TU Delft

Delft University of Technology

Predictive maintenance for aircraft components using proportional hazard models

Verhagen, Wim J.C.; De Boer, Lennaert W.M.

DOI
10.1016/j.jii.2018.04.004

Publication date
2018

Document Version
Accepted author manuscript

Published in
Journal of Industrial Information Integration

Citation (APA)
Verhagen, W. J. C., & De Boer, L. W. M. (2018). Predictive maintenance for aircraft components using proportional hazard models. *Journal of Industrial Information Integration*, *12*, 23-30. https://doi.org/10.1016/j.jii.2018.04.004

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Predictive Maintenance for Aircraft Components using Proportional Hazard Models

Abstract. Unscheduled maintenance can contribute significantly to an airline's cost outlay. Reliability analysis can help to identify and plan for maintenance events. Reliability analysis in industry is often limited to statistically based approaches that incorporate failure times as the primary stochastic variable, with additional strict assumptions regarding independence of events and underlying distributions of failure phenomena. This foregoes the complex nature of aircraft operations, where a whole range of operational factors may influence the probability of occurrence of a maintenance event. The aim of this research is to identify operational factors affecting component reliability and to assess whether these can be used to reduce the number of unscheduled occurrences (i.e. failures). To do so, a data-driven approach is adopted where historical operational and maintenance data is gathered and analysed to identify operational factors with a measurable influence on maintenance event occurrence. Both time-independent and time-dependent Proportional Hazard Models (PHMs), models which incorporate operational factors as covariates, are employed to generate reliability estimates. Results obtained from analysing historical data of a set of nine components with respect to unscheduled removals indicates that adopting new maintenance schedules, derived from the proposed reliability models, can reduce the number of unscheduled occurrences.

Keywords. Predictive maintenance, unscheduled maintenance, Proportional Hazard Model

1. Introduction

Costs associated with maintenance can contribute significantly to an airline's expenditure; historical estimates for maintenance cost range between 10 – 15% of the overall expenditure incurred by airlines [1]. Reliability analysis can help to identify and plan for maintenance events. Reliability analysis in industry is often limited to time-based statistical approaches, incorporating failure times as the primary stochastic variable [2]. Such approaches assume simple binary behaviour in terms of reliability: a component works or it does not. In addition, strict assumptions regarding (in)dependence of events and underlying distributions of associated failure phenomena are frequently made [3], which may be unwarranted in some cases. In relation to the aerospace domain, a major limiting factor of existing statistically-based approaches is that these forgo the complex nature of aircraft operations, where a whole range of operational factors may influence the probability of occurrence of a maintenance event [4, 5]. For instance, aircraft operating from hot, sandy airports or regions have very different conditions of use than aircraft operating from cold, wet airports, which leads to different failure modes and times for components.

The aim of this research is to improve statistical reliability assessment in aircraft maintenance by incorporating the effect of operational factors. To do so, operational

factors affecting component reliability are identified and assessed for their capability to reduce the number of unscheduled occurrences (i.e., failures). A data-driven approach is adopted where historical operational and maintenance data is gathered and analysed to identify operational factors with a measurable influence on maintenance event occurrence. The identification of these explanatory variables constitutes the primary contribution to the state of the art. Additionally, both time-independent and time-dependent Proportional Hazard Models (PHMs) are employed to generate reliability estimates, as these statistical models do have the possibility to incorporate explanatory variables as covariates.

The structure of this paper reflects this focus. In Section 2, a brief theoretical context is given. In Section 3, the modelling approach is given, including a discussion of the method to identify relevant operational factors and formulation of the reliability models used in this research. The next Section provides results for a set of selected components. Finally, some conclusions and indications for future research are presented.

2. Theoretical context

To place the contribution in its context, the state of the art in three related topics is covered. First of all, available maintenance strategies will be discussed, with specific attention towards their applicability and uptake in aircraft maintenance. This feeds into a discussion of the state of the art with respect to reliability modelling, with a particular focus on statistical approaches. Finally, relevant considerations with respect to data are discussed.

2.1. Maintenance strategies in aviation

A variety of strategies are available to guide determination, planning and execution of appropriate maintenance actions for given capital assets [6]. Tinga [7] distinguishes corrective and preventive maintenance as the two dominant options. In corrective maintenance, "parts are only replaced or repaired after they have failed", whereas preventive maintenance is conducted – as the name implies – to prevent the occurrence of failures. The benefits of corrective maintenance mainly accrue due to a full utilization of asset lifetime, but this strategy may have significant downsides with respect to system safety and availability, and is characterized by high spare parts inventory costs, high overtime labor costs, high component downtime, and low production availability [8].
On the other hand, preventive maintenance allows for optimal planning of maintenance activities to assure availability and has obvious benefits in terms of safety. However, for safety reasons, assets are usually replaced long before the end of their lifetime, which is economically inefficient.

Aviation is characterized by its stringent demands on safety. However, it is a highly competitive business environment at the same time [9]. Over time, a systematic approach towards the determination of the appropriate maintenance strategy for specific systems, subsystems, components and parts has been developed, culminating in the Maintenance Steering Group (MSG) – 1 and 2 approaches [10, 11]. These approaches advocated a bottom-up approach towards maintenance task determination, using amongst other criteria the role of hidden versus visible failures as well as safety-

critical versus non-safety-critical failures. Similar, more generalized principles were enunciated as part of the Reliability-Centered Maintenance (RCM) philosophy [12]. In the 1970s, the MSG-3 approach was adopted, which is still the leading approach for aviation stakeholders (including OEMs, operators, maintainers, and regulators) to determine the appropriate initial maintenance program for new aircraft types. Significant changes in MSG-3 versus its predecessors are a top-down approach focusing on system functionality, a shift towards direct maintenance task determination instead of maintenance strategy, and more explicit consideration of economic factors [11].

