Escola d'Enginyeria

Universitat Autònoma de Barcelona

Optimization of safety stock in aeronautical maintenance using demand forecasting models

Thesis submitted for the degree on Aeronautical Management

Author: Oscar López Yáñez

Supervisor: Aleksandar Jovanovic

Sabadell, June 2026

The undersigned, Aleksandar Jovanovic, supervisor of the Final Degree Thesis, professor at the School of Engineering of UAB,

CERTIFIES:

That the work corresponding to this report has been carried out under their supervision by

Oscar López Yáñez

And for the record, signs this document in Sabadell, June of 2026

Signed: Aleksandar Jovanovic

RESUM - TREBALL DE FI DE GRAU DE L'ESCOLA D'ENGINYERIA

Títol del Treball de Fi de Grau: Optimització de l'estoc de seguretat en el manteniment aeronàutic mitjançant models de previsió de demanda	
Autor: Oscar López Yáñez	**Data:** Juny del 2026
Tutor: Aleksandar Jovanovic	
Titulació: Grau en Gestió Aeronàutica	

Paraules clau:

- **Català:** estoc de seguretat, manteniment aeronàutic, previsió de demanda, SARIMA, RMSE, MRO
- **Castellano:** stock de seguridad, mantenimiento aeronáutico, previsión de demanda, SARIMA, RMSE, MRO
- **English:** safety stock, aeronautical maintenance, demand forecasting, SARIMA, RMSE, MRO

Resum:

Castellano:

Este TFG analiza la optimización del stock de seguridad en el mantenimiento aeronáutico mediante la comparación de tres modelos de previsión de demanda (suavizado exponencial, SARIMA y una red neuronal artificial) aplicados a siete patrones sintéticos de consumo de repuestos. A partir del error de pronóstico (RMSE) obtenido por cada modelo en cada patrón, se calcula su impacto directo sobre la fórmula del stock de seguridad. El estudio identifica qué modelo se ajusta mejor a cada tipo de demanda, sin buscar un ganador universal. Finalmente, se traduce la mejora en precisión a términos económicos, cuantificando la reducción de capital inmovilizado. El trabajo combina simulación de datos, modelado estadístico y análisis financiero aplicado a la gestión de inventarios MRO.

Català:
Aquest TFG analitza l'optimització de l'estoc de seguretat en el manteniment aeronàutic mitjançant la comparació de tres models de previsió de demanda (suavitzat exponencial, SARIMA i una xarxa neuronal artificial) aplicats a set patrons sintètics de consum de recanvis. A partir de l'error de pronòstic (RMSE) obtingut per cada model en cada patró, es calcula el seu impacte directe sobre la fórmula de l'estoc de seguretat. L'estudi identifica quin model s'ajusta millor a cada tipus de demanda, sense buscar un guanyador universal. Finalment, es tradueix la millora en precisió a termes econòmics, quantificant la reducció de capital immobilitzat. El treball combina simulació de dades, modelatge estadístic i anàlisi financera aplicada a la gestió d'inventaris MRO.

English:
This thesis analyzes the optimization of safety stock in aircraft maintenance by comparing three demand forecasting models (exponential smoothing, SARIMA, and an artificial neural network) applied to seven synthetic spare parts consumption patterns. Based on the root mean square error (RMSE) obtained by each model for each pattern, the direct impact on the safety stock formula is calculated. The study identifies which model best fits each type of demand, without seeking a universal winner. Finally, the improvement in accuracy is translated into economic terms by quantifying the reduction in fixed capital. The work combines data simulation, statistical modeling, and financial analysis applied to MRO inventory management.

Table of Contents

1 Introduction **10**

1.1 Motivation and background of the problem 10
1.2 Objectives of the study 11
1.3 Scope and limitations of the study 12
1.4 Document structure 12

2 Aeronautical maintenance and the MRO context **14**

2.1 The MRO Sector: Definition, scope, and structure 14
2.2 The AOG concept and its operational impact 17
2.3 The MRO supply chain: Why generic models don't work 18
2.4 The regulatory framework: EASA and FAA 20

3 Stock management and safety stock **22**

3.1 Basic concepts: Demand, Lead Time, and Reorder point 22
3.2 The Classic formula for safety stock 23
3.3 Demand variability in aviation 23
3.4 Classification of aircraft spare parts 24
3.5 The limitations of the classical statistical approach 26

4 Forecast demand models **27**

4.1 Forecasting in inventory management 27
4.2 Exponential smoothing: Excel's FORECAST.ETS 28
4.3 SARIMA 30
4.4 Artificial Neural Networks (ANN) 30

5 Methodology and experimental design **32**

5.1 Demand simulation: Why synthetic and controlled 32
5.2 The seven demand patterns and their operational correspondence 33
5.3 Error metrics (MAE and RMSE) and conversion to a score for the radar . 34
5.4 Implementation: Python (SARIMA and ANN) and Excel (FORECAST.ETS) 35

6 Results and analysis of system error **36**
6.1 Fit for each series: The seven graphs discussed 36
6.2 The results matrix (MAE and RMSE by model and pattern) 37
6.3 Results by model 38
6.4 The effect of the seasonal period 39
6.5 Comparative visualization: The radar chart 39

7 Financial translation **41**
7.1 From error reduction to safety margin reduction 41
7.2 Estimated reduction in capital tied up in safety stock 42
7.3 Impact on inventory without increasing the risk of AOG 44

8 Conclusions **45**
8.1 Achieved and unachieved objectives 45
8.2 Study limitations 47
8.3 Future research directions 47

References **49**

A Python Source Code **51**
A.1 SARIMA Model 51
A.2 ANN Model 54

B Series-Level Diagnostic Plots **59**
B.1 S1 - Tires 59
B.2 S2 - Hydraulic Fluid 61
B.3 S3 - De-icing Fluid 62
B.4 S4 - Brakes 63
B.5 S5 - Engine Oil 64
B.6 S6 - Filters 65
B.7 S7 - Line Consumables 66

List of Figures

2.1 Share of engine maintenance spend by original equipment manufacturer (OEM), segmented by narrowbody and widebody aircraft. 15
2.2 Example of supplier consolidation in component maintenance spend: comparison between the Boeing 777 and the Boeing 787. 16

6.1 Fit of each series over the test period (14 days): actual demand versus Excel, SARIMA, and ANN forecasts. 36
6.2 RMSE by model and series over the test period. 37
6.3 RMSE heatmap by model and series. 37
6.4 Radar chart: score by model and series (based on MAE). 40

B.1 S1 Tires - actual daily demand over the 90-day training period (1 January - 31 March 2026). The upward trend and the recurring weekend peak are clearly visible, driven by leisure-flight traffic concentrating at the end of the week. 59
B.2 S1 Tires - actual demand versus Excel, SARIMA, and ANN forecasts over the 14-day test period (days 91-104). Excel and SARIMA track the actual series closely; the ANN diverges sharply from day 95 onward due to recursive error accumulation. 60
B.3 S1 Tires - RMSE by model over the test period. Excel (1.07) and SARIMA (1.19) are nearly tied; the ANN (6.86) is roughly six times worse. 60
B.4 S2 Hydraulic fluid -- training period. Demand is driven by scheduled maintenance tasks concentrated at the start of the week, producing a Monday--Tuesday peak that repeats throughout the quarter. 61
B.5 S2 Hydraulic fluid - test period. SARIMA tracks the actual series slightly better than Excel here, particularly at the peak on day 103. The ANN underestimates through most of the horizon. 61
B.6 S2 Hydraulic fluid - RMSE by model. SARIMA (2.80) outperforms Excel (2.89) by a narrow margin; the ANN (8.11) is the weakest by a wide margin. 61

B.7 S3 De-icing fluid - training period. Demand starts high in January and declines steadily as winter ends, giving this series the most pronounced downward trend in the set. 62
B.8 S3 De-icing fluid - test period. Excel follows the declining trend and the weekly oscillation well. SARIMA overshoots on several days, and the ANN produces a nearly flat forecast that ignores both the trend and the cycle. . 62
B.9 S3 De-icing fluid - RMSE by model. Excel (1.90) leads clearly; SARIMA (3.24) is penalised by its occasional overshoots; the ANN (5.63) is the least accurate. 62
B.10 S4 Brakes - training period. The series combines moderate growth with a weekend peak similar to tires, reflecting their shared dependence on leisure-flight volume. 63
B.11 S4 Brakes - test period. Excel and SARIMA are nearly indistinguishable and both follow the actual series tightly. The ANN lags the actual peaks by one to two days on several occasions. 63
B.12 S4 Brakes - RMSE by model. Excel (1.65) and SARIMA (1.78) are within 0.13 units of each other; the ANN (4.84) is notably worse. 63
B.13 S5 Engine oil - training period. This is the most level series in the set: the trend is mild and the weekly cycle is the most regular. It represents the steady baseline consumption of a mature narrowbody fleet. 64
B.14 S5 Engine oil - test period. Excel and SARIMA stay close to the actual line throughout. The ANN produces the most aberrant result of any series: despite being the most stable series in the set, the ANN error exceeds 12 units on some days. 64
B.15 S5 Engine oil - RMSE by model. SARIMA (1.43) edges out Excel (1.48) on this series; both are well below the ANN (7.82). 64
B.16 S6 Filters - training period. Demand peaks midweek, tied to planned maintenance windows. The series grows gradually over the quarter as the inspection calendar fills up toward spring. 65
B.17 S6 Filters - test period. Excel captures the midweek peaks reasonably well. SARIMA tends to lag peaks by one day; the ANN again shows wide swings that do not track the actual series. 65
B.18 S6 Filters - RMSE by model. Excel (2.42) leads over SARIMA (2.93); the ANN (5.43) is more than twice as far from the actual series. 65
B.19 S7 Line consumables - training period. Demand rises steadily from January to April, driven by increasing flight activity toward spring. The weekend peak is the sharpest in the set, reflecting the high concentration of leisure departures on Saturdays and Sundays. 66

B.20 S7 Line consumables - test period. Excel and SARIMA follow the actual series almost exactly, including the sharp weekend peak at day 103. The ANN diverges significantly on multiple days, producing the highest absolute errors in the whole test set. 66
B.21 S7 Line consumables - RMSE by model. Excel (1.38) and SARIMA (1.47) are very close; the ANN (8.31) posts its worst result across the seven series here. 66

List of Tables

5.1 The seven demand series: empirical characterization and operational logic . 34

7.1 Safety stock in units by model (Z = 1.65; LT = 7 days) 42
7.2 Capital tied up in safety stock and capital released (EUR) 43

8.1 Capital released per consumable versus the ANN (EUR) 46

Chapter 1

Introduction

1.1 Motivation and background of the problem

Managing an inventory of aircraft spare parts is not like conventional logistics. When a part is missing, the problem is not solved by delaying an order or apologizing to a customer. The aircraft is grounded. What the industry calls Aircraft on Ground (AOG) is a race against the clock. The direct cost of a grounded aircraft can exceed 50.000 per hour depending on the type of plane, and that's without counting contractual penalties with airlines or the damage to the MRO operator's reputation. The pressure is real and the historical response of the industry has been predictable. Overstocking ends up being the most expensive insurance in the world.

The problem is that this inventory has a very specific financial cost. Capital is sent for parts that generate no return and take up warehouse space. In fleets of a certain size, the parts catalog may contain thousands of components with vastly different consumption patterns. Some are used regularly. Others go weeks without being touched and then experience a spike that no one had anticipated.

The central element of the analysis is the safety stock (SS), the inventory buffer that an organization maintains above expected demand to absorb variations that no model can predict. Its classical formulation, under the assumption of a normal distribution, is as follows:

$$SS = Z \cdot \ \sigma_{error} \cdot \ \sqrt{LT}$$

Where Z is the factor associated with the target service level, σ error (RMSE) is the standard deviation of the forecast error, and LT is the supplier's lead time. Out of these three variables, the only one we can directly influence is σ error. If the forecasting model is less inaccurate meaning if the gap between what it predicts and what is actually consumed

is smaller, that deviation decreases. And when it decreases, the SS required to guarantee the same service level also decreases. There is no need to assume more AOG risk. You simply need less inventory.

That is the argument that runs throughout this thesis. Reducing the forecasting error translates into less tied-up capital, without compromising the fleet's operational availability. The question that remains open, and which this thesis attempts to answer, is how much gap for improvement exists by changing the model used to make that forecast, and whether that margin depends on the type of demand for each reference.

1.2 Objectives of the study

The overall objective of the project is to evaluate whether more sophisticated forecasting models improve the accuracy of the method the industry use, and to measure what that improvement would mean for safety stock. The comparison does not seek to identify a single winner. What interests us is understanding which model best fits each type of demand, because a stable consumption pattern and one with a marked trend do not pose the same challenge to the model.

To achieve this overall objective, we have defined five specific objectives:

1. Simulate seven daily demand series that replicate the typical consumption patterns of an MRO warehouse, each with its own combination of trend and weekly cycle, so that we know exactly the actual consumption underlying each series.
2. Implement three forecasting models on these series: the FORECAST.ETS model in Excel, which applies exponential smoothing and represents the industry standard; a SARIMA model, which extends ARIMA with seasonal components; and an artificial neural network, which learns nonlinear relationships without making assumptions about the series' structure.
3. Calculate the MAE and RMSE for each model over the test period and compile the results into a matrix that crosses model and demand pattern.
4. Determine, based on that matrix, where exponential smoothing is sufficient, where SARIMA adds value to it, and where the neural network outperforms the previous ones or fails.
5. Translate the reduction in error into the safety stock formula to quantify the capital that is freed up by adopting the more accurate model.

1.3 Scope and limitations of the study

The experiment is synthetic and controlled. Instead of using actual consumption data from a real fleet, we work with seven demand series generated for the study, recorded daily over a three-month period spanning winter and spring. This approach has a clear advantage: by generating the data, we know the underlying actual consumption of each series, and this allows us to isolate how each model responds to a specific pattern without the noise and gaps inherent in any real operational record. The error comparison is performed on the final segment of each series, set aside as a test period, which is where the three models produce their forecasts.

It is best to acknowledge the limitations from the outset, rather than leaving them for the conclusions. The first is the direct trade-off of working with simulated data. We gain control, but we lose the texture of a real historical record, with its recording errors and one-off events. The second is that the series do not include the unit price of each component, nor the supplier, nor the actual delivery time, so the financial translation relies on representative assumptions rather than cost figures from a specific operator. The third has to do with the nature of the series themselves. None of the seven exhibits intermittent demand, those long periods of zero activity interrupted by isolated spikes that characterize much of the aerospace catalog. The unpredictable demand regime, which we discussed in the theoretical framework, is therefore outside the scope of this study, and we note it as a topic for future research.

This research does not aim to integrate any model into an ERP system or put it into production. The scope is analytical. The goal is to demonstrate that the approach works on controlled data and that the potential financial impact justifies considering its implementation. What comes after that demonstration is beyond the scope of this thesis.

1.4 Document structure

Following this introduction, the second chapter sets the context for aircraft maintenance and the MRO sector, with a special focus on AOG and why the aviation supply chain behaves differently from those in other sectors. The third chapter focus into inventory management and develops the safety stock formula, which serves as the bridge between forecast error and tied-up capital.

The fourth chapter describes the three forecasting models we compare before applying them. The fifth explains the methodology and experimental design: how the seven time series were generated, what each represents in operational terms, how the training period was separated from the test period, and what metrics we used. The sixth chapter is the practical core of the study, where we present the results matrix and analyze the

performance of each model by demand pattern. The seventh chapter translates these results into financial terms. The eighth chapter summarizes the conclusions, the objectives achieved and those not achieved, and the lines of research that the study leaves open.

Chapter 2

Aeronautical maintenance and the MRO context

2.1 The MRO Sector: Definition, scope, and structure

The acronym MRO, which stands for Maintenance, Repair, and Overhaul, means all activities necessary to keep an aircraft airworthy. The term references everything from the daily inspection a technician performs on the tarmac to the procedure in which an engine is completely disassembled in a specialized workshop. What all these activities have in common is that none of them is optional. International regulations, enforced through EASA in Europe and the FAA in the United States, require operators to comply with mandatory airworthiness standards.

