustrated in Fig. 4. Substituting $\mathrm{RUL}_{\text{True}}$ with PH in Eq. (16) for the worst-case bounds derived from $RA = 0$ yields $\mathrm{PA} = -PH$ when $\mathrm{RUL}_{\text{Pred}} = 0$, and $\mathrm{PA} = +PH$ when $\mathrm{RUL}_{\text{Pred}} = 2\,PH$. The two extremes define the admissible range of prediction accuracy, thus deeming a prediction acceptable if and only if PA satisfies

$$-PH \leq \mathrm{PA} \leq PH, \tag{17}$$

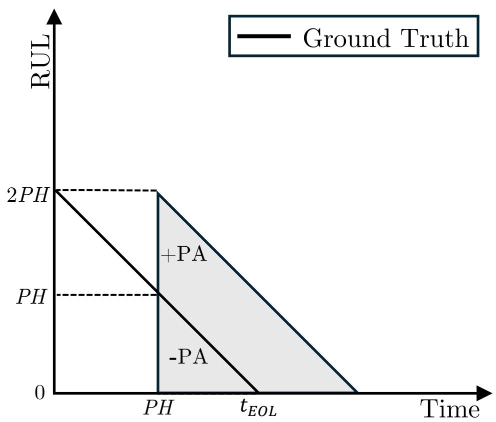


Figure 4. Admissible region for prediction accuracy at the prediction horizon. A prediction is considered acceptable if $\mathrm{PA} \in [-PH, +PH]$, corresponding to the worst-case bounds derived from $RA = 0$.

False alarms. To enhance the realism and incorporate the uncertainties associated with diagnosis, a provision to account for FA is also included. FA occur when a faulty diagnosis may force the operator to conduct an immediate replacement/or repair of the machine (Vachtsevanos & Zahiri, 2022). In other words, irrespective of the of RUL, if a FA is raised, the prognostic system treats it as an imminent failure, which in turn stipulates an immediate repair/replacement of the component. In this study, the FA rate is varied between 0% and 10%.

2.3. Inventory Management

Using the predictions from the previous stage, a failure calendar is generated which represents the spare part demand. That demand is then utilized by the inventory model to determine the schedule for spare part replenishment. Here, the study uses the Wagner-Whitin (WW) dynamic lot sizing algorithm (Wagner & Whitin, 2004). The WW is a classical lot-sizing model used to determine the optimal ordering schedule and production quantities over a finite planning horizon. The objective of this model is to minimize the total inventory, by optimizing for the ordering and holding costs while ensuring that the demand for each period is satisfied. According to WW algorithm, the total cost of placing an order in period t to satisfy demand up to period j is given by

$$C(t, j) = C_{\text{Ordering}} + C_{\text{Holding}} \sum_{k=t}^{j} (k - t)\, d_k, \tag{18}$$

where k and d_k represent the periods being covered and the demand associated with the period respectively. The minimum cumulative cost $F(j)$ to satisfy demand up to period j can then be determined recursively through

$$F(j) = \min_{1 \leq t \leq j} \{F(t-1) + C(t, j)\} \tag{19}$$

with $F(0) = 0$.

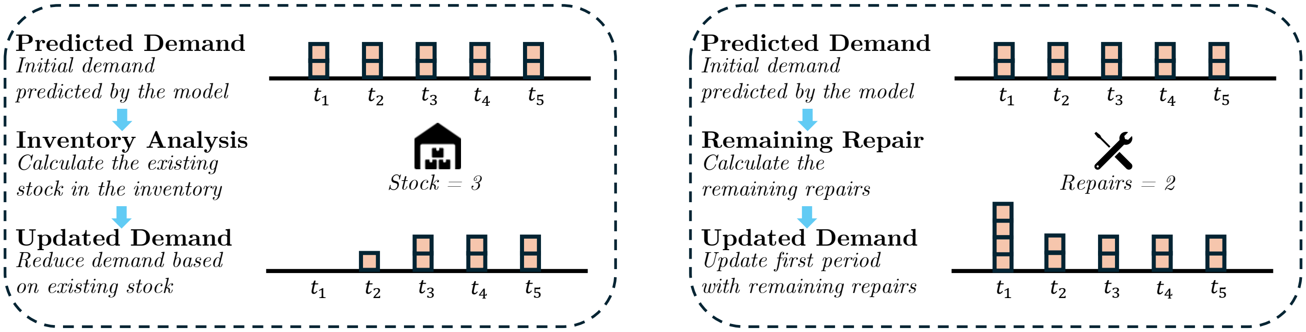
The recursion evaluates all feasible ordering periods t that could satisfy demand up to period j and selects the one that minimizes the total cumulative cost, thereby yielding the optimal replenishment schedule.

