# The influence of incorrect repairs on the unpredictability of the demand of spare parts in aviation

L.M. Heijenrath

TUDelft

# The influence of incorrect repairs on the unpredictability of the demand of spare parts in aviation

by

L.M. Heijenrath

to obtain the degree of

**Master of Science**
in Aerospace Engineering,
specialization Air Transport & Operations,

at the Delft University of Technology,
to be defended publicly on Monday January 25, 2020 at 9:00 AM.



| | | |
|---|---|---|
| Student number: | 4296729 | |
| Thesis committee: | Dr. ir. W.J.C. Verhagen, | Royal Melbourne Institute of Technology, supervisor |
| | Dr. M.D. Pavel, | TU Delft, chair |
| | Ir. J. Sinke, | TU Delft, additional member |





TU Delft

# Preface

The past seven and a half years brought me a lot. The time at student association K.S.V. Sanctus Virgilius has been a valuable addition to the time I spent studying for assignments and exams at TU Delft. I'm grateful for the chances I had to develop myself on personal skills during my time as sales manager at the Commissie Externe Betrekkingen and my internship at ING Bank. I'm looking forward to the first step in my professional career, where I'll be using my knowledge gained during the bachelor program of Civil Engineering and the master program of Air Transport & Operations at the faculty of Aerospace Engineering.

Foremost, I would like to thank my daily supervisor, Wim, for his guidance through this project. As we were both hoping for a collaboration on shorter distance, the current pandemic did not allow us to ever meet in person. Moreover, we've had contact only via the internet, as we've been more than 165000 km separated from each other. Nevertheless, I'm satisfied with the way of collaboration we've had. From day 1, Wim has inspired me to go that extra mile and dive even deeper into literature in order to achieve the maximum.

Finally, I would like to thank all my friends and family that supported me during my thesis. A special thanks to Wouter, Jacco, Mathijs, Louka, Joop, Mariet and Luna.

My time at TU Delft has come to an end, closing a chapter in my life. I'm thrilled for the next one.

*L.M. Heijenrath*
*Rotterdam, December 2020*

# List of Symbols

| | |
|---|---|
| $\hat{x}_t$ | forecast at the beginning of period $t$ of demand in period $t$ |
| $\hat{x}_n(j)$ | forecast at the beginning of origin $n$ of demand in period $n+j$ |
| $d_t$ | actual demand in period $t$ |
| $d_{n+j}$ | actual demand in period $n+j$ |
| $h$ | prediction horizon |
| $\hat{k}_t$ | forecast in period $t$ of number of periods between consecutive positive demands |
| $k_t$ | actual number of periods since the last positive demand at the beginning of period $t$ |
| $\hat{s}_t$ | forecast of demand in period $t$, given that this is a positive number |
| $\hat{x}_t^c$ | forecast of demand in period $t$ used for components of type $c$ |
| ${d_t}^c$ | actual demand in period $t$ used for component type $c$ |
| $\hat{z}_t^c$ | forecast of number of repairs of components of type $c$ in period $t$ |
| ${z_c}^t$ | actual number of repairs of components of type $c$ in period $t$ |
| $\hat{a}_t^c$ | forecast of the average number of spare parts used in period $t$ for the repair of a component of type $c$ |
| $\alpha$, $\beta$ | smoothing constants ($0 \leq \alpha, \beta \leq 1$) |
| $G$ | Number of subsidiary failures corresponding to a single primary failure |
| $S$ | Unconditional number of subsidiary failures |
| $Z_i$ | Time between primary failure $i$ and primary failure $i-1$ |
| ${Y_i}^{(j)}$ | Time between the $i$th failure after the $j$th primary failure and the previous failure |
| $H(t)$ | Expected number of subsidiary failures from a single primary failure in an interval of length $t$ |
| $F_k(t)$ | cdf of the time of the $k$th primary failure |
| $f_k(t)$ | pdf of the time of the $k$th primary failure |
| $N^{(i)}(t)$ | Number of failures in $[0,t]$ due to the $i$th subsidiary process |
| $\lambda$ | Expected value and variance of the stochast X for a Poisson distribution |

# List of Abbreviations

| | |
|---|---|
| ADI | Average Demand Interval. vi, vii, 20, 23, 28, 37, 38, 55–57, 60, 62–65, 68, 70, 71, 74–77, 79, 80, 82–85, 87, 88 |
| ANOVA | Analysis of Variance. 38 |
| ATA | Air Transport Association of America. vi, 22, 37, 41–55, 60, 62, 63, 85 |
| BPP | Branching Poisson Process. 34, 35, 45, 47, 48, 52, 54, 55 |
| cdf | Cumulative distribution function. 38 |
| CI | Confidence Interval. 51, 74 |
| CV | Coefficient of Variance. 23, 28 |
| $CV^2$ | squared Coefficient of Variance. vi, vii, 20, 37, 38, 55–57, 60, 62–65, 68, 70, 71, 74–85, 87, 88 |
| FAA | Federal Aviation Administration. 52 |
| HPP | Homogeneous Poisson Process. 50 |
| KS | Kolmogorov-Smirnov. 38 |
| KW | Kruskal-Wallis. vi, 38, 39, 51, 52, 60, 63, 67, 74, 80, 86 |
| LCC | Low-Cost Carrier. 57 |
| MANOVA | Multivariate Analysis of Variance. 38 |
| MRO | Maintenance, Repair and Overhaul. 21, 23, 24, 29, 34, 36, 40, 58, 59, 76, 87, 88 |
| MTBF | Mean Time Between Failure. 55, 56, 58 |
| MTBR | Mean Time Between Removal. 41, 45, 47, 49, 50 |
| NHPP | Non-Homogeneous Poisson Process. 50 |
| pdf | Probability Density Function. 42 |
| RMST | Restricted Mean Survival Time. 56 |

# List of Figures

# List of Tables

# Contents

# Introduction

In aviation, it is crucial for airlines to maintain their fleet in an airworthy state. Therefore, each individual aircraft is required to meet a high level of technical standards. Hence, maintenance is required to keep the aircraft on the required technical level. As components are replaced, demand of new components is generated. In order to minimize the downtime of an aircraft, Maintenance, Repair and Overhaul (MRO) providers aim to have the spare parts available in the inventory. Many studies have showed that the spare parts demands in aviation are intermittent or lumpy. The problem of these types of demand patterns for the industry comes in the unpredictability of the demands. The result of these detrimental patterns of the demands is that the stocks of components have to be larger than desirable, as companies keep stock buffers in order to ensure the availability of parts, leading to increased holding costs and occupation of extra space.
Current studies do not provide further understanding of the generation of demand due to the use of time-series techniques, do not take multiple demand drivers into account, and assume that the state of installed components is always in a state "as-good-as-new". One of the downside of the use of time-series techniques is that no further understanding of the generation of the demand is provided. Hence, the understanding of the problem of demand forecasting is limited. Therefore, this study is the first to take different demand drivers into account combined with the consideration of the effect of incorrect repairs on the demand patterns. As demand drivers, the fleet size, environmental conditions and component commonality strategies are considered.
The main research question is formulated as:

**What is the influence of incorrect repairs on the predictability of the demand of spare parts in aviation?**

The build-up of the document is presented here. First, the technical paper is presented, starting on page xi. Next, an overview of the investigated literature is provided in chapter 1. Then, the methodology is shortly explained in chapter 2. This is followed by the preprocessing of the data, elaborated in chapter 3. Next, the implementation of the model is presented in chapter 4. Chapter 5 provide the reader with an overview of the simulations. Next, the evaluation and discussion on the results of the model are presented in chapter 6. Finally, the conclusions and recommendation are elaborated in chapter 7.

# Technical paper

# The influence of improving repair quality on the unpredictability of the demand of spare parts in aviation

L.M. Heijenrath

*Section Air Transport & Operations, Department Control and Operations, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands*

**Abstract**

This study provides new insights in the problem of lumpy aircraft spare parts demands by incorporating new drivers that have an impact on the failure patterns of aircraft components. The study introduces a model and presents corresponding results that obtains component failure characteristics based on data from an aircraft manufacturer. A Monte Carlo simulation technique is used to take different repair qualities, fleet sizes, environmental conditions and shared component pool strategies into account. The outcome is evaluated to capture the impact of these parameters. Based on the occurring patterns of the failures, the demand patterns can be inferred. The study confirms the conclusion from previous research that the fleet size is the main contributor to the unpredictability of the demands of spare parts, but notes that this conclusion is not always usable in practice, as practical limitations regarding the extension of fleets are in play. The study concludes that an improvement of the repair quality is beneficial for the variance of the demand and the total amount of failures over time.

## 1. Introduction

In aviation, it is crucial for airlines to maintain their fleet in an airworthy state. Therefore, each individual aircraft is required to meet a high level of technical standards. Hence, maintenance is required to keep the aircraft on the required technical level. As components are replaced, demand of new components is generated. In order to minimize the downtime of an aircraft, Maintenance, Repair and Overhaul (MRO) providers aim to have the spare parts available in the inventory. Many studies have showed that the spare parts demands in aviation are intermittent or lumpy [1],[2],[3]. The problem of these types of demand patterns for the industry comes in the unpredictability of the demands. The result of these detrimental patterns of the demands is that the stocks of components have to be larger than desirable, as companies keep stock buffers in order to ensure the availability of parts, leading to increased holding costs and occupation of extra space.

The study of Ghobbar and Friend [4] concludes that airline operators can improve their forecasts by identifying which drivers induce the intermittent and lumpy behaviour of the demand. Many studies have contributed to the improvement of the understanding of aircraft spare parts demands. Regattieri et al. [2] analyses the accuracy of twenty time-series forecasting techniques and shows that airlines usually do not use these techniques. Instead, in-house experience or suggestions of the suppliers of the components are used.

Current studies do not provide further understanding of the generation of demand due to the use of time-series techniques, do not take multiple demand drivers into account, and assume that the state of installed components is always in a state ”as-good-as-new”.

One of the downside of the use of time-series techniques is that no further understanding of the generation of the demand is provided. As the recent study of Van der Auweraer et al. [5] noted, installed base information can be used to forecast upcoming demand of spare parts. The study of Lowas and Ciarallo [6] uncovers some reasons for the unpredictable behaviour of spare parts demands. The most significant single factor impact is found to be the size of the fleet of aircraft. It is concluded that smaller fleets have higher values for the Coefficient of Variance (CV) and Average Demand Interval (ADI) of the demand compared to large fleets. The authors of the study recommended further study to better understand demand generation drivers, as only some were tested.

One commonly made assumption is that of the state of the component when installed on an aircraft. Stud-

ies using the expected lifetime of components often neglect the fact that errors in the repair process occur and hence repairable components are not restored in an as-good-as-new state. As maintenance personnel face a high level of time pressure and environmental circumstances in the industry, errors in the process will occur. Research has shown that in at least 39% of the cases, a maintenance error is related to installation errors or incomplete repairs [7]. Broken components that are placed back into operation result in subsequent failures due to the broken state of the component [8]. Hence leading to a peak in failures, resulting in a peak in the demand for new spare parts. This drives the lumpiness behaviour of the demands of the spare parts.
Therefore, this study is the first to take different demand drivers into account combined with the consideration of the effect of incorrect repairs on the demand patterns. As demand drivers, the fleet size, environmental conditions and different component commonality strategies are considered.
Section 2 provides the reader with the current state-of-the-arts. Next, section 3 presents the data and explains the working of the model that are used in this study. Section 4 presents the results of the study. Next, section 5 discusses the validity of the results. Finally, section 6.1 presents the conclusions and recommendations for future studies.

## 2. Theoretical background

In this section, relevant research regarding forecasting of spare parts demands and dependent failures is provided.

### 2.1. Demand of spare parts

For MRO providers, the behaviour of the failures of components over time is crucial in order to predict what parts are necessary at a certain point in time. When the demand of spare parts is more predictable, hence leading to smaller errors in forecasts, the margins of additional spare parts in the inventory can be lowered. Currently, the aviation industry carriers around €30 billion yearly to stock spare parts to keep aircraft airworthy [9]. To classify the patterns of the demands of spare parts, two metrics are used; the ADI and the $CV^2$. The demand is set to be intermittent for patterns that have an ADI > 1.32 and a $CV^2 < 0.49$, and lumpy for ADI > 1.32 and $CV^2 > 0.49$. See figure 1 for an overview of the different demand patterns. As stated in section 1, many studies found that the demands of spare parts in aviation tend to be intermittent or lumpy.

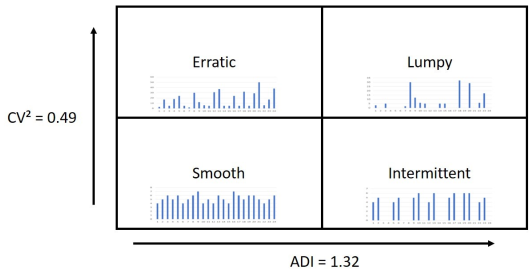



Figure 1: Overview of demand patterns based on ADI and $CV^2$ [10]

The study of Boone et al. [11], which focuses on critical challenges of inventory management in service parts supply, concluded that inaccuracy of spare parts forecasts was the only unanimously selected challenge by panel members in this study. Furthermore, it was ranked second as the most difficult challenge facing spare parts inventory managers. Hence, the intermittent and lumpy behaviour results in a problem for the inventory management of MRO providers.
Over the years, the methods used for forecasting have further developed. In the earlier days, the main focus was to improve the accuracy of the forecast by the use of time-series methods. Time-series methods range from relative simple methods (Naive forecast, where the forecast for the coming period equals the actual demand of the current period), to more advanced methods, such as Croston's method [12]. In this method, the forecasts for the demand size and the demand interval are treated separately in order to minimize the error of the forecast. In 2005, Syntetos and Boylan [13] improved the method developed by Croston by the discovery of the bias of Croston's method. The research came up with a deflation of the forecast factor by a factor $\left(1 - \frac{\alpha}{2}\right)$. Another improvement of the Croston's method is realized by the research of Romeijnders et al. [14]. This research proved that two-step forecasting methods are more accurate than the benchmarked technique. The method showed a 20% reduction of forecasts errors. Hence, current literature is still improving the results of some time-series techniques.
However, current research tends to focus more on causal methods. As multiple factors influence the demand of different aircraft components, it is crucial for deeper understanding to identify these causalities and quantify their impact. The demand is generated when a component is removed from service, mostly due to failure or a strategic decision to replace the component. On aircraft level, the research of Ghobbar and Friend [4] showed the effect of the aircraft utilization and flight hours on the failures of components, as wear and tear of com-

ponents increases with increasing utilization and flying hours, and therefore the demand rate of spare parts increases. On fleet level, Lowas and Ciarallo [6] proved that small fleets have higher demand $CV^2$ and higher ADI than large fleets and that higher buy periods tend to rise the $CV^2$ and ADI as well. In both studies, a correlation between the spare parts demands and certain demand drivers is substantiated. However, both studies mention that there is still little understanding of the causes of fluctuations in spare part demands. According to the best knowledge of the author, all studies tend to leave the correctness of maintenance of the components untouched. The state of the component on installation is crucial for the lifetime and therefore moment of failure of the component.

As the lumpiness of the demand of spare parts is one of the keys to the limitation for improvement of predictability of the demand, better understanding of drivers for lumpy behaviour is needed. Incorrect repairs, leading to components that do not function when put into service, trigger subsequent failures and thus a peak in failures for a group of components. Hence, this results in a peak in the demand for the specific spare parts.

### 2.2. Dependent failures

Despite the goal of MRO providers and airlines to prevent failure of components from happening while extending their lifetime as long as possible within safety margins, failures do occur in aviation. Incorrect repairs cause components to remain in broken state, having a crucial impact on the functioning of the component and interdependent components. When placed back in to service, the broken component will not function as desired and hence, interdependent components will be affected. An example of this is the failure of components due to overloading, as the workload will be fully dependent on the working components, since the broken component can not take any load. These phenomenons are present in the current aviation industry [7].

A process that can account for these incorrect repairs is the Branching Poisson Process [15]. The model finds its origin in 1963, as it is used to model the stop-and-go motion of vehicles as a result of a slowly moving vehicle in front. This phenomenon is similar to a single failure causing a number of subsequent failures. According to the theory, a primary failure might trigger subsequent failures, dependent on the correctness of the repair. Here, $r$ is the probability that a repair is not done correctly. Hence, $1 - r$ represents the chance that the repair is done correctly. In the case that the repair is done incorrectly, the incorrect repair will spawn a finite renewal process of subsidiary failures. The amount of subsidiary failures is a discrete random variable. Given the fact that the time of the first primary failure $(Z_1) = z$, the expected number of failures in the interval $[0, t]$ can be expressed as $H(t - z)$. Then, the contribution of the first event in the subsidiary process for the expected events can mathematically be described as:

$$\begin{aligned} E\left[N^{(1)}(t)\right] &= E\left\{E\left[N^{(1)}(t)|Z_1\right]\right\} \\ &= \int_0^t E\left[N^{(1)}(t)|Z_1 = z\right] f_1(z)\, dz \\ &= \int_0^t H(t-z) f_1(z)\, dz \end{aligned} \tag{1}$$

where $f_1(t)$ represents the probability density function of the primary events. In the case of a Homogeneous Poisson Process (HPP), $f_k(t) = \lambda \exp(-\lambda t)$ The same representation can be made for the expected number of failures $N^{(k)}(t)$ in $[0, t]$ due to the $k$th subsidiary process:

$$E\left[N^{(k)}(t)\right] = \int_0^t H(t-z) f_k(z)\, dz \tag{2}$$

That stated, the expected number of failures of any type in $[0, t]$, $\Lambda(t)$, can be expressed as the sum of the expected number of primary failures in $[0, t]$ and the expected number of subsidiary failures. Leading to:

$$\begin{aligned} \Lambda(t) &= E\left[N(t)\right] \\ &= \Lambda_Z(t) + \int_0^t H(t-z)\, \Lambda_Z^{'}(z)\, dz \end{aligned} \tag{3}$$

In the case that the primary process is a Homogeneous Poisson Process:

$$\begin{aligned} \Lambda(t) &= \lambda t + \int_0^t H(t-z)\, \lambda\, dz \\ &= \lambda\left[t + \int_0^t H(t-z)\, dz\right] \end{aligned} \tag{4}$$

Where $\lambda$ is the rate parameter of the HPP. Lastly, the Rate of Occurrence of Failure (ROCOF), $\mu(t)$, can be stated as the derivative of equation 4:

$$\begin{aligned} \mu(t) &= \lambda + \lambda \int_0^t \frac{\partial}{\partial t} H(t-z)\, dz \\ &= \lambda\left\{1 + \left[-H(t-z)\right]_{z=0}^{z=t} + H(0)\right\} \\ &= \lambda\left[1 + H(t)\right] \end{aligned} \tag{5}$$

For the full derivation of these equations, the interested reader is referred to the work of Rigdon and Basu [15]. A graphical overview of the occurrence of primary and

subsidiary failures is provided in figure 2. As can be seen from the figure, the complete process is the superposition of the primary and subsidiary events of failures. Here, the assumption is made that the two types of events are indistinguishable.
The downside of the BPP method is that the parameter estimation is difficult to perform.

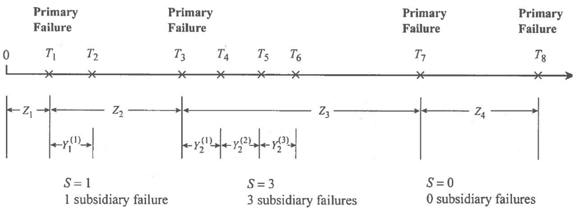



Figure 2: Visual presentation of the BPP [15]

## 3. Experimental framework

In this section, the model formulation and implementation, the input data that is used and the simulation parameters are further elaborated. Section 3.1 provides the explanation of the working of the model and section 3.2 motivates the chosen values for the varying parameters. Section 3.3 provides insight in the data set and the way how input data for the model is generated from this.

### *3.1. Methods*

The model should provide a quantitative answer to the impact of the different repair qualities on the demand of the failures. Therefore, it is decided to capture the impact of changing repair qualities in the ADI, $CV^2$ and the amount of failures. The first two metrics cover the predictability of the demand, the final metric is used for the quantitative quality of the demand.
In order to incorporate these incorrect repairs, the parameter $r$, the chance of incorrect repairs, is implemented in the model. This parameter influences the discrete random variable that represents the spawn of subsequent failures. Hence, when an incorrect repair takes place and the component is placed back into service, the amount of failures during a relative small timespan will peak due to the amount of subsequent failures. The influence of this parameter on the $CV^2$ and ADI is the core attribute of this model. Although it could be argued that the value of $r$ might change over time, hence a function $r(t)$ might be present, this is not incorporated in this study.
Besides the main parameter $r$, additional parameters that influence the pattern of the demand of spare parts are taken into account as an extension. The work of Lowas and Ciarallo [6] showed the influence of the fleet size on the demand patterns. Thijssens and Verhagen [16] showed that the environmental conditions impact the reliability of components for multiple different reasons. Air pollutants, salinity and the salt content in the atmosphere all have an impact on the corrosion process of components. Besides a natural reference climate (temperate), humid and desert climates are taken into account as well, both affecting the Mean Time Between Failure (MTBF). The impact of the incorrect repairs in combination with these other varying circumstances provides a wider view on the impact in general.
For every aircraft in the fleet, failures at the selected ATA locations (see section 3.2) are simulated according to their corresponding failure rate $\lambda$, obtained from the analysis of the data. Based on the number of primary removals that contain subsequent removals in the data set, an estimate of the probability of an incorrect repair $r$ can be made for the specific component location. Subsequently, possible subsequent failures are simulated.
Next, the results of the individual aircraft are summed, resulting in the sum of the failures over time for every ATA location that is selected to be part of the model. This is done for multiple combinations of parameters. The values for the ADI and $CV^2$ are stored. The abovementioned metrics describe the predictability of the failures over time, but do not provide an answer to the quantity of failures. Therefore, this metric is added to the results as well, in order to capture both the behaviour as the sum of the failures. A visualization of the working of the model is provided in figure 3. As randomness is in play with the occurrences of incorrect repairs and the distribution of subsidiary failures, a total of 50 iterations of the model are performed before the analysis of the results. The final results are based on the average of the individual iterations.
The outcomes of the model are in four variants. Each discussed briefly below.

- Variant 1 - Base: In this variant, the level of repair is the only varying parameter. Only temperate environmental conditions are taken into account. All fleet sizes are taken into account, but no distinction is made in the presentation of the results. No increased component commonality across different aircraft is taken into consideration.

- Variant 2 - Incorporation of varying fleet sizes: This variant uses the same set of results as Vari-

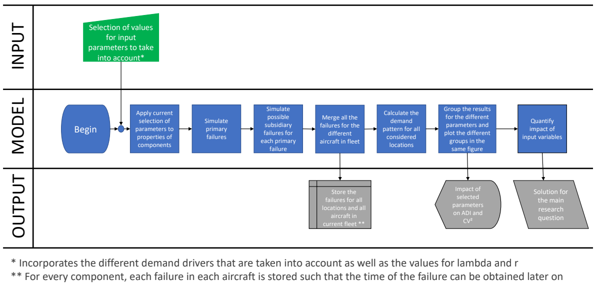



Figure 3: Flowchart of the working of the model

ant 1, but a distinction between the different fleet sizes is made in the presentation of the results.

- Variant 3 - Incorporation of varying fleet sizes and environmental conditions: As different environmental conditions influence the effect of the expected lifetime of components, this will result in varying values for the different $\lambda$s. Here, the results of the humid and desert environments is also taken into account.

- Variant 4: Incorporation of varying fleet sizes, environmental conditions and component commonality strategies. As flag carriers tend to diversify their fleets [17], MRO providers have to deal with different aircraft types. Currently, the component commonality across the aircraft types is limited. However, recent research of Zhang et al. [18] showed promising results regarding the potential gains with respect to costs when the component commonality is increased. Hence, this variant investigates the effect of the increment of component commonality, given the fact that different performing aircraft use the same sets of components. A deviation of 20% for the performance of different types of aircraft is used.

See Appendix C for the pseudocode of the model.

### 3.2. *Parameters*

In order to obtain the influence of different scenarios, multiple parameter values have to be taken into consideration.
The main goal of this research is to reveal the impact of the quality of the repair process for components that are placed back into the aircraft on the $CV^2$ and ADI of the demands of the spare parts. As the initial values of $r$ are retrieved from the data analysis of the dataset, these values are used as reference values (Normal scenario). Scenarios with 100% increased (Worse scenario), 50% decreased (Improved scenario) and 100% decreased (Perfect scenario) values for $r$ are tested. The first mentioned alteration of $r$ represents a scenario in which the amount of incorrect repaired components that is placed back into service is twice as high as the reference scenario.
The second alteration represents the scenario where the chances on an incorrect repair are decreased by 50%. Therefore, less incorrect repaired components are placed back into service.
The latest option represents the scenario where no incorrect repaired components are placed back onto the aircraft, and thus that all components that are placed back function properly.
The work of Thijssens and Verhagen [16] showed the impact of three environmental factors to the Restricted Mean Survival Time (RMST) of components in aviation. The RMST is equal to the mean survival time, except that the RMST is restricted in a time range $[0, \theta]$, to avoid the negative influences of the poorly determined right tail of a survival curve during estimation [19].

In this study, the impact of the environmental factors is directly related to the MTBF of components by the numerical factor that is provided in table 1.
For every aircraft considered in the analysis, the airline can be traced back via the external organization code. In this way, the dominant environmental conditions at the main hub of the airline can be applied and the values for the specific aircraft can be adjusted. The study of

| **Natural climate** | **MTBF-ratio** |
|---|---|
| Temperate | 1 |
| Humid | 0.62 |
| Desert | 0.73 |

Table 1: The effect of environmental factors on the MTBF [16]

Lowas and Ciarallo [6] provided insights into reasons for lumpy spare parts demands. The study found that the parameter with the largest impact of the lumpiness of the demands of spare parts was the fleet size. In order to validate this finding and to extend its scope, it is tested in this research as well. As the referred study clearly motivates the range of selected values of the fleet size, this is not further motivated in this study, as the same range of values are chosen for the model.
Finally, the increment of component commonality across different aircraft types is tested. Here, it is assumed that different aircraft types perform differently, resulting in variations in the average operating time

of components. A deviation of 20% is assumed. The size of the deviation itself is not crucial, as the outcome will be directly compared to variant where no different aircraft are considered. If significant differences occur, it becomes a stepping stone for motivation of future research.

| Variable | Tested values | Number of steps |
|---|---|---|
| **$r$ factor** | 0.0, 0.5, 1, 2 | 4 |
| **Environmental factors** | Natural, Humid, Desert | 3 |
| **Fleet size** | 8, 16, 32, 64, 96, 128, (256), (512) | 6 (8) |
| **Component commonality** | Not present, Present | 2 |

Table 2: Monte Carlo simulation parameters

### 3.3. Materials

In order to provide the model with the right input parameters based on the behaviour of the failures of components on aircraft, data from an anonymous aircraft manufacturer is used. The data comprises removal data spanning over multiple decades. For each data point, the part number, date, aircraft type, ATA location, serial number of the aircraft and operator are used in this research. A selection of the components is made in order to limit the scope. An overview of this analysis can be found in Appendix A. This limits the scope down to components in the following eight ATA chapters: 23 (Communications), 24 (Electrical Power), 27 (Flight Controls), 28 (Fuel), 29 (Hydraulic Power), 32 (Landing gear), 34 (Navigation) and 77 (Engine indicating). Based on the operator, the environmental conditions can be determined for every aircraft in the data set. This has a direct impact on the lifetime of components and therefore influences the Mean Time Between Failure (MTBF), hence the spawn rate of primary failures [16].

From the selected data, primary and subsidiary removals could be identified. In this analysis, a subsidiary removal is defined as a removal that occurs within fourteen days of the primary removal. Here, it is assumed that components are interdependent if and only if they are located in the same ATA chapter. With this information, the spawn rate of primary removals could be made for every ATA location on every aircraft. Adjustments are made with respect to the environmental conditions, in order to be able to correctly quantify the effect of varying environmental conditions later on in the model. Having an overview of the primary and subsidiary removals, the chance of an incorrect repair for each ATA location can be made by reviewing the amount of primary failures that incorporate subsidiary removals. For every location, composition of subsequent failures is reviewed. Through this, the model can vary the offset of a primary failure based on the distribution of the offset from the data. In the model, the amount of subsidiary failures are randomly distributed over the fourteen days after the day of the primary failure.

Next, for all aircraft and ATA locations, a check has to be made regarding the homogeneity of the primary removal rates of the components. The results of this test are presented in table 3. It can be concluded that the spawn rate of primary removals is constant in most of the cases. Therefore, a Homogeneous Poisson Process can be used to simulate these removals. Furthermore, no significant difference was found among the performance of the different aircraft in the data. Hence, the results could be aggregated.

| No Trend | Increasing Trend | Decreasing Trend |
|---|---|---|
| 92.86% | 2.44% | 4.70% |

Table 3: Results of the trend analysis for the failure rate of the primary removals

## 4. Results

The results are presented in order of the four different variants of the model. Visualizations of the results are provided, although only a small selection of all visualizations are provided in this paper. In the figures, each data point represents the average demand characteristics (ADI and $CV^2$) of a unique combination of varying parameters. For each variant, the impact of an improvement in the repair quality is quantified. The motivation presenting the deviations on an improvement finds its origin in the desire of an MRO provider to minimize the errors in the repair quality and to strive for improvement. Therefore, MRO providers can use the outcome of this study to quantify the effect of an improved repair quality on the ADI, $CV^2$ and total amount of failures.

The results of the statistical tests are provided in Appendix B. The results are presented in the form of $p$-values of the Mann-Whitney $U$-test and the Kruskal-Wallis $H$-test [20], [21], [22]. Values that are not significantly different according to these tests ($p > 0.05$), are marked with an asterisk in the tables.

### 4.1. Results of variant 1

The visual representation of the results in figure 4 do not directly provide the reader with a discernible performance difference of the different repair qualities. Tables 4 and 5 provide the reader with a quantitative comparison. Here, comparison is made between the current repair quality (left column) and the desired repair quality (top row). The number provides the ratio of the average value of the metric of the desired repair quality and the current repair quality. Table 6 provides the reader with the total amount of failures of the different repair qualities. Here, it can be seen that an improved repair quality lowers the amount of failures. An improvement of

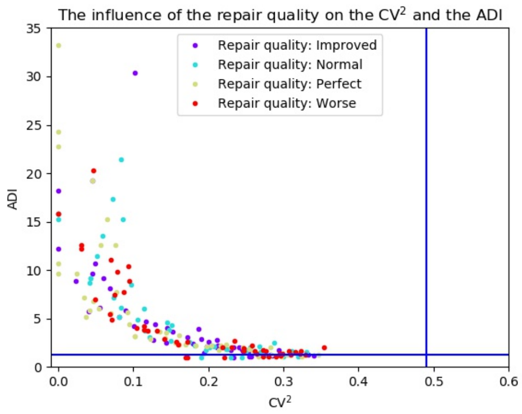



Figure 4: Visual results of variant 1

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 1.000 | 1.151 | 1.278 | 1.462 |
| **Normal** | 0.869 | 1.000 | 1.126 | 1.283 |
| **Improved** | 0.782 | 0.888 | 1.000 | 1.152 |
| **Perfect** | 0.684 | 0.779 | 0.868 | 1.000 |

Table 4: V1: Overview of the influence on the average ADI for changing repair qualities

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 1.000 | 0.970 | 0.916 | 0.842 |
| **Normal** | 1.031 | 1.000 | 1.011* | 0.926 |
| **Improved** | 1.091 | 0.989* | 1.000 | 0.961 |
| **Perfect** | 1.187 | 1.079 | 1.041 | 1.000 |

Table 5: V1: Overview of the influence on the average $CV^2$ for changing repair qualities

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Failures** | 18329 | 15115 | 12763 | 10200 |

Table 6: V1: Average amount of failures for the different repair qualities per iteration

the repair quality from "Worse" to "Normal" increases the ADI by 15.1%, decreases the $CV^2$ by 3.0% and decreases the total amount of failures by 17.5%.

An improvement from "Normal" to "Improved" repair quality leads to a increase of the ADI by 12.6%, an increase of the $CV^2$ by 1.1%, and a reduction of the total amount of failures of 15.6%. Hence, this improvement has a positive effect on the total amount of failures, but deteriorates the predictability of the failures over time. The improvement of the repair quality from the "Improved" to the "Perfect" repair quality situation leads to an improvement of the ADI by 15.2%, a reduction of the $CV^2$ by 3.9%, and a reduction of the total amount of failures of 20.1%.

### *4.2. Results of variant 2*

The visual results of figure 5 show the reader an increased fleet size lowers the ADI and increases the $CV^2$. The results of the varying repair quality for all fleet sizes are provided in table 7. Generally, it can be concluded

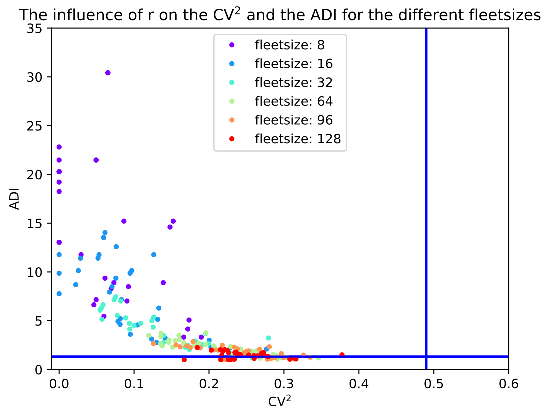



Figure 5: Visual results of variant 2

| Improvement | Fleet size | ADI | $CV^2$ | Failures |
|---|---|---|---|---|
| Worse - Normal | 8 | +18.4% | -17.8% | -19.9% |
| | 16 | +13.2% | -29.7% | -17.2% |
| | 32 | +16.5% | -23.4% | -18.9% |
| | 64 | +11.2% | -9.2% | -17.7% |
| | 96 | +9.2% | -7.3% | -17.7% |
| | 128 | +7.8% | -2.2% | -18.5% |
| Normal - Improved | 8 | +16.8%* | +3.9%* | -13.0% |
| | 16 | +17.1% | +7.8% | -16.0% |
| | 32 | +12.1% | +0.7% | -15.2% |
| | 64 | +9.8% | -4.9% | -16.1% |
| | 96 | +6.8% | +0.5%* | -15.2% |
| | 128 | +5.3% | -0.0% | -15.2% |
| Improved - Perfect | 8 | +18.6% | -10.8% | -21.4% |
| | 16 | +20.4% | -9.1% | -19.4% |
| | 32 | +14.5% | -16.3% | -19.8% |
| | 64 | +11.4% | -11.3% | -19.9% |
| | 96 | +10.8% | -6.9% | -20.7% |
| | 128 | +7.8% | -2.1% | -20.1% |

Table 7: Quantitative overview of evaluation of variant 2

that an improvement in the repair quality results in a

| | Worse | Normal | Improved |
|---|---|---|---|
| 1 | **1** | **1** | **1** |
| 2 | 1 | 0 | 0 |
| 3 | 2 | 0 | 0 |
| 4 | 0 | 0 | 0 |
| 5 | **1** | **1** | **1** |
| 6 | 2 | 2 | 0 |
| 7 | 2 | 2 | 0 |
| 8 | 0 | 0 | 0 |
| 9 | **1** | **1** | **1** |
| 10 | 2 | 2 | 2 |
| 11 | 0 | 0 | 0 |
| 12 | 2 | 2 | 2 |
| $\mu$ | 1.56 | 1.57 | 1.4 |
| $\sigma$ | 0.50 | 0.49 | 0.49 |
| $CV^2$ | 0.103 | 0.097 | 0.122 |

Table 8: Straightforward example of increased $\mathrm{CV}^2$ for better repair quality

higher ADI, a lower $\mathrm{CV}^2$ and a lower amount of total failures. It can be seen from table 7 that the impact of an improvement of repair quality is larger on smaller fleet sizes. However, the impact on the amount of failures is not strongly influenced by the fleet size. Hence, it can be seen that this decrease is somewhat constant for different fleet sizes.
It is interesting to note that the improvement of repair quality from ”Normal” to ”Improved” in most cases does not have a positive effect on the $\mathrm{CV}^2$. See the italic numbers in figure 7. The explanation of this is that although less subsequent failures occur, the variance in demand quantity is increasing with a higher rate compared to the mean value of the demand quantity. An example is given in table 8. For each repair quality, an overview of the amount of failures at each time point is provided. Primary failures are indicated by the bold numbers, subsequent failures are provided in regular text style.

### *4.3. Results of variant 3*

As the combination of fleet size and climate are taken into account here, eighteen different visualizations (six fleet sizes, three environmental conditions) could be shown to the reader. However, contribution of this is limited. Hence, only the quantitative overview in the form of table 9 is provided. The table provides the influence of changing repair qualities on the ADI and $\mathrm{CV}^2$. Note that the results of the temperate environmental conditions are already provided in section 4.2. The table shows the deviations of the improvement displayed in the first column. The main addition of variant 3 to the study is to explore the effect of improving repair qualities for varying environmental conditions, expressed in the ADI, $\mathrm{CV}^2$ and total failures. For the desert and humid environments, patterns similar to the temperate environmental conditions are found. The change in ADI and $\mathrm{CV}^2$ is damped out when the fleet size becomes larger, while the relative losses in total failures remain somewhat constant. Both the improvements of ”Worse” to ”Normal” and ”Improved” to ”Perfect” perform similarly for all metrics. However, the improvement in repair quality from ”Normal” to ”Improved” sees a limited decrease in the $\mathrm{CV}^2$. In fact, many scenarios induce an increase of the $\mathrm{CV}^2$. This is similar to the results of variant 2.
By comparing the temperate and humid environmental scenarios, it can be concluded that in the humid conditions, the ADI is less sensitive to the improvement of the repair quality. This results in smaller increments of the ADI compared to the temperate environmental conditions. The results of the deviation in $\mathrm{CV}^2$ provide no clear winner, as both environmental conditions outperform the other conditions for different values of fleet sizes and improvements. The relative losses in total failures are higher for the temperate environmental conditions, although the difference between the two scenarios is small.
By comparing the temperate and desert environmental scenarios, it becomes clear the desert environments perform slightly better compared to the temperate conditions when it comes to the increase of the ADI. That is to say, the increase of the ADI in the same situation is slightly less compared to the increase of the ADI for temperate environmental conditions. When comparing the deviations for the $\mathrm{CV}^2$, no clear pattern can be found. In some cases, the desert conditions outperform the temperate conditions, but the opposite occurs for the same amount of scenarios. For the decrease of the total amount of failures, the desert conditions profit slightly less compared to the temperate environmental conditions.
Generally, the increment of the ADI for improvement is the most limited for humid conditions, the performance of the decrease of the $\mathrm{CV}^2$ is similar for all environmental conditions, and the relative reduction of the total amount of failures is similar for all environmental conditions, although the temperate environmental conditions perform slightly better in most cases.

