The aim of the paper is to estimate a system technical life. When estimating a residual technical life statistically, a big amount of tribo-diagnostic data is used. This data serves as the initial source of information. It includes the information about particles contained in oil which testify to oil condition as well as system condition. We focus on the particles which we consider to be interesting and valuable. This kind of information has good technical and analytical potential which has not been explored well yet. By modelling the occurrence of particles in oil we expect to find out when a more appropriate moment for performing preventive maintenance might come. The way of modelling and further estimation is based on the specific characteristics of a regression analysis, fuzzy logic and diffusion processes – namely the Wiener process. Following the modelling results we could, in fact, set the principles of “CBM – Condition Based Maintenance”. However, the possibilities are much wider, since we can also plan in service operation and mission. All these steps result in inevitable cost saving which we would like to contribute to.

**Keywords**: Field data assessment, off-line diagnostics, first hitting time, residual life, maintenance optimization.

1. Introduction – motivation

Reliability, safety and availability of complex and time dynamic systems – like mechatronic, communication, space and smart systems – has attracted more and more attention in recent years – see, e.g. [17]. Systems – we would like to present – work in various and mostly on-Line/Off-Line systems. The majority of maintenance procedures – specifically the PM intervals on the other hand would be obsolete and very rigid in terms of current technical needs. Based on the information previously introduced, reliability analysis, evaluation and predictive methods for reliability assessment need to take into account actual, recent and real-time/on-line system conditions during operation. Real-time reliability and availability assessment may act as vital role in condition-based maintenance which may help to form further maintenance and optimisation decisions – some examples see, e.g. [12]. In real processing of data mining there is important need to have also practical applicability of proposed theoretical models and ideas emerging from modelling. We can find some works on tackling the problems of condition-based maintenance in many aspects. For example in [14] there is a problem of predicting the real-time conditional reliability of an individual tool after its performance data was obtained. In [2] there is proposed an on-line reliability estimation method of an individual component based on degradation signals in which the performance was modelled. Products with exponential degradation paths were studied, e.g. in [8], while degradation signal modelling based on exponential smoothing was modelled, e.g. in [3] and [19] – namely the degradation measures with finite duration impulses. In [20] there is considered the existence of multivariate performance measures, while the proposal for the approach which combines degradation process monitoring with environmental variation is presented in [13].

As present there is a tendency to change the format of technical maintenance. Preventive maintenance (PM) at fixed intervals has been
abandoned and condition based maintenance has been introduced instead. This trend might be followed only on condition that high-quality data on system condition is available. In technical literature, e.g. [2, 3, 8, 14], there are different ways of using direct and indirect diagnostic data. There are introduced the possibilities of vibrodiagnostics, thermal radiation, and also tribo-diagnostics there. As for the vibrodiagnostics and the thermal radiation, they frequently appear in existing publications and scientific papers. Regarding the tribo-diagnostic data, it has been assessed mainly empirically, restrictedly and by specialists.

During the operation of the observed technical equipment in previous years, a lot of tribodiagnostic data were obtained. The truth is that these data have not been used efficiently. The authors of this article identified the potential of these data and applied it in further analysis. The operation data we possess are firmly given by order to collect observations on diagnostics in course of in-service operation – we speak about tribo-diagnostic data. These data are obtained thanks to parts syntactic methods (Atomic Emission Spectrometry – AES) and morphology observation (Laser Net Finder – LNF). From these data we are about to present these indicators which are really of use in presenting of the system real deterioration. No such previous observations and assessment of operating object were conducted. Previous works – see, e.g. [33] – do not speak very deeply about some technical observations and tribo-data as to special big systems like diesel locomotives, mine lorries and war ships. No such extensive investigation has been conducted on medium lorries and common off-road vehicles. What we know for sure is the fact that the tribo data have real potential of presenting system condition. It is probably the most accurate way of determining system state using non-direct diagnostics. Therefore we hope that based on our analytical principles – presented here – real optimisation steps in preventive maintenance planning, costing and mission planning will be allowed/possible to be performed.

There were a few reasons for starting this research. The main reason was obviously to find the way of saving costs during the phase of operation and maintenance of the observed technical equipment. It is rather clear that both the operation and the maintenance have the potential to save costs. The question is how the potential might be identified and used further. The technical literature currently available shows us that the condition based maintenance is a right alternative. However, to introduce this type of maintenance, a certain amount of high-quality data as inputs should be available.

