Fault prediction in aircraft tires using Bayesian Networks

Isabela Xavier Castilho
isabela.castilho@ist.utl.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2015

Abstract

Aircraft tire condition and maintenance is of utmost importance since it is a part that withstands a great variety of operational conditions during the different stages of a flight. When a tire displays wear or damage signs that might compromise the aircraft's safety, it is removed and replaced by a repaired tire. However, in order to avoid delays in the operation, it is critical that a replacement tire is readily available for installation, and that relies on the management of the resources of different maintenance sections. The management decisions of these sections are often challenged by unexpected tire failures, and the task schedules are repeatedly modified, resulting in an inefficient use of the available resources. To improve this situation, the present thesis proposes a tire failure prediction tool based on Bayesian networks. This tool outputs the amount of tires of each size that is predicted to fail in a specified three day, week of month period. The analysis is divided into two parts: firstly, an analysis of relevance using ANOVA is performed to investigate which variables impact the most the number of cycles a tire performs between failures; and secondly, a Bayesian network regression model is selected and used to perform the predictions of interest.

Keywords: Bayesian networks, ANOVA, fault prediction, aircraft tire maintenance

1. Introduction

1.1. Context and Motivation

With the increase of the demand for flights over the years, multiple transformations have been observed in the aviation industry. However, since growth of facilities and resources is not always possible, more effort has been put into making any existing facilities and processes more efficient.

At the maintenance level, there are two main challenges: installing and removing parts faster, and having replacement parts ready to be installed in as little time as possible. In this work, the second challenge will be addressed specifically for aircraft wheels.

For each wheel model used in a fleet, there are several spare wheel available in storage. However, these are not necessarily ready for installation. In most cases the wheel needs to be assembled, and this process takes up to a day of work. After the assembly, the tire cannot be in service for 24 hours when it is subject to a pressure test. It becomes evident that if a tire needs to be replaced between flights, not having an assembled wheel ready to install may result in the delay of the flight and direct and indirect increased costs for the company. For this reason, estimating the tires that will fail and need to be replaced in a certain period would allow the workshop resources to be focused in prioritizing work in order to have those tires assembled in advance.

1.2. Wheel maintenance at TAP Portugal

At TAP Maintenance & Engineering, the maintenance of wheels is handled by the components department, which is subdivided into four sections as shown in figure 1. The sections of the Components Maintenance department whose performances would be improved with component failure prediction systems are the production, planning and logistics. The scheduling of tasks to be performed at the workshop by the production section is firstly suggested by the planning section using a FIFO (First In, First Out) strategy, which means priority is given to the repair of the tires that were first

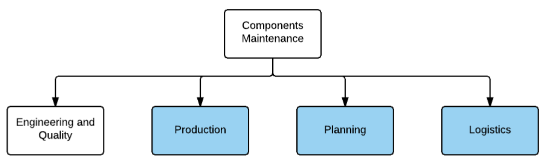


Figure 1: Organisational structure of the components maintenance department. The sections in blue would benefit from component failure prediction systems.

removed. This schedule is then adapted by the production manager taking into account any deviations from the usual task process, and the AOG (Aircraft On Ground) criterion. The AOG criterion dictates that the tasks should always be managed in order to keep a low probability of having an aircraft unable to operate. A 3 day and weekly tire failure prediction would allow for a more suitable schedule to be suggested by the planning section, and would support the decisions made by the production manager to adapt such schedule. On the other hand, the logistics section at the components maintenance department, which ensures the proper supply of auxiliary products and tools for the production section to operate, would mostly benefit from a monthly prediction. By knowing the volume of tires that would be repaired in a specific month, the logistics section could place more accurate orders of products such as O-ring seals, lubricants, and denatured alcohol.

Additionally, there are two departments external to the components maintenance department that could make use of tire failure predictions. The first department is the section of rotatable parts management, which manages the amount of tires stored in different locations, and requests tires from the manufacturers to increase the available stock at specific locations. This section requests tires based on the average amount of tires that were needed in a specific month of previous years. However, this estimation often proves to be inaccurate and additional tires are requested for urgent delivery when the stock at a location drops below an acceptable level. A monthly prediction of wheel failures would greatly improve the accuracy of these requests.

Finally, the operations department could benefit from wheel failure predictions to manage the flight schedule and aircraft. With a prediction of tire failure, this department could manage the fleet so that the aircraft with predicted tire failures would stay in Lisbon overnight and, if needed, have any damaged tires replaced before its operation start the following day.

2. Analysis of Variance - ANOVA

Given two groups of data, if a quantification of the disparity of their distributions is wanted, several methods can be applied, such as a t-test or a Mann-Whitney test. However, given more than two groups of data, comparing their distributions among themselves would require performing methods such as the above for every pair of groups individually. In that case, Analysis of Variance (ANOVA) allows for a more approachable comparison of distributions.

2.1. One-way ANOVA

For this project, ANOVA is applied to evaluate which variables are most relevant to the number of cycles a tire performs between failures. To do so, a measurement variable and a nominal variable are taken into account. The measurement variable will be the number of cycles a tire performs between failures, while the nominal variable will be the variable which relevance is to be determined. Then, multiple observations of the measurement variable are made for each value of the nominal variable and stored in groups.