### *4.4. Results of variant 4*

The individual results of the outcome of this variant are not presented, but directly compared with the results of variant 3. Yn this way, the impact of an increased component commonality index can be evaluated. Hence, the results are discussed with the support

| Improvement | Fleet size | ADI | | | $CV^2$ | | | Failures | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Temperate | Humid | Desert | Temperate | Humid | Desert | Temperate | Humid | Desert |
| Worse - Normal | 8 | +18.4% | +13.9% | +13.9% | -17.8% | -22.7% | -18.8% | -19.9% | -16.7% | -16.8% |
| | 16 | +13.2% | +15.2% | +15.6% | -29.7% | -27.8% | -24.6% | -17.2% | -17.2% | -17.4% |
| | 32 | +16.5% | +12.8% | +13.5% | -23.4% | -10.7% | -15.7% | -18.9% | -16.8% | -16.9% |
| | 64 | +11.2% | +8.7% | +10.1% | -9.2% | -6.1% | -7.8% | -17.7% | -15.8% | -17.4% |
| | 96 | +9.2% | +6.1% | +7.8% | -7.3% | -1.0%* | -1.0% | -17.7% | -16.0% | -17.2% |
| | 128 | +7.8% | +4.8% | +6.0% | -2.2% | +2.7%* | +0.4%* | -18.5% | -16.0% | -16.9% |
| Normal - Improved | 8 | +16.8%* | +17.1% | +15.7% | +3.9%* | +12.9% | +3.4% | -13.0% | -13.3% | -13.8% |
| | 16 | +17.1% | +10.8% | +13.6% | +7.8% | +2.8% | +6.0% | -16.0% | -12.3% | -13.7% |
| | 32 | +12.1% | +8.9% | +11.4% | +0.7% | -5.6% | -3.8% | -15.2% | -13.5% | -14.8% |
| | 64 | +9.8% | +6.9% | +7.2% | -4.9% | -2.0% | -2.8% | -16.1% | -14.0% | -13.4% |
| | 96 | +6.8% | +5.0% | +5.7% | +0.5%* | +1.6%* | -1.6% | -15.2% | -13.5% | -14.0% |
| | 128 | +5.3% | +3.6% | +4.0% | -0.0% | +4.4%* | +3.4%* | -15.2% | -13.1% | -14.2% |
| Improved - Perfect | 8 | +18.6% | +14.2% | +17.0% | -10.8% | -13.4% | -7.3% | -21.4% | -17.9% | -18.3% |
| | 16 | +20.4% | +16.2% | +15.5% | -9.1% | -13.7% | -10.4% | -19.4% | -18.6% | -19.1% |
| | 32 | +14.5% | +13.0% | +12.5% | -16.3% | -12.0% | -13.6% | -19.8% | -17.8% | -18.9% |
| | 64 | +11.4% | +8.0% | +9.8% | -11.3% | -4.3% | -6.4% | -19.9% | -17.8% | -19.4% |
| | 96 | +10.8% | +6.7% | +7.2% | -6.9% | -0.0% | -1.7% | -20.7% | -18.3% | -18.8% |
| | 128 | +7.8% | +4.7% | +5.8% | -2.1% | +3.5%* | +1.2% | -20.1% | -18.1% | -18.7% |

Table 9: Quantitative overview of evaluation of variant 3

| Improvement | Fleet size | ADI | | | $CV^2$ | | | Failures | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Temperate | Humid | Desert | Temperate | Humid | Desert | Temperate | Humid | Desert |
| Worse - Normal | 8 | -8.0% | +1.2% | -0.3% | -1.6% | -4.0% | -12.7% | +3.8% | -0.3% | +0.1% |
| | 16 | +3.7% | +2.1% | -0.8% | +1.6% | +1.1% | +6.9% | -1.7% | -0.9% | +0.4% |
| | 32 | -3.1% | -1.3% | -0.7% | +6.7% | -3.9% | -1.8% | +2.5% | +1.1% | -0.3% |
| | 64 | -0.6% | +0.5% | -0.2% | -1.0% | +1.1% | +0.2% | -0.1% | -0.6% | +0.5% |
| | 96 | +1.1% | +0.2% | +0.2% | +1.0% | +0.3% | -0.9% | -1.4% | -0.0% | -0.2% |
| | 128 | -0.4% | -0.3% | -0.2% | +0.1% | -0.0% | +2.3% | +0.2% | +0.1% | -0.2% |
| Normal - Improved | 8 | +5.4% | -1.3% | +4.9% | +1.9% | +5.0% | +1.3% | -3.0% | +0.5% | -2.9% |
| | 16 | +1.7% | +1.2% | +1.3% | -2.7% | -0.3% | -0.8% | -0.8% | -0.4% | -0.7% |
| | 32 | +0.7% | +2.3% | -1.3% | -6.6% | +0.4% | -1.2% | -0.3% | -0.2% | +0.7% |
| | 64 | -0.4% | -0.7% | +0.2% | +2.1% | -0.2% | -0.2% | +0.8% | +1.0% | +0.7% |
| | 96 | +0.3% | -0.2% | -0.8% | -2.6% | -0.7% | +0.6% | -0.0% | +0.3% | +0.5% |
| | 128 | +0.9% | +0.2% | +0.2% | -0.0% | -0.3% | -2.6% | -0.2% | -0.5% | -0.0% |
| Improved - Perfect | 8 | -5.2% | -1.6% | -6.5% | +2.4% | -5.3% | -1.3% | +3.4% | +1.7% | +2.8% |
| | 16 | -7.4% | +1.7% | -2.6% | -0.7% | -2.1% | -6.7% | +2.0% | +0.1% | +1.2% |
| | 32 | -0.2% | -0.4% | +0.5% | -1.7% | +1.1% | +4.4% | -0.0% | -0.3% | +0.3% |
| | 64 | +0.9% | +0.4% | -0.1% | +0.7% | +0.4% | -0.9% | -0.1% | +0.3% | +0.6% |
| | 96 | -1.5% | -0.2% | +0.7% | +0.8% | +0.1% | +0.6% | +0.3% | +0.6% | -0.2% |
| | 128 | -0.4% | -0.1% | +0.1% | -0.2% | -0.5% | +0.1% | -0.3% | +0.3% | +0.1% |

Table 10: The absolute differences between variant 3 and 4 in percentages

of table 10. Here, the results are presented as the difference in performance of variant 3 and 4. Therefore, if in a certain scenario the ADI of variant 3 is increased by 1.0% and the ADI of variant is increased by 2.0%, the table states a difference of +1.0%.
Although in some cases there are significant differences notable, the majority of the deviations are relatively small. Hence, the level of predictability of failures is not deteriorated by the introduction of components that are operable for multiple aircraft types.

## 5. Discussion

For different fleet sizes, shared component strategies among different aircraft types and environmental conditions, the influence of the repair quality is quantified by capturing the changing values for the ADI and $CV^2$. Table 11 provides the total-effect indices of Sobol's sensitivity analysis of the different variables [23]. The total-effect index translates the contribution to the output variance of the variable. The influence of the fleet size is dominant for both the variance in outcome of the ADI as the $CV^2$. Therefore, it can be concluded that the fleet size is the main influencing factor for both metrics. This implicates that adjusting the fleet size has the largest impact on potentially lowering the ADI and $CV^2$. However, for many reasons, the expansion of the fleet is not always possible. In the cases where this expansion is not feasible and the fleet sizes cannot be increased, the influence of the repair quality on the demand pattern becomes more dominant. This can be seen in table 12, where the fleet size is fixed and the variance in the outcome depends on the repair quality, the performance differences of the aircraft in the fleet and the environmental conditions. A critical note has

| | | Fleet size | Repair quality | Performance difference aircraft | Environment |
|---|---|---|---|---|---|
| S1 | **CV$^2$** | 0.818 | 0.083 | 0.028 | 0.028 |
| | **ADI** | 0.868 | 0.021 | 0.012 | 0.022 |
| ST | **CV$^2$** | 0.885 | 0.136 | 0.078 | 0.090 |
| | **ADI** | 0.942 | 0.052 | 0.049 | 0.063 |

Table 11: Sensitivity analysis: First- and Total-Effect indices

| | | Repair quality | Performance difference aircraft | Environment |
|---|---|---|---|---|
| S1 | **CV$^2$** | 0.427 | 0.182 | 0.204 |
| | **ADI** | 0.293 | 0.286 | 0.367 |
| ST | **CV$^2$** | 0.636 | 0.361 | 0.413 |
| | **ADI** | 0.354 | 0.339 | 0.440 |

Table 12: Sensitivity analysis: First- and Total-Effect indices with a constant fleet size

to be made regarding the values of parameters for the different repair qualities. The obtained results from the data analysis are used as a reference scenario (the "Normal" repair quality), while the other three are based on a multiplication of this scenario.
The values of the parameter for the different repair qualities are exaggerated and are not based on a market research. In practice, the difference in performance will hardly be of this size. However, the advantage of these diverse scenarios is that the impact of changing repair quality is enlightened.

The strong assumption regarding the interdependency among components in the same ATA chapter results in the limitation of the usefulness of the outcome when it comes to the location of the failures and the corresponding failure patterns. Although components in the same ATA chapter could be interdependent, this does not have to be the case. Even more, components could be connected and dependent on components in different ATA chapters. However, when aggregating the results of the failures, the location of the components is not decisive for the outcome of this research.

Another assumption is made in variant 4, where the effect of different performing aircraft in a fleet is tested. Due to the lack of research and data of the performance of different types of aircraft, only a rough estimation of possible deviations in performance could be made by the author. Furthermore, the values obtained from the data set are used as a reference scenario here, presenting the average performing aircraft types. Future research could be conducted in this field to verify this assumption.

As this research quantifies the impact of the repair quality on the different demand metrics, the repair quality is used as varying parameter in the model. However, the influence of the chance of other varying parameters might also influence the metrics. Hence, there is no proof that all changes come from the varying repair quality alone and the variance caused by interaction of the different parameters should be included as well. Table 11 provides the first- and total-effect indices of the sensitivity analysis. As can be seen from the table, the difference among the first- and total-effect indices are relatively small. Hence, the influence of the interaction is limited.

Another important note should be made regarding the statistical outcomes of the Kruskal-Wallis $H$- and the Mann-Whitney $U$-tests. As the commonly chosen 95% interval provides a fair threshold for the rejection of the null hypothesis, $p$-values below 0.05 cause the rejection of the null hypothesis and thus the assumption that different groups of data have different medians. However, this $p$-value is highly dependent on the number of data points in the compared groups. As the number of iterations for the simulation is set to 50, the sizes of the subsets grow by a factor 50. Therefore, the $p$-values become smaller, resulting in a more frequent rejection of the null hypothesis. However, when reviewing only a single iteration, the $p$-values are higher and the null hypothesis is rejected less often. It is however not an option to exclude the iterations from the model, as these iterations provide the stability of the outcome by omitting the random factor.

A final critical note can be made on the limited set of drivers for failures. As stated frequently in previous research, not all drivers of failures are known, resulting in research that includes limited drivers. However, this research provides a broadening to the current knowledge by including the effect of different repair qualities.

## 6. Conclusions & Recommendations

This section comprises the conclusions of the study, followed by recommendations from the author for future research.

### *6.1. Conclusions*

The impact of changing repair quality on the predictability of the failures of components has been

quantified. In general, an improvement of repair quality induces increased ADI, a reduced $CV^2$ and a reduction of the total amount of failures. A note is made regarding larger fleet sizes (more than 64 aircraft of the same type), as the effect of the increased repair quality on the ADI and $CV^2$ becomes less significant, while the effect on the total failures remains the same. Therefore, one could conclude that when facing larger fleets, the improvement of the repair quality has a wider support base, as the downside of the implementation becomes smaller. Ironically, larger fleets have less problems with the grip on the predictability of spare parts, as is proven in this research.
The author can not judge for an individual MRO provider whether or not the repair quality should be improved, as insights in the current failure patterns are required. Insights in the current state of the repair process, the current grip on demand, the acceptable stock levels of spare parts and available budget for the improvement of the repair quality are a selection of aspects that have to be investigated. Hence, the industry should use the outcome of this research as input for case studies to improve the repair quality.

With this research, another step towards the full understanding of the drivers for failures of components is taken. With the influence of the repair quality on the failure behaviour of components, one common-made assumption is taken out.

*6.2. Recommendations*

Future research in the variance of the repair quality among different MRO providers is advised in order to strengthen the outcome of the model. Furthermore, not only incorrect repairs, leaving the component in the same broken state as before, but minimal and imperfect repairs, modelled with Kijima Type I and II techniques, should be considered as well [24],[25]. This enables the model to implement a more realistic representation of the repair process, instead of the current two-sided option.
If the aviation industry decides to further investigate the possibilities regarding the component pooling across multiple aircraft types, a detailed analysis of the performance of the individual aircraft types leads to a more accurate prediction of the performance of a heterogeneous fleet compared to a homogeneous fleet.
Finally, the author advises a follow-up study that reveals the interdependencies among flight safety-critical components for different types of aircraft. The result of this study would contribute to the practical relevance regarding the patterns and locations of the primary and subsidiary failures over time.

## References


[1] A. A. Ghobbar, C. H. Friend, Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model, Computers & Operations Research 30 (2003) 2097–2114. doi:https://doi.org/10.1016/S0305-0548(02)00125-9.

[2] A. Regattieri, M. Gamberi, R. Gamberini, R. Manzini, Managing lumpy demand for aircraft spare parts, Journal of Air Transport Management 11 (2005) 426–431. doi:https://doi.org/10.1016/j.jairtraman.2005.06.003.

[3] A. A. Syntetos, M. Keyes, M. Babai, Demand categorisation in a european spare parts logistics network, International journal of operations & production management (2009).

[4] A. A. Ghobbar, C. H. Friend, Sources of intermittent demand for aircraft spare parts within airline operations, Journal of Air Transport Management 8 (2002) 221–231. doi:https://doi.org/10.1016/S0969-6997(01)00054-0.

[5] S. Van der Auweraer, R. N. Boute, A. A. Syntetos, Forecasting spare part demand with installed base information: A review, International Journal of Forecasting 35 (2019) 181–196. doi:https://doi.org/10.1016/j.ijforecast.2018.09.002.

[6] A. F. Lowas, F. W. Ciarallo, Reliability and operations: Keys to lumpy aircraft spare parts demands, Journal of Air Transport Management 50 (2016) 30–40. doi:https://doi.org/10.1016/j.jairtraman.2015.09.004.

[7] W. Rankin, R. Hibit, J. Allen, R. Sargent, Development and evaluation of the Maintenance Error Decision Aid (MEDA) process, International Journal of Industrial Ergonomics 26 (2000) 261–276. doi:https://doi.org/10.1016/S0169-8141(99)00070-0.

[8] P. A. W. Lewis, A branching Poisson process model for the analysis of computer failure patterns, Journal of the Royal Statistical Society: Series B (Methodological) 26 (1964) 398–441.

[9] H. Wong, D. V. Oudheusden, D. Cattrysse, Cost allocation in spare parts inventory pooling, Transportation Research Part E: Logistics and Transportation Review 43 (2007) 370–386. doi:https://doi.org/10.1016/j.tre.2006.01.001.

[10] A. A. Syntetos, J. E. Boylan, J. Croston, On the categorization of demand patterns, Journal of the operational research society 56 (2005) 495–503.

[11] C. Boone, C. Craighead, J. Hanna, Critical Challenges of Inventory Management in Service Parts Supply: A Delphi study, The Journal of Operations Mangement Research 1 (2008) 31–39. doi:10.1007/s12063-008-0002-2.

[12] J. D. Croston, Forecasting and Stock Control for Intermittent Demands, volume 23, Taylor & Francis, 1972. doi:10.1057/jors.1972.50.

[13] A. A. Syntetos, J. E. Boylan, The accuracy of intermittent demand estimates, International Journal of Forecasting 21 (2005) 303–314. doi:https://doi.org/10.1016/j.ijforecast.2004.10.001.

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

[15] S. E. Rigdon, A. P. Basu, Statistical methods for the reliability of repairable systems, Wiley New York, 2000.

[16] O. W. M. Thijssens, W. J. C. Verhagen, Application of extended cox regression model to time-on-wing data of aircraft repairables, Reliability Engineering & System Safety 204 (2020) 107136. doi:https://doi.org/10.1016/j.ress.2020.107136.

[17] P. Belobaba, A. Odoni, C. Barnhart, The global airline industry, John Wiley & Sons, 2015.

[18] Y. Zhang, Y. Zheng, M. Xianchao, D. Wenjun, D. Dayong, T. Zhaoguang, S. Zhang, Exploration and implementation of commonality valuation method in commercial aircraft family design, Chinese Journal of Aeronautics 32 (2019) 1828–1846.

[19] Y. Zhang, A comparison of methods for estimating Restricted Mean Survival Time, Ph.D. thesis, Leiden University, 2018.

[20] H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, The annals of mathematical statistics (1947) 50–60.

[21] H. B. Mann, Nonparametric tests against trend, Econometrica: Journal of the econometric society (1945) 245–259.

[22] W. H. Kruskal, W. A. Wallis, Use of ranks in one-criterion variance analysis, Journal of the American statistical Association 47 (1952) 583–621.

[23] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and computers in simulation 55 (2001) 271–280.

[24] M. Kijima, U. Sumita, A useful generalization of renewal theory: counting processes governed by non-negative markovian increments, Journal of Applied Probability (1986) 71–88.

[25] M. Kijima, Some results for repairable systems with general repair, Journal of Applied probability (1989) 89–102.

## Appendix A. Component trade-off

| ATA chapter | System | Data points | Occurrences with subsidiary failures | average MTBR of primary failures [days] | Removals in all aircraft | Total unique aircraft | Flight Safety-Critical | Selected |
|---|---|---|---|---|---|---|---|---|
| 21 | Air Conditioning | 2909 | 489 | 131 | Yes | 182 | No | No |
| 22 | Auto Flight | 2892 | 332 | 118 | Yes | 159 | No | No |
| 23 | Communications | 4053 | 595 | 131 | Yes | 195 | Yes | Yes |
| 24 | Electrical Power | 1697 | 397 | 60 | Yes | 192 | Yes | Yes |
| 25 | Equipment / Furnishings | 3483 | 342 | 159 | Yes | 166 | No | No |
| 26 | Fire Protection | 447 | 117 | 112 | Yes | 112 | No | No |
| 27 | Flight Controls | 1761 | 419 | 131 | Yes | 161 | Yes | Yes |
| 28 | Fuel | 1773 | 318 | 164 | Yes | 175 | Yes | Yes |
| 29 | Hydraulic Power | 1290 | 283 | 130 | Yes | 168 | Yes | Yes |
| 30 | Ice & Rain Protection | 1597 | 365 | 153 | Yes | 177 | No | No |
| 31 | Indicating / Recording System | 2002 | 431 | 106 | Yes | 185 | No | No |
| 32 | Landing Gear | 9195 | 448 | 127 | Yes | 217 | Yes | Yes |
| 33 | Lights | 2971 | 416 | 125 | Yes | 189 | No | No |
| 34 | Navigation | 8233 | 982 | 85 | Yes | 225 | Yes | Yes |
| 35 | Oxygen | 4355 | 294 | 155 | Yes | 176 | No | No |
| 36 | Pneumatic | 1110 | 184 | 125 | Yes | 170 | No | No |
| 38 | Water / Waste | 803 | 154 | 107 | Yes | 108 | No | No |
| 49 | Airborne Auxiliary Power | 1533 | 349 | 149 | No | 108 | No | No |
| 52 | Doors | 328 | 73 | 180 | Yes | 68 | No | No |
| 53 | Fuselage | 174 | 47 | 154 | No | 48 | No | No |
| 55 | Stabilizers | 157 | 41 | 193 | No | 41 | No | No |
| 56 | Windows | 666 | 122 | 143 | Yes | 122 | Yes | No |
| 57 | Wings | 304 | 60 | 99 | No | 59 | Yes | No |
| 61 | Propellers / Propulsion | 498 | 98 | 153 | No | 63 | Yes | No |
| 71 | Power Plant General | 345 | 87 | 127 | Yes | 91 | Yes | No |
| 72 | Engine Turbine/Turboprop, Ducted Fan/Unducted Fan | 256 | 64 | 123 | Yes | 72 | Yes | No |
| 73 | Engine Fuel & Control | 868 | 191 | 112 | Yes | 135 | Yes | No |
| 74 | Ignition | 104 | 22 | 204 | No | 24 | Yes | No |
| 75 | Air | 142 | 33 | 133 | No | 34 | Yes | No |
| 76 | Engine Controls | 146 | 37 | 233 | Yes | 37 | Yes | No |
| 77 | Engine Indicating | 1088 | 193 | 149 | Yes | 160 | Yes | Yes |
| 78 | Exhaust | 399 | 107 | 252 | Yes | 107 | No | No |
| 79 | Oil | 161 | 34 | 156 | No | 35 | Yes | No |
| 80 | Starting | 575 | 132 | 107 | Yes | 132 | Yes | No |

Table A.13: Component trade-off on ATA chapter level

## Appendix B. Statistical test results

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 0.500 | 0.008 | 0.000 | 0.000 |
| **Normal** | 0.008 | 0.500 | 0.029 | 0.000 |
| **Improved** | 0.000 | 0.029 | 0.500 | 0.015 |
| **Perfect** | 0.000 | 0.000 | 0.015 | 0.500 |

Table B.14: V1: Overview of the $p$-values of the Mann-Whitney U-test for ADI of the different scenarios

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 0.500 | 0.010 | 0.000 | 0.000 |
| **Normal** | 0.010 | 0.500 | 0.056 | 0.000 |
| **Improved** | 0.000 | 0.056 | 0.500 | 0.003 |
| **Perfect** | 0.000 | 0.000 | 0.003 | 0.500 |

Table B.15: V1: Overview of the $p$-values of the Mann-Whitney U-test for $CV^2$ of the different scenarios

| **Fleet size** | **Repair quality** | **ADI** | | | | **$CV^2$** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.062 | 0.000 | | 0.500 | 0.081 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.003 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.007 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.008 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.289 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.010 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.021 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table B.16: V2: Outcome of the $p$-values of the Mann-Whitney $U$-tests

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **$CV^2$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table B.17: V2: Outcome of the $p$-values of the Kruskal-Wallis $H$-test

| Fleet size | Repair quality | ADI | | | | CV$^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.007 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.005 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.001 | 0.000 | | 0.500 | 0.004 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.001 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.013 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.006 | 0.001 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.262 | 0.001 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.382 | 0.288 | 0.390 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.303 | 0.271 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.149 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table B.18: V3: Outcome of the $p$-values of the Mann-Whitney $U$-tests for humid conditions

| Fleet size | Repair quality | ADI | | | | CV$^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.004 | 0.000 | | 0.500 | 0.049 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.003 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.001 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.001 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.010 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.267 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.008 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.111 | 0.076 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.426 | 0.004 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table B.19: V3: Outcome of the $p$-values of the Mann-Whitney $U$-tests for desert conditions

| **Fleet size** | **8** | **16** | **32** | **64** | **96** | **128** |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **CV$^2$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.780 |

Table B.20: V3: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for humid conditions

| **Fleet size** | **8** | **16** | **32** | **64** | **96** | **128** |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **CV$^2$** | 0.038 | 0.002 | 0.000 | 0.007 | 0.000 | 0.000 |

Table B.21: V3: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for desert conditions

| **Fleet size** | **Repair quality** | **ADI** | | | | **CV$^2$** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.001 | 0.000 | | 0.500 | 0.011 | 0.000 |
| | Improved | | | 0.500 | 0.001 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.003 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.004 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.005 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.014 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.024 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.033 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table B.22: V4: Outcome of the $p$-values of the Mann-Whitney $U$-tests for temperate conditions and mixed fleet composition

| Fleet size | Repair quality | ADI | | | | CV$^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.001 | 0.000 | | 0.500 | 0.093 | 0.000 |
| | Improved | | | 0.500 | 0.001 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.003 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.011 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.116 | 0.001 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.031 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.033 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.417 | 0.479 | 0.106 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.465 | 0.060 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.073 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table B.23: V4: Outcome of the $p$-values of the Mann-Whitney $U$-tests for humid conditions and mixed fleet composition

| Fleet size | Repair quality | ADI | | | | CV$^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.001 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.024 | 0.000 |
| | Improved | | | 0.500 | 0.006 | | | 0.500 | 0.005 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.012 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.005 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.076 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.004 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.019 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.388 | 0.009 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.004 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.008 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table B.24: V4: Outcome of the $p$-values of the Mann-Whitney $U$-tests for desert conditions and mixed fleet composition

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **$CV^2$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table B.25: V4: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for normal conditions and mixed fleet composition

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **$CV^2$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.381 |

Table B.26: V4: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for humid conditions and mixed fleet composition

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| **ADI** | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **$CV^2$** | 0.001 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 |

Table B.27: V4: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for desert conditions and mixed fleet composition

## Appendix C. Pseudocode of the model

**Algorithm 1:** Pseudocode of model

```
input  : All possible different values for r, weather conditions, fleet sizes and component commonality
         strategies
output: The CV^2 and ADI of the demands of spare parts and the total amount of failures

for all different fleet sizes do
    for all different weather conditions do
        for all aircraft in fleet do
            for all different repair qualities do
                for all ATA chapters do
                    Simulate primary failures
                end
                for all simulated primary failures do
                    simulate probability and occurrence of subsidiary failures and apply
                end
            end
            add the failures of the selected aircraft to the matrix with results ;
            for all ATA locations do
                Calculate the CV^2 and the ADI of the demand ;
                Add the fleet size, ATA location, demand pattern variables, weather conditions, component
                  commonality strategy and repair quality in the results
            end
        end
    end
end
```

# 1

# Literature Overview

**Earlier graded under AE4020**

This literature overview is used for the introduction of the graduation process of the Master track of Air Transport and Operations, part of the Master Control & Operations, at the faculty of Aerospace Engineering at TU Delft. This study aims to provide an overview of the current state of the research that has been conducted so far. Combining this with the proposed research question, the gap in the current state of the art research can be evinced. Besides the novelty in the current literature, it clarifies the relevance of the contribution of this research.

The research question of this study is as follows:
**What is the influence of incorrect repairs on the ADI and the $CV^2$ of the demand of spare parts in aviation?**

In order to compass all relevant literature, it is decided to collect the literature from different subjects and combine this to help answer the research question. First, in subsection 1.1, the role of spare parts in aviation is further investigated. The different categories of spare parts are defined as well as the need for spare parts in the industry.
Section 1.2 dives deeper into the demand of spare parts. The behavior of spare parts demands in aviation as well as the methods used for forecasting future demands are further inspected. The final part of this section enlightens the research so far on demand drivers of spare parts in aviation.
As incorrect repairs are directly related to repairable components, the repair process of the component is an important asset in the supply chain of this type of spare parts. Section 1.3 provides an overview of the role of maintenance in aviation, as well possible causes and the relevance in aviation for incorrect repairs. Furthermore, the role of human errors in the maintenance environment of aviation is enlightened and the different types of errors that lead to incorrect repairs are defined as well.
Finally, the possible dependency of failures is explained.
To conclude this chapter, section 1.5 states the positioning and the shortcomings of the current state of the art, as well as the related contributions of the upcoming research.

## 1.1. Spare parts in aviation

In this section, the role of spare parts in aviation is explained. Subsection 1.1.1 exposes the necessity of spare parts in this industry. Subsection 1.1.2 dives deeper into the different types of spare parts.

### 1.1.1. The need of spare parts

Aircraft need to be in a high level technical standard in order to meet up to the expectations of the passengers, the law and flight safety regulations. For airlines, this results in having aircraft to perform as expected and transport the passengers or cargo from origin to destination. Each aircraft has more than 350.000 individual components, such as electronics, engines and wires (Lyte (2016)). The condition of all components

| Aircraft Spares | Economically Recoverable | Authorized Repair | Serial Number | Depreciated | Comparative Unit Cost |
|---|---|---|---|---|---|
| Rotable | Yes | Yes | Yes | Yes | Highest |
| Repairable | Yes | Yes | No, but rarely yes | Yes | Higher |
| Recoverable | Yes | No | No | No | High |
| Expendable | No | No | No | No | Generally low |

Source: Verhagen (2019c)

Table 1.1: Properties of the different spare part classifications

varies over time. It can be the case that a certain part of the aircraft fails and the upcoming planned flight isn't safe anymore.

Component failures can occur, for instance, by overloading or fatigue. When a part of an aircraft fails it might effect the airworthiness of the aircraft. If the failed part is mission critical, immediate replacement is required. In these cases, it is crucial for airlines to minimize the down-time of the aircraft.
Airlines reach out to Maintenance, Repair and Overhaul (MRO) providers to repair the aircraft as soon as possible. Therefore, MRO providers need to have the right spare parts available at that moment. This minimizes the waiting time for repair of the aircraft, which results in lower maintenance costs as the total down-time of the aircraft is reduced. The lack of availability of a certain part can lead to an increase of the time spent on the repair operation, leading to high costs. Hence, MRO providers keep stock buffers to ensure minimal downtime. Having the right amount of spare inventory is crucial in successful airline operations for MRO providers. Contradictory, the relationship between demand drivers and the demand of spare parts, and thus the needed level of inventory stockpiles, is one of the least understood assets, as mentioned by Ghobbar and Friend (2002).
Having the right amount of spare parts available is especially essential for so called medium-range line replaceable units (LRUs), as those units have two important properties; they are not too expensive and not widely available. Engines, for example, won't be stockpiled as much, as they are extremely pricey. Similarly, simple units such as bolts won't be stacked because they are widely available, so there is no need to stack them, and holding costs can be saved. Besides the parts that need immediate replacement is case of a random failure, MRO providers also have to take predefined replacements into account.

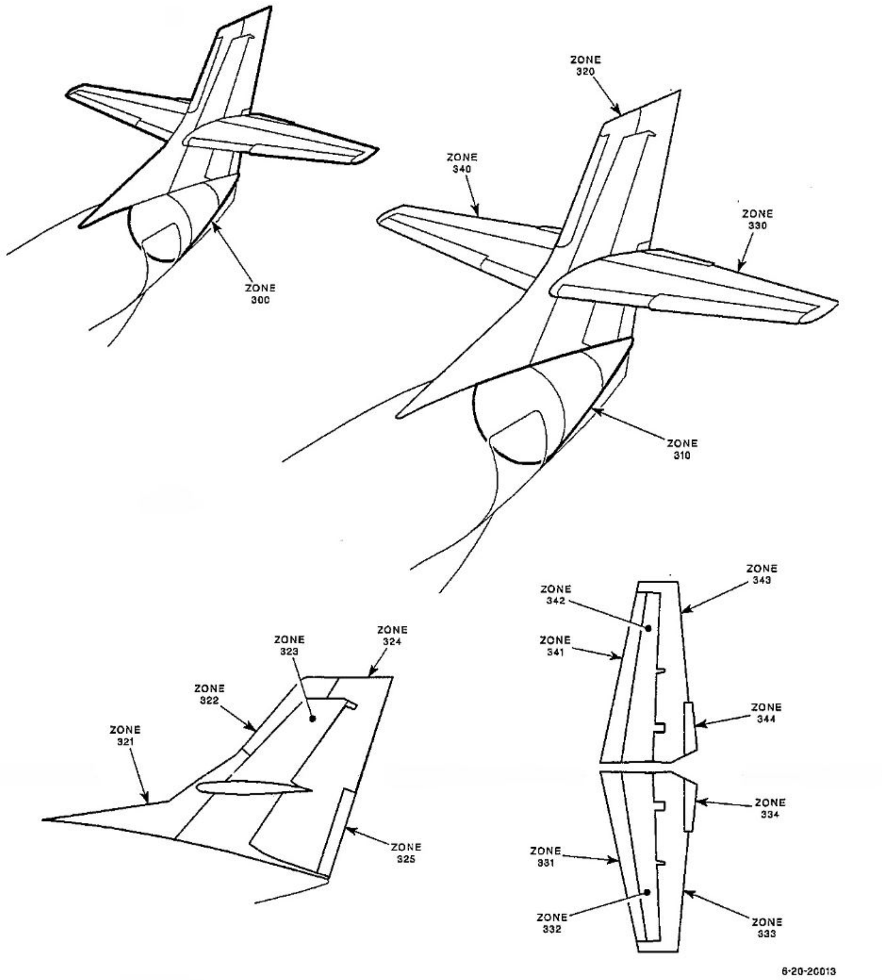
Multiple studies have shown that the optimal inventory quantities depend on many factors, such that it is hard to predict what is needed for a given moment, and that the quantities vary during the seasons. Tracht et al. (2013) showed that current planning methods of MRO providers do not cover the requirements of commercial aviation industry and provide a new method which can be build upon for further optimization. However, this research shows that MRO providers are still struggling to optimize their stocks of spare parts.

### 1.1.2. Classification of spare parts

As stated in the lecture slides of Verhagen (2019c), spare parts can be classified in different groups. Each classification has its own properties. **Expendables** are parts that are thrown away after failure as they can not be re-used. The unit cost of these parts is usually low.
**Recoverables** are, as the name is suggesting, economically recoverable parts. After failure or removal, they can be repaired and re-used. Authorized repair is not needed, and recoverables are expected to survive a single use, such that they are not depreciated.
**Repairables** are economically recoverable, but an authorized repair is necessary. The unit cost are higher than the costs of a recoverable, and repairables are depreciated. The focus of this research will involve repairables only.
Finally, **rotables** are economically recoverable parts tracked with a serial number. Authorized repair is necessary and the parts are depreciated after failure or replacement. The unit costs are the highest of all spare parts. An overview is given in table 1.1.

### 1.1.3. Localizing aircraft parts

In order to standardize the division of aircraft and engines into areas and sub-areas, the ATA-100 Zoning was introduced in 1956 by the Air Transport Association of America. Furthermore, it simplifies the manual users' problems in locating the specific components and areas. A zone is identified by one of three indicators, depending on the size of the zone. The first of three numbers represents the major zone, the second the major sub-zone, and the third the zone. For example, zone 300 represents the empennage, major sub-zone 320 represents the vertical stabilizer and rudder (part of the empennage), and zone 321 represents the vertical stabilizer leading edge. A visualization is provided in figure 1.1. In 2000 a new information standard was pub-




Source: `https://j41-vr-gfs.hk/ata_intro_info.php?id=2&lang=en`

Figure 1.1: ATA zone 300

lished. iSpec 2200, incorporating both the ATA-100 and the later developed ATA Spec 2100, was introduced for the use of digital representation and exchange of technical data.
Besides the zones, the ATA system contains a numbering system which divides the aircraft parts into standardized chapters. The ATA numbering consists of three two-digit sections. The first two digits are the most important and represent the system, the second two represent the subsystem and the final two the unit (Verhagen (2019d)). An example of the numbering convention is given in table 1.2.

| FIRST ELEMENT | SECOND ELEMENT | THIRD ELEMENT | COVERAGE |
|---|---|---|---|
| CHAPTER (SYSTEM) | SECTION (SUBSYSTEM) | SUBJECT (UNIT) | |
| **26 -** | **00 -** | **00** | Material which is applicable to the system as a whole. |
| (SYSTEM) | | | |
| "Fire Protection" | | | |
| **26 -** | **20 -** | **00** | Material which is applicable to the subsystem as a whole. |
| | (SUBSYSTEM) "Extinguishing" | | |
| **26 -** | **22 -** | **00** | Material which is applicable to the sub-subsystem as a whole. This number (digit) is assigned by the manufacturer. |
| | (SUB-SUBSYSTEM) "Engine Fire Extinguishing" | | |
| **26 -** | **22 -** | **03** | Material which is applicable to a specific unit of the sub-subsystem. Both digits are assigned by the manufacturer. |
| | | (UNIT) "BOTTLES" | |

Source: Verhagen (2019d)

Table 1.2: ATA numbering convention

## 1.2. Demand of spare parts

As mentioned in subsection 1.1.1, it is difficult for MRO providers to keep the stocks of the different spare parts on the right level. One of the main reasons for this is the fact that it is difficult to predict the demand of the spare parts. The unscheduled replacement and repair of parts on the fleets of aircraft is the main contributor to this uncertainty. Therefore, it is beneficial to understand the lifetime cycle of the different parts in the aircraft.

### 1.2.1. Categorization of demand patterns

For MRO providers, the nature of the failures of the components is crucial in order to predict what parts are necessary at which moment in time. When the demand of spare parts is more predictable, the margins of additional numbers of spare parts in the inventory can be lowered. The aviation industry carries about €30 billion yearly to stock spare parts to keep the airplanes in the air, according to Wong et al. (2007). The occurrences of failures can be bundled to predict the demand of aircraft spare parts. To classify spare parts demand, two metrics are used:

- *ADI:* $\frac{\sum_{i=1}^{n}(t_i - t_{i-1})}{n}$
- *Coefficient of Variance (CV)$^2$:* $\left(\sqrt{\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{\overline{y}}}\right)^2$

  With $\overline{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$

Based on the values of the CV$^2$ and the ADI, the demand can be classified in in one of the four categories. A graphical overview is given in figure 1.2. The

1. **Smooth** demand has a CV$^2$ < 0.49 and an ADI < 1.32. The demand is stable in both time as in quantity.
2. **Intermittent** demand has a CV$^2$ < 0.49 and an ADI > 1.32. The demand in stable in quantity. However, the time between the demand is straggling.
3. **Erratic** demand has a CV$^2$ > 0.49 and an ADI < 1.32. The time between demand is comparable, but the quantity of the demand differs.
4. **Lumpy** demand has a CV$^2$ > 0.49 and an ADI > 1.32. Both the period between demands and the quantity of the demand differ.

Multiple studies found that aircraft spares demands are lumpy.

