Modelling of oil tribotechnical data

J Kuben\textsuperscript{1}, P Račková\textsuperscript{1}, O Šimon\textsuperscript{2} and M Zajac\textsuperscript{3}

\textsuperscript{1} University of Defence in Brno, Department of Mathematics and Physics, Kounicova 65, 662 10 Brno, Czech Republic
\textsuperscript{2} University of Defence in Brno, Department of Communication Technologies, Electronic Warfare and Radiolocation, Kounicova 65, 662 10 Brno, Czech Republic
\textsuperscript{3} University of Defence in Brno, Department of Informatics, Cybernetic Security and Robotics, Kounicova 65, 662 10 Brno, Czech Republic

jaromir.kuben@unob.cz

Abstract. The paper deals with statistical processing of field data gained from tribotechnical diagnostics. The data originate from an engine oil analysis. The oil was taken at various locations from several hundreds of medium lorries T810 used in the Czech army. The goal was to propose regression models describing dependence of these quantities on kilometres travelled. Such models can be used, among others, to estimate first hitting time for individual quantities indicating the necessity of the oil change. Two types of models have been used – linear regression models and nonlinear regression models – which seem to be the most appropriate choice for the description of processed data.

1. Introduction

Tribotechnical diagnostics provides us with non-destructive methods of finding out information about the state of various mechanisms. In this case, medium lorries T810 used in the Czech army were an object of interest. A lot of precious items of information can be gained from oil of these vehicles as it is assumed that the condition of a vehicle tightly corresponds to composition of its engine oil. A large amount of data originating from an engine oil analysis, involving over 30 components of various importance was at this research’s disposal.

From the point of view of further processing, oil components can be split into three groups containing observed quantities of interest, see [1]:

- Physical and chemical properties of oil, e.g. kinematic viscosity, flash point, total base number (TBN), anti-oxidation and anti-wear particles (AOWP),
- Products of wearing, e.g. Fe, Cu, Pb,
- Outer contaminants, e.g. soot and Si.

In the analysis of oil, various methods were used: Atom Emission Spectrometry Fourier (AES) – the content of iron (Fe), lead (Pb) and other elements in [ppm] and Fourier Transformation Infra-Red (FTIR) – content of soot, anti-oxidant, and anti-wear additives (AOWP) in [%] and normed methods for stating physical and chemical parameters – kinematic viscosity, flash point, total base number etc.

Our goal was to propose appropriate regression models which could be used to determine first hitting time, that is instant when critical values of observed quantities are reached. The acquired outcomes will serve practical approaches such as i) operation optimisation, ii) maintenance planning and iii) rationalisation of life cycle costs.

Similar topics were studied by numerous authors using various mathematical methods for the investigation of oil tribotechnical data originating from various vehicles. In [2–4], the goal was to keep both preventive and corrective maintenance costs as low as possible; while in [5], the importance of preventive maintenance is demonstrated based on the comparison of data from vehicles with and without such a maintenance. In [6], the influence of soot in oil is studied for armoured personnel carriers (APC), heavy off-road trucked vehicles T-72/97 and medium lorry trucks T810. In [7], aggregation of field data
and their assessment is discussed. Analytical methods can be applied to study particles from wear process and oil deterioration, see [8, 9]. In [10], mathematical methods like stochastic processes, diffusion models and Wiener process are applied. In [11, 12], regression models are used for Fe, Pb, AOWP, and soot. Results in [6–12, 19] and further papers referenced therein, indicate that oil change periods could be extended, and considerable expenses saved.

2. Linear regression models
Standard methods of regression analysis were used. The critical step was to propose an appropriate regression function which would fit data well. Extent of samples was relatively big, almost 300 observations. The explanatory variable was each time the distance travelled by the vehicle in [km]. Typically, the dependant variable is changing only mildly, most often it decreases or increases slowly. But there are exceptions. For example, kinematic viscosity can decrease when fuel gets into oil and can increase as a consequence of a primary or secondary pollution. Similarly, soot content decreases (soot dissolves) when the vehicle does not operate while it increases when the vehicle operates. With respect to the above mentioned facts, linear regression models $Y_i = \eta(x) + \epsilon_i$, $\eta(x) = \beta_0 + \beta_1 \varphi(x)$ were used, where error terms $\epsilon_i$ are assumed to be independent and normally distributed with $\epsilon_i \sim \text{Norm}(0, \sigma^2)$. Choices $x$, $\sqrt{x}$, $1/x$, arctan $x$ and other functions for $\varphi(x)$ were tested. Using least squares method the following well-known formulae for point estimates $b_0$, $b_1$ of $\beta_0$, $\beta_1$ can be derived ($n$ is the number of observations, $x$ and $\overline{y}$ are sample means):

$$b_2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) y_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2}, \quad b_1 = \overline{y} - b_2 \overline{x}, \quad \text{where} \quad \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}. \quad (1)$$