The sector economic impact is considerable. The global MRO market was projected to reach $120,000 million by 2025, with forecasts indicating it will exceed $150,000 million by 2030 (Oliver Wyman and IATA, 2025). But beyond the absolute figures, what is relevant to this study is the sector's countercyclical nature. When airlines cannot renew their fleets, they keep older aircraft in service for longer, which generates demand for maintenance. The average age of the global fleet has risen from 14.1 years in 2019 to 16.3 in 2025, a direct consequence of accumulated delays in the delivery of new aircraft (Oliver Wyman and IATA, 2025). Maintaining an older aircraft is not linearly more expensive: industry data show that a narrow-body aircraft over thirteen years old costs approximately 40% more to maintain than one that is three years old (Oliver Wyman and IATA, 2025).

The market is organized into four segments, and the distribution among them matters because it determines who can compete in each one. Engine maintenance accounts for just over 51% of total spending, up from 48% in 2019 (Oliver Wyman and IATA, 2025).

The increase is due to the fact that new-generation engines, such as the LEAP and the GTF, are undergoing their first overhauls earlier than expected, and each overhaul is significantly more expensive than for previous models. Component maintenance accounts for 18% and includes avionics, landing gear, APUs, hydraulic systems, and brakes. Fuselage maintenance, which covers scheduled structural inspections, accounts for 17%, and line maintenance, the most routine and frequent type, rounds out the breakdown at 13%.

This breakdown is not neutral in competitive terms. For engines and the most complex components, the cost of materials accounts for 60% of the total maintenance cost, giving original equipment manufacturers (OEMs) a structural advantage that goes beyond price. They control certified spare parts, repair instructions, and specific test equipment (Oliver Wyman and IATA, 2025). Figure 2.1 illustrates the extent to which the engine market is concentrated among a very small number of manufacturers, and how that concentration varies by aircraft type. In the narrow-body segment, CFM International stands for 56% of the value of new production, followed by Pratt & Whitney with 41%, while in the wide-body segment, General Electric takes the lead, holding the largest market share ahead of Pratt & Whitney. In practice, a handful of manufacturers dominate the engine supply in each category, and with it, access to spare parts and repairs in that segment. In 2024, aftermarket services accounted for 74% of GE Aerospace's total commercial revenue and 66% for Rolls-Royce (Oliver Wyman and IATA, 2025), demonstrating that the real business for engine manufacturers lies not only in selling engines but also in maintaining them. For independent MRO operators, competing in these segments requires access to licenses, repair instructions, and materials that OEMs control and grant on terms they consider appropriate.

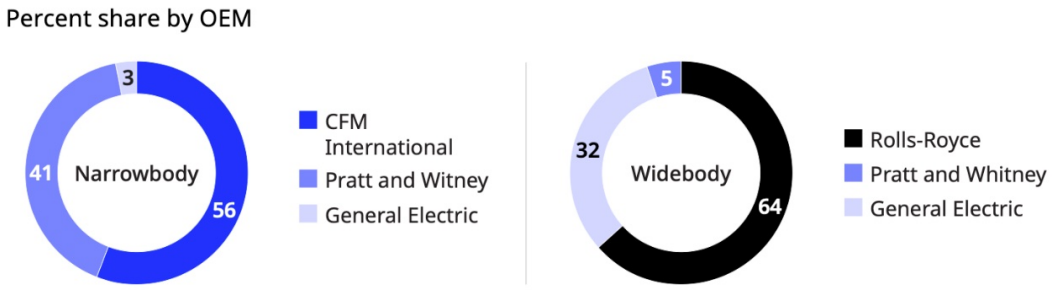


Figure 2.1: Share of engine maintenance spend by original equipment manufacturer (OEM), segmented by narrowbody and widebody aircraft.

Airframe maintenance works differently. Because it needs more man power and less dependent on the manufacturer's intellectual property, the segment is more fragmented and has lower barriers to entry. This is where independent operators and airlines who own maintenance facilities have the most gap to operate. It is no coincidence that OEMs

account for only 5% of airframe maintenance spending, compared to 63% in the engine segment (Oliver Wyman and IATA, 2025).

The aviation value chain remains a structure in which original equipment manufacturers occupy the top tier, with tier-one suppliers assembling large sections and systems, tier-two suppliers manufacturing subassemblies, and tier-three suppliers providing raw materials and basic parts. This architecture has been consolidated over decades. Li (2018) notes that as early as the 787 program, Boeing distributed about 70% of the manufacturing work among external tier-one suppliers, ceasing to be a vertically integrated manufacturer to act instead as a systems integrator. Figure 2.2 illustrates this by comparing component maintenance costs for the B777 (1993) with those for the B787 (2007). While first model distributed those costs among a long list of suppliers, the second model concentrates them among a few major players, such as Collins or SAFRAN. This consolidation improves efficiency under normal conditions but leaves the supply chain significantly more vulnerable to disruptions. A disruption at any point quickly propagates upstream, and recent years, marked by the pandemic, geopolitical conflicts, and a series of certification issues, have demonstrated that this vulnerability was not merely theoretical.

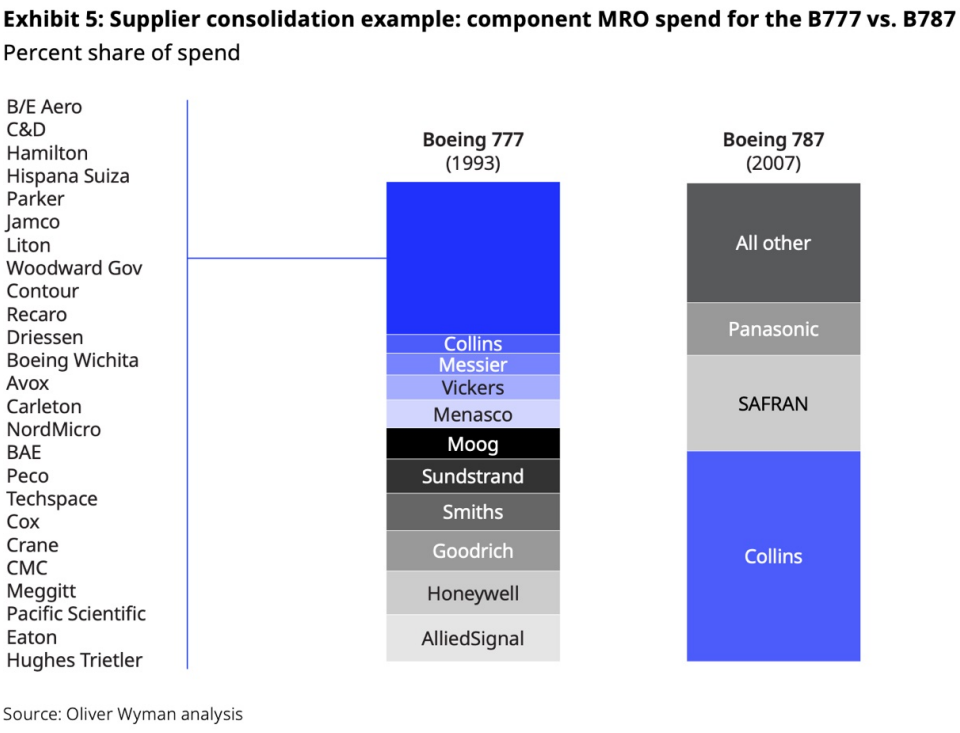


Figure 2.2: Example of supplier consolidation in component maintenance spend: comparison between the Boeing 777 and the Boeing 787.

2.2 The AOG concept and its operational impact

An aircraft grounded due to a missing part or a lack of maintenance is referred to as an Aircraft on Ground, or AOG. This is the situation that the entire MRO system exists to prevent. And it is also why effective spare parts inventory management has implications that go far beyond logistics.

The costs of an AOG begin to accumulate from the very first minute. The direct cost of an inoperative aircraft can exceed 50.000 euros per hour depending on the type of aircraft and the affected route, not to mention the associated costs, like rebooking passengers onto alternative flights, contractual penalties with client airlines, regulatory compensation under Regulation (EC) No. 261/2004, and the reputational damage that a prolonged delay causes with customers.

When all these layers are added together, the actual cost of an AOG can be several times the direct cost of the part that caused it. This asymmetry between the cost of keeping the part in stock and the cost of not having it is what explains the sector's historical behavior, in the face of uncertainty, maintenance departments accumulate inventory. If nothing is ever missing, there will never be an AOG. The problem, which we discussed in the previous chapter, is the financial cost of maintaining that permanent buffer across thousands of part numbers.

Industry data confirms that AOG pressure has intensified in recent years and shows no clear signs of getting softer. Almost eighty percent of respondents in Oliver Wyman's 2025 annual MRO survey identified parts availability as the primary factor behind the increase in turnaround times, ahead of supply-demand imbalances, labor shortages, or OEM control over key repairs (Oliver Wyman & IATA, 2025). We could think this is a residual effect of the pandemic but the problem goes more far as it is a structural problem with multiple causes.

TAT, Turnaround Times, have lengthened across all segments. The historical standard for engine overhauls ranged from 30 to 60 days. Currently, some shop visits exceed 300 days in extreme cases. Landing gear, which typically required 45 days under normal conditions, can now take between 90 and 120 days (Oliver Wyman & IATA, 2025). The 2025 survey reveals that more than 75% of participants reported longer turnaround times for engines and APUs over the past year, and more than 60% reported the same for avionics and landing gear. In all segments measured, the percentage reporting increases exceeds that reporting stability or improvement (Oliver Wyman & IATA, 2025).

This problem is the base that justify this study. Longer TATs keep more parts tied up in the repair cycle and reduce their availability in the warehouse. This increases the frequency and duration of AOGs. To compensate, airlines increase their emergency inventory. This additional inventory ties up capital that is no longer available for other

investments. Spare parts levels in the industry are approximately 8% above the historical average, with increases exceeding 15% at some operators (Oliver Wyman & IATA, 2025). The estimated holding cost for the industry as a whole in 2025 amounts to $1.4 billion (Oliver Wyman & IATA, 2025).

The impact is not just financial. Every grounded aircraft also entails an operational cost in terms of capacity: seats that aren't being used, wet lease agreements that must be activated to compensate, and the airline's image in the eyes of its passengers. And there is an environmental dimension that, in the current regulatory context, is not insignificant: operating older aircraft instead of the new ones that should have replaced them generates additional emissions estimated at approximately 17 million metric tons of CO2 above what would have been emitted had fleet renewal plans been followed (Oliver Wyman & IATA, 2025).

To fix AOG delays problem, we must focus at the beginning of the problem. And the root cause, to a large extent, is the inaccuracy in forecasting spare parts demand. A model that reduces forecasting error allows for more precise sizing of the safety stock, without the need for oversized safety margins that compensate for the lack of confidence in the forecast.

2.3 The MRO supply chain: Why generic models don't work

The inventory of aircraft spare parts cannot be managed like that of a conventional distribution company. There are at least three characteristics that cause standard models to fail when applied without adaptation, and understanding them is a key point for justifying the methodological approach we adopt in this study.

Regulations limit substitution:

In aviation, a part that is functionally equivalent to the original is not a valid substitute unless it is certified by the competent authorities. Every component installed on an aircraft must have complete documentary traceability, a valid airworthiness certificate, and be specifically approved for the aircraft type in question. This eliminates the possibility of rapid substitution that exists in other industrial sectors. If a component is not available when needed, the alternative is not to seek a generic equivalent: it is to wait, or to place an emergency order at spot market prices through the AOG desk, with the additional costs that this causes.

Li (2018) notes that the relationship between major integrators and their suppliers has evolved from a build-to-print model, in which the supplier manufactured according to the blueprints provided. While the efficiency of the system improves, proprietary knowl-

edge becomes in less hands. Certifying an alternative supplier to manufacture or repair a specific component can take years and requires significant investment in engineering and regulatory approval processes. The result is that competition in many component segments is structurally limited, and MRO operators that rely on those components have little room to maneuver when the usual supplier fails.

This rigidity has a direct implication for inventory, if you cannot substitute or quickly obtain the component from an alternative source, the only real protection is to keep it in stock. Which leads, once again, to overstocking as a rational response to a structural constraint in the system.

Demand is volatile in ways that simple statistical models cannot capture:

Unlike the consumer industry, where demand for spare parts exhibits a certain degree of seasonality and predictability, in aviation, the need for a part can go from zero to urgent within hours. Unscheduled maintenance represents a failure detected during a night inspection or a defect identified by the in-flight monitoring system, generates demand that no model can anticipate with perfect accuracy.

There is also an important difference between scheduled and unscheduled maintenance that historical data does not always clearly reflect. Scheduled tasks have known intervals, their demand for spare parts is predictable. Unscheduled maintenance is stochastic by definition. In many consumption records, both sources of demand appear mixed together in the same monthly record, without a label to distinguish them. The model receives the aggregated signal and cannot separate which part corresponds to regular patterns and which part is random noise.

This is known as bullwhip effect, which is a terminology well known in the supply chain. Small variations in actual demand at the operator level generate amplified fluctuations in orders placed with suppliers, which in turn generate even greater fluctuations among their own suppliers. In the aerospace context, where lead times are long and order adjustments are delayed, this effect can generate over demand signals and lead to cycles of excess inventory followed by stockouts.

The cost structure is asymmetric:

The third feature is the asymmetry between the cost of running out of parts and the cost of holding excess inventory. Running out of a high-turnover component can cost tens of thousands of euros per hour of AOG. Holding excess inventory of that same component for several months has a real cost, but one that is limited and predictable: the financial cost of tied-up capital plus storage costs. This asymmetry makes overstocking a rational short-term decision for any materials manager mindful of the penalty of an AOG. However, at the fleet and full-catalog level, this behavior generates aggregate inventory that can represent millions of euros in capital that does not generate a return.

Airlines have responded to the supply uncertainty of recent years by increasing their inventories, with increases that in some cases exceed 15% above normal levels (Oliver Wyman & IATA, 2025). As a short-term defensive tactic, this makes sense. As a permanent strategy, however, the financial cost ultimately becomes unsustainable, especially in an industry where operating margins are structurally tight.

Mocenco (2015) emphasizes that the risks of disruption in the aviation supply chain are caused from both natural causes and capacity constraints, geopolitical tensions, and regulatory changes. The titanium shortage resulting from the Russian invasion of Ukraine, which caused a price increase of nearly 90% and impacted manufacturing lead times for structural components throughout the supply chain (Oliver Wyman & IATA, 2025), is a recent example of the extent to which external shocks can invalidate any inventory plan built on assumptions of stability. No forecasting system anticipates a war. But a more accurate model under normal conditions reduces prior exposure and creates more room to absorb shocks when they arrive.

2.4 The regulatory framework: EASA and FAA

EASA and the FAA are not just administrative regulations for MRO business. They are active factors that influence both the demand for spare parts and the supply available to meet that demand, and any inventory model that ignores them is built on incomplete assumptions.

Both authorities establish the airworthiness requirements that operators must meet, certify repair stations and technicians authorized to perform maintenance, and issue Airworthiness Directives, which are mandatory instructions addressed to aircraft owners and operators. An Airworthiness Directives may require the inspection, modification, or replacement of a specific component within a specified timeframe. When such an Airworthiness Directives affects a low-turnover component that is not part of the regular inventory, the procurement system is suddenly faced with a one-time demand that was not anticipated. The usual consequence is an emergency purchase at spot prices, with costs that can exceed the component's normal list price several times over.

There is, however, a predictable aspect of Airworthiness Directives that is worth knowing. Unlike unscheduled maintenance, for which demand is purely stochastic, Airworthiness Directives are published well in advance, allowing for planning. They are part of the public information released by EASA and the FAA, and operators who actively monitor regulatory publications can anticipate with some accuracy which components will need to be inspected or replaced in the coming months. A well-designed forecasting system should incorporate this signal.

Certification processes also shape the available supply. For an alternative repair shop to

perform a specific repair on a given component, it must obtain approval from the competent authority, a process that can take years of technical and administrative procedures. In segments where OEMs hold an intellectual property advantage, the time it takes for alternative competitors to enter the market is the most effective barrier. It is not just a matter of certification costs, it also involves access to repair manuals, continued airworthiness instructions, and specific test equipment that OEMs do not always share with independent repair shops.

