Component-Based Data-Driven Predictive Maintenance to Reduce Unscheduled Maintenance Events

Verhagen, Wim J.C.; Curran, Richard; de Boer, L.W.M.

DOI
10.3233/978-1-61499-779-5-3

Publication date
2017

Document Version
Final published version

Published in
Transdisciplinary Engineering: A Paradigm Shift - Proceedings of the 24th ISPE Inc. International Conference on Transdisciplinary Engineering, TE 2017

Transdisciplinary Engineering: A Paradigm Shift
C. Chen et al. (Eds.)

doi:10.3233/978-1-61499-779-5-3

Component-Based Data-Driven Predictive Maintenance to Reduce Unscheduled Maintenance Events

Wim J.C. VERHAGEN[1], Lennaert W.M. DE BOER and Richard CURRAN
Faculty of Aerospace Engineering, Technical University of Delft, The Netherlands

Abstract. Costs associated with unscheduled and preventive maintenance can contribute significantly to an airline's expenditure. 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 ten components with respect to unscheduled removals indicates that adopting new maintenance schedules, derived from the proposed reliability models, could reduce the number of unscheduled occurrences by approximately 37%. The potential benefits of adopting the proposed strategy are extensive. Nonetheless, numerous assumptions have been introduced to overcome challenges imposed by the complex nature of the data. To overcome these challenges, recommendations are made for future development of the proposed approach.

Keywords. Predictive maintenance, unscheduled maintenance, Proportional Hazard Model

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 statistically based approaches that incorporate 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

[1] Corresponding Author, Mail: w.j.c.verhagen@tudelft.nl

(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. 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 1, a brief theoretical context is given. In Section 2, 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.

1. Theoretical context

Time-based reliability models use component age (time) to model reliability. In its simplest form its lifetime distribution function *F(t)* and probability density function *f(t)* are based on common statistical distribution functions (e.g. exponential, normal, log-normal). For more complex components where sequences of random variables are involved, such as repairables, statistical models can be reformulated to include a renewal parameter [3]. Research has shown that type II General Renewal Processes (GRP-II) generally provide better estimates than [Non] Homogeneous Poison Processes ([N]HPPs) and Renewal Processes (RPs) [4]. In GRP models the ith failure is formulated using the previous failure time ($(i-1)^{th}$ failure) and with a renewal function, derived from the renewal parameter. Common weaknesses of time-based reliability models include a lack of capability to incorporate explanatory variables and a lack of representation of multiple degradation states.

Proportional Hazard Models, also known as Cox models, extend time-based models by introducing covariates [5]. The standard (time-based) statistical hazard function is reformulated to introduce covariates and corresponding parameters, as given in equation 1.

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

Z are the covariates corresponding to failure t, $\lambda_0(t)$ is the underlying hazard function (e.g. 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, for instance in aerospace domain applications [6,7], but to a limited degree in practice. 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 these models for a structured, automated application towards reliability estimation incorporating operational variables.

2. Modelling approach

The approach used to model and analyse reliability of components, including the effect of operational factors, is highlighted in Section 2.1. Subsequently, two steps in the approach are detailed further: identification of relevant operational factors (Section 2.2) and reliability modeling (Section 2.3).

2.1. General modelling and analysis approach

The general modelling and analysis approach adopted in this study is shown in Figure 1. As visualized, the modelling approach consists of five main blocks:

1) **Program initiation:** this step addresses importing fleet-wide maintenance and flight datasets, the identification of component-specific data in the wider maintenance dataset, the extraction of component-related flight data from the flight dataset, and, as a last and critical step, extraction and characterisation of component-related maintenance events. 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. 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.
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 2.2.
4) **Reliability modelling:** In step 4, a set of reliability models is applied to analyse the component dataset(s). A standard statistical approach (incorporating failure time as the single variable of interest) is employed to give baseline predictions, in accordance with current industry standards. In addition, two variants of the Proportional Hazard Model (PHM) are employed to account for the influence of operational factors, as identified in step 3. These variants are discussed in more detail in Section 3.3.

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.

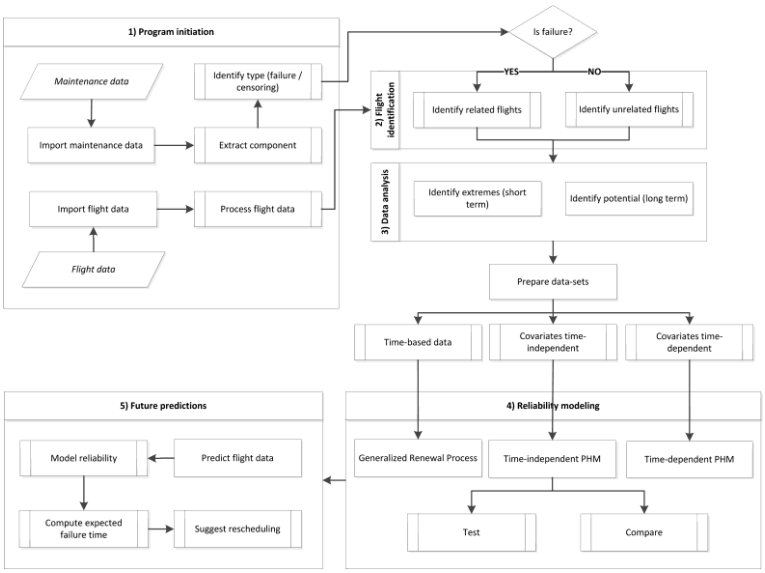

Figure 1. Modelling and analysis approach.