Following from these developments, preventive maintenance is currently the most applied strategy in aviation nowadays, with a significant amount of inspection tasks incorporated in maintenance programs. However, underlying this simple label of preventive maintenance, a wide range of options is available. A further specification of strategies has been devised by Tinga [13] – see Figure 1. He distinguishes between Condition-Based Maintenance (CBM), where the detection of an abnormal condition directly triggers a maintenance task, and predictive maintenance, where the optimal maintenance interval is predicted based on condition, time, usage or loads. On a similar note, Jardine, Lin and Banjevic [14] define Condition-Based Maintenance (CBM) as a policy that uses the information collected through condition monitoring as a basis for recommending maintenance actions. Unnecessary maintenance tasks are avoided by taking maintenance actions only when there is evidence of abnormal behaviour of a physical asset. Predictive maintenance is defined by Mobley [8] as encompassing 'regular monitoring of the actual mechanical condition, operating efficiency, and other indicators of the operating condition', where the indicators can be time, loads, usage hours, etc. For predictive maintenance, relevant predictive approaches are Time-Based Maintenance (TBM), Usage-Based Maintenance (UBM) and Load-Based Maintenance (LBM). UBM and LBM both require continuous monitoring of usage and loads respectively as well as a physical model-based prognostic method to pro-actively predict remaining useful life [13]. In contrast, Time-Based Maintenance (TBM) is the most straightforward approach towards prediction of failure behaviour. TBM typically employs statistical models to evaluate failure time as the single (random) variable of interest, with an underlying assumption of constant usage conditions. However, variations are available that allow for the incorporation of exogenous variables (and fluctuations therein) within TBM models. TBM and these extensions are the core focus of the discussion in Section 2.2.

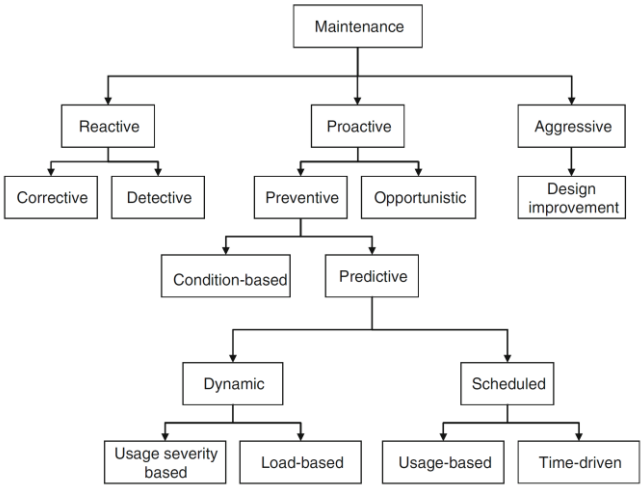


Figure 1. Classification of maintenance strategies [13]

2.2. Statistical reliability modelling approaches

Time-based reliability models use component age (time) to model reliability. Within the maintenance context, time is typically expressed as flight hours, flight cycles or calendar time. For non-repairable assets, the cumulative distribution function *F(t)* – also denoted as failure function in this context – and the probability density function *f(t)* are based on common statistical lifetime distribution functions (e.g. exponential, normal, log-normal). For more complex components such as repairables, where sequences of random variables are involved, stochastic processes are adopted to model reliability behaviour over time. Often-used models include renewal processes (RP), including Homogeneous Poisson Processes (HPP) [3], where perfect repairs are assumed (also known as 'as-good-as-new'). Conversely, Non-Homogeneous Poisson Processes (NHPP) are used to model repairables that undergo minimal repair ('as-bad-as-old'), which is particularly applicable for situations where repairable components are part of a larger, complex system. As an in-between representation, statistical models can be reformulated to include a renewal parameter [3]. This represents situations with imperfect repair, where the repaired condition of the component is better than old, but worse than new. In GRP models the i^{th} failure is formulated using the previous failure time ($(i-1)^{th}$ failure) in combination with a renewal function, derived from the renewal parameter. In effect, a component obtains a virtual age after repair: each repair sets back the clock on component age to some extent, which is determined by the effectiveness of the repair. Research has shown that Kijima type II General Renewal Processes (GRP-II) generally provide better estimates than [Non] Homogeneous Poison Processes ([N]HPPs) and Renewal Processes (RPs) [15, 16].

Common weaknesses of the conventional time-based reliability models covered above include a lack of capability to incorporate explanatory variables and a lack of representation of multiple degradation states. In other words, these conventional models assume constant conditions of use over time, which is not representative for many situations. Furthermore, the degradation behaviour of components is reduced to a simple, binary choice: either a component works, or it has failed.

In response to the first shortcoming, Proportional Hazard Models, also known as Cox models, were proposed in the 1970s to extend time-based models by introducing covariates [17]. These covariates can be utilized to represent exogeneous factors, such as operational conditions. The standard (time-based) statistical hazard function is reformulated to introduce covariates and corresponding parameters, as given in equation 1.

$$\lambda(t, Z \mid \theta, \beta) = \lambda_0(t \mid \theta) e^{\beta^T Z} \quad (1)$$

Z are the covariates corresponding to failure t, $\lambda_0(t)$ is the underlying hazard function (which does not need to be specified, but is usually estimated using parametric distributions such as the exponential, Gamma, or normal distribution), θ ($\theta_0, \theta_1, \ldots, \theta_i$) denotes the unknown parameters of the underlying distribution function, and β ($\beta_0, \beta_1 \ldots, \beta_j$) denotes the unknown parameters corresponding to each covariate. This equation can readily be reformulated to incorporate time dependent covariates $Z(t)$, at the cost of computational complexity.