2.2. Assumptions

One-way ANOVA relies on the assumption that observations within each group are normally distributed. However, this method remains robust when applied to non-normal data [8]. To observe the normality of the data, the differences between each measurement and the mean of its group (i.e. the residuals) should be plotted. In case the obtained distribution is severely non-normal, data transformations should be applied to normalize the residuals.

The second assumption is that the data is homoscedastic (i.e. all groups have the same standard deviations). One-way ANOVA can still produce accurate results for heteroscedastic data when the dataset is balanced. For unbalanced heteroscedastic data, Welch's ANOVA is more accurate than one-way ANOVA.

2.3. Null hypothesis

What is to be tested in this work is how relevant a nominal variable is for the measurement variable, and that implies that the means of the measurements for each group are significantly different. Therefore, the null hypothesis here tested will be that the means of the measurement variable are equal for each group, and this hypothesis will have to be rejected for a nominal variable to be considered relevant. The null hypothesis here tested can be written as

$$H_0 : \forall k, l : \mu_k = \mu_l. \tag{1}$$

2.4. Method

To apply one-way ANOVA, the mean of the measurements within each group is computed and then the variance among these means is compared to the average variance within each group. In order to validate the null hypothesis, the weighted among-group variance would have to be equal to the within-group variance.

Using the designation G_j for each value of the nominal variable that results in a total number of groups of $G = \sum_j G_j$, $y_{i,j}$ for one observation i of n_j within the group j, and μ the overall mean of the measurements, the mean of a group is computed as

$$\mu_j = \frac{\sum_i y_{i,j}}{n_j}. \tag{2}$$

The sample variance is computed by SS/df where SS is the sum of squared deviations from the mean and df is the degrees of freedom. The sample variance will be called "mean square" (MS) here. Accordingly, the mean square between the group means with degrees of freedom $df_{between} = G - 1$ is obtained as

$$MS_{between-groups} = \frac{SS_{between}}{df_{between}} = \frac{\sum_j n_j(\mu_j - \mu)^2}{G-1}. \quad (3)$$

Similarly, the mean square for an individual group j with n_j samples will be given by

$$MS_j = \frac{SS_j}{df_j} = \frac{\sum_i (y_{i,j} - \mu_j)^2}{n_j - 1}, \quad (4)$$

and the mean squares within groups is given by

$$MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{\sum_j \sum_i (y_{i,j} - \mu_j)^2}{\sum_j n_j}. \quad (5)$$

To compare the values obtained by equations 3 and 5 a test statistic is determined through the division of the mean squares among means (3) by the mean squares within groups (5). This test statistic is designated F_s.

$$F_s = \frac{MS_{between-groups}}{MS_{within-groups}} \quad (6)$$

2.5. Interpretation of results

Since F_s has a known F-distribution (also known as Fisher-Snedecor distribution), the probability under the null hypothesis of obtaining the computed value of F_s can be determined. This distribution depends on the degrees of freedom of the among group mean squares ($d_1 = G - 1$) and the degrees of freedom of the within group mean squares ($d_2 = i - G$) and results in the probability density function

$$f(x, d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B(\frac{d_1}{2}, \frac{d_2}{2})}, x \in \Re_0^+ \quad (7)$$

where B is the beta function or Euler integral of the first kind given by

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt, Re(a)Re(b) > 0 \quad (8)$$

The probability obtained for F_s from equation 7 is the p-value. Under the null hypothesis that all groups have the same mean, the p-value is the probability of the null hypothesis being true. In this case, the lower the p-value is, the more heterogeneous the group means will be and more significant the nominal variable will be in the distribution of the measurement variable.

An alternative way to conclude the ANOVA analysis is to compare the obtained F_s with a critical value of F that can be obtained from tables with the degrees of freedom d_1 and d_2 and a significance level α. If $F \geq F_{critical}$, the null hypothesis is rejected and the nominal variable may be considered relevant in the distribution of the measurement variable.

Another popular solution to interpret the results of an ANOVA study and compare the significance of different nominal variables is to compute Eta-squared (η^2), which is a measure of the effect size between a nominal and the measurement variables. This value represents the percentage of the variance in the measurement variable that is explained by a nominal variable and is obtained by

$$\eta^2 = \frac{SS_{between}}{SS_{total}} = \frac{SS_{between}}{SS_{between} + SS_{within}}. \quad (9)$$

3. Analysis of variable relevance using ANOVA

In this work, one-way ANOVA is performed to understand which of the variables available in the provided databases most heavily contribute to the variability of the number of cycles a tire performs between failures. The most relevant variables found will later be used as inputs in the learning algorithm to predict when a tire failure will occur.

The variables which relevance will be analysed and respective domains (which will correspond to ANOVA groups) are presented in table 1.

