

# **Remaining Useful Life Estimation of Caterpillar Vehicle Compartments**

Christopher Page

Melinda Hodkiewicz

School of Mechanical and Chemical Engineering

Edward Cripps

School of Mathematics and Statistics

Martin Weston & Robert Taylor

CEED Client: WesTrac Pty Ltd

## **Abstract**

*WesTrac provide a whole of life management plan through their Equipment Management Centre (EMC) for all customers using Caterpillar machinery, which includes a Scheduled Oil Sampling (SOS) service. Fluid analysts, or interpreters, grade these oil samples and provide feedback accordingly, allowing customers to monitor the overall health of their equipment. This project identifies which elements of the oil are most important in the identification of failures and constructs a statistical model using oil sample data in conjunction with failure history in order to estimate compartment Remaining Useful Life (RUL). The modelling allows WesTrac to predict component failure with higher confidence in order to recommend the correct precautionary actions to their customers. Ultimately, an estimation of asset RUL assists in determining whether a particular compartment should be repaired and, if not, for how much longer it is likely to function.*

## **1. Introduction**

WesTrac's relationship with Caterpillar (CAT) spans approximately twenty years and has made them one of the top five dealers of CAT equipment by sales value globally. In addition to supplying construction and mining equipment, WesTrac provides quality assurance, equipment management and contamination control to their customers. These services are provided for all CAT equipment in WA, NSW, the ACT and Northern China. In particular, the WesTrac site in WA is responsible for analysing oil sample data from the above locations and in 2007 became the first single lab dealership in the world to test over 300,000 machine oil samples in a single year (WesTrac Pty Ltd, 2012). The interpreters in WesTrac's WA office analyse customers' oil data after a series of laboratory tests before grading each sample based on the level of each contaminant.

The oils are classified as A, B C or X, where the worst case X sample classification is given to oil with significant contamination. For each X sample, the customer is advised to stop operating the vehicle immediately otherwise failure will occur. The compartment using the oil should be repaired and serviced as soon as possible. Another oil sample should be taken soon after the repair to determine whether the problem has been resolved (Ratnam, 2011).

These classifications are stored along with the lab results in WesTrac's online Oil Commander system. The purpose of this project is to construct a statistical model using these lab results along with the relevant work order histories in order to predict compartment RUL.

## 1.1 Background

There have been a number of studies conducted in the field of Condition-Based Maintenance (CBM), prognostics and health management in order to predict RUL. The RUL of a system is the time that elapses from its current state until the end of its useful life. This useful life is subjective and based on the context of the system. In engineering, this is generally taken to be the time to failure or when operation is no longer considered viable, given the current asset age and condition, and its operational history (Jardine et al., 2006). Many of the approaches used to calculate RUL are statistical in nature and require suitable historical observational data to accompany the chosen model.

CBM data are generally referred to as direct or indirect. Direct CBM data are able to determine the overall state of a given system using commonly accepted formulae and threshold limits. An example of this is using observed crack size to predict further propagation and eventual structural failure. Conversely, indirect CBM data can only partially describe a system's health, and often require the use of historical failure data to assist in predicting RUL (Si et al., 2011). Common examples include the use of vibration and oil sample data to aid statistical modelling. As this project relies on oil sample data, a review of indirect statistical models was conducted. Of the three typically used indirect methods, a covariate-based modelling approach was selected for this project over stochastic filtering-based models and hidden Markov/semi-Markov models.