PHM models have been employed succesfully in research before [18-23]; for an in-depth review of PHM, the reader is referred to the papers of Si et al. [24] and Kumar and Klefsjö [25]. The identification of covariates that may affect reliability is discussed in more detail in Kalbfleisch and Prentice [26] and Ascher and Feingold [27], respectively. Kumar and Klefsjö summarize these discussions in a succinct manner by noting that explanatory variables (i.e., covariates) "may, for instance, include operating environment (e.g. temperature, pressure, humidity or dust), operating history of a machine (e.g. overhauls, effects of repair or type of maintenance) or the type of design or material" [25]. When considering practical applications, in particular those in the aerospace domain, PHM models are typically set up using a limited number of covariates, due to constraints in availability of (monitoring) data and assumptions regarding causal influence of these covariates on failure behaviour [18, 28-30]. However, current developments in aircraft operations and maintenance – in particular with respect to increased storage and availability of sensor data to characterise operational conditions during flight – open up the possibility to revisit PHM models for a structured, automated application towards reliability estimation incorporating operational variables. This study therefore addresses an 'inverse problem': an extensive set of potential explanatory variables is available. To this end, two methods are proposed in Section 3.2 to help identify relevant operational factors.

2.3. Data considerations in reliability modelling

The theoretical forms of the GRP and PHM models assume the availability of complete data, i.e., a sample where each data point in the sample corresponds to a component failure. In real life, due to the degree of preventive maintenance being practiced in aircraft maintenance, many components will be replaced before having actually failed. Various allowances must be made for this, related to censoring, parameter estimation and goodness-of-fit testing. These are discussed in more detail below.

2.3.1. Data censoring

In statistics an observation is considered censored if its value is only partially known. For example, consider two cases: in case one, a component is operated until failure, and in case two, the component is operated until it has successfully completed

300 Flight Hours (FH). In the first case, the exact time of failure is known, however in the second it is only known that the component has survived beyond 300 FH. The aforementioned example describes Type I right censoring, a censoring type which occurs when the experiment (operation) is stopped at a predetermined time. Censoring is common in the aviation industry as component maintenance is preventively oriented for all safety-critical as well as many operationally and economically important components. The theory and application of other censoring types (e.g. left, interval, type II) is readily available, however in context to this paper only right censoring will be considered.

2.3.2. Parameter estimation

Multiple approaches exist for parameter estimation in reliability applications. Often, the Maximum Likelihood Estimation (MLE) approach is used, as it is the most accurate estimation approach for a variety of sample conditions. To account for censored data, the likelihood function L used in the computation of model coefficients can be reformulated to [31]:

$$L = \prod_{d \in D}^{n_d} f(t_d) \prod_{l \in L}^{n_l} (1 - S(t_l)) \prod_{r \in R}^{n_r} S(t_r) \prod_{i \in I}^{n_i} (S(U_i) - S(V_i)) \tag{2}$$

In equation 2, $S(t)$ refers to the survival function (also known as the reliability function), which describes the probability of survival for a given sample. Subscript d (l, r, i) denotes the d^{th} (l^{th}, r^{th}, i^{th}) observation corresponding to the group of failure truncated (left, right, interval censored) observations D (L, R, I). n_g denotes the total number of observations belonging to observation group $g = \{D, L, R, I\}$.

2.3.3. Goodness-of-fit testing

The testing of models derived from censored data has been an ongoing issue in reliability engineering [32]. Common tests such as Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling can be modified to account for right censored data. Research has shown that these modified tests depend highly on the distribution of censoring times and perform poorly when the data is truly randomly censored [33]. To address these concerns in the context of the case study application (see Section 4), the Nikulin-Rao-Robson (NRR) χ^2 test is adopted in addition to the test referred to above. This test is less subjective to variations in the censoring distribution times, however it is limited to smaller sample sizes [34].

As an additional consideration, the NNR χ^2 test statistic requires the computation of the partial derivatives of the loghazard function. As this function depends on the restoration factor q and on the covariate coefficients in time-(in)dependent PHMs it can become rather complex. Hence to compute the partial derivatives a numerical approach was adopted [35].

In general, all aforementioned tests (Kolmogorov, Cramer-von Mises-Smirnov, Anderson-Darling, and NRR χ^2) have shown to yield poor results for highly censored datasets. Despite this drawback, in cases of high censoring, the tests can still be used for comparing multiple models.

3. Modelling approach

As mentioned previously, the increased storage and availability of sensor data to characterise operational conditions during flight open up the possibility to revisit proportional hazard models to include the effect of operational factors into prediction of component reliability. The approach used to model and analyse reliability of components, including the effect of operational factors, is highlighted in Section 3.1. Subsequently, two steps in the approach are detailed further: identification of relevant operational factors (Section 3.2) and reliability modeling (Section 3.3).

3.1. General modelling and analysis approach

The general modelling and analysis approach adopted in this study is shown in Figure 2. As visualized, this approach consists of five main blocks, which have been implemented in the programming language *R*.