Another reason for assessing and searching for RUL (Residual Useful Life) was to find a lot more adequate model than the ones introduced earlier. The previous models are based on a regression analysis and fuzzy logic which has the potential to support regression models. What we are trying to do in the paper, is to present a new view on the same issue which is supposed to either support the conclusions or disprove them.

2. State-of-the-art and literature survey

Some work in the field of oil data assessment has already been conducted, see, e.g. [10]. In this paper we introduced some fundamental data correlation and characteristics.

In the most recent literature publications a lot of space is devoted to a condition based maintenance. Therefore we have chosen the latest sources dealing with Mean Residual Life (MRL) estimation based on data mining, modelling and other approaches. Deterioration and degradation are other areas we are particularly interested in. For example the work [10] presents the modelling of residual life (MRL – mean residual life) using Proportional Hazards model (PH model) in case of indirect condition monitoring, i.e. the equipment state is not deterministically known. The other work [23] presents possibilities of modelling Remaining Useful Life (RUL) using either a model based approach or a data-driven approach. In [7] we suggested the approach based on a mathematical model for degradation-based signals from a population of components. In work [11] there are methods of estimating the parameters of condition monitored equipment whose failure rate follows the Cox’s time-dependent Proportional Hazards Model. The work [28] presents principles of a non-linear model to estimate the remaining useful life of a system based on monitored degradation signals. Approach looking for balance between costs and preventive periodic maintenance is presented e.g. in [21].

A diffusion process with a non-linear drift coefficient with a constant threshold was transformed to a linear model with a variable threshold to characterize the dynamics and nonlinearity of the degradation process (this new diffusion process contrasts sharply with existing models that use a linear drift, and also with models that use a linear drift based on transformed data that were originally nonlinear). The estimation of remaining useful life, an analytical approximation to the distribution of the first hitting time of the diffusion process crossing a threshold level is obtained in a closed form. An effort to estimate the permanent system deterioration is made in [16]. Therefore the level of true degradation determines the appropriate maintenance actions which are to be carried out. It is another approach to modelling the degradation process by segregating it into manifested (temporary) degradation and true (permanent) degradation – equipment degradation. The estimation of true degradation (with the use of quantitative data + imprecise and vague knowledge) is carried out using fuzzy sets and fuzzy inference system (FIS) on the observed condition indicators and process information. The case study presents steel rolling mill equipment – bearings – degradation. In [9] we focused on the development of a prognostic model to estimate MRL (Mean Residual Life) of Rail Wagon Bearings within certain confidence intervals. This work is concerned with the prognosis of mechanical rotating components. It is about the construction of a survival curve from censored data derived from a nonparametric method introduced by Kaplan and Meier. In the work there is also a construction of the degradation curve using Proportional Hazard Models introduced by COX with censored data used for estimating the survival function.

Our paper, however, is also aimed at looking for the RUL of the equipment, but not in the first instance. We would like to get an optimising coefficient for hard time PM as well as tools for mission planning. For that reason we will use a multivariate function approach when determining an optimal threshold for diagnostic indicators. The paper presents two main approaches to the data assessment. The first one is based on a regression analysis and supported by FIS (Fuzzy Inference System). The FIS is a tool which serves either to accept or reject our decisions when selecting a regression course model. The regression approach shall indicate a possible way of determining the ERL (Estimated Residual Life). The second approach introduces another way based on a diffusion process, namely the Wiener process. This should help with determining the expected distribution of FHT (First Hitting Time) which actually represents the RUL distribution.

