

THE APPLICATION OF DISCRIMINANT ANALYSIS IN TRIBOTECHNICAL DIAGNOSTICS OF OILS

ПРИМЕНЕНИЕ ДИСКРИМИНАЦИОННОГО АНАЛИЗА В ТРИБОТЕХНИЧЕСКОЙ ДИАГНОСТИКЕ МАСЛА

Sejkorová M., Machalíková J.
Jan Perner Transport Faculty – University of Pardubice, Czech Republic

Abstract: This paper presents the possibilities of application of multidimensional statistical methods in tribotechnical diagnostics. The discriminant analysis method was used, which enables evaluating the differences between two or more groups of objects characterized by several indicators - tribodiagnostic parameters. The purpose was to find a predictor model which enables the classification of objects (samples taken of oils) into classes, i.e. to assess whether or not the oil is able to perform its function in the engine.

Keywords: TRIBOTECHNICAL DIAGNOSTICS, DISCRIMINANT ANALYSIS, OIL

1. Introduction

The effectiveness of the operation of vehicles and items of machinery depends on their technical conditions. The most effective methods of monitoring the technical conditions of individual components include no-dismantlement tribotechnical diagnostics (TTD), which, among other things, deals with the evaluation of changes of selected qualitative parameters of the lubricant. The results obtained by measuring the set of physical and chemical parameters provide information not only about substantial changes in the lubricant's properties, connected with the changes of the unit's technical conditions and operating conditions, but also provide information about the quality of the oil fill. This enables an evaluation of the ability to meet all the requirements that are put on the oil fill and to determine the optimum interval of replacement of the oil fill depending on the level of degradation and contamination. The starting point for evaluation of the dynamics of changes of individual parameters is the values of the unused oil.

Effective utilization of tribotechnical diagnostics, in spite of the considerable initial costs of the equipment, brings the possibility of effective management of the lubricants. Diagnostics based on analyses of the lubricant also enable indicating the occurrence of a failure in time, eliminating it at an early stage and planning a remedy so as to prevent the unexpected breakdown of equipment and so that the repair is not untimely and useless.

This paper presents the possibilities of utilization of multidimensional statistical methods in tribotechnical diagnostics. The discriminant analysis method (DA) was used, by means of which the evaluated samples of oils can be classified into one of the already existing classes of engine oils:

- *good oils*, which are able to perform their function in the engine;
- *bad oils*, where the values of the monitored qualitative parameters are out of the range ensuring serviceable condition of the engine with the particular oil fill.

The purpose was to find a predictor model enabling the classification of objects (samples taken of oils) into classes, i.e. to assess whether or not the oil is able to perform its function in the engine. The purpose was to find a predictor model, which enables the classification of new objects (samples taken of oils) in predefined classes. Calculations were done by means of the statistical program *STATISTICA 7.0*.

2. Principle of the DA method

The classic classification discriminant analysis belongs among the methods of examination of the relation between the p

group of independent indicators, called discriminators (columns of the source matrix), and one qualitative dependent variable – output.

During discriminant analysis the resulting value of the *discriminant function* of the respective object is quantified by substituting in it the values of discriminators x_1, x_2, \dots, x_p of the p group. Such quantified values of the function are used for classification of the objects into classes known in advance, which are characterized by their probability density function $f_j(\mathbf{x})$, where $\mathbf{x}^T = [x_1, x_2, \dots, x_p]$. The principle of this method consists in classification, discrimination of the object of vector \mathbf{x} , in class G_j according to the following rule [3]:

$$(1) \quad f_j(\mathbf{x}) = \max_{i \in \{1, \dots, g\}} f_i(\mathbf{x})$$

A priori probability

In the analyzed sample selection we determine the *a priori probability* of the object belonging to the k^{th} class. If π_1 stands for the a priori probability of the object belonging to the 1st class, then $\pi_2 = 1 - \pi_1$ is the probability of the object belonging to the 2nd class.

It depends on whether or not there is:

- an equal number of objects in the classes - the a priori probability results from the reciprocal value of the number of classes j , $C_k = 1/k$. For three classes $k = 3$ and probability $C_k = 0.33$, so $\pi_1 = 0.33$, $\pi_2 = 0.33$, $\pi_3 = 0.33$.
- an unequal number of objects in the classes - the a priori probability is given by the relation

$$(2) \quad \pi_k = \frac{n_k}{n}$$

n_k ... stands for the number of objects in the k class

n ... stands for the total number of objects in the analyzed sample

A posteriori probability

After performance of multi-dimensional observation x it is possible to determine the *a posteriori probability* of belonging to the j^{th} class according to the Bayes formula

$$(3) \quad P(G = j | \mathbf{x}) = \frac{f_j(\mathbf{x})\pi_j}{\sum_{i=1}^k f_i(\mathbf{x})\pi_i}, \quad j = 1, \dots, k$$

$P(G = j | \mathbf{x})$... stands for the conditional probability of phenomenon G

$f_j(x)$... stands for the conditional probability density of the complex of p considered variables for $j = 1, \dots, k$

The object will be classified in the class in which the a posteriori probability is the highest and, at the same time, the probability of erroneous classification is minimum. If there are two classes defined, the denominator will remain the same in both cases. The classification rule will be simplified to the following form:

$$\pi_1 f_1(\mathbf{x}) > \pi_2 f_2(\mathbf{x}) \rightarrow \text{the object will be classified in the first class}$$

$$\pi_1 f_1(\mathbf{x}) < \pi_2 f_2(\mathbf{x}) \rightarrow \text{the object will be classified in the second class}$$