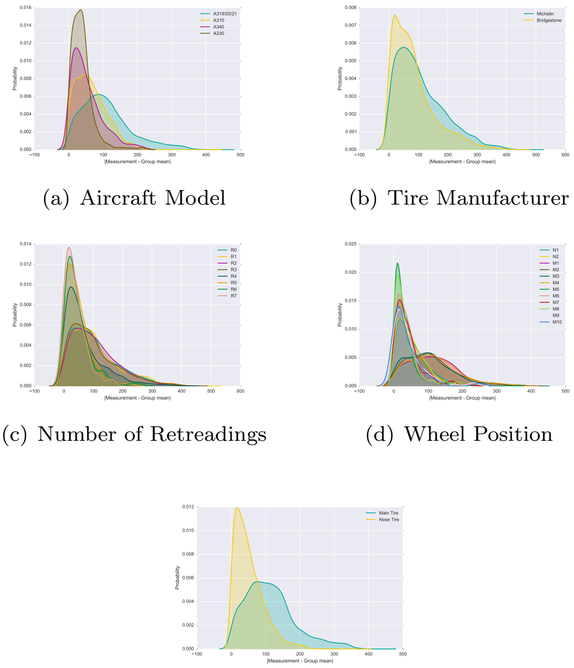
Table 1: Domains of the nominal variables to be used in the relevance study

Variable	Domain
Aircraft model	"A310", "A319/20/21", "A330", "A340"
Manufacturer	"Michelin", "Bridgestone"
Retreadings	1, 2, 3, 4, 5, 6, 7
Wheel position	"N1", "N2", "M1", "M2", "M3", "M4" "M5", "M6", "M7", "M8"; "M9", "M10"
Type of tire	"Nose", "Main"

3.1. Normality of the residuals

The first assumption to be tested in order to apply ANOVA is the normality of the residuals (i.e. the difference between each measurement and its group mean). To do so, the residuals of each variable group were plotted in figure 2. The distributions of such residuals is not explicitly normal and a quantification of such non-normality was performed through a Kurtosis and Skewness test.

The skewness test used was suggested by D'Agostino [4], and the kurtosis test by Anscombe



(a) Aircraft Model (b) Tire Manufacturer

(c) Number of Retreadings (d) Wheel Position

(e) Type of Tire

Figure 2: Validation of the assumption of normality in the residuals of each variable group

and Glynn [1]. Both these tests returned a null probability of these being normal distributions. However, since ANOVA remains robust when applied to non-normal data [8] and this study intends to prioritize variables by relevance and not make direct use of parameters computed with ANOVA, the validity of the assumption of normality will be considered in this case acceptable to apply ANOVA.

3.2. Homoscedasticy

To verify the assumption of homoscedasticy, the standard deviations of each variable group were computed and compared within groups of the same variable. The computed values for the different variable groups significantly varied, and the assumption of homoscedasticy could not be validated.

The results produced by one-way ANOVA are not heavily affected by heteroscedastic data. However, for ANOVA robustness the dataset must be balanced.

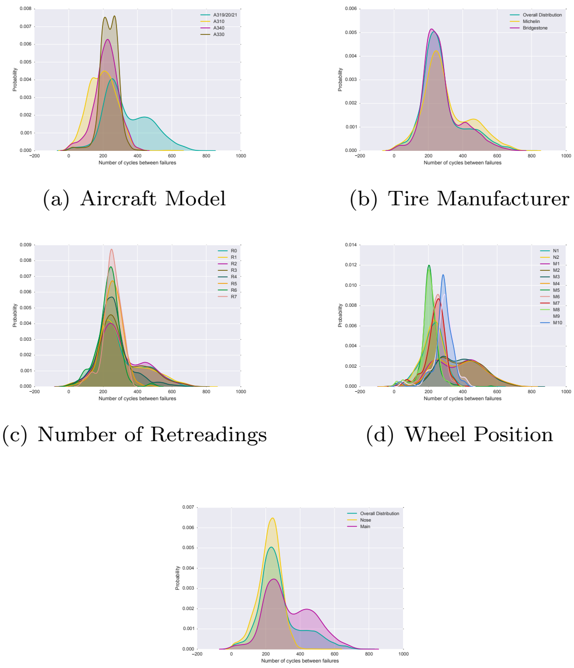
Since the data used in this study is not balanced, the approach taken was to remove samples from each group so that all groups of each variable had the same amount of samples as the group with the smaller sample size.

3.3. Results

The distributions of the variables here studied as shown in figure 3 allow for a visualization of the differences in the means of the groups of each variable and a preliminary estimation of which variables possess groups with distributions that do not significantly overlap. These will be the most relevant variables in the estimation of the number of cycles a tire lasts between failures.

Table 2: Parameter η^2 for variable relevance determination

Variable	η^2
Aircraft model	19.51%
Manufacturer	1.14%
Number of retreadings	1.41%
Wheel position	37.29%
Type of tire	20.89%



(a) Aircraft Model (b) Tire Manufacturer

(c) Number of Retreadings (d) Wheel Position

(e) Type of Tire

Figure 3: Distributions of variable groups for relevance estimation in the distribution of the number of cycles between failures

The distributions presented in figure 3 only allow for a qualitative interpretation. To quantify and rank the variables by relevance in the number of cycles performed by a tire, the parameter η^2 was determined. The obtained values are presented in table 2. A higher η^2 parameter implies a higher percentage of the variability of the number of cycles that may be explained by the variability in a specific variable, and therefore that variable will be of higher relevance for the number of cycles between failures. It is clear that the most relevant variables to the number of cycles a tire performs between failures are the wheel position, the type of tire and the aircraft model. Since the type of tire can be deduced from the wheel position, the variables to be used in the learning algorithm will be the wheel position and aircraft model.

