ISSN 1648-7788 / eISSN 1822-4180

VILNIUS TECH
Vilnius Gediminas Technical University

AVIATION

2025
Volume 29
Issue 4
Pages 253–260
https://doi.org/10.3846/aviation.2025.25363

DRAWBACKS OF DEMAND ACCURACY ASSESSMENT MODELS ON THE EXAMPLE OF SLOW-MOVING SPARE PARTS IN CIVIL AVIATION

Danila LARIN ✉

Transport and Telecommunication Institute, Lauvas Street 2, 1019 Riga, Latvia

Article History:
- received 20 August 2025
- accepted 11 November 2025

Abstract. Lots of researchers worldwide use a big variety of forecast models to predict demand. After running the forecast model, researcher always has a question if received prediction was accurate or not. To do so, a number of methods exist to assess model accuracy. Application of accuracy assessment models itself is not complex. The most difficult part for researcher: interpretation of the result and the understanding of information to take the right decisions. Companies who do demand forecast in 95% of cases use only one accuracy assessment method for their forecast model. In case, companies do it for fast moving items and the business doesn't have any special requirement for the result level, it could be accepted. But in case slow-moving inventory is used and the company requires a certain service level, then there is a space for potential mistakes when running one model only. This work figures out the drawbacks of the current approaches towards forecast accuracy assessment of spare parts with little transaction history and proposes approaches to choose right accuracy assessment models. Experiment on data of existing company A that does aircraft maintenance was run to study the results of various forecast accuracy assessment models.

Keywords: forecast error, mean absolute percentage error, material shortage, intermittent demand, sporadic consumption.

✉Corresponding author. E-mail: *danila.larin1487@gmail.com*

1. Introduction

According to estimation of commercial aviation consultants and press in 2021 aftermarket sales of spare parts have reached almost $37 billion (Kanhere, 2025). The estimation of available stocks only airlines have is about $50 billion, and only 25% of it is used in operations (Feng et al., 2021). It means that available stock is excessive. The companies are overstored and this stresses the relevance to make accurate forecast of spare parts demand.

Demand prediction is an area where lots of research was done and a variety of forecast approaches, methods and models were used and developed. There is a big variety of them: stochastic, machine learning, personal expertise. Main task of the forecast is to give a precise projection for demand. Poor forecasts are a driver for various difficulties: excessive inventory, shortage, sales decrease, loss of goods and etc. In order to understand how precise the prediction model is, every researcher or an analyst uses accuracy assessment tools and there is quite a big variety of them: scaled errors, relative errors, percentage errors etc. And throughout the time there was quite a number of scientific articles and researches done proving that one accuracy assessment tool is better for smooth demand, another is better for only one time series, others suit more multiple time series. But in reality, when a researcher implements the tool for accuracy assessment then comes the two major problems: interpretation of the result and, as it comes from the 1st problem, sufficiency of the quantity of accuracy assessment metrics.

The problem of result interpretation is regarded from two sides: one is how to interpret the result itself and the concrete meaning received and second if the result satisfies the business goals.

The problem of accuracy assessment metrics sufficiency results from the problem of interpretation and in order to understand properly the result of forecast accuracy assessment it is required to have a complex system of accuracy measurement.

Summarizing the abovesaid the following research questions were formulated:

1. Could forecast result be misinterpreted?
2. What are the drawbacks of current accuracy assessment approaches?
3. Is there an approach that avoids misinterpretation of accuracy assessment?

This paper on the example of the data of an aircraft (AC) maintenance organization has revealed the drawbacks of the forecast accuracy assessment tools and stated the observations and proposals to take into consideration when choosing the accuracy assessment model.

This work regards only the question of accuracy assessment that is only one step out of many (demand forecast, lot size determination, order point, historical data analysis, consideration of forecast costs and so on) within forecasting process. Accuracy assessment itself couldn't be regarded and implemented separately as a stand-alone process to perform the forecast. This article is a continuation study from previous research held in the sphere of forecasting of slow-moving aviation spare parts with little transaction history. Current article focuses on forecast accuracy because this step could help a researcher to better interpret the results of the forecast, take right management decisions and better understand model behavior. Forecast models involved for accuracy assessment are the ones that are applied to forecast slow-moving spare parts. Questions concerning forecast models suitable for slow-moving spare parts, its performance comparison, demand and supply process uncertainty are covered in articles (Larin, 2021, 2022). Spare parts involved in simulation are expendable spare parts required defect rectification during heavy maintenance of an aircraft and a referred to lumpy and erratic demand type.