Within the rolling-horizon framework, the WW algorithm is resolved at each decision epoch to generate an updated ordering plan. However, only the replenishment decision corresponding to the current period is implemented, while future order recommendations remain provisional and are revised at subsequent periods as new information (e.g., on failure times) is available. The applicability of the WW algorithm in rolling-horizon settings has been demonstrated in prior studies (Campuzano-Bolarín, Mula, Díaz-Madroñero, & Legaz-Aparicio, 2020), though it calls for some modification in the way assumptions are handled. In the classical setting, the WW formulation relies on several underlying assumptions:

- *A1:* The forecast horizon is finite.
- *A2:* Backorders are not allowed.
- *A3:* The initial and final inventory levels are zero.
- *A4:* In each period, only a single lot may be produced.
- *A5:* Replenishments are instantaneous.

The classical WW formulation assumes that demand is satisfied in each period without shortages and that inventory levels are zero at the beginning and end of the planning horizon. In a rolling-horizon implementation, however, the inventory position at the start of each decision epoch may be non-zero, and unfulfilled demand from previous periods may exist. Consequently, assumptions *A2* and *A3* may be violated and require appropriate adjustments to the failure calendar before applying the WW algorithm.

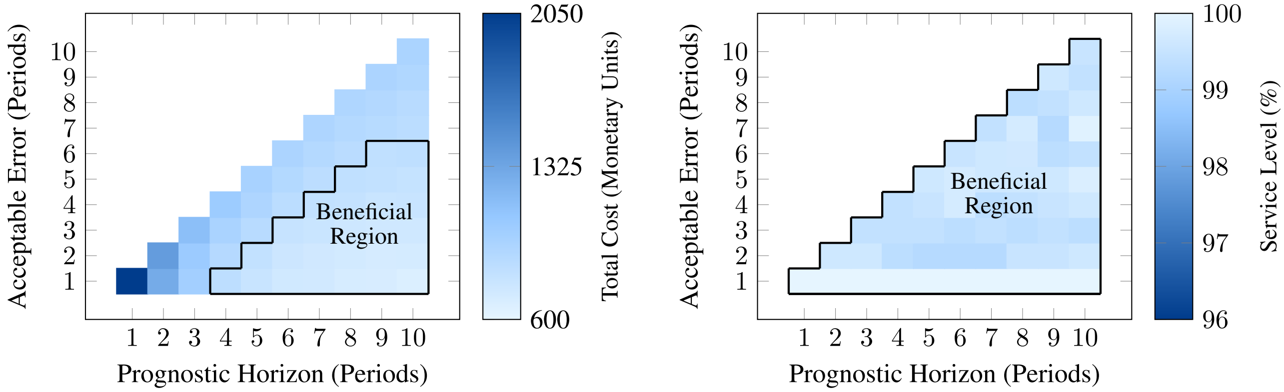
In this study, if positive inventory exists at the beginning of the horizon, the predicted demand must be reduced to account



(a) Existing inventory

(b) Unfulfilled demand

Figure 5. Adjustment to predicted demand based on inventory and repair status



(a) Total Cost

(b) Service Level

Figure 6. Heatmaps visualising the impact of prognostic metrics on performance parameters.

for the components already in stock. Conversely, if previous demand has not been fully satisfied (i.e., existence stock-outs), emergency replenishment are introduced in the initial period of the failure calendar to clear the backlog. The adjusted demand sequence is then passed to the WW algorithm to determine the optimal lot-sizing decisions, as shown in Figure 5.