This study produced results that may be considered unforeseen since the individual characteristics of the tire as the manufacturer and the number of retreadings are not significantly relevant for the durability of the tire between failures. The number of retreadings that is proportional to the age of the tire was expected to have a high impact in its durability.

4. Learning Algorithms

Ideally, to obtain a prediction on the number of cycles a tire performs between failures, a model would be built using one or various experts in this phenomenon, and a set of known equations. However, there are currently no available equations to model the failure of a tire, and the costs to conduct such study would greatly surpass the benefits of the obtained predictions. For this reason, a data driven approach was taken.

Data driven modelling requires less time to produce a working model and less expertise on the phenomena behind the model. On the other hand, this approach has the disadvantage of not directly allowing for a better understanding on the phenomena that affect the variable at study, only producing results and not a proper physical explanation of the phenomena behind them.

These are the foundations for machine learning. Since a machine cannot build physical understanding of phenomena, it "learns" from data and with that knowledge, it builds a model that will allow it to infer something on future inputs.

In this work, the prediction of the number of cycles a tire performs between failures is approached as a regression problem, where from a set of features the algorithm attempts to predict this value.

4.1. Bayesian Networks

There are numerous representations for data analysis such as decision trees or artificial neural networks that encode variable relationships and operations in such a way that allows for a conclusion on a specific event or variable to be made at the output. Bayesian networks are an example of a graphical model with properties that overcome many of the difficulties of other representations.

Firstly, the algorithms applied to a Bayesian network can perform even when incomplete data is provided (i.e. when some of the variables to be studied in a case are not observed). Secondly, unlike other representations, Bayesian networks allow a user to gain understanding about the problem at study because its structure explicitly encodes causal relationships between variables. And finally, this representation facilitates the fusion of prior domain knowledge with posterior observations allowing for a more efficient and fluid learning process.

For a set of variables $\mathbf{X} = X_1, ..., X_n$, a Bayesian network will be composed by the network structure S that represents as much conditional independence knowledge about the variables in $\mathbf{X}$ in an acyclic graph as possible, and a set P of local probability distributions of each variable. These two components allow for a joint probability distribution for $\mathbf{X}$ to be determined and inferences on a variable of interest to be made.

The selection of variables is mainly a choice of the network designer, and may be assisted by techniques as the analysis of variance performed in this work for that purpose. Regarding the choice of network structure, several methods have been suggested to find a structure that properly represents the problem. These are generally search algorithms that use a heuristic method to evaluate how adequately each proposed structure fits the problem and suggest the next structure to be evaluated. However, for most of the problems where a Bayesian network is applied, the designer already understands the problem domain well enough to establish causal relationships between variables, and these causal relationships allow for an appropriate network structure to be defined without resorting to search algorithms.

Once the set of variables $\mathbf{X}$ and the network structure S are defined, the joint probability distribution for $\mathbf{X}$ is then

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\mathbf{pa}_i), \tag{10}$$

where $\mathbf{pa}_i$ is the set of parent nodes of a node x_i, and $p(x_i|\mathbf{pa}_i)$ are the local probability distributions.

4.2. Probability determination in Bayesian networks

The probabilities assigned to each node on a Bayesian network are often not known a priori by the network designer, and although these could be manually estimated through statistical techniques, this approach proves inefficient and time consuming especially for large and complex networks. For this reason, in this section a method for learning probabilities in a Bayesian network from a random sample of data is introduced.

Given the hypothesis S^h that the physical joint probability distribution of $\mathbf{X}$ can be obtained according to a given network structure S, and the vector of parameters $\boldsymbol{\theta}_s = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ where θ_i is the vector of parameters for the local distribution function $p(\mathbf{x}_i|\mathbf{pa}_i, \boldsymbol{\theta}_i, S^h)$, the joint probability distribution for $\mathbf{X}$ can be written as

$$p(\mathbf{x}|\boldsymbol{\theta}_s, S^h) = \prod_{i=1}^{n} p(x_i|\mathbf{pa}_i, \boldsymbol{\theta}_i, S^h). \quad (11)$$

The posterior probability distribution $p(\boldsymbol{\theta}_s|D, S^h)$ is then to be determined given a random sample $D = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$.

The learning of probabilities in this work makes use of the unrestricted multinomial distribution [?]. In this approach, each variable X_i is discrete, and has r_i values $x_i^1, ..., x_i^{r_i}$, and each local probability distribution function is a combination of multinomial distributions, one for each configuration of $\mathbf{Pa}_i$. The expression

$$p(x_i^k|\mathbf{pa}_i^j, \boldsymbol{\theta}_i, S^h) = \theta_{ijk} > 0 \quad (12)$$

can then be written, where θ_{ijk} is the multinomial parameter, $\mathbf{pa}_i^j$ denote the configurations of parents of node X_i and $\boldsymbol{\theta}_i = ((\theta_{ijk})_{k=2}^{r_i})_{j=1}^{q_i}$ are the parameters, with q_i being the number of configurations of $\mathbf{Pa}_i$. The notation $\boldsymbol{\theta}_{ij}$ is used corresponding to the vector of parameters $(\theta_{ij2}, ..., \theta_{ijr_i})$ for all i and j.