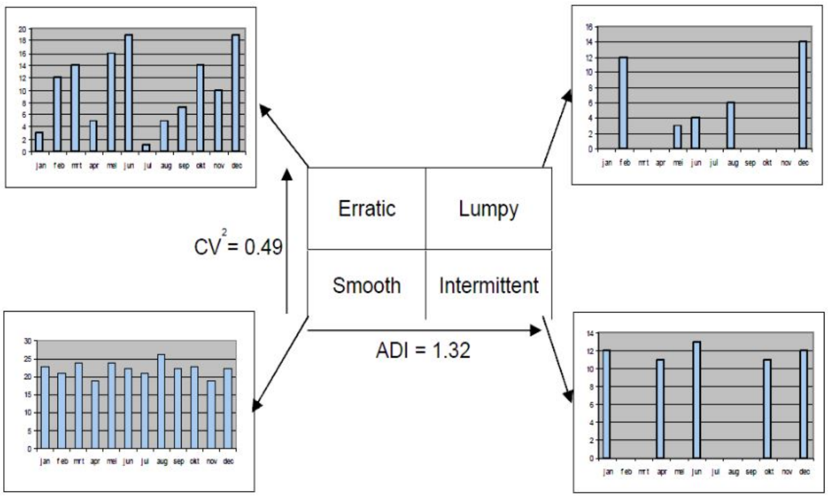



Figure 1.2: Demand classification

### 1.2.2. Forecasting of demand

Forecasting methods can be either qualitative or quantitative. Human judgement is used in qualitative forecasting. An example of a typical qualitative forecasting method is the opinion of an executive or a market survey. As no data is needed up front, this type of method can be very useful when data is not available. A downside to qualitative forecasting methods is the role of subjectivity in the results of the method. As data is available for this research, further explanation of this type of forecasting is omitted.
Quantitative forecasting methods involve numerical information that is obtained from previous periods. These methods do not rely on subjectivity from experts and is therefore objective. The downside of quantitative methods is the necessity of (a lot of) data. Quantitative forecasting can be divided into two categories. First, time-series methods use a single or multiple target values from previous time periods to forecast the future value of this variable. That being said, it is necessary that time is divided into regular intervals in order to abstract useful data from it. Further explanation is given in subsection 1.2.2. Second, causal methods try to find a relation between explanatory variables and the demand of spare parts. Here, it is assumed that the dependent variable has an cause-and-effect relationship with a number of independent variables. This will be further discussed in subsection 1.2.2.

The difficulty of forecasting the needed amount of spare parts has a great impact on the inventories of the MRO providers. A study of critical challenges of inventory management by Boone et al. (2008) concluded that the only challenge that was unanimously selected by the panel members was related to the inaccuracy of service parts forecasts. Furthermore, this challenge was ranked second in the top ten of challenges facing service parts inventory managers. This denotes the fact that this a real problem for MRO providers.

Ghobbar and Friend (2003) reviewed historical demands for multiple components and found most of the spare parts demands to be lumpy. Furthermore, Regattieri et al. (2005) enlightened the difficulty of forecasting for this type of demands by reviewing the performance of twenty different forecasting methods. However, not all the conclusions from this study are in line with the results of other studies. Some of the tested methods, such as Winter's method, perform well in the study of Regattieri et al. (2005), while in other studies, these methods are known for the large errors in forecasting.

Numerous studies provide different methods to forecast this type of demand. Guo et al. (2017) compared five types of individual forecast models with a proposed double-level combination model and found the latter to be more accurate and consistent with the actual demand. It has been proven by Qian and Chan (2015) that combination forecast is more accurate than direct forecast. One of the methods that is commonly used for forecasting intermittent or lumpy demand is Croston's method. The Croston method separates the demand size and the inter-demand interval and builds estimates for each part separately. The demand itself is assumed to occur as a Bernoulli process, and the inter-demand intervals are geometrically distributed. The sizes of the demand are assumed to follow a normal distribution. Syntetos and Boylan (2005) tested Croston's method and compared it to two other commonly used methods, the Single Exponential Smoothing and Simple Moving Averages, and a new, self created method. In this comparison, the Croston method came out on top of the traditional two other methods. However, the formulated new method proved to be more accurate than Croston's method. As Croston's method is commonly used for intermittent or lumpy demands, Romeijnders et al. (2012) proposes a method that uses a two-step forecasting technique which is more accurate than the benchmarked method. This method is tested on 10 years of Fokker Services spare parts demand data and reduces forecasts errors by up to 20%.

**Time-series methods** Five commonly used methods are presented here. They are sorted from relatively simple to more complex forecasting calculations.

**Naive Forecast** The naive method assumes that the upcoming demand is equal to the demand of the previous period. Mathematically, it can be defined as:

$$\hat{x}_{t+1} = x_t \tag{1.1}$$

**Moving Average Forecast** The moving average forecasting method takes the average of the $N$ previous periods and sets this as the forecast for the upcoming period.

$$\hat{x}_{t+1} = \frac{1}{N}\sum_{i=1}^{N} d_{t-N+i} \tag{1.2}$$

**Exponential Smoothing Forecast** The exponential smoothing forecasting method uses two parameters in order to predict the demand for the upcoming period. Both the forecast of the previous month, as well as the actual demand of the previous month is taken into account. A smoothing constant $\alpha$ is added in order to adapt the weight of both parameters to the situation.

$$\hat{x}_{t+1} = (1-\alpha)\hat{x}_t + \alpha d_t \tag{1.3}$$

**Croston's Forecast** The method created by Croston in 1972 lead to an increase in accuracy of the forecasting of intermittent and lumpy demand. Croston (1972) argued that the demand size and the demand interval should be treated separately in order to minimize the error of the forecast.

$$\hat{x}_{t+1} = \frac{\hat{s}_{t+1}}{\hat{k}_{t+1}} \tag{1.4}$$

with

$$\hat{s}_{t+1} = \begin{cases} \hat{s}_t & \text{if} \quad d_t = 0 \\ (1-\alpha)\hat{s}_t + \alpha d_t & \text{if} \quad d_t > 0, \end{cases} \tag{1.5}$$

and

$$\hat{k}_{t+1} = \begin{cases} \hat{k}_t & \text{if} \quad d_t = 0 \\ (1-\beta)\hat{k}_t + \beta d_t & \text{if} \quad d_t > 0, \end{cases} \tag{1.6}$$

**Syntetos-Boylan Forecast** Syntetos and Boylan (2001) discovered the bias of Croston's method and came up with the first steps to an improved method. Later, Syntetos and Boylan (2005) finalized the improvement of the method from Croston by deflating the forecast factor by $(1-\frac{\alpha}{2})$.

$$\hat{k}_{t+1} = \left(1-\frac{\alpha}{2}\right)\frac{\hat{s}_{t+1}}{\hat{k}_{t+1}} \tag{1.7}$$

With $\hat{s}_{t+1}$ and $\hat{k}_{t+1}$ defined in equations 1.5 and 1.6.

**Two-step Forecast** The two-step forecasting method does not directly base its forecasts on the history of the demand of the part, but dives rather deeper into the component level. The number of repairs for each component is updated separate from the average demand per repair.

$$\hat{z}^c_{t+1} = (1-\alpha)\,\hat{z}^c_t + \alpha {z_t}^c \tag{1.8}$$

and

$$\hat{a}_{t+1}^c = \begin{cases} \hat{a}_t^c & \text{if} \quad {z_t}^c = 0 \\ (1-\beta)\hat{a}_t^c + \beta \frac{{d_t}^c}{{z_t}^c} & \text{if} \quad {z_t}^c > 0, \end{cases} \tag{1.9}$$

It is worth mentioning that the average demand per repair is not updated in periods without repairs. The forecast of the demand used only for components of type $c$ is

$$\hat{x}_{t+1}^c = \hat{a}_{t+1}^c \hat{z}_{t+1}^c \tag{1.10}$$

By uniting all relevant forecasts for the components, the final two-step forecast becomes

$$\hat{x}_{t+1} = \sum_{c=1}^{C} \hat{x}_{t+1}^c \tag{1.11}$$

All methods described in this subsection are solely based on historic data and no causal drivers for demand. As mentioned by Van der Auweraer et al. (2019), less studies have focused on causal forecasting methods, while this sort of forecasting deepens the insights into spare parts demand generation.

**Causal methods** Multiple factors have an impact on the failure or removal of different aircraft components. However, recent studies fail to comprise all these factors into one model, as mentioned by Van der Auweraer et al. (2019). One of the important drivers of demand of spare parts is the failure of a component. Installed base information related to the installation date combined with a mathematical model of the lifetime of that part can improve the accuracy of the forecast, as described below.

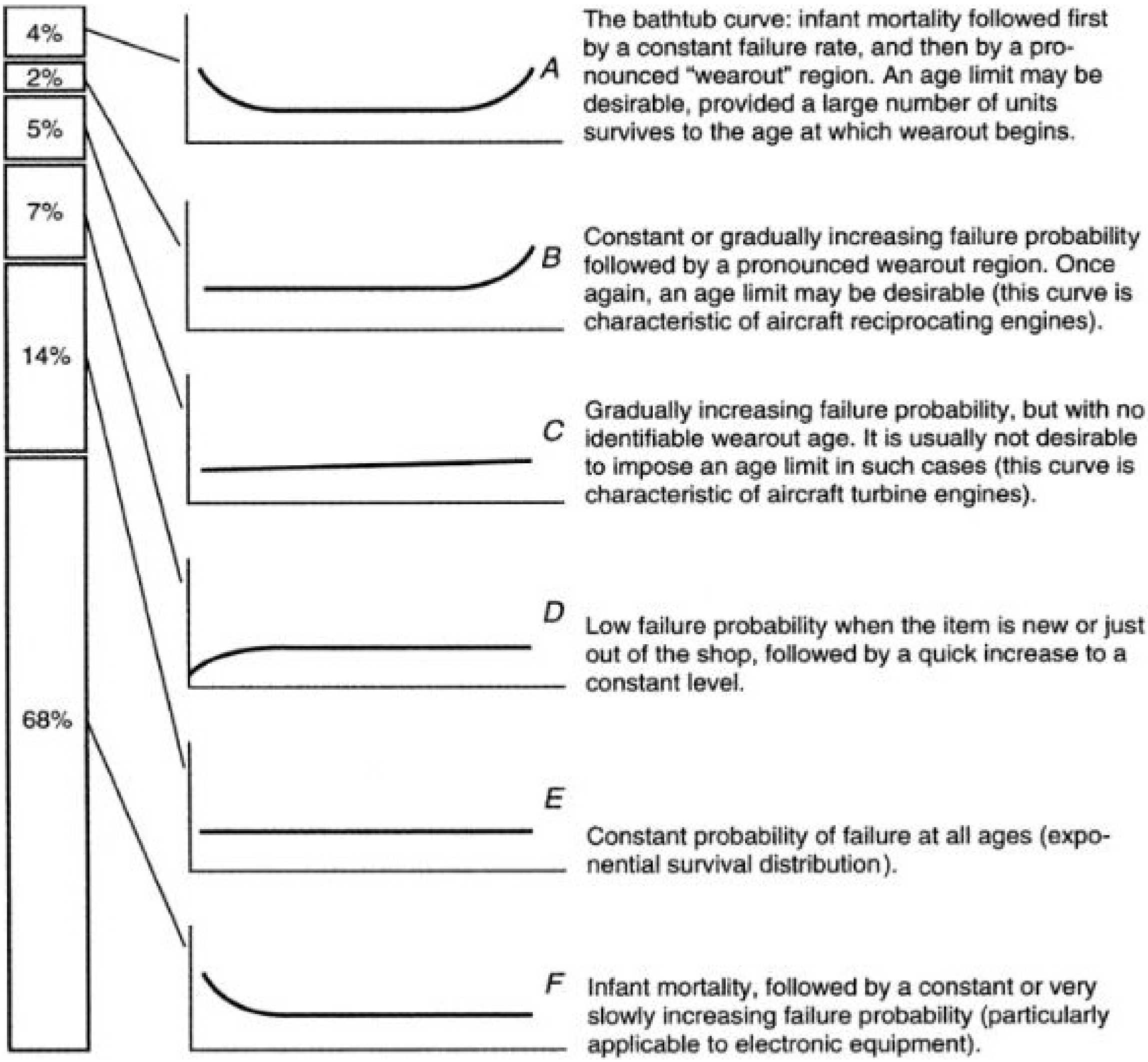
**The lifetime of components** Abernethy et al. (1983) noted that the Weibull life distribution model is a correct representation of the failure rates of (spare) parts, especially in aviation. It can therefore be concluded that the Weibull distribution represents the failures over time for components of aircraft. The Weibull distribution can represent all failure regimes, though not simultaneously.
Although some parts are replaced after they indicate that failure is about to happen by, for example, cracks, the time in function of these parts also matches a Weibull distribution. Secondly, it may occur that parts are replaced after failure. However, given the fact that typical forecast periods are approximately the same frequency as routine inspections, this inability of immediate detecting of failures is irrelevant to the accuracy of the forecasting method.
As noted by Abernethy et al. (1983), different failure regimes in aviation occur. These regimes are visualized in figure 1.3. For every failure regime, the prevalence among aircraft components has been studied by Smith and Hinchcliffe (2003). The single failure regimes have different parameters of the Weibull distribution. An overview is given in table 1.3. The four categories that are defined are: infant mortality, random failures, early

| | **Wear-out characteristic** | **Prevalence among aircraft components** | **Mathematical model** | **Weibull parameters** |
|---|---|---|---|---|
| A | Bathtube curve | 4% | 3 different Weibull distributions & combinations of values for the different stages | |
| B | "Pronounced wear-out region" | 2% | Weibull | $1 < \beta \leq 4$, $x_0$, $\alpha$ define life |
| C | "Gradually increasing" | 5% | Weibull with extended tail | $\beta > 1$ |
| D | "Low... quick increases" | 7% | Weibull | $\beta \leq 4$ |
| E | "Constant probability of failure" | 14% | Exponential | $\beta = 1$ |
| F | "Infant mortality" | 68% | Weibull | $\beta < 1$ |

Source: Smith and Hinchcliffe (2003)

Table 1.3: Reliability patterns for aircraft components

wear-out and old age wear-out. With the given prevalence of the different failure regimes among aircraft components, it is possible to simulate the failures of different aircraft components and thus the demand. By running multiple simulations with different parameters values of the Weibull distribution (Monte Carlo simulation) and by declaring a fixed end-time, it is possible to enlighten the failures over time. That is, by simulating on this way, it is assumed a component is put into service at the starting time and replaced immediately after failure.




Source: Lowas and Ciarallo (2016)

Figure 1.3: Different failure regimes for aircraft components

The Weibull distribution can represent different sorts of failure regimes. Mathematically, it is defined as:

$$F(x) = 1 - \exp\left[-\left(\frac{t - x_0}{\alpha}\right)^{\beta}\right] \tag{1.12}$$

$$f(x) = F'(x) = \frac{\beta}{\alpha}\left(\frac{t - x_0}{\alpha}\right)^{\beta} \exp\left[-\left(\frac{t - x_0}{\alpha}\right)^{\beta}\right] \tag{1.13}$$

Here, $F(x)$ represents the cumulative distribution function. That is the chance that a failure will occur before or at time $x$. $f(x)$ is the probability density function, representing the chance that failure occurs at time $x$. Both functions come with additional parameters; $\alpha$ being the scale parameter, $\beta$ the shape parameter and $x_0$ the offset or the failure-free time (Verhagen (2019a)).
The equations (1.12) and (1.13) approximate the probability density function and the cumulative distribution function of the failures, but not the replacements. Since not all parts are replaced at the exact same moment they fail, another step is needed in order to represent the behavior of replacements/demands. However, it can be proven that these rates are related. As replacement at predefined time intervals is usually costly, it is avoided in aviation.

**Environmental factors** As airlines operate their aircraft in multiple countries and thus in different climates, each aircraft in its fleet face different circumstances related to the environment. As mentioned by Ghodrati (2005), the reliability characteristics of systems and components are influenced by both operating time, as well as by factors related to the environment such as dust, humidity, moisture of the air and temperature. The consumption patterns of the tested set of spare parts significantly differs between the test with environmental factors in play, and without these factors taken into account. Artiba et al. (2005) concludes that a lot of companies do not take environmental effects into account and thus risk extra downtime regarding the unavailability of spare parts due to underestimated demand. Although environmental factors do play an role in the reliability aspects of spare parts, little research has been conducted in this field.

Although a lot of research has been conducted to the predictability of the demand of spare parts, very little is known about the demand drivers. This knowledge is necessary for further understanding the vexing problem of spare parts in aviation industry.

### 1.2.3. Demand drivers

As identified by Lowas and Ciarallo (2016), current studies tend to focus on correlations between aircraft spare parts demands and other factors, but do not uncover causation. The study introduces the novelty of seeking for demand drivers in order to increase the accuracy of predictions. Different drivers for the lumpiness of demand are tested.

On aircraft level, Ghobbar and Friend (2002) showed that aircraft utilization and flying hours are two drivers of the demand, as wear and tear of components increases with increasing utilization and flying hours, and therefore the demand rate of spare parts.
On fleet level, Lowas and Ciarallo (2016) proved that smaller fleets have higher demand $CV^2$ and higher ADI than larger fleets, and that higher buy periods (the interval in time of acquiring new aircraft) tend to rise the $CV^2$ and ADI of spares demands.

Above-mentioned methods provide a correlation between the spare parts demands and certain demand drivers. In both studies the chosen possible drivers are not further substantiated. This however does not rule out the influence of other possible demand drivers. Both studies mention that there still is little understanding of the causes of fluctuations in spare parts demand. Furthermore, it is interesting to see common denominator in all these studies. According to the best knowledge of the author, all tend to leave the correctness of maintenance of the components untouched. However, the condition of the installed part can be of great influence on the behavior and the time-to-failure of the component itself and related components as well.
As the lumpiness of the demand of spare parts is one of the main limitations for the improvement of predictability of demand, better understanding of the drivers of this sort of pattern is needed. Incorrect repairs can cause subsequent failures and thus contribute to the lumpiness of the demand. This is further elaborated in section 1.3.

### 1.2.4. Performance metrics

In order to evaluate the accuracy performance of series of forecasts, standardized metrics are used in literature. As forecasting of intermittent and lumpy series reaches wider than aircraft spare parts demand, findings from studies with different backgrounds but similar patterns of series are useful as well. However, as noted by Van der Auweraer et al. (2019), different measures are used across the studies.
One method that is often used to evaluate a proposed method is the Mean Absolute Percentage Error (*MAPE*, see equation 1.14).

$$MAPE = \frac{100}{h} \sum_{j=1}^{h} \left| \frac{\hat{x}_t(j) - d_{t+j}}{d_{t+j}} \right| \tag{1.14}$$

However, as the demand of spare parts in aviation is lumpy most of the time, and lumpy demand features periods when there is no (or zero) demand, the denominator of the *MAPE* method will result in a zero, leading to a infinite value for the *MAPE*. This property of lumpy (and intermittent) series have been underestimated or are even completely ignored in previous research (Syntetos and Boylan (2005)). In order to overcome this, Hyndman and Koehler (2006) proposed to use the Mean Absolute Scaled Error (*MASE*, equation 1.15) as standard measure. Following the recommendations of the study, scaled errors are used in this study as well.

$$MASE = \frac{1}{h}\sum_{j=1}^{h}\left|\hat{x}_t(j) - d_{t+j}\right| / \frac{1}{n-1}\sum_{t=2}^{n}|x_t - x_{t-1}| \tag{1.15}$$

Next the lack of the consistency of a single evaluation method, different benchmark methods are applied to proposed forecasting methods. Examples that are found in literature are the policy of the company, Syntetos-Boylan forecast method, Croston's method and Exponential Smoothing.

## 1.3. Maintenance in aviation

Maintenance is one of the largest costs facing airlines. According to Hobbs (2008), for every hour of flight time, 12 man-hours of maintenance is required. Not only is maintenance time consuming, it is crucial in order to keep aircraft airworthy. Most significantly, maintenance errors can lead to great implications for flight safety. Furthermore, deficient maintenance has been implicated in a growing percentage of airline accidents. Besides the safety, deficient maintenance can cause delays or cancellations. According to Boeing (2008), a flight cancellation can cost the airline around $140.000, while delay at the gate costs an average of $17.000 per hour.

When a component fails, it needs to be replaced by a component in an operating condition. Based on classification of the involved spare parts (see subsection 1.1.2), the parts are thrown away or recovered. When recovered, the condition of the parts can be restored to different levels. In most research, when talking about repairables, the perfect repair principle is assumed. This is elaborated in subsection 1.3.2.
As the research of this thesis is about incorrect repairs, information regarding possible causes for this incorrectness need to be examined. Assuming that it is never the intention to repair components on an incorrect matter, it happens by error. Subsection 1.3.3 describes the influence of human error in aviation and its consequences.

A maintenance action is the result of some sort of trigger. Maintenance can be categorized into three types (Dhillon (2002)):

1. **Preventive maintenance** is time-triggered maintenance. Actions are carried out on a periodic schedule in order to keep the components in airworthy conditions.

2. **Corrective maintenance** is triggered by the observations of maintenance persons. When a failure is discovered, it is reported and can be repaired. Assumed is that this judgement is always as accurate as possible, see subsection 1.3.3 for elaboration.

3. **Predictive maintenance** is triggered by mathematical models and the use of modern equipment. When a system or calculation alerts that the chance of failure is too high for the coming period, it is replaced prior to the failure itself.

### 1.3.1. Preventive and predictive removals

In order to prevent a component from failure, it can be taken out of use after a predefined time. However, there might be significant lifetime of the component left. This is the downside of preventive maintenance. As in aviation the competition over the last decades increased, In order to predict failure of a component, knowledge about the characteristics of this type of component is necessary. The most basic method that has been used for decades is the following. The lifetime of previously installed components is analyzed and based on an the average survival time, the expected survival time of the specific component can be found. Based on this expected lifetime, the current component can be replaced when a certain percentage of this lifetime is reached, or when the chance of failure on the next operation is found to be too high. The component is replaced without looking at the condition. Over time, the complexity of this analysis increased. Nowadays, weather conditions, the number of take offs and landings, and many more possible influences are taken into account. On this way, MRO providers try to maximize the duration of a component on an aircraft, respecting the safety of the passengers and crew on board.

As mentioned in subsection 1.2.3, the condition of the installed component is important for its lifetime. For expendable parts, this does not have a great impact as all parts are installed with a new condition. However, for parts that are taken out of service, repaired and placed back into service, the condition might vary significantly. This can either be the case due to the level of repair (subsection 1.3.2) or due to human error (subsection 1.3.3).

### 1.3.2. Level of repair

First, the **renewal (or perfect) repair** condition assumes that after the repair, the system is brought to an as-new state. If the case is restricted to conditions of history independence and stationary increments for the number of failures, the system is stable over time and can be modelled by a Homogeneous Poisson Process (HPP).
Another repair condition is the so-called **minimal repair**. In this case, independence and non-stationary increments are coupled with the 'as good as before' assumption: The repair done on the system leaves the system in exactly the same condition as just before the failure. This results in a Non-homogeneous Poisson Process (NHPP) model, which can model systems that are deteriorating or improving.
In between of the previous two repair conditions, the **imperfect repair** condition covers the middle ground. The repair that is done on the system results in a better condition as just before the failure, but a worse condition than new. This leads to imperfect repair models using the concept of virtual, which can also model systems that are improving or deteriorating (Verhagen (2019b)). The three different levels of repair are plotted in figure 1.4. As this research deals with repairable parts, the level of repair plays an important role when

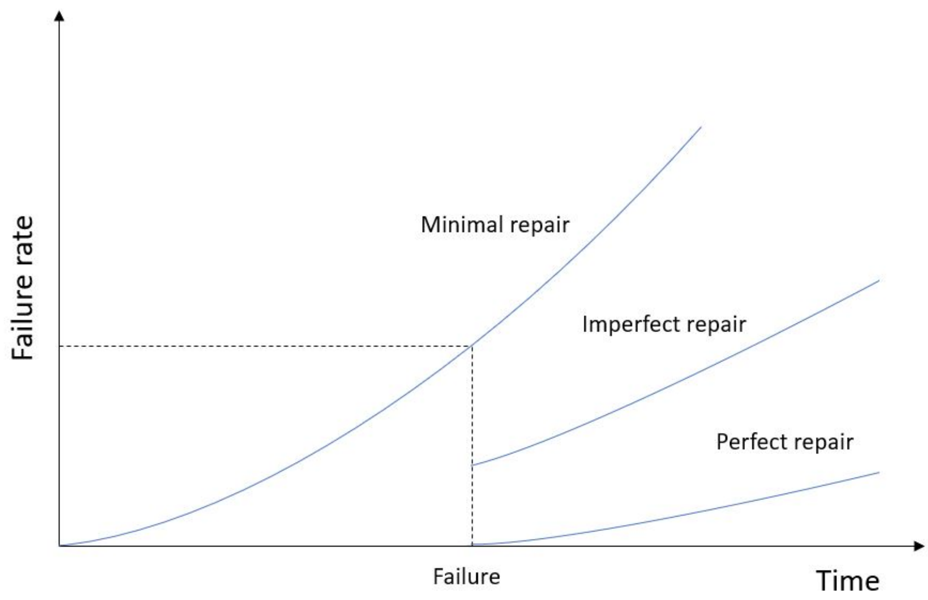



Figure 1.4: Three different levels of repair

placed back onto an aircraft.

### 1.3.3. Human errors in maintenance

During the maintenance phases of a product, humans play an indispensable role. As humans are involved, human errors occur. Following Dhillon and Liu (2006), human errors can be defined as the failure to perform a specified task (or the performance of a forbidden action) that could lead to disruption of scheduled operations or result in damage to property and equipment. Despite the fact that human errors exist for a very long time, only in the last 60 years has it been subject of scientific research. According to Dhillon (2013), in maintenance, human error can be classified into six different categories:

1. Operating errors;
2. Assembly errors;
3. Design errors;
4. Inspection errors;

5. Installation errors; and
6. Maintenance errors

In this research, the focus will lie on the maintenance, inspection and installation errors.

In order to map the effect of human errors in aviation, it is important to enlighten to maintenance tasks and environments involved. Latorella and Drury (1992) defined a model of aviation maintenance and inspection and came up with different interacting components in this system and introduces the fact that these components interact over time as well as within both physical and social, or organizational, environments. The *operator* is the first classification of components and includes the working people that are involved in the repair and inspection process, such as line operators and foreman level operators as well as production foremen and engineers. The second component, *equipment*, comprises all the tools used in inspection and maintenance tasks, such as flashlights and masks (basic) and more elaborate equipment such as that required for non-destructive inspections. Third, the *information environment*, includes both the required and used to perform specific inspections such as certain graphical representations, but also those necessary to coordinate maintenance tasks such as shift turnover forms. The final component, the *physical environment*, comprise the parameters such as noise level, temperature, color, possibility of chemical hazards and so on.

Since the beginning of the twenty-first century, human errors in aviation accidents have been widely studied. The International Air Transport Association, the IATA, assessed that 26% of the accidents in aviation is related to maintenance factors (Hackworth et al. (2007)). In order to get a better view on the errors in maintenance, the Boeing Commercial Airline Group developed an decision aid system called MEDA (Maintenance Error Decision Aid) to map the cause of the incidents (Rankin et al. (2000)). The MEDA consists of two products: a Results Form and a Users Guide. The Results Form consists of five categories of error occurrence: (1) general data, (2) operational event data, (3) maintenance error classification, (4) contributing factors analysis, and (5) corrective actions (Liang et al. (2010)). The first section contains spaces to give details of airline identification information, type of engine and date and time of the error investigation.
The second section lists all possible events which would lead to a MEDA investigation if caused by a maintenance error. Examples are delays of the flight, return to gate events and damage to the aircraft.
The maintenance error section sums up all the errors that could occur and would lead to an event. This includes improper installation, improper servicing, improper repair, improper fault inspection, actions causing foreign object damage, actions causing surrounding equipment damage, and actions causing personal injury.
The fourth section contains the situational variables that could lead to maintenance errors. The ten categories are:

1. Information - used by maintenance personnel in their jobs, for example: manuals, bulletins and tips,
2. tools, equipment and components,
3. airplane design and configuration,
4. task and job,
5. (technical) knowledge and skills,
6. individual performance factors - examples: time constraints/pressure, personal events
7. Facilities and environment
8. Organizational environment issues - examples: company policies and processes, work force stability
9. leadership and supervision - example: planning
10. communication - example: between people and organizations

The fifth and final section consists of three parts: firstly, the intended actions to prevent the error but did not happen need to be addressed. The second part consists of corrective actions that would take place where the technician does his or her work. Finally, corrective actions that take place in other parts of the maintenance organization of the airline are described.

In the study from Liang et al. (2010), 40 maintenance events and their causality relative to human errors from

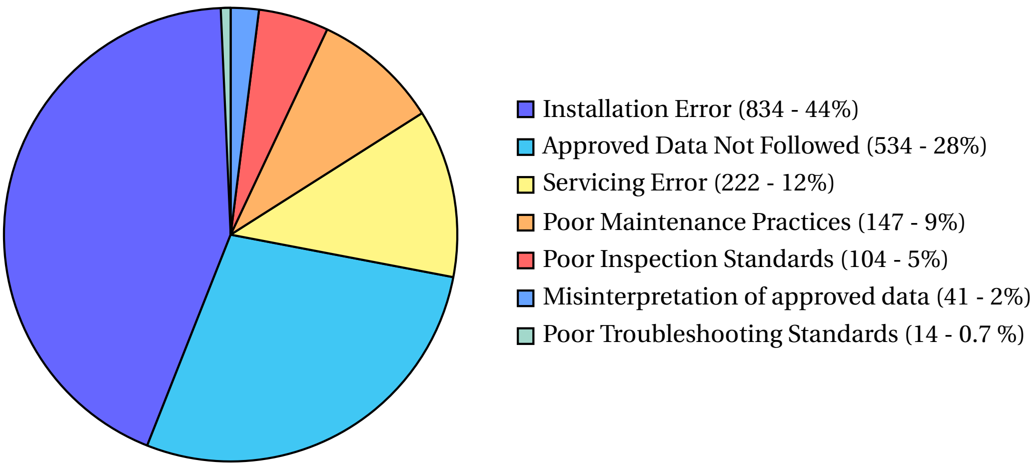
| Maintenance errors | Description | Frequency (times / 24 months) |
|---|---|---|
| Installation error | Equipment/part not installed | 2 |
| | Wrong equipment/part installed | 1 |
| | Wrong orientation | 1 |
| | Incomplete installation | 5 |
| | Access not closed | 1 |
| | Damage on installation | 1 |
| | Other | 1 |
| Fault/isolation/test/inspection error | Not found by inspection | 2 |
| | System/equipment not deactivated/reactivated | 1 |
| Foreign object damage error | Material left in aircraft/engine | 1 |
| | Other | 2 |
| Airplane/equipment damage error | Tools/equipment used improperly | 5 |
| | Pushed/pulled/drove into | 1 |
| | Other | 1 |
| Personal injury error | Struck by/against | 1 |

Source: Liang et al. (2010)

Table 1.4: Maintenance error analysis from Liang et al. (2010)

MEDA tables of an anonymous airline maintenance company between 2006 and 2007 are studied. The results related to maintenance errors are given in table 1.4. The errors that are taken into account in this research are the installation errors, aggregated with the fault/isolation/test/inspection errors. To validate the given occurrence patterns of table 1.4, the research of Airworthiness and Airspace Regulation Group (2015) is used, as well as the earlier mentioned research of Rankin et al. (2000).

The data that was used for the technical report of the Civil Aviation Authority comprised 2733 maintenance occurrence reports covering a period of seven years, from the start of 2005 until the end of 2011. As the data contained information of large aircraft, large helicopters and small aircraft, only the data related to the large aircraft is taken into account as verification. Figure 1.5 provides an overview of the classification of errors from the data. As can be seen, 834 errors are accounted to installation errors. Although it seems relevant to take the category of poor maintenance practices into account as well, this category mainly consists of foreign objects that are left behind after inspection. Examples are torches, spanners and individual bolts. Since this does not have an impact on the condition of the components on the aircraft, this is not pertinent. The installation




Source: Airworthiness and Airspace Regulation Group (2015)

Figure 1.5: Maintenance error types 2005-2011

errors can be associated to multiple factors, stated below in table 1.5. A few random events are not included in these categories. In this analysis, only the key underlying factor that contributed to the failure is assigned to a single failure. That being said, it could be the case that for some errors multiple factors could have been assigned. From here on in this research, all the causal factors from table 1.5 will be combined together and will be treated as one.

The results from the Rankin et al. (2000) research related to the maintenance error types are given in table 1.6. Combining the results of the three above-mentioned researches, the amount of relevant errors for this research are as follows:

- In the research of Liang et al. (2010), the frequency of occurrence of both the installation errors and fault/isolation/test/inspection errors combined results in 15 times / 24 months. By comparing it to all errors that are noted by this research, 15 out of 26 times (%58), the error was related to the earlier mentioned two categories.
- The research of Airworthiness and Airspace Regulation Group (2015) showed that in 834 out of 1890 occurrences (44%), the error is categorized as installation error. Noted is that this research defines the categories a bit different than the classical MEDA categories. However, errors due to poor inspections are also collected under this category, so no other categories have to be taken into account.
- Rankin et al. (2000) again used categories that differ a bit from the previous two, although this research is more in line with the MEDA than the one from Airworthiness and Airspace Regulation Group (2015). In 29 of the 74 (39%) cases, the error was related to installation errors or improper/incomplete repairs.

| **Maintenance error** | **Number of events** |
|---|---|
| Instruction non-adherence | 325 |
| Poor inspection | 158 |
| Wrong part fitted | 96 |
| Part not fitted | 73 |
| Wrong orientation | 54 |
| Cross connection | 35 |
| Poor inspection (independent) | 33 |
| Poor insp. / test | 32 |
| Panel detached in flight | 13 |
| Wrong location | 10 |

Source: Airworthiness and Airspace Regulation Group (2015)

Table 1.5: Associated causal factors of installation errors of the CAA research

| **Maintenance errors** | **Number of errors** |
|---|---|
| Installation error | 26 |
| Fault/isolation/test/inspection error | 11 |
| Improper servicing | 9 |
| Improper/incomplete repair | 3 |
| Foreign object damage error | 2 |
| Personal injury error | 1 |
| Other | 17 |
| No maintenance error reported | 5 |

Source: Rankin et al. (2000)

Table 1.6: Maintenance error analysis from Rankin et al. (2000)

Maintenance personnel face multiple human factors unique in for this industry. The work they perform may be carried in all sorts of unpleasant circumstances, from all sorts of weather types, to physically heavy activities and performing work at heights. However, the work requires attention for details and administrative skills. Furthermore, good communication and coordination is essential, yet verbal communication can be

difficult due to sound levels around the working area and the possible wear of hearing protection.
Besides that, maintenance personnel have to cope with time pressure, both on short term as on the long term. This has all to do with scheduled and unscheduled maintenance. Unscheduled maintenance results in unscheduled tasks for the maintenance personnel and are corrective in nature. In advance, it is unknown what components will fail and so the required knowledge for the maintenance personnel can not be specified up front. Furthermore, as downtime is especially costly at unknown moments due to cancellation or delays of flights, the specific components have to be fixed as soon as possible.
Scheduled maintenance tasks are typically preventative. These tasks are performed on a regular basis, which leads to routine actions for maintenance personnel. The chance of mistakes related to insufficient knowledge is very low. However, mistakes are made due to 'absent mind' tasks or breakdowns in teamwork. Furthermore, time pressure is more related to the long term in this case. The components that are placed and the checks that are done have an impact on the long term, as the aircraft is delivered back to the airline in a state that should guarantee airworthiness for the upcoming period. However, if a mistake is made by the maintenance personnel, the aircraft might not be airworthy and, in the worst scenario, the aircraft crashes during in-flight operation. It is needless to say what the impact of these mistakes can be. More causes for human errors in aviation exist, although for this research it is not necessary to uncover them all. For that reason, not all the possible causes are summed up in this literature study. The interested reader can find more details about the different causes for human error in the research of Hobbs (2008).

