Dynamic human resource management decision support model

Based on tactical aircraft maintenance demand forecasts

B. Slangen

09-07-2020

TU Delft

Dynamic human resource management decision support model based on tactical aircraft maintenance demand forecasts

B. Slangen

to obtain the MSc degree Aerospace Engineering
at the Delft University of Technology,
to be defended publicly at Thursday July 9th, 2020 at 10:00 AM.

Studentnumber:	4278070
Project duration:	September 1,2019 - July 9,2020
Thesis committee:	Prof. dr. ir. J.M. Hoekstra Dr. A.J. Cabo Ir. P.C. Roling Dr. ir. V.S.V. Dhanisetty

A digital version of this thesis is available at http://repository.tudelft.nl
The cover photo is acquired from iStock.com, Stockfoto ID: 905918352

Preface

This thesis concludes my seven years of study at Delft University of Technology. My time here in Delft has been wonderful and enriched me in every way thinkable, inside and outside the faculty of Aerospace Engineering.

This research has been a bumpy ride but therefore also a very educational one. Although I gained quite some knowledge during this process I expect that mostly the skills I have developed will stay with me for the rest of my life.

I certainly would not have been able to complete this journey without the help of many people. First of all my supervisor Viswanath Dhanisetty was of invaluable help. You especially learned me how to communicate my findings in an efficient and clear way. I also would like to thank Wim Verhagen for his advice and his encouragement to think about the broader picture of my research. Lotte, you have been my 'rock in the sea' and made sure that I did not lose myself in this project by reminding me to put this whole thing in perspective once in a while. I want to thank my parents for their unconditional support in the last seven years and all the years before. Ultimately you gave me self confidence, which is I think the greatest gift parents can give their child.

Last but not least I would like to thank "The Boss" Bruce Springsteen. His beautiful music smoothened the bumpy ride a bit and according to my Spotify history I made him a lot of money last year.

Bram Slangen
July, 2020

Introduction

This document is the thesis document in which the thesis deliverables are combined. For the purpose of the defense the paper given in Ch. 1 is most relevant. This paper presents the applied methodology and the most important results of the research. For a more extensive literature study regarding demand forecasting models the reader is referred to Ch. 2. The original project plan is given in Ch. 3. These last two parts are already graded and are there just for the interest of the reader.

The different parts of this report are stand-alone documents and should be regarded as such.

Contents

Chapter 1

Paper

In this chapter the paper is presented.

Dynamic human resource management decision support model based on tactical aircraft maintenance demand forecasts

Bram Slangen

Delft University of Technology

June 23, 2020

Abstract

The long and therefore expensive training of aircraft maintenance technicians underline the need for accurate demand forecasts that allow for dynamic control of acquisition and training rate of personnel. This control enables human resource management to react swiftly to increases in workforce demand at times of technician shortages. To help human resource management a novel decision support model based on tactical demand forecasts in the aircraft maintenance context is proposed in this paper. Additionally, this paper presents a systematic research towards the optimal models to forecast tactical maintenance demand. The analysis is conducted using aggregated structural repair data of a fleet of wide-body passenger aircraft in the first ten years of its introduction. The results of this study show the potential of the proposed model as it is robust for varying amounts of non-constant workforce outflow and different fleet sizes. Furthermore, the model can be applied efficiently from one year after the acquisition of the first new aircraft. The novelty of this study is the direct integration of personnel training and acquisition with workforce demand forecasts. Additional research is recommended to validate the use of this model on other aircraft types, to explore the use of this model in the area of human resource management optimization and to extent this model to an organizational level.

I. Introduction

Workforce management is vital to aircraft maintenance organizations as personnel costs are one of the highest items of expense (Wahyudin et al., 2016). Increasing competition among Maintenance Repair and Overhaul organizations (MROs) forces them to enhance their efficiency hence improving workforce capacity planning (Maintenance Cost Task Force, 2018),(Phillips et al., 2009). Additionally, Original Equipment Manufacturing companies (OEMs) are enlarging their market share in the aircraft aftermarket. These developments complicate human resource management and consequently capacity planning for MROs as they have difficulties to retain and acquire well trained personnel (Prentice et al., 2017). Next to the pressing retention issues the MRO's relatively old workforce shows a prospect of large portions of the staff leaving maintenance industry due to retirement (Bill, 2018),(Constanza et al., 2017). The aforementioned developments cause a global shortage of aircraft technicians.

To cope with the shortage of skilled personnel airlines and MROs heavily invest in the schooling of young people through their own training facilities or in cooperation with maintenance schools (Zuehlke, 2014). Aircraft technicians are bound to strict certification requirements and the training times are long (EASA, 2015). The long training duration and the general shortage on personnel makes it difficult to respond quickly to unexpected increases in maintenance demand.

These unexpected increases in demand often occur when a fleet of novel aircraft is acquired by an airline. Recently, the application of composite materials in the new generation of passenger aircraft was accompanied with a promise of significant reduction in maintenance (Airbus, 2016),(Boeing, 2006). However, after introduction of the aircraft, maintenance was more extensive than anticipated. The susceptibility of composite components for damage by impacts, increased the maintenance demand dramatically (Drew and Mouawad, 2013),(Cohan, 2015). For an MRO organization to efficiently cope with this unexpected demand, forecasts are required that can predict the required resources on a tactical level. Tactical in this context refers to a timescale of two to three years and a fleet level approach. No method that approached this problem on this tactical level could be identified in literature by the author. Additionally, inquiries of various human resource management academics also confirm the lack of knowledge in this particular field of study.

i. Aim of the study

Based on the lack of knowledge and the present need in aircraft maintenance industry the goal of this study is to aid human resource management in their decision making on a tactical level regarding acquisition and training of personnel. A maintenance demand based approach has not been tried yet in literature. Consequently, the research question is as follows:

> "How can dynamic, tactical maintenance demand forecasts aid human resource management regarding personnel acquisition and training in the aircraft maintenance context?"

This paper aims to answer this question by proposing and evaluating a workforce flow model that uses dynamic, tactical workforce demand forecasts as input to produce personnel acquisition and training rates. Analysis of the model characteristics, its sensitivity to different parameter values and its performance then allow for an assessment of its value to human resource management. The tactical workforce demand forecasts are obtained through temporal aggregation. This is a method that is developed to model intermittent demand (Nikolopoulos et al., 2011),(Kourentzes et al., 2014). Subsequently, Ordinary Least Squares, Weighted Least Squares and Feasible Generalized Least Squares models are applied to the aggregated demand in order to forecast it. A forecast performance based model selection procedure is executed to find the optimal forecasting model.

The novelty of this study is the direct integration of a personnel training and acquisition model with workforce demand forecasts. Furthermore, to the authors knowledge, the approach of this problem on a tactical level (2 to 3 years timescale and on fleet level) is not yet attempted in an aircraft maintenance context.

ii. Paper structure

The remainder of the paper is structured as follows. In Sec. II a summarized literature background study is given, which is primarily focused on the modeling of maintenance demand. Sec. III is subdivided into six parts where the first two describe the design of the personnel flow model. The last four parts elaborate on the available data and the selection of the optimal demand forecasting models. In Sec. IV the results of the demand forecasting model selection process are presented after which the selected model is applied in the workforce flow model. The performance of the latter is also evaluated in this section. In Sec. V and VI the discussion of the results and the general conclusion are presented respectively.

II. Background study

To place the proposed model in a literature context, and to find applicable methods to forecast demand, a background study is conducted, which is presented in the following paragraphs.

i. Maintenance capacity planning

Human resource management and workforce demand forecasting are inherently related to capacity planning. The literature on capacity planning in a long-term context is scarce and there are, to the author's knowledge, no readily available methods to forecast workforce demand. However, the importance of maintenance demand predictions is highlighted often in various studies. According to Heimerl and Kolish the first step in human resource capacity planning is the determination of the workforce required per skill and period (Heimerl and Kolisch, 2010). In Ben-Daya et al. (2009), Haroun and Duffuaa emphasize that accurate forecasts for the future maintenance work are essential for determining the workforce capacity. They stress that the critical aspects of maintenance capacity planning are the expected workload size and the required skills of the workforce. They also note that: 'Making long run estimations is one of the areas in maintenance capacity planning that is both critical and not well developed in practice'.

In most literature regarding long-term workforce predictions the timescale is often larger than the timescale intended for in this study. Also, in these studies workforce is often forecasted indirectly using economic indicators. For instance, in Edwards (2010) ten year forecasts for the demand of PhD students in Australia are made using a macro-economic model. Another study uses predictions of sector growth as exogenous variable to compute skilled personnel demand (Woolard et al., 2006). In Sing et al. (2016) the authors present a case study of long-term workforce demand forecasting of building inspectors in Hong Kong. A labor multiplier approach is applied that assumes a known amount of work packages and a constant amount of labor per work package. Additionally, Briscoe and Wilson propose a multitude of co-integrating regressions to establish the existence of long-run equilibrium relationships between employment, output and wages in the context of the engineering industry of Great Britain (Briscoe and Wilson, 1991). The authors emphasize the increase of forecast inaccuracy due to the introduction of these uncertain variables.

The studies above show that for long-term forecasting exogenous information is often used to forecast workforce demand. Regularly this exogenous information is a forecast in itself and therefore adds to the uncertainty of the main model.

ii. Maintenance demand forecasts

Demand forecasting models are described extensively in literature but mostly in an inventory management context. Parallels can be drawn between spare part demand forecasting and workforce demand forecasting which is why a research into this field is worthwhile for this study.

Aircraft component failure is known for its intermittent behavior (Ghobbar and Friend, 2003). Croston was the first to address the intermittent demand problem and Croston's method is one of the most used methods for intermittent demand forecasting in industry (Xu et al., 2012),(Boylan et al., 2008). Croston found that traditional forecasting techniques lead to excessive stock of spare parts in the case of intermittent demand, which is inefficient (Croston, 1972). In Syntetos and Boylan (2001) the authors find that Croston's method is biased and introduce an adjustment to Crostons method called Syntetos Boylan Approximation (SBA), which outperforms Crostons's method in terms of accuracy.

Next to the highly successful SBA method there are other models that attempt to describe intermittent demand. In Willemain et al. (2004) the problem of forecasting intermittent demand is solved by introducing a new type of time series bootstrap while in Hua and Zhang (2006) the problem is approached through Support Vector Machines which allows for the use of explanatory variables. Other studies in the field of intermittent demand forecasting explore the use of Neural Networks, which are promising but are sensitive to overfitting and do not outperform traditional methods consistently (Gutierrez et al., 2008),(Amin-Naseri and Rostami Tabar, 2008).

The aforementioned literature shows that intermittent demand forecasting is still a complex problem and that the traditional methods are relatively inaccurate. In Nikolopoulos et al. (2011) the concept of temporal aggregation was proposed as a solution to intermittent demand forecasting. Temporal aggregation of demand aims at transforming intermittent demand into smooth demand by aggregating demand in lower-frequency bins (Murray et al., 2018). This smoothing of demand is advantageous as it allows for other forecasting methods to be used than the limited selection of methods that are available for intermittent demand. Aggregating might however result in an information-loss as a direct result of a decrease in amount of datapoints (Petropoulos and Kourentzes, 2015). An earlier study emphasizes the opportunities that a combined temporal aggregation model offers in terms of improved demand forecasts (Kourentzes et al., 2014). The main motivation of this study was to mitigate the issue of model selection, which is tedious in general. The premise of the study is that at different aggregation levels different properties or characteristics of the time series can be identified and can therefore improve the forecast accuracy.

The effect of the level of aggregation on forecast accuracy is analyzed empirically in Nikolopoulos et al. (2011). In the study the Aggregate-Disaggregate Intermittent Demand Approach (ADIDA) method is developed. This method aggregates high frequent sampled demand into lower frequently sampled demand where it is subsequently modeled. Depending on the intended goal of the research the forecast can either be used in this latter frequency or be disaggregated to the original frequency. In this way information gathered at a lower temporal frequency can be used at a higher frequency which can be beneficial in terms of forecast accuracy. Combining this approach with a naive forecasting model outperformed traditional methods such as Croston's method and SBA.

Literature shows that there are two approaches on temporal aggregation (Kourentzes et al., 2017). The first one is the usage of an optimal aggregation level and the second one is the usage of multiple aggregation level to retrieve all relevant features of the time series. The limitation of the single aggregation level approach is that a model should be assumed while the multiple aggregation level approach is robust against model choice. In conclusion, there is empirical and theoretical evidence that temporal aggregation is beneficial in terms of demand forecasting but there is no consensus on how that should be optimally achieved.

The background study shows that temporal aggregation of demand and the combination of forecasts at different levels of aggregation is beneficial in terms of forecasting accuracy. Therefore, this aggregation approach will be used in this study.

III. Methods

To answer the research question a model is developed that integrates demand forecasts with the dynamics of personnel or workforce within an aircraft maintenance organization. In this study these workforce dynamics are described using a workforce flow model which, in this paper, is also referred to as personnel flow model. The demand forecasts are obtained through demand forecast models.

In this methodology section of the paper the proposed workforce flow model is presented, followed by a description of the demand forecasting model selection procedure.

i. Workforce flow model description

As outlined in the introduction no workforce flow model as intended in this study could be identified in literature. Consequently, for this research a novel workforce flow model is proposed. The structure of the model is based on the EASA Part-66 document (EASA (2015)) but in addition several assumptions are made, which are elaborated upon in the upcoming paragraphs. The proposed model is shown schematically in Fig. 1.

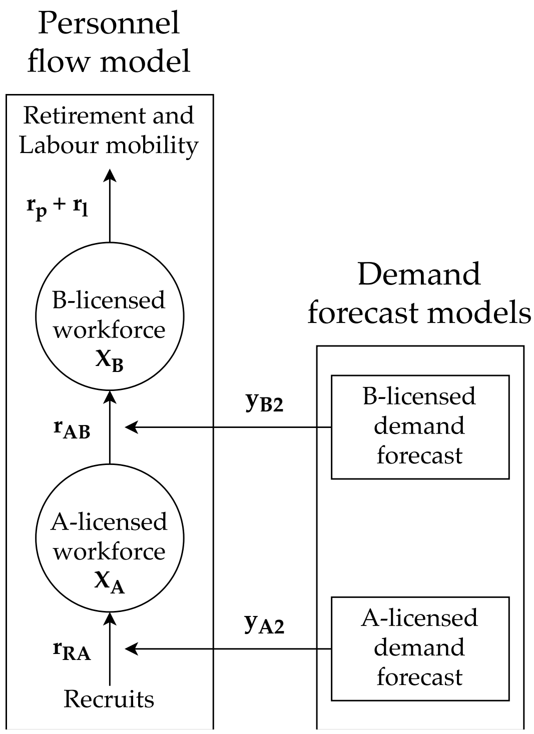

Figure 1: *Schematic representation of the proposed personnel flow model and its relation to the demand forecast models*

In this figure the flow of personnel from acquisition stage to outflow is presented. Recruits are trained into A-licensed technicians and subsequently into B-licensed technicians. The amount of personnel put to training at a certain point in time is given by the variables r_{RA} and r_{AB}. These training rates are determined by workforce demand forecasts and the expected outflow of personnel.

This model requires several assumptions to be made. Firstly, just two types of personnel are assumed, A-licensed and B-licensed. In general C-licensed personnel has a managerial function, which is why they are not included in this model. The privileges and requirements belonging to these two different licenses and the respective training duration are described in the EASA Part-66 document. In short, A-licensed technicians are allowed to perform minor, temporary, maintenance tasks, often associated with line maintenance. B-licensed technicians are certified to perform major, non-temporary, maintenance tasks. Although in reality there are different paths for a technician to achieve these two licenses it is assumed that the training duration (including the required gain of experience) is two years for both licenses. Secondly, it is assumed that only B-licensed personnel is susceptible to retirement or labour mobility as shown in Fig. 1. This is based on the fact that generally B-licensed personnel is older and more experienced. The outflow due to retirement is assumed to be constant. The outflow due to labour mobility (r_l) is dependent on time. It is modeled as being a normally distributed random variable $\mathcal{N}\left(r_{\mu l}, \sigma^2\right)$ where $r_{\mu l}$ is the expected outflow rate due to labour mobility and σ is the standard deviation. Naturally, a realization of $\mathcal{N}\left(r_{\mu l}, \sigma^2\right)$ cannot be smaller than zero as that would mean a random inflow of personnel. Therefore if $r_l(t) < 0, r_l(t) \longrightarrow 0$.

The workforce sizes and the training rates are based on the demand for A-licensed and B-licensed workforce. Due to the long training duration this workforce demand should be forecasted to be able to match the capacity with demand in the future. This leads to the following mathematical model description.

$$r_{AB}(t) = y_B(t) - y_B(t-1) + r_{pp} + r_{\mu l} \tag{1}$$

$$r_{RA}(t) = y_A(t) - y_A(t-1) + r_{AB}(t) \tag{2}$$

Where $r_{AB}(t)$ is the amount of A-licensed personnel put to training for their B-license at timestamp t, $r_{RA}(t)$ is the amount of recruits put to training for their A-license. r_{pp} and $r_{\mu l}$ are the expected amount of personnel retiring and leaving due to labor mobility at timestamp $t + t_{hor}$ respectively where t_{hor} is the forecast horizon. t_{hor} is equal to the training time in this study. $y_B(t)$ is the forecast of the B-license demand at timestamp $t + t_{hor}$ while $y_B(t-1)$ is forecast of B-license demand at timestamp $t - 1 + t_{hor}$. The variables used in Eq. 2 are equivalent to those aforementioned but then for A-license demand. In essence Eq. 1 and Eq. 2 state that the training rate at timestamp t is equal to the increase in demand at timestamp $t + t_{hor}$, taking into account the expected outflow.

Subsequently the workforce size can be determined using the computed training rates as shown in Eq. 3 and Eq. 4:

$$X_A(t) = X_A(0) + \sum_{-t_{hor}}^{t-t_{hor}} r_{RA}(t) - \sum_{-t_{hor}}^{t-t_{hor}} r_{AB}(t) \tag{3}$$

$$X_B(t) = X_B(0) + \sum_{-t_{hor}}^{t-t_{hor}} r_{AB}(t) - \sum_{0}^{t} r_p + r_l(t) \tag{4}$$

Where $X_A(t)$ is the A-licensed workforce size at timestamp t, $X_A(0)$ is the A-licensed workforce size at timestamp $t = 0$. $X_B(t)$ and $X_B(0)$ are their equivalences but then for B-license demand. Implicit to this equations are two assumptions. Firstly, the A-licensed workforce put to training for their license-B are assumed to stay part of the A-licensed workforce until their training to B-licensed technician is fully finished. Secondly, it is assumed that between $t = -t_{hor}$ and $t = 0$ the training rates $r_{AB}(t)$ and $r_{RA}(t)$ were equal to the true outflow ($r_p + r_l(t)$) between $t = 0$ and $t = t_{hor}$. Thirdly, it is assumed that firing of personnel is not permitted. Fourthly, when the B-licensed workforce is larger than the expected demand in two years ($X_B(t) >= y_B(t)$) the training rate $r_{AB}(t)$ becomes 0. Lastly, in order to keep up with the training rate to B-license, the acquisition rate $r_{RA}(t)$ is always at least equal to $r_{AB}(t)$.

This model is applied within an existing maintenance organization where there is a workforce present. This is also the reason why the training rate $r_{AB}(t)$ might exceed $X_A(t)$ as it is assumed that a sufficiently large A-licensed workforce is present within the organization. However, in this study one is interested in the change in maintenance demand which is why $X_B(0)$ and $X_A(0)$ are both set at 0.

The proposed model is a continuous model, which means that the training rates and current workforce sizes are computed for every timestamp t. A pseudo code representation of how the model is implemented in this study is given in App. A.

ii. Workforce flow model evaluation

To answer the research question, the workforce flow model is to be evaluated. From a human resource management perspective one is interested in the following aspects of the model:

1. Performance of model in terms of matching maintenance demand and workforce availability over time.

2. Model performance at different demand aggregation levels.

3. Sensitivity of model performance to varying, constant workforce outflow rates.

4. Sensitivity of model performance to non-constant workforce outflow rates.

5. Sensitivity of model performance to start of demand forecasting.

6. Sensitivity of model performance to different fleet sizes.

The aspects listed above are used to evaluate the model and are elaborated upon in the following paragraphs.

The performance of the model is assessed through evaluating the difference between the actual demand and the workforce size from a fixed timestamp onward, regardless of the start of demand forecasting. As error metric the root mean squared error is used.

The second aspect involves the influence of demand aggregation level on the performance of the model. Literature advises to aggregate maintenance demand as it improves the demand forecast accuracy (Nikolopoulos et al., 2011),(Kourentzes et al., 2017),(Schneider and Cassady, 2015). It is chosen to aggregate up to a monthly and a quarterly level as those are reasonable frequencies to evaluate the workforce size and training rates at. Also, at those aggregation levels there are still redundant data points to apply the demand forecasting model to. In Kourentzes et al. (2014) combining aggregated demand forecasts is advocated as each aggregation level is able to capture certain characteristics of the data. In Spiliotis et al. (2019) it is noted that an averaging approach of demand forecasts at different temporal aggregation levels is often as accurate as more complex approaches. In following of these studies two different combinations of forecasts will be used next to the uncombined quarterly and monthly aggregated demand forecasts.

To combine the aggregated demand forecasts the quarterly demand forecast is disaggregated to a monthly frequency by distributing the demand evenly over the three months that make up that quarter. Subsequently this disaggregated quarterly demand forecast is combined with the monthly aggregated demand forecast. A trivial combining approach would be to average the two forecasts by assigning equal weights to both forecasts. This approach is the first method, which will be referred to as 'Combined, avg.' in the remainder of this study. The second combination approach uses a more complex weighting function for the quarterly and monthly demand forecasts. The weighting function is such that at the start of the analysis a weight of one is assigned to the quarterly demand forecast, which linearly decreases to zero at the end of the analysis while the weight assigned to the monthly method increases to one. The premise of this weighting function is that the demand is still intermittent in the beginning and that a higher aggregation level results in a better forecast performance. In the remainder of the study the second method will be referred to as 'Combined, dyn.'

The third aspect regards the robustness of the model to a larger, constant outflow rate. To test this the performance of the model is assessed across different constant outflow rates.

In reality outflow due to labour mobility is not constant. Therefore the fourth aspect assesses the performance under various degrees of variability in outflow.

The fifth aspect involves the influence of the starting point of analysis on the model performance. As formulated in Eq.1 and Eq.2 the training rate is based on workforce demand forecasts. These forecasts require workforce demand data, specified per workforce type. Assuming the first aircraft acquisition to take place at $t = 0$ the forecasting of demand can only start after a certain period (t_{start}) in order to observe the necessary demand needed for an accurate forecast. In practice, a low t_{start} allows for a quick response to a change in demand but with only a few data points available an accurate forecast is more difficult to obtain. A high t_{start} on the other hand enables a more accurate forecast but the relatively late response influences the overall performance of the model. Regardless of t_{start} value the model performance is evaluated from the same timestamp onward.

The sixth aspect regards the performance of the model for different fleet sizes. Three fleet sizes will be considered in this study. Due to confidentiality exact fleet sizes cannot be provided, which is why ratios are provided.

The robustness of the model for above mentioned aspects is evaluated by comparing the root means squared error values for each parameter combination related to these aspects.

iii. Available demand data

The described personnel flow model requires accurate demand forecasts and therefore it is vital to select the best maintenance demand forecasting model in terms of forecasting performance. In the remaining four parts of the methodology section this aspect of the study is described. Before this model selection procedure can be elaborated upon a description of the available data is given.

Optimally workforce demand data in terms of man-hours would be available for this study. Nowadays however, maintenance is planned using task packages and clustering of maintenance tasks is applied to ensure efficient resource capacity planning. Namely, by clustering tasks set-up times can be shared across them (Van Dijkhuizen and Van Harten, 1997). This clustering means that the workforce demand per maintenance task is hard to determine. An inquiry of MRO engineers also showed that information on workforce demand per maintenance task is not readily available within the organization. The second best option is to use other data and to convert this information into workforce demand.

The data available for this study consists of ten years of structural repair data of a relatively large fleet of widebody passenger aircraft of the same type. Over these considered ten years aircraft are added to the fleet on an irregular basis, representing a realistic scenario of a growing fleet in its early years. In this study the amount of repairs is used as the dependent variable. For confidentiality reasons the exact amount of repairs and the fleet size cannot be provided to the reader.

The conversion to workforce demand requires assumptions to be made. Firstly, it is assumed that the growth of workforce demand is directly proportional to the increase in amount of repairs for both licenses. Secondly, it is assumed that category A and category B repairs are taken together and assumed to be executed by the B-licensed workforce. Repairs belonging to these categories are so called permanent repairs which makes this a valid assumption according to the description of B-licensed personnel in EASA (2015). Category C repairs are temporary repairs, which correspond to the privileges of A-licensed technicians. Thirdly, multiple repairs on a specific component as defined by its ATA, SUB-ATA, SUB-SUB-ATA number and its three digit zone identification of a single aircraft on a single day are considered as one failure. This reduces the impact of clustering as it makes every repair action more equivalent to each other in terms of workforce demand.