The posterior probability distribution $p(\boldsymbol{\theta}_s|D, S^h)$ may then be determined under the assumptions that the sample D is complete (i.e. there are no missing data in the sample), and that the parameter vectors $\boldsymbol{\theta}_{ij}$ are mutually independent, or

$$p(\boldsymbol{\theta}_s|S^h) = \prod_{i=1}^{n}\prod_{j=1}^{q_i} p(\boldsymbol{\theta}_{ij}|S^h). \quad (13)$$

Under these two assumptions, the parameters will remain independent given the random sample D, that is

$$p(\boldsymbol{\theta}_s|D, S^h) = \prod_{i=1}^{n}\prod_{j=1}^{q_i} p(\boldsymbol{\theta}_{ij}|D, S^h), \quad (14)$$

and each vector of parameters $\boldsymbol{\theta}_{ij}$ can therefore be independently updated, and the determination of $p(\boldsymbol{\theta}_s|D, S^h)$ may be sectioned into multiple smaller and simpler problems.

Assigning a Dirichlet distribution $Dir(\boldsymbol{\theta}_{ij}|\alpha_{ij1}, ..., \alpha_{ijr_i})$ as a prior distribution for each vector $\boldsymbol{\theta}_{ij}$, the posterior distribution

$$p(\boldsymbol{\theta}_{ij}|D, S^h) = Dir(\boldsymbol{\theta}_{ij}|\alpha_{ij1} + N_{ij1}, ..., \alpha_{ijr_i} + N_{ijr_i}) \quad (15)$$

is obtained, where N_{ijk} is the number of cases in D in which $X_i = x_i^k$ and $\mathbf{Pa}_i = \mathbf{pa}_i^j$. This way, the probability distribution of a a specific vector of parameters $\boldsymbol{\theta}_{ij}$ given data and the hypothesis that the physical joint probability distribution of $\mathbf{X}$ can be extracted from the structure S, will be weighted by the number of cases observed in D that support that vector of parameters.

Averaging over the possible configurations of $\boldsymbol{\theta}_s$, it is possible to make predictions on other cases after D. For a case $\mathbf{x}_{N+1}$ for which $X_i = x_i^k$ and $\mathbf{Pa}_i = \mathbf{pa}_i^j$, the probability of observing that case given random samples D and the hypothesis S^h can be written as

$$p(\mathbf{x}_{N+1}|D, S^h) = E_{p(\boldsymbol{\theta}_s|D,S^h)}(\prod_{i=1}^{n} \theta_{ijk}). \quad (16)$$

Since the parameters are mutually independent, this expectation can be computed as

$$p(\mathbf{x}_{N+1}|D, S^h) = \prod_{i=1}^{n} \int \theta_{ijk} p(\theta_{ij}|D, S^h) d\boldsymbol{\theta}_{ij}. \quad (17)$$

From the expression for the probability distribution of an observation given the samples D and state of information ξ [?]

$$p(\mathbf{x}_{N+1} = x^k|D, \xi) = \frac{\alpha_k + N_k}{\alpha + N}, \quad (18)$$

the probability of interest becomes

$$p(\mathbf{x}_{N+1}|D, S^h) = \prod_{i=1}^{n} \frac{\alpha_{ijk} + N_{ijk}}{\alpha_{ij} + Nij} \quad (19)$$

where $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$ and $N_ij = \sum_{k=1}^{r_i} N_{ijk}$.

4.3. Model evaluation and selection

The elements and techniques on Bayesian networks above introduced emphasize the fact that for a given problem, numerous Bayesian network representations may be built and used. While this method is flexible enough that a specific choice of network structure and conditional probabilities is never considered incorrect, the performance of the obtained network in predicting the values of interest will vary. Therefore, the need for a formulation of model selection criteria becomes clear.

In this section, a method for model evaluation and selection is introduced.

The most common approach to model selection, and the one used in this work consists in defining a criterion to evaluate how well a model fits observed data and running a search algorithm to find the model that obtains the highest score with this criterion.

When using leave-one-out cross validation, a model is trained with all available cases only leaving one case out. Afterwards, a prediction is made on that case with the obtained model, and this prediction is evaluated with a utility function. This procedure is then repeated for every case, and the rewards are summed.

Resorting to the proper scoring rule $logp(\mathbf{x})$ [2] as utility function and the training samples $D'_n =$

$\{\mathbf{x}_1, ..., \mathbf{x}_{n-1}, \mathbf{x}_{n+1}, ..., \mathbf{x}_n\}$, the cross validation criterion

$$CV(S^h, D) = \sum_{n=1}^{N} \log p(\mathbf{x}_n | D'_n, S^h) \quad (20)$$

can be written. However, this criterion interchanges training and testing cases because for a given case $\mathbf{x}_n$, the cases used for training D'_n will make use of $\mathbf{x}_n$ in their own training sets. This fact is likely to induce the selection of a model that over fits the data [5], and the accuracy of any predictions posteriorly attempted with this model will suffer from this lack of flexibility in the model.