3. Objects of diagnostics, oil field data and methods of oil assessment

The assumed objects of diagnostics are in our case heavy tracked vehicle engines. These engines have not been ready yet in terms of design to use an ON-LINE diagnostic system. In practice similar possibilities of other applications have already been existing. It results from the information stated above that we are still supposed to use an OFF-LINE engine diagnostics system when sampling lubrication fluid at certain intervals, and using known and optimised special tribodiagnostic methods [34]. In our case we use the results and information from atomic emission spectrometry. Following this analysis we can obtain the information about the presence of the individual elements of a specific kind and the amount of elements. When evalu-
ating data, the information is transformed many times and provides only estimated reality which might be different from reality itself. If the vagueness in classes distribution is not given by a stochastic character of measured characteristics, but by the fact that the exact line among states classes does not exists, it will be later on good to apply a fuzzy set theory and adequate multi-criteria fuzzy logic. However, we cannot identify the real origin of the respective elements – e.g. as results of fatigue, cutting or sliding. Therefore in our further research we try to identify where these elements might come from. We base our assumptions on the idea to increase the potential for maintenance optimisation inputs and cost benefit analysis inputs. We can perform a good analysis as we have a statistically significant set of data. Taking into account the amount of data, the results are believed to be valuable and statistically reliable. We concentrate on Fe particles contents and their presence in the engine of a heavy tracked vehicle.

4. Application of the regression approach

In this part we present the outcomes of regression functions utilization to describe the data forming and course development. We concentrate only on Fe particles and one vehicle engine type, namely a heavy off-road tracked vehicle. In some previous works, see, e.g. [15 and 34] several outcomes have already been presented. Therefore our results here are based only on the most likely regression courses.

Consequently, the dependencies such as a linear, parabolic and base function – a square root plus confidence intervals in all instances and 34 several outcomes have already been presented. Therefore our assumptions on the idea to increase the potential for maintenance inputs and cost benefit analysis inputs. We can perform a good analysis as we have a statistically significant set of data. Taking into account the amount of data, the results are believed to be valuable and statistically reliable. We concentrate on Fe particles contents and their presence in the engine of a heavy tracked vehicle.

The outcomes from the regression analysis for group of vehicles of the same type are presented below in figures 1–3.

**Table 1. Input data of Fe particles**

<table>
<thead>
<tr>
<th>Sample / Mh</th>
<th>Fe particles (ppm)</th>
<th>Sample / Mh</th>
<th>Fe particles (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/0</td>
<td>17.57</td>
<td>7/46</td>
<td>15.84</td>
</tr>
<tr>
<td>2/8</td>
<td>20.88</td>
<td>8/57</td>
<td>16.41</td>
</tr>
<tr>
<td>3/11</td>
<td>15.77</td>
<td>9/64</td>
<td>23.15</td>
</tr>
<tr>
<td>4/22</td>
<td>19.58</td>
<td>10/72</td>
<td>23.94</td>
</tr>
<tr>
<td>5/26</td>
<td>20.53</td>
<td>11/84</td>
<td>20.86</td>
</tr>
<tr>
<td>6/35</td>
<td>12.73</td>
<td>12/95</td>
<td>17.59</td>
</tr>
</tbody>
</table>

In view of their random character, a random vector \( X = (X_1, \ldots, X_k) \) represents independent variables and a dependent variable is represented by a random variable \( Y \).

When describing and examining the dependence of \( Y \) on \( X \), we use a regression analysis, and this dependence is expressed by the following regression function:

\[
y = \varphi(x, \beta) = E(Y|X=x),
\]

(1)

where \( x = (x_1, \ldots, x_k) \) is vector of numerical variables, \( y \) is a dependent variable, \( \beta = (\beta_1, \ldots, \beta_m) \) is vector of regression coefficients \( \beta_j \). For our data we will look for a regression function in a linear form and we will apply a linear regression model:

\[
y = \sum_{j=1}^{m} \beta_j f_j(x),
\]

(2)

where \( f_j(x) \) are well-known functions where \( \beta_1, \ldots, \beta_m \) are not involved.

For the data we will select gradually the following regression functions for individual item:

- \( m=2, f_1(x)=1, f_2(x)=x \), regression function: \( y = \beta_1 + \beta_2 x \)
- \( m=3, f_1(x)=1, f_2(x)=x, f_3(x)=x^2 \), regression function: \( y = \beta_1 + \beta_2 x + \beta_3 x^2 \)
- \( m=2, f_1(x)=1, f_2(x)=x^{1/2} \), regression function: \( y = \beta_1 + \beta_2 x^{1/2} \)