The situation intensifies by a shortage of qualified technical personnel both in the maintenance facilities themselves and within the regulatory authorities. In 2023, approximately 40% of FAA certification engineers had less than two years of experience in their roles (Oliver Wyman & IATA, 2025). This fact takes on particular significance when one considers that new composite materials, advanced coatings, and next-generation electronic systems require specific technical expertise to be properly evaluated. The learning curve of regulatory staff is racing against the pace of industry innovation, and this tension results in longer and less predictable certification timelines.

Chapter 3

Stock management and safety stock

3.1 Basic concepts: Demand, Lead Time, and Reorder point

Three concepts will be recurrent for this work. These are defined as the variables that any procurement model must estimate accurately in order to function.

Demand is the starting point. Predicting with some accuracy how many units of each part number will be consumed in a given period allows us to decide how much to order, when to order, and how much to keep in stock. In aviation, this is more complicated than it seems. Spare parts consumption does not follow seasonal patterns as in the consumer goods industry. It has unexpected spikes, periods of inactivity, and an unscheduled maintenance component that, by definition, cannot be anticipated. Forecasting errors in this sector tend to be greater than in others, and those errors have a direct cost.

Lead time is the period from when an order is placed until the goods are available in the warehouse. In an ideal environment, it would be constant and predictable. In practice, it is subject to variability due supplier delays, certification issues, and logistical disruptions. This variability requires constantly recalculating inventory levels, because the system cannot assume that what takes time today will take the same amount of time next month.

The reorder point is the stock level that, when reached, generates a new order. Its calculation depends on the expected demand during the lead time and the safety stock you wish to maintain. If either of these two variables is miscalculated, the system will replenish stock too late or too early.

3.2 The Classic formula for safety stock

Safety stock (SS) is the inventory buffer that an organization maintains above expected demand to absorb variations that no model anticipates. Its classic calculation, under normal distribution assumptions, takes the form:

$$SS = Z \cdot \ \sigma_{error} \cdot \ \sqrt{LT}$$

Z is the factor associated with the target service level: 1.65 for 95%, 1.28 for 90%. It is the only parameter that the organization explicitly chooses; the others are estimated from the data.

σ_{error} is the standard deviation of the forecast error, the average deviation between what the model predicts and what is actually consumed. It is also the only variable that can be influenced. If the model improves, the deviation decreases. And when the deviation decreases, the safety stock required to maintain the same service level also decreases.

LT is the supplier's lead time. It appears under the square root because uncertainty accumulates over time, but at a decreasing rate.

The relationship between forecast accuracy and safety stock is not intuitive, but it is direct and quantifiable. A more accurate model is not just a mathematical improvement: it is capital that is no longer tied up in the warehouse. That is what this study aims to demonstrate.

3.3 Demand variability in aviation

The safety stock formula works reasonably well when demand is stable and forecast errors are distributed more or less symmetrically around zero (Gonçalves et al., 2020). These conditions do not always hold in aviation, and it is important to understand why before choosing a model.

The demand for spare parts covers a wide range. At one extreme are items with continuous consumption, exhibiting a trend and often a seasonal cycle, which are the ones we study in this paper. At the other end are those with intermittent demand, which may go months without activity and then experience a spike concentrated over a few days, triggered by a breakdown, an airworthiness directive, or the coincidence of several maintenance events. For this last one, the historical standard deviation of consumption is a poor estimator of what will happen in the next cycle, and the theory describes them as lumpy demand. Croston (1972) proposed a specific method for this case, which separates the frequency with which demand occurs from the size of each transaction, because methods based on averages underestimate the actual variance by diluting the peaks among the idle periods.

There is also a distinction that the classical statistical model tends to overlook. The one separating scheduled maintenance from unscheduled maintenance. Scheduled tasks, such as line inspections or periodic checks, generate demand that is predictable, because the intervals are known in advance. Unscheduled maintenance responds to unexpected failures, and its demand is stochastic by definition. In a part's consumption history, both sources appear mixed together, without a label to distinguish them, so the model receives the aggregated signal and cannot separate which part follows a pattern and which part is noise.

Fleet age adds another layer. A component that consumed two units per month in the early years of operation may reach five or six after a decade, not because anything has changed in the operation, but because the cumulative degradation of the systems increases the frequency of replacement. If the forecasting model fails to detect this upward trend, it will systematically underestimate future demand. That is why a method that explicitly accounts for the trend is desirable, rather than one that merely averages the most recent periods.

And finally we must count AOG cases. Faced with the risk of grounding an aircraft, the materials manager tends to set conservative safety margins (Alomar and Nikita, 2025). This is a reasonable response, but it has a side effect: inventory grows as a precaution across all part numbers, even those where the actual risk is low, and capital becomes inefficiently tied up. A more accurate forecast reduces this need for a buffer, as it allows inventory levels to be set with the confidence that the conservative margin was meant to provide.

The seven series in this study contains continuous demand with a trend and a weekly cycle. The challenge here is not intermittency, but rather capturing both of these features, the trend and seasonality, which is precisely what distinguishes some models from others in the results section (Tadayonrad and Ndiaye, 2023).

3.4 Classification of aircraft spare parts

A fleet's spare parts catalog includes parts that vary greatly in value and in the logic governing their replacement, so the first step in any management system is to classify them. The most useful criteria for inventory management is the disposition of the part once it is removed from the aircraft, and what happens to it when it is taken out of service.

At one extreme are rotables. These are high-value components that are repaired and reinstalled multiple times over their lifespan. They are identified by serial number and tracked individually. Hydraulic pumps, actuators, avionics units, and landing gear fall into this category. For these, replenishment does not mean buying a new part, but rather managing a closed-loop system, by removing the unit and sent it to the shop, where

is repaired, and returned to the warehouse. What determines inventory levels is not consumption demand, but rather repair time and its variability. Alongside rotables are repairables, which follow a similar logic but have a lower unit value.

Life-limited parts are a special case. They have a legal limit on the number of cycles or flight hours and are discarded once that limit is reached, regardless of their apparent condition. Turbine discs and certain structural components are typical examples. Their procurement is planned based on the fleet's expected usage, not on a reorder point.

At the other end of the spectrum are expendables and consumables. Expendables are low-cost parts that are replaced rather than repaired when removed, such as fasteners, rivets, seals, or filters, and are rarely tracked by serial number. Consumables are materials that are used up during maintenance and cannot be reused, such as oils, greases, hydraulic fluid, de-icing fluid, or sealants. The line between the two groups is blurred, and in many warehouses they are managed the same way, because they share what really matters here, they are replenished by purchasing, not by repairing.

This difference determines the type of inventory model applied to each group. Repairable items follow a closed-loop system, where the safety stock covers the variability in repair time and is calculated using specific models for sizing the repairable item pool. Consumables and expendables follow an open-loop system, governed by a reorder point and a safety stock that depend on forecast demand and lead time. This is precisely the system described by the formula in Section 3.2, and it is the only one in which forecast accuracy directly translates into inventory levels.

For this reason, the experiment focuses on consumables. This is the category where better forecasting reduces safety stock. Some wear and tear items, such as tires or carbon brakes, go through a "remanufacturing" shop, so they are not strictly speaking consumables. Their high turnover means, that in practice they are replenished in line with the same logic of demand and reorder point, so we treat them within the same framework.

There are other ways to classify the catalog that coexist with this one. ABC segmentation ranks items by their weight in the total inventory value and focuses management efforts on the few that account for most of that value. Classification by criticality addresses operational impact, separating parts whose absence causes an AOG from those that allow the flight to depart with restrictions. Both are useful for prioritization, but they address a different question than the one addressed in this study. Since our objective is to measure how each model reacts to each demand pattern, and not to prioritize a catalog, we do not apply segmentation by value or by criticality. The seven series are treated equally.

3.5 The limitations of the classical statistical approach

The safety stock formula has been used in the industry for over half a century. As a conceptual framework, it remains valid, but some of its assumptions fail in contexts such as the aerospace industry, and it is important to bear these in mind before applying it (Gonçalves et al., 2020).

The formula assumes that forecast errors follow a Gaussian distribution, which is symmetric and well-behaved, from which the Z-factor associated with the service level is derived. In intermittent demand scenarios, this distribution is not normal, and applying Gaussian percentiles results in a poorly calibrated safety stock: either too low, with a risk of stockouts, or too high, with excess tied-up capital. In the continuous consumption series we study, the assumption is more reasonable, although it still depends on the model not having a systematic bias (Stoilov and Stoilova, 2023).

The second point concerns how σ_{error} is obtained. The formula treats this deviation as a given, but its value depends entirely on the forecasting model used. A method that does not capture the series structure well will result in a large error, and that error will inflate the safety stock even if the target service level remains unchanged. A simple average of the most recent periods, for example, lags behind the trend and corrects it too late, so that when consumption rises constant, it underestimates demand and leaves a systematic error. That is why the industry standard method is not a simple average, but rather exponential smoothing, which incorporates a trend component and a seasonality component (Rahardja, 2021; Grech, 2024).

The question addressed in this study starts precisely there. It is not a matter of verifying whether a model is perfect, but rather of measuring whether models more sophisticated than the industry standard achieve a smaller error, without sacrificing service level. To do this, we compare Excel's exponential smoothing, which already captures trend and seasonality, with a SARIMA model that explicitly describes the series dependence and with a neural network that learns nonlinear relationships without prior assumptions. Whether the greater complexity of the latter two translates into a real improvement, and for which demand patterns it does so, is what the results chapter addresses (Makridakis et al., 2020).

Chapter 4

Forecast demand models

4.1 Forecasting in inventory management

The safety stock formula from the previous chapter requires two unknown variables, how much demand there will be during the lead time and by how much the forecast will be off when estimating it. Forecasting demand is, therefore, the first step in any procurement decision and the basis upon which purchases are planned and inventory levels are set in the supply chain (Tadayonrad and Ndiaye, 2023). We do not seek to guess the exact daily consumption, which would be impossible, but rather to estimate whose error is as small and stable as possible, because that error is what feeds σ_{error} and, with it, the inventory level. Studies on the determination of safety stock agree on this point, as forecast accuracy is one of the factors that most strongly influence the safety stock an organization needs to maintain (Gonçalves et al., 2020; Stoilov and Stoilova, 2023). In the aviation sector, this inventory exists to prevent the lack of a part from grounding an aircraft, an event the operator seeks to avoid at all costs (Alomar and Nikita, 2025).

The problem we address is a time-series problem. We have the daily consumption of a reference item over a period and want to estimate consumption for the following days, based on the chronological historical data on which all predictions of this type rely. A model takes the history and returns a forecast for the horizon of interest, and the quality of that forecast is measured by comparing it with what actually happened.

The available methods cover a wide range of complexity. At one extreme are smoothing models, which summarize the recent past using very few parameters. At an intermediate level are the statistical models of the ARIMA family, which explicitly describe how an observation depends on previous ones. Beyond these lie machine learning methods, such as neural networks, capable of capturing nonlinear relationships at the cost of requiring more data and more tuning. Complexity does not guarantee accuracy. Highly complex models tend toward overfitting and can become trapped in local optima, while simple

methods typically behave more robustly in the face of real-world uncertainty (Makridakis et al., 2020). That is why the design of this study starts with a simple model as a baseline and tests whether more elaborate models outperform it.

We compare three models, one for each level of complexity: Excel's exponential smoothing as the industry benchmark, a SARIMA model that provides a richer statistical description, and a neural network that tests whether a flexible learner outperforms the previous two. The choice of models at different levels responds to the need to evaluate the actual balance between mathematical sophistication and logistical benefit (Makridakis et al., 2020). All three are trained on the same initial segment of each series and evaluated on the same reserved final segment, ensuring a clean comparison. The following sections describe each model before applying them.

4.2 Exponential smoothing: Excel's FORECAST.ETS

Exponential smoothing is based on a simple idea. The best estimate of the next value is a weighted average of past values, in which recent observations carry more weight than older ones, and that weight decays exponentially as we go back in time (Rahardja, 2021). In its most basic version, which tracks only the level of the series, the model is summarized in an equation:

$$L_t = \alpha\, y_t + (1 - \alpha)\, L_{t-1}$$

Where L_t is the estimated level at time t, y_t is the actual observation, and α is a parameter between 0 and 1 that controls how quickly the model forgets the past. A high α reacts quickly to recent changes; a low α produces a smoother forecast. The forecast for the next period is simply the last estimated level.

This basic form does not work when the series has a trend or seasonality, which is precisely the case for our seven series. Holt's method adds a second equation that estimates the slope of the series, so that the forecast is no longer flat and follows the rise or fall. Holt-Winters takes the next step and incorporates a third component, the seasonal component, which captures the pattern that repeats every certain number of periods (Grech, 2024). With all three components active, the model decomposes the series into level, trend, and seasonality, and updates each one in every period using its own smoothing constant:

$$L_t = \alpha\,(y_t - s_{t-m}) + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \beta\,(L_t - L_{t-1}) + (1-\beta)\,b_{t-1}$$
$$s_t = \gamma\,(y_t - L_t) + (1-\gamma)\,s_{t-m}$$

Here, b_t is the trend, s_t is the seasonal factor, m is the number of periods in the cycle, and β and γ are the constants that smooth the trend and seasonality. The h-period forecast combines the three components using $\hat{y}_{t+h} = L_t + h \cdot b_t + s_{t+h-m}$. This family of models is also known by the acronym ETS, standing for Error, Trend, and Seasonality, which is the name used in the Excel function (Rahardja, 2021).

The FORECAST.ETS function implements this scheme. Based on the historical data, it automatically estimates the smoothing constants and detects the length of the seasonal cycle on its own (Grech, 2024). Specifically, the study uses the additive variant in all three components, known as ETS(A,A,A) or the AAA version, which is appropriate when the amplitude of the seasonal pattern does not increase with the level of the series. The usefulness of this function for time series forecasting is documented in the applied literature (Rahardja, 2021; Grech, 2024). In our series, the cycle is weekly (seven days), and seasonality has a significant impact on the demand we are trying to forecast (Tadayonrad and Ndiaye, 2023). All three models in the study operate with this same seasonality, Excel recognizes it automatically, while SARIMA and the neural network receive it explicitly specified. That means, the comparison neither rewards nor penalizes any model for poorly fitted seasonality.

Exponential smoothing is the method that many MRO planning teams use in practice, because it is integrated into tools they already have and because its results are easy to interpret (Grech, 2024). Its reliability is also supported by evidence. In the M4 competition, organized by Makridakis and his collaborators using 100,000 time series, exponential smoothing methods and their combinations ranked among the most accurate (Makridakis et al., 2020). In fact, the winning model combined exponential smoothing with a neural network, and several pure machine learning methods failed to outperform the statistical benchmarks. Taking this method as a starting point sets the bar where it truly belongs and requires SARIMA and the neural network to justify their greater complexity with a real improvement.

4.3 SARIMA

SARIMA is the seasonal extension of ARIMA models, so it is best to present it in parts. ARIMA combines three ideas (Lin et al., 2023). The autoregressive (AR) part assumes that each observation depends linearly on a number p of previous observations. The moving average (MA) part assumes that it depends on the q most recent prediction errors. Integration (I), between the two, consists of differentiating the series d times, that is, working with the changes from one period to the next rather than with the levels, to eliminate the trend and leave a series that is mean-stable. The complete model is called ARIMA(p, d, q).

ARIMA does not capture cycles that repeat at regular intervals. SARIMA resolves this by adding a seasonal block that replicates the same autoregressive, differencing, and moving average logic, but with respect to the cycle period. The notation becomes SARIMA(p, d, q)(P, D, Q)s, where the uppercase letters are the seasonal orders and s is the cycle length. A seasonal autoregressive term, for example, relates today's observation to that at the same point in the previous cycle. This family of models is commonly used in forecasting logistics volumes, where recent research combines it with neural networks such as LSTM to improve accuracy (Lin et al., 2023), which connects to the last model we present.

The configuration used in this work is SARIMA$(1,1,1)(1,0,1)_7$. The autoregressive part incorporates an autoregressive term, a differencing term, and a moving average term. The seasonal part incorporates an autoregressive term and a moving average term, without seasonal differencing, with a period of seven that coincides with the weekly cycle of the series. The parameters are estimated using maximum likelihood.