Besides the temporal aspect, demand forecasting on this scale also requires a fleet-wide approach due to the shared use of resources (Schneider and Cassady, 2015). Therefore all repairs are aggregated over all aircraft and categorized per license type. This results in a maintenance demand development over time as shown in Fig. 2 and Fig.3.

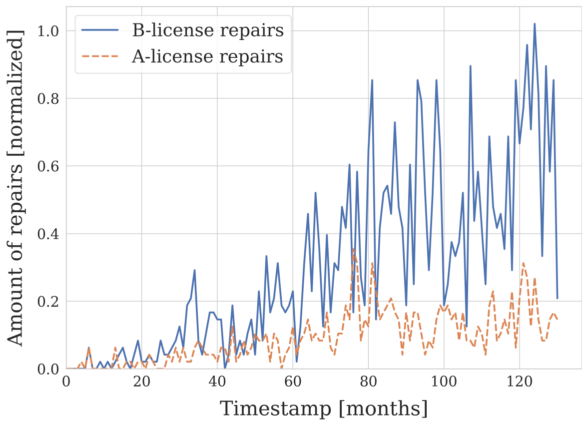

Figure 2: *Monthly aggregated amount of repairs over time for both B-licensed and A-licensed personnel.*

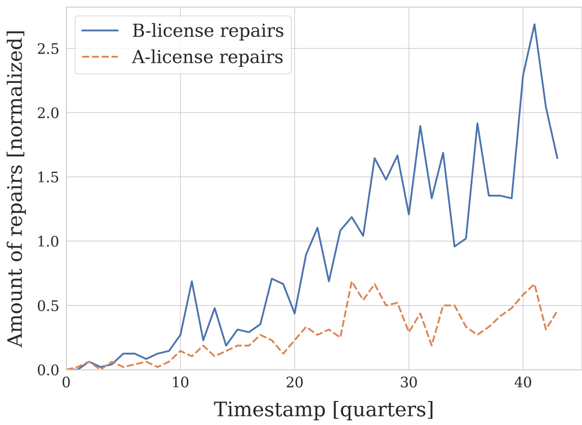

Figure 3: *Quarterly aggregated amount of repairs over time for both B-licensed and A-licensed personnel.*

iv. Demand influencing factors

The application of forecasting models that can explain repair demand on a fleet level requires variables. Variables or factors that influence maintenance demand on a fleet level are not researched extensively in literature. To the authors knowledge only one study applies factors as predictive variables to forecast maintenance demand (Pogačnik et al., 2017). In this study factors that influence fleet maintenance are used to predict fault probability on individual aircraft. Aircraft parameters such as operator, aircraft type, aircraft age, flight hours, flight cycles, engine type and operation location, are taken into consideration. Besides factors that impact maintenance directly there are also factors that influence maintenance costs which can be used as indicators for maintenance demand. According to Wu et al. (2004) factors that impact direct maintenance costs are fleet size, commonality, fleet age, fleet utilization and frequency of check intervals. On single aircraft level, age is used as key factor to describe maintenance costs and forecasting maintenance demand (Saltoglu et al., 2016),(Weckman et al., 2006).

Current airline fleet maintenance is organized around a inspection based maintenance policy (Dupuy et al., 2011). The inspection work packages prescribed by A,B,C and D checks are well known in advance. This scheduled maintenance almost always induces unscheduled maintenance as well (Wagner and Fricke, 2006). Consequently, the A-Check, C-Check and D-Check schedules are not only a good predictor for scheduled maintenance but also an indicator for unscheduled maintenance and repairs.

Exploration of the data shows that just six of the aforementioned potential variables can be used in this study. The variables are retrieved from external data sources and are listed below:

- Fleet size
- Average fleet age
- Median fleet age
- Third quartile fleet age
- Expected demand due to A-Check
- Expected demand due to C-Check

Fleet size, average fleet age and median fleet age are straightforward. Third quartile fleet age is an age of which 75 percent of the fleet is younger. The values of these variables are all well known in advance and are fixed. This is because aircraft have long delivery times. Also, due to the inspection based maintenance policy that airlines apply nowadays, A-Check and C-Check schedules can be determined for years to come (at least in indicative form).

In order to use the schedules of A-Check and C-Check as explanatory variables a continuous model should be made of them that uses these schedules as input. According to the maintenance manuals and expert interviews the A-check interval for the widebody used in this study is 105 days and the check takes 2 days on average. The C-Check interval is 750 days and is assumed to take up 10 days on average. Furthermore, it is assumed that both checks take place at 90 percent of the prescribed interval.

This information leads to a simple model for each check. Both the A-Check and C-Check variable then consist of the aggregated A-Check and C-Check models respectively. Tab. 1 visually explains the build-up of the A-Check variable which is analogous to the C-Check one.

Time (days)	0	1	2	3	...	94	95	96	97
AC1 (intr. day 0)	0	0	0	0	...	0	1	1	0
AC2 (intr. day 1)	n.a.	0	0	0	...	0	0	1	1
AC3 (intr. day 2)	n.a.	n.a.	0	0	...	0	0	0	1
Variable	0	0	0	0	...	0	1	2	2

Table 1: *Example of the A-Check variable formation*

The variables turn out to be heavily collinear, especially the age related variables. This multicollinearity among the variables makes a model selection based on t-tests of regression coefficients not feasible. Therefore the optimal forecasting model is determined by comparing the respective forecast performance. However, the type of model is identified by an insample analysis of the data presented in Fig. 2 and 3. This process is elaborated upon in the next section.

v. Method selection

In this section the type of model is identified which will be referred to as method in the upcoming paragraphs. The data described in the previous section form the basis of this method identification. The graphs given in Fig. 2 and 3 show that a linear model is a potential model and therefore the use of linear regression is the starting point of this selection procedure.

The following procedure is applied. First ordinary least squares regression (OLS) is applied using the variables

'Average fleet age', Fleet size', 'A-Check schedule' and 'C-Check schedule'. This is shown for the monthly aggregated B-licensed repairs in Fig. 4.

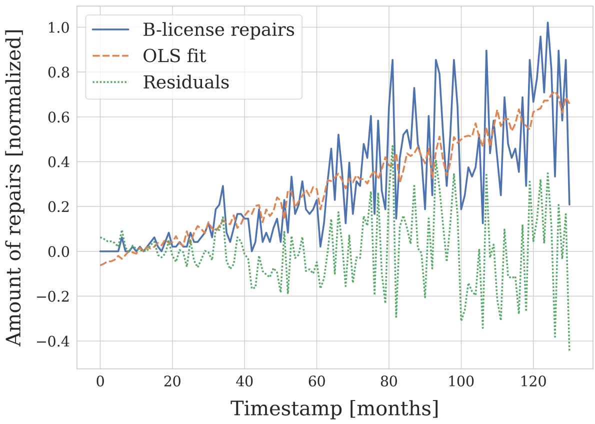

Figure 4: *OLS fit on monthly aggregated B-license repair data using variables 'Average fleet age', Fleet size', 'A-Check schedule' and 'C-Check schedule'.*

The residuals are subsequently analyzed to check for violations of the OLS assumptions and to find out if there is serial correlation:

1. Linearity check using Component-Component plus residual plot
2. Periodicity check
3. Autocorrelation check
4. Homoscedasticity check
5. Normality check

The linearity check showed that all variables are linearly related to the repair data which validates the use of a linear model. The periodicity check and the autocorrelation check showed that there is no seasonality in the data and that there is no statistically significant autocorrelation. This means that there is no serial correlation among the data points. The homoscedasticity check shows that the residuals are heteroscedastic which is a violation of an OLS assumption. The normality check shows that the residuals are normally distributed. Applying these tests on the other repair categories and quarterly sampled data lead to the same conclusion.

OLS is robust for heteroscedasticity up to a certain extent and is therefore not excluded from the model selection. Weighted Least Square regression (WLS) and its generalization Feasible Generalized Least Squares regression (FGLS) do not assume homoscedastic errors and are therefore potential methods. In matrix notation the Generalized Least Squares (GLS) estimator for the coefficient vector β is given in the following equation (Fox, 2016):

$$\mathbf{b}_{GLS} = (\mathbf{X}'\mathbf{\Sigma}_{\epsilon\epsilon}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}_{\epsilon\epsilon}^{-1}\mathbf{y} \tag{5}$$

Where $\mathbf{X}$ is the design matrix, $\mathbf{b}$ is the estimator for the coefficient vector, $\mathbf{\Sigma}_{\epsilon\epsilon}$ is the error covariance matrix and $\mathbf{y}$ is the vector of the response values. In case of an OLS model it is assumed that $\mathbf{\Sigma}_{\epsilon\epsilon} = I$. In the case of WLS and FGLS the diagonal values in $\mathbf{\Sigma}_{\epsilon\epsilon}$ are not equal to one but are weighted. For the estimator to be BLUE (Best Linear Unbiased Estimator) the weight of observation i should be the reciprocal of the variance for that observation i according to the Gauss-Markov theorem (Springer, 2008). The variance of each observation is not readily available which is why in this study the weights for the WLS model are assumed to be the reciprocal of the residual of the OLS model at the same observation.

FGLS assumes that there is an underlying explanation for the heteroscedasticity and consequently a model can be used to determine the weights. Intuitively one recognizes that the increase in variance is due to the increase in fleet size and the aging of the fleet. An OLS model with the variables 'Fleet size' and 'Median fleet age' is used to obtain the weights. FGLS is not BLUE. It is biased but consistent and it is more efficient than OLS when the number of observations go to infinity.

For above mentioned reasons OLS,WLS and FGLS are used in the demand forecast model selection procedure.

vi. Optimal demand forecast model selection

The optimal demand forecasting model is identified through a forecast performance comparison of all potential models. The procedure corresponding to this performance analysis is best explained using the example in Fig. 5.
The data between $t = 0$ and t_{eval} are used to train the model. Subsequently the model is applied to the out-of-sample data and the performance of the model is evaluated between t_{eval} and $t_{eval} + t_{hor}$ where t_{hor} is the forecast horizon. The performance metric used is the root

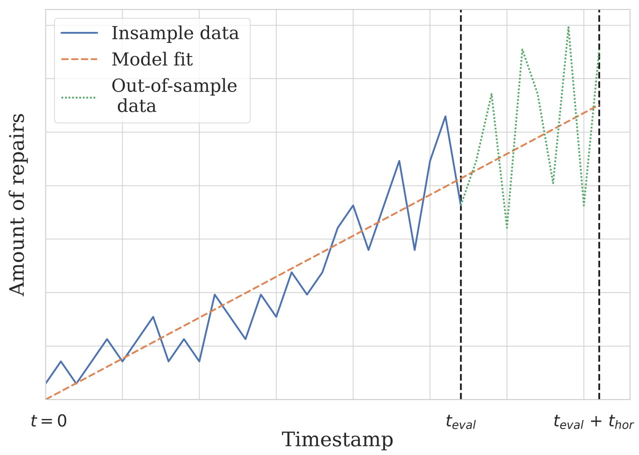

Figure 5: *Example of forecasting procedure using fictitious data. t_{eval} is the timestamp at which the forecast starts. t_{hor} is the forecast horizon.*

mean squared error which is normalized by the difference between the maximum value and minimum value of the response variable between t_{eval} and $t_{eval} + t_{hor}$. The normalized root mean squared error is in the remainder of this study referred to as NRMSE.

This procedure is applied over varying parameters and in different contexts for both license types and both aggregation frequencies. The parameters and contexts are given below:

- 3 fleet sizes:
  1. Small sized fleet
  2. Medium sized fleet
  3. Large sized fleet
- 20 random fleet compositions per fleet scenario to remove the possible bias induced by specific aircraft.
- 3 methods:
  1. OLS
  2. WLS
  3. FGLS
- 13 variable combinations:
  1. Average age
  2. Average age, Fleet size
  3. Average age, Fleet size, A-Check
  4. Average age, Fleet size, A-Check, C-Check
  5. Median age
  6. Median age, Fleet size
  7. Median age, Fleet size, A-Check
  8. Median age, Fleet size, A-Check, C-Check
  9. Third quartile fleet age
  10. Third quartile fleet age, Fleet size
  11. Third quartile fleet age, Fleet size, A-Check
  12. Third quartile fleet age, Fleet size, A-Check, C-Check
  13. Fleet size
- 17 evaluation points (t_{eval}). Starting at one year after first repair, at a half year frequency.

The forecast horizon t_{hor} is set at two years which corresponds to the training duration for both the A-license and B-license. For each of the parameter combinations a forecast as presented in Fig. 5 is executed which results in a total of 159,120 NRMSE values. These NRMSE values are subsequently used to find the optimal model by comparing the means of the NRMSE distributions of each method and variable combination. This is done for each fleet size, each sample frequency and each license type. The model with a NRMSE mean that is significantly lower compared to the other models is the optimal model. There is a possibility that multiple models are optimal if there is not a significant difference between them.

Using an ANOVA model as given in Eq. 6 in combination with a post-hoc analysis, multiple means can be compared.

$$y_{ij} = \mu + \beta_i + \varepsilon_{ij}, \quad i = 1, \ldots, q, j = 1, \ldots, n_i \tag{6}$$

Where y_{ij} denotes the jth observation in group i, μ is the overall average, β_i denotes the main effect in group i and ε_{ij} are random errors (Herberich et al., 2010). A post-hoc Tukey test takes into account the family wise error but compares all possible pairs. In the context of model assessment one is not interested in the simultaneous comparison of all models but only in the performance of models with respect to the apparent best model. This leads to Dunnett's test which tests the difference in group effects using the following partial null hypotheses and corresponding alternative hypotheses:

$$\mathrm{H}^0_{ik} : \beta_i - \beta_k \leq 0 \quad \forall i \neq k, i = 1, \ldots, q \tag{7}$$

$$\mathrm{H}^1_{ik} : \beta_i - \beta_k > 0 \quad \forall i \neq k, i = 1, \ldots, q \tag{8}$$

Where $k \neq i$ and k is the group with the lowest mean. When the null hypothesis is not rejected in favor of the alternative hypothesis it means it cannot be stated that the mean of i is significantly higher than the mean of distribution k. Hence, it cannot be concluded that the model belonging to distribution k is significantly better compared to i. Therefore, if the partial null hypotheses belonging to group i cannot be rejected the corresponding model is one of the optimal models. A family-wise confidence level of 0.95 is used in this study.

Dunnett's test assumes that the distributions are normally distributed and that the variance is the same for each distribution. It is expected that the first assumption will hold. The second assumption however is less likely to be valid due to the fact that some models will tend to overfit which means that their forecast performance will vary substantially compared to other models. Therefore a variation of Dunnett's model is applied that does not assume homoscedasticity of the errors ε_{ij} . The in Herberich et al. (2010) proposed procedure for comparing multiple means under heteroscedasticity is used in this study. In contrast to regular ANOVA it does not assume a constant variance of ε_{ij}.To obtain estimations for β a heteroscedastic consistent covariance matrix estimation technique is applied. In Herberich et al. (2010) it is suggested to use the HC3 method introduced in MacKinnon and White (1985) for this and is therefore used in this study as well.

IV. Results

The results of the workforce model and the outcome of the forecast model selection procedure are presented in this section.

i. Forecasting model selection results

The NRMSE distributions resulting from the forecasting procedure are positively skewed which is why a log transformation is applied. As an example the NRMSE distributions for the monthly sampled A-license repair data of a Small Fleet are presented in Fig. 6.

Each boxenplot in Fig. 6 consists of 340 NRMSE values. On the y-axis the abbreviated model names are given. For example, OLS: MA-FS stands for Ordinary Least Squares model with the variables 'Median Age' and 'Fleet Size'.

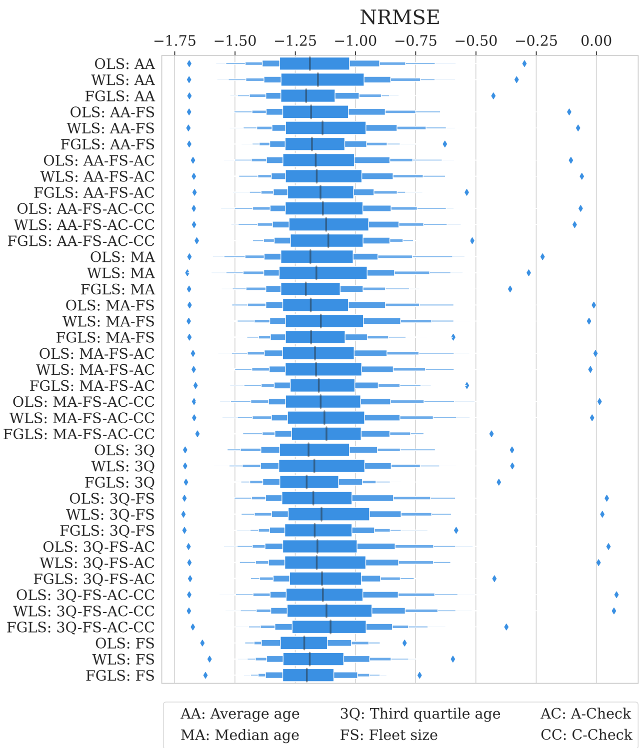

Figure 6: *NRMSE distributions per model for monthly sampled License-A repairs of a Small Fleet*

The results of the subsequent ANOVA modeling and the Dunnett post-hoc test are presented in Fig. 7 and 8. The lower the NRMSE value the better the performance of the respective model.

In these figures the optimal models for a 'Small', 'Medium' and 'Large fleet' and for both monthly and quarterly aggregated demand are highlighted in blue. The 'Small fleet' consists of three-fifths of the amount of aircraft of the 'Large fleet'. The 'Medium fleet' consists of four-fifths of the amount of aircraft of the 'Large fleet'. A general conclusion that can be drawn from Fig. 7 and 8 is that simple models consisting of only one or two variables outperform more extensive models. This can be explained by the strong multicollinearity of the variables which make the extensive models prone to overfitting and thus reduces their forecast performance.

Of the optimal models the model with lowest variance in NRMSE values is selected to be the one used in the

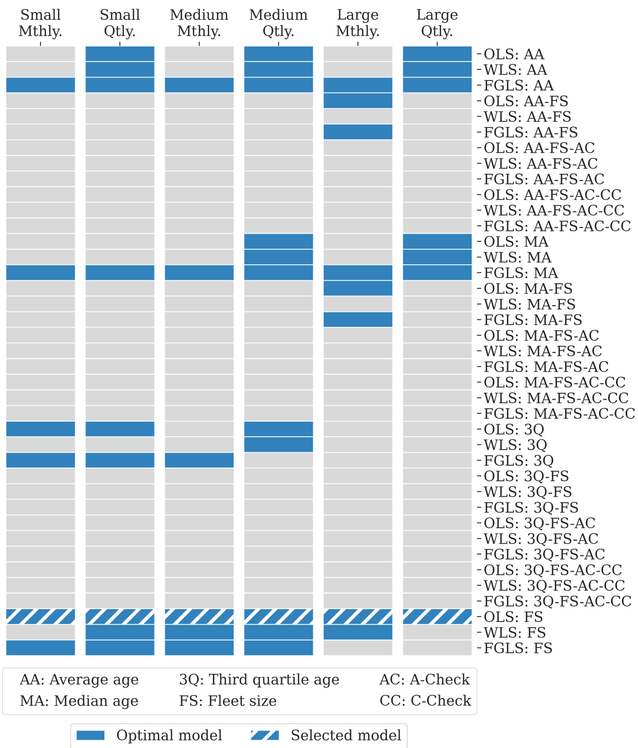

Figure 7: *Optimal forecast models and selected forecast model for A-license repair demand based on lowest mean NRMSE*

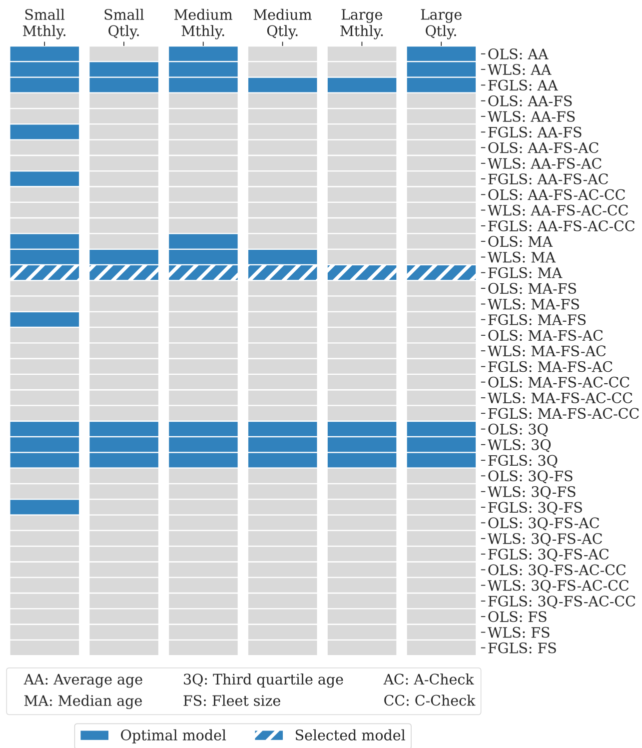

Figure 8: *Optimal forecast models and selected forecast model for B-license repair demand based on lowest mean NRMSE*

personnel flow model. For both the A-licensed and B-licensed workforce the selected model is highlighted by white diagonal stripes in Fig. 7 and 8. The A-license demand is best modeled by an Ordinary Least Squares model using 'Fleet size' as variable. A-license demand has more homoscedastic errors compared to B-license demand which explains the choice for OLS. B-license demand on the other hand is best modeled by a Feasible Generalized Least Squares model using 'Median age' as variable. For both the A-License and B-License demand the selected model is the same for each fleet size and demand aggregation level.

ii. Personnel model results, constant outflow

The selected demand forecast models are used in the personnel flow model. The repair data is obtained from an unseen fleet variant, which means that it is not used in the demand forecast model selection procedure. In Fig. 9 an example of the 2 years, monthly demand forecast is given for a Large fleet and an outflow rate of 0.02 per month.

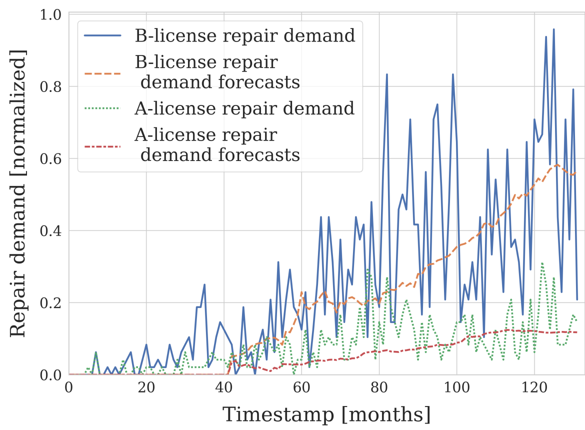

Figure 9: *Demand forecast for both A-licensed and B-licensed workforce. Start of forecast: 1.5 year, Monthly aggregation level forecast, Outflow rate: 0.02 per month, Large fleet.*

The forecasts are subsequently used as input for the personnel flow model and the resulting workforce sizes for this example are shown in Fig. 10.

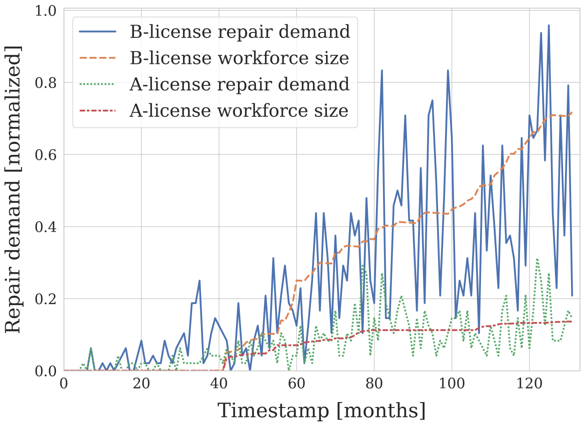

Figure 10: *A-licensed and B-licensed workforce size. Start of forecast: 1.5 year, Monthly aggregation level forecast, Outflow rate: 0.02 per month, Large fleet.*

To assess the performance of the personnel model and its sensitivity to outflow rate, fleet size and forecast start, the error between the true demand and the workforce size is computed and expressed in the root mean squared error. Independent from the forecast start the performance is evaluated between two and a half years after $t = 0$ and the last timestamp. To be able to compare the different RMSE values they are normalized. The results of the personnel model for both the A-license repairs and B-license repairs are presented in Fig. 11 and 12 respectively.

In these figures the results of the personnel model are given for the three considered fleet sizes, four different forecast starts, four demand forecast models and different but constant outflow-rates. The outflow rates are expressed in ratios of the average demand growth per month of the B-license repairs, which is found to be 0.004. Eleven different rates are considered ranging from zero to hundred percent of the average growth per month. The performance of the model is evaluated from 2.5 years onward, regardless of the forecast start. The lower the NRMSE value the better the performance of model. The results are discussed in the following paragraphs.