## 1.2 Cox's Proportional Hazards Model

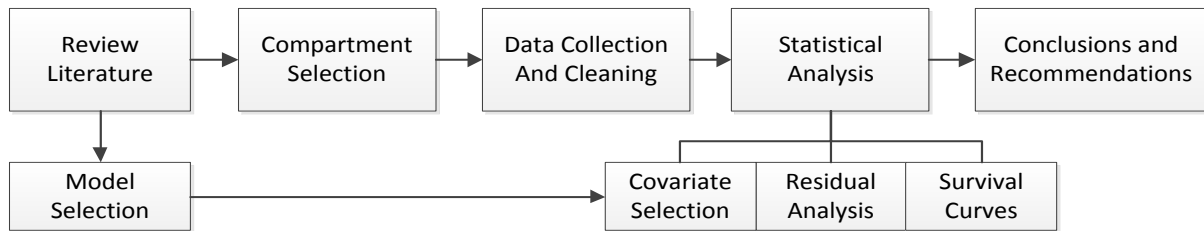
Cox's Proportional Hazards Model (PHM) was chosen for this project due to the wealth of literature relating its use in engineering prognostics using oil sample data. Additionally, these models are less computationally intensive and easier to implement with the data and resources available. Introduced by David Cox (Cox, 1972), the PHM provides a statistical tool widely used in survival analysis. These models utilise historical time data for one or more explanatory variables (covariates) of interest in order to analyse the survival of the associated system. Initially used in social and medical studies, these models have since gained popularity in the field of reliability engineering, where they can greatly assist CBM and improve asset management. The Cox PHM for a subject  $i$  at time  $t$  has a hazard function (Mills, 2011) which takes the form of Equation 1.

$$\begin{aligned} h_i(t) &= h_0(t) \exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}\} \\ &= h_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{X}_i\} \end{aligned} \quad (1)$$

This is the conditional probability of failure at time  $t$  given the values  $\mathbf{X}_i$  at this time (Wong et al., 2010). Changes in the covariates have a multiplicative effect on the hazard function,

which models the degradation of a system as a product of the baseline hazard  $h_0(t)$  and a (positive) exponential function (Sikorska et al., 2011). The baseline hazard indicates how the model behaves in the absence of any covariates.  $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$  is a vector of covariates based on the sample data and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T$  is a column vector of regression coefficients which can be numerically estimated. The above model leads to the formulation of a survival function from which RUL can be deduced.

## 2. Methodology



**Figure 1** Process diagram outlining project methodology

Figure 1 outlines the methodology followed throughout the project. Upon consultation with the client mentors, three customer mine sites were selected for the data analysis. From these, oil sample data were collected for 105 diesel engines from the sites' 793B, C, D and F mine haul trucks. These particular customers were chosen for their generally consistent oil sampling and the potential availability of work orders pertaining to engine maintenance. Data was exported from WesTrac's Oil Commander server before being sorted and cleaned in Microsoft Excel. It was then imported into the statistical software package R (R Development Core Team, 2012) which was used to conduct the statistical analysis.

Work orders containing engine maintenance data were difficult to obtain at the beginning of the project. Attempts were made to gain access to some of these work orders from site on several occasions. However, those that were made available contained very little relevant information regarding engine failures. As such, the statistical analysis conducted so far in the project has only been based on the data collected from Oil Commander by treating each X sample as a failure. While not every X sample will be related to a complete engine failure, they are indicative of a serious problem with the oil being used, resulting in an engine service or repair as previously discussed.

## 3. Results and Discussion

WesTrac has not yet used PHM or any other form of statistical analysis to estimate the RUL of its CAT vehicle compartments. As such, this project is the first of its kind for WesTrac and is not based on any prior work conducted by the company, with the exception of any classification information provided by the EMC and other CEED students.

### 3.1 Covariate Selection

Each of the 30 covariates available was tested in the PHM individually to test their statistical significance. Of these, 12 were found to significantly fit the single covariate PHM. From here, highly correlated covariates were omitted. If two or more covariates are highly correlated,

then there is no need to include more than one in the model since each additional covariate doesn't provide any additional information relating to the probability of failure.