The advantage of SARIMA over exponential smoothing is that it explicitly models the series' dependency structure, with interpretable coefficients and a parameterized weekly cycle, and provides statistically supported confidence intervals. Regular differentiation allows it to track the trend, and the seasonal block with a period of seven captures the weekly cycle, a component that the literature identifies as a key determinant in demand forecasting (Tadayonrad and Ndiaye, 2023). In this way, the model addresses both of the dominant features of these series simultaneously. We compare its accuracy with that of the other two models in the results chapter.

4.4 Artificial Neural Networks (ANN)

An artificial neural network is a function approximator. Instead of starting with an equation that describes the series, it learns the relationship between inputs and an output based on examples, and adjusts a large number of internal parameters until the error is minimized. This flexibility allows it to capture nonlinear relationships that the previous

models, which are linear by construction, cannot represent, and for this reason its use has become widespread in logistics forecasting, often in combination with statistical methods (Lin et al., 2023).

The type of network we use is a multilayer perceptron. It is organized into layers of neurons: an input layer that receives the data, one or more hidden layers that transform it, and an output layer that returns the prediction. Each neuron calculates a weighted sum of its inputs, adds a bias, and passes the result through an activation function. In our configuration, that function is the ReLU, which passes through positive values and sets negative ones to zero, and the network has two hidden layers of ten and five neurons. Training adjusts the weights using backpropagation and the Adam optimizer, an efficient variant of gradient descent. Before entering the network, the variables are standardized so that all operate on a comparable scale and the algorithm converges better.

What determines the behavior of a network of this type is not only its architecture, but also the information it receives as input. Here, the network predicts daily consumption based on three past values: the previous day's, the day before that, and seven days ago. The first two lags capture short-term inertia. The third, the seven-day lag, is what gives it access to the weekly cycle by comparing it to the same day of the previous week. It thus functions as a nonlinear autoregressive model, and unlike smoothing and SARIMA, it imposes no prior form on the relationship between those inputs and the output.

To forecast beyond the first day, the network operates recursively. It predicts the next day, incorporates that prediction into the history, and uses it as input for the day after that. The procedure is practical, but it carries a well-known drawback: if the initial predictions deviate, the error propagates and amplifies across the horizon, because the network ends up feeding on its own mistakes.

We include the neural network for the same reason that the M4 competition tested it. It is reasonable to ask whether the complexity of modern machine learning techniques truly translates into superior performance and tangible inventory savings compared to well-calibrated statistical methods (Makridakis et al., 2020). The results chapter answers that question using the seven time series and compares their accuracy with that of the benchmark models.

Chapter 5

Methodology and experimental design

5.1 Demand simulation: Why synthetic and controlled

Evaluating a forecasting model using real fleet data presents a fundamental problem. The historical data on the consumption of aircraft spare parts is distorted by factors unrelated to the demand pattern itself, such as supply disruptions, variable supplier lead times, and unforeseen events such as AOGs (Alomar and Nikita, 2025). The aerospace supply chain also adds its own complexity, with high-value components, strict traceability, and a highly concentrated supplier network (Mocenco, 2014). Added to all this is the fact that supply reliability and seasonality directly influence the demand that is ultimately recorded (Tadayonrad and Ndiaye, 2023). Real-world data, therefore, mixes model behavior with operational noise.

For this reason, the design relies on a controlled simulation study. By generating the demand series, we know their exact structure, and can measure how far each model deviates from a consumption pattern we fully understand. The simulation isolates each algorithm's response to purely mathematical variations, such as a trend or a seasonal cycle, and prevents real-world biases, for example, the absence of records during a stockout, from contaminating the comparison. With this approach, differences in accuracy between models are attributed to their architecture and not to an operational anomaly.

5.2 The seven demand patterns and their operational correspondence

The whole experiment rests on seven consumable items and their demand. Each of the three forecasting models is trained and tested on the same seven daily consumption series, provided by the tutor and recorded between 1 January and 14 April 2026. Every series stands for a distinct consumable in an MRO operator's warehouse, and the comparison between models is carried out one consumable at a time, because a stable item and a sharply trending one do not pose the same problem to a forecast.

The choice of consumables is deliberate. As we saw in Section 3.4, this group is replenished by purchasing rather than by repair, with continuous outflows that follow the logic of a reorder point and a safety stock, unlike rotables, which move through a closed-loop repair cycle (Mocenco, 2014). Consumables are therefore the category where forecast accuracy translates directly into the inventory level, which is exactly what this study sets out to measure.

Table 5.1 characterizes the seven series. For each one it reports its average daily level, which sets the scale of consumption; the direction and strength of its trend, expressed as the percentage change between the start and the end of the period; its coefficient of variation (CV), which measures relative volatility as the standard deviation divided by the mean; the day of the week on which demand peaks, which captures the weekly seasonality; and a short note on the operational driver behind the pattern. Taken together, these columns describe each demand profile along the four axes that matter for a forecast: level, trend, volatility, and seasonal timing.

The seven series share a common structure and differ in the details. All of them combine a linear trend with a weekly cycle, none contains a single zero-consumption day, and their volatility stays within a moderate range, between a 12.6% and a 27.9% coefficient of variation. What separates one series from another is the particular combination of those four axes, and weekly seasonality is, in fact, one of the features that most strongly influences the forecast for each item (Tadayonrad and Ndiaye, 2023).

The operational logic behind the patterns fits the winter-to-spring quarter in which the data were recorded. Items tied to the volume of leisure flights, such as tyres, brakes, and line consumables, grow as traffic builds toward spring, and their demand peaks at the weekend, when this type of operation concentrates its flights. Items driven by scheduled maintenance, such as hydraulic fluid or filters, spread their demand across the start or the middle of the week, depending on when those tasks are planned. De-icing fluid follows the seasonal calendar and declines as winter ends, while engine oil keeps the high, steady consumption of a mature fleet. By tying each statistical pattern to a real consumable, we can judge which model best fits the nature of the demand for each part.

Table 5.1: The seven demand series: empirical characterization and operational logic

Series	Consumable	Avg. level (u/day)	Trend	CV (%)	Weekly peak / demand logic
S1	Main-wheel tyres	32.5	+77%	24.9	Weekend / Monday. Leisure flights; landing cycles
S2	Hydraulic fluid (Skydrol)	37.5	+59%	22.3	Early week (Mon-Tue). Scheduled line tasks
S3	De-icing / anti-icing fluid	35.9	−20%	12.6	Midweek (Wed-Thu). Winter-driven, fades into spring
S4	Brake pads / discs	25.9	+127%	27.9	Weekend. Ramping route; landing cycles
S5	Engine oil	42.5	−9%	13.0	Monday / Sunday. Mature fleet; flight hours
S6	Filters (oil, air, fuel)	38.1	+54%	17.7	Midweek (Thu-Fri). Scheduled checks
S7	Line turnaround consumables	32.9	+82%	26.7	Strong weekend. Charter turnarounds

Note: CV = coefficient of variation (standard deviation / mean). Trend expressed as percentage change from day 1 to day 90. CV ranges from 12.6% (S3, de-icing fluid) to 27.9% (S4, brake pads / discs).

One point of the design is worth stating plainly. None of the seven series shows intermittent demand, with long stretches of zero broken by isolated spikes, which is common elsewhere in the aviation catalogue. The experiment therefore stays within continuous demand that carries a trend and a seasonal cycle, and we leave the intermittent case as a line for future work.

5.3 Error metrics (MAE and RMSE) and conversion to a score for the radar

The evaluation is based on measuring the difference between the forecasted demand and the simulated actual demand, using two error indicators widely used in demand forecasting (Tadayonrad and Ndiaye, 2023). The first is the MAE (Mean Absolute Error), which averages the absolute value of daily errors. It is an easy-to-interpret measure because it is expressed in the same units as demand and is relatively insensitive to outliers.

The second is the RMSE (Root Mean Square Error), which penalizes large errors more heavily because it squares each deviation before averaging it and taking the square root. This property links it directly to the logistics of the problem. When the forecast does not exhibit a systematic bias, the RMSE coincides with the standard deviation of the forecast

error, σ_{error}, which is precisely the variable that appears in the safety stock formula SS = Z $\cdot$ σ_{error} $\cdot$ $\sqrt{LT}$ (Gonçalves et al., 2020). Reducing the RMSE decrease the required safety stock, which is why it is the metric that links forecast error to inventory savings.

To compare the performance of the three models against the seven patterns at a glance, the raw error values are normalized into a score. The transformation we use converts the MAE into a score ranging from 0 to 10 using the expression score = MAX(0; (8 − MAE) / 8 Œ 10), such that a smaller error yields a higher score. This score is what feeds the radar chart in the results chapter, which visually shows which algorithm performs best for each type of consumable and facilitates a multi-criteria analysis.

5.4 Implementation: Python (SARIMA and ANN) and Excel (FORECAST.ETS)

The construction of the models combines two environments depending on their complexity. The SARIMA model and the neural network are programmed in Python, using standard scientific libraries for data processing and computation, such as pandas and numpy, and statsmodels and scikit-learn for the models themselves. SARIMA, models the linear dependence of the series on its own lags and past errors, with the seven-period seasonal block described in the previous chapter.

The neural network in this study is a multilayer perceptron, not a recurrent network. It consists of two hidden layers of ten and five neurons, with ReLU activation and training using the Adam optimizer, and the data is standardized before entering the network to facilitate convergence. Unlike recurrent architectures such as LSTMs, which incorporate state memory to handle long sequences, the perceptron learns the nonlinear relationship between the inputs we provide, consumption lags, and the next day's output. This choice keeps the model simple and reproducible, in line with the size of the time series.

The third model is implemented in Excel and serves as a statistical baseline. The native FORECAST.ETS function applies Holt-Winters triple exponential smoothing, decomposing the series into an additive model of error, trend, and seasonality (AAA version) and assigning weights that decay exponentially toward the past (Rahardja, 2021; Grech, 2024). Its low computational cost and immediate availability are precisely the reasons why it is the industry's benchmark method.

Chapter 6

Results and analysis of system error

6.1 Fit for each series: The seven graphs discussed

Figure 6.1 shows the seven graphs, one for each series, with actual demand and the three forecasts overlaid on the 14 day test period. A visual inspection confirms what the data later reveals. In all seven time series, the Excel and SARIMA curves stay close to the actual line and reproduce both the trend and the weekly peaks, while the neural network curve clearly diverges in all of them.

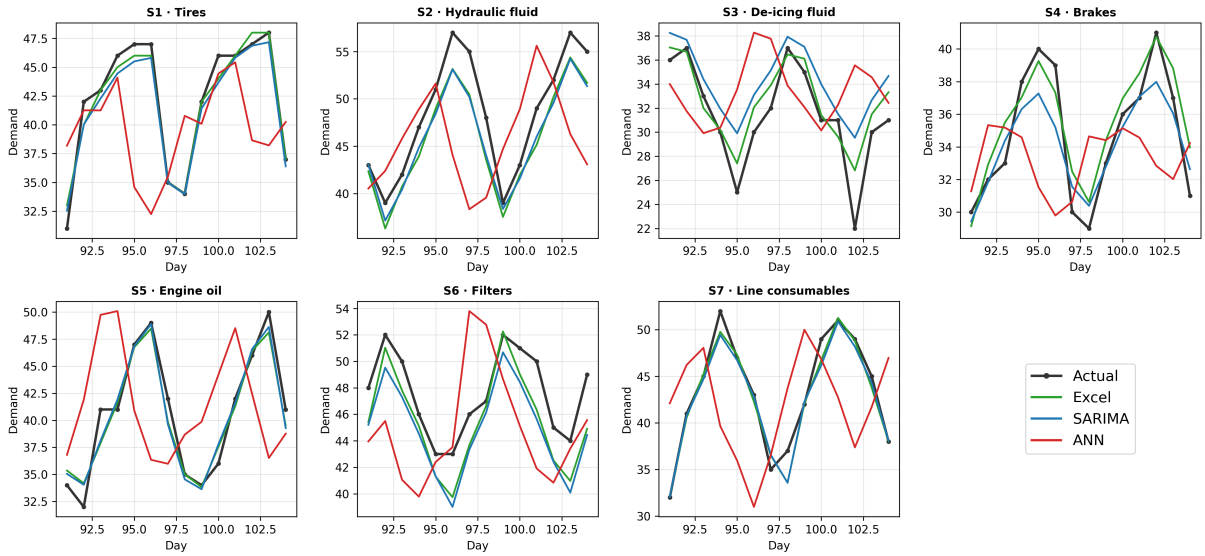


Figure 6.1: Fit of each series over the test period (14 days): actual demand versus Excel, SARIMA, and ANN forecasts.

Among the top two models, the differences are small and depend on the type of series. Excel better fits the smooth decline in de-icing fluid (series 3) and filter consumption (series 6), where the smoothing follows the slope with fewer spikes. SARIMA, on the other hand, narrowly prevails for hydraulic fluid (series 2) and motor oil (series 5). For

weekend-related consumables; tires (series 1) and production line consumables (series 7), both models follow the cycle with almost no lag, with errors of less than one unit.

The neural network tells a different story. Its deviation increases the higher the level of the series or the more pronounced its volatility, as occurs with motor oil (series 5) and line consumables (series 7). It is striking that in series 5, one of the most stable in the set, the network's error even exceeds the series' own variability, indicating that its recursive forecast diverges from actual behavior as the horizon advances rather than tracking it.

6.2 The results matrix (MAE and RMSE by model and pattern)

The following two tables show the error of each model over the test period, measured using the two metrics presented in Section 5.3. The MAE expresses the average error in units of demand, and the RMSE penalizes large deviations more heavily (Tadayonrad and Ndiaye, 2023).

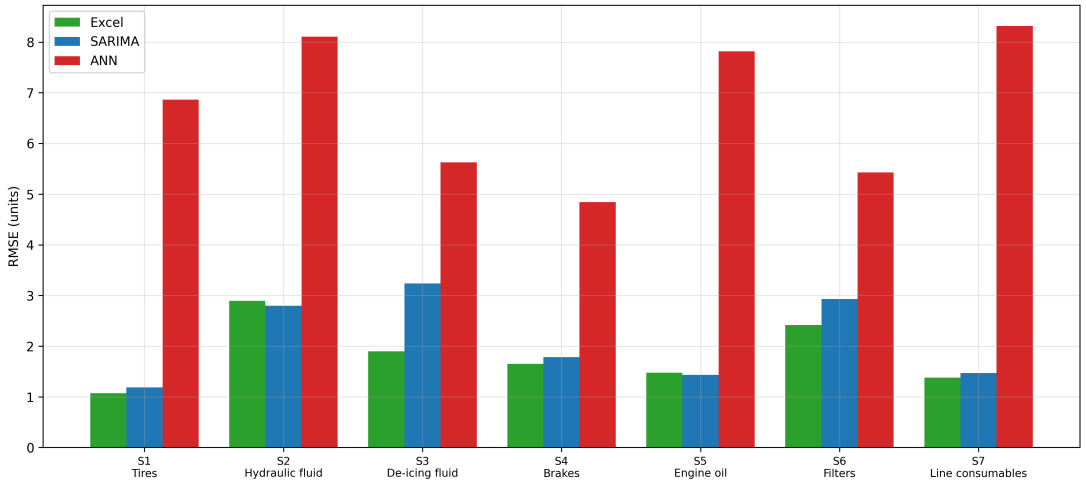


Figure 6.2: RMSE by model and series over the test period.

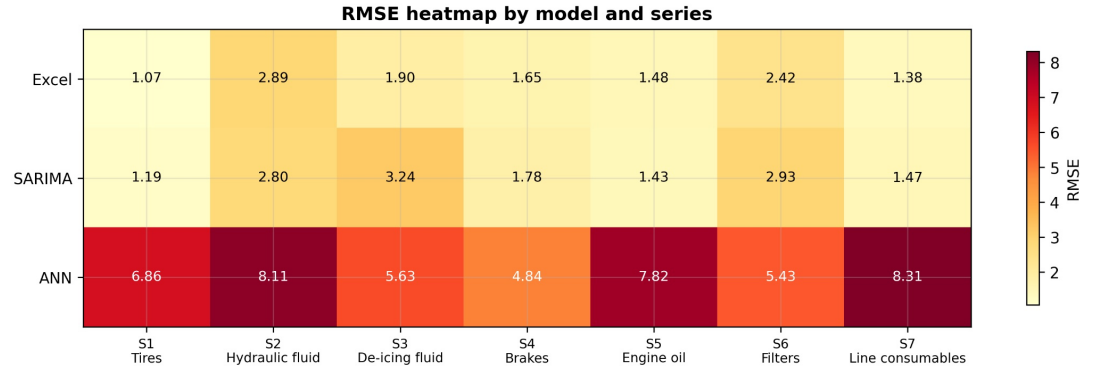


Figure 6.3: RMSE heatmap by model and series.