The impact of the outflow-rate on the model performance

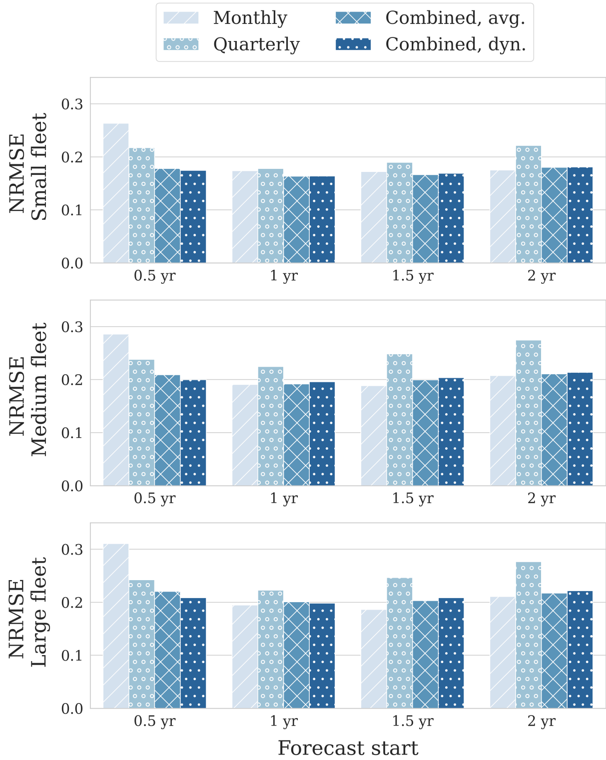

Figure 11: *Personnel model performance for A-licensed workforce under varied but constant outflow.*

is small, which is why it is expressed in the length of a hundred percent confidence interval bar and not given a distinct dimension in the plot. These confidence interval bars can be seen in the B-licensed workforce results presented in Fig. 12. The small impact on the performance is caused by the fact that the model is not able to reduce the workforce size actively due to the assumption that personnel cannot be fired. A positive outflow rate allows for the reduction of personnel, which is why the workforce size differs over time compared to the zero outflow rate case. The A-licensed personnel model performance is not impacted by the outflow-rate.

From the plots in Fig. 11 and 12 it can be concluded that the B-license workforce model performance is not sensitive for forecast start or fleet size. This is not the case for the A-license workforce model performance as it generally decreases with increasing fleet size and increasing forecast start. An optimum is found at a forecast start of one year.

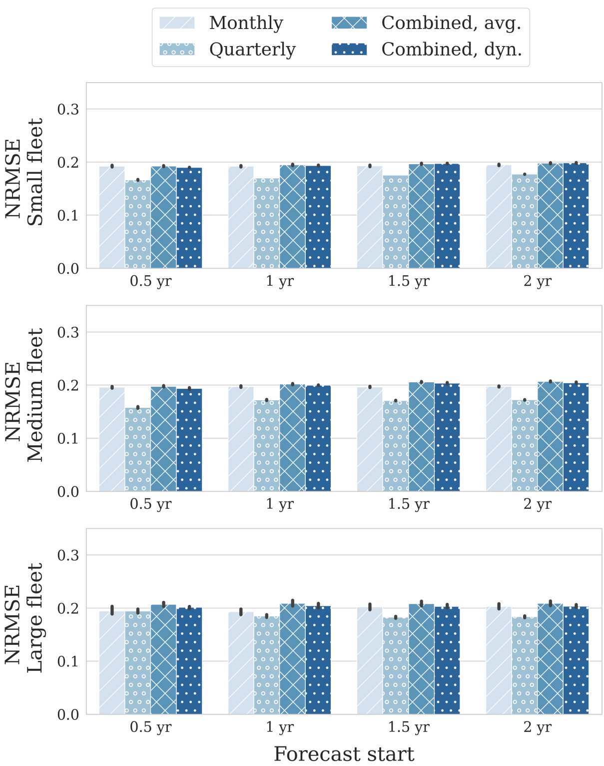

Figure 12: *Personnel model performance for B-license repairs under varied but constant outflow.*

The choice of demand forecast model impacts the personnel model performance. What can clearly be noticed is that the models that are acting on a monthly frequency (Monthly, Combined, avg. and Combined, dyn.) perform similarly, especially at higher forecast starts. The demand model based on a Quarterly sampling performs worst in the A-license workforce model but best in the B-license workforce model. An explanation for the latter is that the Quarterly model is better able to capture the trend of personnel demand due to its higher demand aggregation level.

iii. Personnel model results, non-constant outflow

The personnel outflow for the monthly sampled models due to labor mobility is modeled as a random variable with a $\mathcal{N}\left(r_{\mu l}, \sigma^2\right)$ distribution. The quarterly sampled model uses $\mathcal{N}\left(3 \cdot r_{\mu l}, (3 \cdot \sigma)^2\right)$ as random variable. The results presented in Fig. 11 and 12 are obtained by models that assume a constant outflow of personnel ($\sigma = 0$). To assess the impact of a non-constant outflow on the personnel model the NRMSE value is computed for different standard deviations. It should be noted that if the realization of the random variable is smaller than zero it is set at zero.

For this analysis $r_{\mu l}$ is fixed at 0.001 per month and the forecast start is set at one year. To still be able to compare the quarterly and monthly sampled models directly the standard deviation is normalized by the total mean outflow per timestamp of the model. In Fig. 13, 14 and 15 the results of the analysis are given for a Small fleet, Medium fleet and Large fleet respectively. Just like with the non-constant outflow only the B-licensed workforce is impacted by the outflow. Therefore only the results corresponding to that license are presented in the aforementioned figures.

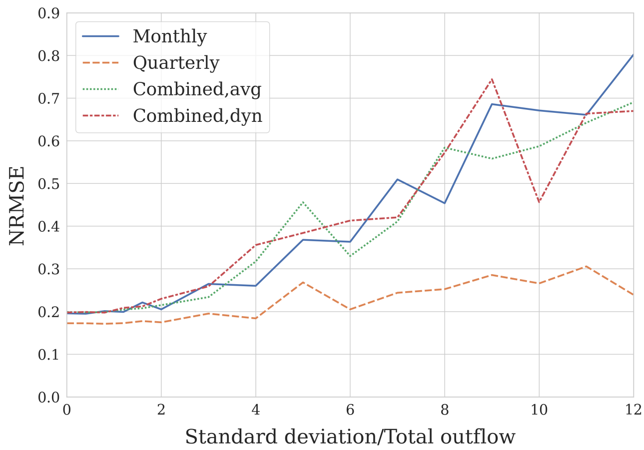

Figure 13: *Performance of personnel model for Small Fleet under non-constant outflow*

These figures show that for the monthly sampled personnel models the NRMSE increases up to values between 0.6 and 0.8. On the other hand, the performance of the quarterly sampled model stays the same. It must be stated however, that the range of standard deviations as shown in the figures is extreme. To give an example, a standard deviation/total outflow of two means for the models that extreme values of 12 times $r_{\mu l}$ fall within the 99.8% interval. Up till a standard deviation over total outflow value of two

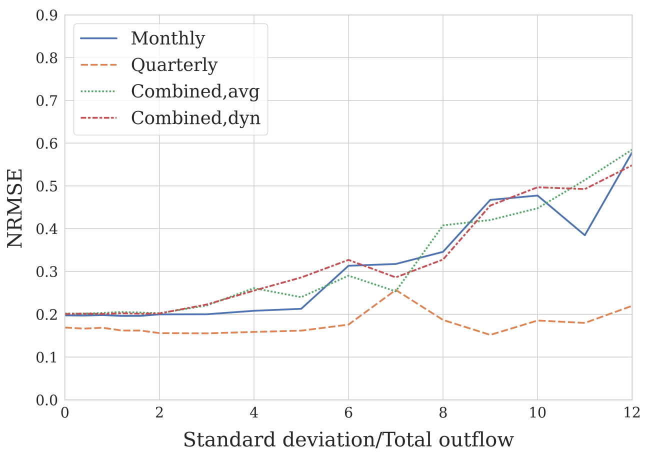

Figure 14: *Performance of personnel model for Medium Fleet under non-constant outflow*

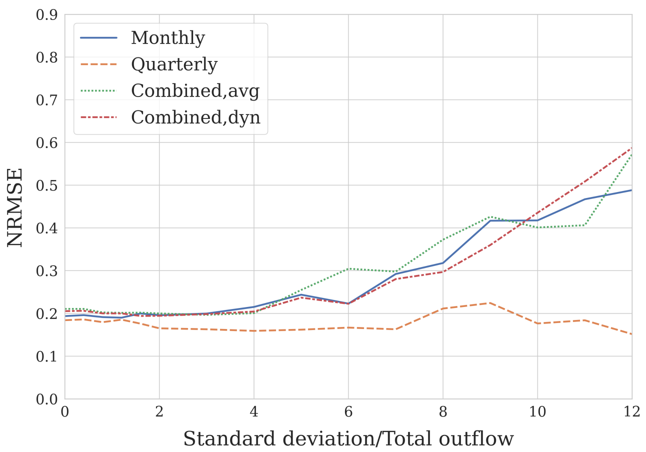

Figure 15: *Performance of personnel model for Large Fleet under non-constant outflow*

the performance of the models stays constant.

V. Discussion

The personnel flow model proposed in this study is based on assumptions and has its limitations. In the following paragraphs these will be discussed. Furthermore, potential improvements and recommendations for further research are provided. Lastly, unexpected results are discussed in this section as well.

i. Personnel flow model

The proposed personnel flow model should be evaluated as being supplementary to other workforce demand forecasting methods as it does not directly offer an organization wide or strategic solution to human resource management. Also, it does not cover the demand for inspection related maintenance tasks. As proposed in this paper the model is limited to be used during the introduction of a new aircraft fleet to an existing MRO organization or airline.

The assumption that the growth of workforce demand is directly proportional to the increase in amount of repairs is critical in this study. It is not possible to validate this assumption explicitly but, as stated in the methodology section of this paper, it is likely that information on required workforce demand per repair task is not available within MRO organizations. When true workforce demand data per repair task is available it is strongly recommended to study the effectiveness of the proposed personnel flow model using that information, or to validate the assumption of direct proportionality.

The inability of the model to fire personnel causes it to be positively biased. For this reason it is important to limit the use of this model to the case of a growing fleet which implies an increasing workforce demand over time. When assuming non-constant outflow the model becomes negatively biased due to the fact that the outflow cannot become smaller than zero. This bias causes the models poor performance at high standard deviation values. This bias impacts the quarterly sampled model less than the monthly sampled ones. This is explained by the fact that due to its lower frequency the quarterly sampled model experiences this extreme value less often over the considered time frame, which results in a smaller decrease in workforce size and performance.

The computation of $r_{AB}(t)$ and $r_{RA}(t)$ does not use the workforce sizes $X_A(t)$ and $X_B(t)$ but is solely based on the difference between the forecasted demand at $t-1+t_{hor}r$ and $t+t_{hor}$. This means that the model cannot cope with large, unexpected decreases in workforce size. This issue is only problematic when the variability of the amount of personnel leaving due to labor mobility is high. Namely, only with high variability many negative values of $r_l(t)$ are forced to zero which skew the distribution of $r_l(t)$.

ii. Demand forecasting

Critical remarks can be made regarding the used repair data itself. The observed amount of repairs is prone to large variation, especially on a monthly aggregation level, as can be seen in Fig. 2. Consequently the residuals of the demand forecast models are large as well. Potentially these residuals can be explained by additional explanatory variables, which can be constructed from information on the usage of the considered aircraft. However, extensive analysis of the data shows that part of the variability in amount of repairs is caused by random events such as bird strikes, which are unpredictable by nature. Another explanation for the high variability is the fact that this data is acquired from a large MRO which provides maintenance service to multiple aircraft types and fleets of various airlines. Hence, the planning of repairs of the fleet used in this study might be impacted by the organization wide maintenance planning. However, it is important to realize that from a human resource management perspective the trend and the average workforce demand over time is of primary interest. Eventually, allocation of workforce to specific maintenance tasks at specific moments in time is part of capacity planning and maintenance scheduling.

iii. Recommendations for further research

Apart from investigating the effectiveness of the proposed personnel flow model using true workforce demand data, other recommendations for further research can be made. The first would be the application of this model on a fleet of a different aircraft type to see if the same results are achieved. Also, it would be interesting to investigate the usage of a variation of this model on an organizational or multi-fleet level. An extension of the model that incorporates the current size of the workforce to determine the training rate could reduce the models sensitivity to sudden decreases in workforce size. Furthermore, the model assumes that B-licensed personnel cannot perform A-license repairs which is not realistic. An extension of the model is recommended to let the model balance the two workforces in an optimal way and cope with the repair demand more effectively. Lastly, in practice MROs also make use of a flexible workforce and independent technicians that are hired temporarily. The proposed model can be applied as input to studies that aim to optimize a combination of the different types of workforce.

VI. Conclusion

This study aims to answer the question of how dynamic, tactical maintenance demand forecasts can aid human resource management in the aircraft maintenance industry regarding personnel acquisition and training. To answer this question a personnel flow model is proposed of which its performance and robustness for varying degrees of outflow of personnel, different fleet sizes, different forecast aggregation levels and starting points of analysis are assessed. The novelty of this study lies in the direct integration of repair demand with personnel acquisition and training. Furthermore, this study contains a systematic assessment of repair demand forecast models, which, on this tactical level, is also novel.

Analysis of the personnel flow model shows that its performance is not affected by an increase in constant outflow. Also, the model is robust for non-constant outflow of personnel if the standard deviation does not become larger than four times the average total monthly outflow. In general the model performs best at an early starting point, with an optimum at one year after the acquisition of the first aircraft. A-Licensed personnel acquisition and training rates are best evaluated at a monthly frequency while for B-licensed personnel a quarterly frequency is preferred.

The demand forecasting model selection procedure showed that the repair demand is best forecasted using linear regression models that are robust for heteroscedasticity of the errors. Furthermore, simple models that use one or two variables are found to outperform more extensive models. For A-license repairs an Ordinary Least Squares model using 'Fleet size' as variable is found to be optimal while for B-license repairs a Feasible Generalized Least Squares regression model using 'Median age' as variable is best.

This study shows that the proposed personnel flow model is useful to human resource management due its favorable characteristics and the fact that it can be used in various, realistic contexts. The results presented in this study encourage further research into quantitative methods that can help the decision making regarding training of personnel within the aircraft maintenance industry.

References

Airbus (2016). Experience and lessons learned of a Composite Aircraft. Technical report.

Amin-Naseri, M. R. and Rostami Tabar, B. (2008). Neural network approach to lumpy demand forecasting for spare parts in process industries. *Proceedings of the International Conference on Computer and Communication Engineering 2008, ICCCE08: Global Links for Human Development*, pages 1378–1382.

Ben-Daya, M., Duffuaa, S. O., Raouf, A., Knezevic, J., and Ait-Kadi, D. (2009). *Handbook of Maintenance Management and Engineering*. Springer International Publishing.

Bill, J. (2018). Another Look at the Aviation Maintenance Personnel Shortage and the Solutions.

Boeing (2006). Boeing 787 from the Ground Up. Technical report, Boeing.

Boylan, J. E., Syntetos, A. A., and Karakostas, G. C. (2008). Classification for forecasting and stock control: A case study. *Journal of the Operational Research Society*, 59(4):473–481.

Briscoe, G. and Wilson, R. (1991). Explanations of the demand for labour in the United Kingdom engineering sector. *Applied Economics*, 23(5):913–926.

Cohan, P. (2015). American Airlines' 787 Dreamliner Nightmare. *Forbes*.

Constanza, D., Prentice, B., and Smiley, J. (2017). Aviation Growth Is Outpacing Labor Capacity. Technical report.

Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly*, 23(3):289–303.

Drew, C. and Mouawad, J. (2013). New Challenges for the Fixers of Boeing's 787.

Dupuy, M. J., Wesely, D. E., and Jenkins, C. S. (2011). Airline fleet maintenance: Trade-off analysis of alternate aircraft maintenance approaches. *2011 IEEE Systems and Information Engineering Design Symposium, SIEDS 2011 - Conference Proceedings*, pages 29–34.

EASA (2015). AMC and GM to Part-66. Technical Report 2, European Aviation Safety Agency.

Edwards, D. (2010). The future of the research workforce: Estimating demand for PhDs in Australia. *Journal of Higher Education Policy and Management*, 32(2):199–210.

Fox, J. (2016). *Applied regression analysis & generalized linear models*. SAGE publications.

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 and Operations Research*, 30(14):2097–2114.

Gutierrez, R. S., Solis, A. O., and Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111(2):409–420.

Heimerl, C. and Kolisch, R. (2010). Scheduling and staffing multiple projects with a multi-skilled workforce. *OR Spectrum*, 32(2):343–368.

Herberich, E., Sikorski, J., and Hothorn, T. (2010). A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS ONE*, 5(3):1–8.

Hua, Z. and Zhang, B. (2006). A hybrid support vector machines and logistic regression approach for forecasting intermittent demand of spare parts. *Applied Mathematics and Computation*, 181(2):1035–1048.

Kourentzes, N., Petropoulos, F., and Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2):291–302.

Kourentzes, N., Rostami-Tabar, B., and Barrow, D. K. (2017). Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels? *Journal of Business Research*, 78(April):1–9.

MacKinnon, J. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29:305–325.

Maintenance Cost Task Force (2018). Airline Maintenance Cost Executive Commentary Public Version. (November):14.

Murray, P. W., Agard, B., and Barajas, M. A. (2018). Forecast of individual customer's demand from a large and noisy dataset. *Computers and Industrial Engineering*, 118(July 2017):33–43.

Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., and Assimakopoulos, V. (2011). An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: An empirical proposition and analysis. *Journal of the Operational Research Society*, 62(3):544–554.

Petropoulos, F. and Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66(6):914–924.

Phillips, P., Diston, D., Starr, A., Payne, J., and Pandya, S. (2009). A review on the optimisation of aircraft maintenance with application to landing gears. *Engineering Asset Lifecycle Management - Proceedings of the 4th World Congress on Engineering Asset Management, WCEAM 2009*, (September):68–76.

Pogačnik, B., Duhovnik, J., and Tavčar, J. (2017). Aircraft fault forecasting at maintenance service on the basis of historic data and aircraft parameters. *Eksploatacja i Niezawodnosc*, 19(4):624–633.

Prentice, B., Costanza, D., and Smiley, J. (2017). When Growth Outpaces Capacity: A Labor Shortage and Out-of-Date Technology May Raise MRO Costs for an Expanding Global Fleet. Technical report.

Saltoglu, R., Humaira, N., and Inalhan, G. (2016). Maintenance stop time influence on aircraft total maintenane cost with downtime integrated cost model. *Proceedings of 2016 7th International Conference on Mechanical and Aerospace Engineering, ICMAE 2016*, pages 502–506.

Schneider, K. and Cassady, C. R. (2015). Evaluation and comparison of alternative fleet-level selective maintenance models. *Reliability Engineering and System Safety*, 134:178–187.

Sing, C. P., Chan, H. C., Love, P. E., and Leung, A. Y. (2016). Building Maintenance and Repair: Determining the Workforce Demand and Supply for a Mandatory Building-Inspection Scheme. *Journal of Performance of Constructed Facilities*, 30(2):1–8.

Spiliotis, E., Petropoulos, F., and Assimakopoulos, V. (2019). Improving the forecasting performance of temporal hierarchies. *PLoS ONE*, 14(10):1–21.

Springer (2008). *The Concise Encyclopedia of Statistics*. Springer, New York.

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

Van Dijkhuizen, G. and Van Harten, A. (1997). Optimal clustering of frequency-constrained maintenance jobs with shared set-ups. *European Journal of Operational Research*, 99(3):552–564.

Wagner, M. and Fricke, M. (2006). Estimation of daily unscheduled line maintenance events in civil aviation. *ICAS-Secretariat - 25th Congress of the International Council of the Aeronautical Sciences 2006*, 6:3905–3912.

Wahyudin, R. S., Sutopo, W., Hisjam, M., and Hardiono, R. S. (2016). Resource allocation model to find optimal allocation of workforce, material, and tools in an aircraft line maintenance. In *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, volume 2, pages 782–787.

Weckman, G. R., Marvel, J. H., and Shell, R. L. (2006). Decision support approach to fleet maintenance requirements in the aviation industry. *Journal of Aircraft*, 43(5):1352–1360.

Willemain, T. R., Smart, C. N., and Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, 20:375–387.

Woolard, I., Kneebone, P., and Lee, D. (2006). Forecasting the Demand for Scarce Skills , 2001-2006. *Review Literature And Arts Of The Americas*, pages 2001–2006.

Wu, H., Liu, Y., Ding, Y., and Liu, J. (2004). Methods to reduce direct maintenance costs for commercial aircraft. *Aircraft Engineering and Aerospace Technology*, 76(1):15–18.

Xu, Q., Wang, N., and Shi, H. (2012). A Review of Croston's method for intermittent demand forecasting. In *Proceedings - 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2012*, number July, pages 1456–1460.

Zuehlke, B. (2014). Now Hiring! *Aviation Pros*.

A. Personnel flow model pseudo code

Algorithm 1: ForecastFunc()

Input:
t = timestamp
$Data_B$ = maintenance demand data for license B uptil time t
$Data_A$ = maintenance demand data for license A uptil time t
$y_A(t)$ = two years ahead A-license demand forecast at time t
$y_B(t)$ = two years ahead B-license demand forecast at time t
$model_A$ = regression model for A-license demand
$model_B$ = regression model for B-license demand
r_{pp} = predicted rate of B-licensed workforce flowing out due to retirement (constant)
$r_{\mu l}$ = predicted mean rate of B-licensed workforce flowing out due to labour mobility (constant)
$X_B(t)$ = B-licensed workforce size at time t
$X_A(t)$ = A-licensed workforce size at time t
Output:
$r_{AB}(t)$ = rate of workforce put to training for license B at time t
$r_{RA}(t)$ = rate of workforce put to training for license A at time t
$y_B(t)$ = two years ahead B-license demand forecast at time t
$y_A(t)$ = two years ahead A-license demand forecast at time t

$y_B(t) \longleftarrow regression(Data_B, model_B)$
$y_A(t) \longleftarrow regression(Data_A, model_A)$
$r_{AB}(t) \longleftarrow y_B(t) - y_B(t-1) + r_{pp} + r_{\mu l}$
if $r_{AB}(t)$<0 **then**
  $r_{AB}(t) = 0$
if $X_B(t-1)$ >= $y_B(t)$ **then**
  $r_{AB}(t) = 0$
$r_{RA}(t) \longleftarrow y_A(t) - y_A(t-1) + r_{AB}(t)$
if $r_{RA}(t)$<$r_{AB}(t)$ **then**
  $r_{RA}(t) = r_{AB}(t)$
if $X_A(t-1)$ >= $y_A(t)$ **then**
  $r_{RA}(t) = r_{AB}(t)$

Algorithm 2: Personnel model

Input:
t_{start} = time of start forecast
t_{hor} = forecast horizon
$X_B(0)$ = B-licensed workforce size at time t_0
$X_A(0)$ = A-licensed workforce size at time t_0
r_p = rate of B-licensed workforce flowing out due to retirement (constant)
$r_l(t)$ = rate of B-licensed workforce flowing out due to labour mobility at time t
Output:
$r_{AB}(t)$ = rate of workforce put to training for license B at time t
$r_{RA}(t)$ = rate of workforce put to training for license A at time t
$X_B(t)$ = B-licensed workforce size at time t
$X_A(t)$ = A-licensed workforce size at time t

while t<=t_{final} **do**
  else if t<t_{start} **then**
    $r_{AB}(t) \longleftarrow r_p + r_l$
    $r_{RA}(t) \longleftarrow r_p + r_l$
    $X_A(t) \longleftarrow X_A(0)$
    $X_B(t) \longleftarrow X_B(0)$
  else if t>=t_{start} **and** t < ($t_{start} + t_{hor}$) **then**
    $r_{AB}(t), r_{RA}(t) \longleftarrow ForecastFunc()$
    $X_A(t) \longleftarrow X_A(0)$
    $X_B(t) \longleftarrow X_B(0)$
  else if t>=($t_{start} + t_{hor}$) **and** t <= ($t_{final} - t_{hor}$) **then**
    $r_{AB}(t), r_{RA}(t) \longleftarrow ForecastFunc()$
    $X_A(t) \longleftarrow X_A(0) + \sum_{0}^{t-t_{hor}} r_{RA}(t) - \sum_{0}^{t-t_{hor}} r_{AB}(t)$
    $X_B(t) \longleftarrow$
      $X_B(0) + \sum_{0}^{t-t_{hor}} r_{AB}(t) - \sum_{t_{hor}-1}^{t-1} r_p + r_l(t)$
    $r_l(t) \longleftarrow \mathcal{N}\left(r_{\mu l}, \sigma^2\right)$
    if $r_l(t) < 0$ **then**
      $r_l(t) \longleftarrow 0$
  else
    $X_A(t) \longleftarrow X_A(0) + \sum_{0}^{t-t_{hor}} r_{RA}(t) - \sum_{0}^{t-t_{hor}} r_{AB}(t)$
    $X_B(t) \longleftarrow$
      $X_B(0) + \sum_{0}^{t-t_{hor}} r_{AB}(t) - \sum_{t_{hor}-1}^{t-1} r_p + r_l(t)$
    $r_l(t) \longleftarrow \mathcal{N}\left(r_{\mu l}, \sigma^2\right)$
    if $r_l(t) < 0$ **then**
      $r_l(t) \longleftarrow 0$

B. Software use

For this research various software is used. Primarily Python 3.6 is used to perform the data analysis. Its library Statsmodels is used for the OLS, WLS and FGLS models. The figures are made using Matplotlib and Seaborn. The Dunnett's test is performed via R. To be specific, its library Multcomp is used as described in Herberich et al. (2010).