The coefficient of determination \( (R^2) \) will show its suitability for approximation / data spacing with a relevant regression function. With the coefficient getting bigger, the regression analysis reflects the assessed data better. The form of the coefficient of determination calculation is as follows:

\[
R^2 = 1 - \frac{S_{\text{min}}}{\sum_i^m x_i^2 - n(Y)^2}, \text{ where } S_{\text{min}} = \sum_i^n \left( x_i - \sum_j^m b_j f_{ji} \right)^2
\]

(3)

The outcomes from the regression analysis for group of vehicles of the same type are presented below in figures 1–3.
5. Utilisation Of Fuzzy Inference System (FIS) to support and to compare with Regression Approach

A Fuzzy Inference System (FIS) is based on the terms a fuzzy set and a fuzzy relation which were introduced by Lotfi A. Zadeh in 1965. The fuzzy set is one of possible generalizations of the term set. The fuzzy set is a pair \((U, \mu_A)\) where \(U\) is a universal set and \(\mu_A : U \rightarrow (0,1)\) is a membership function assigning the elements from \(U\) to a fuzzy set \(A\). The membership is marked with \(\mu_A(x)\).

Nowadays one of the most widely used applications is a Fuzzy Inference System – FIS (once used as a “fuzzy regulator” term). Two basic types of the FIS are used, and they are Mamdami and Sugeno [22 and 31]. Each FIS consists of input and output variables and FIS rules. For each FIS we specify:

- the number of input and output variables,
- for each input and output the number of predefined values (linguistics values) in the form of fuzzy sets,
- FIS rules described by predefined values.

We do not often expect a fuzzy set to be the FIS output, but we wish to get a single value \(z_0\), i.e. we want to defuzzify the FIS output. The centroid method is one of the most frequently used defuzzification methods. The FIS specified this way is called Mamdani FIS [22].

If we do not know how the process works (i.e. the FIS rules cannot be set), but the sufficient amount of input and output data is available, we can use the modification of Mamdani-FIS Sugeno (Takagi-Sugeno FIS) [22 and 31].

When looking for FIS correlation between output values and input ones as for an unknown process, the method used a lot more frequently is the Sugeno FIS method which is in fact a Mamdani FIS modification. In order to find a relevant FIS, we use the data that serves as a background for the input and output values of the process. In most cases these values are a subset of real numbers, and therefore the inputs and outputs are in a numerical form. The input variables are similar to Mamdani FIS. The output variables \(Z_j\) are in constant or linear forms.

\[
Z_j = a_j + \beta_{1j} f_{1j} + \beta_{2j} f_{2j} + \ldots + \beta_{nj} f_{nj},
\]

where \(a_j, \beta_{ij}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, k\) are suitable constants, \(k\) is the number of rules in the FIS model, and \(n\)-tuple \((f_{1j}, f_{2j}, \ldots, f_{nj})\) consists of \(n\) input variables to the FIS (model). Sugeno FIS output is the value of weighted average \(Z_j\) where the weight is obtained by comparing the input \((x_{1j}, x_{2j}, \ldots, x_{nj})\) with predefined input values [22 and 31].

To find a suitable Sugeno FIS, which describes the selected data, it is appropriate to divide the data into tuning and checking data. We find the FIS that corresponds to the tuning data best. The tuning part of data is divided into smaller parts, and predefined input (output) values and the rules describing relationship between relevant inputs and outputs are assigned to each part. There are two basic ways of dividing the data:

- dividing the area (which includes tuning data) into smaller parts. A fuzzy set is assigned to each part, and their combination is used for creating rules,
- applying clustering methods to find clusters in data. One rule is made for each cluster.

After selecting the number of fuzzy sets (linguistics values) and rules, we search for appropriate parameters \((a_j, \beta_{ij})\) using output variables \(Z_j\). These parameters were found through a neural network. The tuning itself results in setting parameters for the FIS to describe assigned tuning data as well as possible. The accuracy is verified by calculating the output values from the test data by the FIS, and then they will be compared with the original output of the test data. The design, tuning and selection of the FIS were performed in MATLAB (Version 5.3) – FuzzyToolbox.

The FIS is applied here in order to support our regression courses and principles. When looking for a correlation between data, fuzzy model results and regression results, we concentrated on quadratic and first base function courses only. Following the results of the regression analysis, we decided in favour of these two models, since they have higher \(R^2\) coefficient of determination and therefore the most suitable correlation and inference. We also choose a more strict condition in the form of a mean value of the observed object (vehicles group). The outcomes are presented in Figures 4 and 5.