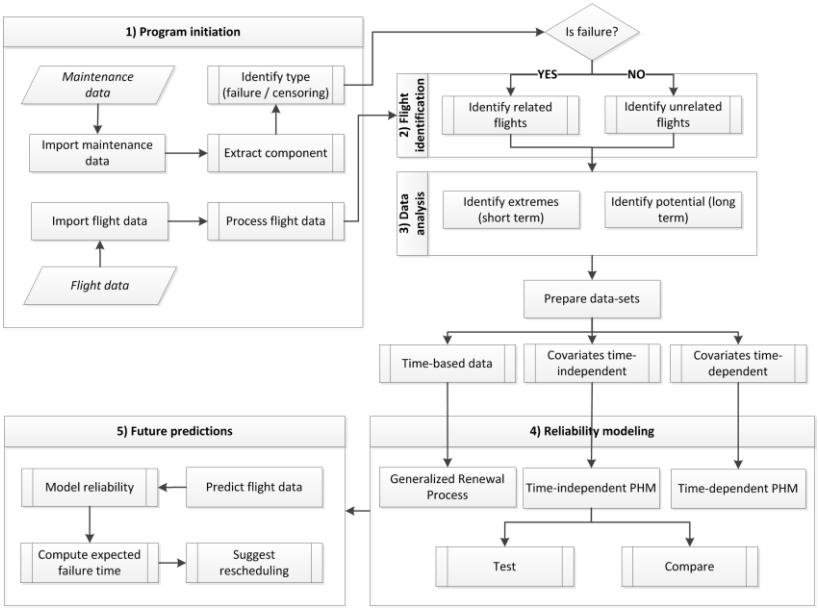


Figure 2. Modelling and analysis approach

1) **Program initiation:** to initialize the program and prepare for subsequent reliability and modelling steps, the first step comprises the import of fleet-wide maintenance and flight datasets. This is followed by the identification of component-specific data in the wider maintenance dataset, as well as the extraction of component-related flight data from the flight dataset. As a last and critical step, component-related maintenance events are extracted and characterised. These events can be of type *Failure*, in which a component has failed unexpectedly and has required unscheduled corrective maintenance, and *Censored*, in which a component has been replaced according to schedule at a specified time interval.

2) **Flight identification:** flight identification helps to address the following hypothesis: *the heavier the operational use of components, the higher the probability of component failure*. Flight identification identifies flights which *may* have had an influence on unexpected component failure. This addresses a problem in the underlying dataset used in this study, namely that the exact time of an event is not registered, only the day of occurrence. As multiple flights are conducted each day, a heuristic has been developed to identify *a set of flights* which can be associated with a particular failure event on a particular day, instead of an individual flight. In essence, the heuristic specifies that if unscheduled maintenance occurred between 24:00 and 05:30, it is associated with flights that occurred on the previous day. For any unscheduled maintenance occurring after 05:30 – the time of the first flight on a typical day of operations – it is assumed that the event can be associated with the same day of operations.
3) **Data analysis:** in step 3, two distinct approaches are used to identify operational factors of influence towards component failure: extreme value analysis and maximum difference analysis. These approaches are discussed in more detail in Section 3.2.
4) **Reliability modelling:** In step 4, a set of reliability models is applied to analyse the component dataset(s). A standard statistical approach (the Generalized Renewal Process (GRP), incorporating failure time as the single variable of interest) is employed to give baseline predictions, in accordance with current industry standards. Assuming the available datasets have succesfully given rise to identification of operational factors that may have influenced reliability behaviour over time, two variants of the Proportional Hazard Model (PHM) are employed to account for this influence. These variants are discussed in more detail in Section 3.3. In this step, it is possible to adopt simpler or more involved reliability models, e.g., parametric models other than the GRP for estimation of the baseline hazard function, or other variants of the PHM model such as the mixture Weibull proportional hazard model [23].
5) **Future predictions:** The final step in the approach concerns the generation of expected failure times using the reliability models established in the previous step. By predicting flight utilization and conditions, it is possible to estimate expected values for failure times for specific components, which can be used to adjust maintenance scheduling.

3.2. Identification of relevant operational factors

As mentioned in step 3 of the overall approach, it is critical to know which operational factors can have a measurable influence on component reliability behaviour over time. Given historical data regarding this behaviour, and operational data which can be linked to the component utilization over time, it becomes possible to identify which operational factors influence reliability. In line with the general approach, it is assumed that the identification of related flights towards a component maintenance event has been succesful, leading to a small subset of flights with potential relation to the event. To reiterate, it is then hypothesized that *the heavier the operational use of components, the higher the probability of component failure*. The Extreme Value Analysis (EVA)

and Maximum Difference Analysis (MDA) techniques use this hypothesis to identify operational factors which were abnormally high during one or more flights leading up to component failure.

3.2.1. Extreme Value Analysis (EVA)

The focus of this module is to further narrow down the number of potentially related flights and assign a flight per failure event based on the occurrence of extreme values. In general terms, this module assesses (to a certain significance level) which operational factors were abnormally high. Extreme Value Analysis (EVA) optimises one flight variable at a time, searching for optimals in both the positive and negative direction. When optimising in the positive (negative) direction, flights with observation values x below (above) the mean μ are penalised by assigning a negative p value. This increases the probability that the selected flights experienced similar extremities in the operational variables. The optimisation problem is formulated as given in equations 3 and 4.

$$\text{Maximise } z_v^D = \sum_{i \in N} \sum_{j \in M_i} f_{ij,v}^D \times p_{ij,v}^D, \qquad \forall v \in V,\ D \in \{-,+\} \tag{3}$$