Even though aviation industry has lots of prescriptions from the manufacturer to ensure operations, regarded type of spares in not covered. Manufacturer by default provides information on routine expendables that are needed to be changed and its periodicity in structure repair manual (SRM), item parts catalogue (IPC) and maintenance plan. Examined expendable spare parts in this work required to fix structural damage found during heavy maintenance event are not covered by the list of proposed expendable spare parts. Required material to be changed is proposed by manufacturer (Boeing, Airbus) on request basis, especially when the damage is out of SRM scope, and delivery of such spare parts could take up to six month and even more. One day of an aircraft on the ground could cost more than $15k a day. The definition of stock list required for such cases becomes the know-how of maintenance organization both from engineering and procurement side to predict the demand.

This work is structured with five main sections, introduction and conclusion. Section 2 covers related scientific resources and literature. Based on Section 2, methodological part, forecast and accuracy assessment models are described and simulation scenario for numerical experiment on real MRO data of Company A was defined. Data framework is described in Section 4 and the simulation, run on the data, is represented in Section 5. Limitations of methodological approach and limitations of accuracy assessment model application are shown in Section 6. Conclusions provides brief summary on the experiment held and states results of this work.

2. Related works

Accuracy assessment of demand forecast is the factor of the highest importance that is needed to be taken into consideration to take right decisions. There was quite a number of papers and researchers done on this issue. The first researchers who suggested as structured accuracy assessment models that existed at the time, compared them and used on the different data scale were Makridakis and Hibon (1979). They revealed that the models perform differently under various conditions.

Contemporary vision and categorization of accuracy assessment measures were proposed by Hyndman and Koehler (2006) and stated 4 categories: scale-dependent measures, measures based on percentage errors, measures based on relative errors and measures based on scaled errors.

Among different accuracy measures the most general and popular in practical use and business was considered to be Mean Absolute Percentage Error (MAPE) (Ahlburg, 1992; Kolassa & Martin, 2011). MAPE makes possible to compare different forecast models, multi item series (not dependent on the scale) and easy in interpretation (Byrne, 2012) but still have several drawbacks: it is very sensible to outliers, could become infinite with the division on 0 and put a different penalty on positive errors than on negative errors (Koutsandreas et al., 2021; Kim & Kim, 2016). Solutions to overcome the drawbacks of MAPE were proposed in the works of Tayman et al. (1999) to avoid the influence of outliers with MAPE-R model and (Markidakis & Hibon, 2000) to get rid of 0 in denominator with symmetric MAPE model.

Summarizing the latest articles, it could be stated that there is no ideal solution for every case and depending on the business or research goal, forecast period, required service level, type of demand, type of forecast model or ease in interpretation, various accuracy assessment models could be applied and used (Pan & Montreuil, 2021). Apart from classical popular models like MAPE, Mean Absolute Scaled Error (considered to be robust for intermittent demand (Hyndman & Koehler, 2006)) lots of researchers and analysts are using additional metrics to understand the accuracy from business point of view like not on stock (NOS) model, cumulative forecast error (CFE) and part in store (PIS) accuracy (Wallstrom & Segerstedt, 2010). NOS model was proved to be effective to calculate results of forecast model and manage service level of the warehouse for slow moving items depending on the business goal (Sen, 2023). In order to eliminate disadvantages of classical accuracy assessment approaches researches are trying to adopt new approaches and bring modifications (St-Aubin & Agard, 2022).

In order to study the behavior and analyze forecast model and proper interpret it artificial increase/decrease is used. It could be used for various purposes: to verify model sensitivity analysis, stress-test of forecast models, validation and debugging of model and analysis on outliers and anomalies. Verification and validation of simulation models is described in the works (Sargent, 2013) and Schuermann (2014) where the author covers stress testing of banking forecast models.