2.4. Failure Simulation

The final stage in the DES for a given period is the realization of failures. If a component on a particular period reaches the end of its RUL, based on the ground truth, it is considered to have failed. If the necessary spares exist in the inventory, the component is going be replaced and a new RUL will be generated for the replacement item. In contrast, if a replacement cannot be performed due to insufficient spare parts in inventory, a stock-out penalty cost is incurred and added to the total inventory cost for each unmatched demand. Additionally, the event is recorded as unfulfilled demand and is incorporated into the service level calculation.

3. RESULTS

With the methodology described, a number of parameter studies are performed to assess the failure prediction approaches and compare their respective performances with respect to the simulation parameters highlighted in Table 1.

Table 1. Simulation Parameters

Parameter	Value
Total Simulations	10,000
Simulation Duration (Days)	30
Ordering Cost (Units)	100
Holding Cost (Units)	1
Stock-out Cost (Units)	10
MTBF (Days)	10
Beta	2

3.1. Impact of Prognostic Horizon and Prediction Accuracy

The results illustrating the effects of PH and PA on total cost and service level are presented in Figures 6a and 6b respectively. The shaded region within the figure indicate combinations of PH and PA that outperform baseline values obtained from the statistical method of prediction shown in Table 2.

Table 2. Total cost and service level obtained from statistical demand prediction.

Total Cost (Monetary Units)	Service Level (%)
824	96.4

3.1.1. Effect on Total Cost

The analysis of the effect of prognostic variation on total cost highlights two trends. First, increasing the value of PH leads to a continuous reduction in total cost. This is due to the fact that a longer prognostic horizon window provides the WW algorithm with enhanced visibility into future demand, enabling more optimal lot-sizing decisions. An increased foresight facilitates better order consolidation and reduces unnecessary inventory replenishment. In most real-world settings, where ordering costs are significantly higher than holding costs, such consolidation leads to substantial total cost reductions.

Secondly, for a fixed PH, decreasing the PA further reduces total cost. Lower PA values correspond to tighter prediction error bounds and thus more accurate failure forecasts. Reduced uncertainty in failure timing enables the WW inventory model to generate more reliable replenishment schedules, limiting both premature ordering and costly reactive decisions.

Hence, improvements in prognostic horizon and predictive accuracy directly translate to reductions in total cost.

3.1.2. Effect on Service Level

In contrast to the behaviour observed with total cost, the service level remains consistently close to 100% across nearly all the combinations of PH and PA. This indicates that, within the current logistics framework, service level performance is mostly insensitive to prognostic parameters. This outcome is primarily because of the assumptions associated with the implementation of WW algorithm. The WW formulation does not incorporate lead-times. Consequently, even delayed predictions are sufficient to initiate replenishment before failure realization, such that stock-outs arise only under consistent positive PA errors. Given that stock-outs are the exclusive determiners of the service levels, the absence of lead-time constraints on a simulation standpoint suppresses stock-out risk.

In a bid to understand the effect of lead-time on inventory performance, further analysis was performed by introducing stochastic replenishment lead-times into the simulation framework. Two lead-time scenarios were considered: $L \in \{0, 1\}$ and $L \in \{0, 1, 2\}$, corresponding to discrete uniform sampling over the respective ranges for the time periods. Figures 7a and 7b present heatmaps of service level across combinations of PH and PA under these lead-time conditions, with highlighted regions indicating parameter settings that outperform the reliability-based benchmark, represented in Table 3.

The results indicate that increasing lead-time variability leads to a deterioration in service level performance. In addition, longer lead-times reduce the region of the prognostic parameter space for which the prognostic policy outperforms the reliability-based benchmark, indicating that the effectiveness of prognostic information becomes increasingly constrained by lead-time dynamics.

Table 3. Service Level obtained from reliability based prediction under different lead-time variability.