To avoid over fitting the data, it would suffice to use a training sample restricted to all cases "before" the case $\mathbf{x}_n$ to be evaluated, $D''_n = \{\mathbf{x}_1, ..., \mathbf{x}_{n-1}\}$, and this would result in a cross validation criterion

$$CV(S^h, D) = \sum_{n=1}^{N} \log p(\mathbf{x}_n | D''_n, S^h). \quad (21)$$

However, as noted in [5], this criterion is equivalent to the log marginal likelihood

$$\log p(D|S^h) = \sum_{n=1}^{N} \log p(\mathbf{x}_n | \mathbf{x}_1, ..., \mathbf{x}_{n-1}, S^h), \quad (22)$$

which will therefore be used as criterion in this study. This criterion may be interpreted as the log of the probability of predicting the cases in D given a specific network structure and parameters.

5. Prediction of tire failure with a Bayesian network

5.1. Prediction of cycles between failures

In order to predict the number of tires of each size that fail in a 3 day, week or month period, an algorithm that predicts the number of cycles a tire performs before its following failure was built.

From the analysis of variance study in section 3 it was established that the variables to be used as features to perform this prediction are the aircraft model and the wheel position. Since these variables are both categorical, it is possible to insert each value these may assume as a binary feature in the Bayesian network as shown in figure 4.

To observe the performance of the obtained network, the available dataset was randomly separated in two groups of samples: the first group of samples is solely used to train the model, and the second group of samples is later used to test the resulting model. This separation is imperative to prevent biased results from being extracted. If all data was to be used to train the model, a latter prediction on any of the samples would benefit from the fact that the model was trained with that specific case.

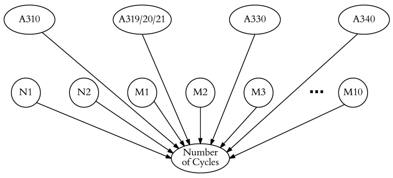


Figure 4: Bayesian network structure to predict the number of cycles a tire performs between failures.

Multiple separations of data were tested in order to confirm that the obtained results would not benefit from a specific random selection of the training and testing datasets, and having this hypothesis been confirmed, the resulting relative errors for the predictions of number of cycles of a testing dataset are displayed in figure 5.

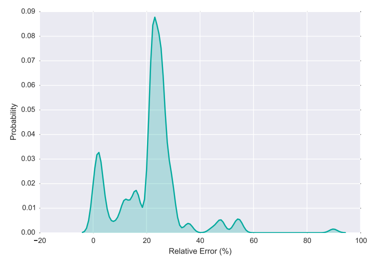


Figure 5: Relative errors obtained in the prediction of number of cycles between failures

The prediction errors obtained and shown in figure 5 result in an average error of 21.172% in number of cycles. Since this algorithm is optimized through a cross-validation process, this error may be originated in data incoherences, lack of data on a relevant variable and a suboptimal feature selection.

The incoherences in data collection were in part corrected by eliminating explicit outliar events from the dataset before training the model, however many data collection errors cannot be distinguished from the natural oscillations observed in the number of cycles a tire performs, and these will remain.

As for the lack of data on relevant variables, the failure of a tire is a phenomenon caused by such a large number of factors that it would not be possible to collect data on all of those factors. One of the factors that might be added to the feature set is the vertical speed observed at touch-down for the landings occurred since the last failure of a tire. It is clear that a higher vertical speed prior to a landing would increase the strain in the tire structure and result in an earlier failure. However, a tire typically endures at least 200 cycles before failing, and the vertical speed of each landing as many other environmental factors are likely to be constant in

Table 3: Set of features based on expert knowledge

A310 main landing gear tire
A310 nose landing gear tire
A319/20 main landing gear tire
A321 main landing gear tire
A319/20/21 nose landing gear tire
A330 main landing gear tire installed in positions from 1 to 8
A340 central landing gear tire installed in positions 9 and 10
A330/40 nose landing gear tire

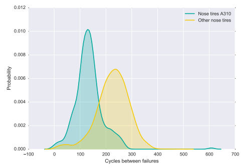
average through that number of cycles.

Although the remaining data incoherences cannot be removed, and additional data on different variables is not available to be added, the feature selection may be improved and result in a lower average prediction error. From the most relevant variables obtained in the analysis of variance, the selected features were the possible values of these variables. This approach is the most intuitive for a designer that does not have direct experience in tire maintenance. Yet, when an engineer with experience in this field was consulted, a different set of features was suggested as follows:

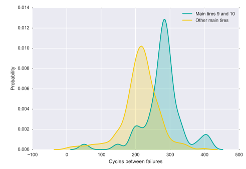
The isolation of the A310 nose gear tire from other nose gear tires stands on the fact that these tires have specifications that differ significantly from other nose gear tires. Similarly, the A321 main tire also differs from A319 and A320 main gear tires. The A340 central gear tires typically perform a higher amount of cycles before failure because these do not have brakes and therefore do not endure as much friction as other tires. Additionally, it was suggested that the left side tires were compared with the right side tires to test the relevance of this separation, since due to the northwest winds frequently present in the Lisbon airport, the differential braking required to overcome these winds and keep aircraft from turning are expected to increase wear on the left tires.