As for forecast models themselves, that are examined on accuracy, till today simple exponential smoothing (SES) and its modifications are the mostly used approach for different type of demand: lumpy, erratic, smooth and intermittent (Yapar, 2016; Luiz de Biazzi, 2019; Svetunkov et al., 2022). Croston and bootstrap models are widely used as well because of computational ease, approaches don't require specific knowledge and suit best for intermittent demand application (Zhu, 2021). Definitely there are other approaches that were introduced recently with the application of artificial intelligence, neural network and machine learning to be applied but the difficulty is that these approaches are good for the data that has lots of transaction history and other parameters available for prediction. This was shown in the work of Kiefer et al. (2021) where Croston method outperformed machine learning applied for the forecast of intermittent data. For lumpy and erratic type of demand Synthetos-Boyal approximation (SBA), croston and bootstrap are preferable (Feng et al., 2021).

3. Simulation scenario and methodology

In order to answer research questions raised in Introduction of this work, it was decided to run numerical experiment and take data from MRO organization A with results received from demand forecast models applied and run accuracy assessment models and after check model behavior by means of stress test. This numerical experiment has a main goal to improve interpretation of model behavior and receive forecast results of slow-moving aviation spare parts with little transaction history. Explication of applied forecast and accuracy assessment models is given below in this section.

For the research pipeline three main tasks were set:

Task 1: to reveal the optimal accuracy assessment model under condition that it was not only the most precise but the most profitable solution for an airline.

Task 2: to reveal the behavior of the accuracy models after a significant artificial change of 25% in demand forecast.

Task 3: compare obtained results.

Description of the defined tasks is the following:

Task 1. The main purpose of the company is to earn profit, and this is a typical problem for operation research. In aviation industry profit of an airline depends on the availability of the company to fly and not to lose flight hours because of the aircraft on the ground due to spare parts shortage. The general view of the profit function F(*p*) that is used generally by airline companies is the following:

$$F(p) = R\sqrt{Sl} - Ce^{Sl}, \tag{1}$$

where: p – profit, expressed with money value; R – coefficient of commercial effectiveness of an airline showing the growth of revenue when the total quantity of available flight hours changes (service level of flight hours); Sl – service level, expressed as a coefficient taken from normal distribution table defined according to desired % of spare parts availability; C – coefficient of cost effectiveness of an airline showing the level of expenses when service level of available flight hours changes; F(p) is expressed in money value (USD, EUR etc).

The conditions for optimization task were stated the following way:

$$F(p) \rightarrow max; \tag{2}$$

$$Sl \rightarrow max. \tag{3}$$

It should be noted that Sl will have the optimum and it could be not 100%. In other studies, in the same case the optimal service level for available flight hours was reached at level 98%. But in this case, as far as slow-moving spares and intermittent demand involved, then it is assumed for this simulation that Sl won't be able to overcome 90%.

Having assumed that $Sl \rightarrow max$ then the following condition is set that the optimal quantity of spare parts is available on stock. In other words, not on stock (*NOS*) metric shows the percentage of spare parts availability in case it is required (or service level for spare parts) and it is defined the following way:

$$NOS = \frac{N_{CFE>0}}{N_{CFE}} \cdot 100\%; \tag{4}$$

$$CFE = \sum_{i=1}^{t} (X_t - \widehat{X}_t) Sl \rightarrow \max, \tag{5}$$

where: $N_{CFE>0}$ – the number of material shortage; N_{CFE} – the number of all forecast errors; CFE – cumulated forecast error during the period t; X_t – calculated error.

Having taken the cost function out of F(p) then this part is regarded through material consumption and the accuracy assessment was defined accordingly like an assessment of spare part demand forecast.

As for the accuracy assessment models, three widest spread were chosen that suit MAPE, Mean Absolute Scaled Error (MASE), scaled Mean Absolute Error (sMAE). This option was done in favor of scale independent measures because the data is available with different meanings and series. As well the option for these three models is described in literature review.

MAPE was chosen because it is considered to be the most popular and robust and easy from interpretational and computational point of view (Hoover, 2006). MAPE model looks the following way:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{Q_{tfact} - Q_{tforecast}}{Q_{tfact}} \right|, \tag{6}$$

where: Q_{tfact} – real observation/actual value of demand; $Q_{tforecast}$ – forecast value of demand.