Based on the interpreters' comments to the customers for each X sample it is possible to determine which covariates should have the largest impact on obtaining an X from a physical perspective, rather than a statistical one. This then prompted the re-addition of some covariates to the potential model despite initially being statistically insignificant. This led to a preliminary model containing five covariates. These are iron (Fe), Particle Quantification Index (PQI), sodium (Na), soot (ST) and viscosity at 40 °C (V40). The PHM was built using data for these covariates from the past two years of service, with engines time-censored if they had not yet received an X sample.

### 3.2 Failures Associated With Chosen Covariates

Potential engine failure modes associated with the chosen covariates are outlined in table 1. Maintenance actions are recommended by WesTrac's interpreters for each X sample resulting from the levels of these contaminants in the oil.

<b>Covariate</b>	<b>Potential Failure Mode</b>	<b>Recommended Action</b>
<b>Na</b>	Elevated sodium levels often indicate a coolant leak in the engine.	Monitor coolant usage. Pressurise the cooling systems and check for a pressure drop to identify a leak in the system.
<b>Fe</b>	High levels of iron in the oil are associated with wear particles formed from the degradation of engine components.	Cut the filter open and inspect the media for wear particles.
<b>ST</b>	High soot can indicate problems with restricted air flow and dusty operating conditions.	Check the induction and fuel system components and make sure that the engine boost pressure is within the manufacturer's specifications.
<b>V40</b>	Low viscosities (oil thinning) generally pose a problem for the oil more often than higher ones (oil thickening). This can generally be attributed to fuel dilution.	The source of the leak causing fuel dilution should be located and remedied before resuming operation.
<b>PQI</b>	A non-zero PQI indicates the presence of metallic wear debris in the oil. These are visible in the oil and found during SOS.	High PQI is generally rectified in the same manner as for iron.

**Table 1** Potential failure modes and recommended actions for covariates

### 3.3 Residual Analysis

Several graphical residual tests were conducted to assess the adequacy of the model in order to address and solve any shortcomings encountered due to the covariate data. Testing the assumption of proportional hazards revealed that iron may have a time-varying effect in the model. As a result, iron was stratified into low (0-35 ppm) and high (36+ ppm) levels. This

allows a re-interpretation of the model with the non-proportional effect removed, while seeing how the survival differs for different strata. Although the covariates appear in the exponential part of equation 1 linearly, it was found that a natural logarithm transformation provided a better description of the model.

$$h_i(t) = h_0(t) \exp\{\beta_1 \ln(\mathbf{Na}) + \beta_2 \ln(\mathbf{ST}) + \beta_3 \ln(\mathbf{V40}) + \beta_4 \ln(\mathbf{PQI}) + \text{strata}(\mathbf{Fe})\} \quad (2)$$

Equation 2 was found to obey the assumption of proportional hazards with these transformations applied and each covariate retained statistical significance. The exponential of the  $\beta_k$  for Na, ST, and PQI were all found to be greater than 1, indicating that an increase in each of these covariates leads to an increase in the hazard rate (Mills, 2011). Conversely, the coefficient for V40 was less than 1, implying that a decrease in viscosity is synonymous with increasing the hazard. These results agree with the physical causes for the potential failure modes outlined in table 1 and are displayed in table 2.

Covariate	Transformation	$\beta$	$\exp(\beta)$	Change in hazard per 50% increase in covariate level
Na	log	0.45835	1.58147	20%
ST	log	0.38938	1.47606	17%
V40	log	-2.35257	0.09512	-61%
PQI	log	0.27622	1.31813	12%