Chapter 2

Literature study

In this chapter the literature study is presented. This study focuses on literature regarding maintenance demand forecasting and also provides a description of the methodology for the optimal predictive model selection. This thesis deliverable is already graded.

GRADED

Strategic maintenance demand forecasting

Improving skill based workforce planning

A literature study for the MSc. Aerospace Engineering thesis

Bram Slangen, 4278070
Air Transport Operations
Delft Technical University
The Netherlands
05-02-2020

Executive summary

The aim of this literature study is to provide the necessary theoretical background for the MSc. Aerospace Engineering thesis. This implies building a strong foundation for the addressed need, assessing the state of the art literature regarding the subject, finding the gap in scientific knowledge and exploring the available methods of analysis.

Due to the increasing competition in aircraft maintenance industry, cost reduction is of high priority for Maintenance Repair and Overhaul organizations (Maintenance Cost Task Force, 2018),(Phillips et al., 2009). This cost reduction can be achieved by optimizing the strategic capacity planning. This is especially true for the planning of technical specialists as their costs are high and their education extensive (Wahyudin et al., 2016).

Study of literature shows clearly that there is a need for accurate demand forecasting to aid this strategic workforce capacity planning. However, what fleet characteristics should be used to efficiently forecast this maintenance demand is not known. In an attempt to address this lack of knowledge the research aims to answer the following research question:

> How can fleet properties efficiently aid maintenance demand forecasting in order to improve strategic skill-based workforce planning?

To help answering of this research question three sub-questions are developed:

1. How can strategic aircraft maintenance workforce planning be improved?
2. What fleet properties influence strategic maintenance demand forecasting?
3. How can the impact of the fleet properties on maintenance demand forecasting be assessed efficiently?

This report aims to either fully or partly answer these research questions by assessing the relevant literature and presenting the proposed methodology which will be used for the quantitative analysis.

Capacity planning for aircraft maintenance

The first step of workforce capacity planning is the determination of the workforce required per skill and period (Heimerl and Kolisch, 2010). The second step is the scheduling of personnel or the actual assignment of tasks to people. In contrast to the workforce capacity planning this latter step is researched extensively in literature (Firat and Hurkens, 2012), (Li and Womer, 2009),(Dinis et al., 2019a),(Wahyudin et al., 2016). Although these papers focus on the scheduling part of capacity planning all authors emphasize the need for workforce demand forecasting, especially regarding specialized personnel.

Literature on workforce demand forecasting in aircraft maintenance industry is scarce and, to the authors knowledge, even non-existent in the case of strategic demand forecasting . Strategic capacity demand forecasting is applied in other industries but here again the amount of literature is limited. The available literature suggests the usage of external information or exogenous variables to augment the demand forecasts

(Lee et al., 1998),(Spetz, 2017),(Briscoe and Wilson, 1991), (Edwards, 2010),(Woolard et al., 2006),(Sing et al., 2016).

Before strategic demand forecasting can be assessed the temporal requirements have to be defined. According to literature and experts a forecast horizon of 1-3 years should be adhered to with a temporal resolution of a month to a quarter (Duffuaa and Alfares, 2009). Besides the temporal aspect strategic maintenance demand forecasting also requires a fleet-wide approach due to the shared use of resources (Schneider and Cassady, 2015). This suggests the aggregation of all relevant components up to fleet level.

Industry and literature show a clear need for a fleet-wide and long-term approach to skill-based workforce demand forecasting by making use of exogenous variables. This will be the base for the remainder of this study.

Fleet maintenance demand

Literature on fleet properties that might influence strategic maintenance demand is scarce. The predictive maintenance policy adhered to by MRO organizations suggests including maintenance check schedules in the analysis. A study into fault probability prediction of individual aircraft uses the parameters age, type, operator, flight hours, flight cycles, engine type and operation location to determine this probability (Pogačnik et al., 2017). A study on the cost of maintenance on fleet level argues that factors fleet size, fleet commonality, fleet age, fleet utilization and frequency of check intervals influence the direct maintenance costs (Wu et al., 2004). Besides these factors it is also expected that fleet composition and fleet acquisition can influence the demand forecasting performance. The selection of factors that will be assessed to answer the research question is dependent on the availability of data and is therefore discussed in the methodology section of this summary.

Demand forecasting

The need for demand forecasting is found and the potential factors that influence this demand or the forecast performance are identified. To analyze the effects of these factors quantitatively models that forecast the maintenance demand have to be found.

Maintenance demand forecasting is a challenging issue due to intermittent failure characteristics of aircraft components (Ghobbar and Friend, 2003). Modeling of intermittent demand is generally done through applying variations of the Single Exponential Smoothing method. Well used variations are Croston's method and the Syntetos and Boylan Approximation (SBA) but their forecasting performance is poor (Syntetos and Boylan, 2005),(Croston, 1972). Novel methods consisting of neural networks and support vector machines are promising but they are also relatively inaccurate (Gutierrez et al., 2008),(Amin-Naseri and Rostami Tabar, 2008),(Carmo and Rodrigues, 2004).

An approach that aims to solve the issue of inaccurate intermittent demand forecasting applies temporal aggregation. As defined by Nikolopoulos et al.: 'Temporal aggregation refers to aggregation in which a low frequency time series is derived from a high frequency time series' (Nikolopoulos et al., 2011). Temporal aggregation changes the demand category from intermittent to smooth which can generally be forecasted more accurately. Studies show that temporal aggregated methods outperform Croston's method and SBA in terms of intermittent demand forecasting (Nikolopoulos et al., 2011),(Babai et al., 2012),,(Kourentzes et al., 2014),(Petropoulos and Kourentzes, 2015). Among experts there is still debate about whether or not a single, optimal temporal aggregation level or a combined, multiple aggregation level model is better (Kourentzes et al., 2014).

As argued in the previous section strategic maintenance demand forecasting forces the aggregation of different components up to fleet level. Analogous to temporal aggregation this cross-sectional aggregation reduces intermittent behavior which improves demand forecasting. Although it seems logical from a practical strategic demand forecasting point of view to assume this level of aggregation, it might be that clustering of components with similar characteristics offers an improved forecasting performance as studies from other fields suggest (Kalchschmidt et al., 2006),(Misiti et al., 2010),(Zotteri and Verganti, 2001). In these studies demand generators are clustered based on their similarity of properties. Next to individual properties, time series patterns can be used to cluster components however this requires distinguishable patterns which are often not found when intermittency is present (Martínez-Álvarez et al., 2011),(Jin et al., 2014),(Bokde et al., 2018),, (Dastidar, 2017), (Viswanathan et al., 2008).

Methodology

The literature presented in the past sections provides the theoretical foundation which now can be used to formulate the methodology. Besides this theoretical background the methodology is also driven by the availability of data. Therefore a preliminary data analysis is performed.

Data

The data available is Boeing 777 maintenance log data. This data contains information on failure times of parts per aircraft. No information on the required man hours for the related maintenance task is given. Consequently, a conversion from the amount of failures to the actual workforce demand cannot be made. Therefore, the demand that will be forecasted is the amount of failures of a selection of components for the entire fleet. The selection of components is based on the need for skill based planning. Composite components generally require specialized personnel which is why for the component selection all components consisting fully or partly of composite material are chosen.

Variables

Based on the availability of data and their feasibility the following four predictor variables are used for further analysis:

- Fleet size
- Average fleet age
- Expected demand due to A-Check
- Expected demand due to C-Check

Next to these variables there are factors that might influence demand characteristics or the forecasting performance. These are aircraft type, airline, aircraft usage and acquisition policy. The influence of these factors is controlled for through the methodology.

Data pre-processing

No individual properties are known and no patterns in the demand behavior of the individual components could be found. This means that clustering is not a valid option and the aggregation will be on a fleet level. From a temporal perspective the aggregation will be on a monthly and quarterly basis. As combined forecasts of multiple temporal aggregation levels might be more accurate this option will be analyzed as well.

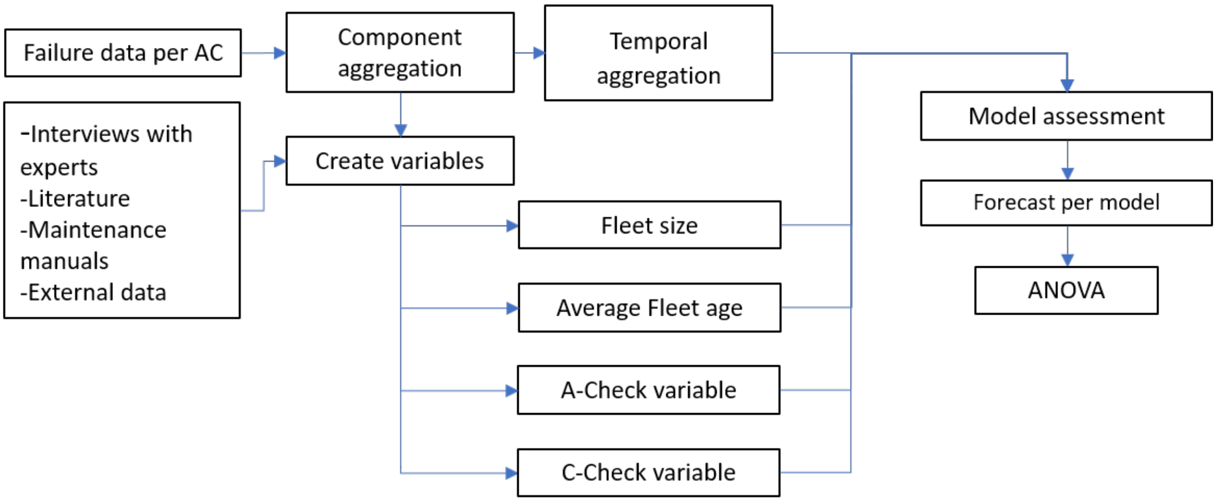

Figure 1: Methodology overview

Model assessment and forecasting

To evaluate the influence of fleet size, fleet age, A-Check schedule and C-Check on strategic maintenance demand forecasting a regression model approach is chosen. The methodology includes a model selection phase where properties of the data such as serial dependence, seasonality and heteroskedasticity determine which variation of regression is most suitable. It is expected that multiple models result from that selection phase and all of these are then used in the forecasting phase where the performance of each model is assessed. Each model is applied in different fleet settings, using multiple variable combinations, using multiple groups of aircraft (to ensure cross-validation) and by using multiple split-up points and forecast horizons.

The RMSE values resulting from the forecasting phase are then analyzed using an ANOVA analysis. This analysis can assess the significance of the difference between means of RMSE groups. A summarized flowchart representing the methodology is shown in Fig. 1. More detailed flowcharts are given in Fig. 5.1, 5.2 and 5.3.

Contents

Chapter 1

Introduction

The aim of this literature study is to provide the necessary theoretical background for the MSc. Aerospace Engineering thesis research. This implies building a strong foundation for the addressed need, assessing the state of the art literature regarding the subject, finding the gap in scientific knowledge and exploring the available methods of analysis. The goal of these literature study objectives is to fully or partly answer the research questions at hand.

During the last decades the aviation market has become a competitive one which forces airlines to reduce costs wherever they can. Next to the airlines evident interest in maintenance cost reduction, Maintenance Repair and Overhaul organizations (MROs) on their turn are interested in more efficient maintenance operations as the competition in the MRO market has increased tremendously as well (Phillips et al., 2009).

Part of the maintenance costs are due to the use of resources. In the aviation industry the main resources are spare parts, hangar space/equipment and personnel. To make optimal use of these resources a strategic perspective on capacity planning is required. Strategic capacity planning is a form of capacity planning that determines the appropriate level of maintenance resources on a strategic level (Duffuaa and Alfares, 2009). The resource that is of particular interest from a strategic perspective is personnel. Nowadays skilled personnel is namely scarce in aerospace industry due to technicians leaving airlines and MRO organizations for various reasons. Next to this, the training of technicians is time consuming and expensive due to strict aviation regulations (Johnson, 2018). This means that strategic information on the expected required workforce is vital for an optimized capacity planning. Accurate strategic maintenance demand forecasts are therefore critical but are challenging to obtain due to the stochastic nature of maintenance.

Based on the aforementioned needs and on a preliminary literature study the following research questions are formulated. During the formulation process of the research questions and the literature study itself the data set, that is available for the research, is also used to provide direction. The main research question is as follows:

> How can fleet properties efficiently aid maintenance demand forecasting in order to improve strategic skill-based workforce planning?

To assist in answering the main research question, multiple sub-questions are formulated:

1. How can strategic aircraft maintenance workforce planning be improved?
   (a) What is the current need in the MRO industry regarding workforce demand forecasting?
   (b) What is the temporal aggregation level at which maintenance demand should be evaluated from a strategic perspective?
   (c) What is the time-horizon which should be applied while forecasting?

(d) What is the level of component aggregation at which maintenance demand should be evaluated from a strategic workforce demand forecasting perspective?

(e) What is the current state of literature of workforce demand forecasting in general and which methods are often used?

2. What fleet properties influence strategic maintenance demand forecasting?

   (a) What are the influential factors present in current fleet maintenance practice?

   (b) What are fleet properties that influence fleet maintenance demand and demand forecasting?

   (c) What aspects of fleet planning are of influence on fleet maintenance demand forecasting?

3. How can the impact of the fleet properties on maintenance demand forecasting be assessed efficiently?

   (a) What is the optimal temporal aggregation level from a model efficiency of view?

   (b) What is the optimal component aggregation level from a model efficiency point of view?

   (c) Given the fleet porperties what are potentially efficient models to forecast maintenance demand?

   (d) How can the performance of the predictive models be assessed?

As stated before, these questions are used as guidelines for the upcoming literature study. Each chapter in this literature study is related to a sub question. After the literature study most of these questions can be answered either fully or partly.

In Ch. 2 the definition of capacity planning in an aviation context is provided. Thereafter the state of the art literature regarding strategic capacity planning is presented after which an assessment of workforce management in general is given. This chapter is followed by Ch. 3 where the fleet properties that influence demand or demand forecasting are identified. The bulk of this literature study is given in Ch. 4 where first demand classification is analyzed after which the literature on intermittent demand modeling and prediction is studied. Consequently the information obtained from the literature is translated into a methodology which is elaborated upon in Ch. 5. In the conclusion in Ch. 6 the results of the literature study are summarized.

Chapter 2

Capacity planning of aircraft maintenance

As argued in the introduction, improved capacity planning can drastically reduce costs. To find out how this can be achieved an extensive literature study is presented in this chapter.

In this study the definition of maintenance planning by Duffuaa and Alfares will be used: 'The determining of the appropriate level and workload assignment of different maintenance resources in each planning period' (Duffuaa and Alfares, 2009). Capacity planning can be subdivided into maintenance planning and maintenance scheduling. There is no clear agreement on these two definitions but to stay consistent in this research the definitions according to Duffuaa and Al-Sultan are used (Duffuaa and Al-Sultan, 1997). They state that maintenance planning consists of:

1. Identification of work to be planned.
2. Determination of work complexity and composition.
3. Estimation of manpower requirements.
4. Identification of spare parts and material requirements and their availability.
5. Identification of special tools required.

Maintenance scheduling on the other hand is the process of making schedules of the workforce and related tasks. This study focuses on strategic capacity planning which inherently directs the research towards maintenance planning. However, literature on maintenance scheduling often gives interesting insights in the needs and practices of capacity planning in general and is therefore not explicitly excluded from the discussion in the upcoming paragraphs.

2.1 Strategic capacity planning

Schneider and Cassady stress that capacity planning should be performed at fleet level as the units within the fleet have to share the available resources (Schneider and Cassady, 2015). The most important resources related to strategic capacity planning are spare parts, personnel and hangar space. Besides a fleet wide approach strategic capacity planning also relates to long-term planning (2-3 years) as stated by (Duffuaa and Alfares, 2009). Especially personnel or workforce management benefits from long-term planning as personnel training often takes years and personnel is difficult to come by in general. Next to this it is shown in Sec. 2.2 that there is hardly any literature on this subject and no known application in the aerospace sector, indicating the gap of knowledge in this field. Due to these reasons this research will focus on the strategic capacity planning of personnel.

2.2 Workforce management

In the aviation industry, literature regarding workforce capacity planning is mostly focused on the scheduling of personnel and the optimization of the schedule. Literature on resource planning is scarce but found to become more and more relevant, especially in the area of skill based workforce planning.

2.2.1 Workforce planning in aircraft maintenance industry

According to Heimerl and Kolish the first step in human resource capacity planning is the determination of the workforce required per skill and period (Heimerl and Kolisch, 2010). The second step is the actual assignment of people to work-packages.

Although the literature on the first step is scarce there are studies that describe the need of good workforce planning. Haroun and Duffuaa emphasize in their contribution to the Handbook of Maintenance that accurate forecasts for the future maintenance work are essential for determining the workforce capacity (Ben-Daya et al., 2009). They especially stress that the critical aspects of maintenance capacity are the expected workload size and the required skills of the workforce. Haroun and Duffuaa also note that: 'Making long run estimations is one of the areas in maintenance capacity planning that is both critical and not well developed in practice'. The observations of Haroun and Duffuaa are confirmed by Marquez as he argues that there are three main problems regarding workforce planning (Màrquez, 2007):

1. Determination of the workload, classified by skills
2. Determination of the ideal number of maintenance workers
3. Determination of the maintenance schedule

Marquez also states that an increasing amount of maintenance tasks in maintenance organizations require highly qualified and highly specialized personnel

The importance of accurate, skill based workforce planning is stressed by literature regarding the second step; the assignment of people to workpackages. Firat and Hurkens propose a method to schedule tasks which includes an inhomogeneous set of skilled personnel (Firat and Hurkens, 2012). They emphasize the importance of including skill management and state: 'Especially when activities require skills from several specialization fields at different levels, skill management becomes more challenging'. Other papers that describe this issue of multi skilled personnel scheduling in the maintenance context are Li and Womer (2009) and Dinis et al. (2019b). Wahyudin et al. propose an integrated Mixed Integer Linear Model model that involves not only the workforce but also the material use and tool use (Wahyudin et al., 2016). The authors stress the importance of skill based personnel planning as the costs of each of these specialists differ, suggesting that improvement of skill based workforce planning reduces costs significantly. In (Firat and Hurkens, 2012), (Li and Womer, 2009), (Dinis et al., 2019b), (Wahyudin et al., 2016) a known workforce demand is assumed but this demand should be forecasted during the planning phase.

The literature mentioned above signals a clear need for strategic workforce demand forecasts of specialized personnel. The literature on this subject is limited and therefore academic papers of other fields of study are consulted.

2.2.2 Workforce demand forecasting

Most literature regarding long-term workforce predictions revolves around macro economic problems and approach it using macro economic or demographic models that augment the main model.

Often workforce is forecasted indirectly by forecasting some indicator and then transforming it into demand afterwards. In (Edwards, 2010) forecasts for the demand of PhD students in Australia using an economic model is developed for policy makers to assess the ability of a country to build or maintain an innovation driven economy. The model uses external data sources such as economic outlooks, expected tourist numbers and forecasts of changes in technology. Another study uses predictions of sector growth as exogenous variable to compute skilled personnel demand (Woolard et al., 2006). In Sing et al. (2016) the authors present a case study of long-term workforce demand forecasting of building inspectors in Hong Kong. A labor multiplier approach is applied that assumes a known amount of work packages and a constant amount of labor per work package.

On the other hand, direct forecasts are also present in literature (Lee et al., 1998),(Spetz, 2017). Next to this Briscoe and Wilson propose a multitude of co-integrating regressions to establish the existence of long-run equilibrium relationships between employment, output and wages in the context of the engineering industry of Great Britain (Briscoe and Wilson, 1991). Forecasts of factor prices and interest rates are used as predictive variables. The authors emphasize the increase of forecast inaccuracy due to the introduction of these uncertain variables.

The studies above show that for long-term forecasting exogenous information is often used to forecast workforce demand. Regularly this exogenous information is a forecast in itself and therefore adds to the uncertainty of the main model.

Chapter 3

Fleet maintenance demand

As stated in the previous chapter the workforce demand should be assessed from a fleet wide perspective. Consequently, the factors that influence fleet maintenance should be looked for at this level.The literature analyzed in Ch. 2 showed that the use of exogenous predictors is a good method to eventually forecast maintenance demand. Therefore the aim of this chapter is to find the predictors that can aid in forecasting strategic maintenance workforce demand. Besides, it is expected that not all factors that influence strategic maintenance demand can be used as a predictor from a modeling perspective due to model or data restrictions. In this case these factors should be controlled for during the analysis.

Current airline fleet maintenance is organized around a predictive maintenance policy (Dupuy et al., 2011). In practice this means that inspections or so called checks are scheduled during which a set of pre-determined work packages are executed in order to inspect the integrity of the aircraft. Next to line maintenance the most commonly used checks nowadays are the A, C and D check as the B check tends to be combined with A checks (Qantas, 2016). Depending on the type of aircraft different inspection intervals are used. To give an indication the A, C and D check intervals of the B-747 are respectively 600-1000 flight hours, 7500-10000 flight hours and six years of service (Eggink and Bateman, 2010). The inspection work packages prescribed by the checks are well known in advance. This scheduled maintenance almost always induces unscheduled maintenance as well (Wagner and Fricke, 2006). Consequently, the A-Check, C-Check and D-Check schedules are not only a good predictor for scheduled maintenance but also an indicator for unscheduled maintenance.

The maintenance check schedules are implicit predictors as they do not describe the cause of maintenance demand. Factors that influence maintenance demand explicitly are not researched extensively in literature. To the authors knowledge only one paper applies factors as predictive variables to forecast maintenance demand. In Pogačnik et al. (2017) factors that influence fleet maintenance are used to predict fault probability on individual aircraft. Aircraft parameters such as operator, aircraft type, aircraft age, flight hours, flight cycles, engine type and operation location, are taken into consideration. Factors that impact direct maintenance costs are fleet size, commonality, fleet age, fleet utilization and frequency of check intervals (Wu et al., 2004). On single aircraft level, age is used as key factor to describe maintenance costs and forecasting maintenance demand (Saltoglu et al., 2016),(Weckman et al., 2006).

In short, not much is known about the factors influencing maintenance demand on a strategic level. Above mentioned studies suggest that the factors fleet age, check frequency, aircraft type, operator and fleet size are important fleet characteristics that should be taken into account when forecasting maintenance on fleet level.

Chapter 4

Demand forecasting

The previous chapters clearly identified the need for strategic workforce demand prediction and the potential factors that are of influence on this demand. This chapter aims to provide potential methodologies that are able to forecast maintenance demand. In Ch. 2 the preferred aggregation levels from a organizational perspective are found. However, from a forecasting point of view these aggregation levels might not be optimal. Therefore, a systematic analysis of aircraft maintenance demand modeling is presented, starting at the lowest level of aggregation which is the single component level.

At the basis of maintenance personnel demand is the failure and inspection of individual aircraft components and systems. Ghobbar and Friend state that demand forecasting at component level is one of the most challenging issues in the airline industry (Ghobbar and Friend, 2003). This is due to the intermittent demand characteristics of these assets. The dual source of variation, namely both the inter-demand intervals and demand sizes, is stochastic in nature and therefore difficult to forecast (Wang and Syntetos, 2011). The literature on intermittent demand forecasting is ample but before the techniques of intermittent demand forecasting can be assessed, demand classification should be treated as demand aggregation is strongly related to it.

4.1 Demand classification

According to Boylan et al. (2008) literature on demand classification is minimal. To the authors knowledge demand classification was first mentioned in literature by Williams (Williams, 1984). Williams reacted on the to that point standard approach of demand classification:

High sporadicity One demand at least ten times the average weekly demand.

Low sporadicity Average demand during a lead time less than 10.

No sporadicity Neither of the above.

Williams identified a number of problems with this classification method:

- No dimensionless conditions
- The definitions do not suggest what approach to use
- Sensitive to outliers
- Slow moving products are often forced into the highly sporadic category

To solve this issue Williams suggested a variance partition method that is able to cluster demand in one of four categories using two metrics. In reaction on this system an improvement was presented in (Syntetos and Boylan, 2005) which is based on the findings presented in (Johnston and Boylan, 1996). This new classification system is given in Fig. 4.1.