Subject to

$$\begin{aligned} &\sum_{j \in M_i} f_{ij,v}^D = 1, \quad \forall i \in N, D \in \{-,+\} \\ &f_{ij,v}^D \in \{0,1\}, \quad \forall i \in N,\ j \in M_i,\ v \in V,\ D \in \{-,+\} \end{aligned} \tag{4}$$

With $f_{ij,v}^D$ being a decision variable which represents optimal flight selection for variable v in optimization direction D, where it should be noted that $f_{ij,v}^D$ is 1 if flight j corresponding to event i is the cause of failure, and is 0 if flight j corresponding to event i is not the cause of failure. Furthermore, $p_{ij,v}^D$ expresses a probability that variable v (representing an operational factor) in flight f_{ij} belongs to group C, which is the set of censored events (i.e., the events without failure). D is the optimization direction for variable v, where if D is negative (-), p values of variables v during flight f_{ij} are penalized if observed value $x_{ij,v}$ is above mean value μ_v, and if D is positive (+), p values of variables v during flight f_{ij} are penalized if observed value x_{ij} is below mean value μ_v .Furthermore, some sets are involved: N being a set of unscheduled maintenance events (i.e, failures), M_i being flights potentially related to failure event i, and V being a set of operational factors.

Finally, note that in the equations above, p is a positive value in the interval [0 1]. To specify an optimization direction D, all p values are computed such that, depending on the direction, observations $x_{ij,v}$ below (or above) μ_v are penalized. Hence,

$$p_{ij,v}^{D} = \begin{cases} D \times (1 - 2P(z > \left| \frac{x_{ij,v} - \mu_v}{\sqrt{\frac{\sigma_v^2}{n}}} \right|)) & \text{if } x_{ij,v} \geq \mu_v \\ -D \times (1 - 2P(z > \left| \frac{x_{ij,v} - \mu_v}{\sqrt{\frac{\sigma_v^2}{n}}} \right|) & \text{if } x_{ij,v} < \mu_v \end{cases} \quad (4)$$

3.2.2. Maximum Difference Analysis (MDA)

The maximum difference module is important for time-independent PHM models, which focus on mean values during a component's fail cycle (see Section 3.3). Its application is straightforward:

1. Compute mean (per operational factor) of all flights related to failure events (Group F).
2. Extract mean and standard deviation (per operational factor) of all flights related to censored events (Group C).
3. Compute probability (per operational factor) of F belonging to C using Z-test (large population size and known standard error).
4. Extract operational factors that are least likely to belong to Group C.

Successful execution of EVA and MDA produces a selection of flights associated with failure events along with a reduced list of operational factors that are likely to be the root cause of failures. Examples are given in Section 4.

3.3. Reliability modeling approach

The failure events and associated operational factors constitute essential input for the reliability models, as described next. In total, three distinct models are employed to estimate reliability:

1) **Generalized Renewal Process (GRP):** a GRP type II model is formulated to serve as a baseline estimate using failure and censor times only. No operational factors are included into this model formulation. GRP-II models employ the concept of virtual age. Various underlying distributions have been tested and assessed for goodness-of-fit, including the normal, log-normal, logistic, gamma, exponential and Weibull distributions. This is justified as multiple components with multiple failure modes have been analysed. For parameter estimation, maximum likelihood estimation (MLE) has been employed. The MLE routines have been adjusted to take into account censored data, as noted in section 2.3, and multiple serial numbers per governing part number. To maximize the likelihood function, numerical algorithms have been employed as a closed-form solution to the likelihood

function was not available. In particular, the Nelder-Mead and BFGS algorithms have been used [36].

2) **Time-independent Proportional Hazard Model (PHM):** a time-independent PHM has been employed according to the formulation as given in equation 1. Again, underlying distributions for the hazard function include the normal, log-normal, logistic, gamma, exponential and Weibull distributions. In essence, the GRP time-based reliability models described above are extended by introducing time-independent covariates. These covariates – represented as mean values over one flight - are taken from the MDA analysis, but are limited in number using forward selection to keep the standard error (and associated confidence intervals) within reasonable bounds. MLE is used again to perform parameter estimation.
3) **Time-dependent Proportional Hazard model (PHM):** Equation 1 is adjusted to take into account operational factors which vary over time; *Z* becomes *Z(t)*. EVA analysis yields operational covariates, with values that can vary as a function of time. In time-dependent models the hazard rate for all flights related to a maintenance event is computed. Each observation (flight) is subject to some error. It follows that the error of the computed reliability increases cumulatively. To limit the total error and omputational time, a forward selection approach has been implemented with a maximum of two iterations (two covariates).

4. Results

The method outlined in Section 3 is applicable to any component for which failure times, censored event and operational data is available. In this study, maintenance and operations data has been collected from an airline operator in the Asia-Pacific region. The effective range of the dataset spans a period from 2011-2015: the maintenance data extends over the period 2005-2015, whereas the operational data covers the period from 2011-2015.

By comparison of unscheduled removal rates with an industry benchmark, nine components with under-average performance have been selected for further investigation. This is described in more detail in Section 4.1. In Section 4.2, results for a single component are highlighted, giving a representative example for application of the approach highlighted in Section 3. In Section 4.3, validation with respect to the full set of nine components is performed.

4.1. Input data

Results have been derived for nine components. The basis for selection is the observed unscheduled removal rate (URR) in the fleet of aircraft contained in the dataset. This observed URR has been compared to the world wide fleet average, which is tracked and shared by the manufacturer. The nine components with the largest unfavourable deviation from the world wide fleet average have been selected for further evaluation. In other words, these components have higher-than-expected unscheduled removals, when compared to an industry benchmark. The components are given in Table 1, along with descriptive statistics showing the number of maintenance events, the related amount of flights and the percentage of censoring per component over the range of the

dataset. The source of information for this data is the maintenance dataset, which spans from 2005 to 2015.