MAPE is calculated for every position and then average percentage for all items is calculated.

To overcome the disadvantages of MAPE, MASE is of a use. Actually MASE is the relationship between MAE to MAE (naïve):

$$MASE = \sum_{i=1}^{n} \frac{|e_i|}{\frac{1}{n-1} \sum_{i=2}^{n} |Y_i - Y_{i-1}|}, \tag{7}$$

$$e_i = Q_{ifact} - Q_{iforecast}, \tag{8}$$

where: e_i – forecast error; Q_{ifact} – real observation/actual value; $Q_{iforecast}$ – forecast value.

Y_i – value from the prior period (for example real demand figure from the previous period (naïve)).

sMAE (scaled error) is easy to be interpreted in comparison to MAE because it has the relation to mean value of demand. sMAE has the following equation:

$$sMAE = \frac{MAE}{Mean} = \frac{T}{h} \frac{\sum_{j=1}^{h} |e_{T+j}|}{\sum_{t=1}^{T} |y_t|}, \tag{9}$$

where: e_{T+j} – error on the forecast step j; h – forecast horizon; T – number of the last observation in the time series.

sMAE error evaluation approach has almost no disadvantages. The only drawback is that it has no limits on the big values. For example, the meaning could be of 1000%.

Scale dependent measures haven't been regarded because multiple series needed to be compared. Relative errors and relative measures haven't been regarded because it has the same disadvantages like percentage errors (Chen et al., 2017).

As far as spares with little transaction history involved accuracy assessment will be done on example of Croston, SBA, Bootstrap and SES forecasting methods. These methods are among the best to forecast intermittent, lumpy and sporadic demand (Melnikova, 2019; Syntetos & Boylan, 2001; Ord et al., 2010). Applied bootstrap (nonparametric) in this work is repeated random sampling based on historical transactions of spare parts.

There are other forecast models that exist like auto regressive integrated moving average (ARIMA), aggregate-disaggregate intermittent demand approach (ADIDA), AI, but in their work, they were not used for the reasons that were described in the work (Hyndman & Athanasopoulos, 2021):

- ARIMA has a limitation of at least thirty observations during the period. Otherwise, the accuracy and reliability of the model will be low. As well ARIMA requires stationary rows. In this case of intermittent demand, it is not possible. Research could differentiate to make the row stationary but this complicates the calculations and requires from a researcher certain skill. Validation methods become non-informative.
- ADIDA is difficult in use because it is not clear to which level the temporal aggregation needs to be done. In this case only two years of observation are available with less than thirty transactions during that time. Mr. Spithourakis proved as well that high aggregation levels suffer from excessive smoothing (Spithourakis et al., 2014).
- AI and machine learning methods like XGBoost and random forest are not used in this work because it requires minimum thirty transactions to learn from. And in this case only expendables and consumables are regarded that have a very limited data and number of parameters.

Task 2. As it was indicated in literature review in order to check the behaviour of the model as one of the means to do it stress test (or forecast value change) is of a help. In order to determine percent to use for the simulation as a change to forecast, moderate option of stress test was defined (Ouabouch, 2015; Schlegel & Trent, 2015). For the experiment it was set to be 25% of increase and is considered to be moderate stress. This figure is supported as well basing on previous research (Larin, 2021) where it was defined with correlation analysis on exactly the same data of spare parts that consumption of observed inventory depends on AC quantity on maintenance. The growth of AC on maintenance by 25% was confirmed with the maintenance schedule. Severe change of 100% was not regarded because it is not general industry practice and hardly possible to be commercially accepted by MRO company that constantly runs business.

Example of stress test options and frames is represented in Table 1. In Table 1 basic value stands for the initial forecast value. Columns "moderate stress" and "severe stress" show examples of deviation frames from basic value that could be introduced to the model for stress test depending on stress type.

Table 1. Example of stress options and frames

Parameter	basic value	moderate stress	severe stress
Demand (spare parts forecast quantity)	200	160 / 240	50 / 400

So, the task is to introduce a change of 25% in demand forecast and compare the results with the initial meanings in terms of accuracy assessment. That is to say $Q_{forecast}$ will be artificially grown by 25%.