$\mathbf{L} \in \{\mathbf{0}, \mathbf{1}\}$	$\mathbf{L} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}$
92 %	90 %

A secondary observation is that, under stochastic lead-times, configurations with higher PA may occasionally exhibit improved service levels relative to scenarios with lower PA or permissible error. This effect arises because increased prediction tolerance can induce earlier or more frequent replenishment triggers, which in some cases partially compensate for delays introduced by non-zero lead-times. However, this behaviour is strictly regime-dependent and should not be interpreted as evidence that higher prognostic error is beneficial. Instead, it reflects an interaction between prediction variability and lead-time uncertainty that affects the timing of replenishment decisions. In practical applications, higher prognostic error would generally be expected to degrade performance due to reduced reliability of failure anticipation and increased risk of inappropriate decision triggering..

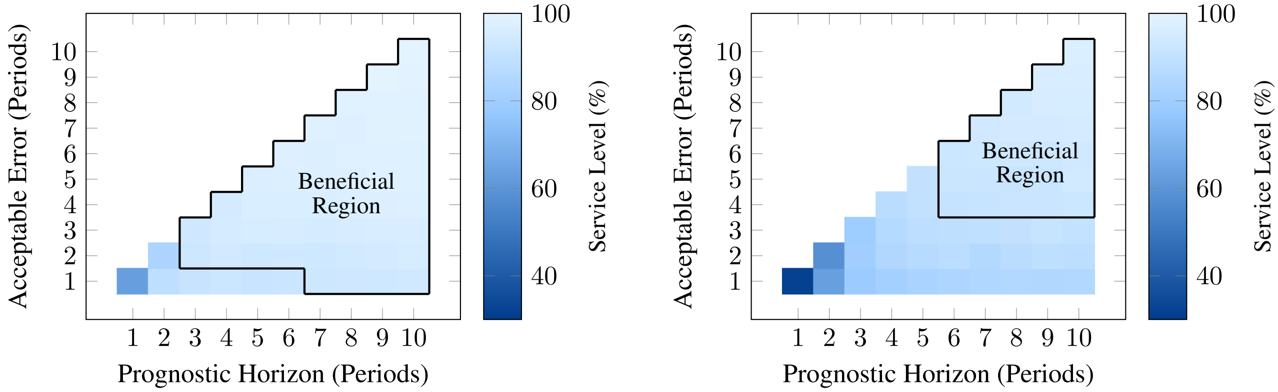
3.2. Impact of Skewed Predictions

The subsequent analysis is conducted by introducing a skew to the PA predictions, with the mean of the left and right skews set at $-{}^{\mathrm{PA}}/_{2}$ and $+{}^{\mathrm{PA}}/_{2}$, respectively. For the representative case, PH and PA are fixed at 10 and 4, while all other simulation parameters are maintained as in the previous case.

3.2.1. Effect on Total Cost

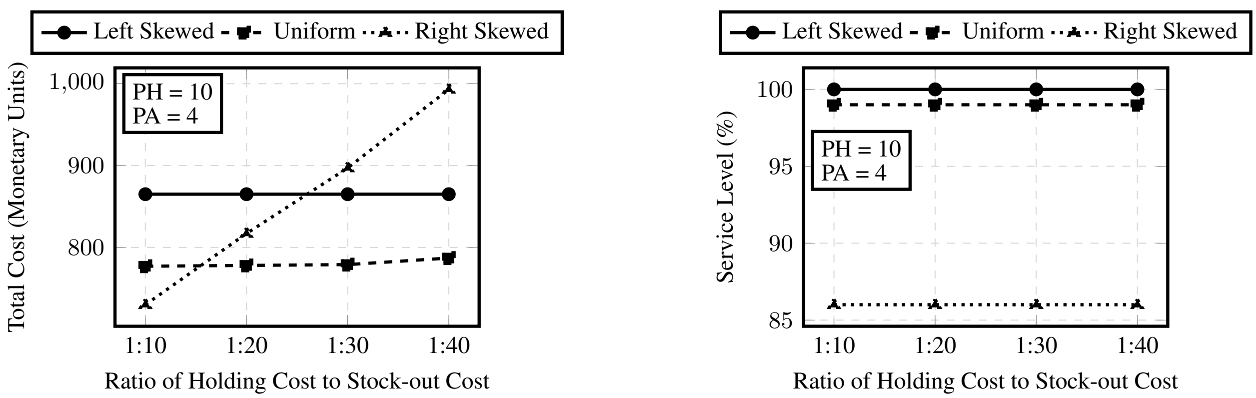
Figure 8a depicts the result on total cost arising from prediction skewness and the ratio of holding to stock-out cost. As can be prominently observed in the figure, the performance of left and right skewed predictions differ sharply as the stock-out costs increase relative to holding cost. In particular, total cost under right-skewed errors (delayed predictions) increases significantly, whereas in the case of left-skewed errors (early prediction), the total cost remains comparatively stable.