Moreover, we also want to find confidence and prediction intervals for the response. A confidence interval for the mean response is

$$\left[ \eta(x) - t^* \sigma \sqrt{\frac{1 + \frac{(x - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{n}}, \quad \eta(x) + t^* \sigma \sqrt{\frac{1 + \frac{(x - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{n}} \right]. \quad (2)$$

where $t^* = t_{n-2} \left(1 - \frac{\alpha}{2}\right)$ is a quantile of Student distribution, $\alpha$ is a level of significance and an unknown parameter $\sigma$ is estimated with the sample standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n} (\eta(x_i) - y_i)^2}{n - 2}}.$$

Replacing in (2) the quantile $t_{n-2} \left(1 - \frac{\alpha}{2}\right)$ with $\sqrt{F_{2,n-2}(1 - \alpha)}$, where $F_{2,n-2}(1 - \alpha)$ is a quantile of Fisher distribution we get a confidence band for the whole regression function.

Finally, a confidence interval for an individual response, which is usually called prediction interval, is

$$\left[ \eta(x) - t^* \sigma \sqrt{\frac{1 + \frac{(x - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{n}}, \quad \eta(x) + t^* \sigma \sqrt{\frac{1 + \frac{(x - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{n}} \right]. \quad (3)$$

For the derivation of the above formulae and the explanation of concepts confidence and prediction intervals and bands see, e.g., [13-16].
3. Evaluation of regression functions
As an example, we present the evaluation of regression functions for three oil components which are
generally considered to be significant for the assessment of vehicles condition. The formulae were ob-
tained using Computer Algebra System Maple 2018 and the results are listed in table 1.

Table 1. Linear regression models.

| Component                                      | Regression function |
|------------------------------------------------|---------------------|
| Iron (Fe)                                      | $10.55 + 0.1599\sqrt{x}$ |
| Anti-oxidant and anti-wear particles (AOWP)    | $99.83 - 0.00054x$   |
| Total base number (TBN)                        | $10.79 - 0.0101\sqrt{x}$ |

The option of $\phi(x)$ was chosen to get smaller residual sum. In all cases a null hypothesis $H_0: \beta_1 = 0$
was tested and rejected even for very small level of significance $\alpha$. Likewise, normality of error terms
was tested using $\chi^2$-test, Shapiro-Wilk’s test, and d’Agostino et al. tests based on skewness and kurtosis
coefficients but the results were negative. Nevertheless, as the number of observations is quite high this
is not a serious problem due to the central limit theorem.

In figures 1, 2 and 3 graphs of residual functions are displayed including confidence bands for mean
response, for regression function and prediction bands for individual response for $\alpha = 0.05$.

**Figure 1.** Amount of iron depending on distance travelled.

**Figure 2.** Amount of anti-oxidation and anti-wear particles depending on distance travelled.
4. Nonlinear regression models

Data for some oil components are considerably inhomogeneous and it is difficult to find a linear regression model that fits them well. Since the extrapolation to the right outside the range of observed values of an independent variable is required to estimate first hitting time (instants when critical values of observed variables are reached) we have also tested simple nonlinear models. We considered smooth piecewise linear or quadratic functions or their combinations with one break point. For such a type of nonlinear regression models see e.g. [17].

\[
\eta(x) = \begin{cases} 
  a_1(x - x_0)^2 + a_2(x - x_0) + y_0 & \text{for } x \leq x_0, \\
  a_2(x - x_0) + y_0 & \text{for } x > x_0,
\end{cases}
\]

where \( x_0 \) is a break point. The parameters \( a_1, a_2, x_0, y_0 \), coefficients of the regression function, are to be determined using nonlinear least squares method.

As an example, let us present the results for total base number (TBN) data. Appropriate initial values for searched parameters must be set because the algorithm for numerical solution of nonlinear least squares method only guarantees to find local minima. We obtain:

\[
a_1 = 3.161 \cdot 10^{-7}, \quad a_2 = -8.175 \cdot 10^{-5}, \quad x_0 = 1030, \quad y_0 = 10.42.
\]

The corresponding graph is displayed in figure 4. Residual sum of squares 8.991 is smaller than the one given by the linear model above, which is 9.387 [km²].

---

**Figure 3.** Value of total base number depending on distance travelled.

**Figure 4.** Value of total base number depending on distance travelled.
5. Conclusions

We will estimate first hitting times for the three considered oil components. Critical values for T810 requiring oil exchange are listed in table 2.

| Component                                              | Critical value              |
|--------------------------------------------------------|----------------------------|
| Iron (Fe)                                              | 50 [ppm] (upper bound)     |
| Anti-oxidant and anti-wear particles (AOWP)            | 20 [%] (lower bound)        |
| Total base number (TBN)                                | 6 [mg KOH/g] (lower bound)  |

Using regression functions from table 1 we evaluated that the estimates for first hitting time are 60 881, 147 649, 223 617 [km], respectively. Nonlinear regression function (piecewise polynomial function) described by (4) gives the value 55 133 [km]. These values are much bigger than those recommended for vehicles T810, which are prescribed (depending on the type of oil) to about 720 days or 200 Mh corresponding to 6 000 km.