Fig. 4. Comparison of quadratic Fe course and fuzzy model

[Image]

Fig. 5. Comparison of first base function Fe course and fuzzy model

[Image]

It is remarkable that the expected regression courses of Fe – quadratic and first base – suit also the FIS.

6. Estimation of RUL based on a regression function

It is worthy of attention that a Fe particles generation based on oil field data might have a linear, a quadratic and a base functions course. This correlation and inference outcomes are based on the analysis performed above while using a regression and fuzzy approach. The linear course seems to be the best inference for this particular vehicle engine type. Therefore we take into consideration the linear independence course. Moreover, we have available data obtained from a similar engine laboratory life test. Such test simulated real engine operation and was performed as an accelerated test until the engine was destroyed. These data and the engine serve as a reference item for a further analysis and comparison. Therefore we decided to try to estimate the residual life based on the reference engine and all the outcomes presented above. We believe that the capability to “read the diagnostic data” might help with mission planning, maintenance optimisation, or, e.g. in a cost benefit analysis. Some sources of inspiration can be found, for example, in [1, 4, 5, 16, 25, 26, 27, 29, 30, 32, 35]. The interval estimation of residual life was made on the basis of
the Fe particles analysis and the results of the accelerated life test. Its value is rated on a scale of 220 Mh to 257 Mh. It results from the set interval and the real value declared for oil change that in practice the oil is changed when its durability is higher than 50%. In Figure 6 there is a graphical presentation of the RUL estimation.

![Graphical presentation of RUL estimation](image)

### 7. Utilization of the Wiener process

As it has been mentioned before, the data are available in a big amount – it is a statistically important set. The data is collected at intervals and under conditions determined by a methodology which includes:
- homogenous time intervals given by technical equipment mileage,
- oil temperature and right oil mixing,
- the same place of sample collection,
- the same way of performing the tribo analysis after the sampling.

When applying the Wiener process, we use some outcomes from a regression analysis. These regression outcomes are re-calculated into a usable form. The example of the data form is in Table 2.

<table>
<thead>
<tr>
<th>Sample / Mh</th>
<th>Fe particles (ppm)</th>
<th>Standard deviation mean value</th>
<th>Standard deviation individual value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/0.15</td>
<td>13.71158</td>
<td>0</td>
<td>4.753368309</td>
</tr>
<tr>
<td>2/0.30</td>
<td>13.72016</td>
<td>0.001014608</td>
<td>4.753368417</td>
</tr>
<tr>
<td>3/0.45</td>
<td>13.73055</td>
<td>0.002092216</td>
<td>4.753368742</td>
</tr>
<tr>
<td>4/0.60</td>
<td>13.74004</td>
<td>0.003043824</td>
<td>4.753369284</td>
</tr>
<tr>
<td>5/0.75</td>
<td>13.74953</td>
<td>0.004058431</td>
<td>4.753370042</td>
</tr>
<tr>
<td>6/0.90</td>
<td>13.75902</td>
<td>0.005073039</td>
<td>4.753371016</td>
</tr>
</tbody>
</table>

We assume that the case we observe is a stochastic process with time dependence. The generation of Fe particles is time dependent. Therefore the application of a diffusion process seems to be perfectly adequate. Due to the normal distribution of a random variable and are put in Figure 2. The critical value of particles amount is 50.

![Graphical presentation of Wiener process simulation for an individual value](image)

Next, it applies that:
1. $E[W(t)] = 0$ for $t \geq 0$
2. $\text{Var}[W(t)] = t$ for $t \geq 0$

The Wiener process represents one possible form of diffusion processes. It is actually the integral of what in practical applications is called a white noise. The Wiener process with a drift will be used in our application. The initial mean value (drift) is $\beta_1$ and standard deviations for each time increment have been previously calculated – see Table 1. For our model we apply the Wiener process with a drift given by a stochastic differential equation:

$$dY(t) = \mu \cdot dt + \frac{\sigma \cdot dW(t)}{\sqrt{t}} \quad (4)$$

where $dW(t)$ is increment of the Wiener process and $dt$ is increment of time, $\sigma$ is a standard deviation (either of an individual or a mean value), $\mu$ is a mean value, $t$ is an instant of time, process initial value $Y(0) = \beta_1$. Time increment for modelling was 0.15 Mh. When modelling and simulating, we apply the course of an individual and a mean value, as shown in Fig. 7 and 8. The critical value of particles amount is 50.