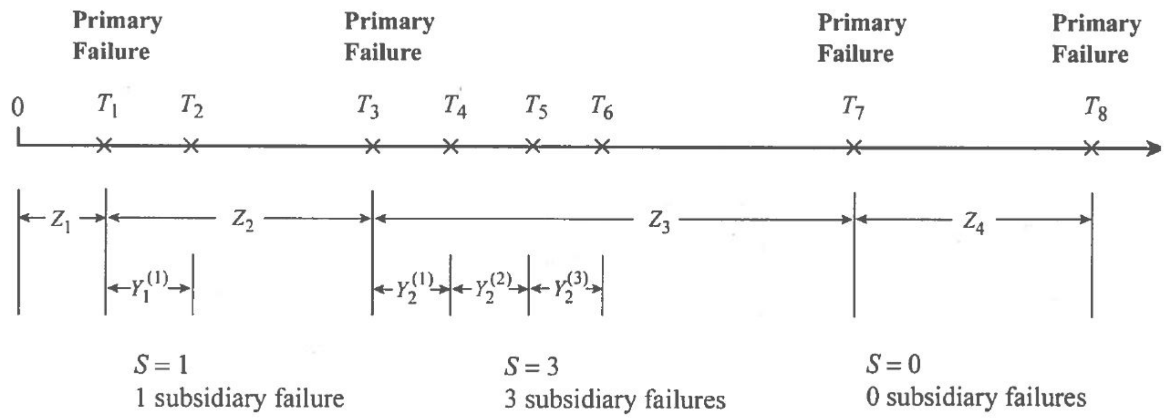
## 1.4. Dependent failures

Despite the goal of MRO providers and airlines to prevent failure of components from happening while extending their lifetime as long as possible, in practice they do occur. In case of incorrect repairs, the unrestored component is placed back into operation and is expected to operate properly. However, due to the broken state it is still in, it will not function as desired and might cause subsequent failures for much the same reason as the original failure. As shown in section 1.3, this is relevant for the aviation industry. A model that can account for these incorrect repairs is the Branching Poisson Process (BPP) (Rigdon and Basu, 2000).
The model finds it origin in 1963, as it is used to model the stop-and-go motion of vehicles as a result of a slowly moving vehicle in front. This is quite similar to a single failure causing a number of subsequent failures. In order to fit these failures in the model, it is assumed that primary failures occur with a Poisson process rate of $\lambda$. This primary failure might lead to subsequent failures or not, dependent on the correctness of the repair. Assumed is that there is a probability of $1-r$ that the repair is done correctly, and thus a probability of $r$ that this repair is not done correctly.
In the case that the repair is not done correctly, this incorrect repair will spawn a finite renewal process of subsidiary failures. The amount of subsidiary failures is a discrete random variable. Cox (1966) and Lewis (1964) described the cases where this variable has a geometric, negative binomial, and Poisson distribution. Given the fact that the time of the first primary failure $(Z_1) = z$, the expected number of failures in the interval $[0, t$ can be expressed as $H(t-z)$. Then mathematically;

$$\begin{aligned} E\left[N^{(1)}(t)\right] &= E\left\{E\left[N^{(1)}(t)|Z_1\right]\right\} \\ &= \int_0^t E\left[N^{(1)}(t)|Z_1 = z\right] f_1(z)\ dz \\ &= \int_0^t H(t-z) f_1(z)\ dz \end{aligned} \tag{1.16}$$

The same analysis can be done for the expected number of failures $N^{(k)}(t)$ in $[0, t]$ due to the $k$th subsidiary process:

$$E\left[N^{(k)}(t)\right] = \int_0^t H(t-z) f_k(z)\ dz \tag{1.17}$$




Source: Rigdon and Basu (2000) figure 3.8, page 79

Figure 1.6: Branching Poisson Process

That stated, the expected number of failures of any type in $[0, t]$ is thus:

$$\begin{aligned}\Lambda(t) &= E[N(t)] \\ &= E\big(\text{number of primary failures in}[0,t]\big) \\ &+ \sum_{k=1}^{\infty} E\big(\text{number of subsidiary failures in } [0,t] \text{ from the } k\text{th primary failure}\big) \\ &\vdots \\ &= \Lambda_Z(t) + \int_0^t H(t-z)\ \Lambda'_Z(z)\ dz\end{aligned} \tag{1.18}$$

The full derivation of the $\Lambda(t)$ function, together with the other derivations of the following equations, can found on section 3.4 of Rigdon and Basu (2000). In this document, this is shortened by the use of three vertical dots.
The Rate of Occurence of Failure (ROCOF) for the BPP equals the derivate of the $\Lambda(t)$ function over time:

$$\begin{aligned}\mu(t) &= \Lambda'(t) \\ &= \Lambda'_Z(t) + \frac{d}{dt}\int_0^t H(t-z)\ \Lambda'_Z(z)\ dz \\ &\vdots \\ &= \mu_Z(t) + \int_0^t H(t-z)\mu_Z(z)\ dz\end{aligned} \tag{1.19}$$

In the case that the primary process is a Homogeneous Poisson Process (a Poisson process with a constant intensity function), then equation 1.18 becomes:

$$\begin{aligned}\Lambda(t) &= \lambda t + \int_0^t H(t-z)\ \lambda\ dz \\ &= \lambda\left[t + \int_0^t H(t-z)\ dz\right]\end{aligned} \tag{1.20}$$

Next, the ROCOF can be obtained from equation 1.19 or by differentiating equation 1.18.

$$\begin{aligned}\mu(t) &= \lambda + \lambda \int_0^t \frac{\partial}{\partial t} H(t-z)\ dz \\ &= \lambda \Big\{ 1 + [-H(t-z)]_{z=0}^{z=t} + H(0) \Big\} \\ &= \lambda\,[1 + H(t)]\end{aligned} \tag{1.21}$$

It is worth noticing that $H(t)$ equals the expected value of failures in a finite renewal process. Assuming that it is known that there are to be exactly $s$ failures in the subsidiary process, the expectation $E(H(t)|s)$ should tend to $s$ when taking $t \to \infty$. Mathematically:

$$\lim_{t\to\infty} E\left[N^{(1)}(t)|s\right] = s \tag{1.22}$$

Therefore

$$\begin{aligned}H(t) &= E\left[N^{(1)}(t)\right] \\ &= E\Big\{E\left[N^{(1)}(t)|s\right]\Big\} \\ &= \sum_{s=0}^{\infty} E\left[N^{(1)}(t)|s\right] P(S=s)\end{aligned} \tag{1.23}$$

By taking the limit $t \to \infty$ on both sides:

$$\begin{aligned}\lim_{t\to\infty} H(t) &= \lim_{t\to\infty} \sum_{s=0}^{\infty} E\left[N^{(1)}(t)|s\right] P(S=s) \\ &\ \ \vdots \\ &= E(S)\end{aligned} \tag{1.24}$$

As the ROCOF, defined in equation 1.21, is dependent of $\lim_{t\to\infty} H(t)$, the new ROCOF function is:

$$\lim_{t\to\infty} \mu(t) = \lim_{t\to\infty} \lambda\,[1 + H(t)] = \lambda\,[1 + E(S)] \tag{1.25}$$

One of the downsides of the BBP method is that the parameter estimation is difficult to perform. An extension of BPP analysis is done by Lewis (1967), where the primary failure process is assumed to be a non-homogeneous Poisson process.

## 1.5. Conclusions of the literature review

From the previous chapters, it can be concluded that the knowledge of the generation of demand of spare parts in aviation is not sufficient. When better understanding the demand, the inventory levels of MRO providers can be lowered and maintenance tasks can be better prepared and scheduled. This leads to optimization of the industry and thus cost reduction, which is an important improvement area in aviation.

Multiple studies tried to either improve the accuracy of time-series models, or to find one or more drivers for the demand. The time-series models however do not contribute to a better understanding of the demand generation. Furthermore, in the search for the possible drivers, the relation between incorrect repairs and lumpy demand patterns has not been made, and is often assumed to have no effect to take into account. Furthermore, the current research regarding the effect of human errors in the industry is still very limited. The current state of the art is more focused towards the categorizing of different human errors than it is to the effect of these errors. Although this research is not mainly focused on the effect of human error, it can be used as a stepping stone to give potential directions for further research.

Since the lumpy behavior is one of the key factors of the inaccuracy of forecasts, and incorrect repairs strengthen the lumpy behavior of the demand of spare parts, the commonly made assumption regarding the influence of incorrect repairs must be further investigated in order to make the next step to better understanding demand generation of spare parts in aviation.

# 2
# Methodology

In this chapter, the methodology of the research is explained to the reader. Section 2.1 provides the overview of the experimental setup. Section 2.2 summarizes the important parameters of the research and provides glimpse into the usability of the outcome. Section 2.3 provides the reader with an overview of the used statistical tests in order to provide significance to the results obtained from the model.
Chapter 4 dives deeper into the working of the model itself.

## 2.1. Experimental setup

In order to simulate and measure the impact of incorrect repairs on the behavior of the demands of spare parts, Rigdon and Basu (2000) showed that the Branching Poisson Process simulates phenomenons with primary and subsidiary occurrences the best. As mentioned in chapter 1, time-series models do not provide deeper insight in potential drivers of lumpiness of demand, so could immediately be left out of consideration. For further motivation of the choice of method, the reader is referred to subsection 1.2.2.

The model generates primary failures with a certain occurrence rate. For each primary failure, the possibility of subsidiary failure(s) exists. The number of subsidiary failures that follow from a primary failure is a discrete random variable. See figure 1.6 for a graphical overview. In order to implement the correct parameters related to this method, an analysis on a data set, containing removal data from multiple aircraft over a time span of multiple decades, is used. The dataset provides the occurrence rate parameter $\lambda$ for each ATA location. Furthermore, an initial estimate of the incorrect repair probability $r$ will be acquired by looking into the occurrences of subsidiary removals. These $\lambda$ and $r$ values are used as input for the model.
Subsequently, a model is built to simulate the failures over time. In the model, the level of incorrect repairs in aviation, stated in Airworthiness and Airspace Regulation Group (2015), Rankin et al. (2000) and Liang et al. (2010) is implemented. Multiple scenarios are taken into account with different parameters. As the main goal of the research is to find the role of incorrect repairs to the ADI and $CV^2$ of the demands of spare parts, a simulation in Python (Van Rossum and Drake Jr, 1995) is used to measure its impact. In order to do so, the Branching Poisson Process is implemented. Python is chosen since this programming language has been adapted in previous projects related to courses of the Bachelor and Master program. The working of the model is explained in chapter 4.

## 2.2. Results, outcome and relevance

The variables with a high level of importance in this research are:

- $\lambda$, the occurrence rate of primary failures that differs for each set of related components;
- $r$, the chance of occurrence of an incorrect repair ($0 \leq r \leq 1$);
- fleet size, as an extension to the model;
- weather conditions, as an extension to the model

The result of the impact of the incorrect repairs on the demand of spare parts will be captured in

- The Average Demand Interval (ADI);
- The squared Coefficient of Variance ($CV^2$)
- The total amount of failures over the period

As high values for the first two of these output variables result in the lumpy behavior of the demand and thus its unpredictability, the added value of the incorrect repairs to this behavior is unveiled, and a step towards better understanding the causes of this lumpiness is provided. The amount of failures over time is taken into consideration as the first two metrics do not provide insight into the quantity of failures. For future research, the understanding of factors that impact the lumpy demand patterns can lead to increased accuracy in the forecasts of the demand.

## 2.3. Statistical significance

The impact of varying values for parameters can only be presented if the outcome varies significantly. In order to test this, multiple statistical tests exist.
One of the commonly used is the classic one-way Analysis of Variance (ANOVA) test (Field, 2013). This ANOVA test tests if two or more populations have the same mean value. An extension of this classic one-way ANOVA test is the Multivariate Analysis of Variance (MANOVA) test and is used when multiple continuous dependent variables come into play (Hair et al. (1998)). The MANOVA tests whether or not the earlier mentioned parameters explain a statistically significant amount of variance in the $CV^2$ and ADI of the demands. One of the advantages of using a MANOVA over multiple ANOVA tests is that the impact of a Type-I error (reject correct null hypothesis, false positive) is much smaller. The characteristic F-value is calculated according to equation 2.1. Here, $MS$ stands for the mean square.

$$\begin{aligned} F &= \frac{\text{Variance between populations}}{\text{Variance within populations}} \\ &= \frac{MS_{\text{populations}}}{MS_{\text{error}}} \end{aligned} \tag{2.1}$$

However, in order to obtain valid results from these tests, all assumptions regarding the underlying data have to be satisfied.

- The first assumption in order to obtain valid results from these tests, reads that the groups all share the same variance or standard deviation. This assumption can be tested with the use of the Levene test (Levene, 1960). Its validates the null hypothesis that all groups have equal variance.
- The second assumption is that underlying data must fit a normal distribution when it comes to the Cumulative distribution function (cdf) of the data. This can be tested with the Kolmogorov-Smirnov (KS) test (Kolmogorov, 1933). If the underlying does not meet the requirements for the one-way ANOVA test, alternative tests are used.

To deal with the first assumption, an alternative to the classic one-way ANOVA is used. The test that is commonly used in literature and serves the same goal as the classic one-way ANOVA test, is Welch's T-test (Welch, 1951). Welch's T-test is an adaption of the student's T-test, which was used in 1908 to monitor the quality of the stout from the Guinness Brewery (Student, 1908). The student's T-test however is not applicable to data that do not have equal variances. Therefore, the Welch T-test is the more applicable form in this case. See equation 2.2. Here $\overline{X}_n$, $s_n$ and $N_n$ represent the sample mean, standard deviation and the size of the $n$th sample.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{{s_1}^2}{N_1} + \frac{{s_2}^2}{N_2}}} \tag{2.2}$$

However, Welch's T-test is assuming that the underlying data is normally distributed. A test that does not require the underlying data to be normally distributed or have equal variances and is often used in literature is the KW H-test (Kruskal and Wallis, 1952). The KW H-test validates or rejects the hypothesis that the median of all considered populations are equal. When considering multiple groups at once, a rejection of the null hypothesis does not indicate which of the groups differ, but it rather points out that not all the groups have the same median. The H statistical value of the KW H-test is mathematically elaborated in equation 2.3. Here, $n$,$c$,$T_j$ and $n_j$ stands for the sum of sample sizes of all samples, the number of samples, the sum of ranks in the $j$th sample and the size of the $j$th sample. Subsequently, the H value is compared to the critical chi-squared value with $c-1$ degrees of freedom. If the chi-squared value is less than the H value, the null hypothesis is rejected. Otherwise, the evidence is not sufficient to conclude that the null hypothesis is incorrect (Glen, 2016).

$$H = \left[ \frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{{T_j}^2}{n_j} \right] - 3(n+1) \tag{2.3}$$

When considering only two populations, instead of the KW test, The Mann-Whitney U test is commonly used in literature (Mann and Whitney, 1947). The null hypothesis of this test is that for randomly selected values from both populations, the probability of the selection from the first population being greater than selection from the second population is equal to the probability of the selection from the second population being greater than the selection from the first population (equal medians). The calculation for the $U$ statistic, typical for this method, is left out of consideration here. The full calculation can be found in the research of Mann and Whitney (1947).

The Kruskal-Wallis test and Mann-Whitney U test both take out the two assumptions that influence the quality of the analysis performed by the tests. For this reason, these two tests are used in this research for the validation and rejection of null hypotheses. When the amount of populations exceeds the value of two, the Kruskal-Wallis test is used first in order to verify statistical differences among the group. Next, each pair of populations is tested with the use of the Mann-Whitney U test. An overview of the methods and the properties is provided in table 2.1

| **Method** | **Test for** | **Assumes equal variance** | **Assumes normal distribution** | **Applicable amount of populations** |
|---|---|---|---|---|
| One-way ANOVA | Equal means | Yes* | Yes** | >2 |
| Student's t-test | Equal means | Yes* | Yes** | 2 |
| Welch's t-test | Equal means | No | Yes** | 2 |
| Kruskal-Wallis H-test | Equal medians | No | No | >2 |
| Mann-Whitney U-test | Equal medians | No | No | 2 |

* Levene test required, ** Kolmogorov-Smirnov test required

Table 2.1: Overview of validation tests

# 3

# Data preprocessing

This chapter describes the content and the structure of the data set that is used in this research. Furthermore, the selection of useful data is elaborated. Finally, the role of the retrieved data from the data set is substantiated.

## 3.1. Content of the main data set

The used main data set comprises more than 45 years of removal data of different types of aircraft from an MRO provider. In total, almost 500.000 data points are collected in this set.
Each data point represents a removal of a component. Besides the component ID, an identification number for internal use within the organization of the MRO provider, part number, quantity, date of removal, aircraft type ID, External Organization Code, ATA number, ATA category, aircraft serial number, part serial number, time since installation, cycles since installation, time since overhaul, cycles since overhaul, times since new, cycles since new, reason for removal and the position are stored. Although the first data points originate from the early 1930's, the provided information for these data points is limited. From 2008 onwards, the aircraft serial number, part serial number, time since installation, cycles since installation, time since new, cycles since new and the reason for removal are stored more frequently. The reason for removal provides a short explanation why the component is removed, but is not consistent in language, clarity or description level. An overview of some data points is given in table 3.1. Here, only relevant columns are displayed. In order to load

| **PartNumber** | **Date** | **AtaNumber** | **AtaCat** | **AircraftSerialNumber** | **ExtOrgCode** | **PartSerialNumber** |
|---|---|---|---|---|---|---|
| HTE960034 | 7-2-2012 00:00 | 381202 | 381 | 11583 | 1-HN0 | 17081593 |
| G6937-18 | 9-2-2012 00:00 | 235101 | 235 | 11583 | 1-HN0 | 639 |
| 72015727 | 9-2-2012 00:00 | 253101 | 253 | 11583 | 1-HN0 | 4516 |
| 5011809-3 | 12-2-2012 | 324201 | 324 | 11583 | 1-HN0 | AUG92-0782 |

Table 3.1: Small example of relevant columns of the dataset

the data set into Python, the pandas package is used. This package is commonly used for data analysis of large data sets. See code 3.1.
A second data set is loaded in. This data set comprises the translation from the external organisation code (see table 3.1, column ExtOrgCode) to the name of the organisation. This is used in a later phase of the data analysis. Since this data set is confidential, only a single example of the content is given in table 3.2. This should provide the reader with sufficient understanding of the data set. Here, the ExtOrgCode2 has similar values to the ExtOrgCode column, but deletes the white space that is present in the ExtOrgCode column, such that it is easy accessable in Python.

```
file_location = r'C:\Users\larsh\Documents\TU Delft\Master\Air Transport &
↪  Operations\Overig\Afstuderen\Data\ComponentRemovals.xlsx'
df = pd.read_excel(file_location, header = 0)

file_location2 = r'C:\Users\larsh\Documents\TU Delft\Master\Air Transport &
↪  Operations\Overig\Afstuderen\Data\External_organisations.xlsx'
externalorganisationdata = pd.read_excel(file_location2, header = 0)
externalorganisationdata["ExtOrgName"] = externalorganisationdata["ExtOrgName"].str.rstrip()
```

Source Code 3.1: Obtainment of data

## 3.2. Data cleaning

In order to perform a proper data analysis, the data has to be cleaned. The following data cleansing methods are applied to the dataset.

### Data type conversion

The data of the "Date" column has to be transformed into a date and time format, such that Python will not see the value as an integer. This conversation makes it possible for later calculations to derive the amount of days between a set of removals. Furthermore, the ATA number of all data points is converted to a numeric type.

### Missing values

All data points that have no aircraft serial number are let out of further consideration as this property is required in order to find the Mean Time Between Removal (MTBR) on the specific aircraft. Furthermore, the data points that do not come with an ATA number of the removed component, representing the location, are left out of further consideration, as this is required in the first analysis where it is assumed that the ATA location is decisive for the possibility of interdependencies between components.

## 3.3. Impossible values

The first recorded date in the provided dataset is from the year 1900, while the latest date is from the year 3013. Clearly, some mistakes are made regarding the dates. These registration errors are made related to the date of the removal. The time-frame that is taken into further consideration comprises the years 2007 until 2020. These years are chosen as 2007 is the first year in the data set that contains data points from more than one day only. The final year to take into consideration is 2019, as that year is the most recent year that can include correct data. Although some data points belong to a date that extends 2019, this cannot be correct, as those data points lie in the future.
See source code 3.2 for the implementation in Python. The data is now prepared for further analysis. By running the code, the reading of the Excel can be manually turned on and off with the binary variable `getdata`.

```
getdata = 1

if getdata == 1:
    file_location = r'C:\Users\larsh\Documents\TU Delft\Master\Air Transport &
    ↪  Operations\Overig\Afstuderen\Data\ComponentRemovals2.xlsx'
    df = pd.read_excel(file_location, header = 0)
    df["Date"] = pd.to_datetime(df["Date"], errors = 'coerce')
    df = df[df.Date.dt.year > 2006]
    df = df[df.Date.dt.year < 2020]
    df = df[df["AircraftSerialNumber"].notnull()]
    df = df[df["TimeSinceInstallation"].notnull()]
    df = df[df["AtaNumber"].notnull()]
    df = df[df["AtaNumber"].str.isnumeric()]
    df = df.fillna(0)

    df["ExtOrgCode"] = df["ExtOrgCode"].str.rstrip()
```

Source Code 3.2: Data cleaning

## 3.4. Data analysis

Before diving deeper into the calculations, some preparation work is necessary. First, the length of the data set with the remaining data is calculated as this makes it easier in a later stage of to recall the length of the data set, instead of using a formula for this over and over. Next, an array is created with all serial numbers of the aircraft, followed by an array with all the unique values of the serial numbers of the aircraft, thus representing the number of aircraft that is taken into further consideration with its length. Next, an empty dictionary is created to later store the data per aircraft. Finally, four functions are defined. The first is used to calculate the number of days between two dates, the second to enable the pop-up of message boxes, the third to calculate the Probability Density Function (pdf) of a Poisson process, and the final code to calculate the time between removals of two subsequent removals. See code 3.3 for the exact statements in Python.

| ExtOrgCode | ExtOrgName | ExtOrgAbbr | ExtOrgNumber | ExtOrgCode2 | Climate | Factor |
|---|---|---|---|---|---|---|
| 1-HN0 | KLM Cityhopper NL | KLM CITY | 189 | 1-HN0 | Temperate | 1 |

Table 3.2: Sample of the data set converting external organisation data

```
nrows = df.shape[0]

AircraftSerialNumber = np.zeros(nrows)
AircraftSerialNumber[:] = df.iloc[:,9]
SerialNumbers = np.unique(AircraftSerialNumber)

Aircrafts = {}

#%%
def Mbox(title, text, style):
    return ctypes.windll.user32.MessageBoxW(0, text, title, style)
def diff_day(d1, d2):
    return abs((d1.year - d2.year)*365.25 + (d1.month - d2.month)/12*365.25 + (d1.day - d2.day))
def poisson(k, lamb):
    """poisson pdf, parameter lamb is the fit parameter"""
    return (lamb**k/factorial(k)) * np.exp(-lamb)
def HPPcheck(inputlist):
    array = np.zeros(len(list(HPPtest[int(float(title)),title2]))-1)
    for j in range(len(array)):
        array[j] = (list(HPPtest[int(float(title)),title2])[j+1] -
        ↪  list(HPPtest[int(float(title)),title2])[j]).days
    return array
```

Source Code 3.3: Dimension calculation & preparation of storage for aircraft data

Next, the data can be loaded into the desired format. For every aircraft, a separate dictionary entry is filled with its data. In order to overcome possible longer processing times of Python, this process of data loading can be manually switched off by changing the value for the binary variable `calculations` to 0. Finally, the column corresponding to the ATA location is shortened to a two digit number.

```
calculations = 1

if calculations == 1:
    for i in range(len(SerialNumbers)): #looping over all the aircraft in data
        Aircrafts["Aircraft{0}".format(i+1)] = df[AircraftSerialNumber == SerialNumbers[i]]
        ↪  #creating dictionary entries
        Aircrafts["Aircraft{0}".format(i+1)].AtaCat =
        ↪  (Aircrafts["Aircraft{0}".format(i+1)].AtaCat).astype(str).str[:2].astype(int)
```

Source Code 3.4: Loading the data in desired format

### Clustered removals

Subsequently, the clustered removals can be determined. First, a new DataFrame, `Cluster`, is made, consisting of columns that store the aircraft number, the ATA location, the date, the primary failure and its possible subsequent removals.

Here it is assumed that removals are interdependent if and only if the first two digits of the ATA code are equal. For example, a failure with ATA identification number 792151 (the air-oil heat exchanger) and 792243 (the pressure filter of the oil distribution) are assumed to be related and thus candidates for possible clustered failures.
Another condition to be a candidate for a clustered (and thus interdependent) failure, is the time span between (at least) two removals. In this example, it is chosen to set a threshold to this value of 14 days. This value is chosen in the analysis as this time span (possibly) incorporates a number of flights. This is crucial as the load on a component during takeoff, flight and landing might trigger the failure.
Next, each subset of a specific aircraft is looped through. For each subset, the first and the second data point are the starting point. Multiple scenarios can occur, and they are explained one by one:

- If both conditions mentioned above are `True` (within seven days and corresponding ATA locations) for these points, there are two options. In both cases, after the actions, the index of the candidate point is increased with one.
  - If the first of the two data points (the reference point) has not yet been added to the new `Cluster` DataFrame, it is added first. The serial number of the aircraft, the ATA category, the date and the serial number of the part are stored. What follows is that the second data point (the candidate point) is added to the same row in `Cluster`, only it is saved to the first empty `Subsequent` column. Here, only the serial number of the part is stored.
  - if the reference point has already been added to the DataFrame due to an earlier check, only the candidate point is added to the DataFrame, at the first free spot in the list of `Subsequent` removals.
- If the reference point and the candidate point are interdependent but are not within the threshold related to the maximum time span:
  - And the reference point has the same serial number of the part as the current final primary serial number that is added to the DataFrame, the new reference point becomes the last tested candidate point, and the new (first) candidate point becomes the data point after the new reference point.
  - Else, the current reference point is added to the `Cluster` DataFrame and the new reference point becomes the last tested candidate point, and the new (first) candidate point becomes the data point after the new reference point.
- If the reference point and the candidate point do not meet both conditions and the candidate point is the first candidate point for this reference point (the candidate point is one row below the reference point in the data set), the reference point is added to the `Cluster` DataFrame and the new reference point becomes the last tested candidate point, and the new (first) candidate point becomes the data point after the new reference point.
- If the reference point and the candidate point do not meet both conditions and the candidate point is not the first candidate point for this reference point (so other candidate points have been tested and are already saved to the `Cluster` DataFrame), the new reference point becomes the last tested candidate point, and the new (first) candidate point becomes the data point after the new reference point.
- The final possibility that should be considered is that when the new reference point has a index that is larger than the length of the subset of the current aircraft, then the last data point in the subset of the current aircraft should be added to the `Cluster` DataFrame and the analysis should continue to the next aircraft. This step might seem trivial, but due to the fact that in this analysis only the reference point is continually added to the `Cluster` DataFrame, the final candidate point might be skipped due to the fact that the index of the candidate point is larger than the dimensions of the subset of the selected aircraft.

The precise implementation of this analysis in Python is given in appendix A, the corresponding algorithm is provided in algorithm 1. After this step, data comprising 265 aircraft is left.

**Algorithm 1:** Algorithm for clustered removals

```
input  : dictionary entries for all considered aircraft
output: Primary and subsidiary removals
for All aircraft do
    set the first removal as reference point and the following removal as first candidate point
    if the tested removals occur within 14 days and have the same location in the aircraft then
        if the reference point is already added to the new DataFrame then
            Add only the candidate point to the first free spot of subsequent removals for the
              reference point
        else
            Add the reference point as primary removal and store corresponding data (time, location,
              aircraft serial number, aircraft type)
        end
    end
    if the reference point and the candidate point have the same ATA location but are not within 14
      days then
        if The current reference point is already in the new DataFrame then
            The last tested candidate point will become the new reference point and the data point
              after the new reference point becomes the new candidate point
        else
            the current reference point is added to the DataFrame and the last tested candidate point
              become the new reference point. The data point after the new reference point becomes
              the new candidate point
        end
    end
    if the reference and candidate point do not meet both conditions then
        if the candidate point is the first candidate point for the reference point then
            the reference point is added to the DataFrame and the current candidate point becomes
              the new reference point
        else
            the reference point is added to the DataFrame and the current candidate point becomes
              the new reference point
        end
    end
end
```

## Offset

Another important property of the data is the offset that is generated from the primary failures. In order to calculate the amount of removals that produce an of offset of $i$ subsidiary removals, for each ATA location, the amount of values in each column is counted and stored in the `offset2` dictionary. Furthermore, a probability distribution is added to the third column of the array as well, in order to provide the model with an empirical distribution.

```
offset2 = {}
for titleoffset, groupoffset in Cluster.groupby("ATA location"):
    offset2[titleoffset] = np.zeros([groupoffset.shape[1]-5,3])
    for j in range(1,groupoffset.shape[1]-5):
        if j == groupoffset.shape[1]-5:
            offset2[titleoffset][j,1] = (groupoffset["Subsequent
            ↪  {}".format(groupoffset.shape[1]-5)].isna().sum())
        else:
            offset2[titleoffset][j,1] = (groupoffset["Subsequent {}".format(j+1)].isna().sum() -
            ↪  groupoffset["Subsequent {}".format(j)].isna().sum())
        offset2[titleoffset][j,0] = j

for titleoffset, groupoffset in Cluster.groupby("ATA location"):
    for j in range(1,groupoffset.shape[1]-5):
        offset2[titleoffset][j,2] = offset2[titleoffset][j,1] / sum(offset2[titleoffset][:,1])
```

Source Code 3.5: Storing the offset of the primary failures

## Climate

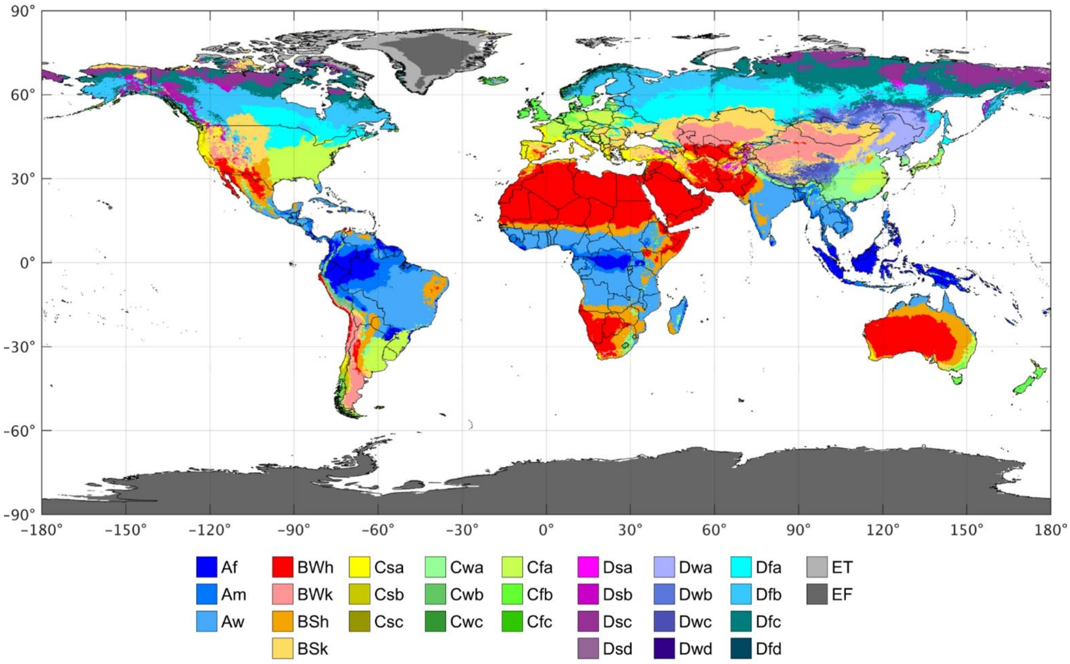
Next, for all considered aircraft, the climate related to the main hub of the operator is provided in the `ExtOrg` DataFrame. Here, the information from the `externalorganisationdata` data set is used. For each aircraft, the aircraft serial number is used in order to track the corresponding operator and the corresponding climate. The criteria for the classification of the three different climates is based on the research of Thijssens and Verhagen (2020). The Köppen Climate Classification System is used for the classification of the climates. Its categories are based on the average of temperature and precipitation over time (Chen and Chen, 2013). An overview of the different climate codes is given in figure 3.1. According to Thijssens and Verhagen (2020), three different climates that influence the MTBR of components are *Temperate, Humid* and *Desert*. The *Desert* climate is defined as the regions with Köppen climate codes BWh, BWk and Cwb. The *Humid* climate is defined as the regions with Köppen climate codes Af, Am, Aw, As, Cfa, Cwa, Dsa, Dsb, Dwa, Dwb, Dfa and Dfb. All other climate codes are considered to represent the *Temperate* climate group. The quantification of the influence of the climate is provided in table 4.2, as chapter 4 dives deeper into the simulation parameters. Here, the *Temperate* climate is taken as a reference and both values of the *Humid* and *Desert* climate are expressed as a percentage of the reference climate.
By adjusting the data from the data set for the climate that was operated in, all data can be calculated as values for the reference climate (temperate climate). Hence, on this way, in the simulations, different climates can be incorporated by adjusting the climate factor by the provided factors.
This distribution of climate zones is used in table 3.2, as well as in code 3.6.

```
ExtOrg = df[['AircraftSerialNumber','ExtOrgCode']].drop_duplicates()
ExtOrg.reset_index(drop = True, inplace = True)
ExtOrg["Operator"] = ""
ExtOrg["Climate"] = ""
ExtOrg["Factor"] = ""
for i in range(len(ExtOrg)):
    ExtOrg.iloc[i,2] = (externalorganisationdata[externalorganisationdata.ExtOrgCode2 ==
    ↪  ExtOrg.iloc[i,1]]["ExtOrgName"]).tolist()[0]
    ExtOrg.iloc[i,3] = (externalorganisationdata[externalorganisationdata.ExtOrgCode2 ==
    ↪  ExtOrg.iloc[i,1]]["Climate"]).tolist()[0]
    ExtOrg.iloc[i,4] = (externalorganisationdata[externalorganisationdata.ExtOrgCode2 ==
    ↪  ExtOrg.iloc[i,1]]["Factor"]).tolist()[0]
```

Source Code 3.6: Operator and climate information for all aircraft

## Primary removals

Next, a new DataFrame `Primaries` is created, referring to the name of the removals that occur with a Poisson rate of $\lambda$ according to the theory of the BPP. The DataFrame consists of seven columns; the aircraft number, the ATA location, the aircraft type, the date of the first removal, the date of the last removal, the number number of removals and the lambda value. For each combination of aircraft number and ATA location, it is checked how many removals occur and what the timeframe of these removals covers. In the case that only one removal is found that for a certain combination of ATA location, aircraft type and aircraft number, the date




Source: Beck et al. (2018)

Figure 3.1: Köppen climate classification map

of the first and the date of the last removal will be equal and the lambda value, which is calculated according to equation 3.1, will be infinity. As the number of removals for this combination is only one, they are filtered out.