The figures paint a clear picture. Looking at the RMSE, the metric linked to safety stock (Gonçalves et al., 2020), Excel achieves the lowest error in five of the seven time series (1, 3, 4, 6, and 7), while SARIMA does so in the remaining two (2 and 5). With the MAE, the tally changes slightly, as SARIMA regains series 4 and ends up with three wins compared to Excel's four. The neural network is not the most accurate in any series using either metric, and its average error is easily three times that of the other two models.

6.3 Results by model

Excel's exponential smoothing is confirmed as the most accurate model in the set, with a mean MAE of 1.51 and a mean RMSE of 1.83. Its advantage does not stem from a single correct prediction, but from consistent performance: it wins or comes very close to first place in all series, without a single poor result. For the method the industry uses by default, this result reinforces its role as the benchmark (Grech, 2024).

SARIMA is not far behind, with a mean MAE of 1.77 and a mean RMSE of 2.12. It clearly outperforms in hydraulic fluid and engine oil, where its autoregressive structure better captures the series' inertia. The case of brakes (series 4) warrants a comment, as it is the only one where the two metrics differ: SARIMA achieves the lower MAE but the higher RMSE of the two leading models, meaning it commits a slightly lower average error at the cost of some larger occasional errors. From an inventory perspective, where extreme deviations carry significant weight, this difference tips the balance in favor of Excel for that specific series.

The neural network comes in last across all seven series, and by a wide margin. The explanation lies not in seasonality, since the model includes a seven-day lag that captures the weekly cycle, but in two features of its operation. The first is that a perceptron trained on a range of values tends to extrapolate poorly when the series grows and the test period takes it outside that range, a known limitation of these models compared to methods that explicitly account for the trend. The second is the recursive nature of its forecast, which feeds back into its own predictions and accumulates error over the two-week period. Series 5 illustrates this well: despite being one of the most stable, the network's error exceeds the series' natural variability, a sign that the forecast deviates rather than tracking the level. The result aligns with observations from the M4 competition, where several machine learning methods failed to outperform simpler, well-calibrated statistical benchmarks (Makridakis et al., 2020).

6.4 The effect of the seasonal period

A key factor influencing the entire comparison is the choice of seasonal period. The time series have a weekly cycle, and the three models address it with a period of seven: Excel detects this automatically, while SARIMA and the neural network are explicitly set to this period. This alignment is what allows for a fair comparison, because none of the differences in the previous matrix are due to a model operating with the wrong seasonality. The importance of setting the cycle correctly is consistent with the literature, which identifies seasonality as one of the factors that most affect demand forecasting (Tadayonrad and Ndiaye, 2023).

It is worth noting that an initial configuration set the period to twelve, inherited from templates designed for monthly data. Applying that value to a daily series with a weekly cycle would have obscured the pattern the models were meant to capture, which is why it was corrected to seven before generating the results presented here. The experiment, therefore, does not numerically contrast the seven-period with the twelve-period, but rather starts from the correct configuration. Quantifying how much each model performs worse with a poorly adjusted period would be an interesting sensitivity analysis and is proposed as a possible extension of this work. What the results do show is that, with the cycle properly set, the weight of the decision shifts to how each model handles the trend and the forecast several steps ahead, which is where the neural network loses ground.

6.5 Comparative visualization: The radar chart

To compare the three models at a glance, the MAE values are transformed into a score from 0 to 10 using the expression presented in Section 5.3, such that a higher score indicates a smaller error. Figure 6.4 represents them in a radar chart, a multi-criteria visualization that makes it easy to see which model is best suited for each consumable (Tadayonrad and Ndiaye, 2023).

On the radar chart, Excel and SARIMA form two large, highly overlapping polygons that cover almost the entire surface of the graph, while the neural network is confined to a small, highly irregular polygon toward the center. The figure visually illustrates what the tables show: two leading models practically tied, with Excel slightly ahead, and a third that lags behind in all demand patterns. The network's scores drop particularly sharply for motor oil and line consumables, the series where its recursive forecast deviates most from actual consumption.

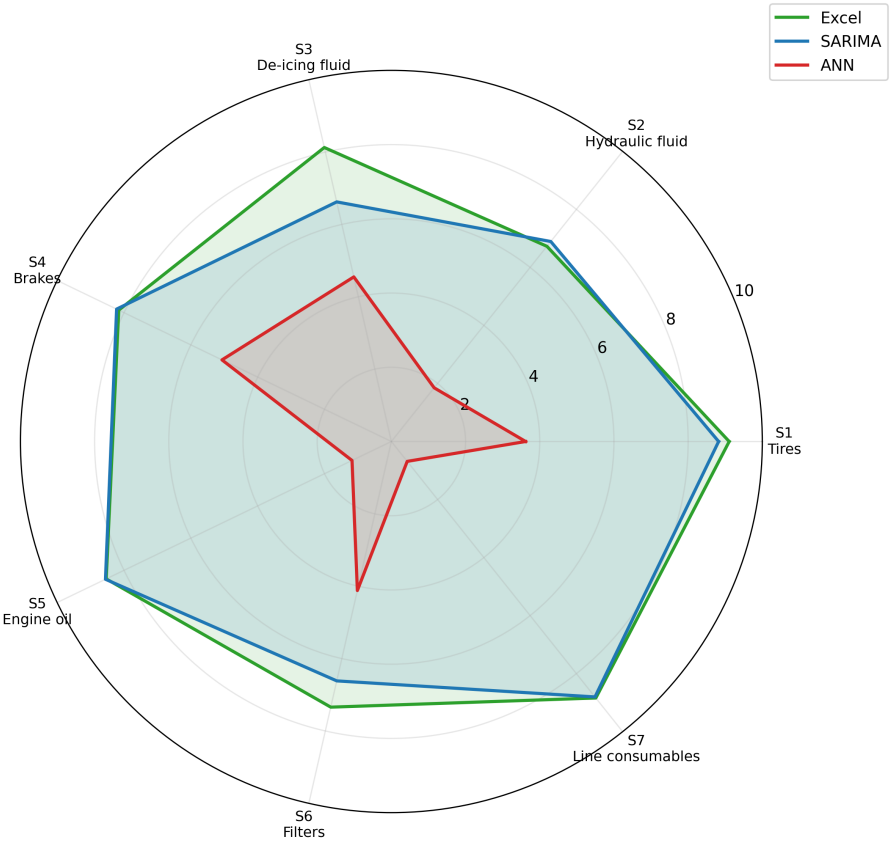


Figure 6.4: Radar chart: score by model and series (based on MAE).

Chapter 7

Financial translation

7.1 From error reduction to safety margin reduction

The analysis in the previous chapters falls short if it stops at the forecast error. What really matters to an operator is how much capital they can free up by improving their forecasts, and that link is provided by the safety stock formula presented in Chapter 3:

$$SS = Z \cdot \sigma_{\text{error}} \cdot \sqrt{LT} \tag{7.1}$$

The factor Z depends on the service level to be guaranteed, σ_{error} is the standard deviation of the forecast error, and LT is the supplier's lead time. From the results in Chapter 6, we obtain precisely σ_{error}, because the RMSE of each model coincides with that deviation when the forecast does not exhibit a systematic bias (Gonçalves et al., 2020). We can, therefore, substitute σ_{error} with the RMSE of each model and calculate the safety stock required by each.

There is a property of the formula worth noting before plugging in numbers. For a given item, Z and LT are common to all three models, so the safety stock is directly proportional to the RMSE. In other words, the ratio of the safety stock of two models is equal to the ratio of their RMSEs, regardless of the service level or lead time we assume. This relationship makes the conclusion robust, because even if we change the assumptions, the relative advantage of the more accurate model remains intact.

Let's look at this with the de-icing fluid (series 3), where the three models are clearly distinct. If we set a service level of 95%, which corresponds to Z = 1.65, and a representative lead time of 7 days, so that $\sqrt{LT} = 2.65$, the safety stock for each model is as follows:

$$\begin{aligned} \text{Excel:} &\quad SS = 1.65 \cdot 1.90 \cdot 2.65 = 8.3 \text{ units} \\ \text{SARIMA:} &\quad SS = 1.65 \cdot 3.24 \cdot 2.65 = 14.1 \text{ units} \\ \text{ANN:} &\quad SS = 1.65 \cdot 5.63 \cdot 2.65 = 24.6 \text{ units} \end{aligned}$$

The most accurate model needs to set aside 8 units of thawed stock to cover the same probability of not running out of stock, while the neural network requires nearly 25, that is, practically three times as much. That factor of three is no coincidence, but rather an exact reflection of the fact that the network's RMSE is three times that of Excel (5.63 versus 1.90). The worst forecast does not provide better protection; it merely forces us to hold more inventory to compensate for its lack of accuracy.

7.2 Estimated reduction in capital tied up in safety stock

The next step is to price these units. The experiment's dataset does not include actual costs or lead times, so we work with representative assumptions that we make explicit so the reader can rescale the result to their own case. We maintain the 95% service level (Z = 1.65) and the 7-day lead time for the seven PNs, and we assign each consumable a unit cost commensurate with its nature, ranging from 8 EUR for a de-icing container or 25 EUR for a line kit to 2.800 EUR for a brake or 3.500 EUR for a tire.

Based on the RMSE for each model, we first calculate the safety stock in units, shown in Table 7.1, and then the capital tied up by that stock by multiplying it by the unit cost, as shown in Table 7.2.

Table 7.1: Safety stock in units by model (Z = 1.65; LT = 7 days)

Consumable	Unit cost (EUR)	SS SARIMA	SS ANN	SS Excel
Tires (S1)	3.500,00	5.2	30.0	4.7
Hydraulic fluid (S2)	45,00	12.2	35.4	12.6
De-icing fluid (S3)	8,00	14.1	24.6	8.3
Brakes (S4)	2.800,00	7.8	21.1	7.2
Engine oil (S5)	35,00	6.3	34.1	6.4
Filters (S6)	120,00	12.8	23.7	10.5
Line consumables (S7)	25,00	6.4	36.3	6.0

The difference between the models is of such magnitude that it can hardly be ignored.

Table 7.2: Capital tied up in safety stock and capital released (EUR)

Consumable	Cost SARIMA	Cost ANN	Cost Excel	Savings vs. ANN
Tires (S1)	18.133,56	104.847,67	16.334,17	88.513,50
Hydraulic fluid (S2)	549,68	1.592,38	568,11	1.042,70
De-icing fluid (S3)	113,03	196,59	66,22	130,37
Brakes (S4)	21.803,06	59.202,27	20.166,55	39.035,72
Engine oil (S5)	219,22	1.194,14	225,46	974,92
Filters (S6)	1.533,06	2.843,49	1.265,34	1.578,15
Line consumables (S7)	160,55	907,40	150,34	757,06

Excel's exponential smoothing ties up approximately 38.800 EUR while SARIMA comes close at 42.500 EUR. And the neural network drives the figure up to 170.800 EUR, that is four and a half times the capital required by the most accurate model. Choosing the best of the three for each reference instead of the neural network would free up 132.032 EUR that would otherwise be tied up without providing protection. Since Excel is the most accurate in almost all series, that figure is equivalent, in practice, to replacing the neural network with the baseline.

The distribution of savings is not uniform, and there lies much of the value of analyzing reference by reference rather than aggregating the entire catalog into a single figure (Tadayonrad and Ndiaye, 2023). Almost all the capital at stake is concentrated in two references. Tires contribute 88.514 EUR in savings and brakes another 39.036 EUR, so together they account for about 97% of the total. The reason is that they combine a high unit cost with a very high RMSE from the network, and the product of these two factors skyrockets. At the opposite end of the spectrum, a low-cost consumable like de-icing fluid or line kits barely makes a dent in euros, no matter how large its relative error may be. The operational lesson is straightforward: the effort to refine the forecast pays off most where a high price intersects with demand that is difficult to predict, and it is advisable to prioritize those items rather than spreading the effort evenly.

The gap between the top two models is small but consistent. Excel ties up about 3.700 EUR less than SARIMA overall, representing nearly 9% less capital, and is the cheaper option in five of the seven items. SARIMA only outperforms it in hydraulic fluid and motor oil, the same series where it achieved the lowest RMSE, and by margins of just a few euros.

7.3 Impact on inventory without increasing the risk of AOG

There remains one obvious objection that should be addressed. Reducing the safety stock could be interpreted as cutting the cushion that protects against a breakdown, and in aircraft maintenance, a breakdown can result in an aircraft being grounded, with the associated costs (Alomar and Nikita, 2025). The answer lies in the formula itself. The reduction is not achieved by lowering the service level, but by maintaining it. The Z-factor, which determines the probability of not running out of stock, is the same for all three models, so that the smaller safety stock in Excel provides exactly the same protection as that of the neural network.

What changes is not the risk assumed, but the inventory needed to sustain it. An accurate forecast comes closer to actual consumption, leaves a small margin of error, and therefore needs to hold fewer units to achieve 95% service. An inaccurate forecast is wrong more often, and the only way to continue guaranteeing that same 95% is to inflate the buffer. The extra capital tied up by the neural network does not buy more availability; it merely pays for its own lack of accuracy (Stoilov and Stoilova, 2023).

This leads to the study's financial conclusion. In this case, the model that is cheapest to implement and operate is also the one that ties up the least capital, without sacrificing a single point of service level. The added complexity of the neural network does not translate into savings, but rather the opposite, bringing the same lesson to the monetary realm that the M4 competition demonstrated in the realm of accuracy (Makridakis et al., 2020). And it is worth remembering the scale of this example. We are talking about seven references and a single quarter. In a fleet with a catalog of hundreds of part numbers, the figure of 132.032 EUR would multiply, and the decision on which forecasting model to use would cease to be a technical matter and become a direct lever on the organization's working capital.

Chapter 8

Conclusions

8.1 Achieved and unachieved objectives

The study aimed to determine whether more sophisticated forecasting models improve the accuracy of the method the industry uses by default, and to measure what that improvement would mean for safety stock. Here we summarize the extent to which the five objectives outlined in the introduction were met and what the results reveal when considered as a whole.

The first four objectives, which were experimental in nature, were fully achieved. We generated seven daily demand series with a trend and weekly cycle, each associated with an MRO warehouse consumable and with known actual consumption. We implemented the three planned models, Excel's exponential smoothing as a baseline, a SARIMA model, and a neural network, all aligned to a seven-day weekly period. We calculated the MAE and RMSE for each combination of model and pattern, and compiled them into a matrix that allows us to assess the performance of each method against the baseline.

The key result, however, contradicts the initial expectation. The more complex models did not outperform the baseline. Excel's exponential smoothing achieved the lowest error in five of the seven series according to the RMSE, with an average MAE of 1.51 and an average RMSE of 1.83. SARIMA came very close, with 1.77 and 2.12, and outperformed the others in the remaining two time series, hydraulic fluid and motor oil, where its autoregressive structure better captures the inertia. The neural network was the least accurate across all seven series, with an average error three times that of the other two models. The added complexity did not translate into accuracy, a finding consistent with observations in the M4 competition (Makridakis et al., 2020) and one that should be viewed as a result of the study rather than a setback.