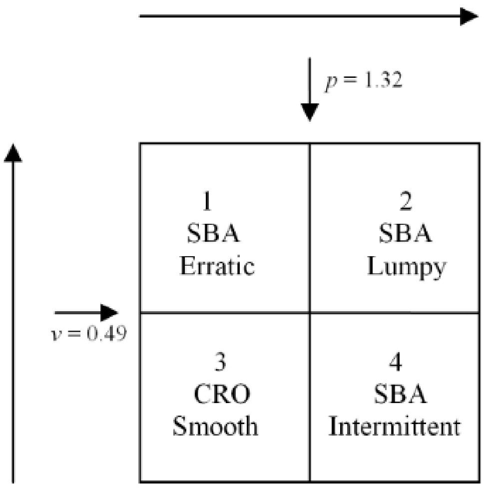

Figure 4.1: Demand categorization according to Syntetos and Boylan (2005) (SBC categorization). Figure from (Kostenko and Hyndman, 2006)

v is the coefficient of variation of demand sizes and p is the mean inter-demand interval. The abbreviations CRO and SBA are Croston's method and Syntetos Boylan Approximation method respectively, both of which are explained in Sec. 4.2. In a note on (Syntetos and Boylan, 2005) It should be noted that all the classification schemes mentioned in this section are based on empirical studies.

In this study the categorization scheme as presented in Fig. 4.1 is used.

4.2 Intermittent demand modeling

As stated in the introduction of this chapter aircraft component failure is known for its intermittent behavior (Ghobbar and Friend, 2003). According to the categorization scheme given in Fig. 4.1 intermittent demand has a relatively high mean inter-demand interval length but a small coefficient of variation. Especially the high mean inter-demand interval complicates the forecasting procedure.

Croston's method is one of the most used methods for intermittent demand forecasting in industry and Croston was the first to address the intermittent demand problem (Xu et al., 2012),(Boylan et al., 2008). He found that traditional forecasting techniques lead to excessive stock of spare parts in the case of intermittent demand which is inefficient (Croston, 1972).

It is useful to compare Croston's method with the Single Exponential Smoothing method as given in Eq. 4.1.

$$\begin{aligned} s_0 &= x_0 \\ s_t &= \alpha x_t + (1-\alpha)s_{t-1}, t > 0 \end{aligned} \tag{4.1}$$

Where s_t is the forecast for the next period, x_t is the demand at time t, α is the smoothing factor and s_{t-1} is the forecast for time t. α should be chosen to be between 0 and 1.

The Croston method applies two Single Exponential Smoothing forecasts models (SES) on both the size of the demand and the time period between demands (Croston, 1972). Another difference with SES is that Croston's method only updates when demand is found. The decomposition of the original series into two series of non-zero demand and inter-demand intervals is shown in Fig. 4.2. The mathematical description is given in Eq. 4.2.

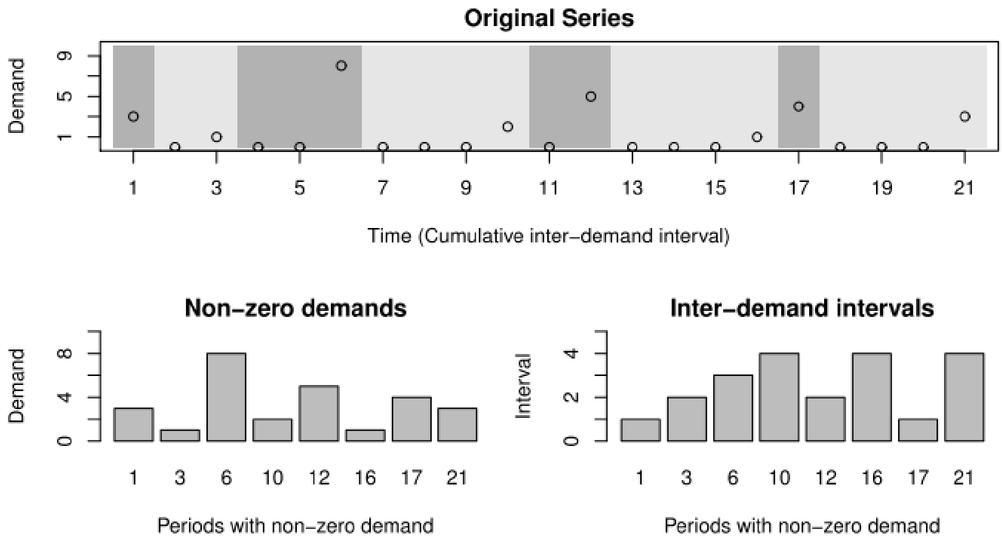

Figure 4.2: Croston's decomposition (Petropoulos et al., 2016)

$$
\begin{aligned}
\hat{z}_t &= \hat{z}_{t-1} + \alpha_z \left(z_{t-1} - \hat{z}_{t-1}\right) \\
\hat{p}_t &= \hat{p}_{t-1} + \alpha_p \left(p_{t-1} - \hat{p}_{t-1}\right)
\end{aligned} \tag{4.2}
$$

Where $\hat{z}$ is the estimate of the demand volume and $\hat{p}$ is the estimate of the intervals between demand. z_{t-1} is the last non-zero demand value, p_{t-1} is the last inter-demand interval. α_z and α_p are the smoothing factors for the non-zero demand and non-zero demand intervals respectively.

To get the original demand the two estimates are divided:

$$
\hat{y}_t = \frac{\hat{z}_t}{\hat{p}_t} \tag{4.3}
$$

Despite Croston's success there are some known issues regarding the method. Croston stated that the following assumptions are valid for his method:

1. The distribution of non-zero demand sizes is normally identically and independently distributed (i.i.d.)
2. The distribution of the inter-arrival times is also iid but then geometrically distributed
3. The demand size and the inter-arrival times are mutually independent

However the first two assumptions are not true (Shenstone and Hyndman, 2005). Shenstone and Hyndman propose a non-stationary variation of the Croston's method after comparing four versions of the model. Also it has tractable expressions for the forecast mean and variance. Due to the non-stationarity all four models forecasts tend to diverge at long forecast horizons which causes inefficient forecasts

In Syntetos and Boylan (2001) the authors find that Croston's method is biased. A follow-up study compares Simple Moving Average, Single Exponential Smoothing, Crostons Method and a method introduced by the authors called the Syntetos Boylan Approximation (SBA) (Syntetos and Boylan, 2005). The latter method mitigates the bias by adjusting the estimate given in Eq. 4.3 which results in:

$$\hat{y}_t = \left(1 - \frac{\alpha}{2}\right)\frac{\hat{z}_t}{\hat{p}_t} \tag{4.4}$$

Next to the highly successful SBA method there are other models that attempt to describe intermittent demand. In (Schultz, 1987) a very similar model to the Croston's model that splits up the forecasting procedure into a forecast for the demand size and a forecast for the period between demands is presented. It is extended upon by adding an estimate for the replenishment rate based on the forecast which includes a safety margin. In Willemain et al. (2004) the problem of forecasting intermittent demand is solved by introducing a new type of time series bootstrap. Another comparing study by Hua and Zhang compares this method with exponential smoothing, integrated forecasting technique and Logistic Regression Support Vector Machines (Hua and Zhang, 2006). The latter method allows for the introduction of explanatory variables which, according to the author, improves the forecasting accuracy significantly, outperforming the other methods.

A relatively new study regarding the modeling of lumpy demand is presented in Gutierrez et al. (2008). In this study a comparison of SBA single exponential smoothing, Croston's method and Neural Network models (NN models) is performed. Using a three layer perceptron with three nodes they were able to predict the demand with a superior accuracy compared to the other three models. The authors note however that the NN forecasts perform significantly worse than the traditional time series methods in the case of a decrease in average of nonzero demand sizes between the training set and the test set. Another paper applies the same methodology as presented in Gutierrez et al. (2008) but includes a Recurrent Neural Network (RNN) (Amin-Naseri and Rostami Tabar, 2008). The latter turns out to be most accurate for 21 of the 30 time-series that were considered. A study from 2004 compares different variations of Neural Networks with Croston's method.: The Radial Basis Function Neural Network (RBF), the Elliptical Basis Function Neural Network (EBF) and the Normalized Radial Basis Function Neural Network (NRBF) by (Carmo and Rodrigues, 2004). It turns out that the RBF models were adequate models for short-term forecasts of irregularly spaced time series, due to its ability to take into account non-linear correlations in the data. Carmo and Rodrigues recommend that the RBF model should be assessed using other performance indicators such as longer prediction horizons, or for time series with low-frequency periodicities.

Aforementioned literature shows that intermittent demand forecasting is still a complicated problem. Croston's method and its variations are mainly used in practice nowadays and neural networks are a promising alternative but are not proven to be adequate for long-term forecasts and are relatively difficult to apply. Due to the poor performance of intermittent demand forecasting techniques, other methodologies are tried in academia as presented in the upcoming section.

4.3 Temporal aggregation, cross-sectional aggregation and clustering

As stated in the previous chapters aircraft components generally exhibit this intermittent failure behavior. A fleet wide or strategic perspective on resource demand points towards the aggregation of demand in some way. In literature two forms of demand aggregation are mentioned: temporal aggregation and cross-sectional aggregation. Temporal demand aggregation is defined as follows: 'Temporal aggregation refers to aggregation in which a low frequency time series (e.g. quarterly) is derived from a high frequency time series (e.g. monthly) and is used for forecasting purposes' (Nikolopoulos et al., 2011). It is appealing to aggregate intermittent demand in lower-frequency bins as it causes a reduction in zero-values, forcing the demand characteristics towards a more smooth demand. In other words, it changes the demand category as presented in Fig. 4.1 (Murray et al., 2018).

The other form of aggregation is cross-sectional aggregation. This type of aggregation is related to the issue of heterogeneous demand which means that multiple sources of demand that have different behavior might be aggregated. A good example is an aircraft system which consists of multiple components. All components have different demand characteristics and the total system demand is therefore not homogeneous. Clustering or disaggregation into smaller clusters with similar failure behavior might be beneficial in terms of forecasting performance.

4.3.1 Temporal aggregation

As stated in the introduction of this section temporal aggregation might transform intermittent demand into smooth demand. This is advantageous as it allows for other forecasting methods than the limited selection of methods that are available for intermittent demand. Aggregating might however result in an information-loss as a direct result of a decrease in amount of datapoints (Petropoulos and Kourentzes, 2015). The effect of the level of aggregation on forecast accuracy is analyzed empirically in Nikolopoulos et al. (2011). In the study the ADIDA (Aggregate-Disaggregate Intermittent Demand Approach) method is developed which is well summarized in Fig. 4.3. The original data is aggregated in a lower frequency and a forecast is performed

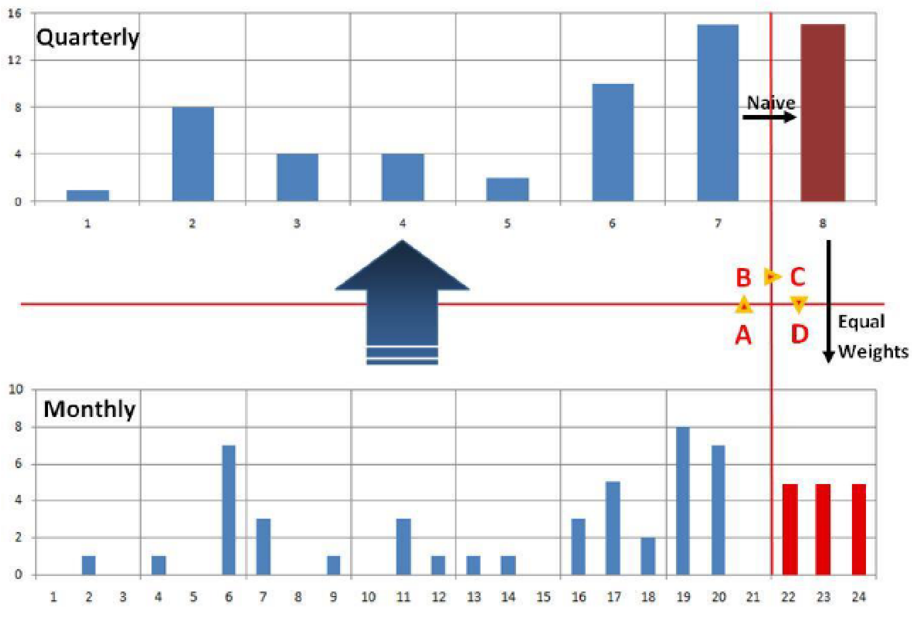

Figure 4.3: ADIDA method. A: Original data, B: Aggregated data, C: Forecast using aggregated data, D: Disaggregation of forecast using equal weights. (Nikolopoulos et al., 2011)

using the aggregated data. The disaggregation step from C to D is optional. In the study the ADIDA method was applied on the component demand data and compared with Crostons method and the SBA method for different aggregation levels. Thereafter an optimal aggregation level per part was determined and surprisingly for almost all parts the ADIDA method in combination with the naive estimator outperformed the SBA method, especially for relatively low aggregation levels. The optimal aggregation level was not the same for each part. However, a series wide optimal aggregation level of 9 could be identified. As a recommendation (Nikolopoulos et al., 2011) states that the issue of trend and seasonality should be assessed. Namely, these demand characteristics can potentially drop out when demand is aggregated to a lower frequency.

Nowadays there is a debate within academia about whether a single, optimal temporal aggregation level is preferred over the use of multiple combined aggregation levels. Namely, research showed that using single methods on multiple aggregation levels perform as well as the ADIDA method (Petropoulos and Kourentzes, 2015). An earlier study emphasizes the opportunities that a combined temporal aggregation model offers in terms of improved demand forecasts (Kourentzes et al., 2014). The main motivation of this study was to mitigate the issue of model selection which is tedious in general. The premise of the study is that at different aggregation levels different properties or characteristics of the time series can be identified and can therefore improve the forecast accuracy. The study also stresses the need for a theoretical background for the case of an optimal aggregation level which is not present in Nikolopoulos et al. (2011).

In Athanasopoulos et al. (2017) the multi temporal aggregation level approach is formalized, calling it reconciled forecasting. The following benefits of this approach are stated in this study:

1. Alignment of forecasts of different planning horizons (operational, tactical and strategic)
2. Increased forecast accuracy
3. Mitigating modeling risks

However, in Spiliotis et al. (2019a) limitations of the temporal hierarchies approaches introduced by Athanasopoulos et al. (2017) and Kourentzes et al. (2014) are identified and three strategies to mitigate these are proposed:

1. Combining methods for each base forecast
2. Adjusting base forecasts to mitigate bias
3. Avoid shrinking seasonality in combined forecast (induced by averaging of base forecasts)

In Spiliotis et al. (2019a) the authors suggest based on a literature study that a simple averaging approach of forecasting methods produces the most accurate forecast. Also from a practical consideration (computation time) a simple approach is more advantageous.

The studies presented above show that there are two approaches on temporal aggregation (Kourentzes et al., 2017). The first one is the usage of an optimal aggregation level and the second one is the usage of multiple aggregation level to retrieve all relevant features of the time series. The limitation of the single aggregation level approach is that a model should be assumed while the multiple aggregation level approach is robust against model choice. There is empirical and theoretical evidence that temporal aggregation is beneficial in terms of demand forecasting but there is no consensus on how that should be optimally achieved (Kourentzes et al., 2017).

4.3.2 Cross-sectional aggregation and clustering

Cross-sectional aggregation improves forecast accuracy analogous to temporal aggregation (Babai et al., 2012). The aggregation of components allows intermittent demand to become more smooth and therefore easier to model. From a strategic workforce capacity planning point of view, one would say that the maintenance demand generating components and aircraft should be aggregated up to fleet level. This means the summing of all parts of all aircraft in the fleet. However, it might be that from a modeling perspective other aggregation levels are more optimal. Clustering of components allows for shifting of heterogeneous demand generators into multiple homogeneous clusters. The premise is that these homogeneous clusters can be more accurately modeled than the fully aggregated demand.

In Kalchschmidt et al. (2006) aggregation is used to cluster heterogeneous customer groups into homogeneous groups which are less complex to model. In Misiti et al. (2010) a method to optimize clusters for dis-aggregated electricity load forecasting is proposed. The main idea behind both studies is to disaggregate the global demand into smaller clusters based on individual customer characteristics in such a way that the sum of the clusters improves the prediction of the whole global signal. Another study that uses the commonalities among individual demand generators to aggregate demand also concludes that aggregation of demand improves forecast accuracy (Zotteri and Verganti, 2001). In Dastidar (2017) it is stated that when dealing with volatile and intermittent demand, segmenting the products based on their demand characteristics is important.

Another approach to clustering looks at demand patterns. A relatively new technique in this field is Pattern Sequence-based Forecasting (PSF) which is applied in Martínez-Álvarez et al. (2011),Jin et al. (2015),Jin et al. (2014),Bokde et al. (2018). This method requires that at an individual level distinguishing patterns are present. The method is only applied on individual, smooth demand where a clear demand pattern can be identified at the lowest level of aggregation. Another attempt that tries to solve for intermittent demand forecasting is presented in Venkitachalam et al. (2003). In this study a clustering-bootstrap method is introduced that classifies individual components into groups that have similar failure patterns. The authors identify two benefits namely the provision of better statistics for forecasting models and a reduction in total amount of models.

The studies above show that when individual characteristics can be found, demand clustering is promising. However when similarities among individual demand generators are not there the clustering is not beneficial as is shown in Viswanathan et al. (2008). In this study a comparison is made between top-down (disaggregating demand from a fully aggregated time series) and a bottom-up approach (using the disaggregated data for the forecast). The authors concluded that when variability of the inter-order times is high a top-down approach is best in terms of model performance.

Shortly, it can be concluded that effective clustering of data can only be performed when characteristics of individual time series are known.

A logical continuation of this treatise is the application of hybrid models that combine cross-sectional aggregation with temporal aggregation. In Spiliotis et al. (2019b) the hybrid model implementation is primarily done to improve forecasting performance and to minimize the effect of modeling uncertainty. On top of that a cross-sectional hierarchical method is applied to ensure reconciliation across the levels of aggregation. It turns out that this hybrid aggregation method outperforms a method based on temporal aggregation only. An application of this hybrid approach on the Australian tourism sector showed slight benefits in terms of forecasting performance (Kourentzes and Athanasopoulos, 2019). However, the authors state that the largest benefit is from a organizational perspective as the reconciled nature of the forecast allows for alignment of local, regional and national tourism forecasts.

Both temporal and cross-sectional aggregation methods outperform traditional demand forecast models. Also from an strategic capacity planning perspective the aggregation of demand is desirable. Therefore this approach will be assumed in the methodology section.

Chapter 5

Methodology

The literature study provided the theoretical foundation which now can be used to formulate the methodology. As argued in Sec. 4.3 the aggregation of demand causes intermittent demand to become more smooth. Smooth demand can generally be forecasted more accurately but requires other models compared to intermittent demand. Preliminary data analysis shows a clear linear relationship between the potential variables and the demand data which directs the model selection process towards linear models. This, in combination with the research question at hand makes the General Least Square regression model (GLS) a valid starting point for the model selection process. Namely, the GLS offers an efficient way to assess the influence of the explanatory factors. A further refinement of model selection is not possible during this stage of the research. Therefore the methodology incorporates a model selection procedure that is steered by data characteristics.

When not stated explicitly the source of the statistical theory is from Fox (2016).

5.1 Generalized Least Squares Regression

The generalized linear model can be described as follows:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \tag{5.1}$$

Where $\mathbf{y}$ is a $(n \times 1)$ vector of the dependent variable values, $\mathbf{X}$ is the $(n \times k + 1)$ matrix of the values of the independent variables. β is the $(k + 1 \times 1)$ vector of parameters and ϵ is the $(n \times 1)$ vector containing the error term. In the GLS case the error term is not independently distributed but can be described as:

$$\epsilon \sim N_n(\mathbf{0}, \mathbf{\Sigma}_{\epsilon\epsilon}) \tag{5.2}$$

Where $\mathbf{\Sigma}_{\epsilon\epsilon}$ is the covariance matrix. Using a log-likelihood estimation the estimation of β becomes:

$$\mathbf{b}_{GLS} = (\mathbf{X}'\mathbf{\Sigma}_{\epsilon\epsilon}^{\mathbf{-1}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}_{\epsilon\epsilon}^{\mathbf{-1}}\mathbf{y} \tag{5.3}$$

This estimator requires $\mathbf{\Sigma}_{\epsilon\epsilon}$ to be known but in practice this is never the case. Also, simultaneous estimation of $\mathbf{\Sigma}_{\epsilon\epsilon}$ and β is not realistic as the amount of unknowns is too large. To be able to estimate β, restrictions for $\mathbf{\Sigma}_{\epsilon\epsilon}$ have to be found. In the following sections applicable restrictions are elaborated on. These particular models will form the input for the general methodology, graphically described in Fig. 5.2.

In least square regression an important concept is the Best Linear Unbiased Estimator (BLUE) concept which is based on the Gauss-Markov theorem. This theorem states: 'When the error probability distribution is unknown in a linear model, then, amongst all of the linear unbiased estimators for the parameters of the linear model, the estimator obtained using the method of least squares is the one that minimizes the variance. The mathematical expectation of each error is assumed to be zero, and all of them have the same (unknown) variance'. (Springer, 2008). Where best means lowest variance and no bias. The requirements for the model posed by the BLUE property of no autocorrelation, constant variance and mean of zero are by definition true for the Ordinary Least Squares regression (OLS) which is why this is an interesting restriction of the General Least Squares model.

5.1.1 Ordinary Least Squares Regression

OLS poses the most stringent restriction on the covariance matrix. It assumes that all diagonal entries are equal and positive. The rest of the entries are zero. This model thus assumes no auto-correlation (serial dependency) and homoscedasticity. Due to this assumption of constant variance the β estimator reduces to:

$$\mathbf{b}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{5.4}$$

The OLS is in essence a special form of the Weighted Least Squares regression as all weights are equal to one. To cope with heteroscedastic errors the more general Weighted Least Squares regression can be used.

5.1.2 Weighted Least Squares Regression

If no auto-correlation of the error is assumed but when the errors are heteroscedastic the OLS model is consistent but not efficient. A solution to this heteroscedasticity is the application of weighted least squares (WLS). This rationale behind this method is to apply higher weights to the data points with low variance and lower weights to data points with high variance.

$$\mathbf{b}_{WLS} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \tag{5.5}$$

Where $\mathbf{W}$ is the matrix of weight values. In principle any weight matrix can be constructed however only if each weight is equal to the reciprocal of the variance of the measurement the $\mathbf{b}_{WLS}$ is also BLUE. When a physical explanation can be given for the heterogeneity, explanatory variables can be used to form the weight matrix. WLS is especially valuable for small data-sets.

5.1.3 Transformation of dependent variable

A transformation of the dependent variable is also a way to reduce heteroscedasticity. However, transforming variables can have large consequences for the interpretability of the model and while it may increase the in-sample model fit it does not automatically mean that the out-of-sample predictions after back-transformations are more accurate. Therefore, transformations should be applied with care.

5.2 Time series regression

When the errors are auto-correlated and the errors are heteroscedastic a time series regression is a valid option. In this section classical time-series forecasting methods are described mathematically. For the sake of brevity extensive derivations and proofs of the statistical properties of the models are omitted. When not specified the model descriptions including the mathematical derivations and the assumptions are obtained from Palma (2016) and Kendall and Ord (1990).

The models described below are of the ARIMA family which is a special case of the GLS.

5.2.1 ARIMA model family

The Autoregressive Integrated Moving Average (ARIMA) model family is a well-known type of model. It is build up from different components and all these components are elaborated upon in this section.

Autoregressive model

The autoregressive model (AR) is a regression model which regresses on its past values. It was first introduced by Yule in (Yule, 1927). In a modern mathematical representation the AR(p) model is described in the following way:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \tag{5.6}$$

Where X_t is the dependent variable of interest, ϕ_i is the model parameter and ϵ_t is the error term ($\epsilon_t \sim i.i.d.N(0, \sigma^2)$). Finally p designates the order of the model. For the AR process to be stationary $|\phi| < 1$. A shock at time t (induced by ϵ_t) affects the outcome of the variable infinitely long. However, if the AR process is stationary the effect of the shock diminishes to 0.

Important model characteristics in general are the mean, the variance and the covariance. If assumed that both statistics are independent of time, simple expressions for them can be found. For sake of brevity the AR(1) model is used in the description below:

$$\begin{aligned}
X_t &= c + \phi X_{t-1} + \epsilon_t && (5.7)\\
E(X_t) &= E(c) + E(\phi X_{t-1}) + E(\epsilon_t) && (5.8)\\
\mu &= c + \phi\mu && (5.9)\\
\mu &= \frac{c}{1-\phi} && (5.10)
\end{aligned}$$

The variance is given by:

$$\begin{aligned}
var(X_t) &= var(c) + var(\phi X_{t-1}) + var(\epsilon_t) && (5.11)\\
var(X_t) &= \phi^2 var(X_t) + \sigma^2 && (5.12)\\
var(X_t) &= \frac{\sigma_\epsilon^2}{1-\phi^2} && (5.13)
\end{aligned}$$

$$Cov(X_t, X_{t+h}) = Cov(X_t, \phi X_t) = \phi^h var(X_t) = \frac{\phi^h \sigma^2}{1-\phi^2} \tag{5.14}$$

The correlation of the AR(p) process depends on ϕ^h which causes the model to be weakly dependent and have a memory. This is also the reason why the model is affected by a shock ϵ_t for a theoretical infinite amount of periods. Due to the property of constant mean and variance the AR model can be used to model non-seasonal, non-trend time series.