Next, the `Primaries` DataFrame is sorted by ATA location and subsequently by the date of the first removal. A boxplot for each ATA location is made in order visualize the spread of the $\lambda$ value. Furthermore, a DataFrame `ATA repair performance` that stores the number of removals, the number of removals without subsidiary removals and the initial estimate of the value for $r$ for all possible combinations of ATA location and aircraft type. The estimate of $r$ is calculated by the number of primary removals that do have subsidiary removals, divided by the total number of removals. See code 3.7 for the exact implementation in Python. After this step, data comprising 244 aircraft is left.

$$\lambda = \frac{\#\text{removals} - 1}{\#\text{days in between}} \tag{3.1}$$

```
#%%
plt.close("all")
Primaries = pd.DataFrame(columns = ['Aircraftnumber', 'ATA_location','ACtype', 'firstdate',
↪  'finaldate', 'removals', 'lambda'])
ACNATA = 0
while ACNATA < len(Cluster):
    Primaries = Primaries.append(pd.Series([Cluster.Aircraftnumber[ACNATA],Cluster["ATA
    ↪  location"][ACNATA],Cluster.ACtype[ACNATA],Cluster.Date[ACNATA], Cluster.Date[ACNATA +
    ↪  np.sum((Cluster["Aircraftnumber"] == Cluster["Aircraftnumber"][ACNATA])&(Cluster["ATA
    ↪  location"]==Cluster["ATA location"][ACNATA]))-1 ],np.sum((Cluster["Aircraftnumber"] ==
    ↪  Cluster["Aircraftnumber"][ACNATA])&(Cluster["ATA location"]==Cluster["ATA
    ↪  location"][ACNATA])),(np.sum((Cluster["Aircraftnumber"] ==
    ↪  Cluster["Aircraftnumber"][ACNATA])&(Cluster["ATA location"]==Cluster["ATA
    ↪  location"][ACNATA]))-1) / diff_day(Cluster.Date[ACNATA],Cluster.Date[ACNATA +
    ↪  np.sum((Cluster["Aircraftnumber"] == Cluster["Aircraftnumber"][ACNATA])&(Cluster["ATA
    ↪  location"]==Cluster["ATA location"][ACNATA]))-1])],index = Primaries.columns),ignore_index
    ↪  = True)
    ACNATA += np.sum((Cluster["Aircraftnumber"] ==
    ↪  Cluster["Aircraftnumber"][ACNATA])&(pd.to_numeric(Cluster["ATA location"]) ==
    ↪  pd.to_numeric(Cluster["ATA location"][ACNATA])))
    print('done with number', ACNATA)
Primaries = Primaries.replace([np.inf, -np.inf], np.nan)

Primaries = Primaries[Primaries["lambda"].notna()]
Primaries["ATA_location"] =(Primaries["ATA_location"].astype(str).str[:2]).astype(int)
Primaries = Primaries.sort_values(["ATA_location","firstdate"])
Primaries.boxplot(column = ["lambda"], by = ["ATA_location"])
plt.xticks(rotation=90)

ATA_repair_performance = pd.DataFrame(columns = ["ATA","ACtype","removals","no_sub_removals","r"])
↪  #The chance for every ATA location that subsidiary removals occur, based on the dataset. This
↪  value for r is used later on.

for title0, group0 in Cluster.groupby("ACtype"):
    for title, group in group0.groupby("ATA location"):
        ATA_repair_performance =
        ↪  ATA_repair_performance.append(pd.Series([float(title),float(title0),
        ↪  float(group.shape[0]), group["Subsequent 1"].isna().sum(), 1- group["Subsequent
        ↪  1"].isna().sum()/group.shape[0]], index = ATA_repair_performance.columns),
        ↪  ignore_index = True)

Primaries["firstdate"] = pd.to_datetime(Primaries["firstdate"])
Primaries["finaldate"] = pd.to_datetime(Primaries["finaldate"])
```

Source Code 3.7: Finding and storing the primary removals

### Homogeneous removal rate

Next, a check is done whether or not the lambda value increases over time, thus leading to more time between removals. This hypothesis originates from the fact that the quality of components might increase over time due to new technologies or the use of better materials, and the fact that maintenance personnel might have more information related to possible errors related to the installation process and resulting in them making less mistakes. In order to do so, for each ATA location, the date of the first removals of all the clustered removals is plotted against its corresponding lambda value. If the plotted value of the lambda increases over time, it can be concluded that the MTBR improves over time. Although this is a gross assumption, since the slope of the fitted line is positive in 31 of the 34 cases, it can be concluded that there are factors in play that increase the $\lambda$, and thus decrease the MTBR over time. Furthermore, a new column `r_initial` is added to `Primaries`. In this column, the $r$ value, the same as in the BPP theory, is added to the DataFrame for each data point. It is assumed that the same ATA locations have equal values for `r_initial`, as the same components have the same level of repair difficulty.
Next the `lambda_adjusted` column is created. Here, the effect of the corresponding climate is taken into account and the lambda value is adjusted to the reference climate.
In the `locationcounter` DataFrame, the occurrences of a set of removals is counted and stored with its ATA location and aircraft type. Furthermore, the total amount of removals that occur in these occurrences is stored as well. Besides the amount of the occurrences of removals, the mean value for lambda (adjusted

to the reference climate) and the value for the parameter $r$, as used in the BPP theory, is provided per ATA location and aircraft type as well. Lastly some more data is left out of further consideration. Aircraft type 12 is removed from further analysis, as this aircraft type comprises too little data points. Furthermore, the total amount of removals for each ATA location for each aircraft type should exceed 20 in order to be considered in further analysis as well. The final variable that is declared in code 3.8, `atas`, stores the ATA locations that occur at least once in every aircraft type considered. See figure 3.3 for a graphical overview.

```
#%%
improvementcounter = 0
deteriorationcounter = 0

for title, group in Primaries.groupby('ATA_location'):
    group.plot(x='firstdate', y='lambda',style='.')
    x = mdates.date2num(Primaries["firstdate"][(Primaries["ATA_location"]).astype(float) ==
    ↪  int(title)])
    y = Primaries["lambda"][(Primaries["ATA_location"]).astype(float) == int(title)]

    RMSE = math.sqrt(np.sum((np.polyval(np.polyfit(x, y, 1), x) - y)**2)/len(x))
    slope, intercept, r_value, p_value, std_err = scipy.stats.linregress(x, y)
    plt.title("Overview of lambda over time of ATA location: {}".format(title))
    if slope < 0:
        plt.plot(x, slope*x + intercept,color = 'green', label = 'linear regression fit with an
        ↪  RMSE of %0.2f' %RMSE)
        improvementcounter +=1
    else:
        plt.plot(x, slope*x + intercept,color = 'red' ,label = 'linear regression fit with an RMSE
        ↪  of %0.2f' %RMSE)
        deteriorationcounter +=1
    plt.legend(loc = 'best')
Mbox('Results', 'For the %i ATA systems that were investigated, %i showed an improvement over time
↪  related to the MTBR, while %i showed deterioration'
↪  %(deteriorationcounter+improvementcounter,improvementcounter,deteriorationcounter), 1)

Primaries["r_initial"] = ""

for title0,group0 in Primaries.groupby("ACtype"):
    for title,group in group0.groupby("ATA_location"):
        group = group.assign(r_initial =
        ↪  float(ATA_repair_performance[(ATA_repair_performance["ATA"] == title) &
        ↪  (ATA_repair_performance["ACtype"] == title0)]["r"]))
        Primaries[(Primaries["ATA_location"]==int(title)) & (Primaries["ACtype"]==title0)] = group

Primaries["lambda_adjusted"] = ""
for i in range(len(Primaries)):
    Primaries.iloc[i,8] = Primaries.iloc[i,6] * np.mean(ExtOrg.loc[ExtOrg["AircraftSerialNumber"]
    ↪  == Primaries.iloc[i,0]]["Factor"])

locationcounter =
↪  Primaries.groupby(['ATA_location','ACtype']).size().reset_index().rename(columns={0:'occurrences'})
locationcounter["removals"] = ""
locationcounter["lambdamean"] = ""
locationcounter["r_initial"] = ""

for i in range(len(locationcounter)):
    locationcounter["removals"][i] = sum(Primaries[(Primaries["ATA_location"] ==
    ↪  locationcounter["ATA_location"][i]) & (Primaries["ACtype"] ==
    ↪  locationcounter["ACtype"][i])]["removals"])
    locationcounter["lambdamean"][i] = sum(Primaries[(Primaries["ATA_location"] ==
    ↪  locationcounter["ATA_location"][i]) & (Primaries["ACtype"] ==
    ↪  locationcounter["ACtype"][i])]["lambda_adjusted"] * Primaries[(Primaries["ATA_location"]
    ↪  == locationcounter["ATA_location"][i]) & (Primaries["ACtype"] ==
    ↪  locationcounter["ACtype"][i])]["removals"])/ locationcounter["removals"][i]
    locationcounter["r_initial"][i] = np.mean(Primaries[(Primaries["ATA_location"] ==
    ↪  locationcounter["ATA_location"][i]) & (Primaries["ACtype"] ==
    ↪  locationcounter["ACtype"][i])]["r_initial"])

locationcounter = locationcounter[locationcounter["ACtype"] != 12] # ACtype 12 left out of further
↪  analysis due to too little data points
locationcounter = locationcounter[locationcounter["removals"] > 20]
ATAcounter = Counter(list(locationcounter.ATA_location))
atas = [k for k, v in ATAcounter.items() if v == len(np.unique(locationcounter.ACtype))] #count
↪  the number of ATA chapters involved in the analysis
```

Source Code 3.8: Progress of MTBR over time

An example of one of the ATA locations is given in figure 3.2. Clearly, the fitted first order polynomial has

| ATA_location | ACtype | occurrences | removals | lambdamean | r_initial |
|---|---|---|---|---|---|
| 21 | 1 | 56 | 290 | 0.03644757646803493 | 0.11392405063291138 |
| 21 | 2 | 85 | 1465 | 0.10357794622279731 | 0.20226969292389882 |
| 21 | 3 | 39 | 1095 | 0.052828579478834285 | 0.21004566210045647 |
| 22 | 1 | 31 | 111 | 0.005233628155813836 | 0.1417322834645668 |
| 22 | 2 | 89 | 1703 | 0.04630855884333103 | 0.168126094570928 |
| 22 | 3 | 39 | 1051 | 0.028115107760627894 | 0.28544243577545164 |
| 23 | 1 | 68 | 851 | 0.04857344581881648 | 0.13821138211382136 |
| 23 | 2 | 86 | 1931 | 0.09918972645494244 | 0.13021101389603726 |
| 23 | 3 | 39 | 1247 | 0.05657678330076932 | 0.11547714514835594 |
| 24 | 1 | 67 | 451 | 0.04989797081311786 | 0.2623655913978492 |

Table 3.3: Small selection of the `locationcounter` DataFrame

a positive slope, corresponding to an increasing $\lambda$, and thus a decreasing MTBR over time. This implies the use of a Non-Homogeneous Poisson Process (NHPP) over an Homogeneous Poisson Process (HPP) to model the MTBR. However, in the basic form of the model explained in chapter 4, the HPP is used. This is backed by the finding that for the vast majority of considered data points, there is no reason to deviate from the null hypothesis that the removals can be considered without any trends. The Mann-Kendall Test (Mann (1945),Kendall (1975)) is used to statistically assess whether or not there is a upward or downward trend in the underlying data. This is done for all intervals between the removals for the different ATA locations for each considered aircraft. The analysis is provided in code 3.9. The ATA locations taken into consideration for this analysis are derived from table 3.6. The method that is used, `mk_original_test`, is built by Hussain and Mahmud (2019). The `mk_original_test` is part of the *pyMannKendall* project, which is an open source project, maintained by publicly funded academic researchers. The method derives the trend for the specific subset of data that is provided in `HPPtest2`. In `HPPtest2`, the intervals (in days) between the primary removals is calculated. The `HPPcheck` function is used for this and also provided in code 3.9.

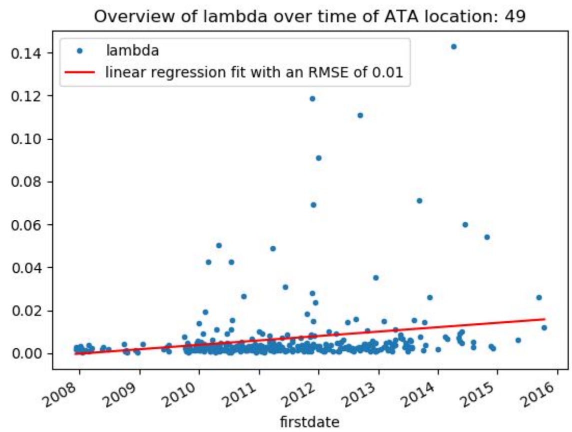



Figure 3.2: Example of plot showing detoriated MTBR over time

```
HPPtest = {}
HPPtest2 = {}
HPPtest3 = {}
for title,group in Cluster.groupby('Aircraftnumber'):
    for title2, group2 in group.groupby('ATA location'):
        if len(group2.Date) > 5 and title2 in [23,24,27,28,29,32,34,77]:
            HPPtest[int(float(title)),title2] = group2.Date
            HPPtest2[int(float(title)),title2] = HPPcheck(HPPtest[int(float(title)),title2])
            HPPtest3[int(float(title)),title2] =
             ↪  mk.original_test(HPPtest2[int(float(title)),title2]).trend

print('In', sum(value == 'no trend' for value in HPPtest3.values())/len(HPPtest3)*100,"% of the
 ↪  occassions, there is no trend in the time between primary failures")
print('In', sum(value == 'increasing' for value in HPPtest3.values())/len(HPPtest3)*100,"% of the
 ↪  occassions, there is an increasing trend in the time between primary failures")
print('In', sum(value == 'decreasing' for value in HPPtest3.values())/len(HPPtest3)*100,"% of the
 ↪  occassions, there is a decreasing trend in the time between primary failures")
```

Source Code 3.9: Algorithm checking for the distribution of the primary removals

The result is provided in table 3.4

| No Trend | Increasing Trend | Decreasing Trend |
|---|---|---|
| 92.86% | 2.44% | 4.70% |

Table 3.4: Results of the trend analysis

## Fleet composition

As can be seen in table 3.3, the aircraft type (either 1, 2 or 3) is provided in the data set. These numbers represent three different type of aircraft with different components. Hence, it is checked whether the behavior of similar components differs among the types of aircraft. For this, the Kruskal-Wallis H-test is used (see section 2.3). For the different aircraft, it is tested whether or not the lambda and $r$ values differ significantly. In the case that there is a significant difference, the model has to take the effect of different fleet compositions into account. In the other case, it can be assumed that the components from the different aircraft types behave similar, resulting in an aggregation of the data across the different types of aircraft. The code is provided in code 3.10. The outcome of the KW test is given in table 3.5. A Confidence Interval (CI) of 95% is chosen here. The null hypothesis is that the $r$ and lambda value for the different aircraft have equal medians. As both $H_0$ hypothesis are validated, it can be concluded that the components in the different aircraft have similar values for lambda and $r$. Hence, there is no need in the model to differentiate for aircraft type, and thus can the fleets be represented by the average value of the three different types combined. The ATA locations that are selected in the code are further motivated in section 3.5.

```
locationcounter2 =
 ↪  locationcounter[locationcounter["ATA_location"].isin([23,24,27,28,29,32,34,77])].
 ↪  reset_index(drop=True)

print(stats.kruskal(locationcounter2[locationcounter2.ACtype == 1]["r_initial"],
 ↪  locationcounter2[locationcounter2.ACtype == 2]["r_initial"],
 ↪  locationcounter2[locationcounter2.ACtype == 3]["r_initial"]))
print(stats.kruskal(locationcounter2[locationcounter2.ACtype == 1]["lambdamean"],
 ↪  locationcounter2[locationcounter2.ACtype == 2]["lambdamean"],
 ↪  locationcounter2[locationcounter2.ACtype == 3]["lambdamean"]))
```

Source Code 3.10: Kruskal-Wallis H-test for the different aircraft types

| | **statistic-value** | **p-value** | $H_0$ **rejected** |
|---|---|---|---|
| **lambda** | 0.215 | 0.898 | No |
| $r$ | 0.215 | 0.898 | No |

Table 3.5: Aircraft type similarity KW test results

## 3.5. Component scope

In order to bring focus to the components that have the most impact on the added value of this research, it is important to narrow the scope of the components in this research. As provided in table 3.6, a total of 34 different ATA chapters have been identified from the data set. All these groups of components will be tested on different criteria, which will lead to a smaller selection of components that allow this research to set its scope to. For the breakdown of these criteria, the paper of Thijssens and Verhagen (2020) is used as a reference. The implementation in Python for the selection of ATA locations is given in code 3.11

### Availability of data points

In order to meet statistical standards and ensure representative values for further calculations, the size of the subset of each ATA chapter should be large enough to support the found values from each subset. Furthermore, each ATA location taken into further consideration should appear in each aircraft type that is taken into account. The threshold is set to 1000 data points here.

### Occurrences of subsequent failures

The higher the amount of removals after the data cleansing, the more reliable the estimates of parameters for the BPP are. This results in the priority of removals from prominent ATA chapters from the dataset. Since most of the time low numbers will be the result of few data points or theoretically high correct repair rates, no lower boundary for this value is set in order to keep it into consideration.

### Flight Safety-Critical

As defined by the Federal Aviation Administration (FAA), Flight Safety Critical Aircraft Parts can be defined as "*parts, assemblies, or installations containing a critical characteristic whose failure, malfunction or absence could cause a catastrophic failure resulting in loss of serious damage to the aircraft on an uncommanded engine shutdown resulting in an unsafe condition*" (FAA, 2010). Therefore, the ATA locations where many of these type of parts are placed, are top priority when it comes to better understanding the timing of these failures. Therefore, ATA locations that have few components that are flight safety-critical are left out of further consideration.

```python
atas = [23,24,27,28,29,32,34,77]
for key in list(offset2.keys()):
    if key not in atas:
        del offset2[key]
```

Source Code 3.11: Selection of ATA locations

| ATA chapter | System | Data points | Occurrences with subsidiary failures | average MTBR of primary failures [days] | Removals in all aircraft | Total unique aircraft | Flight Safety-Critical | Selected |
|---|---|---|---|---|---|---|---|---|
| 21 | Air Conditioning | 2909 | 489 | 131 | Yes | 182 | No | No |
| 22 | Auto Flight | 2892 | 332 | 118 | Yes | 159 | No | No |
| 23 | Communications | 4053 | 595 | 131 | Yes | 195 | Yes | Yes |
| 24 | Electrical Power | 1697 | 397 | 60 | Yes | 192 | Yes | Yes |
| 25 | Equipment / Furnishings | 3483 | 342 | 159 | Yes | 166 | No | No |
| 26 | Fire Protection | 447 | 117 | 112 | Yes | 112 | No | No |
| 27 | Flight Controls | 1761 | 419 | 131 | Yes | 161 | Yes | Yes |
| 28 | Fuel | 1773 | 318 | 164 | Yes | 175 | Yes | Yes |
| 29 | Hydraulic Power | 1290 | 283 | 130 | Yes | 168 | Yes | Yes |
| 30 | Ice & Rain Protection | 1597 | 365 | 153 | Yes | 177 | No | No |
| 31 | Indicating / Recording System | 2002 | 431 | 106 | Yes | 185 | No | No |
| 32 | Landing Gear | 9195 | 448 | 127 | Yes | 217 | Yes | Yes |
| 33 | Lights | 2971 | 416 | 125 | Yes | 189 | No | No |
| 34 | Navigation | 8233 | 982 | 85 | Yes | 225 | Yes | Yes |
| 35 | Oxygen | 4355 | 294 | 155 | Yes | 176 | No | No |
| 36 | Pneumatic | 1110 | 184 | 125 | Yes | 170 | No | No |
| 38 | Water / Waste | 803 | 154 | 107 | Yes | 108 | No | No |
| 49 | Airborne Auxiliary Power | 1533 | 349 | 149 | No | 108 | No | No |
| 52 | Doors | 328 | 73 | 180 | Yes | 68 | No | No |
| 53 | Fuselage | 174 | 47 | 154 | No | 48 | No | No |
| 55 | Stabilizers | 157 | 41 | 193 | No | 41 | No | No |
| 56 | Windows | 666 | 122 | 143 | Yes | 122 | Yes | No |
| 57 | Wings | 304 | 60 | 99 | No | 59 | Yes | No |
| 61 | Propellers / Propulsion | 498 | 98 | 153 | No | 63 | Yes | No |
| 71 | Power Plant General | 345 | 87 | 127 | Yes | 91 | Yes | No |
| 72 | Engine Turbine/Turboprop, Ducted Fan/Unducted Fan | 256 | 64 | 123 | Yes | 72 | Yes | No |
| 73 | Engine Fuel & Control | 868 | 191 | 112 | Yes | 135 | Yes | No |
| 74 | Ignition | 104 | 22 | 204 | No | 24 | Yes | No |
| 75 | Air | 142 | 33 | 133 | No | 34 | Yes | No |
| 76 | Engine Controls | 146 | 37 | 233 | Yes | 37 | Yes | No |
| 77 | Engine Indicating | 1088 | 193 | 149 | Yes | 160 | Yes | Yes |
| 78 | Exhaust | 399 | 107 | 252 | Yes | 107 | No | No |
| 79 | Oil | 161 | 34 | 156 | No | 35 | Yes | No |
| 80 | Starting | 575 | 132 | 107 | Yes | 132 | Yes | No |

Table 3.6: Component trade-off on ATA chapter level

# 4

# Model implementation

This chapter comprises the working of the model. First, the main assumptions leading to simplifications are summed up. These simplifications are made in order to limit the scope of the research and the workability of the model. Furthermore, the combination of substantiations of the assumptions provides the reader of a clear motivation that the limitations on account of these assumptions are limited.
Subsequently, the methodology of the model is explained.

## 4.1. Assumptions

In the model, some general assumptions are made. This section provides an overview of these assumptions, substantiated with the motivation for each individual assumption.

### 4.1.1. Rates of failure and removal

The provided data set comprises removal data of components over multiple decades. However, this research is focused on the failures rather the removals, and thus a conversion has to be made. Although it could be presumed that this conversion cannot be made, it is relatively simple. In some cases, components are replaced after the outcome of an inspection resulted in an perception that a certain safety limit is exceeded. In other cases, the component is replaced after it is found to have failed due to, for example, redundancies. As there are cases where removals take place on a preventive basis and some removals take place after the failure, the removal of a component is spread over time around the actual failure. For this reason, it is assumed in this research that the average rate of removal is equal to the average rate of failure.

### 4.1.2. From failure to demand, replacement and repair

Another time-related assumption for the current model is the direct influence of a failure to the demand of the spare part(s) that have to replace the broken component. In many cases, it is be the case that there is a (small) time difference in the moment that the failure is detected and the moment it is replaced. In a best case scenario, the failure is detected shortly after it occurs, and quickly replaced after the flight. However, multiple reasons for a delay in this process exist. Possibilities for this are the unavailability of the component at the airport of arrival or the fact that the failure is detected later. In this research however, it is assumed that when a component fails, it is detected and the demand of the spare component is directly generated.

### 4.1.3. Interdependencies of components

The theory of the BPP is based on subsequent failures that occur for much the same reason as the original failure. This implies that the failed components have some sort of relationship regarding its function in the whole machine. In the case of this research, this means that the components are somehow related with regard of their function for the specific aircraft. As airplane manufacturers build different aircraft, there is not one default build up. On component level, it might result in different subsets of interdependent components. One assumption that is made in this research is that the interdependency of different components is based on their ATA system location. This means that the first two-digit combination of the ATA location code is decisive whether or not components are interdependent.
A more detailed analysis can lead to the rejection of certain interdependencies and furthermore bring new

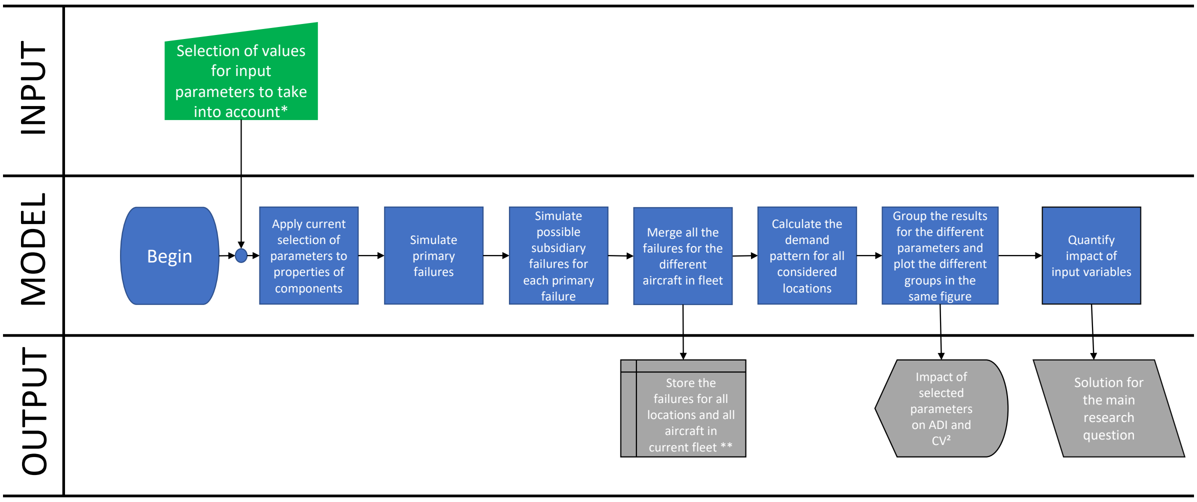
interdependencies to light. This would fortify the outcome of the current model. When this knowledge is on hand, it can easily be implemented in the model. This might be interesting for aircraft manufacturers, as these companies know the best how their aircraft are constructed.

### 4.1.4. Consequence of an incorrect repair

Incorrect repairs do not necessarily trigger subsequent failures. For example, although the repair might be done incorrectly and the component is not placed properly to its original position, it might still function as desired. However, in this research, the theory of the BPP is followed. Hence, each incorrect repair and placement is assumed to spawn a finite number of subsidiary failures.

## 4.2. Methodology of the model

The output of the model should provide an answer to the research question. That is, the quantitative influence of incorrect repairs on the two demand pattern characteristics, the squared coefficient of variance ($CV^2$) and the average demand interval (ADI) should be clear. In order to incorporate these incorrect repairs, the parameter $r$, the chance of incorrect repairs, is implemented in the model. This parameter influences the discrete random variable that represents the spawn of subsequent failures. Hence, when an incorrect repair takes place and the component is placed back into service, the number of failures during a relative small timespan will peak due to the number of subsequent failures. The influence of this parameter on the $CV^2$ and ADI is the core attribute of this model.
Besides the main parameter $r$, additional parameters that influence the pattern of the demand of spare parts are taken into account as an extension. The work of Lowas and Ciarallo (2016) showed that smaller fleet size leads to higher $CV^2$ and ADI compared to larger fleets. Thijssens and Verhagen (2020) showed that the humidity of the environment impacts the reliability of components for multiple different reasons. Air pollutants, salinity and the salt content in the atmosphere all have an impact on the corrosion process of components. Besides a natural reference climate, humid and desert climates are taken into account as well, both effecting the Mean Time Between Failure (MTBF). The impact of the incorrect repairs in combination with these other varying circumstances provides a wider view on the impact in general.
For every aircraft in the selected fleet, failures at the selected ATA locations (see section 3.5) are simulated according to their corresponding failure rate $\lambda$, obtained from the analysis of the data set. Based on the number of primary removals that contain subsequent removals in the data set, an estimate of the probability of an incorrect repair $r$ can be made for the specific component location. Subsequently, possible subsequent failures are simulated.
Then, the results of the individual aircraft are summed, resulting in the sum of the failures over time for every ATA that is selected to be part of the model. This is done for multiple combinations of parameters. The values for the ADI and $CV^2$ of the failures are plotted in a coordinate system such that the impact is visualized. A visualization of the working of the model is provided in figure 4.1. The above-mentioned metrics describe the predictability of the failures over time, but do not provide an answer to the quantity of failures. Therefore, this metric is added to the results as well, in order to capture both the behavior as the sum of the failures.
The earlier mentioned tests (section 2.3) are used in order to conclude whether or not the impact of the varying parameters is statistically significant.




* Incorporates the different demand drivers that are taken into account as well as the values for lambda and r
** For every component, each failure in each aircraft is stored such that the time of the failure can be obtained later on

Figure 4.1: Flowchart of the working of the model

## 4.3. Parameter selection

The different values for the parameters are provided in table 4.1. These values are further elaborated in the following subsections.

| Variable | Tested values | Number of steps |
|---|---|---|
| ***r* factor** | 0.0 0.5 1 2 | 4 |
| **Environmental factors** | Natural Humid Desert | 3 |
| **Fleet size** | 8 16 32 64 96 128 (256) (512) | 6 (8) |
| **Component commonality** | Not present Present | 2 |

Table 4.1: Monte Carlo simulation parameters

### Incorrect repair rate

The main goal of this research is to reveal the impact of the quality of the repair process for components that are placed back into the aircraft on the $CV^2$ and ADI of the demands of the spare parts. As the initial value of $r$ is retrieved from the data analysis of the dataset, this value is used as reference value. Scenarios with a 100% increased, a 50% decreased and a 100% decreased value for $r$ are tested. The first mentioned alteration of $r$ represents a scenario in which the amount of incorrect repaired components that is placed back into service is twice as high as the reference scenario.
The second alteration represents the scenario when half of the incorrect repaired components are placed back into service.
The latest option represents the scenario where no incorrect repaired components are placed back onto the aircraft, and thus that all components that are placed back function properly.

### Environmental factors

The work of Thijssens and Verhagen (2020) showed the impact of three environmental factors to the Restricted Mean Survival Time (RMST) of components in aviation. The RMST is equal to the mean survival time, except that the RMST allows longer observations to be sensored without preventing the survival function from never dropping to zero. In this research, the impact of the environmental factors is directly related to the MTBF of components by the numerical factor that is provided in table 4.2. For every aircraft considered in the analysis, the airline can be traced back via the external organization code. On this way, the dominant environmental conditions can be applied and the values for the specific aircraft can be adjusted.

| Natural climate | MTBF-ratio |
|---|---|
| Reference | 1 |
| Humid | 0.62 |
| Desert | 0.73 |

Source: Thijssens and Verhagen (2020)

Table 4.2: Environmental factors

### Fleet size and composition

The study of Lowas and Ciarallo (2016) provided insights into reasons for lumpy spare parts demands. The study found that the parameter with the largest impact of the lumpiness of the demands of spare parts was the fleet size. In order to test this finding and to extend its scope, it is tested in this research as well. As the study of Lowas and Ciarallo (2016) writes a clear motivation for the range of selected values of the fleet size, this is not further motivated in this study, as the same range of values are chosen for the model.

According to Air Transport World (2019), the top ten largest airlines, based on their fleet size, are given in table 4.3. For each of these airlines, the most used aircraft type is given. Data is obtained from Planespotters.net (2020). One of the things that stand out of table 4.3 is the fact that both LCCs have fleets with high

| Airline | Fleet size | Most common type | Amount of most common type | Comment |
|---|---|---|---|---|
| American Airlines | 875 | Airbus A321 | 240 | - |
| Delta Airlines | 835 | Boeing 737 | 208 | - |
| United Airlines | 804 | Boeing 737 | 352 | - |
| Southwest Airlines | 735 | Boeing 737 | 735 | Low-Cost Carrier (LCC) |
| China Southern Airlines | 609 | Boeing 737 | 213 | - |
| China Eastern Airlines | 567 | Airbus A320 | 220 | - |
| SkyWest | 536 | Bombardier CRJ-100 | 197 | - |
| FedEx Express | 450 | Boeing 757 | 119 | Freight |
| Ryanair | 271 | Boeing 737 | 200 | LCC |

Sources: Air Transport World (2019), Planespotters.net (2020)

Table 4.3: Overview of the fleet composition of top ten airlines based on fleet size

commonality when it comes to aircraft type. As can be seen from the table, Southwest Airlines, the largest LCC, is the leader when it comes to the use of a single aircraft type in its fleet. Multiple studies claim that a more diverse fleet results in higher costs across multiple disciplines, among which maintenance is one (West and Bradley, 2008). The study of Lowas and Ciarallo (2016) shows this as well. For a fleet size larger than 128, the variations in demand, both in ADI and $CV^2$, minimizes and approach their theoretical minimum. Hence, this proves that the statement that one of the reasons for LCCs to use fleet commonality is to decrease the costs related to maintenance, as spare parts demands become more predictable, leading to lower safety boundaries of stocks of spare parts. Belobaba et al. (2015) also backs this statement.
However, on the other side of the market, flag carriers such as KLM, United Airlines and Delta Airlines, have a much more diversified fleet. An overview of the fleets of these carriers is given in table 4.4. As can be seen, medium-sized flag carriers such as KLM only have 52 Boeing 737s, which is their most common used aircraft. For the larger American flag carriers it can be seen that they have even more aircraft types in their fleet comprising not more than 100 aircraft. Therefore, the focus of this research will be on these (smaller) amounts of aircraft, as the striking problem of lumpy demand patterns is more significant. Furthermore, a diversified fleet with different aircraft rises another extension to the model. As research of Zhang et al. (2019) showed that an increase of component commonality across aircraft types is beneficial for the reduction of the costs of spare parts, this might drive the industry to further investigate other outcomes of the implementation of increased component commonality. The final extension of the model investigated the influence on the demand of spare parts when different aircraft types used the same components. As no sound data of the performance difference among the aircraft types is provided, an estimation for the variance in performance is used for this extension. This is further motivated in section 4.4. Therefore, the final selection of fleet sizes will contain the values shown in table 4.1. As table 4.4 indicates that multiple airlines also make use of a small amount of a certain aircraft type, a fleet size of 16 is added to this research. This value was not taken into consideration in the work of Lowas and Ciarallo (2016). The values in parenthesis are checked on their smoothness. If both val-

| Aircraft Type | In fleet |
|---|---|
| Airbus A330 | 13 |
| Boeing 737 | 52 |
| Boeing 747 | 7 |
| Boeing 777 | 29 |
| Boeing 787 | 18 |

| Aircraft Type | In fleet |
|---|---|
| Airbus A319 | 86 |
| Airbus A320 | 97 |
| Boeing 737 | 352 |
| Boeing 757 | 61 |
| Boeing 767 | 54 |
| Boeing 777 | 96 |
| Boeing 787 | 60 |

| Aircraft Type | In fleet |
|---|---|
| Airbus A220 | 31 |
| Airbus A319 | 57 |
| Airbus A320 | 62 |
| Airbus A321 | 103 |
| Airbus A330 | 49 |
| Airbus A350 XWB | 15 |
| Boeing 717 | 91 |
| Boeing 737 | 208 |
| Boeing 757 | 127 |
| Boeing 767 | 75 |
| Boeing 777 | 17 |

Source: Planespotters.net (2020)

Table 4.4: Overview of the fleet compositions KLM (left), United Airlines (middle), and Delta Airlines (right)

ues indeed represent mainly smooth demand patterns, these fleet sizes are left out of further consideration, as the unpredictability of these demands is not a striking problem in the industry.

## 4.4. Scenario selection

In order to maximize the applicability and to approximate the reality as close as possible, multiple scenarios are worked out and simulated in order to obtain the effect of changing parameters on the demand characteristics. The tested scenarios are described below. The selection of each of the scenarios is motivated and substantiated. The scenarios are divided into four main scenarios, each with its own sub-scenarios. This is done in order to maintain the main focus on the impact of the incorrect repairs on the demand characteristics of the spare parts. The sub-scenarios dive deeper into the impact of the other specified drivers, provided in table 4.1.

### 4.4.1. Variant 1 - Base

This scenario will be used as reference simulation, as the level of correctness of the repairs is the only varying parameter. In this simulation, the reference value for the weather conditions is taken into account. Furthermore, all fleet sizes are taken into account, but no separation of results is provided and all data is aggregated. This simulation should answer the main research question, but leaves out any extensions regarding varying fleet size or weather conditions. Regarding the provided algorithm (see algorithm 2), the first two **for** statements comprise only one value. Hence, for this simulation, these loops do not add depth to the model.

### 4.4.2. Variant 2 - Extension: Incorporating different fleet sizes

The work of Lowas and Ciarallo (2016) proved that, for the investigated parameters in their model, the fleet size was the largest contribution to the lumpiness of the demands of spare parts. An aggregation of these influences together with the different rates of incorrect repairs form the first extension to the model. With this extension, it is possible to specify the effect of incorrect repairs on different fleet sizes.

### 4.4.3. Variant 3 - Extension: Incorporating different fleet sizes and environmental conditions

The findings of Thijssens and Verhagen (2020) form the underlying motive for this extension. As that work showed the effect of environmental conditions on the MTBF, an aggregation of these influences together with the rates of incorrect repairs and the varying fleet sizes provides a synergy when it comes to accuracy of the quantification of the influence of the incorrect repairs. From the perspective of MRO providers, this leads to insights in the effect of incorrect repairs for specific combinations of fleet size and environmental factors.

### 4.4.4. Variant 4 - Extension: Incorporating different fleet sizes, environmental conditions and higher component commonality

As can be seen in table 4.4, flag carriers tend to diversify their fleet composition. Main reasons for that are stated in Belobaba et al. (2015) and are not discussed here. Different aircraft have different types of components and thus different failure rates. However, research has shown that component commonality can lead to a reduction of the operating costs or airlines (Zhang et al., 2019). Therefore, airlines and MRO providers might adopt this into their business models. This results in components that are used in different types of aircraft. As different aircraft perform differently when it comes the use of components, it is considerable that components have different lifetimes on different aircraft types. When differentiating the performance of the aircraft artificially by a margin of 20%, different performing aircraft can be simulated. This forms the basis for this extension of the model.

## 4.5. Implementation in Python

The implementation of the model in Python is provided in appendix TOEWIJZEN . For a rough overview of the model, the pseudo code can be found in algorithm 2 and a graphical overview is displayed in figure 4.1.

**Algorithm 2:** Pseudocode of model

```
input  : All possible different values for r, weather conditions, fleet sizes and component
         commonality strategies
output: The CV^2 and ADI of the demands of spare parts and the total number of failures

for all different fleet sizes do
    for all different weather conditions do
        for all aircraft in fleet do
            for all different repair qualities do
                for all ATA chapters do
                    Simulate primary failures
                end
                for all simulated primary failures do
                    simulate probability and occurrence of subsidiary failures and apply
                end
            end
            add the failures of the selected aircraft to the matrix with results ;
            for all ATA locations do
                Calculate the CV^2 and the ADI of the demand ;
                Add the fleet size, ATA location, demand pattern variables, weather conditions,
                  component commonality strategy and repair quality in the results
            end
        end
    end
end
```

# 5
# Simulations

In this chapter the results of the different variants are presented. For each variant, both graphical as quantitative support is given for the results. Furthermore, it is clearly mentioned what values for the specific parameters are selected. As mentioned in section 2.2, the impact will be captured in the ADI, $CV^2$ and the number of failures over time. For the stabilization of the outcome of the model, 50 iterations are applied to the model. This is done to take the effect of randomness out of play. Randomness finds its way into the model in the following ways:

- In the spawn of primary failures. As the time of occurrence of a series of primary removals is created by a Poisson distribution with rate $\lambda$ corresponding to the properties in play, this varies the outcome;
- The chance of subsidiary failures is based on a test whether or not the corresponding $r$ value is smaller or larger then a random number between 0 and 1;
- The distribution of subsidiary failures over the given interval in which they occur after a primary failure is chosen randomly;
- In variant 4, where the fleet is composed of different performing aircraft types. Here, the performance of each aircraft is chosen randomly from the set of possibilities.

The figures in this chapter display the results of only one iteration, as otherwise too many data points are displayed in the figure, resulting in too many overlaps. However, the data in the tables takes all the iterations into consideration. For the statistical tests, the KW test and the Mann-Whitney $U$-test, as well as the comparison of the $CV^2$ and ADI, the results of all iterations are combined before they are compared.
Chapter 6 dives deeper into the evaluation of the results. In all figures that provide the reader with the visualization of data points, a data point represents the performance of a single ATA location.

## 5.1. Overview of simulated parameter values

Table 4.1 presents the values to take into consideration. However, as the study of Lowas and Ciarallo (2016) indicates, the variations in demand of spare parts of fleet sizes above 128 minimize and the demand patterns become smooth. To validate this for this research and to check whether to take these values into further consideration, a variant containing these specific fleet sizes is used. The results are given in figure 5.1. From this figure it can be concluded that the ADI edges its lower boundary for a fleet size of 512. Furthermore, the $CV^2$ reaches a maximum value of 0.3 for this fleet size. For a fleet size of 256, only three points have an ADI higher than 1.32, while the $CV^2$ of all points never exceed the value of 0.40.
For a fleet size of 128, the results are spread. Although the value of the $CV^2$ stays below the critical value of 0.49 before it becomes lumpy, the ADI varies widely. Most of the data points are considered to be intermittent. Hence, the fleet sizes of 256 and 512 are left out of further consideration in this research.

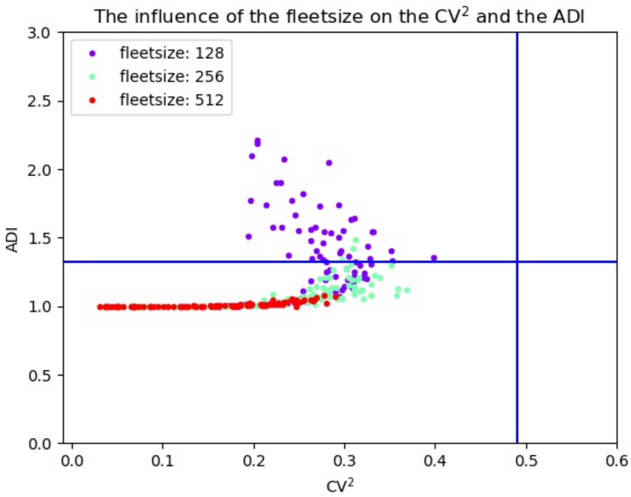



Figure 5.1: Failure patterns for large fleets

## 5.2. Results of variant 1

In the base variant, no environmental effects or selections of fleet size and compositions is considered. Instead, only the reference environmental condition, as defined by Thijssens and Verhagen (2020), and an aggregation of the data for fleet sizes, is considered. The graphical results are shown in figure 5.2, 5.3 and 5.4. As can be seen in figure 5.2, all points represent either intermittent or smooth demand. Although figure 5.2 suggests that most of the data points near the bound between intermittent and smooth demand are corresponding to the 'worse' repair quality, this is not the case. When zooming in on this area (figure 5.3 and 5.4), it is clear that the data points of this cluster are a mix of all repair qualities. The reason for this graphical deception finds its origin in the way the points are plotted into the figure. The set of points corresponding to the 'worse' repair quality conditions are plotted in the figure after the other three sets are, thence they are overlapping some other data points.