Task 3. Compare results of one forecast accuracy assessment model to several models applied. Compare results of demand forecast to forecast modified with stress test. State increase, decrease or stability of model accuracy.

Note. Within the simulation the forecast results regarded were done for the period of three month. The option in favor for this period was described in the previous work where it was revealed that forecast models show different accuracy within different time interval (Larin & Tolujevs, 2020). 3 month period was proved to be the most effective for the regarded slow-moving spare parts with little transaction history and data environment.

4. Data and applied tools description

The simulation was run on the real data of company A that provides aircraft maintenance. The data set taken was exactly the same as it was regarded in Larin (2021). Description of a data set is provided in Table 2.

Chosen data contains only expendable types of spare parts with transaction history less thirty and period of observation equals two years. All the other spare parts

Table 2. Summary of the key spare parameters used to define the scope

AC type	Maintenance type	Type of material	Category of material	Designation of material
Airbus 320	Heavy Maintenance (C, D checks)	Expendables	"Must be" "If required" "As required"	Used only at HM events

Table 3. Spare parts categorization depending on the quantity of transactions

Quantity of Transactions	Spares Category
Over 100	Fast mover
Between 30 and 100	Middle mover
Less than 30	Slow mover

from download were eliminated. Applied categorization of spare parts is represented in Table 3.

Regarding the experiment task, only two pieces of information available were used for the experiment and considered to be appropriate and minimal for demand forecast:

- item specification,
- transaction history/consumption during the time.

Available information is very limited and don't have any other parameter to precise the forecast.

Collection of data was performed by download of excel file from AMOS Software that MRO company A was using. Even though demand forecast tools for SES, SBA and Croston models were available in AMOS, bootstrap was missing. Calculations were done in exported excel file from AMOS. Correlation analysis of AC quantity to spare parts was done in StatSoft Statistica (ver. 12.5.192.7) software. Cross check of demand forecast results received in Excel file was performed in Statistica software.

5. Simulation results

Having done the simulation using the scenario and methodology above, the following received results were presented in Table 4. Table shows results received after running accuracy assessment of four popular forecast technic. In Table 4 indication "with 25% growth" means that received forecast value was grown by 25% and used for calculation. For all forecast methods MAPE accuracy assessment was reduced after the forecast value was grown by 25%. MASE gave controversy results with the decrease of forecast accuracy for Croston and SES methods and increase for SBA and Bootstrap. sMAE showed decrease for Croston, growth for Bootstrap and SES and no change for SBA. For NOS assessment all the forecast technics showed growth. Croston method showed almost no growth (3%).

For more simple visual representation the change (when 25% is added to forecast value) for every forecast

Table 4. Accuracy assessment of different forecast models for normal demand forecast and demand forecast grown by 25%

Type of forecast model	MAPE	MASE	sMAE	NOS
Croston	103%	0.98	99%	77%
Croston with 25% growth	130%	1.74	192%	80%
SBA	84%	1.78	178%	27%
SBA with 25% growth	96%	1.14	179%	38%
Bootstrap	81%	2.69	269%	30%
Bootstrap with 25% growth	93%	2.24	187%	38%
SES with	253%	4.46	447%	44%
SES with 25% growth	331%	6.69	328%	59%

method is represented in Table 5. In Table red arrow means the decrease in accuracy, green arrow shows the growth in accuracy and dash stands for no change. In this case the change of demand forecast has led to growth in accuracy: MASE for SBA and Bootstrap and sMAE for Bootstrap and SES. The growth is explained by the absolute values used in calculation of accuracy models.

Ranking wise the information is provided in Tables 6 and 7.