Table 1. Components for analysis, with associated dataset characteristics

Component	# maintenance events	# flights	% censoring	# failures
697071003: Blade assembly and bearing	1597	548353	75.2	396
174260-08: Crew oxygen mask	373	120599	60.5	147
1152106-3: DC starter generator	1618	314916	78.9	342
903-1342: Hand microphone	199	52862	5.5	188
3-1573-1: MLG wheel & tire assembly	18809	776321	93.9	1132
3-1574: NLG wheel & tire assembly	19504	394788	94.9	989
728809-1: Thermal actuator	732	248162	36.0	468
10-105-31A-N-2: VHF antenna	253	68824	5.5	239
EVR716-11-0350A: VHF transceiver	368	96496	23.0	283

It is clear from Table 1 that there are considerable differences from component to component. The number of maintenance events, related flights and degree of censoring vary considerably. This can be associated with the relevant maintenance strategy: some of the featured components are both safety-critical and highly expensive (e.g. the wheel and tire assemblies for main and nose landing gear). Safety considerations have led to a strict preventive maintenance policy with conservative replacement intervals, leading to a very high number and contribution of censored events (i.e., events where replacements have been carried out according to schedule, so without the component having actually failed). For non-safety critical and less expensive components such as the hand microphone, corrective maintenance is the strategy of choice. In these cases, replacements were typically only carried out after unscheduled failures, leading to a very low amount of events as well as overall degree of censoring.

After matching the maintenance data with the available operational data (leading to a consolidated dataset for the period 2011-2015), the dataset characteristics as given in Table 2 are used for subsequent analysis.

Table 2. Components for analysis, filtered by registered events in MIS and FDR

Component	# maintenance events	% censoring	# failures
697071003: Blade assembly and bearing	282	77.3	64
174260-08: Crew oxygen mask	62	40.3	37
1152106-3:	273	77.7	61

DC starter generator			
903-1342: Hand microphone	33	6.1	31
3-1573-1: MLG wheel & tire assembly	3082	93.8	191
3-1574: NLG wheel & tire assembly	3092	95.05	153
728809-1: Thermal actuator	135	36.3	86
10-105-31A-N-2: VHF antenna	39	5.1	37
EVR716-11-0350A: VHF transceiver	60	25	45

Given the extent of the datasets and the available range of components, the next subsection focuses on a single representative example: the blade assembly and bearing component. This allows for a concise representation of application results for the reliability analysis methods detailed in Section 3.

4.2. Blade assembly and bearing output

Table 3 gives some key characteristics of the dataset for one particular component (blade assembly and bearing), as well as output of the flight identification, EVA and MDA modules. In short, 1597 unique components (being blade assembly and bearing) have been considered since 2004. Together, these components flew on roughly 550.000 flights. During the 2011-2015 period, 1531 operational variables were captured and stored, representing output from the full range of aircraft sensors, instrumentation and monitoring systems. Using the EVA analysis, 38 time-variant operational variables have been identified that have a measurable influence on reliability behaviour over time. In other words, these variables are represented using values that vary as a function of time, i.e., during flight. Using MDA analysis, 78 time-invariant variables have been identified that have a measurable influence on reliability behaviour over time. In other words, these variables are represented using a single, constant value per flight.

Table 3. Key dataset characteristics for blade assembly and bearing example

Dataset attribute	**Value**
Number of components	1597
Number of flights (total)	548353
Number of operational variables	1531
EVA output (relevant operational variables)	38
MDA output (relevant operational variables)	78

Figure 3 shows a visual example of the top operational factor influencing failures of blade assembly and bearing, following from MDA analysis.A significant difference can be observed between the probability density functions of the mean ambient pressure operating on failure event associated flights (as given in red) and the censored event associated flights (as given in blue). For EVA analysis, the longitudinal acceleration (Accn_long_mean) is more severe for failure-associated flights when compared to flights associated with censored events (i.e., no failures).

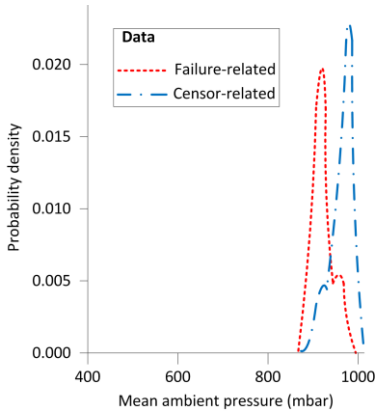


Figure 3. Top operational factor as identified by MDA

In terms of effectiveness of time-based reliability models versus PHM variants, Table 2 shows MLE output, computational time and goodness-of-fit characteristics (according to the NRR χ^2 test) for an underlying exponential distribution and varying number of operational factors. For this particular component, the underlying exponential distribution had the best goodness-of-fit. The four operational factors taken into account were the mean ambient pressure, longitudinal acceleration, maximum roll rate, and the commanded rudder force.

A few observations can be made. Firstly, the MLE estimator value is the best for time-dependent PHM, outperforming the other two model types. However, goodness-of-fit is best for time-independent PHM, with the additional observation that incorporation of additional operational factors increases accuracy of forecasts. However, this comes with a computational cost, as evidenced in the time taken for analysis. Especially for time-dependent PHM models, where the input matrices are considerable in size, the computational time is substantial.