![Graphical presentation of Wiener process simulation for mean value](image)
line), Normal (dash and dotted line) or Weibull’s were not proved. The courses of these tested distributions are shown in Fig. 7 and Fig. 8. We expect then to obtain the FHT distribution only in the form of an empiric distribution function.

8. Estimation of RUL based on the Wiener process

The tested values reached for the individual value the following limits: a minimum value = 134 Mh, a lower confidence interval 2.5% = 265 Mh, an upper confidence interval 97.5% = 524 Mh, a mean value = 382 Mh, median = 378 Mh, maximum = 887 Mh.

As to the mean value, the following limits were achieved: a minimum value = 258 Mh, a lower confidence interval 2.5% = 315 Mh, an upper confidence interval 97.5% = 480 Mh, a mean value = 385 Mh, Median = 381 Mh, Maximum = 746 Mh.

It follows from the results stated above that the mean value is more or less the same, but the lower threshold of confidence intervals is not. However, the lower confidence intervals are interesting for us in order to determine the possible beginning of the PM interval. But if we used the mean value, it would be sufficiently far from the original/fixed PM interval. On the basis of the results we could work with a conservative version and set a new PM interval using the lowest value of a lower threshold of a confidence interval. This would be 265 Mh (an individual value). If we were for a benevolent alternative, we could rely on the mean value and set the PM interval at 380 Mh (more or less the same for both the individual and the mean value).

When planning a mission, we could work with versions that if common operating conditions were observed, a vehicle could be operated with one oil filling theoretically up to the upper confidence limit 97.5%. Considering conservative or benevolent versions, it would be 480 Mh, or 520 Mh.

9. Discussion

Since we have worked with two approaches to one problem, it is always interesting to compare the results. As we can see, the obtained values of RUL estimations do not differ when working with conservative estimations. The procedure based on the Wiener process, however, is somewhat clearer and brings better analytical results. The form of RUL estimation is better to determine. It is also much easier to introduce other forms of a particle production course, not only the linear one. Following this assumption – applying a quadric, or a base function course, we can develop our further procedure and research.

Assessing available oil data, however, has a lot greater potential.

10. Conclusion

In this article we have introduced possible approaches to modeling indirect diagnostic measures. Our intention was to introduce a coherent research form when dealing with indirect oil diagnostic data and analyze it with different approaches/ways. Owing to the different approaches, we were able to present quantitative RUL values which are in view of PHM or CBM very important.

The achieved results complement the set of approaches to the indirect observation of a technical condition. Following the conclusions of modeling with the Wiener process, the results of previous approaches might be completed when searching for:

- optimum interval PM,
- recommended/allowed time for mission completion,
- optimizing dependencies of life cycle cost analyzing.

The approach introduced above opens the possibilities of analyzing other essential diagnostic indicators. The setting of the time derived from the histogram and pdf course should be as accurate as possible and undistorted.
In our further analysis we are going to develop the Wiener approach even more and complement it with other approaches like ARIMA or ARMA methods.

**Acknowledgement:** This paper has been prepared with the great support of the project for the institutional development of K-202 University of Defence and with support of the project of Ministry of Defence of the Czech Republic – BOPROS nr: OFVTUV2013002.

**Bibliography**


**David VALIS**
Department of Combat and Special Vehicles  
Faculty of Military Technologies  
University of Defence  
Kounicova 65, 662 10 Brno, Czech Republic  
E-mail: david.valis@unob.cz

**Libor ZAK**
Department of Applied Mathematics  
Faculty of Mechanical Engineering  
Brno University of Technology  
Technicka 2896/2, 616 69 Brno, Czech Republic  
E-mail: zak.l@fme.vutbr.cz

**Ondrej POKORA**
Department of Mathematics and Statistics  
Faculty of Natural Sciences  
Masaryk University  
Kotlarska 267/2, 611 37 Brno, Czech Republic  
E-mail: pokora@math.muni.cz