Moving Average Model

The moving average model (MA) states that a variable is linearly dependent on the current and selected previous values of a stochastic term. Mathematically the MA(q) is described as given below:

$$X_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{5.15}$$

Where ϵ_t is the error term ($\epsilon_t \sim i.i.d.N(0, \sigma^2)$) and θ is the model parameter and μ is a constant. In the equations below the most important statistics of the MA model are given:

$$E(X_t) = E(\mu + \epsilon_t + \theta\epsilon_{t-1}) = \mu \tag{5.16}$$

$$var(X_t) = var(\mu + \epsilon_t + \theta\epsilon_{t-1}) = \sigma^2 + \theta^2\sigma^2 = (1+\theta^2)\sigma^2 \tag{5.17}$$

$$Cov(X_t, X_{t-1}) = Cov(\mu + \epsilon_t + \theta\epsilon_{t-1}, \mu + \epsilon_{t-1} + \theta\epsilon_{t-2}) = \theta Cov(\epsilon_{t-1}, \epsilon_{t-1}) = \theta\sigma^2 \tag{5.18}$$

The latter is true because ($\epsilon_t \sim i.i.d.(0, \sigma^2)$) and therefore covariance is equal to the variance. Contrarily to the AR model the shock induced by ϵ_t only affects the resulting variable for q+1 periods. This is due to the fact that $Corr(X_t, X_{t-\tau}) = 0$ if τ is larger than q. On the contrary the AR(p) process has a correlation function that is dependent on h which is analogous to τ. Similar to the AR model the MA model is able to model non-seasonal and non-trend time series.

Autoregressive Moving Average Model

Combining the AR and MA models is an evident next step. Autoregressive Moving Average Models (ARMA) models are widely used in literature. The general ARMA model was introduced by Peter Whittle in 1953 and made popular by the book of Box and Jenkins in 1970. The ARMA model is formulated as follows:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{5.19}$$

Autoregressive Integrated Moving-Average Model

A generalization of the ARMA model is the Autoregressive Integrated Moving-Average Model (ARIMA). In contrast with ARMA, ARIMA is able to model a trend in the data. Introducing the lag operator (also called backshift operator) B the ARMA model can mathematically be described as:

$$(1 - \sum_{i=1}^{p'} \alpha B^i)X_t = (1 + \sum_{i=1}^{q} \theta_i B^i)\epsilon_t \tag{5.20}$$

Where $B^i X_t = X_{t-k}$. If one assumes that the left part of Eq. 5.20 has a unit root $(1 - B)$ an ARIMA process can be described as follows:

$$(1 - \sum_{i=1}^{p} \phi B^i)(1 - L)^d X_t = (1 + \sum_{i=1}^{q} \theta_i B^i)\epsilon_t \tag{5.21}$$

Where, compared to the ARMA model in Eq. 5.20, $p = p' - d$. Informally stated: the crucial part of the ARIMA model is that by differencing a random variable with its past value a possible trend is 'removed' after which regular ARMA model is applied. By removing the trend a stationary process is created which can be described by an ARMA model.

An ARIMA model is suitable for modelling time series data with a trend but without seasonality. To include seasonality effects SARIMA is developed. SARIMA is an extension of the ARIMA model. It includes multiple differencing terms which account for the seasonality effects. To include exogenous variables an ARIMAX model or ARMAX model can be applied which can mathematically be described as follows:

$$X_t = \beta Y_t + \epsilon_t \tag{5.22}$$

$$\epsilon_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{5.23}$$

This representation is often used as it allows for a straightforward interpretation of the correlation coefficient β.

5.2.2 Autoregressive Conditional Heteroskedasticity models

In the previous section various methods that can deal with changing mean have been described. Especially in the area of volatility analysis there are various studies where a time-series is analyzed that is non-stationary due to a non-constant variance. The class of models that can model changing variances are (G)ARCH models. This model type is introduced by Engle in (Engle, 1982).

First the ARCH model is explained. For the sake of brevity the ARCH(1) model is selected to serve as an example. The same but slightly more tedious analysis can be done for the general case of an ARCH(q) model.

$$y_t = \sigma_t \epsilon_t \tag{5.24}$$

Where y_t are the so called return residuals (errors of a mean process such as ARIMA or ARMA), σ_t is the time dependent standard deviation while ϵ_t is the white noise random variable. σ_t is modeled as follows:

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 \tag{5.25}$$

Which is analogous to the AR process.

A generalization of the ARCH(q) model is the GARCH(p,q) model introduced by (Bollerslev, 1986). This model includes an MA process which results in the following mathematical representation:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{5.26}$$

Note the similarity with the ARMA(p,q) model given in Eq. 5.19. The GARCH model adds a term that includes past values of the standard deviation.

For econometric purposes the prediction of variance (volatility in jargon) could be a goal on itself as it gives insight in market or stock behavior. However, from a demand forecasting point of view one is interested in combining the variance estimate with the mean estimate to achieve a better overall model fit and prediction. These so called hybrid models are applied extensively in literature such as in the field traffic flow modeling (Chen et al., 2011),(Zhou et al., 2006), in agriculture (Paul, 2015) and electricity price forecasting (Tan et al., 2010), (Liu and Shi, 2013).

5.3 Proposed methodology

To be able to answer the main research question the influence of fleet characteristics on maintenance demand forecasting performance has to be assessed. As stated in the beginning of this chapter regression analysis is a suitable method for this problem.

5.3.1 Data description

The data available is Boeing 777 maintenance log data. This data contains information on failure times of parts per aircraft. There is no information given on the required man hours for the related maintenance task. Consequently, a conversion from the amount of failures to the actual workforce demand cannot be made. Therefore, the demand that will be forecasted is the amount of failures of a selection of components for the entire fleet. The selection of components is based on the need for skill based planning as described in Ch. 2. Composite components generally require specialized personnel which is why as the component selection all components consisting fully or partly of composite material are chosen. After this selection about 3000 data points from 30 aircraft are left for analysis, spread over 10 years.

5.3.2 Predictor variables and controlled factors

In Ch. 3 it was shown that there were several predictor variables of interest for strategic maintenance demand forecasting purposes. Preliminary data exploration and analysis show that four of them are feasible. The variables are retrieved from external data sources and are listed below:

- Fleet size
- Average fleet age
- Expected demand due to A-Check
- Expected demand due to C-Check

The values of these variables are all well known in advance and assumed to be fixed. This means that they are not prone to uncertainty. This assumption can be made due to the fact that aircraft have long delivery times. Also, due to the inspection based maintenance policy that airlines apply nowadays, A-Check and C-Check schedules can be determined for years to come. In order to use the schedules of A-Check and C-Check as explanatory variables a continuous model should be made of them with these schedules as input. The four factors will be assessed in different combinations.

The variables have properties that are of influence on the methodology. It is expected that the variables show collinearity due to the fact that they are all related to time and the amount of aircraft in one way or the other. This has some consequences for statistical test interpretability which should be taken into account when constructing the methodology. Next to the collinearity issue the variables are physically not continuous but ordinal. From a modeling perspective this has consequences as ordinal variables should be modeled as categorical variables. However, one can consider ordinal variables to be continuous when the amount of categories is larger than eight and therefore the variables will be assumed to be continuous.

Next to the predictor variables there are factors that might influence demand characteristics which cannot be formatted into a variable. These factors should be controlled for. These factors are listed below:

- Aircraft type
- Airline
- Aircraft usage
- Acquisition policy

There is only one aircraft type described in the data but the available aircraft belong to two different airlines. According to the literature given in Ch. 3 the latter might be of influence which is why it should be controlled for. Due to the unbalanced data (one airline owns significantly more aircraft than the other) it is not wise to include this factor as a variable which is why the aircraft of the smaller airline are excluded from the analysis. This also minimizes the influence on the maintenance demand due to the usage of the aircraft as both airlines might assign their aircraft to different types of routes that have different operational characteristics or environmental conditions. Although usage data of the individual aircraft are not available it can be assumed that the aircraft belonging to one airline are used in similar conditions which eliminates that influencing factor automatically. This assumption is even more valid considering the long-term perspective of this research and the fact that aircraft are assigned to different routes continuously during their lifetime. Another factor is the fleet planning policy. It is possible that the demand forecast on fleets that are ordered in batches instead of more spread out over time have a different forecast performance. Also fleet size might have an influence on the forecasting performance. This variation cannot easily be transformed into a explanatory variable which is why it is included in the model in the form of various fleet scenarios.

5.3.3 Data pre-processing and model assessment

In Fig. 5.1 the proposed methodology is shown graphically. The need for a strategic perspective on specialized workforce planning requires strategic demand forecasts. Therefore, the aggregation should be done at fleet level and over all components that require the same, specific skill to maintain them. Preliminary data analysis showed that there are no demand characteristics known at the lowest aggregation level which means that clustering is not beneficial.

The strategic perspective on maintenance demand makes that a monthly or quarterly temporal aggregation is most suitable. The variable set is re-sampled accordingly and per temporal aggregation level an OLS is applied using all four variables. Thereafter the residuals are then used to assess the potential models. This process is explained in Fig. 5.2. First the linearity assumption is checked visually followed by a visual check of the presence of monthly periodicity. Also autocorrelation and homoscedasticity are checked visually. Depending on the residual characteristics, a single or a multitude of models can be found. The residuals of all the potential models are assessed using the full data-set in both the monthly and quarterly temporal aggregation level.

5.3.4 Forecasting and result analysis

After the model assessment the forecasting performance of the potential models will be investigated. An extensive description of the forecasting analysis is given in Fig. 5.3. Each potential model that comes out the model assessment phase is put through the flowchart presented in this figure. Eight different fleet scenarios are formed from the available data set. In order to validate the model ten variations of the fleet are constructed within the limits that the fleet scenario poses. Both temporal aggregation levels are used for analysis. Six evenly spaced split up points are chosen, the first one being after two years of operations and the last one three years before the last datapoint. Next to this, five different forecast horizons are used to vary among. Using the split up point and the forecast horizon the training-set and the test set are constructed. The model is trained on the training set and applied on the test set. If the respective model applied a transformation on the data, the values are back transformed to receive the true demand forecast. Also, the performance of the combined forecast (using the prediction of both the monthly and quarterly forecast) is assessed. According to literature it cannot be determined beforehand if a combined temporal aggregation level forecast is more accurate than a single optimal aggregation level forecast and therefore the combination will also be analyzed. The disaggregation step as presented in the ADIDA method (see Fig. 4.3) will be used for this purpose. The used performance metric is the Root Mean Squared Error (RMSE)

The output of the forecasting module is a multi-labeled list of RMSE values. To assess the influence of the variable sets a multi-way Analysis Of Variance (ANOVA) will be performed. This method focuses on the difference among group means, in this case the grouped RMSE value means. Using this method the difference in groups and thus the difference due to the variable selection can be assessed quantitatively.

This flexible approach also allows for nuanced and systematic analysis of the results which is useful when differences in forecast performance are not directly noticeable. For example, it might be that at the highest level there is not a large difference in performance while for specific fleet scenarios, or specific splitting-points there is a difference. On top of that, this approach allows for a thorough analysis of the reasons for the potential differences.

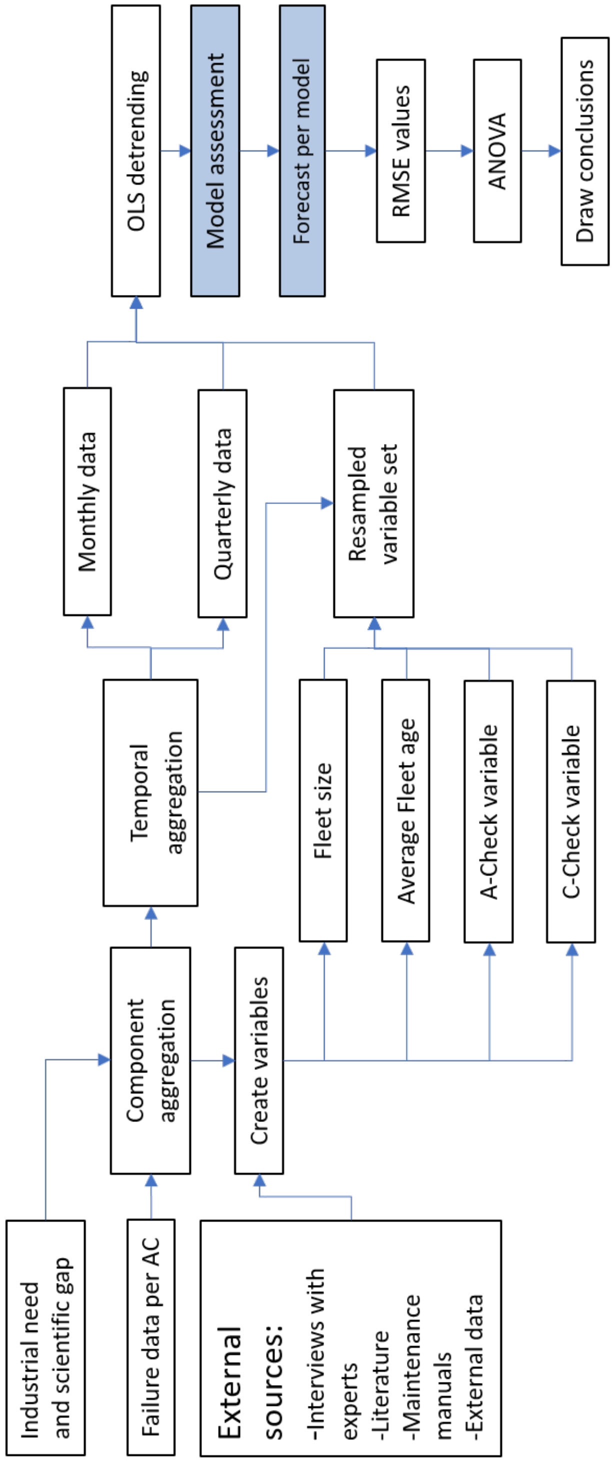

Figure 5.1: Methodology overview. The content of the blocks model assessment, forecasting and analysis of RMSE data are given in figures 5.2 and 5.3

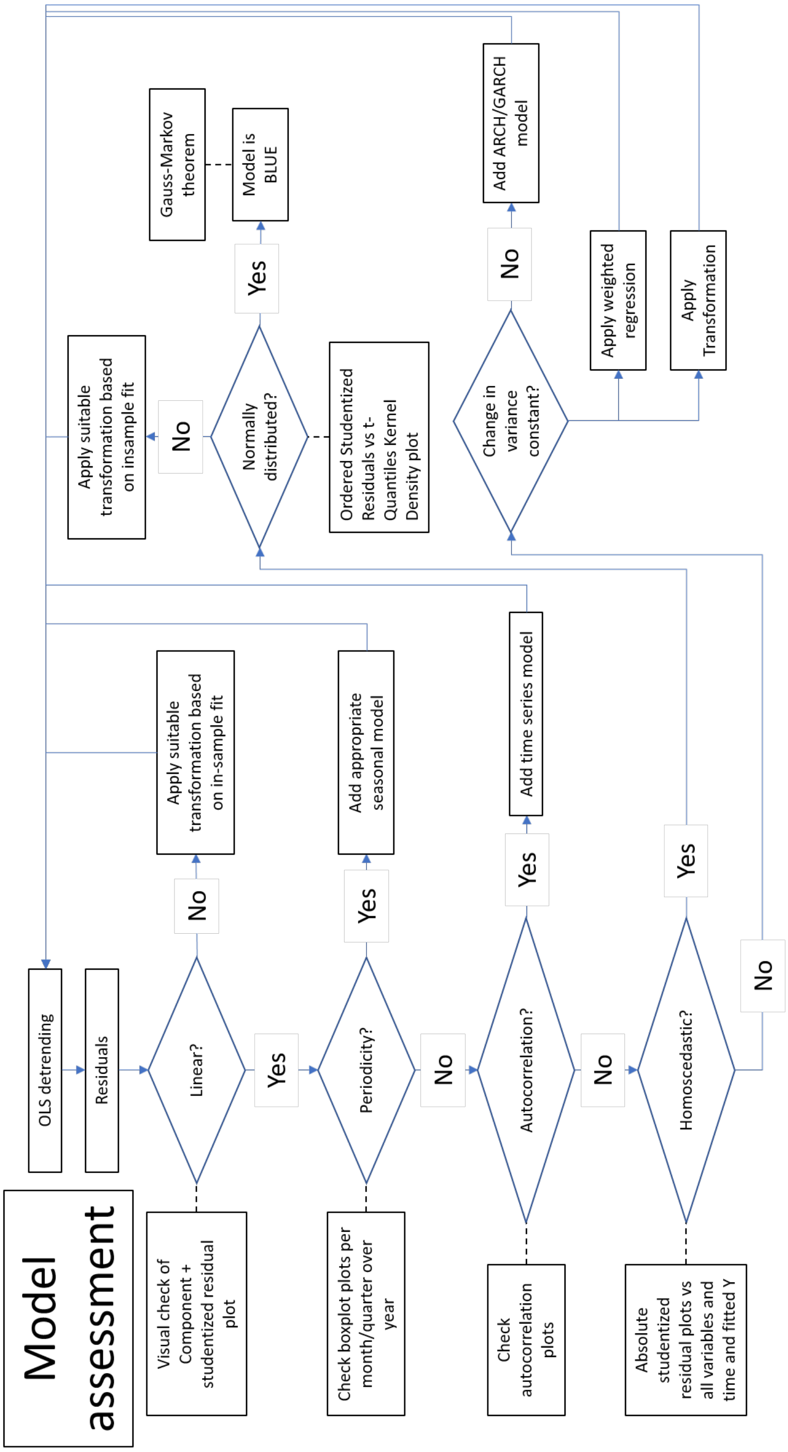

Figure 5.2: Model assessment phase

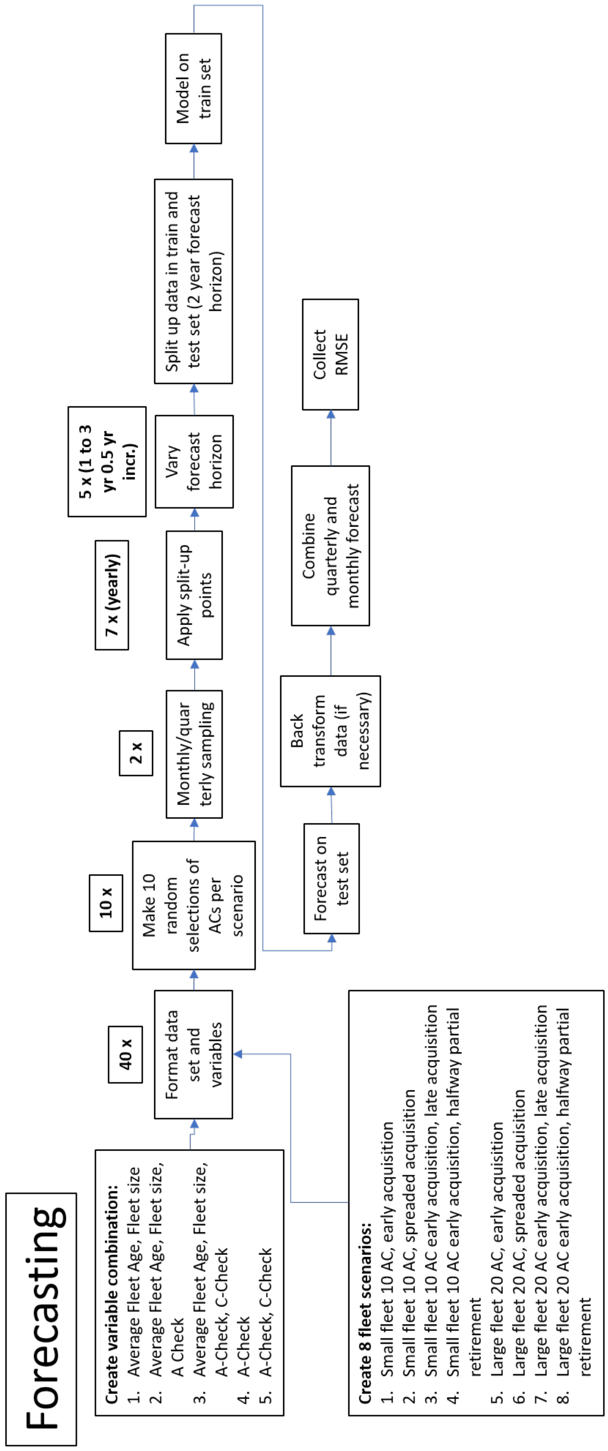

Figure 5.3: Forecasting phase

Chapter 6

Conclusion

The aim of this literature study is to provide the necessary theoretical background for the thesis research. This entails building a strong foundation for the in the thesis addressed need, assessing the state of the art literature regarding the subject, finding the gap in scientific knowledge and exploring the available methods of analysis. The goal of these literature study objectives is to fully or partly answer the research questions at hand. These questions, presented in the introduction (Ch. 1), are answered in this report and are summarized in this conclusion. The statements and conclusions are all based on sources and observations presented in the chapters of this report. For the sake of legibility the corresponding references are not repeated in here.

Strategic capacity planning aims to plan the resources hanger space, spare parts and workforce. At a strategic level especially workforce is relevant due to the long training times and high costs involved. In the field of aircraft maintenance no literature on strategic workforce planning could be found. However, literature from other sources suggest that there is a large need for the skill based planning of workforce as maintenance and manufacturing of high-end products such as aircraft require specialized personnel. Due to the clear need for further research, the focus of this research is on strategic demand forecasting to improve workforce capacity planning.

Strategic demand forecasting requires the appropriate levels of aggregation and a suitable forecast horizon/temporal frequency. The latter two are not clearly defined in literature. However, experts suggest a 1-3 year forecast horizon with a quarterly or monthly frequency. The demand of structural aircraft components is intermittent. This causes difficulties as intermittent demand forecasting models are often inaccurate. The literature study shows that temporal aggregation of demand provides the best forecasting performance as it transforms intermittent demand into smooth demand which makes other models available. The choice of temporal aggregation level can either be optimized for or a combined aggregation level approach can be applied. During the current stage of the research a choice between them cannot be made which is why both approaches are researched.

From a strategic standpoint aggregation of components up to fleet level is most optimal. Preliminary data analysis showed that there are no component characteristics available at the lowest level of aggregation. Therefore clustering is not a valid option.

The influence of fleet properties and fleet planning on maintenance is scarcely described in literature. Textbooks on fleet planning state that the most important factor that is taken into account regarding maintenance is fleet commonality as it impacts the maintenance costs considerably. Influential fleet properties are fleet size, fleet age, airline, maintenance check schedule and aircraft usage. Data availability makes fleet size, fleet age, A-Check schedule and C-Check schedule feasible variables and these are the ones that will be assessed during the research. Fleet commonality, aircraft usage and fleet planning are factors that can have an influence on either the demand itself or the demand forecasting and are therefore controlled for.

To evaluate the influence of fleet size, fleet age, A-Check schedule and C-Check on strategic maintenance demand forecasting a regression model is a logical choice. The methodology includes a model selection phase where properties of the data such as serial dependence, seasonality and heteroskedasticity determine which model is most suitable. It is expected that multiple models result from that selection phase and all of these are then used in the forecasting phase where the performance of each model with different variable selections is assessed. Each model is applied in different fleet settings, using multiple variable combinations, using multiple groups of aircraft (to ensure cross-validation) and by using multiple split-up points and forecast horizons. The RMSE values resulting from the forecasting phase are then analyzed using an ANOVA analysis. This analysis can assess the significance of the difference between means of RMSE groups.

The findings in this literature studies enabled the answering of the sub questions. The methodology extracted from the preliminary data analysis and the theoretical background is solid and will allow for a good quantitative analysis. This analysis will be the continuation of the master thesis which will ultimately answer the research question and add to the body of science on strategic maintenance demand prediction.

Bibliography

Amin-Naseri, M. R. and Rostami Tabar, B. (2008). Neural network approach to lumpy demand forecasting for spare parts in process industries. *Proceedings of the International Conference on Computer and Communication Engineering 2008, ICCCE08: Global Links for Human Development*, pages 1378–1382.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., and Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74.

Babai, M. Z., Ali, M. M., and Nikolopoulos, K. (2012). Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *Omega*, 40(6):713–721.

Ben-Daya, M., Duffuaa, S. O., Raouf, A., Knezevic, J., and Ait-Kadi, D. (2009). *Handbook of Maintenance Management and Engineering.* Springer International Publishing.

Bokde, N., Feijóo, A., Villanueva, D., and Kulat, K. (2018). A novel and alternative approach for direct and indirectwind-power prediction methods. *Energies*, 11(11).