The visual results displayed in figure 5.2 and 5.4 do not directly provide a conclusion. For every set of points

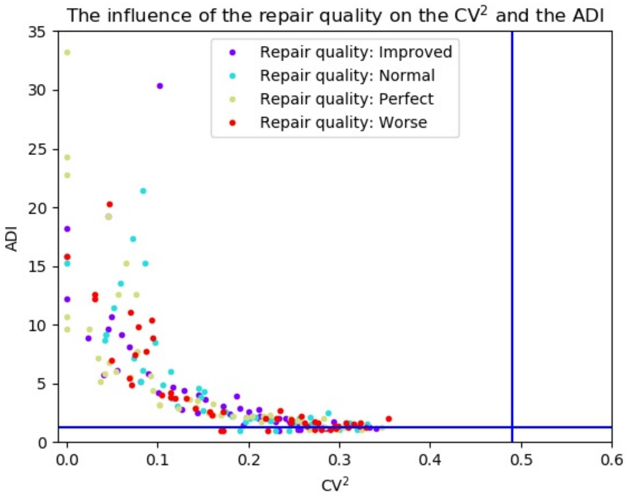



Figure 5.2: Visual results of variant 1

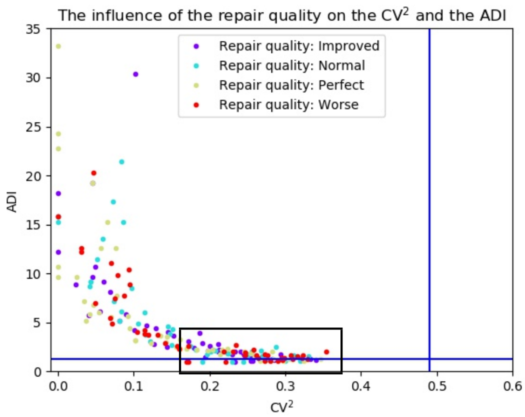



Figure 5.3: Zoomed area for figure 5.4

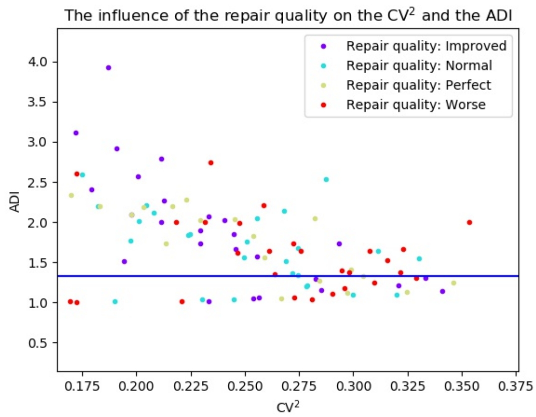



Figure 5.4: Overview of the boxed area of figure 5.3

corresponding to a certain value of the tested parameter (in this case the repair quality), the value of both the $CV^2$ and the ADI are listed. This results in eight lists (four repair qualities, two performance metrics). For all possible combinations of repair qualities, both the ADI as well as the $CV^2$ values are directly compared with each other. Each value in these lists represents the result (either ADI or $CV^2$) of an ATA location for a certain fleet size. As eight ATA locations and six fleet sizes are considered, a total of 48 data points is captured in every

list.
By comparing the ADI and $CV^2$ of the different repair qualities one-on-one, a relative comparison can be made. By dividing all the numbers of a certain scenario by the reference scenario, the relative gain or loss in ADI and $CV^2$ can be measured for each ATA location. Then, by taking the mean value of the whole array, the difference for both the ADI and $CV^2$ can be expressed. This is quantified in tables 5.1 and 5.2. In both tables, the row headers represent the reference scenario, where the column headers represent the alternative scenario. The following example is provided for the understanding of the reader: The average ADI and the $CV^2$ increased by 15.1% and decreased by 3.0% respectively, when the repair quality improves from "Worse" to "Normal". When checking for similarity across the different repair qualities, the Mann-Whitney U-test is used. The $p$-value of this test is provided in the tables 5.3 and 5.4. The $p$-value related the the null hypothesis of the KW test is $\mathbf{2.951 \cdot 10^{-5}}$ for the groups representing the ADI and **0.019** for the groups representing the $CV^2$ of the outcome.

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 1.000 | 1.151 | 1.278 | 1.462 |
| **Normal** | 0.869 | 1.000 | 1.126 | 1.283 |
| **Improved** | 0.782 | 0.888 | 1.000 | 1.152 |
| **Perfect** | 0.684 | 0.779 | 0.868 | 1.000 |

Table 5.1: V1: Overview of the influence on the average ADI for changing repair qualities

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 0.500 | 0.008 | 0.000 | 0.000 |
| **Normal** | 0.008 | 0.500 | 0.029 | 0.000 |
| **Improved** | 0.000 | 0.029 | 0.500 | 0.015 |
| **Perfect** | 0.000 | 0.000 | 0.015 | 0.500 |

Table 5.3: V1: Overview of the $p$-values of the Mann-Whitney U-test for ADI of the different scenarios

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 1.000 | 0.970 | 0.916 | 0.842 |
| **Normal** | 1.031 | 1.000 | 1.011 | 0.926 |
| **Improved** | 1.091 | 0.989 | 1.000 | 0.961 |
| **Perfect** | 1.187 | 1.079 | 1.041 | 1.000 |

Table 5.2: V1: Overview of the influence on the average $CV^2$ for changing repair qualities

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 0.500 | 0.010 | 0.000 | 0.000 |
| **Normal** | 0.010 | 0.500 | 0.056 | 0.000 |
| **Improved** | 0.000 | 0.056 | 0.500 | 0.003 |
| **Perfect** | 0.000 | 0.000 | 0.003 | 0.500 |

Table 5.4: V1: Overview of the $p$-values of the Mann-Whitney U-test for $CV^2$ of the different scenarios

Looking at the total number of failures over time, table 5.5 shows the impact of the changing repair quality.

- An improvement from "Worse" to "Normal" repair quality reduces the number of failures by 17.5%;
- An improvement from "Normal" to "Improved" repair quality reduces the number of failures by 15.6%;
- An improvement from "Improved" to "Perfect" repair quality reduces the number of failures by 20.1%.

The opposite occurs for deteriorating repair qualities; the number of failures increases.

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Failures** | 18329 | 15115 | 12763 | 10200 |

Table 5.5: V1: average number of failures for the different repair qualities per iteration

## 5.3. Results of variant 2

As the first variant aggregated all the data related to the fleet size, this variant shows the impact of the fleet size on the failure patterns, as well as the influence of the incorrect repair rate for every fleet size. Clearly, from figures 5.5 - 5.10 it becomes clear that the ADI decreases as the fleet sizes increases. Furthermore, an increasing fleet size tend to increase the $CV^2$ of the failures. As it is hard to directly see the impact of the

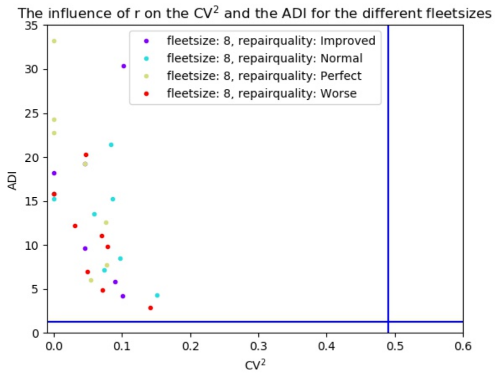



Figure 5.5: V2: Visual results - fleet size 8

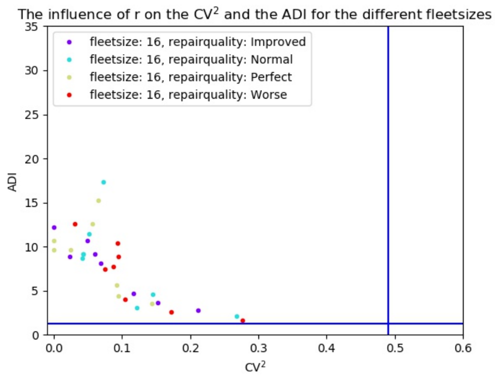



Figure 5.6: V2: Visual results - fleet size 16

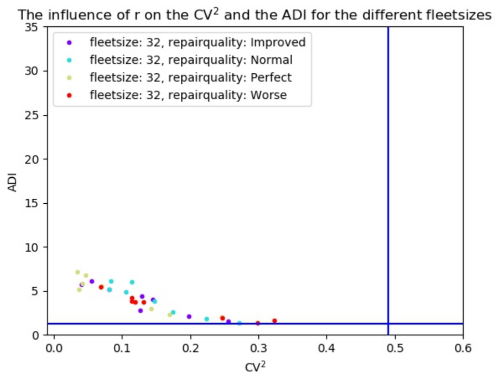



Figure 5.7: V2: Visual results - fleet size 32

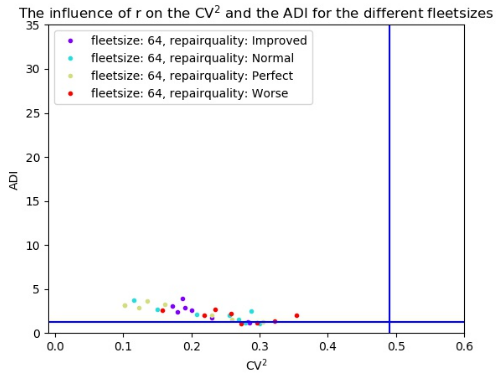



Figure 5.8: V2: Visual results - fleet size 64

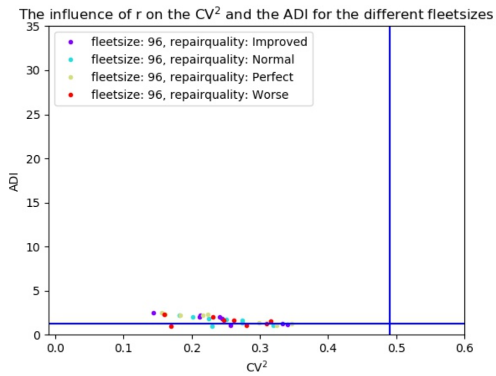



Figure 5.9: V2: Visual results - fleet size 96

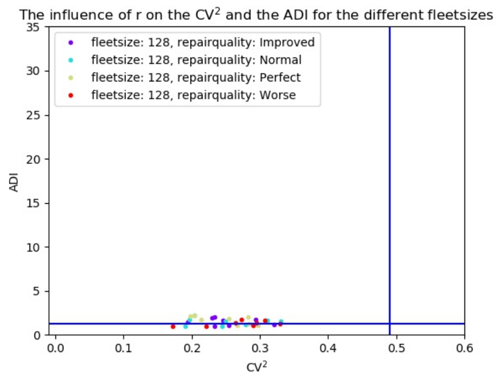



Figure 5.10: V2: Visual results - fleet size 128

varying repair quality on the two key metrics, tables 5.6 and 5.7 provide a better insight. Both tables are a relative comparison to the reference scenario (where the "Normal" repair quality is in play).
For every flight size, a Kruskal-Wallis $H$-test is performed that contains the information regarding the potential rejection of the null hypothesis and thus the significant differences among the different repair qualities for the given fleet size regarding the ADI and $CV^2$. This is displayed in table 5.8. Furthermore, a set of Mann-Whitney $U$-tests provided the statistical substantiation of the results from table 5.6 and 5.7 is used to add the statistical significant information to the content. Results can be found in table 5.9.

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 0.844 | 1.000 | 1.168 | 1.385 |
| **Fleet size 16** | 0.883 | 1.000 | 1.171 | 1.350 |
| **Fleet size 32** | 0.858 | 1.000 | 1.121 | 1.284 |
| **Fleet size 64** | 0.900 | 1.000 | 1.098 | 1.223 |
| **Fleet size 96** | 0.916 | 1.000 | 1.068 | 1.184 |
| **Fleet size 128** | 0.928 | 1.000 | 1.053 | 1.135 |

Table 5.6: V2: Overview of the influence on the average ADI for changing repair qualities

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 1.217 | 1.000 | 1.039 | 0.927 |
| **Fleet size 16** | 1.423 | 1.000 | 1.078 | 0.987 |
| **Fleet size 32** | 1.306 | 1.000 | 1.007 | 0.843 |
| **Fleet size 64** | 1.101 | 1.000 | 0.951 | 0.844 |
| **Fleet size 96** | 1.079 | 1.000 | 1.005 | 0.936 |
| **Fleet size 128** | 1.023 | 1.000 | 0.998 | 0.977 |

Table 5.7: V2: Overview of the influence on the average $CV^2$ for changing repair qualities

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **$CV^2$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 5.8: V2: Outcome of the $p$-values of the Kruskal-Wallis $H$-test

As can be seen in table 5.10, for all fleet sizes, the number of failures decreases as the repair quality improves.

| Fleet size | Repair quality | ADI | | | | $CV^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.062 | 0.000 | | 0.500 | 0.081 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.003 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.007 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.008 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.289 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.010 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.021 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table 5.9: V2: Outcome of the $p$-values of the Mann-Whitney $U$-tests

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 21600 | 17307 | 15049 | 11828 |
| **Fleet size 16** | 42594 | 35269 | 29635 | 23892 |
| **Fleet size 32** | 85900 | 69675 | 59086 | 47415 |
| **Fleet size 64** | 171233 | 141009 | 118273 | 94682 |
| **Fleet size 96** | 256339 | 210985 | 178936 | 141855 |
| **Fleet size 128** | 343242 | 279839 | 237333 | 189596 |

Table 5.10: V2: number of failures for the different fleet sizes and corresponding repair qualities

## 5.4. Results of variant 3

The results presented in sections 5.2 and 5.3 show the results of presented model without the incorporation of the environmental effects. In these outcomes, only the reference climate is taken into account, representing the environment that does not influence the behavior of the individual components. As different airlines operate in different continents and thus face different climates, the scope of the results is broadened by the incorporation of the environmental effects. Note that the results of the temperate environmental conditions are not displayed in this section, as section 5.3 already showed the outcome of this environmental condition. The graphical overview of the different environmental scenarios is given in the figures 5.11, 5.12 and 5.13. As the main goal of this research is to find the impact of incorrect repairs on the failure patterns of components, the repair quality is the varying parameter in the figures. Hence, for each environmental option, the impact of the incorrect repair is investigated. As the combination of fleet size and climate are taken into account

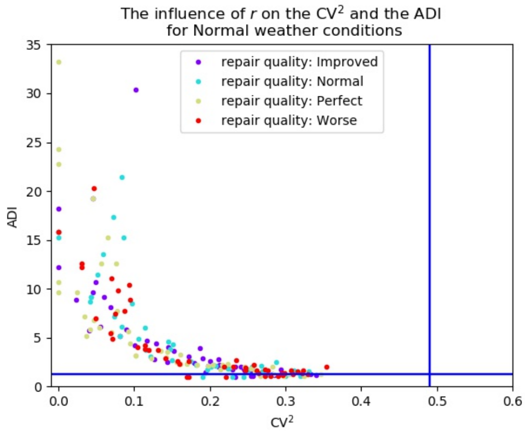



Figure 5.11: V3: Visual results - temperate climate

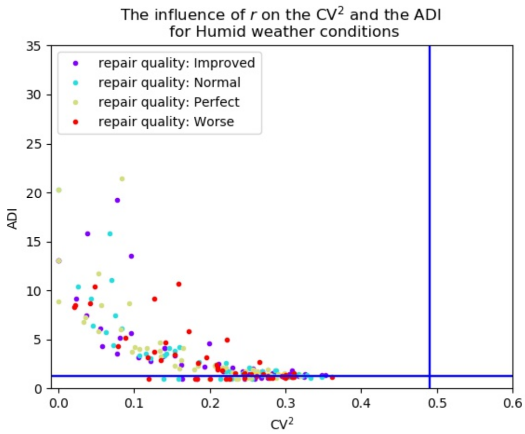



Figure 5.12: V3: Visual results - Humid climate

in this scenario, eighteen different figures (combinations of six different fleet sizes and three environmental situations) could be displayed to the reader. However, as the contribution of displaying all this figures is limited when it comes to the outcome of the research, this is left out of this chapter. For the interested reader, the figures can be found in appendix B. Note that the graphics related to the reference climate (the temperate climate) are already provided in figures 5.5 - 5.10. The results for the humid and desert environment are provided in tables 5.11 - 5.14. For every considered flight size, a KW $H$-test is provided to validate or reject the null hypothesis that groups of different repair qualities and corresponding fleet size have the same median. Table 5.15 provides the $p$-values corresponding to this test for the humid environmental conditions and table 5.16 provides the same information for desert environmental conditions. For the individual comparison of the results, the Mann-Whitney $U$-test is used again. The results are given in tables 5.17 and 5.18, for the

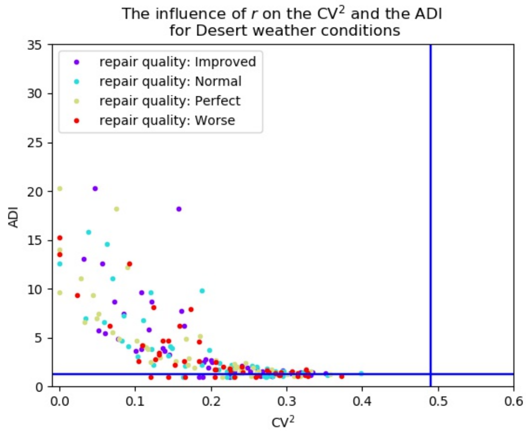



Figure 5.13: V3: Visual results - Desert climate

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 0.878 | 1.000 | 1.171 | 1.338 |
| **Fleet size 16** | 0.868 | 1.000 | 1.108 | 1.288 |
| **Fleet size 32** | 0.886 | 1.000 | 1.089 | 1.230 |
| **Fleet size 64** | 0.920 | 1.000 | 1.069 | 1.154 |
| **Fleet size 96** | 0.942 | 1.000 | 1.050 | 1.120 |
| **Fleet size 128** | 0.954 | 1.000 | 1.036 | 1.085 |

Table 5.11: V3: Overview of the influence on the average ADI for changing repair qualities for humid conditions

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 1.293 | 1.000 | 1.129 | 0.978 |
| **Fleet size 16** | 1.385 | 1.000 | 1.028 | 0.887 |
| **Fleet size 32** | 1.120 | 1.000 | 0.944 | 0.831 |
| **Fleet size 64** | 1.065 | 1.000 | 0.980 | 0.937 |
| **Fleet size 96** | 1.010 | 1.000 | 1.016 | 1.013 |
| **Fleet size 128** | 0.974 | 1.000 | 1.044 | 1.080 |

Table 5.13: V3: Overview of the influence on the average $CV^2$ for changing repair qualities for humid conditions

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 0.878 | 1.000 | 1.157 | 1.354 |
| **Fleet size 16** | 0.865 | 1.000 | 1.136 | 1.313 |
| **Fleet size 32** | 0.881 | 1.000 | 1.114 | 1.254 |
| **Fleet size 64** | 0.908 | 1.000 | 1.072 | 1.177 |
| **Fleet size 96** | 0.928 | 1.000 | 1.057 | 1.133 |
| **Fleet size 128** | 0.943 | 1.000 | 1.040 | 1.100 |

Table 5.12: V3: Overview of the influence on the average ADI for changing repair qualities for desert conditions

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 1.232 | 1.000 | 1.034 | 0.958 |
| **Fleet size 16** | 1.326 | 1.000 | 1.060 | 0.950 |
| **Fleet size 32** | 1.187 | 1.000 | 0.962 | 0.832 |
| **Fleet size 64** | 1.084 | 1.000 | 0.972 | 0.910 |
| **Fleet size 96** | 1.010 | 1.000 | 0.985 | 0.968 |
| **Fleet size 128** | 0.996 | 1.000 | 1.034 | 1.046 |

Table 5.14: V3: Overview of the influence on the average $CV^2$ for changing repair qualities for desert conditions

| **Fleet size** | **8** | **16** | **32** | **64** | **96** | **128** |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **$CV^2$** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.780 |

Table 5.15: V3: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for humid conditions

| **Fleet size** | **8** | **16** | **32** | **64** | **96** | **128** |
|---|---|---|---|---|---|---|
| **ADI** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **$CV^2$** | 0.038 | 0.002 | 0.000 | 0.007 | 0.000 | 0.000 |

Table 5.16: V3: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for desert conditions

humid and desert environmental conditions respectively.
Tables 5.20 and 5.19 provide the reader with an overview of the number of failures for changing repair qualities and fleet sizes for humid and desert environmental conditions.

| Fleet size | Repair quality | ADI | | | | CV$^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.007 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.005 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.001 | 0.000 | | 0.500 | 0.004 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.001 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.013 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.006 | 0.001 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.262 | 0.001 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.382 | 0.288 | 0.390 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.303 | 0.271 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.149 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table 5.17: V3: Outcome of the $p$-values of the Mann-Whitney $U$-tests for humid conditions

| Fleet size | Repair quality | ADI | | | | CV$^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.004 | 0.000 | | 0.500 | 0.049 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.003 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.001 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.001 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.010 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.267 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.008 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.111 | 0.076 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.426 | 0.004 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table 5.18: V3: Outcome of the $p$-values of the Mann-Whitney $U$-tests for desert conditions

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 32126 | 26749 | 23185 | 19029 |
| **Fleet size 16** | 64006 | 53022 | 46518 | 37848 |
| **Fleet size 32** | 128418 | 106867 | 92459 | 76009 |
| **Fleet size 64** | 256738 | 216099 | 185919 | 152860 |
| **Fleet size 96** | 384933 | 323429 | 279864 | 228751 |
| **Fleet size 128** | 511731 | 429693 | 373221 | 305830 |

Table 5.19: V3: number of failures for the different fleet sizes and corresponding repair qualities in humid conditions

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 27820 | 23135 | 19938 | 16281 |
| **Fleet size 16** | 56402 | 46561 | 40190 | 32527 |
| **Fleet size 32** | 112521 | 93557 | 79753 | 64666 |
| **Fleet size 64** | 224494 | 185527 | 160598 | 129464 |
| **Fleet size 96** | 336657 | 278752 | 239857 | 194759 |
| **Fleet size 128** | 447623 | 372019 | 319092 | 259455 |

Table 5.20: V3: number of failures for the different fleet sizes and corresponding repair qualities in desert conditions

## 5.5. Results of variant 4

The previous variants presented the results of incorrect repairs for different fleet sizes and different environmental conditions, but left another extension out of scope. In this extension, the influence of performance differences among aircraft types sharing the same pool of components is discussed. As stated in section 4.4, a variance of 20% of the performance of the results from the data set is incorporated. Hence, the fleets in this simulation are composed of different performing aircraft. For every aircraft in a given fleet size, the model picks one of the performance options, representing a aircraft model, randomly. The first option is an aircraft that performs 20% better compared to the found benchmark from the data analysis. The second option is performing 20% worse, and the final option is performing similar to the found benchmark. Note that the results for non-mixed fleets is provided in the previous variants.

Figures 5.14 - 5.16 provide the visual overview of the results for the different environmental conditions. In these figures, all fleet sizes are aggregated. For the graphical results of the individual fleet sizes, the reader is referred to appendix C. As the graphical results do not directly provide a sufficient outcome to the influence

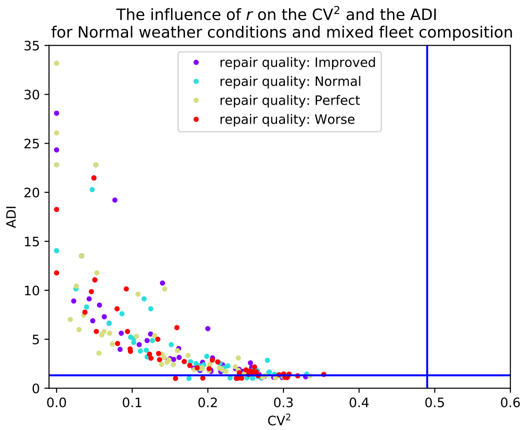



Figure 5.14: V4: Visual results - Temperate climate

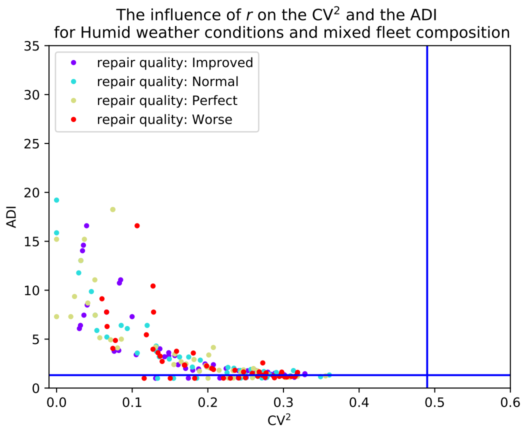



Figure 5.15: V4: Visual results - Humid climate

of the incorrect repair rate, tables 5.21-5.26 quantify the outcome of the model. Again, as in the previous variants, the "Normal" repair quality is taken as a benchmark, and the results are presented relative to this benchmark. Statistical results regarding the differences among the sets of results are given in tables 5.27 - 5.32. The first three of these tables show the $p$-values for the similarity test of the ADI and CV$^2$ of subsets with corresponding variables, but varying repair qualities. The latter three present the $p$ values of the Mann-Whitney $U$-test, which compares the similarity of two subsets with varying repair quality parameter, but equal values for all other parameters in play. Next, the total number of failures for the different scenarios are displayed in tables 5.33-5.35.

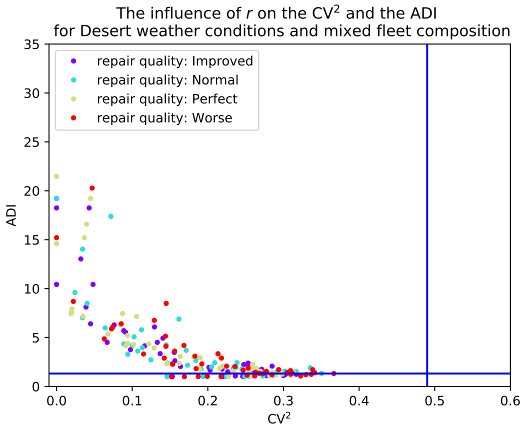



Figure 5.16: V4: Visual results - Desert climate

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 0.906 | 1.000 | 1.222 | 1.385 |
| **Fleet size 16** | 0.856 | 1.000 | 1.188 | 1.342 |
| **Fleet size 32** | 0.882 | 1.000 | 1.128 | 1.290 |
| **Fleet size 64** | 0.904 | 1.000 | 1.094 | 1.229 |
| **Fleet size 96** | 0.907 | 1.000 | 1.071 | 1.170 |
| **Fleet size 128** | 0.931 | 1.000 | 1.063 | 1.142 |

Table 5.21: V4: Overview of the influence on the average ADI for changing repair qualities for temperate conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 0.869 | 1.000 | 1.158 | 1.304 |
| **Fleet size 16** | 0.853 | 1.000 | 1.120 | 1.283 |
| **Fleet size 32** | 0.897 | 1.000 | 1.104 | 1.244 |
| **Fleet size 64** | 0.916 | 1.000 | 1.062 | 1.152 |
| **Fleet size 96** | 0.941 | 1.000 | 1.048 | 1.116 |
| **Fleet size 128** | 0.957 | 1.000 | 1.038 | 1.086 |

Table 5.22: V4: Overview of the influence on the average ADI for changing repair qualities for humid conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 0.881 | 1.000 | 1.216 | 1.343 |
| **Fleet size 16** | 0.871 | 1.000 | 1.149 | 1.298 |
| **Fleet size 32** | 0.886 | 1.000 | 1.101 | 1.244 |
| **Fleet size 64** | 0.910 | 1.000 | 1.074 | 1.178 |
| **Fleet size 96** | 0.926 | 1.000 | 1.049 | 1.132 |
| **Fleet size 128** | 0.945 | 1.000 | 1.042 | 1.103 |

Table 5.23: V4: Overview of the influence on the average ADI for changing repair qualities for desert conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 1.241 | 1.000 | 1.058 | 0.969 |
| **Fleet size 16** | 1.392 | 1.000 | 1.051 | 0.948 |
| **Fleet size 32** | 1.201 | 1.000 | 0.941 | 0.772 |
| **Fleet size 64** | 1.114 | 1.000 | 0.972 | 0.869 |
| **Fleet size 96** | 1.067 | 1.000 | 0.979 | 0.919 |
| **Fleet size 128** | 1.022 | 1.000 | 1.001 | 0.977 |

Table 5.24: V4: Overview of the influence on the average $CV^2$ for changing repair qualities for temperate conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 1.372 | 1.000 | 1.179 | 0.959 |
| **Fleet size 16** | 1.364 | 1.000 | 1.025 | 0.863 |
| **Fleet size 32** | 1.172 | 1.000 | 0.948 | 0.844 |
| **Fleet size 64** | 1.053 | 1.000 | 0.978 | 0.939 |
| **Fleet size 96** | 1.007 | 1.000 | 1.009 | 1.009 |
| **Fleet size 128** | 0.974 | 1.000 | 1.041 | 1.072 |

Table 5.25: V4: Overview of the influence on the average $CV^2$ for changing repair qualities for humid conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 1.235 | 1.000 | 1.047 | 0.957 |
| **Fleet size 16** | 1.460 | 1.000 | 1.052 | 0.872 |
| **Fleet size 32** | 1.216 | 1.000 | 0.950 | 0.863 |
| **Fleet size 64** | 1.082 | 1.000 | 0.970 | 0.899 |
| **Fleet size 96** | 1.019 | 1.000 | 0.990 | 0.979 |
| **Fleet size 128** | 0.973 | 1.000 | 1.008 | 1.021 |

Table 5.26: V4: Overview of the influence on the average $CV^2$ for changing repair qualities for desert conditions and mixed fleet composition

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| ADI | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $CV^2$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 5.27: V4: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for temperate conditions and mixed fleet composition

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| ADI | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $CV^2$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.381 |

Table 5.28: V4: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for humid conditions and mixed fleet composition

| Fleet size | 8 | 16 | 32 | 64 | 96 | 128 |
|---|---|---|---|---|---|---|
| ADI | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $CV^2$ | 0.001 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 |

Table 5.29: V4: Outcome of the $p$-values of the Kruskal-Wallis $H$-test for desert conditions and mixed fleet composition

| Fleet size | Repair quality | ADI | | | | $CV^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.001 | 0.000 | | 0.500 | 0.011 | 0.000 |
| | Improved | | | 0.500 | 0.001 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.003 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.004 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.005 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.014 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.024 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.033 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table 5.30: V4: Outcome of the $p$-values of the Mann-Whitney $U$-tests for temperate conditions and mixed fleet composition

| Fleet size | Repair quality | ADI | | | | $CV^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.001 | 0.000 | | 0.500 | 0.093 | 0.000 |
| | Improved | | | 0.500 | 0.001 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.003 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.011 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.002 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.116 | 0.001 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.031 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.033 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.417 | 0.479 | 0.106 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.465 | 0.060 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.073 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table 5.31: V4: Outcome of the $p$-values of the Mann-Whitney $U$-tests for humid conditions and mixed fleet composition

| Fleet size | Repair quality | ADI | | | | $CV^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Worse | Normal | Improved | Perfect | Worse | Normal | Improved | Perfect |
| 8 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.001 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.024 | 0.000 |
| | Improved | | | 0.500 | 0.006 | | | 0.500 | 0.005 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 16 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.012 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 32 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.000 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.001 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 64 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.005 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.000 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 96 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.076 | 0.000 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.004 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.019 |
| | Perfect | | | | 0.500 | | | | 0.500 |
| 128 | Worse | 0.500 | 0.000 | 0.000 | 0.000 | 0.500 | 0.388 | 0.009 | 0.000 |
| | Normal | | 0.500 | 0.000 | 0.000 | | 0.500 | 0.004 | 0.000 |
| | Improved | | | 0.500 | 0.000 | | | 0.500 | 0.008 |
| | Perfect | | | | 0.500 | | | | 0.500 |

Table 5.32: V4: Outcome of the $p$-values of the Mann-Whitney $U$-tests for desert conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 20919 | 17558 | 14742 | 12086 |
| **Fleet size 16** | 42943 | 34815 | 28962 | 23920 |
| **Fleet size 32** | 84550 | 70649 | 59706 | 47868 |
| **Fleet size 64** | 169776 | 139508 | 118177 | 94529 |
| **Fleet size 96** | 259024 | 209581 | 177661 | 141471 |
| **Fleet size 128** | 341965 | 279328 | 236176 | 188856 |

Table 5.33: V4: number of failures for the different fleet sizes and corresponding repair qualities in temperate conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 32049 | 26606 | 23210 | 19441 |
| **Fleet size 16** | 65296 | 53445 | 46638 | 38007 |
| **Fleet size 32** | 127381 | 107349 | 92656 | 75890 |
| **Fleet size 64** | 255005 | 213128 | 185321 | 152965 |
| **Fleet size 96** | 382840 | 321422 | 278986 | 229643 |
| **Fleet size 128** | 510826 | 429559 | 371071 | 305167 |

Table 5.34: V4: number of failures for the different fleet sizes and corresponding repair qualities in humid conditions and mixed fleet composition

| | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Fleet size 8** | 28078 | 23380 | 19481 | 16454 |
| **Fleet size 16** | 55974 | 46466 | 39760 | 32642 |
| **Fleet size 32** | 111385 | 92918 | 79857 | 64966 |
| **Fleet size 64** | 223273 | 185618 | 159431 | 129452 |
| **Fleet size 96** | 335699 | 277089 | 239598 | 194039 |
| **Fleet size 128** | 447596 | 371117 | 318353 | 259111 |

Table 5.35: V4: number of failures for the different fleet sizes and corresponding repair qualities in desert conditions and mixed fleet composition

# 6

# Evaluation and discussion of model results

The results of the model described in chapter 5 are evaluated and discussed in this chapter. For each variant, the provided results are elaborated, construed and discussed. The hypothetical improvements of repair quality are tested for all variants, measuring the effect on the changing ADI, $CV^2$ and total failures. For the evaluation of all variants, the reader is referred to the figures and tables from chapter 5.

## 6.1. Statistical evaluation and discussion

In this section, the statistical outcomes of the KW $H$-test and the Mann-Whitney $U$-test are evaluated. As stated earlier, the $p$-value of the Mann-Whitney $U$-test and the KW $H$-test describe the statistical significance to reject the null hypothesis. With a 95% CI, values smaller than 0.05 are statistically significant and therefore it can be concluded that the null hypothesis can be rejected in these cases. With the null hypothesis of the KW $H$-test and Mann-Whitney $U$-test, it translates in the conclusion that the compared sets of data are different in median. In the results, the KW $H$-test is used to compare the set of four different repair quality with each other. The underlying question is whether the changing repair quality has an effect on either the ADI or $CV^2$. Next, the same is done for the set of a single repair quality. This set is compared to another set representing one repair quality. All other parameters are the same. Note that, when comparing the same sets, the result of the $p$-value is not equal to 1, although it massively exceeds the CI threshold.

A critical note that has to be made is the fact that the $p$-values for both tests drastically fall when increasing the amount of iterations. This is the consequence of the fact that more iterations result in larger (sub)sets of data. When the compared sets of data grow, the acceptable deviation between the sets becomes smaller. Therefore, small deviations in the sets can lead to statistical significant differences, although in practice the sets are similar.

## 6.2. Evaluation of variant 1

The goal of variant 1 is to quantify the effect of incorrect repairs on the ADI and $CV^2$ for spare parts. Tables 5.1 and 5.2 provide the measured effect on both metrics when the repair quality is changed. Table 5.5 provides an overview of the total number of failures over time. As can be concluded from the latter table, an improvement of the repair quality decreases the total number of failures for all improvement steps. From these tables, it can be concluded that:

- An improvement of the repair quality from "Worse" to "Normal" increases the ADI by 15.1%, decreases the $CV^2$ by 3.0% and decreases the total number of failures by 17.5%;
- An improvement from the "Normal" to the "Improved" repair quality leads to a increase of the ADI by 12.6%, an increase of the $CV^2$ by 1.1%, and a reduction of the total number of failures of 15.6%. Hence, this improvement has a positive effect on the total number of failures, but deteriorates the predictability of the failures over time;
- The improvement of the repair quality from the "Improved" to the "Perfect" situation leads to an improvement of the ADI by 15.2%, a reduction of the $CV^2$ by 3.9%, and a reduction of the total number of failures of 20.1%.

The relative gain and losses of the total number of failures is given in table 6.1. Note that the absolute numbers of failures for each repair quality as well. The relative numbers are measured from this number. Table 5.1 and

| Repair quality | Worse | Normal | Improved | Perfect |
|---|---|---|---|---|
| **Worse** | 916444 | 0.825 | 0.696 | 0.556 |
| **Normal** | | 755740 | 0.844 | 0.675 |
| **Improved** | | | 638149 | 0.799 |
| **Perfect** | | | | 509981 |

Table 6.1: Ev1: Absolute and relative difference in failures for different repair qualities

5.2 present the change of the ADI and $CV^2$ for the changing repair qualities. As can be seen, an improvement of the repair quality results in an increase of the ADI. This is the result of fewer subsequent failures, leaving more days without failures. Hence, the ADI increases.
The opposite occurs for the $CV^2$. However, the decrease of $CV^2$ is smaller compared to the increase of the ADI. Therefore, an improvement in the area of the repair quality causes a trade-off between decreasing $CV^2$ versus increasing ADI. In the case that an increase of the ADI is acceptable, the improvement in repair quality is desirable, as the total number of failures decreases.

### Conclusion on variant 1

Variant 1 provides the general overview of the impact of changing repair qualities on the different metrics. The above-mentioned results are to a certain extend in line with expectations. In every situation, the improvement of the repair quality reduces the total number of failures, which is a positive on the economical side of the problem. However, an improvement of the repair quality raises the ADI in all situations, making it harder to predict when a failure will occur. Paired with this increasing ADI comes a decreased $CV^2$ in two of the three situations. For the improvement from the "Normal" to the "Improved" scenario however, the $CV^2$ increases as well. Therefore, this improvement in repair quality is not desirable if the intermittent or lumpy demand patterns are a striking problem.

## 6.3. Evaluation of variant 2

In the second variant the different fleet sizes are treated separately. By looking at the visual results of figures 5.5 - 5.10 it can be seen that the fleet size is the main contributor to the change in both demand metrics. As mentioned before, the larger the fleet size becomes, the smoother the demand patterns are looking. MRO providers are therefore advised to enlarge their fleet to have more grip on the demand of spare parts. However, for MRO providers that can not enlarge the fleet, this strategy is useless. Tables 5.6, 5.7 and 5.10 are used in the following subsections.