Table 2. Overview of GRP-II, time-independent and time-dependent PHM model results (MLE estimates; NRR goodness-of-fit test; computational time) for underlying exponential distribution

	GRP-II	**Time-independent PHM**				**Time-dependent PHM**			
# oper. Factors	N/A	1	2	3	4	1	2	3	4
MLE	-643.6	-632.3	-624.5	-618.6	-614.7	-582.3	-502.0	-425.1	-387.1
NRR	44.45	37.32	32.43	35.41	27.21	104.25	76.37	46.38	60.14
time [min]	<< 1	6.35	8.7	11.34	13.51	46.1	89.7	152.2	217.9

4.3. Validation

For other components and other underlying distributions, the findings represented in the previous section will naturally vary. In general however, either time-independent or time-dependent PHM models will outperform time-based models, to various levels of accuracy. As mentioned, for the given example, the model accuracy increases when more operational factors are included into the analysis. While this is true in general, the optimum number of operational factors will vary from component to component, and typically lies between two to five.

Analysis of the top 9 components has indicated that a multitude of improvements can be made to current reliability practices to reduce the unscheduled removal rate

(URR). During time-based modelling it became evident that a majority of components was better represented using a normal, log-normal, logistic, Weibull, and (or) gamma distribution opposed to the standard exponential distribution used in current practice at the operator involved in this research. Further investigation showed that more complex reliability models, including covariates, could effectively reduce the number of failures by 10 to 90% without accumulating any additional costs. This can be achieved by adjusting scheduled maintenance event intervals, and by using predicted values for covariates for a future time window of operations to assess the probability of failure over the considered time. In doing so, it may become evident that addition of scheduled maintenance may prevent unscheduled failures from occurring. This logic is represented and generalized over the nine studied components in Figure 4, which shows failure characteristics of the sample per component if time-dependent PHM models (with varying number of operational factors per component) would have been adopted to predict component reliability and adjust maintenance schedules accordingly.

To interpret this figure, there are three necessary points:

1) In real life operations, the operator involved in this study did not predict any failures related to these nine components. As such, the baseline historical 'prediction' for failures has been zero for all components considered. This is not explicitly shown in Figure 4, but highly relevant for interpretation.
2) Figure 4 shows the predicted failures. This is output from the application of time-dependent PHM models (with varying number of operational factors per component) towards the nine components being studied. Note that these predictions have been made by imputing average operational conditions for the period 2011-2015. I.e., historical averages for operational parameters were adopted, which reduces the accuracy of the prediction. This is a relevant but basic strategy to apply when forecasting failures for a future timeframe, even though more informed strategies can be considered. For instance, if one knows which routes a particular aircraft (and by extension, its systems and components) will fly in a future time period, average operational parameter values *for this particular set of routes* can be adopted to increase prediction accuracy.
3) Figure 4 next shows the historical failures. These are the actual number of failures that occurred in the period 2011-2015. The deviations between these numbers and the predictions from the time-dependent PHM models can be explained by two factors: the use of historical averages for operational parameters (as explained previously), plus the fact that only operational parameters with a significant effect on reliability have been adopted as covariates. Inaccuracy in MDA and EVA as well as exclusion of statistically insignificant covariates explain part of the deviation in accuracy.

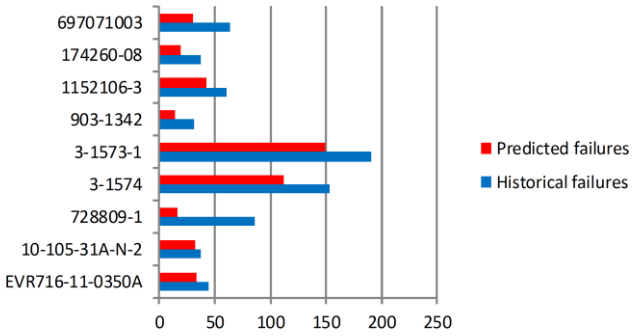


Figure 4. Number of failures per component (historical observed failures versus model predictions)

It must be noted that this is an in-sample validation effort; the model has been developed *and* tested using the dataset from 2011-2015. Separation of the dataset into separate samples for development and testing has not proven to be feasible in the scope of this research, as the resulting sample sizes would have become too small. In real-life operations, unexpected operational conditions may be encountered at any stage, which influences the accuracy of model predictions negatively. As such, a major initiative for follow-on research is to pursue out-of-sample validation (i.e., generating predictions for a separate time window, and using a dataset from that time window to validate how close the predictions were to actual outcomes).

5. Conclusions

Results derived from analysing and modelling the nine components covered in this research show that it is feasible to identify operational factors that have a significant influence on failure probability. The subsequent use of statistical models incorporating operational covariates (the time-dependent and time-independent versions of the Proportional Hazard Model) are suitable to incorporate the effects of these identified operational factors into reliability estimation. Results show that these models tend to outperform time-based models in terms of accuracy.

Limitations to this study are as follows. First, it is difficult to make an a priori assessment of which underlying distribution and which number of operational factors should be included into relability analysis. Analysis of historical data should be executed regularly to verify any choices with respect to these model settings. Furthermore, for some components, there are indications that multiple failure modes may be involved in generating reliability behaviour over time. This must be investigated in more detail to allow for more accurate reliability analysis, i.e., per component failure mode. Finally, any reliability forecasts should be validated using a separate set of maintenance event data to be able to quantify the efficacy of the proposed method in terms of failures prevented and costs saved.