Before inserting the suggested set of features in a new model, the relevance of each feature was tested in order to eliminate any feature that does not in fact exhibit a significant statistical relevance. In figure 6, comparisons between the distributions of the number of cycles tires with specific qualities allow for a relevance analysis of the suggested features.

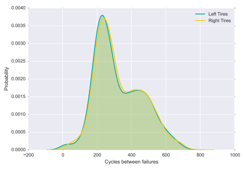
From the distributions shown in figure 6, it is possible to confirm that both the A310 nose gear , the A321 main tire, and A340 central gear tire separation is relevant for the estimation of the number of cycles between failures since these distributions possess notably different means from other tires of the same type. On the other hand, the assumption that left side tires would perform fewer cycles between failures than right side tires is not valid and this distinction will not be made in the set of features to be used.



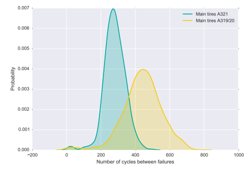
(a) A310 nose gear tire and other aircraft nose gear tires



(b) A340 central gear tires and A340 main gear tires



(c) Left and right side tires



(d) A321 main gear tires and A319/20 main gear tires

Figure 6: Comparison of distributions for the suggested second set of features for relevance analysis

The resulting set of features to be tested will then be A310 main gear, A310 nose gear, A321 main gear, A330 main gear, A319/20/21 nose gear, A319/20 main gear, A340 central gear, A330/40 nose gear, A340 main gear. This set of features originates the Bayesian network in figure 7.

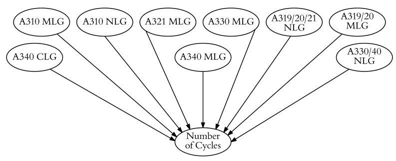


Figure 7: Bayesian network structure to predict the number of cycles a tire performs between failures using a set of features based on expert knowledge. MLG, NLG and CLG denote "main gear", "nose gear" and "central gear".

Training the model with this set of features, the relative prediction errors were plotted in figure 8. The average prediction error for this set of features decreased to 18.681% from the average of 21.172% obtained with the first set of features. It is evident that the performance of the model benefits from a set of features based on prior domain knowledge, and this second model will be used in the remaining of this work.

5.2. Prediction of number of tires by size in a specified time period

For the purposes of maintenance management, the prediction of the number of cycles until the next

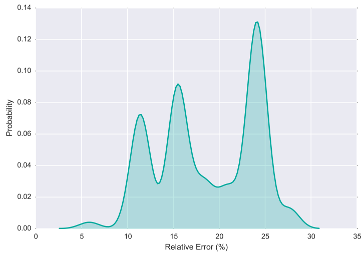


Figure 8: Relative errors in the prediction of number of cycles between failures with a set of features based on expert knowledge.

failure of a tire does not allow for a direct interpretation and action. Instead, from the workshop perspective it is more useful to predict the number of tires of each size that fail in a specific time period, and manage the workshop resources according to that insight.

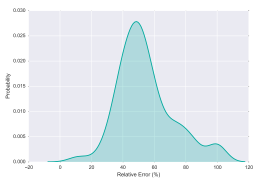
The conversion of number of cycles into a specific date requires the knowledge of the history of landings a tire performs (or the schedule for future flights if a future prediction is to be made), and the date of the tire's last failure. However, since a landing history is not available for all instances, an alternative method for estimating the date of failure of a tire is needed.

To estimate the date of failure of a tire without a landing history, an analysis on the number of cycles and dates was performed using the cycle information and the dates of failure. Analysing pairs of events for each tire, it is possible to extract a monthly average of cycles per day for each aircraft model from the number of cycles registered in the second event of each pair. Since the monthly averages do not significantly vary between two consecutive years, the date of failure prediction is obtained incrementing days with the average cycles per day of the starting month from the previous year, and incrementing the average cycles per day of the following month once that month is reached.

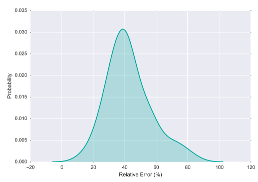
Once the prediction of the date of failure of a tire is available, it is possible to build an algorithm that predicts all tires that fail on a specified day. This algorithm collects the last registered failures of each position in each aircraft tail number and predicts the date of failure of each of these tires. Finally, these dates are compared to the date specified by the user and only the failures that match that date are kept. The tires are then counted by size and a list is presented to the user.

With the prediction algorithm for number of tires by size that fails in a specified date, three similar algorithms were then built: a three day prediction, a weekly prediction and a monthly prediction. In the first two algorithms, the user specifies a starting date in format 'yymmdd' (ex: 151224 for December 24th 2015) and whether a three day or weekly prediction is intended, and the algorithms predict the tires that fail in that period, then counting them by size in the end. The monthly prediction only differs from the first two in the type of input the user inserts, which instead of a day, is a month and a year in format 'yymm'.