### General evaluation

Generally, it can be concluded that an improvement in the repair quality results in a higher ADI, a lower $CV^2$ and a lower amount of total failures. It can be seen from tables 5.6 and 5.7 that the impact of an improvement of repair quality has the most effect on smaller fleet sizes. In general, the smaller the fleet size, the larger the impact on both the ADI and $CV^2$ is.
Table 5.10 provides a different conclusion for the number of failures. For this metric, it is beneficial for every fleet size and every base repair quality to improve this repair quality. The following subsections dive deeper into the individual fleet sizes.

### Evaluation of fleet size 8

By increasing the repair quality for a fleet of 8 aircraft from "Worse" to "Normal" has a positive impact on the behavior of the failures. Paired with a small increase of 18.4% of the ADI comes a decrease of 17.8% for the $CV^2$. Furthermore, the total number of failures decreases by 19.9%.
The improvement of the repair quality from the "Normal" level to the "Improved" level leads to an increase of the ADI of 16.8%, an increase of the $CV^2$ of 3.9% and a reduction of the total number of failures of 13.0%. Hence, this improvement is undesirable when it comes to the grip on predictability of the failures. However, the decrease of total failures adds a positive to this improvement.
The improvement of the repair quality from the "Improved" to the "Perfect" scenario leads to an increase of the ADI of 18.6%, a reduction of the $CV^2$ of 10.8%, and a reduction of the number of failures of 21.4%.

### Evaluation of fleet size 16

By increasing the repair quality for a fleet of 16 aircraft from "Worse" to "Normal" has a huge positive impact on the behavior of the failures. An increase of the ADI with 13.2%, an decrease of the $CV^2$ by 29.7% and a total decrease of 17.2% of the total failures are the result of this improvement.
On the other hand, an improvement from the "Normal" to an "Improved" repair quality has no positive effects on the predictability of the failures of the components (ADI increased by 17.1%, $CV^2$ increased by 7.8%), but does lower the total number of failures by 16.0%.
Finally, an improvement from the "Improved" to the "Perfect" repair quality leads to an increase of 20.4% for the ADI, an decrease of 9.1% for the $CV^2$ and a 19.4% decrease for the total number of failures over time.

### Evaluation of fleet size 32

The improvement of the repair quality from "Worse" to "Normal" leads to a increase of the ADI of 16.5%, a reduction of the $CV^2$ of 23.4%, and a reduction of total failures of 18.9%. As two of the three metrics shift in the positive direction with a higher percentage than the third, negatively shifted, metric, this improvement is desirable.
The improvement from a "Normal" to an "Improved" repair quality has a unfavorable effect on both the ADI and $CV^2$, both increasing with 12.1% and 0.7% respectively. The number of failures decreases by 15.2%.
The improvement to a "Perfect" repair quality leads to an increment of the ADI by 14.5%, a reduction of the $CV^2$ by 16.3% and a reduction of failures of 19.8%. Therefore, this improvement is desirable if the current ADI is not the bottle neck for the striking problem of the demand forecasts.

### Evaluation of fleet size 64

An improvement of the repair quality from "Worse" to "Normal" leads to a increase of the ADI 11.2%, a reduction of the $CV^2$ of 9.2%, and a reduction of the total number of failures of 17.7%. This improvement is desirable if the increase of the ADI is acceptable.
Improving the repair quality from "Normal" to "Improved" leads to an increase of the ADI of 9.8%, a decrease of the $CV^2$ of 4.9%, and a decrease in the number of failures of 16.1%. As mentioned above, the difference

in performance for predictability becomes smaller for larger fleet sizes. As the total number of failures is decreasing with similar percentages, an improvement of the repair quality is desirable.
The last improvement possibility, from "Improved" to "Perfect" repair quality, leads to an increase of the ADI of 11.4%, a decrease of the $CV^2$ of 11.3%, and a decrease in total failures of 19.9%. Again, this improvement is desirable if the increase of the ADI is acceptable.

### Evaluation of fleet size 96

For the improvement of the repair quality from "Worse" to "Normal", an improvement of the ADI of 9.2%, an decrease of the $CV^2$ of 7.3%, and a decrease of total failures of 17.7% is measured. In this scenario, the increase in ADI is larger than the decrease of the $CV^2$, and has therefore less potential situations where this improvement would suit into.
When improving the repair quality from "Normal" to "Improved", the ADI increases by 6.8%, the $CV^2$ increases by 0.5%, and the total failures decrease by 15.2%.
Finally, improving the repair quality from "Improved" to "Perfect" leads to an increase of the ADI of 10.8%, a decrease of the $CV^2$ of 6.9%, and a decrease of the total number of failures of 20.7%.

### Evaluation of fleet size 128

When improving the repair quality from "Normal" to "Improved", the ADI is increased by 7.8%, the $CV^2$ is decreased by 2.2%, and the total number of failures is decreased by 18.5%.
When improving the repair quality from "Normal" to "Improved", the ADI increases by 5.3%, the $CV^2$ decreases by 0.0%, and the total number of failures decreases by 15.2%. Hence, this scenario has a limited positive impact on the predictability of the failures, but does lead to a significant reduction of the total number of failures.
When the repair quality is improved from "Improved" to "Perfect", the ADI is increased by 7.8%, the $CV^2$ is decreased by 2.1%, and the total number of failures is decreased by 20.1%.

### Conclusion on variant 2

It can be concluded that for smaller fleet sizes (8, 16 and 32 aircraft) the effect of improving the repair quality to the "Normal" level is highly beneficial. Along with the decrease of the $CV^2$ overshadowing the increase of the ADI, the total number of failures is reduced significantly. For larger fleet sizes, these effects fade away, as the increase in ADI predominates the decrease of the $CV^2$. However, the improvement still leads a reduction of total failures.
When improving the repair quality from the "Normal" level to the "Improved" level, the impact on the ADI and $CV^2$ is mainly undesirable. The increasing ADI overshadows the gain in $CV^2$. For fleet sizes 8, 16, 32 and 128, the $CV^2$ increases as well.
The improvement of the repair quality from the "Improved" to the "Perfect" condition has a more positive effect compared to the previous improvement, but is still not very effective in overcoming the difficulty of the predictability of the failures of components. The increasing ADI comes in pairs with a reduced $CV^2$, although the effect on the $CV^2$ is less compared to the increase of the ADI. Again, for all fleet sizes, the total number of failures drops significantly.
It can be concluded that there is a negative correlation between the fleet size and the effect of an improvement in repair quality on the ADI and $CV^2$. The larger the fleet size, the smaller the relative potential gains are. See table 6.2 for the quantitative overview.

| **Improvement** | **Fleet size** | **ADI** | **$CV^2$** | **Failures** |
|---|---|---|---|---|
| Worse - Normal | 8 | +18.4% | -17.8% | -19.9% |
| | 16 | +13.2% | -29.7% | -17.2% |
| | 32 | +16.5% | -23.4% | -18.9% |
| | 64 | +11.2% | -9.2% | -17.7% |
| | 96 | +9.2% | -7.3% | -17.7% |
| | 128 | +7.8% | -2.2% | -18.5% |
| Normal - Improved | 8 | +16.8% | *+3.9%* | -13.0% |
| | 16 | +17.1% | +7.8% | -16.0% |
| | 32 | +12.1% | +0.7% | -15.2% |
| | 64 | +9.8% | -4.9% | -16.1% |
| | 96 | +6.8% | +0.5% | -15.2% |
| | 128 | +5.3% | -0.0% | -15.2% |
| Improved - Perfect | 8 | +18.6% | -10.8%* | -21.4% |
| | 16 | +20.4% | -9.1% | -19.4% |
| | 32 | +14.5% | -16.3% | -19.8% |
| | 64 | +11.4% | -11.3% | -19.9% |
| | 96 | +10.8% | -6.9% | -20.7% |
| | 128 | +7.8% | -2.1% | -20.1% |

Table 6.2: Quantitative overview of evaluation of variant 2

## 6.4. Evaluation of variant 3

The desert and humid conditions are taken into account in this variant. The results presented in section 5.4 provide the reader with a better understanding of the influence of the different climates. As both climates have a negative effect on the lifetime of components, more primary failures occur compared to the reference climate.

### General evaluation

In most of the cases, the increase of the ADI predominates the positive effect of the decreasing $CV^2$. Some improvements in repair quality show increases in both the ADI and the $CV^2$, which is not desired.
Again, the larger fleet sizes seem to profit less from the improvement in repair quality when observing the ADI and $CV^2$. However, the absolute number of failures is reduced the most for these fleet sizes, as the percentages of decrease are similar compared to the smaller fleet sizes.

### Evaluation of fleet size 8

The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 13.9% and 13.9%, the $CV^2$ is decreased by 22.7% and 18.8%, and the total failures is decreased by 16.7% and 16.8% for humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 17.1% and 15.7%, the $CV^2$ is increased by 12.9% and 3.4%, and the total failures is decreased by 13.3% and 13.8% for humid and desert environmental conditions respectively. Hence, this improvement has a negative influence on the predictability of the failures.
- From "Improved" to "Perfect", the ADI is increased by 14.2% and 17.0%, decreases the $CV^2$ by 13.4% and 7.3%, and the total failures is decreased by 17.9% and 18.3% for humid and desert environmental conditions respectively.

### Evaluation of fleet size 16

The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 15.2% and 15.6%, the $CV^2$ is decreased by 27.8% and 24.6%, and the total failures is decreased by 17.2% and 17.4% for humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 10.8% and 13.6%, the $CV^2$ is increased by 2.8% and 6.0%, and the total failures is decreased by 12.3% and 13.7% for humid and desert environmental conditions respectively. Hence, this improvement has a negative influence on the predictability of the failures for the desert environmental conditions.
- From "Improved" to "Perfect", the ADI is increased by 16.2% and 15.5%, the $CV^2$ is decreased by 13.7% and 10.4%, and the total failures is decreased by 18.6% and 19.1% for humid and desert environmental conditions respectively.

### Evaluation of fleet size 32

The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 12.8% and 13.5%, the $CV^2$ is decreased by 10.7% and 15.7%, and the total failures is decreased by 16.8% and 16.9% for humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 8.9% and 11.4%, the $CV^2$ is decreased by 5.6% and 3.8%, and the total failures is decreased by 13.5% and 14.8% for humid and desert environmental conditions respectively.
- From "Improved" to "Perfect", the ADI is increased by 13.0% and 12.5%, the $CV^2$ is decreased by 12.0% and 13.6%, and the total failures is decreased by 17.8% and 18.9% for humid and desert environmental conditions respectively.

### Evaluation of fleet size 64

For this fleet size, it is clear to see that the deviations of ADI and $CV^2$ for changing repair qualities become smaller compared to the smaller fleet sizes. The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 8.7% and 10.1%, the $CV^2$ is decreased by 6.1% and 7.8%, and the total failures is decreased by 15.8% and 17.4% for humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 6.9% and 7.2%, the $CV^2$ is decreased by 2.0% and 2.8%, and the total failures is decreased by 14.0% and 13.4% for humid and desert environmental conditions respectively.
- From "Improved" to "Perfect", the ADI is increased by 8.0% and 9.8%, the $CV^2$ is decreased by 4.3% and 6.4%, and the total failures is decreased by 17.8% and 19.4% for humid and desert environmental conditions respectively.

### Evaluation of fleet sizes 96 and 128

For both fleet sizes, the relative gains or losses of the ADI and $CV^2$ are small when varying the repair quality. Furthermore, the $p$-values of the KW $H$ test on the $CV^2$ for fleet size 128 severely exceed the statistical confidence interval of 95% to accept the hypothesis that the different repair qualities perform differently.
However, the reduction in the total number of failures is worth mentioning.

- For humid conditions, the improvements in repair quality from "Worse" to "Normal", "Normal" to "Improved", and "Improved" to "Perfect", result in a decrease of 16.0%, 13.5% and 18.3% for a fleet size of 96, and 16.0%, 13.1% and 18.1% for fleet size 128 respectively.
- For desert conditions, the improvements in repair quality from "Worse" to "Normal", "Normal" to "Improved", and "Improved" to "Perfect", result in a decrease of 17.2%, 14.0% and 18.8% for a fleet size of 96, and 16.9%, 14.2% and 18.7% for fleet size 128 respectively.

### Conclusion on variant 3

The main addition of variant 3 to the analysis is to see what the effect of varying environmental conditions is on the changing ADI, $CV^2$ and total failures.
For the desert and humid environments, patterns similar to the temperate environmental conditions are found. The change in ADI and $CV^2$ is damped out when the fleet size become larger, while the relative losses in total failures remain somewhat constant. Both the improvements of "Worse" to "Normal" and "Improved" to "Perfect" perform similarly for all metrics. However, the improvement in repair quality from "Normal" to "Improved" sees a limited decrease in the $CV^2$. In fact, many increases of the $CV^2$ are observed.
By comparing the temperate and humid environmental scenarios, it can be concluded that in the humid conditions, the ADI is less sensitive to the improvement of the repair quality. This results in smaller increments of the ADI compared to the temperate environmental conditions. The results of the deviation in $CV^2$ provide no clear winner, as both environmental conditions outperform the other condition for different values of fleet sizes and improvements. The relative losses in total failures are higher for the temperate environmental conditions, although the difference between the two scenarios is small.
By comparing the temperate and desert environmental scenarios, it becomes clear the desert environments perform slightly better compared to the temperate conditions when it comes to the increase of the ADI. That is to say, the increase of the ADI in the same situation is slightly less compared to the increase of the ADI for temperate environmental conditions. When comparing the deviations for the $CV^2$, no clear pattern can be found. In some cases, the desert conditions outperform the temperate conditions, but the opposite occurs for the same amount of scenarios. For the decrease of the total number of failures, the desert conditions profit slightly less compared to the temperate environmental conditions.
Generally, the increment of the ADI for improvement is the most limited for humid conditions, the performance of the decrease of the $CV^2$ is similar for all environmental conditions, and the relative reduction of the total number of failures is similar for all environmental conditions, although the temperate environmental conditions perform slightly better in most cases. See table 6.3 for the full quantitative overview.

| Improvement | Fleet size | ADI | | | $CV^2$ | | | Failures | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Temperate | Humid | Desert | Temperate | Humid | Desert | Temperate | Humid | Desert |
| Worse - Normal | 8 | +18.4% | +13.9% | +13.9% | -17.8% | -22.7% | -18.8% | -19.9% | -16.7% | -16.8% |
| | 16 | +13.2% | +15.2% | +15.6% | -29.7% | -27.8% | -24.6% | -17.2% | -17.2% | -17.4% |
| | 32 | +16.5% | +12.8% | +13.5% | -23.4% | -10.7% | -15.7% | -18.9% | -16.8% | -16.9% |
| | 64 | +11.2% | +8.7% | +10.1% | -9.2% | -6.1% | -7.8% | -17.7% | -15.8% | -17.4% |
| | 96 | +9.2% | +6.1% | +7.8% | -7.3% | -1.0% | -1.0% | -17.7% | -16.0% | -17.2% |
| | 128 | +7.8% | +4.8% | +6.0% | -2.2% | +2.7% | +0.4% | -18.5% | -16.0% | -16.9% |
| Normal - Improved | 8 | +16.8% | +17.1% | +15.7% | +3.9% | +12.9% | +3.4% | -13.0% | -13.3% | -13.8% |
| | 16 | +17.1% | +10.8% | +13.6% | +7.8% | +2.8% | +6.0% | -16.0% | -12.3% | -13.7% |
| | 32 | +12.1% | +8.9% | +11.4% | +0.7% | -5.6% | -3.8% | -15.2% | -13.5% | -14.8% |
| | 64 | +9.8% | +6.9% | +7.2% | -4.9% | -2.0% | -2.8% | -16.1% | -14.0% | -13.4% |
| | 96 | +6.8% | +5.0% | +5.7% | +0.5% | +1.6% | -1.6% | -15.2% | -13.5% | -14.0% |
| | 128 | +5.3% | +3.6% | +4.0% | -0.0% | +4.4% | +3.4% | -15.2% | -13.1% | -14.2% |
| Improved - Perfect | 8 | +18.6% | +14.2% | +17.0% | -10.8% | -13.4% | -7.3% | -21.4% | -17.9% | -18.3% |
| | 16 | +20.4% | +16.2% | +15.5% | -9.1% | -13.7% | -10.4% | -19.4% | -18.6% | -19.1% |
| | 32 | +14.5% | +13.0% | +12.5% | -16.3% | -12.0% | -13.6% | -19.8% | -17.8% | -18.9% |
| | 64 | +11.4% | +8.0% | +9.8% | -11.3% | -4.3% | -6.4% | -19.9% | -17.8% | -19.4% |
| | 96 | +10.8% | +6.7% | +7.2% | -6.9% | -0.0% | -1.7% | -20.7% | -18.3% | -18.8% |
| | 128 | +7.8% | +4.7% | +5.8% | -2.1% | +3.5% | +1.2% | -20.1% | -18.1% | -18.7% |

Table 6.3: Quantitative overview of evaluation of variant 3

## 6.5. Evaluation of variant 4

For the evaluation of the results of variant 4, the reader is referred to the presented data in section 5.5. In this section the influence of the incorrect repair rate for varying fleet sizes, environmental conditions, repair qualities and shared component pool strategies is clarified. The results of the simulations are used as input to quantify the effects of varying repair qualities.

### 6.5.1. General evaluation

From tables 5.21-5.26 it can be seen that the larger the fleet size becomes, the smaller the impact of a changing repair quality is. However, tables 5.33-5.35 show that for all circumstances, an improvement in repair quality significantly reduces the total number of failures. Therefore, it can be concluded that larger fleet sizes benefit the most in the field of failure reduction when it comes to improved repair quality, while for smaller fleet sizes there is another potential gain when it comes to the predictability of failures over time.

#### Evaluation of fleet size 8

The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 10.4%, 15.1% and 13.6%, the $CV^2$ is decreased by 19.4%, 26.7% and 31.5%, and the total failures is decreased by 16.1%, 17.0% and 16.7% for temperate, humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 22.2%, 15.8% and 21.6%, the $CV^2$ is increased by 5.8%, 17.9% and 4.7%, and the total failures is decreased by 16.0%, 12.8% and 16.7% for temperate, humid and desert environmental conditions respectively. Hence, from predictability perspective, this improvement is not desirable as both the ADI and $CV^2$ increase for all environmental conditions.
- From "Improved" to "Perfect", the ADI is increased by 13.4%, 12.6% and 10.5%, the $CV^2$ is decreased by 8.4%, 18.7% and 8.6%, and the total failures is decreased by 18.0%, 16.2% and 15.5% for temperate, humid and desert environmental conditions respectively.

#### Evaluation of fleet size 16

The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 16.9%, 17.3% and 14.8%, the $CV^2$ is decreased by 28.1%, 26.7% and 31.5%, and the total failures is decreased by 18.9%, 18.1% and 17.0% for temperate, humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 18.8%, 12.0% and 14.9%, the $CV^2$ is increased by 5.1%, 2.5% and 5.2%, and the total failures is decreased by 16.8%, 12.7% and 14.4% for temperate, humid and desert environmental conditions respectively. Hence, this improvement is not improving the predictability of the demand of the failures.
- From "Improved" to "Perfect", the ADI is increased by 13.0%, 14.5% and 12.9%, the $CV^2$ is decreased by 9.8%, 15.8% and 17.1%, and the total failures is decreased by 17.4%, 18.5% and 17.9% for temperate, humid and desert environmental conditions respectively.

#### Evaluation of fleet size 32

The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 13.4%, 11.5% and 12.8%, the $CV^2$ is decreased by 16.7%, 14.6% and 17.5%, and the total failures is decreased by 16.4%, 15.7% and 16.6% for temperate, humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 12.8%, 10.4% and 10.1%, the $CV^2$ is decreased by 5.9%, 5.2% and 5.0%, and the total failures is decreased by 15.5%, 13.7% and 14.1% for temperate, humid and desert environmental conditions respectively.
- From "Improved" to "Perfect", the ADI is increased by 14.3%, 12.6% and 13.0%, the $CV^2$ is decreased by 18.0%, 10.9% and 9.2%, and the total failures is decreased by 19.8%, 18.1% and 18.6% for temperate, humid and desert environmental conditions respectively.

### Evaluation of fleet size 64

The three improvement possibilities lead to the following deviations in the ADI, $CV^2$ and total failures:

- From "Worse" to "Normal", the ADI is increased by 10.6%, 9.2% and 9.9%, the $CV^2$ is decreased by 10.2%, 5.0% and 7.6%, and the total failures is decreased by 17.8%, 16.4% and 16.9% for temperate, humid and desert environmental conditions respectively.
- From "Normal" to "Improved", the ADI is increased by 9.4%, 6.2% and 7.4%, the $CV^2$ is decreased by 2.8%, 2.2% and 3.0%, and the total failures is decreased by 15.3%, 13.0% and 14.1% for temperate, humid and desert environmental conditions respectively.
- From "Improved" to "Perfect", the ADI is increased by 12.3%, 8.4% and 9.7%, the $CV^2$ is decreased by 10.6%, 3.9% and 7.3%, and the total failures is decreased by 20.0%, 17.5% and 18.8% for temperate, humid and desert environmental conditions respectively.

### Evaluation of fleet sizes 96 and 128

As the influence of the repair quality shrinks for larger fleet sizes and hence the possible improvements on the predictability of the failures become smaller, the individual results of the impact on the ADI and $CV^2$ are not discussed here. Furthermore, the outcome of the statistical tests in tables 5.27-5.29 show that for a fleet size of 128, there is insufficient evidence to conclude that the the different repair qualities have different outcomes when it comes to the ADI and $CV^2$. For the interested reader, tables 5.21-5.26 provide the relative gains and losses.
Contradictory, significant reductions in the total number of failures are noted for large fleet sizes. Furthermore, in absolute numbers, the largest gains are obtained.

- For temperate conditions, the improvements in repair quality from "Worse" to "Normal", "Normal" to "Improved", and "Improved" to "Perfect", result in a decrease of 19.1%, 15.2% and 20.4% for a fleet size of 96, and 18.3%, 15.4% and 20.4% for fleet size 128 respectively.
- For humid conditions, the improvements in repair quality from "Worse" to "Normal", "Normal" to "Improved", and "Improved" to "Perfect", result in a decrease of 16.0%, 13.2% and 17.7% for a fleet size of 96, and 15.9%, 13.6% and 17.8% for fleet size 128 respectively.
- For desert conditions, the improvements in repair quality from "Worse" to "Normal", "Normal" to "Improved", and "Improved" to "Perfect", result in a decrease of 17.4%, 13.5% and 19.0% for a fleet size of 96, and 17.1%, 14.2% and 18.6% for fleet size 128 respectively.

### Conclusion on variant 4

First, the results of this variant are concluded here. Next, the results are compared to the results of variant 3, as this comparison provides the reader with information of the effect of a fleet size with varying performing aircraft types. Table 6.4 is used to support the statements made below.

Comparing the three different environmental scenarios, it can be seen that the same patterns occur for all of them. By an improvement of the repair quality, the increase of the ADI and the decrease of the $CV^2$ damp out as the fleet size becomes larger. Furthermore, the relative reduction of the total number of failures is kept on a constant level when increasing the fleet size. For the improvements "Worse"-"Normal" and "Improved"-"Perfect" of repair qualities, the deviations in the three metrics (ADI, $CV^2$ and total failures) is similar. However, the improvement "Normal"-"Improved" performs different. The gains in the $CV^2$ are limited, while in many cases there even is an undesired increase of the $CV^2$.
Temperate environmental conditions lead to the highest deviation in ADI, $CV^2$ and the failures for all improvements.
In order to compare variant 3 and 4, the two tables are compared and the differences are calculated. Table 6.5 shows the differences in performance of variant 4 compared to variant 3. Although is some cases there are significant differences notable, the majority of the deviations are relatively small. Hence, the level of predictability of failures is not deteriorated by the introduction of components that are operable for multiple aircraft types.

| **Improvement** | **Fleet size** | **ADI** | | | **CV$^2$** | | | **Failures** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Temperate | Humid | Desert | Temperate | Humid | Desert | Temperate | Humid | Desert |
| Worse - Normal | 8 | +10.4% | +15.1% | +13.6% | -19.4% | -26.7% | -31.5% | -16.1% | -17.0% | -16.7% |
| | 16 | +16.9% | +17.3% | +14.8% | -28.1% | -26.7% | -31.5% | -18.9% | -18.1% | -17.0% |
| | 32 | +13.4% | +11.5% | +12.8% | -16.7% | -14.6% | -17.5% | -16.4% | -15.7% | -16.6% |
| | 64 | +10.6% | +9.2% | +9.9% | -10.2% | -5.0% | -7.6% | -17.8% | -16.4% | -16.9% |
| | 96 | +10.3% | +6.3% | +8.0% | -6.3% | -0.7% | -1.9% | -19.1% | -16.0% | -17.4% |
| | 128 | +7.4% | +4.5% | +5.8% | -2.1% | +2.7% | +2.7% | -18.3% | -15.9% | -17.1% |
| Normal - Improved | 8 | +22.2% | +15.8% | +21.6% | +5.8% | +17.9% | +4.7% | -16.0% | -12.8% | -16.7% |
| | 16 | +18.8% | +12.0% | +14.9% | +5.1% | +2.5% | +5.2% | -16.8% | -12.7% | -14.4% |
| | 32 | +12.8% | +10.4% | +10.1% | -5.9% | -5.2% | -5.0% | -15.5% | -13.7% | -14.1% |
| | 64 | +9.4% | +6.2% | +7.4% | -2.8% | -2.2% | -3.0% | -15.3% | -13.0% | -14.1% |
| | 96 | +7.1% | +4.8% | +4.9% | -2.1% | +0.9% | -1.0% | -15.2% | -13.2% | -13.5% |
| | 128 | +6.2% | +3.8% | +4.2% | +0.0% | +4.1% | +0.8% | -15.4% | -13.6% | -14.2% |
| Improved - Perfect | 8 | +13.4% | +12.6% | +10.5% | -8.4% | -18.7% | -8.6% | -18.0% | -16.2% | -15.5% |
| | 16 | +13.0% | +14.5% | +12.9% | -9.8% | -15.8% | -17.1% | -17.4% | -18.5% | -17.9% |
| | 32 | +14.3% | +12.6% | +13.0% | -18.0% | -10.9% | -9.2% | -19.8% | -18.1% | -18.6% |
| | 64 | +12.3% | +8.4% | +9.7% | -10.6% | -3.9% | -7.3% | -20.0% | -17.5% | -18.8% |
| | 96 | +9.3% | +6.5% | +7.9% | -6.1% | +0.1% | -1.1% | -20.4% | -17.7% | -19.0% |
| | 128 | +7.4% | +4.6% | +5.9% | -2.3% | +3.0% | +1.3% | -20.4% | -17.8% | -18.6% |

Table 6.4: Quantitative overview of evaluation of variant 4

| **Improvement** | **Fleet size** | **ADI** | | | **CV$^2$** | | | **Failures** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Temperate | Humid | Desert | Temperate | Humid | Desert | Temperate | Humid | Desert |
| Worse - Normal | 8 | -8.0% | +1.2% | -0.3% | -1.6% | -4.0% | -12.7% | +3.8% | -0.3% | +0.1% |
| | 16 | +3.7% | +2.1% | -0.8% | +1.6% | +1.1% | +6.9% | -1.7% | -0.9% | +0.4% |
| | 32 | -3.1% | -1.3% | -0.7% | +6.7% | -3.9% | -1.8% | +2.5% | +1.1% | -0.3% |
| | 64 | -0.6% | +0.5% | -0.2% | -1.0% | +1.1% | +0.2% | -0.1% | -0.6% | +0.5% |
| | 96 | +1.1% | +0.2% | +0.2% | +1.0% | +0.3% | -0.9% | -1.4% | -0.0% | -0.2% |
| | 128 | -0.4% | -0.3% | -0.2% | +0.1% | -0.0% | +2.3% | +0.2% | +0.1% | -0.2% |
| Normal - Improved | 8 | +5.4% | -1.3% | +4.9% | +1.9% | +5.0% | +1.3% | -3.0% | +0.5% | -2.9% |
| | 16 | +1.7% | +1.2% | +1.3% | -2.7% | -0.3% | -0.8% | -0.8% | -0.4% | -0.7% |
| | 32 | +0.7% | +2.3% | -1.3% | -6.6% | +0.4% | -1.2% | -0.3% | -0.2% | +0.7% |
| | 64 | -0.4% | -0.7% | +0.2% | +2.1% | -0.2% | -0.2% | +0.8% | +1.0% | +0.7% |
| | 96 | +0.3% | -0.2% | -0.8% | -2.6% | -0.7% | +0.6% | -0.0% | +0.3% | +0.5% |
| | 128 | +0.9% | +0.2% | +0.2% | -0.0% | -0.3% | -2.6% | -0.2% | -0.5% | -0.0% |
| Improved - Perfect | 8 | -5.2% | -1.6% | -6.5% | +2.4% | -5.3% | -1.3% | +3.4% | +1.7% | +2.8% |
| | 16 | -7.4% | +1.7% | -2.6% | -0.7% | -2.1% | -6.7% | +2.0% | +0.1% | +1.2% |
| | 32 | -0.2% | -0.4% | +0.5% | -1.7% | +1.1% | +4.4% | -0.0% | -0.3% | +0.3% |
| | 64 | +0.9% | +0.4% | -0.1% | +0.7% | +0.4% | -0.9% | -0.1% | +0.3% | +0.6% |
| | 96 | -1.5% | -0.2% | +0.7% | +0.8% | +0.1% | +0.6% | +0.3% | +0.6% | -0.2% |
| | 128 | -0.4% | -0.1% | +0.1% | -0.2% | -0.5% | +0.1% | -0.3% | +0.3% | +0.1% |

Table 6.5: The absolute differences between variant 3 and 4

## 6.6. Discussion of the results

The four different variants provide an extensive answer to the main research question. For different fleet sizes, shared component strategies among different aircraft types and environmental conditions, the influence of the repair quality is quantified by capturing the changing values for the ADI and $CV^2$. When looking at table 6.6, the total-effect indices of Sobol's sensitivity analysis of the different variables are displayed (Sobol, 2001). The total-effect index translates the contribution to the output variance of the variable. Hence, the influence of the fleet size is dominant for both the variance in outcome of the ADI as the $CV^2$. Therefore, it can be concluded that the fleet size is the main influencing factor for both metrics. This implicates that adjusting the fleet size has the largest impact on potentially lower the ADI and $CV^2$. However, for many reasons, the expansion of the fleet is not always possible. In the cases where this expansion is not feasible and the fleet sizes cannot be increased, the influence of the repair quality on the demand pattern becomes more dominant. This can be seen in table 6.7, where the fleet size is assumed to be constant, hence variation in the ADI and $CV^2$ are not powered by the fleet size.

| | | **Fleet size** | **Repair quality** | **Performance difference aircraft** | **Environment** |
|---|---|---|---|---|---|
| S1 | **$CV^2$** | 0.818 | 0.083 | 0.028 | 0.028 |
| | **ADI** | 0.868 | 0.021 | 0.012 | 0.022 |
| ST | **$CV^2$** | 0.885 | 0.136 | 0.078 | 0.090 |
| | **ADI** | 0.942 | 0.052 | 0.049 | 0.063 |

Table 6.6: Sensitivity analysis: First- and Total-Effect indices

| | | **Repair quality** | **Performance difference aircraft** | **Environment** |
|---|---|---|---|---|
| S1 | **$CV^2$** | 0.427 | 0.182 | 0.204 |
| | **ADI** | 0.293 | 0.286 | 0.367 |
| ST | **$CV^2$** | 0.636 | 0.361 | 0.413 |
| | **ADI** | 0.354 | 0.339 | 0.440 |

Table 6.7: Sensitivity analysis: First- and Total-Effect indices with a constant fleet size

Another critical note has to be made regarding the values for the different repair quality scenarios. The obtained results from the data analysis are used as a reference scenario (the "Normal" repair quality), while the other three are based on a multiplication of this scenario.
The value of the parameter for the different repair qualities are exaggerated and are not based on a market research. In practice, the difference in performance will hardly be of this size. However, the advantage of these diverse scenarios is that the impact of changing repair quality is enlightened.

The assumption regarding the interdependency among components in the same ATA chapter results in the limitation of the usefulness of the outcome when it comes to the location of the failures and the corresponding failure patterns. Although components in the same ATA chapter could be interdependent, this does not have to be the case. Even more, components could be connected and dependent on components in different ATA chapters. However, when aggregating the results of the failures, the location of the components is not decisive for the outcome of this research.

As this research quantifies the impact of the repair quality on the different demand metrics, the repair quality is used as varying parameter in the model. However, the influence of the chance of other varying parameters might also influence the metrics. Hence, there is no proof that all changes come from the varying repair quality alone and the variance caused by interaction of the different parameters should be included as well. Table 6.6 provides the first- and total-effect indices of the sensitivity analysis. As can be seen from the table, the difference among the first- and total-effect indices are relatively small. Hence, the influence of the interaction is limited.

Another assumption is made in variant 4, where the effect of different performing aircraft in a fleet is tested. Due to the lack of research and data of the performance of different types of aircraft, only a rough estimation of possible deviations in performance could be made by the author. Furthermore, the values obtained from the data set are used as a reference scenario here, presenting the average performing aircraft types. Future research could be conducted in this field to verify this assumption.

Another important note should be made regarding the statistical outcomes of the KW $H$- and the Mann-Whitney $U$-tests. As the commonly chosen 95% interval provides a fair threshold for the rejection of the null hypothesis, $p$-values below 0.05 cause the rejection of the null hypothesis and thus the assumption that different groups of data have different medians. However, this $p$-value is highly dependent on the number of data points in the compared groups. As the number of iterations for the simulation is set to 50, the size of the subsets grow by a factor 50. Therefore, the $p$-values become smaller, resulting in a more frequent rejection of the null hypothesis. However, when reviewing only a single iteration, the $p$-values are higher and the null hypothesis is rejected less often. It is however not an option to exclude the iterations from the model, as these iterations provide the stability of the outcome by omitting the random factor.
A final critical note can be made on the limited set of drivers for failures. As stated frequently in previous research, not all drivers of failures are known, resulting in research that includes limited drivers. However, this research provides an broadening to the current knowledge by including the effect of different repair qualities.

# 7
# Conclusions and recommendations

The intermittent and lumpy demand for spare parts in aviation remains to be one of the striking problems for the industry to tackle. These demand patterns are challenging for MRO providers, as it is more difficult to keep their stocks of spare parts on the right level. High holding costs are the result, driving the maintenance costs of the industry.

Multiple studies have come up with new forecasting methods to improve the accuracy of the forecasts. Time-series methods have proven to increase their performance over the years, but these methods do not provide further understanding of underlying processes that drive the intermittent or lumpy demand for components. Therefore, the focus is shifting towards the comprehension of the drivers for intermittent and lumpy demand. These studies already showed the influence of some major demand drivers, such as the fleet size. However, the recommendations for future work are still focused on finding more and more of these drivers, and to take out the assumptions in the current models.

This formed the foundation of this research. The demand of spare parts is driven by the failures or removals of these parts. As the author noted that the repair process is not taken into account in these models despite of being of a huge influence of the characteristics of the lifetime of a component, the foundation for the research was built. The theory of the Branching Poisson Process is used to describe the phenomenon of subsequent failures as a result of a defect component placed back into service, resulting in failures of interdependent components (subsequent failures). These subsequent failures occur within a short time span after the first failure. Therefore, this drives the clustering of failures, resulting in peaks of failures and hence the demand, one of the properties of lumpy demand patterns.
Data containing removal data of aircraft build by a single manufacturer is used to find the occurrence rates of these peaks of primary and subsidiary failures. By using varying repair qualities, fleet sizes and environmental conditions, the influence of these factors are taken into account. Finally, an extension with an eye on the future has been made. As recent studies showed that an increasing component commonality across different aircraft types is beneficial for costs related to maintenance, this shared component pool in a heterogeneous fleet is considered in the model.
Based on the four defined repair qualities, the impact of an improving repair quality is further examined. The outcome of the model can be used by MRO providers to quantify the changes in ADI, $CV^2$ and total number of failures when improving the maintenance department of the company.

By observing the results from the model without the manipulation for different environmental conditions or shared component strategy, the results show the impact of the solely varying repair quality. This scenario is simulated in variant 1. The results show that an improvement of repair quality rises the ADI by a value in the range of 12.6%-15.2%, while decreasing the $CV^2$ by a value in the range of 1.1%-3.9%. Although this does not seem to have a positive effect on the predictability of the failures, the total number of failures over time drops significantly. The drop of failures over time decreases by a value in the range of 15.6%-20.1%. Therefore, it might be considerable in some cases to accept the growth in ADI as the $CV^2$ and the total number of failures are of major importance.

As previous research has showed that the fleet size is one of the major drivers of the change of the demand pattern for spare parts, a breakdown of the results for varying fleet sizes has been investigated. From this breakdown, it could be concluded that smaller fleet sizes have higher deviations for the ADI (positive when improving the quality of the repairs) and the $CV^2$ (lower when improving the quality of repairs). Moreover, the decrease in total number of failures occurs in all situations and drops with a rate between 10.0%-21.6%, dependent on the fleet size and the improvement of the repair quality. As the problem of unpredictable behavior of spare parts demands become smaller for increasing fleet sizes, the small deviations in ADI and $CV^2$ should therefore be no holdback to increase the repair quality for a MRO provider.
Remarkable is that the decrease of the $CV^2$ are negligible or even not present for the improvement from the reference repair quality to the "Improved" repair quality. An explanation for this odd phenomenon is that the amount of subsequent failures is lowered, resulting in a less failures. As less failures occur, the $CV^2$ is more sensitive for a (possible) large offset from a primary failure, leading to an increased standard deviation, and thus an increased $CV^2$. For this improvement, the increased standard deviation is not sufficiently compensated by the higher mean number of failures.

Next, the influence of the environmental conditions is quantified. As operating aircraft in humid and desert conditions reduced the expected lifetime of components, this has an impact on the amount of subsequent failures and hence the peaks in the failure patterns over time.
It is concluded that for humid and desert conditions, similar patterns are obtained for the ADI, $CV^2$ and total failures. Hence, the change in ADI and $CV^2$ damps out when the fleet size increases and the relative reduction of the total number of failures remains somewhat constant over varying fleet sizes, repair qualities and environmental conditions. Again, the repair quality improvement from the reference scenario to the "Improved" scenario has limited positive to in some cases negative effects for the $CV^2$.
Overall, the increment of the ADI for the different fleet sizes and improvements is the most limited for the humid environmental conditions, the performance of the $CV^2$ and (to some extend) the reduction of failures is similar across the different environmental conditions. However, the reference environmental scenario performs slightly better on the reduction of the number of failures in most cases.