Table 5. Visual representation of forecast accuracy assessment change after artificial introduction of 25% growth to forecast quantity

Type of forecast model	MAPE	MASE	sMAE	NOS
Croston	↓	↓	↓	—
SBA	↓	↑	—	↑
Bootstrap	↓	↑	↑	↑
SES	↓	↓	↑	↑

Table 6. Accuracy assessment ranking depending on the model with the forecast value from prediction model

Type of forecast model	MAPE	MASE	sMAE	NOS
Croston	3	1	1	1
SBA	2	2	2	3–4
Bootstrap	1	3	3	3–4
SES	4	4	4	2

Table 7. Accuracy assessment ranking depending on the model with the forecast value grown artificially by 25% from prediction model

Type of forecast demand model	MAPE	MASE	sMAE	NOS
Croston with 25% growth	3	2	3	1
SBA with 25% growth	2	1	1	3–4
Bootstrap with 25% growth	1	3	2	3–4
SES with 25% growth	4	4	4	2

Ranking assessed the accuracy of assessment model for every forecast type. Ranking in Table 6 shows different results for MAPE, MASE and sMAE accuracy assessment models. Bootstrap outperformed other forecast models in accuracy. MASE and sMAE showed the same ranking results for every forecast model. After the growth of forecast value by 25% the ranking has changed. MAPE has shown no change but for MASE and sMAE ranking results were different. NOS ranking for outstocking hasn't changed with the introduction of artificial 25%. But as far as from business perspective and as it was set in methodology part of this paper NOS was defined to be the key parameter than the performance and ranking of accuracy assessment models has changed. With normal demand forecast Croston model has showed best results for MASE and sMAE and 3d for MAPE at the best service level result (NOS). With the introduction of 25% growth to demand forecast the ranking has changed (Table 7). Of course, in real live it is not always the issue and sometimes there is no necessity to ensure the service level. But in case there is then performance of accuracy models could change.

6. Limitations

Applied approach to check the behavior of forecast model was only examined on aviation spare parts that refer to defect rectification during AC heavy maintenance.

Interpretation of model behavior and means of model verification could vary according to inventory management goal and available data parameters. In this work stock availability was in the main focus.

MAPE accuracy assessment have limitation when real consumption during the period was 0. In case other accuracy assessment models are chosen then limitations have to be carefully studied. Limitations of accuracy assessment approaches are indicated in section 3 of this work "Simulation Scenario and Methodology". Scale dependent measures haven't been regarded because multiple series needed to be compared.

7. Conclusions

Slow-moving spare parts with little transaction history is a complex task for analyst and received forecast accuracy could have significant deviations from factual meaning. That is why slow-moving spare parts are often neglected. In aviation inventory goal and basis of economical operations are different from other industries and very dependent on spare parts availability in order to ensure flight operations. In this case even small achievement in accuracy optimization could bring company that deals with aviation spare parts significant gain and ensure robust operations.

This work has showed that introduction of change to forecast value has proved to have a positive effect in understanding of forecast result and helps a researcher properly interpret and verify the model and understand better the accuracy of applied forecast model. This in turn helps a researcher to work out more precise model, increase forecast accuracy and optimize inventory. This work has regarded aviation specificity, aviation data and inventory goals. Numerical experiment held proved effectiveness of proposed approach.

Having done the simulation, the following was summarized:

Task 1: None of the accuracy assessment models has showed the same level of results of all three forecast models chosen. Service level was the best for Croston model (NOS ranking 1) and accuracy assessment models of MASE and sMAE were ranked 1. While NOS ranked SBA model at 3–4 place has showed 2nd accuracy for all of three types of accuracy assessment (MAPE, MASE and sMAE). In other words, the most optimal solution from service level point of view doesn't mean that the analyst or researcher could rely only on one accuracy assessment model and choose the best one.

Task 2: The simulation was performed and as $Q_{forecast}$ was artificially grown by 25%. This has helped to reveal that model accuracy result behaves differently in some cases improving the result and, in some case, decreasing the accuracy. This different behavior is explained by the use of absolute value (module) in calculations of accuracy assessment. But anyway, this has brought to the change in rankings of accuracy assessment. For example, with modified demand by 25% SBA showed best ranking but the NOS parameter was the lowest.

The forecast quantity was grown artificially in this work, but as far as the company or researcher will have a condition to have a certain service level and stock availability then manipulations could be done in real.