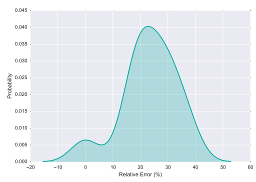
With the intent of evaluating the accuracy of these predictions, a set of predictions for every day (or month in the monthly prediction) of the last three years was made and the obtained errors are shown in figures 9. For each three day, week or month period, these errors consist of the sum of the number of tires by size that were not predicted with the number of tires by size that were predicted and did not fail (i.e. the sum of the number of extra tires that would be required with the number of tires that would remain unused at the end of that period). This error is then normalized with the total number of tires that fail.



(a) 3 day prediction



(b) Weekly prediction



(c) Monthly prediction

Figure 9: Tires by size prediction error normalized to the actual number of tires failed

Inspecting the error distributions obtained in figure 9, it is evident that monthly predictions perform better than three day and weekly predictions. This result was expected since the extra tires predicted for a specific date may compensate the tires lacking in the prediction of another date within the same month or week.

The three day prediction error distribution possesses a maximum density value at 48.913% and a mean of 54.397%, the weekly distribution has a maximum density value of 38.982% and a mean of 42.809% and the monthly distribution a maximum density value of 22.847% and a mean error of 23.359%.

6. Conclusions

The main purpose of this work was to design a tool to predict the amount of tires by size that would fail in a specified three day, week or month period. This tool would allow for improvements to be made at a maintenance management level, especially on human, material and time resources.

An analysis of variance was first performed to determine which variables impacted the most the number of cycles a tire lasted. From this study, the only variables that significantly affected this number of cycles were the aircraft model and the position where the tire was installed. Remarkably, variables directly associated with properties of the tire itself, as the number of retreadings and the manufacturer, proved not to be of significant relevance, and instead the properties associated to external factors influence the life of the tire the most.

With the variables of higher relevance, an algorithm based on Bayesian networks was built to predict the number of cycles a tire would perform until its following failure. This algorithm would receive as inputs a set of binary variables for each possible aircraft model and tire position and return the number of cycles that tire lasts. Testing the predictions made with this algorithm for a sample of tires, an average error of 21.17% in number of cycles was obtained.

An attempt at reducing this error was then made by reformulating the algorithm input while still only making use of the aircraft model and tire position variables. To do so, an engineer with experience in tire maintenance was consulted, and another set of algorithm inputs was suggested. This set of inputs also consisted of binary variables, but instead of introducing an input for each possible variable value, it grouped specific aircraft models and positions for which tires were known to perform similarly, into inputs. Unsurprisingly, when the second set of inputs was tested, a lower average error of 18.68% was attained.

These results emphasize the advantage of encoding relevant prior knowledge into a model, and how useful the insight of an expert may be even to a learning algorithm.

The prediction of number of cycles was then converted into a date using flight history data, and with this algorithm, a tool was built to predict which tires would fail in a specified 3 day, week or month period, and return a list with the amount of failed tires by size predicted for that period.

After testing this tool for a set of time periods, the maximum density errors of 48.91%, 38.98% and 22.85% were obtained respectively for the three day, week and month prediction. These errors consist of the sum of the number of predicted tires that would remain unused at the end of that period with the number of extra tires that would have to be prepared, normalized with the total number of tires that actually fail.

Even with the current errors obtained, these predictions were considered useful as an additional decision support tool. The planning department may resort to the three day and weekly predictions to build a preliminary task schedule that prioritizes the tires by size to be repaired and then apply a FIFO strategy to select the specific tires to be repaired for each size.

Nevertheless, the prediction that proves to be of highest impact is the monthly prediction. The logistics section of the components maintenance department was placing fixed monthly orders of auxiliary products and would now be able to adapt such orders according to the predicted overall amount of tires to be repaired. Additionally, the rotatable components management section can sustain the tire requests at the manufacturer with the monthly prediction, which greatly surpasses the estimations currently made based on previous years' averages.

References

[1] F. J. Anscombe and W. J. Glynn. Distribution of the kurtosis statistic b2 for normal samples. *Biometrika*, 70(1):227–234, 1983.

[2] J. M. Bernardo. Expected information as expected utility. *The Annals of Statistics*, 7(3):686–690, 1979.

[3] R. B. D'Agostino. Transformation to normality of the null distribution of g1. *Biometrika*, 57(3):679–681, 1970.

[4] R. B. D'Agostino, A. Belanger, and J. D'Agostino, Ralph B. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321, 1990.

[5] A. P. Dawid. Statistical theory. the prequential approach (with discussion). *Journal of the Royal Statistical Society A*, 147:178–292, 1984.

[6] R. E. Kirk. *Experimental Design: Procedures for Behavioral Sciences*. Brooks/Cole, California, USA, 3rd edition, 1995.

[7] T. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, USA, 1 edition, 1997.

[8] G. Norman. Likert scales, levels of measurement and the laws of statistics. *Advances in Health Sciences Education*, 15(5):625–632, 2010.

[9] H. J. Seltman. Experimental design and analysis. *Online at: http://www. stat. cmu. edu/, hseltman/309/Book/Book. pdf*, 2012.