This divergence is caused by the way WW algorithm optimizes the demand. Right-skewed errors postpone replenishment



(a) $L \in \{0, 1\}$ (b) $L \in \{0, 1, 2\}$

Figure 7. Heatmaps visualising the impact of lead time on service level performance.



(a) Total cost impact (b) Service level impact

Figure 8. Influence of prediction skewness and cost ratio

decisions, increasing the probability of unfulfilled demand. When stock-out penalties are low, this strategy may remain economically beneficial, as it is able to evade extensive holding cost which may occur when the inventory is prematurely replenished in the event of early predictions. However, as the stock-out cost parameter increases, the cost of delayed replenishment grows disproportionately, causing total cost to escalate. Conversely, left-skewed errors shift predicted demand forward in time, prompting earlier replenishment. Although this increases average inventory on hand and therefore holding cost, the probability of demand being unfilled on the other hand reduces as well. This is beneficial in scenarios where disruptions in operation arising from stock-out is not favorable.

3.2.2. Effect on Service Levels

In contrast, the service level, as shown in Figure 8b, exhibits a consistent trend: early predictions (left skew) always trump over delayed ones (right skew), irrespective of the underlying cost ratio. Delayed predictions often prompt the WW algorithm to postpone replenishment decisions, thereby reducing the stock on hand and increasing the exposure to stock-out events. As a result, the organization becomes more vulnerable to stock-outs under positively skewed prediction errors. Since service level performance is driven exclusively by the ability of an organization to successfully satisfy the incoming demand, any delay in spare part replenishment directly deteriorates service performance. This behavior reflects the strategic difference between proactive and reactive decisions in safety critical systems, where maintaining inventory levels is typically favored over reacting close to the actual point of failure.

3.3. Influence of False Alarms

The robustness of the framework against external shocks is further analyzed by incorporating FAs. To study only the effect of FAs, the simulation was isolated from the effects of PH and PA; therefore, these values are fixed and set to 10 and 0, respectively. As illustrated in Figure 9, the total cost increases approximately proportionally with the false alarm rate. This

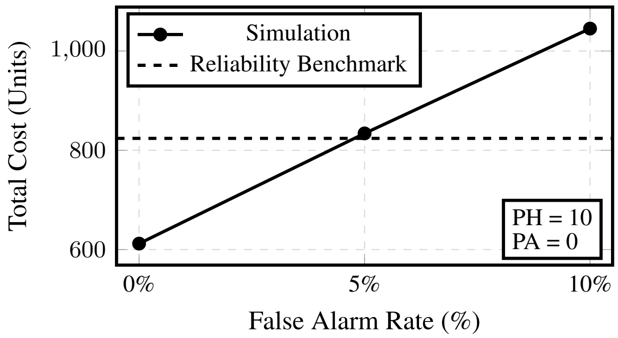


Figure 9. Total cost for different FA rates with PH set to 10.

behavior is primarily driven by premature replacement decisions triggered by incorrect failure predictions, which lead to unnecessary replenishment orders and consequently higher ordering costs. A similar analysis to study the effects of FA on service level is not performed, due to the constraints associated with the current setup. Since lead-time are not considered, replenishment occur instantaneously. As a result, the service level are not affected, unless the predictions are delayed.