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(1):107–118.

Boylan, J. E., Syntetos, A. A., and Karakostas, G. C. (2008). Classification for forecasting and stock control: A case study. *Journal of the Operational Research Society*, 59(4):473–481.

Briscoe, G. and Wilson, R. (1991). Explanations of the demand for labour in the United Kingdom engineering sector. *Applied Economics*, 23(5):913–926.

Carmo, J. L. and Rodrigues, A. J. (2004). Adaptive forecasting of irregular demand processes. *Engineering Applications of Artificial Intelligence*, 17(2):137–143.

Chen, C., Hu, J., Meng, Q., and Zhang, Y. (2011). Short-time traffic flow prediction with ARIMA-GARCH model. *Econometrica*, 100084(Iv):607–612.

Croston, J. D. (1972). Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly*, 23(3):289–303.

Dastidar, A. G. (2017). Intermittent demand forecasting for long tail SKUs. In *2017 International Conference on Data Management, Analytics and Innovation, ICDMAI 2017*, pages 334–339.

Dinis, D., Barbosa-Póvoa, A., and Teixeira, Â. P. (2019a). A supporting framework for maintenance capacity planning and scheduling: Development and application in the aircraft MRO industry. *International Journal of Production Economics*, 218(February):1–15.

Dinis, D., Barbosa-Póvoa, A., and Teixeira, Â. P. (2019b). A supporting framework for maintenance capacity planning and scheduling: Development and application in the aircraft MRO industry. *International Journal of Production Economics*, 218(April):1–15.

Duffuaa, S. O. and Al-Sultan, K. S. (1997). Mathematical programming approaches for the management of maintenance planning and scheduling. *Journal of Quality in Maintenance Engineering*, 3(3):163–176.

Duffuaa, S. O. and Alfares, H. (2009). *Handbook Maintenance Management*. Number January 2009.

Dupuy, M. J., Wesely, D. E., and Jenkins, C. S. (2011). Airline fleet maintenance: Trade-off analysis of alternate aircraft maintenance approaches. *2011 IEEE Systems and Information Engineering Design Symposium, SIEDS 2011 - Conference Proceedings*, pages 29–34.

Edwards, D. (2010). The future of the research workforce: Estimating demand for PhDs in Australia. *Journal of Higher Education Policy and Management*, 32(2):199–210.

Eggink, R. and Bateman, P. (2010). 747 - 8 Offers Operational Improvements and Cross-Model Commonality. Technical report, The Boeing Company.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007.

Firat, M. and Hurkens, C. A. (2012). An improved MIP-based approach for a multi-skill workforce scheduling problem. *Journal of Scheduling*, 15(3):363–380.

Fox, J. (2016). *Applied regression analysis & generalized linear models*. SAGE publications.

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 and Operations Research*, 30(14):2097–2114.

Gutierrez, R. S., Solis, A. O., and Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111(2):409–420.

Heimerl, C. and Kolisch, R. (2010). Scheduling and staffing multiple projects with a multi-skilled workforce. *OR Spectrum*, 32(2):343–368.

Hua, Z. and Zhang, B. (2006). A hybrid support vector machines and logistic regression approach for forecasting intermittent demand of spare parts. *Applied Mathematics and Computation*, 181(2):1035–1048.

Jin, C. H., Pok, G., Paik, I., and Ryu, K. H. (2015). Short-term electricity load and price forecasting based on clustering and next symbol prediction. *IEEJ Transactions on Electrical and Electronic Engineering*, 10(2):175–180.

Jin, C. H., Pok, G., Park, H. W., and Ryu, K. H. (2014). Improved pattern sequence-based forecasting method for electricity load. *IEEJ Transactions on Electrical and Electronic Engineering*, 9(6):670–674.

Johnson, B. (2018). Another look at the aviation maintenance personnel shortage and the solutions.

Johnston, F. and Boylan, J. E. (1996). Forecasting for Items with Intermittent Demand. *The Journal of the Operational Research Society*, 47(1):113–121.

Kalchschmidt, M., Verganti, R., and Zotteri, G. (2006). Forecasting demand from heterogeneous customers. *International Journal of Operations and Production Management*, 26(6):619–638.

Kendall, M. and Ord, K. (1990). *Time Series*. Edward Arnold ltd., 3 edition.

Kostenko, A. V. and Hyndman, R. J. (2006). A Note on the Categorization of demand patterns. *Journal of the Operational Research Society*, 57(10):1256–1257.

Kourentzes, N. and Athanasopoulos, G. (2019). Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*, 75(2018):393–409.

Kourentzes, N., Petropoulos, F., and Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2):291–302.

Kourentzes, N., Rostami-Tabar, B., and Barrow, D. K. (2017). Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels? *Journal of Business Research*, 78(April):1–9.

Lee, P. P., Jackson, C. A., and Relles, D. A. (1998). Demand-based assessment of workforce requirements for orthopaedic services. *Journal of Bone and Joint Surgery - Series A*, 80(3):313–326.

Li, H. and Womer, K. (2009). Scheduling projects with multi-skilled personnel by a hybrid MILP/CP benders decomposition algorithm. *Journal of Scheduling*, 12(3):281–298.

Liu, H. and Shi, J. (2013). Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, 37:152–166.

Maintenance Cost Task Force (2018). Airline Maintenance Cost Executive Commentary Public Version. (November):14.

Màrquez, A. (2007). *Models to deal with maintenance capacity planning*, volume 14.

Martínez-Álvarez, F., Troncoso, A., Riquelme, J. C., and Aguilar Ruiz, J. S. (2011). Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, 23(8):1230–1243.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J.-M. (2010). Optimized clusters for disaggregated elecricity load forecasting. *Revstat*, 8(2):105–124.

Murray, P. W., Agard, B., and Barajas, M. A. (2018). Forecast of individual customer's demand from a large and noisy dataset. *Computers and Industrial Engineering*, 118(July 2017):33–43.

Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., and Assimakopoulos, V. (2011). An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: An empirical proposition and analysis. *Journal of the Operational Research Society*, 62(3):544–554.

Palma, W. (2016). *Time Series Analysis*. John Wiley & Sons, Ltd.

Paul, R. K. (2015). ARIMAX-GARCH-WAVELET model for forecasting volatile data. *Model Assisted Statistics and Applications*, 10(3):243–252.

Petropoulos, F. and Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66(6):914–924.

Petropoulos, F., Kourentzes, N., and Nikolopoulos, K. (2016). Another look at estimators for intermittent demand. *International Journal of Production Economics*, 181(April):154–161.

Phillips, P., Diston, D., Starr, A., Payne, J., and Pandya, S. (2009). A review on the optimisation of aircraft maintenance with application to landing gears. *Engineering Asset Lifecycle Management - Proceedings of the 4th World Congress on Engineering Asset Management, WCEAM 2009*, (September):68–76.

Pogačnik, B., Duhovnik, J., and Tavčar, J. (2017). Aircraft fault forecasting at maintenance service on the basis of historic data and aircraft parameters. *Eksploatacja i Niezawodnosc*, 19(4):624–633.

Qantas (2016). The A, C and D check. In *https://www.qantasnewsroom.com.au/roo-tales/the-a-c-and-d-of-aircraft-maintenance/*.

Saltoglu, R., Humaira, N., and Inalhan, G. (2016). Maintenance stop time influence on aircraft total maintenane cost with downtime integrated cost model. *Proceedings of 2016 7th International Conference on Mechanical and Aerospace Engineering, ICMAE 2016*, pages 502–506.

Schneider, K. and Cassady, C. R. (2015). Evaluation and comparison of alternative fleet-level selective maintenance models. *Reliability Engineering and System Safety*, 134:178–187.

Schultz, C. R. (1987). Forecasting and Inventory Control for Sporadic Demand under Periodic Review. *The Journal of the Operational Research Society*, 38(5):453–458.

Shenstone, L. and Hyndman, R. J. (2005). Stochastic models underlying Croston ' s method for intermittent demand forecasting. *Journal of Forecasting*, (February 2005).

Sing, C. P., Chan, H. C., Love, P. E., and Leung, A. Y. (2016). Building Maintenance and Repair: Determining the Workforce Demand and Supply for a Mandatory Building-Inspection Scheme. *Journal of Performance of Constructed Facilities*, 30(2):1–8.

Spetz, J. (2017). Forecasts of the Registered Nurse Workforce in California. Technical report, Institute for Health Policy Studies & Healthforce Center at UCSF.

Spiliotis, E., Petropoulos, F., and Assimakopoulos, V. (2019a). Improving the forecasting performance of temporal hierarchies. *PLoS ONE*, 14(10):1–21.

Spiliotis, E., Petropoulos, F., Kourentzes, N., and Assimakopoulos, V. (2019b). Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. *Applied Energy*, 261(November 2019):1–23.

Springer (2008). *The Concise Encyclopedia of Statistics*. Springer, New York.

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

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

Tan, Z., Zhang, J., Wang, J., and Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy*, 87(11):3606–3610.

Venkitachalam, G. H. K., Pratt, D. B., Deyong, C. F., Morris, S., and Goldstein, M. L. (2003). Forecasting and Inventory Planning for Parts with Intermittent Demand - A Case Study. In *IIE Annual Conference Proceedings*, pages 1–5.

Viswanathan, S., Widiarta, H., and Piplani, R. (2008). Forecasting aggregate time series with intermittent subaggregate components: Top-down versus bottom-up forecasting. *IMA Journal of Management Mathematics*, 19(3):275–287.

Wagner, M. and Fricke, M. (2006). Estimation of daily unscheduled line maintenance events in civil aviation. *ICAS-Secretariat - 25th Congress of the International Council of the Aeronautical Sciences 2006*, 6:3905–3912.

Wahyudin, R. S., Sutopo, W., Hisjam, M., and Hardiono, R. S. (2016). Resource allocation model to find optimal allocation of workforce, material, and tools in an aircraft line maintenance. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 2, pages 782–787.

Wang, W. and Syntetos, A. A. (2011). Spare parts demand: Linking forecasting to equipment maintenance. *Transportation Research Part E: Logistics and Transportation Review*, 47(6):1194–1209.

Weckman, G. R., Marvel, J. H., and Shell, R. L. (2006). Decision support approach to fleet maintenance requirements in the aviation industry. *Journal of Aircraft*, 43(5):1352–1360.

Willemain, T. R., Smart, C. N., and Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, 20:375–387.

Williams, T. M. (1984). Stock Control with Sporadic and Slow-Moving Demand. *The Journal of the Operational Research Society*, 35(10):939–948.

Woolard, I., Kneebone, P., and Lee, D. (2006). Forecasting the Demand for Scarce Skills , 2001-2006. *Review Literature And Arts Of The Americas*, pages 2001–2006.

Wu, H., Liu, Y., Ding, Y., and Liu, J. (2004). Methods to reduce direct maintenance costs for commercial aircraft. *Aircraft Engineering and Aerospace Technology*, 76(1):15–18.

Xu, Q., Wang, N., and Shi, H. (2012). A Review of Croston's method for intermittent demand forecasting. In *Proceedings - 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2012*, number July, pages 1456–1460.

Yule, G. U. (1927). On a Method of Investigating Periodicities in Disturbed Series. *Philosophical Transactions of the Royal Society of London*, 226(Series A):167–298.

Zhou, B., He, D., Sun, Z., and Ng, W. (2006). Network traffic modeling and prediction with ARIMA/GARCH. In *HET-NETs' 06 Conference*, pages 1–10.

Zotteri, G. and Verganti, R. (2001). Multi-level approaches to demand management in complex environments: An analytical model. *International Journal of Production Economics*, 71(1-3):221–233.

Chapter 3

Project plan

In this chapter the project plan is presented. This thesis deliverable is already graded.

GRADED

Maintenance demand prediction of structural component groups

Bram Slangen, 4278070
Air Transport Operations

November 4, 2019

Abstract

Due to ever increasing competition in the MRO market there is a clear need for optimizing strategic capacity planning. In this research it is argued that at a strategic level one is not interested in the failure of a specific component but rather in the failure of a component with specific characteristics that influence the amount of resources required or the capacity planning in general. This novel approach towards maintenance demand prediction has not been applied before. An advanced Cox model is proposed that incorporates covariates on failure data that categorize failures in their respective groups. The resulting hazard functions offer the failure information which can be used to optimize the strategic capacity planning. Altogether it is expected that this new approach to strategic demand will open up new ways towards modeling failure data in a strategic maintenance planning context.

1 Introduction

Commercial aircraft maintenance is a growing industry that, according to IATA, comprised of $US\$76$ Bil. worldwide in 2017 (Maintenance Cost Task Force, 2018). From an airline perspective the maintenance costs represent about 11% of the total operational costs. During the last decades the aviation market has become a competitive market which forces airlines to reduce costs wherever they can. Next to the airlines evident interest in maintenance cost reduction, Maintenance Repair and Overhaul organizations (MROs) on their turn are interested in more efficient maintenance operations as the competition in the MRO market has increased tremendously as well (Phillips et al., 2009). The Above stated observations show a general need for maintenance cost reduction in aviation and this is therefore an active field of research.

Maintenance costs are generally subdivided into Direct Maintenance Costs (DMC) and Indirect Maintenance Costs (IMC) (Dupuy et al., 2011). DMC are defined by Wu as "The labor and material costs directly expended in performing maintenance on an aircraft or related equipment" (Wu et al., 2004). Inefficient use of resource capacity has an adverse effect on the DMC. Saltoglu et al. defines the IMC as downtime, which is the costs incurred by the airline due to being unable to operate the aircraft (Saltoglu et al., 2016b). To determine the downtime costs they propose a model that incorporates Labour Capacity, clearly indicating that capacity planning is also vital in reducing the IMC (Saltoglu et al., 2016a).

This research mainly considers strategic capacity planning which is a form of capacity planning that determines the appropriate level of maintenance resources, long-term workload assignment on a strategic level(Duffuaa and Alfares, 2009). The main resources under consideration are mainly spare parts, hangar space and labor force . To make optimal use of these resources strategic, or long-term capacity planning is required. Strategic capacity planning does, however, require long-term demand forecasts, which are challenging to obtain due to the stochastic nature of unscheduled maintenance. Therefore finding new methods that forecast maintenance demand is a large subject within academia and industry due to the high costs involved (Cook et al., 2012).

To scope this research we argue the following. For strategic capacity planning one is not interested in the failure of a specific component but more so in the failure of any component of which the required maintenance action has certain characteristics. This is particularly true for resources like personnel and hangar space or equipment availability.
The general objective of this research is to investigate the improving strategic capacity planning by modeling group based aggregated failure data. The groups are based on criteria that originate from aircraft maintenance practice.
In Sec. 2 the literature review is presented. This is followed by Sec. 3, Sec. 4 and Sec. 5 in which the objective, theory and experimental set up are elaborated upon respectively. Furthermore the expected results, project planning and conclusions are given in Sec. 6, 7 and 8.

2 Literature review

In this section the literature review is presented.

2.1 Strategic capacity planning

Strategic capacity planning in aircraft maintenance industry does not have a common definition in academic literature, even though it is of large importance as stated by Duffuaa and Alfares in (Duffuaa and Alfares, 2009). Without defining it explicitly Duffuaa and Alfares suggest assuming a time frame of months to years (Duffuaa and Alfares, 2009). They also give a definition for capacity planning: "Capacity planning aims to find the optimal balance between two kinds of capacity: available capacity, and required capacity". Due to strict safety regulations in the airline industry, required capacity is leading in the capacity planning process. This means that to optimize the capacity planning, future knowledge on the required capacity is crucial.

According to Heimerl and Kolish the first step in resource capacity planning is the determination of the number of resources required per skill and period (Heimerl and Kolisch, 2010). The second step is the actual assignment of people to work-packages. As stated in the introduction one of the resources that are of particular interest for MROs and airlines is personnel. Nowadays skilled personnel is hard to come by in aerospace industry due to technicians leaving airlines and MRO organisations for various reasons, retirement being the largest one, and due to young people not joining these organizations. Next to the low influx of new personnel, the training of technicians is time consuming and expensive due to strict aviation regulations (Johnson, 2018). From a capacity planning perspective this means that strategic information on required personnel is vital for an optimized capacity planning. It would therefore be interesting to predict maintenance demand based on required skill. However, to the authors knowledge there is no literature on skill based maintenance demand prediction.

Clustering of maintenance jobs at MROs is now performed mostly based on the experience of the decision maker but is a cost saver (Van Dijkhuizen and Van Harten, 1997), (Li et al., 2016),(Dinis et al., 2019). In railroad and highway maintenance industry, optimization of maintenance schedule taking into account spatial distance between failures is well represented in literature according to Peng (Peng et al., 2011). In this field the need for models that take into account spatial distance is feeded by the long travel times of maintenance crews. In aircraft industry the spatial distances between maintenance tasks are obviously smaller but locations of defect are often harder to reach, and require extensive preparation works such as scaffolding or the use of advanced equipment. A strategic capacity approach that takes into account the aforementioned is not found in literature.

2.2 Univariate reliability models

To support the maintenance demand determination, reliability models have been developed that can model the failure characteristics of systems.

Currently the vast majority of reliability models in the MRO industry are so called univariate reliability models. Univariate reliability analysis considers only one variable which is often a time variable of some sort. The most common univariate models are the Homogeneous Poisson Process (HPP), Non-homogeneous Poisson Process and the RP model. All three models assume different failure behaviors of the system or part under consideration (Basu and Rigdon, 2000).

The HPP can be used to model systems that have a constant intensity function which means that the system under consideration cannot be improving or deteriorating. Contrarily the NHPP model has a intensity function that is not constant and often the power law process is chosen as intensity function. The non-constant intensity function allows for modeling of improving and deteriorating systems. The Renewal Process is a process that assumes perfect repair after each failure. To accommodate for more accurate modeling of maintenance demand extensions of the aforementioned univariate models have been developed such as non-perfect repair.

The advantages of univariate reliability modeling is that it is relatively easy to implement but on the other hand it restricts the model to one variable which may induce oversimplification of the problem. Also, the assumptions belonging to each of the aforementioned models can become stringent and reduce the flexibility of the model.

2.3 Multi-component system reliability modeling

Aircraft are multi-component systems. From a strategic capacity planning point of view it makes sense to look at the failure characteristics of these multi-component systems instead of the components seperately as one is generally not interested in the failure of a specific component but in the failure of a component in general.

Modeling of multi-component systems is complicated due to stochastic dependencies between the components (Shi and Zeng, 2016),(Scarf, 1997),(Song et al., 2014). Shi and Zeng argue: "Interactions between these components should not be neglected, and should be taken into account in prognostics and maintenance decisions." (Shi and Zeng, 2016). In the same paper they state that models that incorporate stochastic dependency offer opportunities to optimize the maintenance policies as they can include joint maintenance of multiple components. Song et al. developed a model that incorporates competing failure processes of stochastic dependent components by implementing competing failure process analysis (Song et al., 2014). Roberts applied an adjusted NHPP model (Crow NHPP model) to model the failure of a multi component system (Roberts, 1993). Consequently he compared the Crow NHPP model with a Weibull model fitted to each component separately. He concluded that the latter fit was more accurate.

Philip Scarf offers a word of warning in (Scarf, 1997). On the one hand he acknowledges and encourages the development of advanced mathematical models that can incorporate dependencies but stresses that from an practical perspective less complex models are preferred.

Recent literature combines multi-component system analysis with condition monitoring techniques (Ge et al., 2012),(Aizpurua et al., 2017). A recent survey paper by Wang and Chen stresses that Condition Monitoring can improve the analysis of deteriorating, multi-component systems as sensors are more and more applied in complex systems (Wang and Chen, 2016).

2.4 Prognostic Health Management

Nowadays the main trend in maintenance demand prediction is towards condition monitoring which falls under the umbrella of Prognostic Health Management as defined by Pecht and Kumar (Pecht and Kumar, 2008). They identify four different PHM models categories: Statistical reliability based approach, Life cycle load based approach, State estimation based approach and feature extraction based approach. The statistical reliability based approach is a method that does not require failure specific knowledge or systems operational conditions. Life cycle load based approach takes into account the external loads that can affect the system during its entire life cycle. The state estimation method requires real-time estimates of the present state of the component or system. The complete state cannot always be observed but by using state estimation sensor data can be combined to determine the underlying behavior of the system at any point in time. Feature extraction based approach is often used as it is usually difficult or impossible to implement a physics-of-failure based approach for prediction purposes. This approach is derived directly from routinely monitored system operating data and uses statistical and learning techniques from the theory of pattern recognition.

Condition monitoring is currently a widely researched topic in structural maintenance, also in the aerospace industry (Haider, 2019),(Dragan et al., 2020). However, the lack of sensors in aircraft air frames of aircraft in operation nowadays prohibits the use of the life cycle load approach and state estimation methods and the first pilot projects in commercial aviation have been introduced only recently (Korvesis et al., 2018),(Cheung et al., 2020). The statistical reliability based approach is a valid option and builds on the aforementioned univariate methods. The feature extraction method is a promising method as no sensors are required which means that it can be implemented immediately. An evident type of data that can be included in this method is the aircraft usage information or operational data. Examples of usage data are the types of routes the aircraft has flown, take-off weight and amount of hard touch downs.

2.5 Cox Proportional Hazard Model

From the aforementioned developments it can be extracted that the current trend in aircraft maintenance demand prediction is towards multivariate, multi-component system reliability analysis. The most used multivariate reliability analysis model is the Proportion Hazard Model (PHM), also known as the Cox model which was first introduced by David Cox in 1972 (Cox, 1972). Since its introduction the original paper has been cited numerous times, primarily in medical papers (Yazdi et al., 2002),(Deng et al., 2019),(Papier et al., 2019). In the area of maintenance the Cox model and its extensions have been used less often compared to medical field but especially in the last years the model has gained in popularity (Tian and Liao, 2011).

The proportional hazard model is in the basis a regression model that uses a hazard function as dependent variable:

$$\lambda(t; \mathbf{z}) = \exp(\mathbf{z}\beta)\lambda_0(t) \tag{1}$$

In Eq. 1 $\lambda(t; \mathbf{z})$ is the resulting hazard function depending on time and the value of the covariates $\mathbf{z}$, $\lambda_0(t)$ is the baseline hazard function when $\mathbf{z} = \mathbf{0}$ and β is a vector of unknown parameters. The Cox model has the following important characteristics:

1. The model adheres to the proportionality principle
2. The baseline hazard function can be chosen arbitrarily
3. The model cannot deal with tied events
4. Right censored

The first aspect is the most important characteristic of the model. The survival functions belonging to the different samples (defined by the explanatory variables) must adhere to this principle of proportionality which means that the relative risk of an event does not change over time (Cox, 1972). The principle of proportionality can be checked using both quantitative and qualitative methods. A well known qualitative method is the Andersen method, firstly introduced by Kay (Kay, 1977) and popularized by Andersen in (Andersen and Gill, 1982). It is based on a comparison of the cumulative hazard functions between two groups. Another method is the Schoenfeld residual method, firstly introduced by Schoenfeld in (Schoenfeld, 1980). This method makes use of the so called Schoenfeld residuals which are the difference between the covariate vector and the partial mean of the estimated covariate vector at the same time instant. Finally the Kaplan-Meier estimates for the survival function of each group can also indicate whether or not the survival functions show proportionality (Kaplan and Meier, 1958).

Next to the graphical methods quantitative methods that check for proportionality are available. Commonly used methods are the Gill and Schumachers test which uses a generalized rank statistics and the Grambsch Thernau method which is a score test based on weighted Schoenfeld residuals (Gill and Schumacher, 1987),(Grambsch and Thernauw, 1994).

To model non-proportional hazards four general methods are available (Schemper, 1992),(Dunkler et al., 2009),(Schemper et al., 2009):

1. Stratification of a model by a covariate with non-proportional hazards
2. Separate models for disjunct time periods
3. Implementing time-dependent covariates
4. Implementing weighted Cox regression

Stratification of the proportional hazard function is shortly the exemption of a non-proportional covariate out of the model which basically induces k different models (Kleinbaum and Klein, 2012). The disadvantage of using stratification is that the influence of the covariate cannot be assessed immediately. Another method to cope with non-proportionality is by creating different models for disjunct time periods. This uses the fact that for some time intervals the hazards are proportional while for the total considered time it might not. The disadvantage of implementing this method is that it induces sudden changes at the cutpoint between two interval. This is often not a valid assumption.

The most advanced way of allowing non-proportionality in the Cox model is by using time-dependent covariates. This is done by including an interaction of a covariate with time which is technically the product of a value of a covariate with a pre-specified function of time: $\gamma(t)$. This option is the most flexible one, however there are consequences induced by this method. It is only useful with larger sample sizes and if a concise description of the time-dependent effect is of interest. Furthermore, it is not always possible to draw clear conclusions from such models.