The fifth objective, financial translation, was also met, with the caveat of working with representative assumptions regarding cost and lead time. Since the safety stock is propor-

tional to the RMSE, a more accurate forecast directly reduces the safety stock and the capital it ties up (Gonçalves et al., 2020). Calculated on a per-consumable basis, with a 95% service level and a seven-day lead time, the capital freed up by choosing the more accurate model instead of the neural network is distributed as follows:

Table 8.1: Capital released per consumable versus the ANN (EUR)

Consumable	Capital released versus the ANN
Tires	88.514
Brakes	39.036
Filters	1.578
Hydraulic fluid	1.043
Engine oil	975
Line consumables	757
De-icing fluid	130
TOTAL	***132,032***

Overall, the baseline model ties up approximately 38.800 EUR in safety stock for the seven PN's. SARIMA comes close at 42.500 EUR, and the neural network drives the figure up to 170.800 EUR, four and a half times more. The difference of 132.032 EUR is not distributed evenly. Almost all the capital at stake is concentrated in tires and brakes, which alone account for about 97% of the total, because they combine a high unit cost with a very high network error. This finding encapsulates one of the most practical lessons of the study: the effort to refine the forecast pays off most where high prices intersect with demand that is difficult to predict, and it is advisable to prioritize these items rather than spreading the effort evenly.

From all this, we derive a guiding principle for the work. For this type of demand, continuous and with weekly seasonality, a simple, inexpensive, and well-configured method is hard to beat, and it is also the option that ties up the least capital. The value lay not so much in making the model more sophisticated as in getting the configuration right, starting by adjusting the seasonal period to the actual seven-day cycle, and in analyzing each item separately rather than aggregating the entire catalog into a single figure, which would have obscured exactly where the savings lay.

It is worth concluding this section by placing the result in its operational context. For a materials manager, the conclusion is not that the neural network is useless, but that its complexity must be justified by an improvement that did not occur here. A planning team that already uses Excel's exponential smoothing does not need to switch tools to adjust its safety stock. What it needs is to feed that method with the correct seasonal period and work item by item, especially for expensive parts with irregular demand, where

a single point of precision translates into thousands of euros in savings without affecting service levels.

8.2 Study limitations

The scope of these conclusions is limited by several constraints, which we already anticipated in the introduction. The first stands from working with synthetic data. The simulation gave us control over the fundamental truth of each series and allowed us to isolate the behavior of each model, but in exchange it lacks the texture of a real historical record, with its recording errors and one-off events.

The second limitation concerns the type of demand. None of the seven time series is intermittent, so the conclusions apply to continuous demand with trends and seasonality, and not to the cases with long periods of zero demand that are common in the aviation catalog. The third limitation is that we did not have access to actual prices or delivery times, so the financial quantification illustrates the order of magnitude of the savings but does not replace a calculation using figures from a specific operator. A fourth limitation stands due each model was evaluated with a single configuration and over a two-week horizon, so we cannot rule out the possibility that a neural network with more data or a different architecture might improve its results.

8.3 Future research directions

These limitations give rise to future lines of research. The most immediate is to extend the study to intermittent demand, using methods designed for that regime, and to verify whether the order of the models holds when the time series is no longer continuous. A second approach would be to repeat the experiment using real-world data from a fleet, with actual costs and timeframes, to convert the financial estimate into a closed-form calculation that an operator can incorporate into their management system.

In terms of configuration, a sensitivity analysis of the seasonal period remains pending, which would numerically compare the seven-day fit against the twelve-day fit and measure how much each model loses when that period does not coincide with the actual cycle. This would be the way to quantify the correction that was applied from the outset in this study. Regarding the models, the results of the M4 competition suggest that combinations of statistical methods with neural networks perform better than each one separately, and recent research confirms this in logistics with frameworks such as ARIMA-LSTM, an approach worth exploring before ruling out machine learning for this problem.

There are also two extensions that connect to the sector's context. The first is to incorporate into the forecast the signal from airworthiness directives, which are published in

advance and allow for anticipating part of the non-routine demand, information that few systems systematically leverage. The second is to extend the analysis to rotables, which require a different framework focused on repair time rather than consumption forecasting, and which would complete the view of the aircraft inventory that this work has addressed only in terms of consumables.

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), 5129. https://doi.org/10.3390/app15095129

Gonçalves, J. N. C., Carvalho, M. S., & Cortez, P. (2020). Operations research models and methods for safety stock determination: A review. *Operations Research Perspectives, 7*, 100164. https://doi.org/10.1016/j.orp.2020.100164

Grech, V. (2024). Forecasting with Excel. *Acta Medica (Hradec Králové), 67*(3), 87--90. https://doi.org/10.14712/18059694.2025.3

International Air Transport Association. (2025). *Guidance material and best practices for aircraft leases* (4.1 Edition).

International Air Transport Association. (2026). *Maintenance cost data eXchange (MCX) - Executive report FY2024 data*.

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

Li, Y. (2018). Supply chain management of aircraft industry. En *8th International Conference on Management, Education and Information (MEICI 2018)*. Atlantis Press.

Lin, N., Yuan, Y., & Li, H. (2023). Research on logistics volume prediction model based on ARIMA-LSTM. *Highlights in Business, Economics and Management ECRM 2023, 18*, 98-104.

Mocenco, D. (2014). Supply chain features of the aerospace industry: Particular case Airbus and Boeing. *Scientific Bulletin -- Economic Sciences, 14*(2), 17-25.

Rahardja, D. (2021). Statistical time-series forecast via Microsoft Excel (FORECAST.ETS)

built-in function. *Quest Journals Journal of Research in Applied Mathematics, 7*(11), 69-73.

Stoilov, T., & Stoilova, K. (2023). Inventory and safety stock optimization. En *Proc. XXXII International Scientific Conference Electronics - ET2023*. Sozopol, Bulgaria.

Tadayonrad, Y., & Ndiaye, A. B. (2023). A new key performance indicator model for demand forecasting in inventory management considering supply chain reliability and seasonality. *Supply Chain Analytics, 3*, 100026. https://doi.org/10.1016/j.sca.2023.100026

Wandelt, S., Antoniou, C., Birolini, S., Delahaye, D., Dresner, M., Fu, X., Gössling, S., Hong, S.-J., Odoni, A. R., Zanin, M., Zhang, A., Zhang, H., Zhang, Y., & Sun, X. (2024). Status quo and challenges in air transport management research. *Journal of the Air Transport Research Society, 2*, 100014

Appendix A

Python Source Code

The two scripts below were used to generate all SARIMA and ANN results reported in Chapter 6. Both read the same input file (DATABASE.xlsx), write their numerical output to an Excel workbook, and save the diagnostic chart as a PNG. The only external dependencies are `pandas`, `numpy`, `matplotlib`, `statsmodels`, and `scikit-learn`, all of which ship with a standard Anaconda or pip installation.

A.1 SARIMA Model

```
# -*- coding: utf-8 -*-
import os
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX

input_excel_path = r"C:\Users\lopez\Desktop\Re_ TFG - FIRST REPORT ASSIGNMENT\
    DATABASE.xlsx"

df_input = pd.read_excel(
    input_excel_path,
    sheet_name=0,
    header=None
)

data = df_input.iloc[:, 0].dropna().astype(float).tolist()

print("Number of loaded observations:", len(data))
print("Loaded data:")
print(data)
```

The script reads the first column of the first sheet, drops any empty rows, and casts

everything to float. SARIMAX will raise a type error on integer input, so the cast is not optional. The `header=None` argument prevents pandas from treating the first observation as a column name.

```
output_folder = r"C:\Users\lopez\Desktop\Re_ TFG - FIRST REPORT ASSIGNMENT\
    EXPERIMENTS"
os.makedirs(output_folder, exist_ok=True)

excel_output_path = os.path.join(output_folder, "SARIMA_results_full_period.xlsx")
figure_output_path = os.path.join(output_folder, "SARIMA_results_plot.png")
```

`os.makedirs` with `exist_ok=True` creates the output folder if it does not already exist, and does nothing if it does. Both output paths are built with `os.path.join` so the script works on Windows and Mac without changes to the path separators.

```
y = pd.Series(data)
months_observed = range(1, len(y) + 1)

model = SARIMAX(
    y,
    order=(1, 1, 1),
    seasonal_order=(1, 0, 1, 7),
    enforce_stationarity=False,
    enforce_invertibility=False
)

results = model.fit(disp=False)
print(results.summary())
```

The list is converted to a `pd.Series` so statsmodels can work with it. The model is `SARIMA(1,1,1)(1,0,1)[7]`: one autoregressive term, one differencing step, and one moving average term in the non-seasonal part; one autoregressive and one moving average term in the seasonal block with period 7, which matches the weekly cycle of daily data. Both `enforce_stationarity` and `enforce_invertibility` are disabled so the optimizer can explore the full parameter space without hard boundaries. `disp=False` suppresses the verbose convergence log.

```
fitted_values = results.fittedvalues

forecast_steps = 14
forecast_results = results.get_forecast(steps=forecast_steps)
forecast_values = forecast_results.predicted_mean
forecast_ci = forecast_results.conf_int()

sarima_full = pd.concat([fitted_values, forecast_values], ignore_index=True)
months_full = range(1, len(sarima_full) + 1)
```

`get_forecast` returns a results object; `predicted_mean` is the point estimate and `conf_int()` holds the 95% confidence bounds. Fitted values and forecast are concatenated into one continuous series so the chart and the output table share a single consistent index.

```
output_df = pd.DataFrame({
    "Month": list(months_full),
    "Observed_Data": list(y) + [None] * forecast_steps,
    "SARIMA_Fitted_and_Forecast": sarima_full
})

print("\nSARIMA fitted values and forecast:")
print(output_df)

output_df.to_excel(excel_output_path, index=False)
print("\nResults saved as:")
print(excel_output_path)
```

The output table aligns observed data and model output on the same row index. Forecast rows carry `None` in the `Observed_Data` column since no actual value exists for those periods.

```
plt.figure(figsize=(13, 4.8))

plt.plot(months_observed, y, linestyle="--", linewidth=2, label="Data")
plt.plot(months_full, sarima_full, linestyle="-", linewidth=2, label="SARIMA")

plt.title("RESULTS - SARIMA", fontsize=18)
plt.xlabel("Month", fontsize=14)
plt.ylabel("Value", fontsize=14)
plt.xlim(0, len(y) + forecast_steps + 1)
plt.grid(True)
plt.legend(fontsize=13, loc="lower center", ncol=2)
plt.tight_layout()

plt.savefig(figure_output_path, dpi=300)
print("Figure saved as:")
print(figure_output_path)

plt.show()
```

The dashed line is the observed series and the solid line is the model output (fitted period followed by the 14-day forecast). Saving at 300 dpi produces a print-quality PNG. `plt.tight_layout()` adjusts margins automatically so labels do not get clipped.

A.2 ANN Model

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.neural_network import MLPRegressor
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline

input_excel_path = r"C:\Users\lopez\Desktop\Re_ TFG - FIRST REPORT ASSIGNMENT\
    DATABASE.xlsx"

df_input = pd.read_excel(
    input_excel_path,
    sheet_name=0,
    header=None
)

data = df_input.iloc[:, 0].dropna().astype(float).tolist()

print("Number of loaded observations:", len(data))
print("Loaded data:")
print(data)

y = np.array(data, dtype=float)
```

Same data loading logic as the SARIMA script. The only difference is that the result is stored as a `numpy` array rather than a `pd.Series`, since scikit-learn operates on numpy arrays internally. The cast to `float` is explicit to avoid silent integer arithmetic later in the loop.

```
output_folder = r"C:\Users\lopez\Desktop\Re_ TFG - FIRST REPORT ASSIGNMENT\
    EXPERIMENTS"
os.makedirs(output_folder, exist_ok=True)

excel_output_path = os.path.join(output_folder, "ANN_results_full_period.xlsx")
figure_output_path = os.path.join(output_folder, "ANN_results_plot.png")
```

Output folder and file paths follow the same pattern as in the SARIMA script so both experiments write to the same directory and results are easy to compare side by side.

```
seasonality = 7
forecast_steps = 14

X = []
```

```
target = []
month_for_training = []

for t in range(seasonality, len(y)):
    X.append([
        y[t-1],
        y[t-2],
        y[t-7]
    ])
    target.append(y[t])
    month_for_training.append(t + 1)

X = np.array(X)
target = np.array(target)
```

Three lagged values are used as features for each observation: `y[t-1]` and `y[t-2]` capture short-term inertia, and `y[t-7]` gives the network access to the same weekday of the previous week, which is what encodes the weekly cycle. The loop starts at `t = 7` because the `t-7` lag is not available for the first seven days. Those rows are skipped entirely rather than filled with zeros.

```
ann_model = Pipeline([
    ("scaler", StandardScaler()),
    ("ann", MLPRegressor(
        hidden_layer_sizes=(10, 5),
        activation="relu",
        solver="adam",
        max_iter=5000,
        random_state=42
    ))
])

ann_model.fit(X, target)
```

`StandardScaler` is placed inside the `Pipeline` so the network never sees raw demand values, only zero-mean unit-variance inputs. This is what allows Adam to converge without manual learning-rate tuning. Two hidden layers of 10 and 5 neurons keep the architecture simple relative to the series length; going wider or deeper here tends to overfit. `random_state=42` makes the weight initialization reproducible across runs.

```
fitted_part = ann_model.predict(X)
ann_fitted = [None] * seasonality + list(fitted_part)
```

The first seven positions are set to `None` because no fitted value exists for them: the feature matrix starts at `t = 7`, so observations 0--6 have no prediction. Padding with `None` keeps the fitted series aligned with the original index.

```
extended_series = list(y.copy())
ann_forecast = []
forecast_input_table = []

for step in range(forecast_steps):
    t = len(extended_series)

    input_y_t_1 = extended_series[t-1]
    input_y_t_2 = extended_series[t-2]
    input_y_t_7 = extended_series[t-7]

    input_features = np.array([[
        input_y_t_1,
        input_y_t_2,
        input_y_t_7
    ]])

    predicted_value = ann_model.predict(input_features)[0]

    ann_forecast.append(predicted_value)
    extended_series.append(predicted_value)

    forecast_input_table.append({
        "Forecast_Month": t + 1,
        "Input_y_t_minus_1": input_y_t_1,
        "Input_y_t_minus_2": input_y_t_2,
        "Input_y_t_minus_7": input_y_t_7,
        "Forecast_Value": predicted_value
    })
```

Each predicted value is appended to `extended_series` and immediately used as an input for the next step. This recursive approach is standard for single-step models but comes with a known drawback: any deviation in the early predictions feeds forward and can compound over the horizon. The `forecast_input_table` records exactly which lagged values were used at each step, which makes it possible to audit the forecast inputs after the fact.

```
ann_full = ann_fitted + ann_forecast
months_observed = list(range(1, len(y) + 1))
months_full = list(range(1, len(ann_full) + 1))

output_df = pd.DataFrame({
    "Month": months_full,
    "Observed_Data": list(y) + [None] * forecast_steps,
    "ANN_Fitted_and_Forecast": ann_full
})
```

```
output_df["Period"] = (
    ["Observed period"] * len(y) + ["Forecast period"] * forecast_steps
)

training_df = pd.DataFrame({
    "Month": month_for_training,
    "Input_y_t_minus_1": X[:, 0],
    "Input_y_t_minus_2": X[:, 1],
    "Input_y_t_minus_7": X[:, 2],
    "Target_y_t": target,
    "ANN_Fitted_Value": fitted_part
})

forecast_input_df = pd.DataFrame(forecast_input_table)

print("\nANN FITTED VALUES AND FORECAST")
print(output_df.to_string(index=False))

print("\nANN TRAINING INPUTS")
print(training_df.to_string(index=False))

print("\nANN FORECAST INPUTS")
print(forecast_input_df.to_string(index=False))

with pd.ExcelWriter(excel_output_path, engine="openpyxl") as writer:
    output_df.to_excel(writer, sheet_name="Diagram_Data", index=False)
    training_df.to_excel(writer, sheet_name="Training_Data", index=False)
    forecast_input_df.to_excel(writer, sheet_name="Forecast_Inputs", index=False)

print("Excel file saved to:")
print(excel_output_path)
```

Three separate DataFrames are written to three sheets in the same workbook. `Diagram_Data` contains the aligned observed and predicted series used for the chart. `Training_Data` stores the feature matrix and targets so the training inputs can be inspected. `Forecast_Inputs` records the exact lag values fed to the model at each forecast step, which is useful for debugging divergence in the recursive loop.