Figure 10, further illustrates that total cost increases with deteriorating prognostic accuracy and for a given FA rate. This is consistent with earlier observations highlighted in figure 18, where PA degradation leads to delayed replenishment decisions and increased exposure to stock-outs. However, the marginal impact of PA degradation becomes less pronounced in the presence of FA. This does not imply that FA improves system performance, but rather reflects the interaction within the simulation where false alarms lead to earlier triggers for ordering that partially compensate for mistimed or delayed predictions under the modeled assumptions, which might lead to the assumption that FA and early predictions may appear beneficial by reducing the severity of replenishment timing errors and mitigating stock-out exposure.

But, it is important to highlight that this effect is regime-dependent and should not be interpreted as universally desirable. FA does not reduce prognostic error nor alter the underlying failure process; it only shifts decision timing. Under the zero lead time assumption adopted in this study, premature actions may incidentally offset delayed predictions. However, in real-world MRO systems, where lead-times, maintenance costs, maintenance capacity constraints, and operational disruptions are present, excessive false alarms would likely introduce significant penalties. These may include unnecessary maintenance actions, increased labor and replacement costs, premature disposal of components with remaining useful life, higher material waste, and broader environmental impacts.

3.4. Modeling Limitations

The findings portrayed above, should however be interpreted with caution. While the analysis has been extended to include

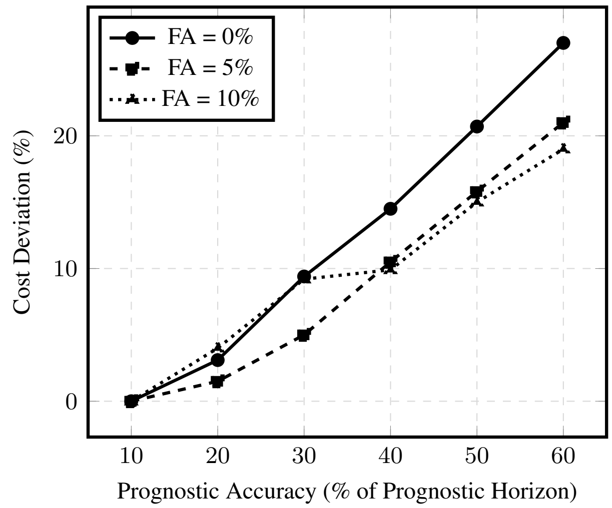


Figure 10. Cost deviation as a function of prognostic accuracy for different false alarm rates.

stochastic lead-times for replenishment, these are represented using simplified discrete distributions and do not fully capture the complexity of real-world supply chain behaviour, such as correlated delays, supplier disruptions, or capacity-dependent lead-times. Furthermore, maintenance actions are assumed to be immediately executable upon decision triggering, without accounting for maintenance crew availability or operational scheduling constraints. These simplifications may influence the magnitude of the observed effects, although the qualitative trends are expected to remain valid.

4. CONCLUSION

This study presents a methodology to evaluate the performance of prognostics against traditional reliability-based approaches in the context of spare parts inventory management. By representing prognostic performance through metrics such as prognostic horizon, prognostic accuracy, prediction bias, and false alarm rate, the framework provides a systematic means of assessing the impact of prognostic quality on inventory decisions and operational outcomes. The results demonstrate that increasing prognostic horizon and tightening prediction error bounds generally reduce total costs by enabling more effective replenishment timing and lot-sizing decisions. Furthermore, prediction bias was found to significantly influence inventory performance, with left-skewed predictions improving service levels at the expense of higher holding costs, while right-skewed predictions reduce inventory costs but increase the risk of stock-outs. False alarms also contribute directly to higher costs through premature replenishment actions and elevated inventory levels.

The introduction of stochastic lead-times significantly alters the relationship between prognostic performance and inventory outcomes. Increasing lead-time variability leads to a deterioration in service levels and reduces the region of prognostic parameter combinations for which the prognostic policy outperforms the reliability-based benchmark. The findings further indicate that the interaction between prediction uncertainty, false alarms, and supply-side variability can produce non-intuitive behaviour in specific operating regimes, highlighting that the value of prognostic information is strongly dependent on the logistics environment in which it is applied. Consequently, prognostic performance should be evaluated within the context of realistic inventory and replenishment constraints rather than in isolation.