Task 3: Statement of all observations and rules is as following:

1. The best performing forecast model doesn't guarantee the best business result. Model could show accurate result, but the company could be stocked out.
2. Forecast task should be properly set depending on the research or business needs.
3. Interpretation of the results should be done with the understanding on mathematical limitations of the accuracy assessment and demand forecast model.
4. To have a better understanding of the forecast model accuracy it is reasonable to use several accuracy assessment models for better interpretation and understanding of model behaviour. It is impossible to understand if you forecast correctly without usage of NOS.
5. Accuracy assessment models are sensible to the type of the demand forecast model chosen. Different categories of accuracy assessment methods perform differently for different types of forecasts.
6. In case of intermittent demand, the fluctuations in accuracy assessment figures are high depending on the model. So it is better to use several accuracy assessment models to compare.
7. Artificial change of forecast results for accuracy assessment purpose is of a help to understand the sensitivity and behaviour of the model.

Practical value of this work is that it could be used by a researcher or an analyst to check the behavior of forecast model and significantly improve understanding and interpretation of calculated value and take right management decisions. Applied approach showed effectiveness for slow moving inventory with little transaction history. A combination of accuracy assessment tools to assess accuracy and deviations could be used on a systematics basis and be used to assess forecast result from different angle and measurement.

As a recommendation for analyst it could be stated that depending on the research task and data availability different accuracy assessment tools could be used. Drawbacks of forecast models and accuracy assessment tools have to be carefully examined and applied. The value of stress test and its frames that will bring a change to demand forecast figure could be taken as it is represented in methodological part of this article, but depending on available data and its parameters it could be modified according to the situation and data specificity.

In order to expand the research work that was held in this work the following questions and topics have to be covered and examined in *future works*:

1. Reveal if other types of forecast error assessment work the same way and could be used as a system.
2. Reveal if for every different type of demand (smooth, lumpy, erratic, intermittent) used in this paper the simulation will work the same way.

Answers to these questions will enable analysts to more confidently select forecasting models and accuracy assessment tools.

References

Ahlburg, D. (1992). A commentary on error measures: Error measures and choice of a forecast method. *International Journal of Forecasting*, *8*(1), 99–111. https://doi.org/10.1016/0169-2070(92)90010-7

Byrne, R. (2012). Beyond traditional time-series: Using demand sensing to improve forecasts in volatile times. *The Journal of Business Forecasting*, *31*(2).

Chen, C., Twycross, J., & Garibaldi, J. (2017). A new accuracy measure based on bounded relative error for time series forecasting. *PloS One*, *12*(3), Article e0174202. https://doi.org/10.1371/journal.pone.0174202

Feng, Y.-W., Chen, J.-Y., Cheng, L., & Zhu, S.-P. (2021). Civil aircraft spare parts prediction and configuration management techniques: Review and prospect. *Advances in Mechanical Engineering*, *13*(6), 1–17. https://doi.org/10.1177/16878140211026173

Hoover, J. (2006). Measuring forecast accuracy: Omissions in today's forecasting engines and demand-planning software. *Foresight*: *The International Journal of Applied Forecasting*, *4*, 32–35.

Hyndman, R., & Athanasopoulos, G. (2021). *Forecasting*: *Principles and practice*. O Texts.

Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Kanhere, M. (2025). *Commercial aircraft aftermarket part market report 2025* (Report ID CMR119923) (Global ed.). Cognitive Market Research & Consulting.

Kiefer, D., Bauer, M., & Grimm, F. (2021). Univariate time series forecasting: Machine learning prediction of the best suitable forecast model based on time series characteristics. In A. Zimmermann, R. J. Howlett, L. C. Jain, & R. Schmidt (Eds), *Human Centred Intelligent Systems*. *KES-HCIS 2021*. *Smart Innovation*, *Systems and Technologies* (Vol. 244, pp. 152–162). Springer. https://doi.org/10.1007/978-981-16-3264-8_15

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kolassa, S., & Martin, R. (2011). Percentage error can ruin your day (and rolling the dice shows how). *The International Journal of Applied Forecasting*, *23*, 21–27.

Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2021). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, *73*(5), 937–954. https://doi.org/10.1080/01605682.2021.1892464

Larin, D. (2022). Optimizing the level of slow moving aviation spare parts inventory with economic order quantity model using bootstrap method. In I. Kabashkin, I. Yatskiv, & O. Prentkovskis (Eds), *Reliability and Statistics in Transportation and Communication*. *RelStat 2021*. *Lecture Notes in Networks and Systems* (Vol. 410, pp. 237–245). Springer. https://doi.org/10.1007/978-3-030-96196-1_21

Larin, D. (2021). Precising the accuracy of aviation spare parts forecast. In I. Kabashkin, I. Yatskiv, & O. Prentkovskis (Eds), *Reliability and statistics in transportation and communication*. RelStat 2020. Lecture Notes in Networks and Systems (Vol. 195, pp. 205–212). Springer. https://doi.org/10.1007/978-3-030-68476-1_19

Larin, D., & Tolujevs, J. (2020). Defing the proper model for aviation spare parts forecast. In *Reliability and statistics in transportation and communication*. *Lecture Notes in Networks and Systems* (pp. 71–79). Springer. https://doi.org/10.1007/978-3-030-44610-9_8

Luiz de Biazzi, J. (2019). Exponential smoothing for intermittent demand with demand basis updated more frequently than seasonality factors. *Gestão* & *Produção*, *26*(1), Article e1297. https://doi.org/10.1590/0104-530x1297-19

Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society, 142,* 97–145. https://doi.org/10.2307/2345077

Markidakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Melnikova, O. (2019). Crosoton's method in the modification of Syntetos and Boylan for forecasting inventories. Scientific result. *Economic Research*, *5*, 3–9. https://doi.org/10.18413/2409-1634-2019-5-2-0-1

Ord, K., Beaumont, A., & Snyder, R. (2010). *Forecasting the intermittent demand* (Working papers 2010-003). Research Program on Forecasting.

Ouabouch, L. (2015). Overview on supply chain resilience. *Materials Management Review*, *July*, 16–18.

Pan, X., & Montreuil, B. (2021). Assessing multi-purpose forecasting accuracy. In *Proceedings of the 2021 IISE Annual Conference* (pp. 728–733). ProQuest.

Sargent, R. G. (2013). An introduction to verification and validation of simulation models. In *Proceedings of the 2013 Winter Simulation Conference* (*WSC*) (pp. 321–327). IEEE. https://doi.org/10.1109/WSC.2013.6721430

Schlegel, G. L., & Trent, R. J. (2015). *Supply chain risk management*. CRC Press. https://doi.org/10.1201/b17531

Schuermann, T. (2014). Stress testing banks. *International Journal of Forecasting*, *30*(3), 717–728. https://doi.org/10.1016/j.ijforecast.2013.10.003

Sen, E. G. (2023). *Inventory management and plannning of slow-moving spare parts* [Master Thesis, Eindhoven University of Logistics]. Eidhoven University of Technology.

Spithourakis, G., Petropoulus, F., Nikolopoulos, K., & Assimakopoulos, V. (2014). A systemic view of ADIDA framework. *IMA Journal of Management Mathematics*, *25*(2), 125–137. https://doi.org/10.1093/imaman/dps031

St-Aubin, P., & Agard, B. (2022). Precision and reliability of forecast performance metrics. *Forecasting*, *4*(4), 882–903. https://doi.org/10.3390/forecast4040048

Svetunkov, I., Kourentzes, N., & Ord, K. (2022). *Complex exponential smoothing*. *Naval Research Logistics*, *69*(8), 1108–1123. https://doi.org/10.1002/nav.22074

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

Tayman, J., Swanson, D., & Barr, C. (1999). In search of the ideal measure of accuracy for subnational demographic forecasts. *Population Research and Policy Review*, *18*(5), 387–410. https://doi.org/10.1023/A:1006317430570

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

Yapar, G. (2016). Modified simple exponential smoothing. *Hacettepe Journal of Mathematics and Statistics*, *47*(143), 1–1. https://doi.org/10.15672/HJMS.201614320580

Zhu, Z. (2021). *Spare parts demand forecasting and inventory management*: *Contributions to intermittent demand forecasting*, *installed base information and shutdown maintenance* [Doctoral Thesis, Erasmus University Rotterdam]. Erasmus Universiteit Rotterdam (EUR).