The fourth option of analysis is the weighted Cox regression, first introduced by Schemper (Schemper, 1992). He argues that due to the non-proportionality the average hazard ratios are estimated in an overestimated or underestimated way. This is due to the fact that in the classic Cox model the average hazard ratio is computed giving equal weights to each separate hazard ratio, as shown in Eq. 2:

$$\theta = \int_0^\infty \frac{h_1(t)}{h_2(t)} \mathrm{d}t \tag{2}$$

This differs from the proposed weighted Cox model where a weighting function $W(t)$ is introduced as shown in Eq. 3:

$$\theta(W) = \int_0^\infty \frac{h_1(t)}{h_2(t)} \mathrm{d}W(t) \tag{3}$$

This weighting function can be defined in various ways, but the most common choice is presented in (Kalbfleisch and Prentice, 1981). Kalbfleisch and Prentice propose a weighting function that depends on the survival functions of the compared hazard functions. The method introduced by Schemper is primarily focused on the estimator of the average hazard ratio and the interpretation of it in a non-proportional case. He argues that if the amount of information and scientific interest make more detailed modeling of the time dependence of a covariate's effect possible, then a Cox model analysis using one or more time-dependent covariate terms is preferred.

The origin of the allowed arbitrariness of the baseline function lies in the parameter estimation technique that Cox proposed. He introduced the partial likelihood which he defined in his original paper (Cox, 1972). This partial likelihood function is derived from the full likelihood function. Cox states that the first part of the full likelihood contained almost all information about β while the last terms contained the information about the baseline hazard function $\lambda_0(t)$. To optimize the estimate of β Cox claimed that the regular rules of maximum likelihood apply. In medicine often the hazard ratio is used as a performance indicator of a treatment and while computing this ratio the baseline hazard function disappears.

As a result of this partial likelihood estimation there is no estimation of the baseline hazard function, which makes it impossible to compute the absolute measures of effect such as the survival probability or hazard rate. However, estimating these two measures is possible and the two most common methods are the Breslow estimator, which estimates the cumulative hazard, and the Kalbfleisch-Prentice estimator, which is an estimator for the survival function Ng et al. (2018a). These methods however give a sub-optimal estimate compared to methods that use the fully specified baseline hazard function.

2.6 Parameterized baseline hazard function models

From a prognostic point of view absolute measures of either the survival function and intensity/hazard function are important. As aforementioned the baseline hazard functions are not of the original Cox model is not estimated formally and consequently the hazard function cannot be described. Also, these hazard functions are prone to overfitting and are usually erratic of nature. To cope with these issues parameterized Cox models have been developed.

A commonly used parameterized Cox model is the Weibull proportional hazard model. This model assumes a Weibull baseline hazard function which includes a shape and scale parameter. These two parameters are estimated using a full maximum likelihood. The Weibull Cox model can model deteriorating and improving systems.

Another method that uses a parameterized baseline hazard function is the Royston-Parmar model, introduced by Royston and Parmar in (Royston and Parmar, 2002). According to an independent scoping review paper this model is a promising alternative for the Cox model, especially in the prognostic field of research (Ng et al., 2018b).

The Royston-Parmar model assumes a smoothed baseline log cumulative hazard function using natural cubic splines. Royston and Parmar chose to smooth the transformed survival function rather than the hazard function to anticipate for end effects that would be more severe for the hazard function.

2.7 On reparability

When considering maintenance on aircraft systems it is evident to include repairability or so called recurring events in the analysis. Multiple papers have been written about the subject both in the field of medicines as in the field of maintenance (Prentice et al., 1981),(Fuqing and Kumar, 2013),(Kumar, 1995).

A variation of the proportional hazard model often applied in the context of repairable systems is the proportional intensity model (PIM) (Fuqing and Kumar, 2013) and (Kumar, 1995). Analogous to the proportional hazard function the proportional intensity function is defined as follows:

$$\lambda(t, z; q) = \lambda_0(t; q)\exp(\beta^{\mathrm{T}}\mathbf{z}) \tag{4}$$

Where q is the so called repair factor which is a variable of the baseline intensity function.

A similar but stratified model is the PWP model firstly introduced by Prentice, Williams and Peterson in (Prentice et al., 1981). They introduced two extensions to the Cox model. In one the baseline function is dependent on the total time since initial startup while in the other the baseline function for stratum S is dependent on the time since last failure. For each stratum the risk set is different as the first one consists of all the subjects that are at risk to fail for the first time, the second one consists of all the subjects that are at risk to fail for the second time and so forth.

A third model that incorporates recurring events is the Andersen-Gill model (Andersen and Gill, 1982). Amorim and Cai list the assumptions of the Andersen-Gill model clearly in (Amorim and Cai, 2015). The baseline hazard function is the same for all events per subject. Furthermore, any correlation of time increments between events are conditionally uncorrelated, given the covariates. Correlations between events can be described using the appropriate covariates. In the field of maintenance a suitable covariate would be the amount of previous repairs. As a final note Amorim and Cai state: "The Andersen-Gill model is commonly applied when the interest is in the overall effect on the intensity of the occurrence of a recurrent event."

2.8 Synthesis from literature

In general there is a need for improved strategic capacity planning due to increased competition in both the aviation and MRO market. Nowadays skilled personnel is scarce and hard to come by which means that strategic information on required personnel is vital. Prediction of maintenance demand based on required skill offers an opportunity to solve this issue. Furthermore, spatial information of defects is not yet accounted for in academic literature.

The current trend leads towards the implementation of multivariate reliability analysis. The success of the Cox model in the medical field and the interest from the maintenance field of research indicates the potential of the Cox model. To the authors knowledge a multivariate approach towards strategic capacity planning using the Cox model has not been conducted before.

3 Research question, aims and objectives

Based on the literature review and the assumed perspective the main research question is the following:

'What is the strategic maintenance demand for groups of B777 structural components based on capacity planning using the proportional hazard model?'

The following sub-questions are acquired from the main question:

1. 'What are the relevant B777 structural component groups?'
   (a) 'Which factors are important to distinguish components with from a resource planning perspective?'
   (b) 'What are the consequences of grouping components from a modeling perspective?'

2. 'Which proportional hazard model is best to model the strategic maintenance demand of the B777 structural component groups?'
   (a) 'What are the relevant limitations of the standard proportional hazard model?'
   (b) 'Which extensions of the proportional hazard model are available and what are their limitations and assumptions?'
   (c) 'What limitations on the Cox model are induced by the B777 structural data?'
   (d) 'What is the best method to incorporate repairability in the Cox model?'
   (e) 'Which of the potential models gives the best fit?'

The objective distilled from the research question is the following:

'Improve strategic maintenance capacity planning by identifying maintenance demand for groups of B777 structural components based on resource planning practice by using the proportional hazard model'

This objective introduces sub-goals. The first sub-goal is to establish the time variable. In maintenance demand analysis this is often the amount of flight cycles or aircraft age. Secondly, feature engineering must take place to be able to create relevant features. The choice of features is based on the literature study on capacity planning. The next sub goal is the determination of the best model. This is done firstly based on the literature study. If several models are candidate, the best model will be chosen based on the best fit on the data. The last sub-goal is the generation of the hazard functions with which the different component groups can be compared.

The novelties of this research are listed below:

1. Apply resource centered approach to maintenance demand prediction
2. Apply novel implementation of Cox model in maintenance field

4 Theoretical Content and Methodology

Most of the models described in the literature study have not been applied in the field of maintenance prediction. Some of the proposed models are available in toolboxes, others are only available in literature. The choice of model will first be based on a theoretical study. If after this study still some models are applicable the models will be fitted to the data and based on that fit the best model will be selected.

Another step in the modeling is the feature selection. This will be primarily based on the available data in the B777 data and the practice of capacity planning (skill, spatial distance etc.). Due to the fact that some of the features are in a descriptive form some categorizing has to take place beforehand.

Due to the large amount of data a validation of the method is available using some kind of k-fold method. A scheme of the methodology is presented in Fig. 1

5 Experimental Set-up

In this section the experiment set-up is explained. Although there are no physical experiments involved of course computer models are constructed and these do need testing.

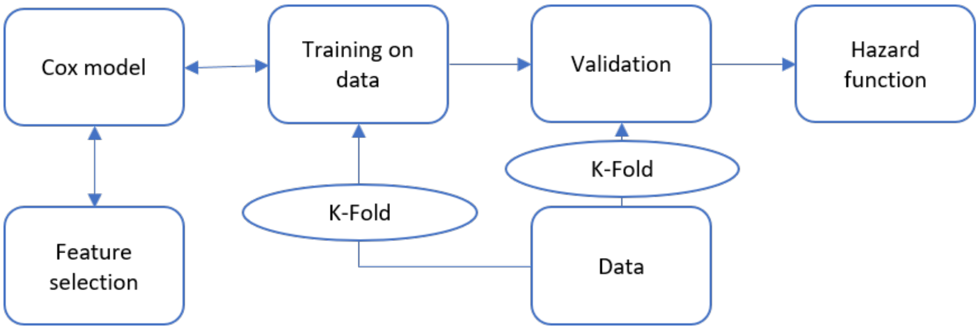

Figure 1: Schematic representation of methodology

5.1 Apparatus

As said in the beginning of this section no physical apparatus is required for this research. On the other hand a variety of software is available that offers the models described in the literature study. Python, R, Stata and SPSS all offer advanced statistics packages that can be used for the research project. However, a short study showed that not all models described in the literature study are available in all software. This means that a combination of software should probably be used.

5.2 Testing

While developing models good testing is mandatory. The tests that will be performed are listed below:

- Verification tests. These tests are there to check if the program code is working as it should be. These tests are performed during the development.
- Sensitivity tests. The sensitivity of the features is important to know. During the feature engineering phase certain choices are made that can influence the end result significantly. The sensitivity of these choices should be evaluated.

6 Results, Outcome and Relevance

The available data are Structural Damage Report (SDR) data of a varied B-777 fleet. The data consists of 89 variables and 9985 data points. Those variables contain mainly information on location of defect, moment of defect, responsible engineer, description of defect, tail number and amount of flight hours. A large part of that information is not relevant for the research objective at hand as most variables do not have any predictive value.

To identify maintenance demand for groups of B777 components, the hazard function per considered group of components is of most interest. The sketch given in Fig. 2 shows an example of how the final results of the research could look like.
The relevance of these results is multitude. Firstly the hazard functions serve as input for the amount of resources that are needed for each group of components. Based on these hazard functions MROs have more detailed information about when which skilled personnel is expected to be needed. On top of that the spatial information offers information about the potentially required equipment.

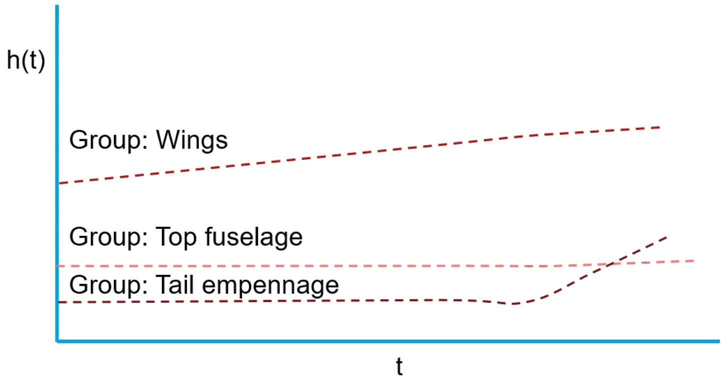

Figure 2: Sketch of hazard functions per component group.

7 Project Planning and Gantt Chart

In Fig. 3 in the appendix the Gantt chart of the thesis research is presented. The reddish colored descriptions indicate the deadlines.

8 Conclusions

Due to ever increasing competition in the MRO market there is a clear need for reducing maintenance costs. Strategic capacity planning is a vital part of the maintenance process and therefore interesting from a cost reducing perspective. Namely, by optimizing strategic capacity planning less resources are spilled which reduces the costs considerably. Resources that are of main interest in aviation industry are personnel, hangar/equipment use and spare parts.

In this research proposal it is argued that at a strategic level one is not interested in the failure of one specific component but rather in the failure of a component with specific characteristics that influence the required resources in a particular way or the capacity planning in general. For example, due to the fact that skilled personnel is hard to come by it is vital for MROs to know when those skills are needed so that they are able to plan accordingly. Also spatial information about failures is important for equipment and hanger usage and in order to cluster maintenance tasks.

In this project plan an advanced Cox model is proposed that incorporates covariates on failure data that categorize failures in their respective groups. The Cox model results in a hazard function per component group. These hazard functions offer the failure information which then can be used to optimize the strategic capacity planning.

To the authors knowledge the above described method of maintenance demand prediction from a strategic capacity planning perspective has never been applied before. Also a never before applied proportional hazard model will be used to model the hazard functions.

Altogether it is expected that this new approach to strategic demand will open up new ways towards modeling failure data from a capacity planning point of view.

References

Aizpurua, J. I., Catterson, V. M., Papadopoulos, Y., Chiacchio, F., and D'Urso, D. (2017). Supporting group maintenance through prognostics-enhanced dynamic dependability prediction. *Reliability Engineering and System Safety*, 168(April):171–188.

Amorim, L. D. and Cai, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1):324–333.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics*, 14(2):590–606.

Basu, A. and Rigdon, S. (2000). Statistical methods for the reliability of repairable systems.

Cheung, C., Sehgal, S., and Vald, J. J. (2020). A Machine Learning Approach to Load Tracking and Usage Monitoring for Legacy Fleets. 1:922–937.

Cook, A., Tanner, G., and Lawes, A. (2012). The hidden cost of airline unpunctuality. *Journal of Transport Economics and Policy*, 46(2):157–173.

Cox, B. D. R. (1972). Regression Models and Life-Tables Author (s): D . R . Cox Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 34 , No . 2 Published by : Blackwell Publishing for the Royal Statistical Society Stable URL : http://www.js. *Journal of the Royal Statistical Society*, 34(2):187–220.

Deng, Y., Zhang, Z., Jia, X., Cheng, W., Zhou, X., Liu, Y., and Wang, M. (2019). Oral bisphosphonates and incidence of cancers in patients with osteoporosis: a systematic review and meta-analysis. *Archives of Osteoporosis*, 14(1).

Dinis, D., Barbosa-Póvoa, A., and Teixeira, Â. P. (2019). Valuing data in aircraft maintenance through big data analytics: A probabilistic approach for capacity planning using Bayesian networks. *Computers and Industrial Engineering*, 128(October 2018):920–936.

Dragan, K., Dziendzikowski, M., and Kurnyta, A. (2020). Perspective of Structural Health Monitoring for Military Aviation in Poland. pages 1065–1081.

Duffuaa, S. O. and Alfares, H. (2009). *Handbook Maintenance Management*. Number January 2009.

Dunkler, D., Ploner, M., Schemper, M., and Heinze, G. (2009). Weighted cox regression using the R package coxphw. *Journal of Statistical Software*, 84(2).

Dupuy, M. J., Wesely, D. E., and Jenkins, C. S. (2011). Airline fleet maintenance: Trade-off analysis of alternate aircraft maintenance approaches. *2011 IEEE Systems and Information Engineering Design Symposium, SIEDS 2011 - Conference Proceedings*, pages 29–34.

Fuqing, Y. and Kumar, U. (2013). Proportional Intensity Model considering imperfect repair for repairable systems. *International Journal of Performability Engineering*, 9(2):163–174.

Ge, E. S., Li, Q. M., and Li, H. (2012). Condition-based maintenance for multi-component systems using proportional hazards model. *Chinese Control Conference, CCC*, 96(5):5418–5422.

Gill, R. and Schumacher, M. (1987). A Simple Test of the Proportional Hazards Assumption. *Biometrika*, 74(2):289–300.

Grambsch, P. and Thernauw, T. (1994). Proportional hazards test and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526.

Haider, S. (2019). Overview of Prognostics and Health Management for Landing Gear Maintenance. *2019 Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–7.

Heimerl, C. and Kolisch, R. (2010). Scheduling and staffing multiple projects with a multi-skilled workforce. *OR Spectrum*, 32(2):343–368.

Johnson, B. (2018). Another look at the aviation maintenance personnel shortage and the solutions.

Kalbfleisch, J. D. and Prentice, R. L. (1981). Estimation of the average hazard ratio. *Biometrika*, 68(1):105–112.

Kaplan, E. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282):457–481.

Kay, R. (1977). Proportional Hazard Regression Models and the Analysis of Censored Survival. *Journal of the Royal Statistical Society*, 26(3):227–237.

Kleinbaum, D. G. and Klein, M. (2012). Survival analysis, a self learning text. In *Survival Analysis; A Self Learning Text*, chapter 5, pages 173–256. 3 edition.

Korvesis, P., Besseau, S., and Vazirgiannis, M. (2018). Predictive maintenance in aviation: Failure prediction from post-flight reports. *Proceedings - IEEE 34th International Conference on Data Engineering, ICDE 2018*, pages 1423–1434.

Kumar, D. (1995). Proportional hazards modelling of repairable systems. *Quality and Reliability Engineering International*, 11(5):361–369.

Li, H., Zuo, H., Liang, K., Xu, J., Cai, J., and Liu, J. (2016). Optimizing combination of aircraft maintenance tasks by adaptive genetic algorithm based on cluster search. *Journal of Systems Engineering and Electronics*, 27(1):140–156.

Maintenance Cost Task Force (2018). Airline Maintenance Cost Executive Commentary Public Version. (November):14.

Ng, R., Kornas, K., Sutradhar, R., Wodchis, W. P., and Rosella, L. C. (2018a). The current application of the Royston-Parmar model for prognostic modeling in health research: a scoping review. *Diagnostic and Prognostic Research*, 2(1):1–15.

Ng, R., Kornas, K., Sutradhar, R., Wodchis, W. P., and Rosella, L. C. (2018b). The current application of the Royston-Parmar model for prognostic modeling in health research: a scoping review. *Diagnostic and Prognostic Research*, 2(1):1–15.

Papier, K., Appleby, P. N., Fensom, G. K., Knuppel, A., Perez-Cornago, A., Schmidt, J. A., Tong, T. Y., and Key, T. J. (2019). Vegetarian diets and risk of hospitalisation or death with diabetes in British adults: results from the EPIC-Oxford study. *Nutrition and Diabetes*, 9(1).

Pecht, M. and Kumar, S. (2008). Data Analysis Approach for System Reliability, Diagnostics and Prognostics. *Pan pacific microelectronics symposium*, pages 22–24.

Peng, F., Kang, S., Li, X., Ouyang, Y., Somani, K., and Acharya, D. (2011). A Heuristic Approach to the Railroad Track Maintenance Scheduling Problem. *Computer-Aided Civil and Infrastructure Engineering*, 26(2):129–145.

Phillips, P., Diston, D., Starr, A., Payne, J., and Pandya, S. (2009). A review on the optimisation of aircraft maintenance with application to landing gears. *Engineering Asset Lifecycle Management - Proceedings of the 4th World Congress on Engineering Asset Management, WCEAM 2009*, (September):68–76.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379.

Roberts, W. T. (1993). Failure predictions in repairable systems. *Science*, 29.

Royston, P. and Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15):2175–2197.

Saltoglu, R., Humaira, N., and Inalhan, G. (2016a). Aircraft Scheduled Airframe Maintenance and Downtime Integrated Cost Model. *Advances in Operations Research*, 2016.

Saltoglu, R., Humaira, N., and Inalhan, G. (2016b). Maintenance stop time influence on aircraft total maintenane cost with downtime integrated cost model. *Proceedings of 2016 7th International Conference on Mechanical and Aerospace Engineering, ICMAE 2016*, pages 502–506.

Scarf, P. A. (1997). On the application of mathematical models in maintenance. *European Journal of Operational Research*, 99(3):493–506.

Schemper, M. (1992). Cox Analysis of Survival Data with Non-Proportional Hazard Functions Author (s): Michael Schemper Source : Journal of the Royal Statistical Society . Series D (The Statistician), Vol . 41 , No . 4 Published by : Wiley for the Royal Statistical Society. *Journal of the Royal Statistical Society*, 41(4):455–465.

Schemper, M., Wakounig, S., and Heinze, G. (2009). The estimation of average hazard ratios by weighted Cox regression. *Statistics in Medicine*, (May 2009):3385–3397.

Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67(1):145–153.

Shi, H. and Zeng, J. (2016). Real-time prediction of remaining useful life and preventive opportunistic maintenance strategy for multi-component systems considering stochastic dependence. *Computers and Industrial Engineering*, 93:192–204.

Song, S., Coit, D. W., Feng, Q., and Peng, H. (2014). Reliability analysis for multi-component systems subject to multiple dependent competing failure processes. *IEEE Transactions on Reliability*, 63(1):331–345.

Tian, Z. and Liao, H. (2011). Condition based maintenance optimization for multi-component systems using proportional hazards model. *Reliability Engineering and System Safety*, 96(5):581–589.

Van Dijkhuizen, G. and Van Harten, A. (1997). Optimal clustering of frequency-constrained maintenance jobs with shared set-ups. *European Journal of Operational Research*, 99(3):552–564.

Wang, R. and Chen, N. (2016). A survey of condition-based maintenance modeling of multi-component systems. *IEEE International Conference on Industrial Engineering and Engineering Management*, 2016-December:1664–1668.

Wu, H., Liu, Y., Ding, Y., and Liu, J. (2004). Methods to reduce direct maintenance costs for commercial aircraft. *Aircraft Engineering and Aerospace Technology*, 76(1):15–18.

Yazdi, M. H., Visscher, P. M., Ducrocq, V., and Thompson, R. (2002). Heritability, reliability of genetic evaluations and response to selection in proportional hazard models. *Journal of Dairy Science*, 85(6):1563–1577.

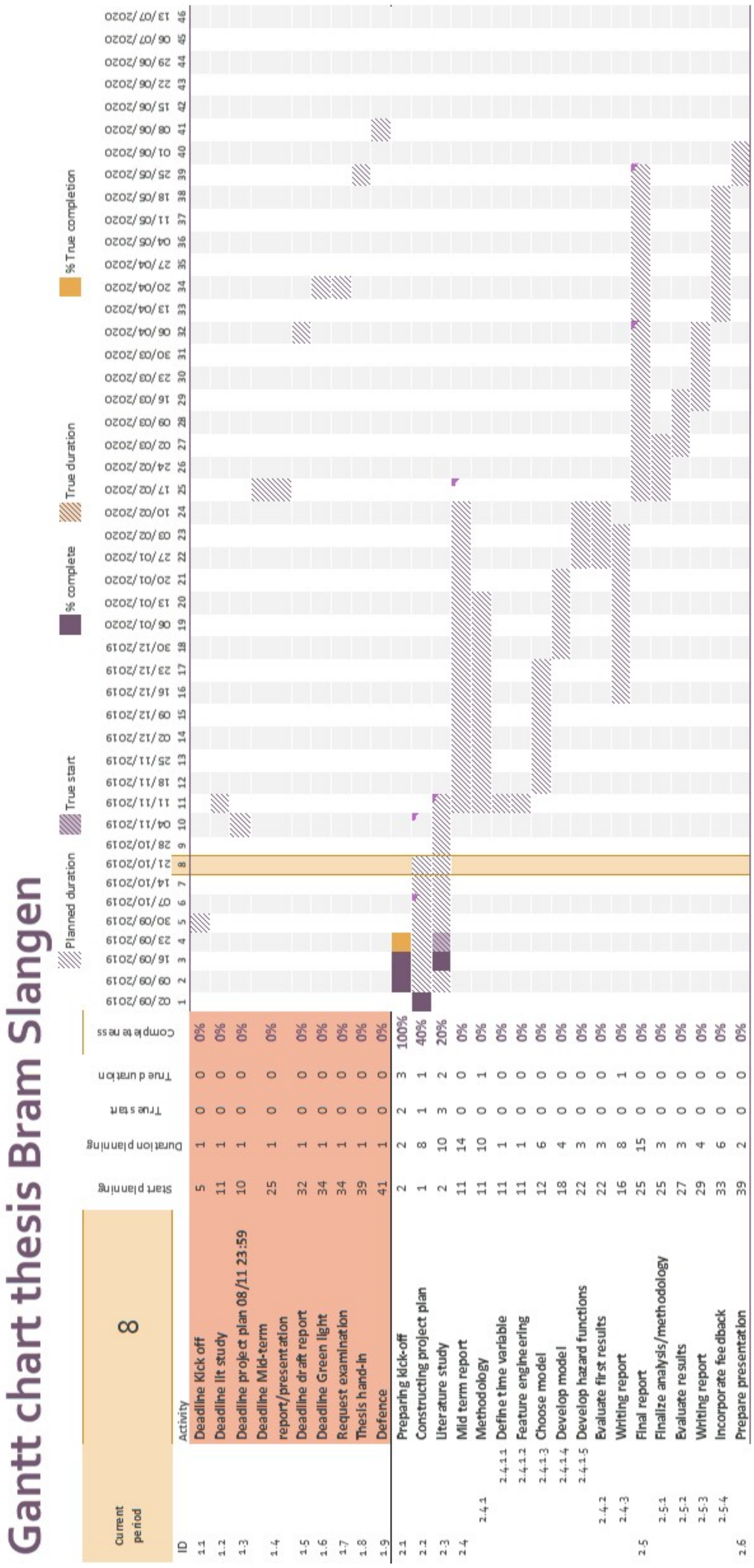

Figure 3: Thesis Gantt chart