Future work will focus on extending the framework to incorporate more realistic operational conditions, including correlated lead-times, manufacturing constraints, and disruption-driven variability in supply processes. In addition, the framework will be expanded to consider multi-item spare parts systems, multi-echelon inventory networks, and maintenance scheduling constraints. These extensions will enable a more comprehensive assessment of prognostics-driven inventory policies and support the development of more resilient and cost-effective maintenance planning strategies in aerospace and other asset-intensive industries.

NOMENCLATURE

AOG Aircraft on Ground.
CBM Condition Based Maintenance.
CDF Cumulative Distribution Function.
CNN Convolutional Neural Network.
DES Discrete Event Simulation.
EOL End of Life.
FA False Alarm.
IOT Internet of Things.
LRU Line Replaceable Unit.
LSTM Long Short-Term Memory.
MRO Maintenance, Repair and Overhaul.
MTBF Mean Time Between Failure.
PA Prognostic Accuracy.
PDF Probability Density Function.
PH Prognostic Horizon.
PHM Prognostic Health Management.
RUL Remaining Useful Life.
SBA Syntetos and Boylan Approximate.
TSB Teunter-Syntetos-Babai.
WW Wagner-Whitin.

REFERENCES

Alomar, I., & Nikita, D. (2025). Managing operational efficiency and reducing aircraft downtime by optimization of aircraft on-ground (aog) processes for air operator. *Applied Sciences*, *15*(9). doi: 10.3390/app15095129

Bousdekis, A., Papageorgiou, N., Magoutas, B., Apostolou, D., & Mentzas, G. (2017). A proactive event-driven decision model for joint equipment predictive maintenance and spare parts inventory optimization. *Procedia CIRP*, *59*, 184–189. doi: 10.1016/j.procir.2016.09.015

Cai, B., Wang, Y., Zhang, Y., Liu, Y., Ge, W., Li, R., ... Liu, G. (2022). Condition-based maintenance method for multi-component system based on rul prediction: Subsea tree system as a case study. *Computers & Industrial Engineering*, *173*, 108650. doi: https://doi.org/10.1016/j.cie.2022.108650

Campuzano-Bolarín, F., Mula, J., Díaz-Madroñero, M., & Legaz-Aparicio, Á.-G. (2020). A rolling horizon simulation approach for managing demand with lead time variability. *International Journal of Production Research*, *58*(12), 3800–3820. doi: 10.1080/00207543.2019.1634849

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

Deng, Y., Bucchianico, A. D., & Pechenizkiy, M. (2020). Controlling the accuracy and uncertainty trade-off in rul prediction with a surrogate wiener propagation model. *Reliability Engineering & System Safety*, *196*, 106727. doi: https://doi.org/10.1016/j.ress.2019.106727

Fang, X., Paynabar, K., & Gebraeel, N. (2017). Multistream sensor fusion-based prognostics model for systems with single failure modes. *Reliability Engineering & System Safety*, *159*, 322-331. doi: https://doi.org/10.1016/j.ress.2016.11.008

Hoff, T., Sprecher, B., Pohya, A. A., Wende, G., & Peck, D. (2025). The impact of circular strategies on titanium supply and demand in the aviation industry. *Resources, Conservation and Recycling*, *222*, 108476. doi: https://doi.org/10.1016/j.resconrec.2025.108476

Kordestani, M., Orchard, M. E., Khorasani, K., & Saif, M. (2023). An overview of the state of the art in aircraft prognostic and health management strategies. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-15. doi: 10.1109/TIM.2023.3236342

Lee, J., & Mitici, M. (2023). Deep reinforcement learning for predictive aircraft maintenance using probabilistic remaining-useful-life prognostics. *Reliability Engineering & System Safety*, *230*, 108908. doi: https://doi.org/10.1016/j.ress.2022.108908