Similar results were obtained for other oil components as well. This indicates that recommended oil change periods could be significantly extended, which would lead to substantial savings of oil expenses. This fact is very important from the economical point of view.

Recently we obtained new large data for this lorry involving even more detailed pieces of information. We plan to continue with this study and use further statistical methods as for example quantile regression, see [18].

Acknowledgement

This work was supported by the project of Specific University Research Preparation, assessment and mathematical processing of tribotechnical diagnostics field data SV17-215 financed by MŠMT of the Czech Republic.

6. References

[1] Stodola, J and Novotný P 2018 Tribodiagnostika BSV (Brno: Univerzita obrany) ISBN 978-80-7231-984-8 (in Czech)
[2] Glos J, Vališ D and Žák L 2014 Possibilities for optimizing maintenance of military vehicles on the results of tribological diagnoses In: International Scientific – technical Conference trans&MOTOAUTO’14. Bulgaria, Varna: Scientific-technical union of mechanical engineering pp. 17–19 ISSN 1310-3946
[3] Glos J, Žák L and Vališ D 2012 Possibilities of Tribo-diagnostics Regarding Maintenance Optimisation. In: Intelligent technologies in logistics and mechatronics systems ITELM’S2012, Kaunas: Kaunas University of Technology pp. 73–78 ISSN 2345-0088
[4] Glos J, Žák L and Vališ D 2013 Possibilities of mathematical modelling of tribo-diagnostics data In: Transport Means Kaunas (Lithuanian Republic: Kaunas Univerzity of Technology) pp. 216–219
[5] Stodola J and Stodola P 2009 Tribology’s Contribution to Efficient Maintenance of Military Engines In: ICMT’09. Oprox, Brno. 10 p. ISBN 978-80-7231-649-6
[6] Vališ D, Hasilová K and Žák L 2017 Combustion engine condition estimation based on indirect diagnostics data In: Risk, Reliability and Safety: Innovating Theory and Practice (London: Taylor & Francis Group) pp. 1930–1936 ISBN 978-1-138-02997-2 doi: 10.1201/9781315374987-293.
[7] Vališ D, Žák L, Glos J and Walek A 2012 Possible inputs to maintenance optimisation based on non-direct diagnoses In: 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012 (PSAM11 ESREL 2012) (New York: Curran Associates, Inc) pp. 143–153
[8] Vališ D, Žák L, Walek A and Glos J 2013 On approaches of assessment of tribo data from heavy tracked vehicle Applied Mechanics and Materials 436 pp. 303–310
[9] Vališ D, Žák L and Chaloupka J 2013 Prediction of further Operation based on Vehicle Tribo Data. In: Proceedings of the 2013 IEEE IEM (Bangkok: IEEE) ISBN 978-1-4799-0986-5
[10] Vališ D and Žák L 2017 Contribution to prediction of soft and hard failure occurrence in combustion engine using oil tribo data Engineering Failure Analysis 82(4) 583–598
[11] Žák L and Vališ D 2015 Oil Additives Used as Indicator and Input for Preventive Maintenance Optimisation In: International Conference on Military Technologies 2015 (ICMT) (Brno: IEEE) pp. 1–6, ISBN 978-80-7231-976-3. doi: 10.1109/MILTECHS.2015.7153659

[12] Žák L and Vališ D 2016 Finding the optimal time for preventive maintenance using linear regression. In: MENDEL 2016, 22nd International Conference on Soft Computing (Brno: University of Technology) pp. 235–240

[13] Anděl J 2007 Statistické metody 4th edition (Praha: MATFYZPRESS) (in Czech)

[14] Faraway J J 2014 Linear Models with R. 2nd edition (New York: Boca Raton, CRC Press)

[15] Pruim R 2018 Foundations and Applications of Statistics 2nd edition (Providence, Rhode Island: American Mathematical Society)

[16] Zvára K 2008 Regrese Praha: MATFYZPRESS. 254 p. ISBN 978-80-7378-041-8. (in Czech)

[17] Seber G A F and Wild C J 2013 Nonlinear regression. Hoboken (New Jersey: John Wiley & Sons)

[18] Zhou K Q and Portnoy S L 1996 Direct use of regression quantiles to construct confidence sets in linear models The Annals of Statistics 24(1) pp. 287–306

[19] Woch M, Kurdelski M and Matyjewski M 2015 Reliability at the checkpoints of an aircraft supporting structure Eksploatacja i niezawodnosc-maintenance and reliability 17(3) pp. 457-462