References

[1] IATA, Airline Maintenance Cost Executive Commentary - Exclusive Benchmark Analysis (FY2014 data), 2014.
[2] D.M. Louit, R. Pascual, A.K.S. Jardine, A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data, Reliability Engineering & System Safety, 94 (2009) 1618-1628.
[3] S.E. Rigdon, A.P. Basu, Statistical methods for the reliability of repairable systems, Wiley, New York, 2000.
[4] Z. Li, J. Guo, R. Zhou, Maintenance scheduling optimization based on reliability and prognostics information, 2016 Annual Reliability and Maintainability Symposium (RAMS), 2016, pp. 1-5.
[5] B. Pogačnik, J. Duhovnik, J. Tavčar, Aircraft fault forecasting at maintenance service on the basis of historic data and aircraft parameters, Eksploatacja i Niezawodnosc – Maintenance and Reliability, 19 (2017) 624–633.
[6] S. Premaratne, K. Senevi, Aircraft maintenance planning and scheduling: an integrated framework, Journal of Quality in Maintenance Engineering, 18 (2012) 432-453.
[7] T. Tinga, Application of physical failure models to enable usage and load based maintenance, Reliability Engineering & System Safety, 95 (2010) 1061-1075.
[8] R.K. Mobley, An Introduction to Predictive Maintenance, Butterworth-Heinemann2002.
[9] J. Kilpi, J. Töyli, A. Vepsäläinen, Cooperative strategies for the availability service of repairable aircraft components, International Journal of Production Economics, 117 (2009) 360-370.
[10] T.D. Matteson, Airline experience with reliability-centered maintenance, Nuclear Engineering and Design, 89 (1985) 385-390.
[11] H.A. Kinnison, T. Siddiqui, Aviation maintenance management, McGraw-Hill2012.
[12] M. Rausand, Reliability centered maintenance, Reliability Engineering & System Safety, 60 (1998) 121-132.
[13] T. Tinga, Principles of Loads and Failure Mechanisms, Springer-Verlag, London, 2013.
[14] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing, 20 (2006) 1483-1510.
[15] M. Kijima, Some results for repairable systems with general repair, Journal of Applied Probability, 26 (2016) 89-102.
[16] H. Pham, System Software Reliability, Springer2006.
[17] D.R. Cox, Regression Models and Life-Tables, in: S. Kotz, N.L. Johnson (Eds.) Breakthroughs in Statistics: Methodology and Distribution, Springer New York, New York, NY, 1992, pp. 527-541.
[18] A.K.S. Jardine, J.A. Buzacott, Equipment reliability and maintenance, European Journal of Operational Research, 19 (1985) 285-296.
[19] J. Arasan, S. Ehsani, Modeling repairable system failures with repair effect and time dependent covariates, IAENG International Journal of Applied Mathematics, 41 (2011) 218-222.
[20] A. Barabadi, J. Barabady, T. Markeset, Application of reliability models with covariates in spare part prediction and optimization – A case study, Reliability Engineering & System Safety, 123 (2014) 1-7.
[21] K. Barker, H. Baroud, Proportional hazards models of infrastructure system recovery, Reliability Engineering & System Safety, 124 (2014) 201-206.
[22] L. Jafari, F. Naderkhani, V. Makis, Joint optimization of maintenance policy and inspection interval for a multi-unit series system using proportional hazards model, Journal of the Operational Research Society, 69 (2018) 36-48.
[23] Q. Zhang, C. Hua, G. Xu, A mixture Weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data, Mechanical Systems and Signal Processing, 43 (2014) 103-112.
[24] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, Remaining useful life estimation – A review on the statistical data driven approaches, European Journal of Operational Research, 213 (2011) 1-14.
[25] D. Kumar, B. Klefsjö, Proportional hazards model: a review, Reliability Engineering & System Safety, 44 (1994) 177-188.
[26] J.D. Kalbfleisch, R.L. Prentice, The statistical analysis of failure time data, John Wiley & Sons2011.
[27] H. Ascher, H. Feingold, Repairable systems reliability: modeling, inference, misconceptions and their causes, New York, 1984.
[28] H.F. Zuo, H.-J. Zhang, X. Rong, Condition Based Aero-Engine Maintenance Decision Method Using Proportional Hazards Model, Hangkong Dongli Xuebao, 21 (2006) 716-721.
[29] H.-J. Shyur, A quantitative model for aviation safety risk assessment, Comput. Ind. Eng., 54 (2008) 34-44.
[30] W. Wang, W. Zhang, A model to predict the residual life of aircraft engines based upon oil analysis data, Naval Research Logistics (NRL), 52 (2005) 276-284.

[31] N. Limnios, D.C. Ionescu, Statistical and probabilistic models in reliability, Springer Science & Business Media2012.
[32] R.B. D'Agostino, Goodness-of-fit-techniques, CRC press1986.
[33] M. Nikulin, B. Lemeshko, E. Chimitova, A. Tsivinskaya, Nonparametric goodness-of-fit tests for censored data, Proc. of the 7th Int. Conf.«Mathematical Methods in Reliability: Theory. Methods. Applications». Beijing, China, 2011, pp. 817-823.
[34] N. Balakrishnan, E. Chimitova, M. Vedernikova, An Empirical Analysis of Some Nonparametric Goodness-of-Fit Tests for Censored Data, Communications in Statistics - Simulation and Computation, 44 (2015) 1101-1115.
[35] J. Nocedal, S.J. Wright, Sequential quadratic programming, Springer2006.
[36] J.C. Nash, On best practice optimization methods in R, Journal of Statistical Software, 60 (2014) 1-14.