Mathew, V., Toby, T., Singh, V., Rao, B. M., & Kumar, M. G. (2017). Prediction of remaining useful lifetime (rul) of turbofan engine using machine learning. In *2017 ieee international conference on circuits and systems (iccs)* (p. 306-311). doi: 10.1109/ICCS1.2017.8326010

Meissner, R., Pohya, A. A., Weiss, O., Piotrowski, D., & Wende, G. (2025). Regulatory pathways to certifiable condition based maintenance solutions in aviation: A comprehensive review. *Progress in Aerospace Sciences*, *158*, 101143. doi:

https://doi.org/10.1016/j.paerosci.2025.101143

Meissner, R., Rahn, A., & Wicke, K. (2021). Developing prescriptive maintenance strategies in the aviation industry based on a discrete-event simulation framework for post-prognostics decision making. *Reliability Engineering & System Safety*, *214*, 107812. doi: https://doi.org/10.1016/j.ress.2021.107812

Nelson, W. (1982). *Applied life data analysis*. John Wiley & Sons Inc.

Nielsen, L. K., Kroon, L., & Maróti, G. (2012). A rolling horizon approach for disruption management of railway rolling stock. *European Journal of Operational Research*, *220*(2), 496–509. doi: 10.1016/j.ejor.2012.01.037

Oliver Wyman, & International Air Transport Association. (2025, October). *Aircraft supply chain: What's holding back the commercial aircraft supply chain — and where to go from here* (Technical Report). International Air Transport Association.

Reis, M. S., & Gins, G. (2017). Industrial process monitoring in the big data/industry 4.0 era: from detection, to diagnosis, to prognosis. *Processes*, *5*(3). doi: 10.3390/pr5030035

Romeijnders, W., Teunter, R., & van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, *220*(2), 386-393. doi: https://doi.org/10.1016/j.ejor.2012.01.019

Ross, S. M. (2014). *Introduction to probability and statistics for engineers and scientists*. Academic Press.

Sahin, F., Narayanan, A., & Robinson, E. P. (2013). Rolling horizon planning in supply chains: review, implications and directions for future research. *International Journal of Production Research*, *51*(18), 5413–5436. doi: 10.1080/00207543.2013.775523

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management*, *1*(1). doi: 10.36001/ijphm.2010.v1i1.1336

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, *21*(2), 303-314. doi: https://doi.org/10.1016/j.ijforecast.2004.10.001

Teubert, C., Pohya, A. A., & Gorospe, G. (2023). *An analysis of barriers preventing the widespread adoption of predictive and prescriptive maintenance in aviation.*

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, *214*(3), 606–615. doi: 10.1016/j.ejor.2011.05.018

Vachtsevanos, G., & Zahiri, F. (2022). Prognosis: Challenges, precepts, myths and applications. In *2022 ieee aerospace conference (aero)* (p. 1-13). doi: 10.1109/AERO53065.2022.9843823

Wagner, H. M., & Whitin, T. M. (2004). Dynamic version of the economic lot size model. *Management Science*, *50*(12), 1770–1774.

Wang, J., & Yin, H. (2019). Failure rate prediction model of substation equipment based on weibull distribution and time series analysis. *IEEE Access*, *7*, 85298-85309.

Wang, Z., Hu, C., Wang, W., Kong, X., & Zhang, W. (2015). A prognostics-based spare part ordering and system replacement policy for a deteriorating system subjected to a random lead time. *International Journal of Production Research*, *53*(15), 4511–4527. doi: 10.1080/00207543.2014.988892

Wolf, J.-A., Meissner, R., Pohya, A. A., & Wende, G. (2026). Flying electric: A comparative analysis of spare part demands and material cost for all-electric, hybrid-electric, and conventional aircraft propulsion systems. *Applied Energy*, *403*, 127036. doi: https://doi.org/10.1016/j.apenergy.2025.127036

Yongquan, S., Xi, C., He, R., Yingchao, J., & Quanwu, L. (2016). Ordering decision-making methods on spare parts for a new aircraft fleet based on a two-sample prediction. *Reliability Engineering & System Safety*, *156*, 40-50. doi: https://doi.org/10.1016/j.ress.2016.07.017