```
plt.figure(figsize=(13, 4.8))

plt.plot(months_observed, y, linestyle="--", linewidth=2, label="Data")
plt.plot(months_full, ann_full, linestyle="-", linewidth=2, label="ANN")

plt.title("RESULTS - ANN", fontsize=18)
plt.xlabel("Month", fontsize=14)
plt.ylabel("Value", fontsize=14)
plt.xlim(0, len(y) + forecast_steps + 1)
```

```
plt.grid(True)
plt.legend(fontsize=13, loc="lower center", ncol=2)
plt.tight_layout()

plt.savefig(figure_output_path, dpi=300)
print("Figure saved to:")
print(figure_output_path)

plt.show()
```

Same chart structure as the SARIMA script. The dashed line is the observed series and the solid line covers both the fitted period and the 14-day forecast. Keeping the chart format identical across both scripts makes it straightforward to compare the two visually.

Appendix B

Series-Level Diagnostic Plots

Each of the seven demand series is presented below with three plots: the full 90-day training period that the models learned from, the 14-day test period showing actual demand against all three forecasts, and a bar chart comparing the RMSE achieved by each model on that series. The colour convention is the same throughout: black for actual demand, green for Excel, blue for SARIMA, and red for the ANN.

B.1 S1 - Tires

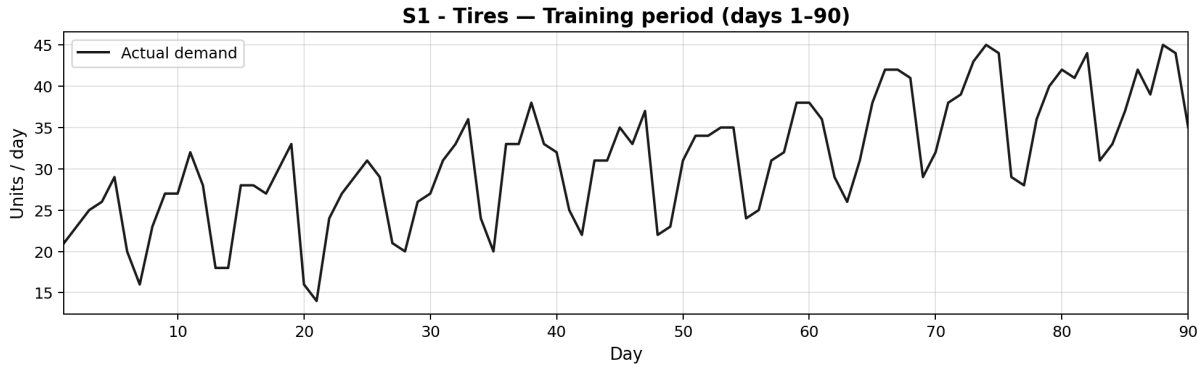


Figure B.1: S1 Tires - actual daily demand over the 90-day training period (1 January - 31 March 2026). The upward trend and the recurring weekend peak are clearly visible, driven by leisure-flight traffic concentrating at the end of the week.

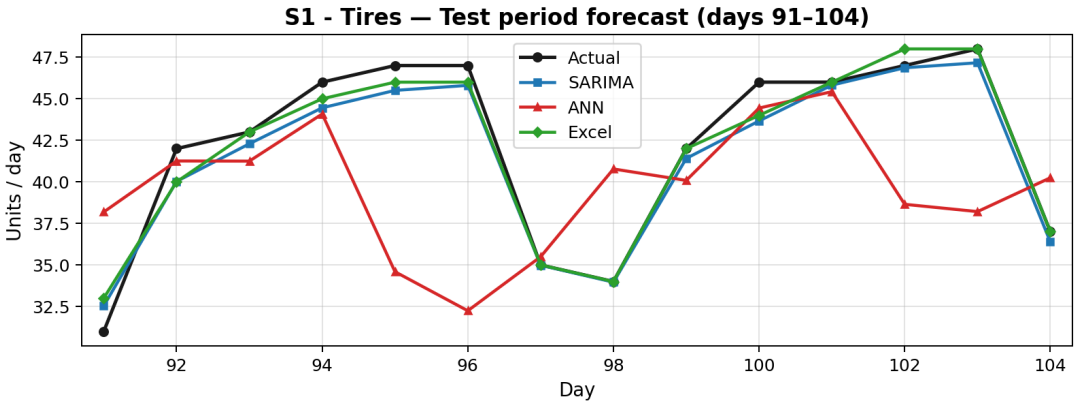


Figure B.2: S1 Tires - actual demand versus Excel, SARIMA, and ANN forecasts over the 14-day test period (days 91-104). Excel and SARIMA track the actual series closely; the ANN diverges sharply from day 95 onward due to recursive error accumulation.

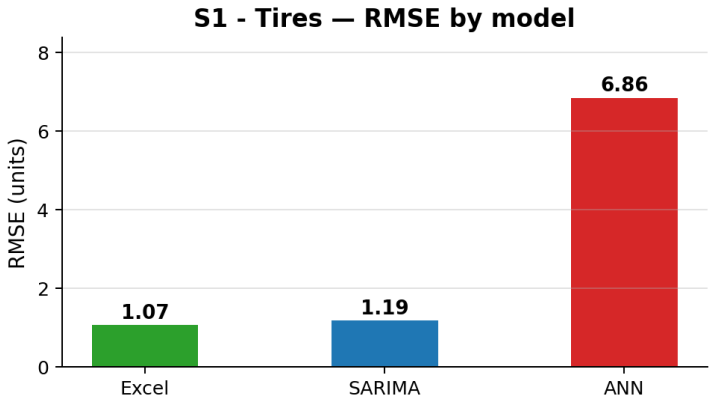


Figure B.3: S1 Tires - RMSE by model over the test period. Excel (1.07) and SARIMA (1.19) are nearly tied; the ANN (6.86) is roughly six times worse.

B.2 S2 - Hydraulic Fluid

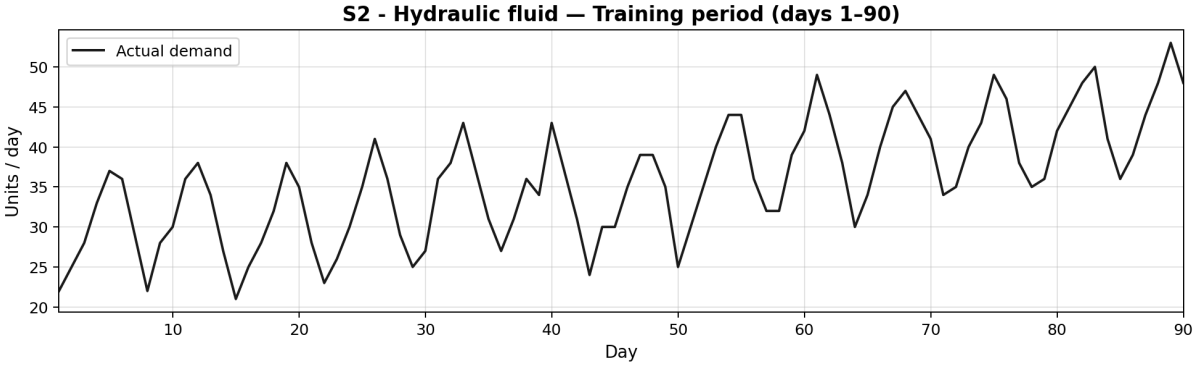


Figure B.4: S2 Hydraulic fluid -- training period. Demand is driven by scheduled maintenance tasks concentrated at the start of the week, producing a Monday--Tuesday peak that repeats throughout the quarter.

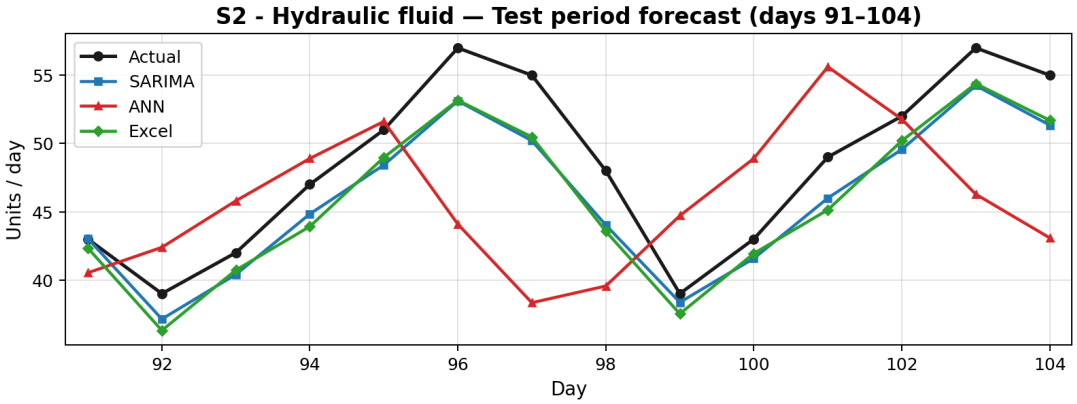


Figure B.5: S2 Hydraulic fluid - test period. SARIMA tracks the actual series slightly better than Excel here, particularly at the peak on day 103. The ANN underestimates through most of the horizon.

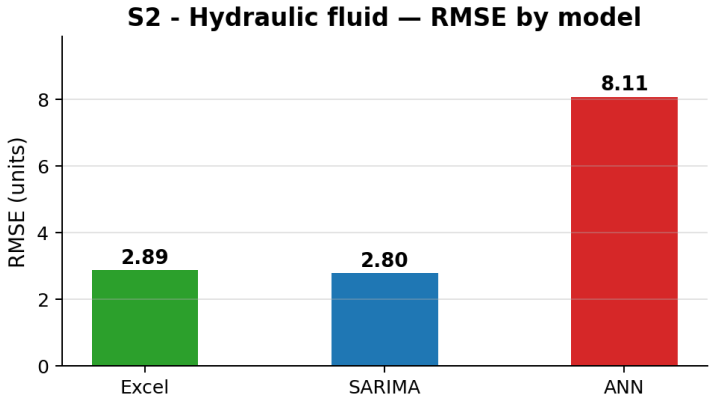


Figure B.6: S2 Hydraulic fluid - RMSE by model. SARIMA (2.80) outperforms Excel (2.89) by a narrow margin; the ANN (8.11) is the weakest by a wide margin.

B.3 S3 - De-icing Fluid

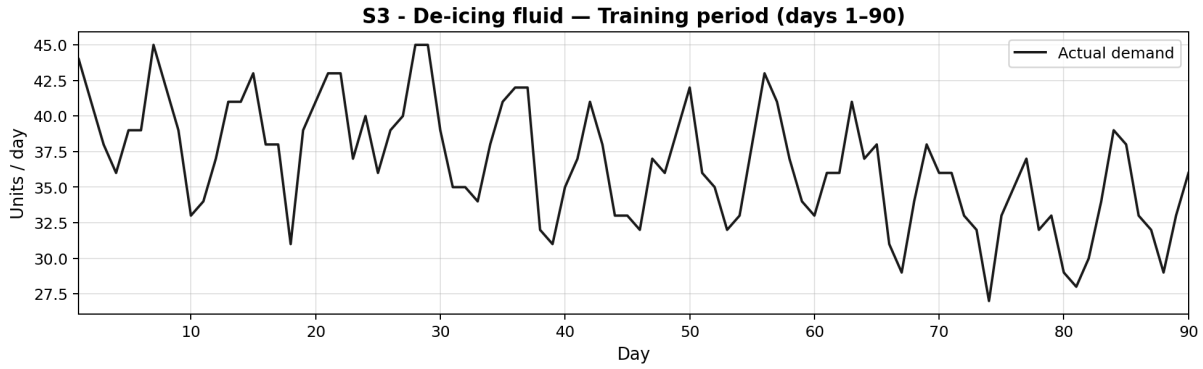


Figure B.7: S3 De-icing fluid - training period. Demand starts high in January and declines steadily as winter ends, giving this series the most pronounced downward trend in the set.

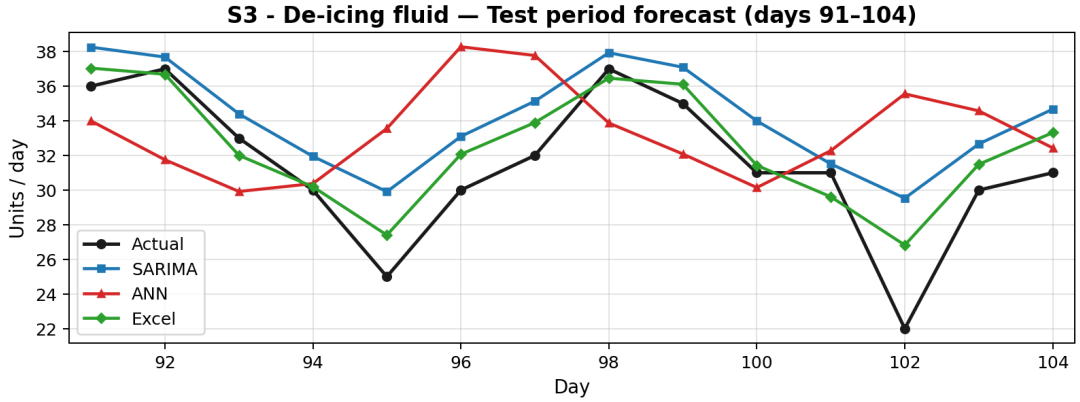


Figure B.8: S3 De-icing fluid - test period. Excel follows the declining trend and the weekly oscillation well. SARIMA overshoots on several days, and the ANN produces a nearly flat forecast that ignores both the trend and the cycle.