Table 2 Results for each covariate

### 3.4 Survival Curves

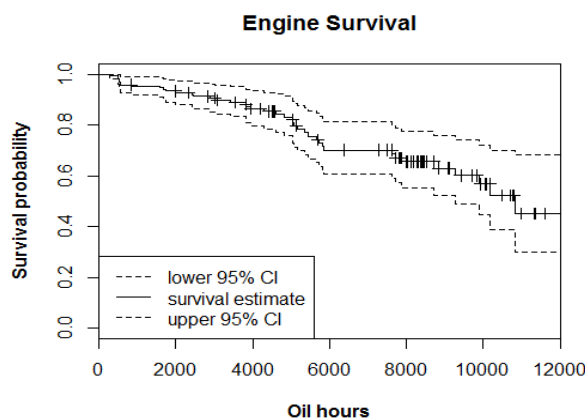


Figure 2 Survival curve - 95% confidence interval

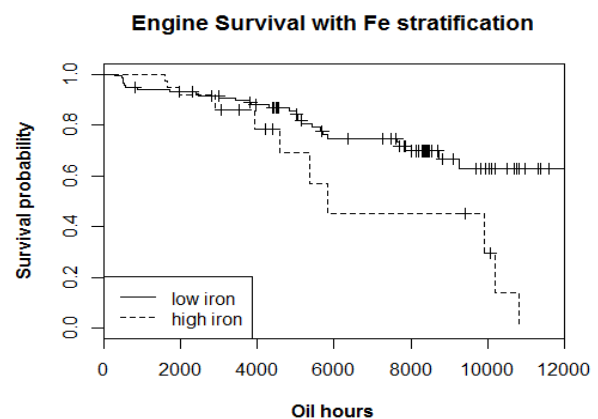


Figure 3 Survival curve - Fe stratification

Figure 2 shows that this survival model predicts approximately 40 per cent survival of the engine population after 12,000 oil hours of service. The term “survival” is taken to mean “not receiving an X oil sample”. The EMC believe that an achieved engine life of 15,000-18,000 oil hours is not unreasonable. The results therefore indicate that 60 per cent of engines from this cohort will receive an X classification (at 12,000 oil hours) before reaching this life. Figure 3 shows two survival curves for low and high iron levels, and adequately depicts a greater tendency towards failure for the upper strata. It can be seen that 40 per cent survival is reached at approximately 6,000 oil hours for high iron, only half the time achieved by the original model (figure 2) for average iron content.

Further work is required as a number of the engines analysed have meter readings exceeding the expected engine life. This suggests that they have been replaced at least once. Hence, work orders pertaining to engine maintenance will assist the statistical analysis, as they should indicate when and why these replacements occurred.

## 4. Conclusions and Future Work

To date, oil sample data for 105 mine haul truck engines have been used to construct a PHM containing five significant covariates in order to estimate RUL. Work orders for two of the three mine sites have been collected and will be used to improve and validate the results obtained so far. Beyond the results discussed for this project, future work could include the investigation of time-varying covariates in the Cox PHM, as iron has already been suspected to have such an effect within this data set. The same principles could also be applied to CAT vehicle compartments other than engines. There is also the potential to conduct a cost analysis based on work order histories and RUL estimation in order to propose an optimal maintenance policy for the future.

## 5. References

- Cox, D., 1972. Regression models and life-tables. *Journal of the Royal Statistics Society, Series B* 34, 187–220.
- Jardine, A.K.S., Lin, D., Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* vol. 20, 1483–1510.
- Mills, M., 2011. *Introducing Survival and Event History Analysis*. SAGE Publications Ltd.
- R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Ratnam, M., 2011. *Automatic “A Sample” Acknowledgement*. Honours thesis, The University of Western Australia.
- Si, X.-S., Wang, W., Hu, C.-H., Zhou, D.-H., 2011. Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research* vol. 213, 1–14.
- Sikorska, J.Z., Hodkiewicz, M., Ma, L., 2011. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing* vol. 25, 1803–1836.
- WesTrac Pty Ltd 2012, *Equipment Management Centre Information*. Available from: <http://www.westrac.com.au/> [26<sup>th</sup> August 2012]
- Wong, E.L., Jefferis, T., Montgomery, N., 2010. Proportional hazards modeling of engine failures in military vehicles. *Journal of Quality in Maintenance Engineering* vol. 16, 144–155.