2.2. 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.

2.2.1. Extreme Value Analysis (EVA)

The focus of this module is to further narrow down the number of potentially related flights and assign one 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 were 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 2 and 3.

$$\text{Maximise } z_v^D = \sum_{i \in N} \sum_{j \in M_i} f_{ij,v}^D \times p_{ij,v}^D, \qquad \forall v \in V,\ D \in \{-,+\} \tag{2}$$

Subject to

$$\begin{aligned} &\sum_{j \in M_i} f_{ij,v}^D = 1, \quad \forall i \in N, D \in \{-,+\} \\ &f_{ij,v}^D \in \{0,1\}, \quad \forall i \in N,\ j \in M_i,\ v \in V,\ D \in \{-,+\} \end{aligned} \tag{3}$$

With $f_{ij,v}^D$ being a decision variable which represent 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,

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

2.2.2. 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 2.3). 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 3.

2.3. 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-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 considered. For parameter estimation, maximum likelihood estimation (MLE) has been employed. The MLE routines have been adjusted to take into account censored data 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 [8].
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. Ergo, 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).

3. Results

The method outlined in Section 2 is applicable to any component. In this study, results are derived from ten components with the highest impact in terms of unscheduled removal rate. Maintenance data with respect to these components was collected and spanned a period from 2004 – 2015. In addition, operational data was collected, spanning a period from 2011 – 2015. Table 1 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.

Table 1. 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 (total)	1531
EVA output (relevant operational variables)	38
MDA output (relevant operational variables)	78

Figure 2 shows a visual example of the top operational factors influencing failures of blade assembly and bearing, following from MDA and EVA analysis. In the case of MDA analysis, a significant difference can be observed between the mean ambient pressure operating on failure event associated flights (as given in red) and the censored event associated flights (as given in blue). Similarly, 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).

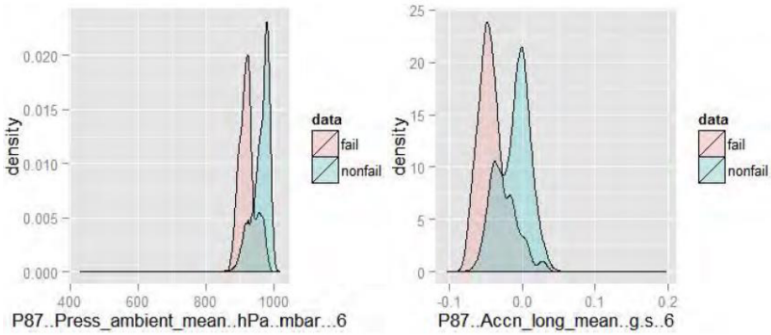

Figure 2. Top operational factor as identified by MDA (left) and EVA (right)

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 test) for a subset of underlying distributions and varying number of operational factors. A few observations can be made. Firstly, the MLE estimator value is most optimal 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. For other components and other underlying distributions, these findings will vary, but in general, 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. This is however not a generalizable statement: the optimum number of operational factors will vary from component to component, and typically lies between two to five.

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. Conclusions

Results derived from analysing and modelling the top 10 components, in terms of URRs, 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, 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, 2014, *Airline Maintenance Cost Executive Commentary - Exclusive Benchmark Analysis (FY2014 data)*, Accessed 20-01-2017. [Online]. Available: https://www.iata.org/whatwedo/workgroups/Documents/MCTF/AMC-Exec-Comment-FY14.pdf

[2] D.M. Louit, R. Pascual and 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*, Vol. 94(10), 2009, pp. 1618-1628. doi: http://dx.doi.org/10.1016/j.ress.2009.04.001

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

[4] M. Kijima and U. Sumita, A useful generalization of renewal theory: counting processes governed by non-negative markovian increments, *Journal of Applied Probability*, Vol. 23(1), 1986, pp. 71-88.

[5] D.R. Cox, Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, Series B*, Vol. 34, 1972, pp. 187-220.

[6] A.K.S. Jardine and P.M. Andersen, Use of concomittant variables for reliability estimation, In: *Proceedings of the 8th symposium on Advances in Reliability Technology*, Bradford University, 1984.

[7] A.K.S. Jardine and J.A. Buzacott, Equipment reliability and maintenance, *European Journal of Operational Research*, Vol. 19(3), 1985, pp. 285-296.

[8] J.C. Nash, On best practice optimization methods in R, *Journal of Statistical Software*, Vol. 60(2), 2014, doi:http://dx.doi.org/10.18637/jss.v060.i02.