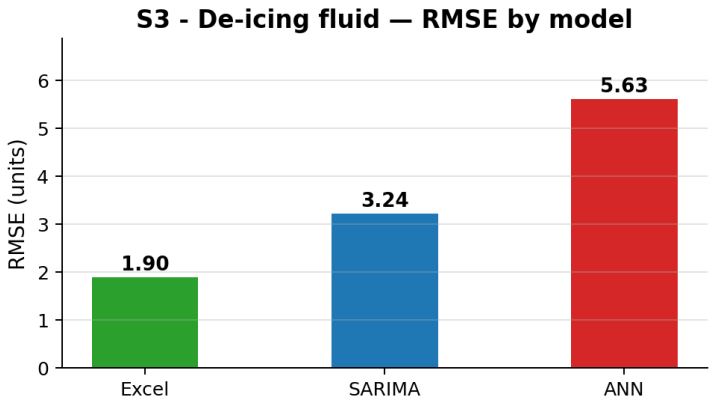


Figure B.9: S3 De-icing fluid - RMSE by model. Excel (1.90) leads clearly; SARIMA (3.24) is penalised by its occasional overshoots; the ANN (5.63) is the least accurate.

B.4 S4 - Brakes

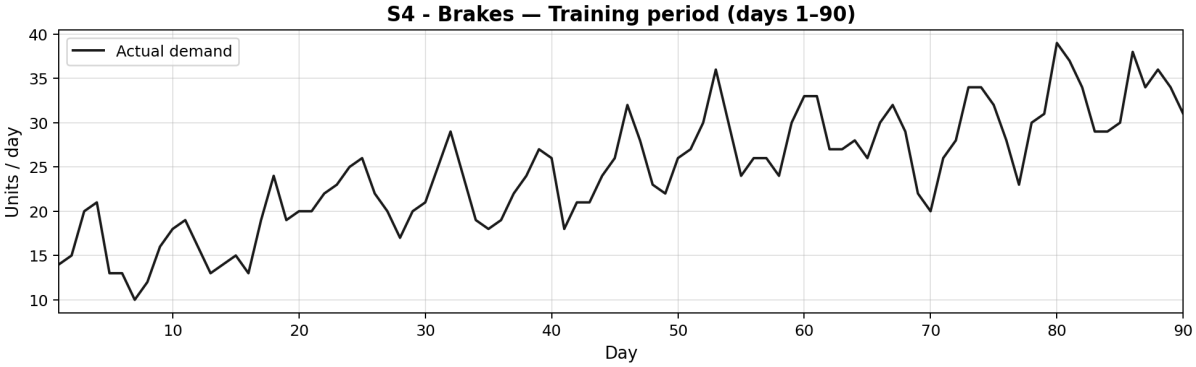


Figure B.10: S4 Brakes - training period. The series combines moderate growth with a weekend peak similar to tires, reflecting their shared dependence on leisure-flight volume.

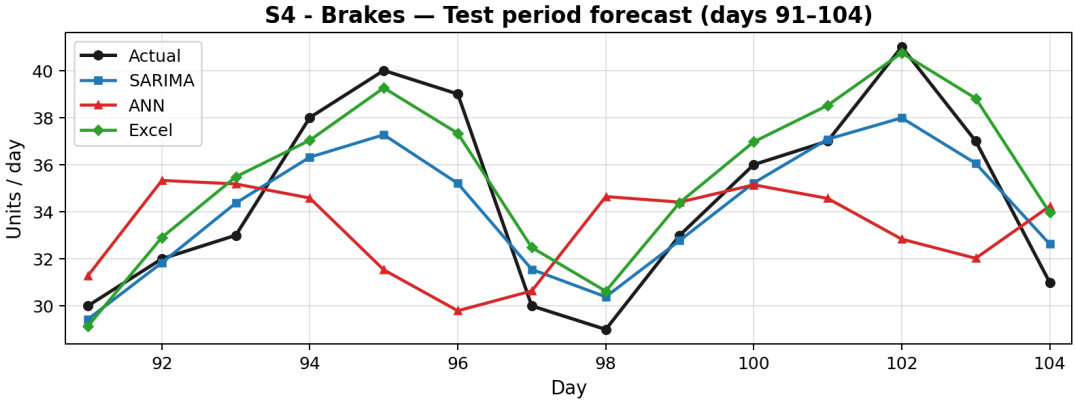


Figure B.11: S4 Brakes - test period. Excel and SARIMA are nearly indistinguishable and both follow the actual series tightly. The ANN lags the actual peaks by one to two days on several occasions.

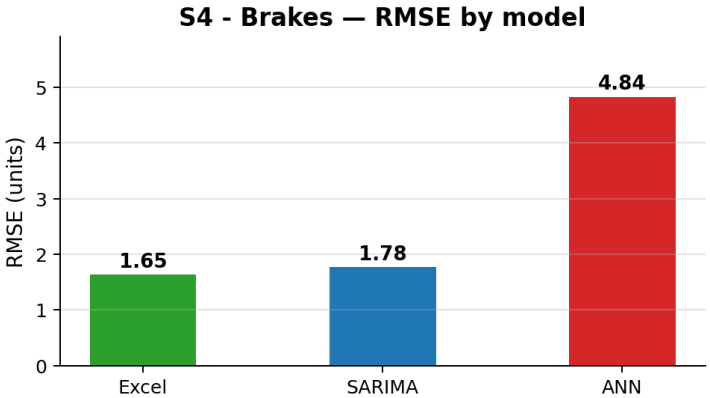


Figure B.12: S4 Brakes - RMSE by model. Excel (1.65) and SARIMA (1.78) are within 0.13 units of each other; the ANN (4.84) is notably worse.

B.5 S5 - Engine Oil

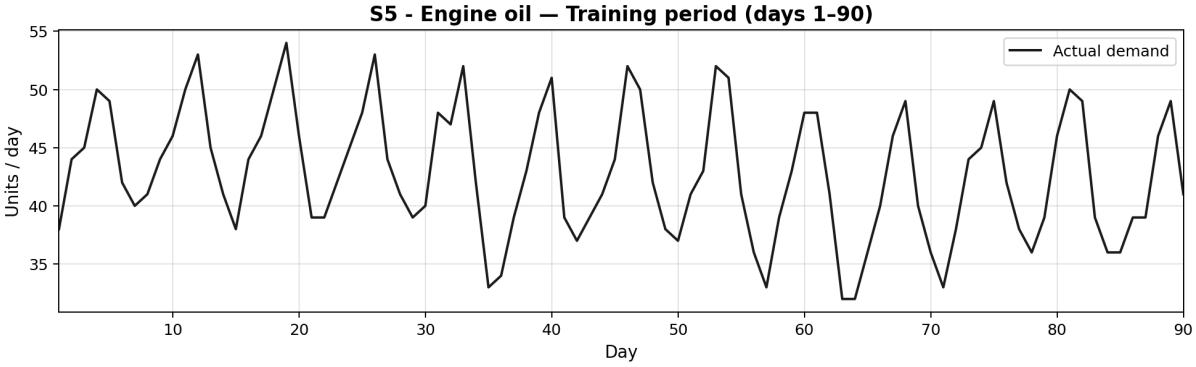


Figure B.13: S5 Engine oil - training period. This is the most level series in the set: the trend is mild and the weekly cycle is the most regular. It represents the steady baseline consumption of a mature narrowbody fleet.

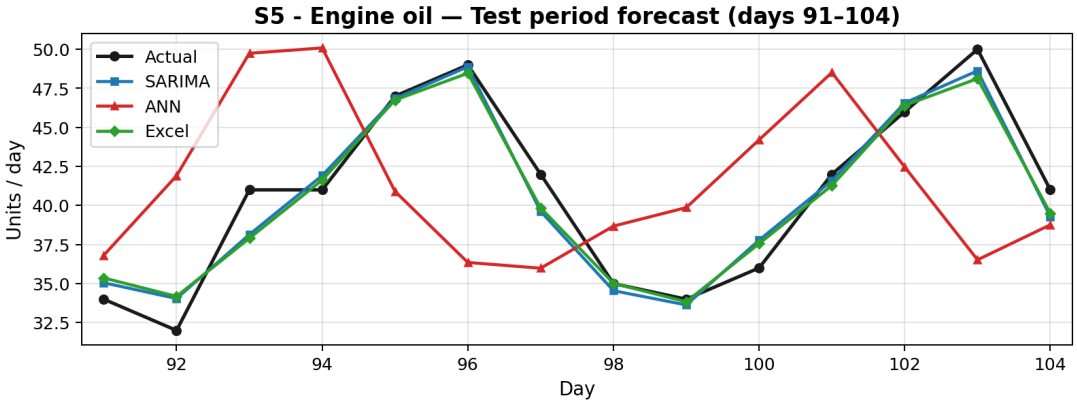


Figure B.14: S5 Engine oil - test period. Excel and SARIMA stay close to the actual line throughout. The ANN produces the most aberrant result of any series: despite being the most stable series in the set, the ANN error exceeds 12 units on some days.

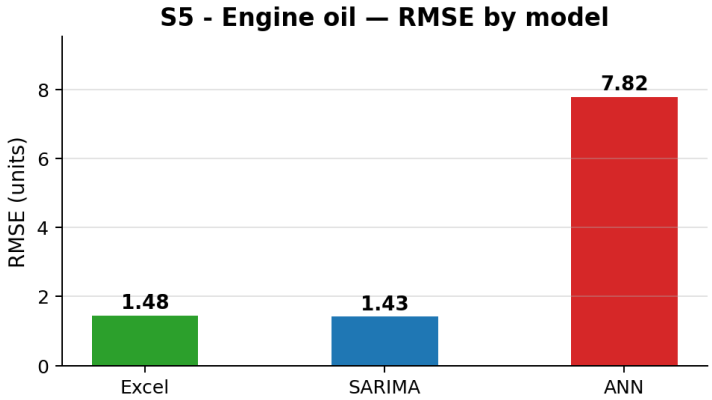


Figure B.15: S5 Engine oil - RMSE by model. SARIMA (1.43) edges out Excel (1.48) on this series; both are well below the ANN (7.82).

B.6 S6 - Filters

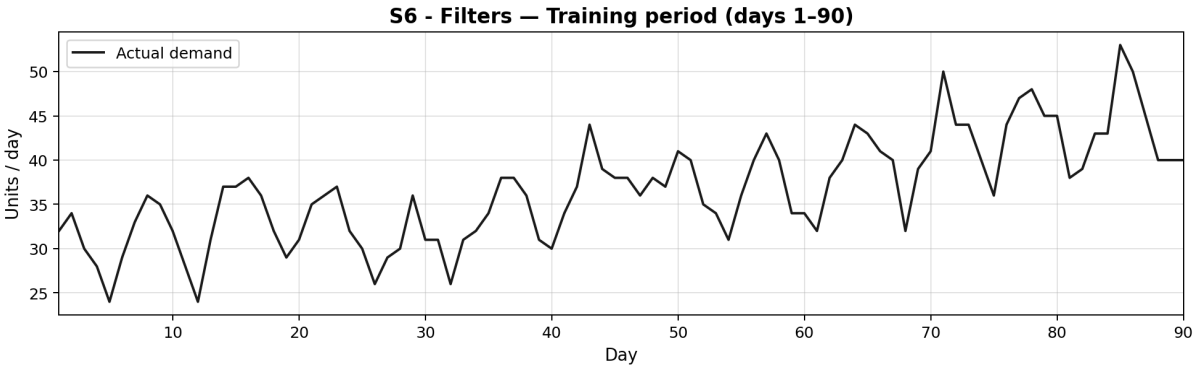


Figure B.16: S6 Filters - training period. Demand peaks midweek, tied to planned maintenance windows. The series grows gradually over the quarter as the inspection calendar fills up toward spring.

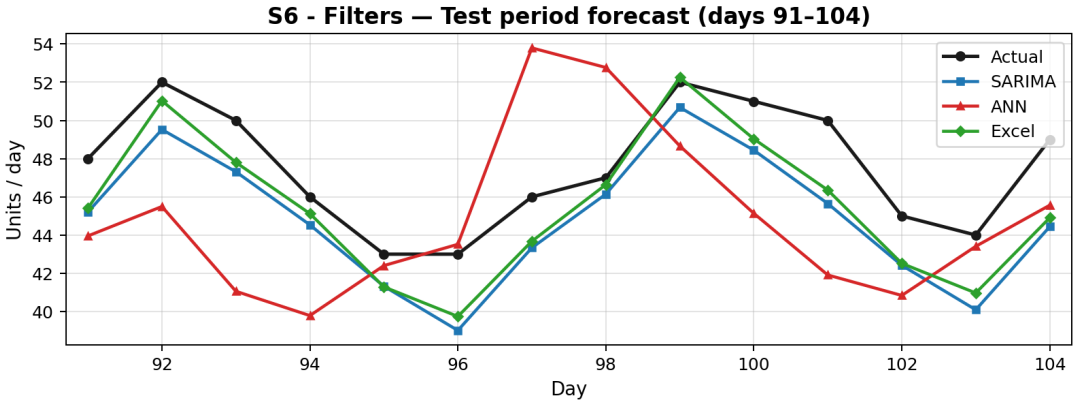


Figure B.17: S6 Filters - test period. Excel captures the midweek peaks reasonably well. SARIMA tends to lag peaks by one day; the ANN again shows wide swings that do not track the actual series.

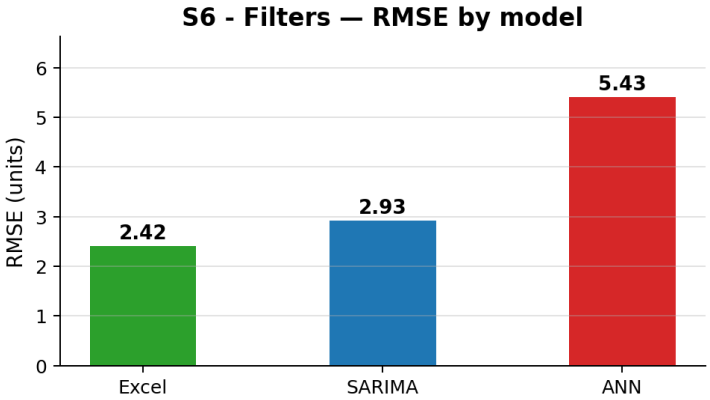


Figure B.18: S6 Filters - RMSE by model. Excel (2.42) leads over SARIMA (2.93); the ANN (5.43) is more than twice as far from the actual series.

B.7 S7 - Line Consumables

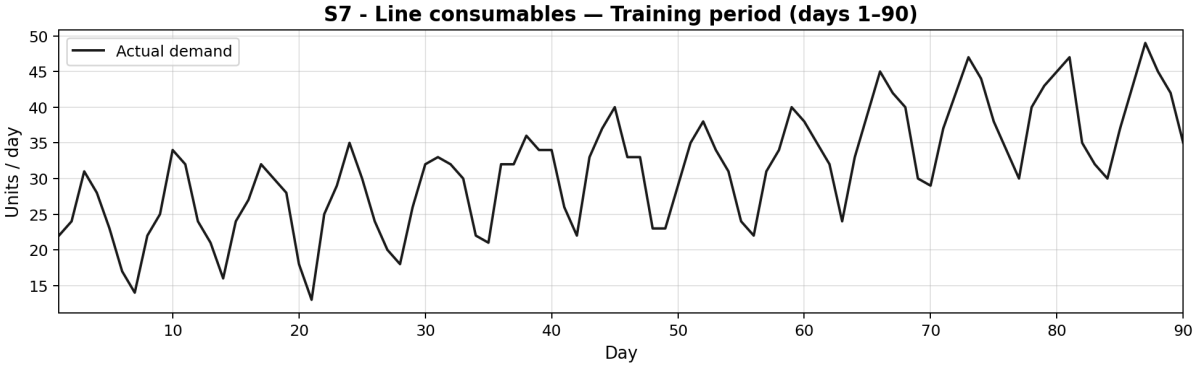


Figure B.19: S7 Line consumables - training period. Demand rises steadily from January to April, driven by increasing flight activity toward spring. The weekend peak is the sharpest in the set, reflecting the high concentration of leisure departures on Saturdays and Sundays.

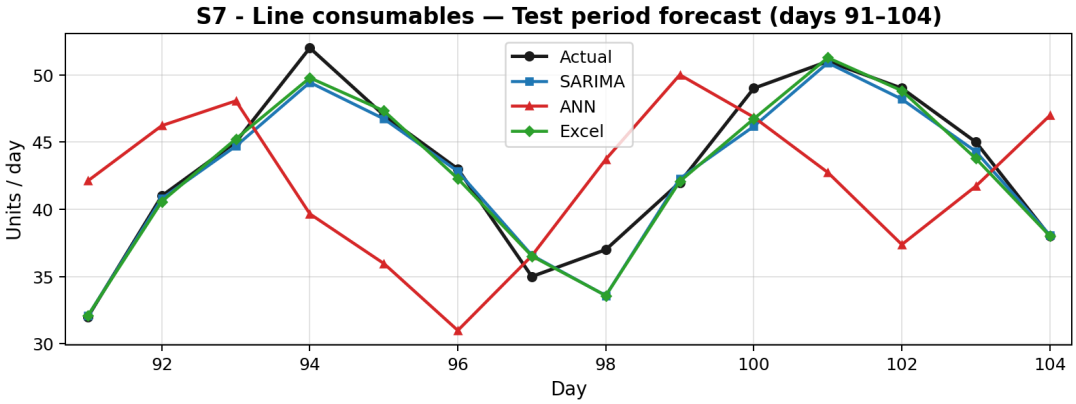


Figure B.20: S7 Line consumables - test period. Excel and SARIMA follow the actual series almost exactly, including the sharp weekend peak at day 103. The ANN diverges significantly on multiple days, producing the highest absolute errors in the whole test set.

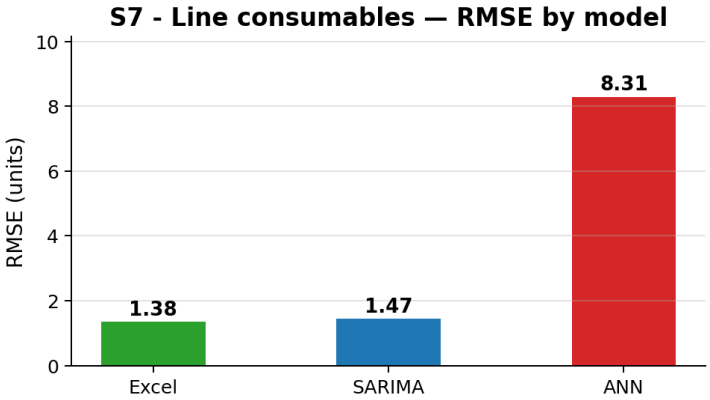


Figure B.21: S7 Line consumables - RMSE by model. Excel (1.38) and SARIMA (1.47) are very close; the ANN (8.31) posts its worst result across the seven series here.