Finally, the model quantifies the impact of a hypothetical new strategy for the aviation industry. By enlarging the usability of components across different aircraft, significant cost reductions related to the purchase and holding of components can be realized. When this strategy would be implemented in the industry, different types of aircraft can use the same sets of components, resulting in artificially increasing the fleet size of aircraft that can use the set of components. The question arises whether different performing aircraft have a significant effect on the failure patterns of the components when compared to a homogeneous fleet.
The outcome of the model showed that, although there are notable deviations, no new patterns emerge. Hence, the increase of ADI and decrease of $CV^2$ damp out as the fleet size becomes larger, while the increasing fleet size has less impact on the relative deviation of the total number of failures. Hence, the introduction of components that can be operated on different types of aircraft with different performances has no significant positive or negative effect on the predictability of the failures of these components compared to a homogeneous fleet.

Summarizing the above-mentioned outcome of this research, the impact of changing repair quality on the predictability of the failures of components has been quantified. In general, an improvement of repair quality induces increased ADI, a reduced $CV^2$ and a reduction of the total number of failures. A critical note should be made regarding larger fleet sizes (more than 64 aircraft of the same type), as the effect of the increased repair quality on the ADI and $CV^2$ become less significant, while the effect on the total failures remains the same. Therefore, one could conclude that when facing larger fleets, the improvement of the repair quality has a wider support base, as the downside of the implementation become smaller. Ironically, larger fleets have less problems with the grip on the predictability of spare parts, as is proven in this research.
The author can not judge for an individual MRO provider whether or not the repair quality should be improved, as insights in the current failure patterns is required. Hence, the industry should use the outcome of this research as input for case studies to improve the repair quality.

With this research, another step towards the full understanding of the drivers for failures of components is taken. With the influence of the repair quality on the failure behavior of components, one common-made assumption is taken out. Future research in the variance of the repair quality among different MRO providers

is advised in order to strengthen the outcome of the model. Furthermore, not only incorrect repairs, leaving the component in the same broken state as before, but minimal and imperfect repairs should be considered as well. This enables the model to implement a more realistic representation of the repair process, instead of the current two-sided option.
If the aviation industry decides to further investigate the possibilities regarding the component pooling across multiple aircraft types, a detailed analysis of the performance of the individual aircraft types leads to a more accurate prediction of the performance of a heterogeneous fleet compared to a homogeneous fleet.
Finally, the author advises a follow-up study that reveals the interdependencies among flight safety-critical components for different types of aircraft. The result of this study would contribute to the practical relevance regarding the patterns and locations of the primary and subsidiary failures over time.

# Bibliography


Abernethy, R. B., Breneman, J. E., Medlin, C. H., and Reinman, G. L. (1983). Weibull analysis handbook. Technical report, Pratt and Whitney West Palm beach fl Government Products DIV.

Air Transport World (2019). World airline fleets. `https://aviationweek.com/2019-world-airline-report-world-airline-fleets`. Accessed: 2020-09-30.

Airworthiness, S. and Airspace Regulation Group, A. H. (2015). Aircraft Maintenance Incident Analysis. Technical report, CIVIL AVIATION AUTHORITY.

Artiba, A., Riane, F., Ghodrati, B., and Kumar, U. (2005). Reliability and operating environment-based spare parts estimation approach. *Journal of Quality in Maintenance Engineering*.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific data*, 5:180214.

Belobaba, P., Odoni, A., and Barnhart, C. (2015). *The global airline industry*. John Wiley & Sons.

Boeing (2008). Flight Interruption Cost Cstimates. Technical report, Boeing Commercial Airplanes.

Boone, C., Craighead, C., and Hanna, J. (2008). Critical Challenges of Inventory Management in Service Parts Supply: A Delphi study. *The Journal of Operations Mangement Research*, 1:31–39.

Chen, D. and Chen, H. W. (2013). Using the köppen classification to quantify climate variation and change: An example for 1901–2010. *Environmental Development*, 6:69–79.

Cox, D. R. (1966). The statistical analysis of series of events. *Monographs on Applied Probability and Statistics*.

Croston, J. D. (1972). *Forecasting and Stock Control for Intermittent Demands*, volume 23. Taylor & Francis.

Dhillon, B. S. (2002). *Engineering maintenance: a modern approach*. cRc press.

Dhillon, B. S. (2013). *Human reliability: with human factors*. Elsevier.

Dhillon, B. S. and Liu, Y. (2006). Human error in maintenance: A review. *Journal of Quality in Maintenance Engineering*, 12:21–26.

FAA (2010). Eligibility and Evaluation of U.S. Military Surplus Flight Safety Critical Aircraft Parts, Engines, and Propellers. Technical report, FAA.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage. Chapter 11.

Ghobbar, A. A. and Friend, C. H. (2002). Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8(4):221–231.

Ghobbar, A. A. and Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30(14):2097–2114.

Ghodrati, B. (2005). *Reliability and operating environment based spare parts planning*. PhD thesis, Luleåtekniska universitet.

Glen, S. (2016). Kruskal Wallis H Test: Definition, Examples and Assumptions. `https://www.statisticshowto.com/kruskal-wallis/`. Accessed: 2020-10-10.

Guo, F., Diao, J., Zhao, Q., Wang, D., and Sun, Q. (2017). A double-level combination approach for demand forecasting of repairable airplane spare parts based on turnover data. *Computers & Industrial Engineering*, 110:92–108.

Hackworth, C., Holcomb, K., Dennis, M., Goldman, S., Bates, C., Schroeder, D., and Johnson, W. (2007). An international survey of maintenance human factors programs. Technical report, FEDERAL AVIATION ADMINISTRATION OKLAHOMA CITY OK CIVIL AEROMEDICAL INST.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., et al. (1998). *Multivariate data analysis*, volume 5. Prentice hall Upper Saddle River, NJ. Chapter 8.

Hobbs, A. (2008). An Overview of Human Factors in Aviation Maintenance. Technical report, Australian Transport Safety Bureau - Australian Government.

Hussain, M. and Mahmud, I. (2019). pymannkendall: a python package for non parametric mann kendall family of trend tests. *Journal of Open Source Software*, 4(39):1556.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.

Kendall, M. (1975). Rank correlation methods. 2nd impression. *Charles Griffin and Company Ltd. London and High Wycombe*.

Kolmogorov, A. (1933). Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, 4:83–91.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

Latorella, K. A. and Drury, C. G. (1992). A framework for human reliability in aircraft inspection. In *Proceedings of the Seventh Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection*, pages 71–82.

Levene, H. (1960). Contributions to probability and statistics. *Essays in honor of Harold Hotelling*, pages 278–292.

Lewis, P. A. W. (1964). A branching Poisson process model for the analysis of computer failure patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(3):398–441.

Lewis, P. A. W. (1967). Non-homogeneous branching Poisson processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 29(2):343–354.

Liang, G.-F., Lin, J.-T., Hwang, S.-L., Wang, E. M.-y., and Patterson, P. (2010). Preventing human errors in aviation maintenance using an on-line maintenance assistance platform. *International Journal of Industrial Ergonomics*, 40(3):356–367.

Lowas, A. F. and Ciarallo, F. W. (2016). Reliability and operations: Keys to lumpy aircraft spare parts demands. *Journal of Air Transport Management*, 50:30–40.

Lyte, B. (2016). How Are Planes Decommissioned, and How Much Value Can Be Salvaged From Their Parts?

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, pages 245–259.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60.

Planespotters.net (2020). Global airline fleet size overview. `https://www.planespotters.net/airlines`. Accessed: 2020-09-30.

Qian, C. and Chan, C.-Y. (2015). A new production approach for compensating forecast error and customer loss in waiting. *International Journal of Production Research*, 53(5):1325–1336.

Rankin, W., Hibit, R., Allen, J., and Sargent, R. (2000). Development and evaluation of the Maintenance Error Decision Aid (MEDA) process. *International Journal of Industrial Ergonomics*, 26(2):261–276.

Regattieri, A., Gamberi, M., Gamberini, R., and Manzini, R. (2005). Managing lumpy demand for aircraft spare parts. *Journal of Air Transport Management*, 11(6):426–431.

Rigdon, S. E. and Basu, A. P. (2000). *Statistical methods for the reliability of repairable systems.* Wiley New York.

Romeijnders, W., Teunter, R., and Jaarsveld], W. v. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, 220(2):386–393.

Smith, A. M. and Hinchcliffe, G. R. (2003). *RCM–Gateway to world class maintenance.* Elsevier.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280.

Student (1908). The probable error of a mean. *Biometrika*, pages 1–25.

Syntetos, A. A. and Boylan, J. E. (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics*, 71(1):457–466.

Syntetos, A. A. and Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21(2):303–314.

Thijssens, O. W. M. and Verhagen, W. J. C. (2020). Application of extended cox regression model to time-on-wing data of aircraft repairables. *Reliability Engineering & System Safety*, 204:107136.

Tracht, K., von der Hagen, F., and Schneider, D. (2013). Applied Repairable-item Inventory Modeling in the Aviation Industry. *Procedia CIRP*, 11:334–339.

Van der Auweraer, S., Boute, R. N., and Syntetos, A. A. (2019). Forecasting spare part demand with installed base information: A review. *International Journal of Forecasting*, 35(1):181–196.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python tutorial.* Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Verhagen, W. (2019a). AE4465: Maintenance Modeling & Analysis, Lecture 3 - Modelling Non-Repairables.

Verhagen, W. (2019b). AE4465: Maintenance Modeling & Analysis, Lecture 5 - Modelling Repairables.

Verhagen, W. (2019c). AE4465: Maintenance Modeling & Analysis, Lecture 7 - Supportability Engineering - basic concepts & forecasting spare parts demand.

Verhagen, W. (2019d). AE4468: Airline Maintenance Operations, Lecture 1 - Introduction to airline maintenance operations.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4):330–336.

West, D. and Bradley, J. (2008). Airline flight networks, cycle times, and profitability: 2004–2006. *Operations Management Research*, 1:129–140.

Wong, H., Oudheusden, D. V., and Cattrysse, D. (2007). Cost allocation in spare parts inventory pooling. *Transportation Research Part E: Logistics and Transportation Review*, 43(4):370–386.

Zhang, Y., Zheng, Y., Xianchao, M., Wenjun, D., Dayong, D., Zhaoguang, T., and Zhang, S. (2019). Exploration and implementation of commonality valuation method in commercial aircraft family design. *Chinese Journal of Aeronautics*, 32(8):1828–1846.

# A

# Clustered removals analysis code

```
clustercalculations = 1

if clustercalculations == 1:
    location = -1
    counter = 0
    Cluster = pd.DataFrame(columns = ['Aircraftnumber','ATA␣location','ACtype','Date','Primary','Subsequent␣1','
        Subsequent␣2','Subsequent␣3', 'Subsequent␣4', 'Subsequent␣5','Subsequent␣6', 'Subsequent␣7', 'Subsequent␣8','
        Subsequent␣9', 'Subsequent␣10','Subsequent␣11','Subsequent␣12','Subsequent␣13', 'Subsequent␣14', 'Subsequent␣15
        ','Subsequent␣16', 'Subsequent␣17', 'Subsequent␣18','Subsequent␣19', 'Subsequent␣20','Subsequent␣21','
        Subsequent␣22','Subsequent␣23', 'Subsequent␣24', 'Subsequent␣25','Subsequent␣26', 'Subsequent␣27', 'Subsequent␣
        28','Subsequent␣29', 'Subsequent␣30','Subsequent␣31','Subsequent␣32','Subsequent␣33', 'Subsequent␣34', '
        Subsequent␣35','Subsequent␣36', 'Subsequent␣37', 'Subsequent␣38','Subsequent␣39', 'Subsequent␣40','Subsequent␣
        41','Subsequent␣42','Subsequent␣43', 'Subsequent␣44', 'Subsequent␣45','Subsequent␣46', 'Subsequent␣47', '
        Subsequent␣48','Subsequent␣49', 'Subsequent␣50','Subsequent␣51','Subsequent␣52','Subsequent␣53', 'Subsequent␣54
        ', 'Subsequent␣55','Subsequent␣56', 'Subsequent␣57', 'Subsequent␣58','Subsequent␣59', 'Subsequent␣60'],index =
        range(0,nrows))

    for i in range(len(SerialNumbers)): #looping over all the aircraft in data
        Aircrafts["Aircraft{0}".format(i+1)] = Aircrafts["Aircraft{0}".format(i+1)].sort_values(["AtaCat","Date"])
        # Start with reference point j = 0 and candidate point t = j + 1
        j=0
        t = j + 1
        while j < len(Aircrafts["Aircraft{0}".format(i+1)])+1:
            if j > 0: # in order to compensate for j+=1 a few lines down
                j = j - 1
            if t < len(Aircrafts["Aircraft{0}".format(i+1)]) and j-1 < len(Aircrafts["Aircraft{0}".format(i+1)]) and list
                (Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j] == list(Aircrafts["Aircraft{0}".format(i+1)]["
                AtaCat"])[t]: #Check if reference and candidate point have same ATA chapter
                if abs((list(Aircrafts["Aircraft{0}".format(i+1)]["Date"])[t] - list(Aircrafts["Aircraft{0}".format(i+1)
                    ]["Date"])[j]).days) < 14:    #Matching ATA and within 14 days, so at the end of the step: move to
                    the next candidate point (t = t + 1)
                    if counter == 0: #if this is the first candidate point, the reference point is not yet added to the
                        Cluster dataframe and should therefore be added.
                        location += 1
                        Cluster.iloc[location,:] = pd.Series([list(Aircrafts["Aircraft{0}".format(i+1)]["
                            AircraftSerialNumber"])[j],list(Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j],list(
                            Aircrafts["Aircraft{0}".format(i+1)]["AircraftTypeId"])[j],
                        list(Aircrafts["Aircraft{0}".format(i+1)]["Date"])[j], list(Aircrafts["Aircraft{0}".format(i+1)][
                            "PartSerialNumber"])[j], np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np
                            .nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.
                            nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan
                            , np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan,
                            np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.
                            nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan], index = Cluster.columns)
                        counter = 1 #Now, the counter is set to 1 in order to avoid overwriting of the reference point
                            that is just added before
                        Cluster.iloc[location]["Subsequent␣{0}".format(t-j)] = list(Aircrafts["Aircraft{0}".format(i+1)][
                            "PartSerialNumber"])[t] # And the serialnumber of the first candidate point that satisfies
                            the conditions is added here
                    else:
                        Cluster.iloc[location]["Subsequent␣{0}".format(t-j)] = list(Aircrafts["Aircraft{0}".format(i+1)][
                            "PartSerialNumber"])[t] #if this is not the first candidate point that satisfies the
                            conditions, it is added as the (t-j)th subsequent removal
                    t += 1 # the next candidate point can be checked now
                elif str(list(Aircrafts["Aircraft{0}".format(i+1)]["PartSerialNumber"])[j]) == str(Cluster.iloc[location
                    ]["Primary"]): # if the reference point is already added, due to the fact that is had a candidate
                    point that matched the conditions, move forward to the next reference point and corresponding
                    candidate point
                    j = t
```

```
                t = j + 1
                counter = 0
            else: #same ATA but not within 14 days, so added the reference point and move forward with a new
                  reference point and candidate points
                location += 1
                Cluster.iloc[location,:] = pd.Series([list(Aircrafts["Aircraft{0}".format(i+1)]["AircraftSerialNumber
                    "])[j],list(Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j],list(Aircrafts["Aircraft{0}".
                    format(i+1)]["AircraftTypeId"])[j],
                list(Aircrafts["Aircraft{0}".format(i+1)]["Date"])[j], list(Aircrafts["Aircraft{0}".format(i+1)]["
                    PartSerialNumber"])[j], np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan,
                     np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan
                    , np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.
                    nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np
                    .nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan,
                    np.nan, np.nan, np.nan, np.nan], index = Cluster.columns)
                j = t
                t = j + 1
                counter = 0
        elif t < len(Aircrafts["Aircraft{0}".format(i+1)]) and j-1 < len(Aircrafts["Aircraft{0}".format(i+1)]) and
              list(Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j] != list(Aircrafts["Aircraft{0}".format(i+1)]["
              AtaCat"])[t] and t-j == 1: #non matching ATAs, so add reference point and move on to next reference and
               candidate points
            location += 1
            Cluster.iloc[location,:] = pd.Series([list(Aircrafts["Aircraft{0}".format(i+1)]["AircraftSerialNumber"])[
                j],list(Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j],list(Aircrafts["Aircraft{0}".format(i+1)
                ]["AircraftTypeId"])[j],
            list(Aircrafts["Aircraft{0}".format(i+1)]["Date"])[j], list(Aircrafts["Aircraft{0}".format(i+1)]["
                PartSerialNumber"])[j], np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.
                nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan
                , np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan,
                np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.
                nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan
                , np.nan], index = Cluster.columns)
            j = t
            t = j + 1
            counter = 0
        elif t < len(Aircrafts["Aircraft{0}".format(i+1)]) and j-1 < len(Aircrafts["Aircraft{0}".format(i+1)]) and
              list(Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j] != list(Aircrafts["Aircraft{0}".format(i+1)]["
              AtaCat"])[t] and t-j != 1: #non matching ATAs and since t-j !=1, the reference point is already added,
              so just move on to the next points
            #print("ongelijke ATA", j,t)
            j = t
            t = j + 1
            counter = 0
        else: #candidate points out of range, set final point
            #print("te lang", j,t)
            j = len(Aircrafts["Aircraft{0}".format(i+1)])
            counter = 0
        j+=1 #move to next point which is later undone, this in order to let the while-loop running and checking
        if j == len(Aircrafts["Aircraft{0}".format(i+1)]) and Cluster.iloc[location]["Subsequent␣1"] != list(
              Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j-1]: #if the last reference point is not yet added
              since it was not the subsequent point of the previous check, then add it to the Cluster DataFrame
            location += 1
            Cluster.iloc[location,:] = pd.Series([list(Aircrafts["Aircraft{0}".format(i+1)]["AircraftSerialNumber"])[
                j-1],list(Aircrafts["Aircraft{0}".format(i+1)]["AtaCat"])[j-1],list(Aircrafts["Aircraft{0}".format(
                i+1)]["AircraftTypeId"])[j-1],
            list(Aircrafts["Aircraft{0}".format(i+1)]["Date"])[j-1], list(Aircrafts["Aircraft{0}".format(i+1)]["
                PartSerialNumber"])[j-1], np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan,
                np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.
                nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan
                , np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan,
                np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.nan, np.
                nan, np.nan], index = Cluster.columns)

    print('Done␣with␣aircraft', i)

Cluster = Cluster[Cluster["Aircraftnumber"].notna()]
```

# B
# Figures of variant 3

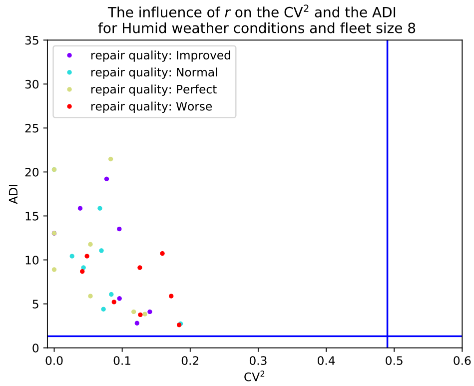



Figure B.1: V3: Visual results - Humid, fleet size 8

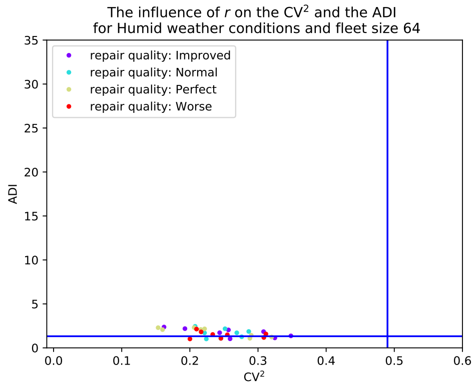



Figure B.4: V3: Visual results - Humid, fleet size 64

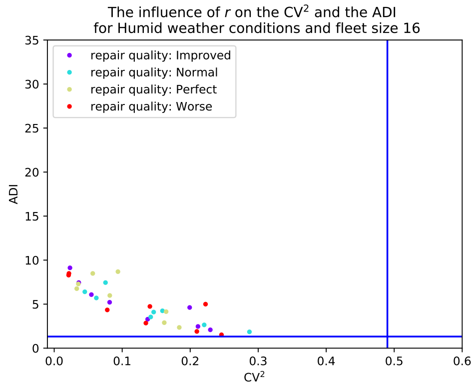



Figure B.2: V3: Visual results - Humid, fleet size 16

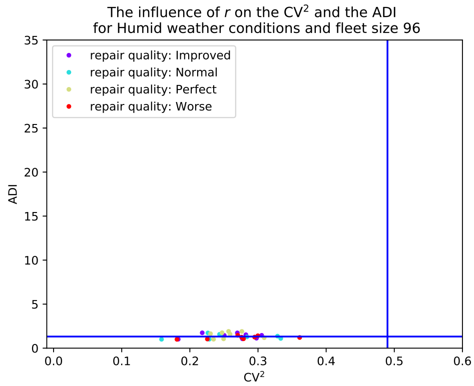



Figure B.5: V3: Visual results - Humid, fleet size 96

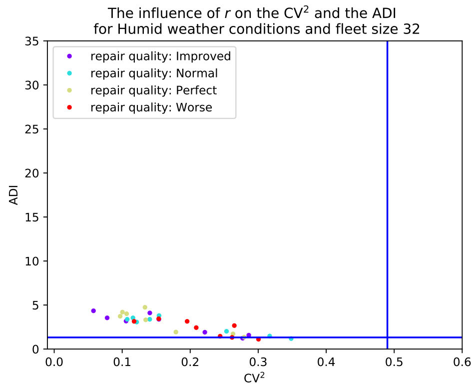



Figure B.3: V3: Visual results - Humid, fleet size 32

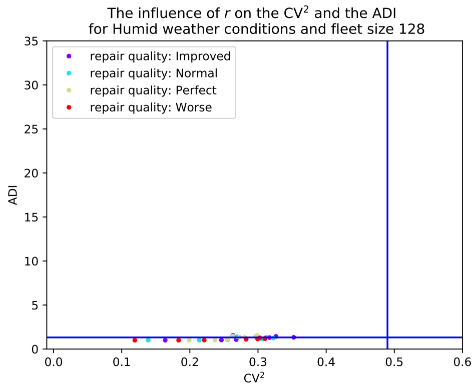



Figure B.6: V3: Visual results - Humid, fleet size 128

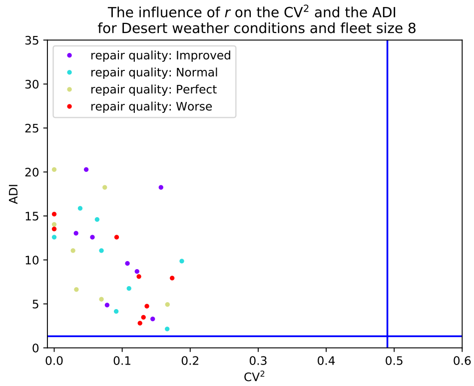



Figure B.7: V3: Visual results - Desert, fleet size 8

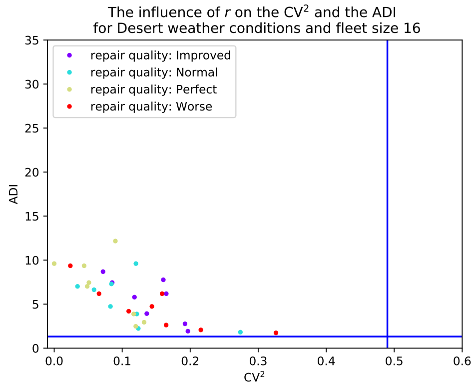



Figure B.8: V3: Visual results - Desert, fleet size 16

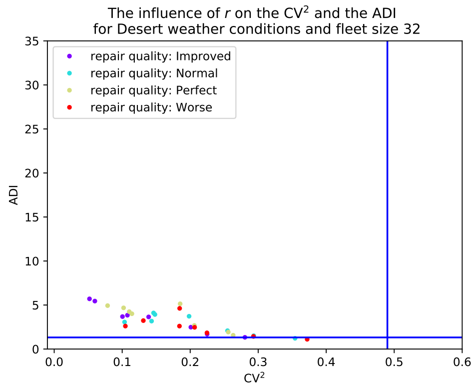



Figure B.9: V3: Visual results - Desert, fleet size 32

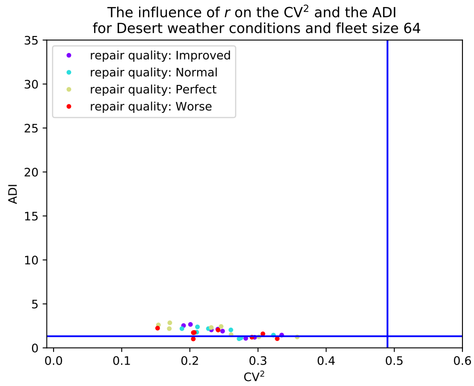



Figure B.10: V3: Visual results - Desert, fleet size 64

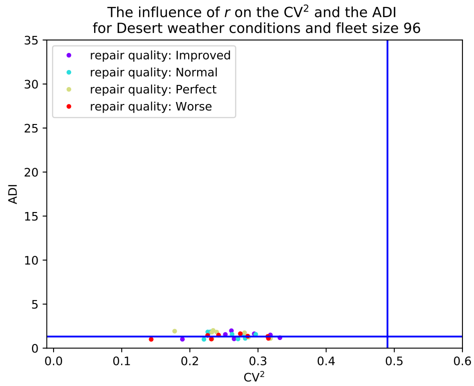



Figure B.11: V3: Visual results - Desert, fleet size 96

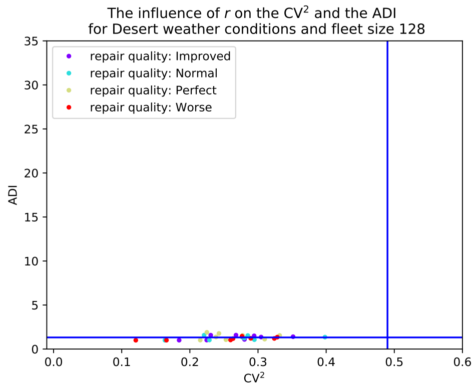



Figure B.12: V3: Visual results - Desert, fleet size 128

# C
# Figures of variant 4

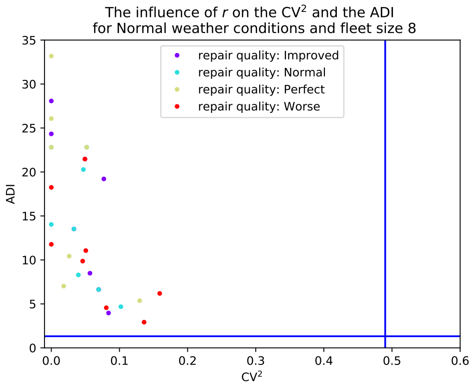



Figure C.1: V4: Visual results - Normal, fleet size 8

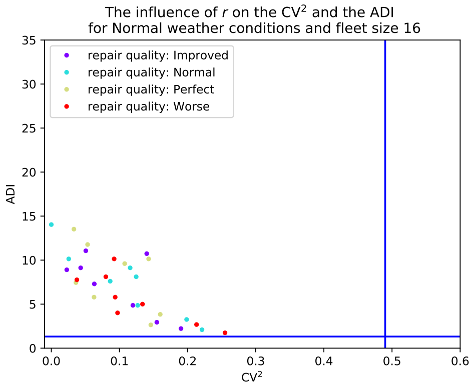



Figure C.2: V4: Visual results - Normal, fleet size 16

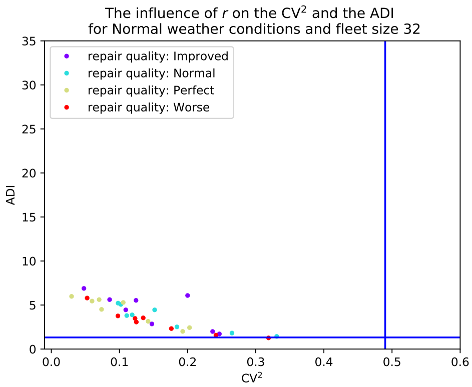



Figure C.3: V4: Visual results - Normal, fleet size 32

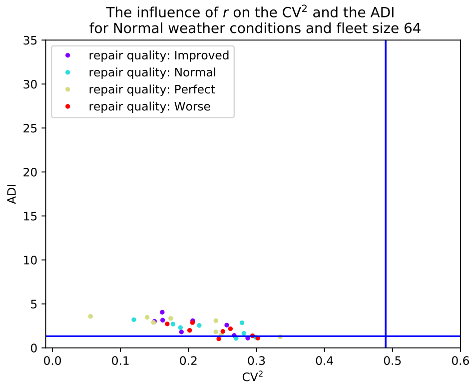



Figure C.4: V4: Visual results - Normal, fleet size 64

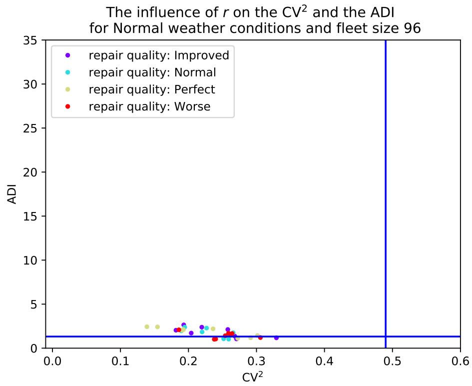



Figure C.5: V4: Visual results - Normal, fleet size 96

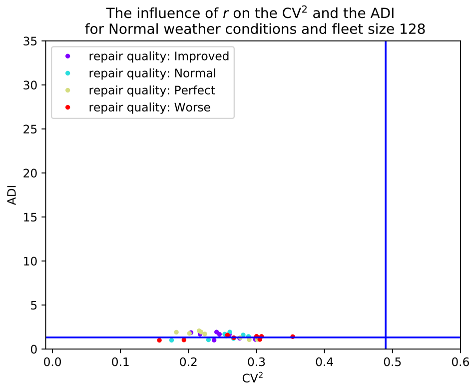



Figure C.6: V4: Visual results - Normal, fleet size 128

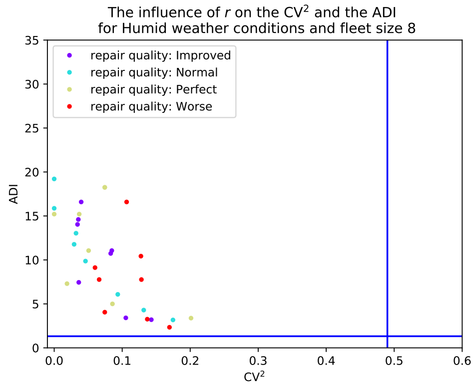



Figure C.7: V4: Visual results - Humid, fleet size 8

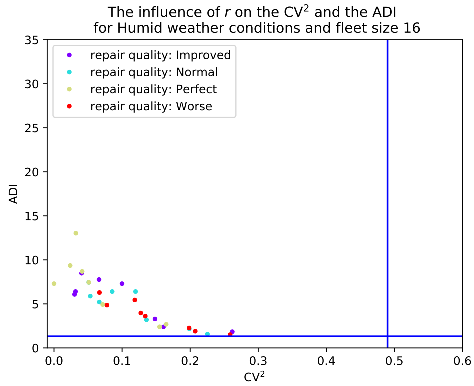



Figure C.8: V4: Visual results - Humid, fleet size 16

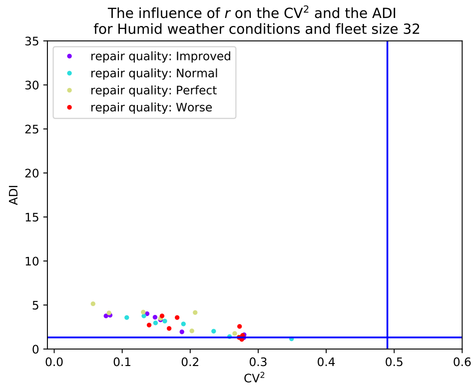



Figure C.9: V4: Visual results - Humid, fleet size 32

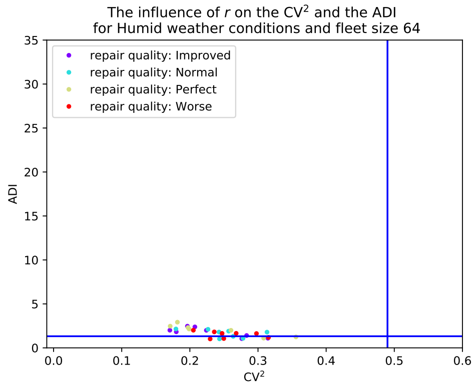



Figure C.10: V4: Visual results - Humid, fleet size 64

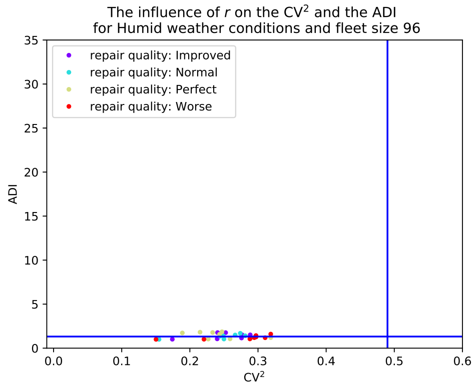



Figure C.11: V4: Visual results - Humid, fleet size 96

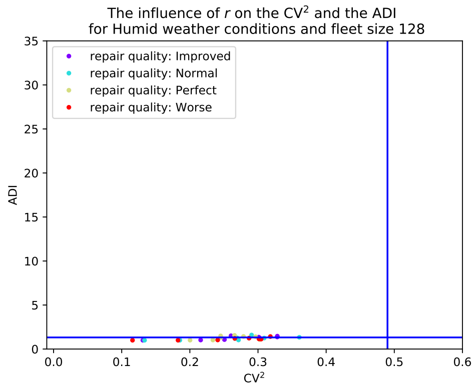



Figure C.12: V4: Visual results - Humid, fleet size 128

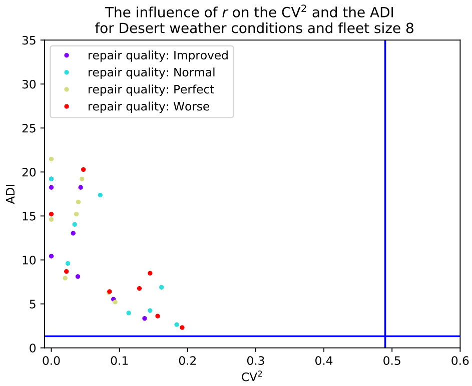



Figure C.13: V4: Visual results - Desert, fleet size 8

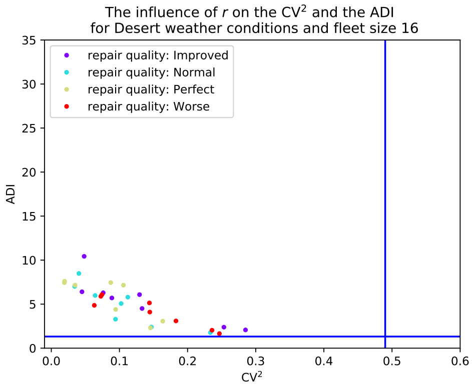



Figure C.14: V4: Visual results - Desert, fleet size 16

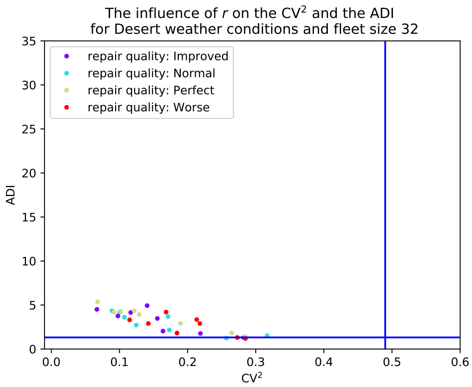



Figure C.15: V4: Visual results - Desert, fleet size 32

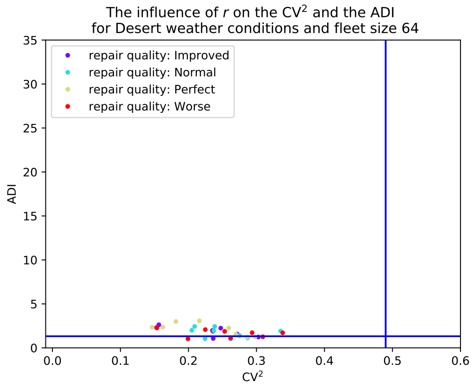



Figure C.16: V4: Visual results - Desert, fleet size 64

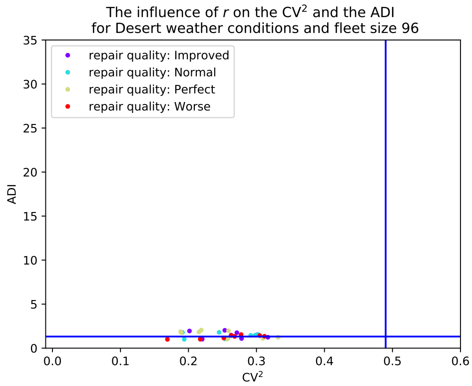



Figure C.17: V4: Visual results - Desert, fleet size 96

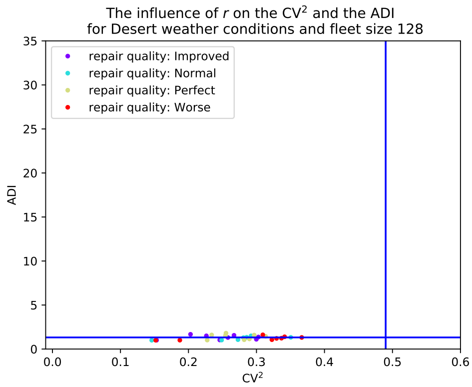



Figure C.18: V4: Visual results - Desert, fleet size 128