3. Implementation of the probability model

In the first step the source data matrix was constructed, where the indicators (columns) represent the monitored tribodiagnostic parameters of the used engine oil OA-M6 ADS II – distance the vehicle travelled in kilometres (PMK), distance the vehicle travelled in kilometres with the oil fill (POK), concentration of abrasive metals Fe, Cu, Pb, the value of Conradson Carbon Residue (CCR), the value of the total contamination index (TCI), the values of the Total Base Number (TBN) and Total Acid Number (TAN), the value of the kinematic viscosity at 100°C (KV) and the flash point (FP); the objects (rows) represent individual samples of oils taken from heavy off-road lorries TATRA 815.

On the basis of the knowledge of all the indicators (tribodiagnostic parameters), the entire group of the tested objects (samples taken of oils) were sorted in two classes, where oil in serviceable condition was marked *good*, oil with limit values of some indicator was marked *bad*. The analyzed selection included 71 samples of oil in total, of which 48 were classified in the *good* class and 23 in the *bad* class.

Each indicator in all objects was examined separately by means of exploratory analysis diagnosing.

Owing to the fact that the monitored tribodiagnostic parameters show considerable variability and they are also measured in various units, data standardization was performed and the standardized values were used thenceforth. Fig. 1 presents the exploratory analysis of indicators by means of box diagrams.

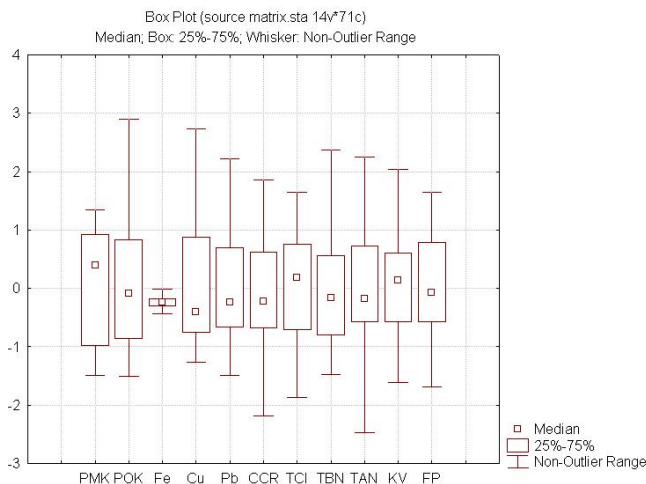


Fig. 1: Box diagrams of monitored tribodiagnostic parameters

Discriminant analysis is sensitive to the normality of the discriminators. Data, which do not show multidimensional normality, may cause problems in the estimation of the discriminant function. Indicators *Fe*, *PMK* were excluded from the set of indicators, because they showed strong abnormality. Another perturbing influence, which may strongly affect the results, is the multicollinearity in data (multicollinearity means that two or more discriminators are highly correlated). The table below shows that all indicators are of a linear nature (Table 1).

Table 1: Correlation coefficients of pairs of all indicators of the source data matrix

Correlations (source matrix)	
Marked correlations are significant at $p < .05000$	
N=71 (Casewise deletion of missing data)	
Variable	
POK	1,00 0,12 0,33 0,52 0,35 -0,27 0,57 -0,18 -0,16
Cu	0,12 1,00 0,38 0,44 0,32 0,02 0,40 -0,30 -0,03
Pb	0,33 0,38 1,00 0,56 0,17 -0,30 0,42 -0,13 -0,07
CCR	0,52 0,44 0,56 1,00 0,52 -0,19 0,47 -0,01 -0,15
TCI	0,35 0,32 0,17 0,52 1,00 -0,38 0,46 -0,24 -0,38
TBN	-0,27 0,02 -0,30 -0,19 -0,38 1,00 -0,18 0,12 0,17
TAN	0,57 0,40 0,42 0,47 0,46 -0,18 1,00 -0,14 0,02
KV	-0,18 -0,30 -0,13 -0,01 -0,24 0,12 -0,14 1,00 0,60
FP	-0,17 -0,02 -0,06 -0,16 -0,39 0,19 0,02 0,53 1,00

The step method was used for derivation of the discriminant function. This method selects the indicators, which contribute most to the discrimination, and eliminates those, which are not suitable for analysis. The following table contains an examination of the influence of individual discriminators included in the model.

Table 2: Results of discriminant functional analysis

Discriminant Function Analysis Summary (source matrix)						
Step 7, N of vars in model: 7; Grouping: Stav (2 grps)						
Wilks' Lambda: ,11516 approx. F (7,63)=69,152 $p < ,0000$						
N=71	Wilks' Lambda	Partial Lambda	F-remove (1,63)	p-level	Toler.	1-Toler. (R-Sqr.)
Cu	0,2866	0,4019	93,7716	0,0000	0,7463	0,2537
TAN	0,2228	0,5169	58,8886	0,0000	0,5283	0,4717
CCR	0,1529	0,7530	20,6675	0,0000	0,5384	0,4616
Pb	0,1242	0,9269	4,9714	0,0293	0,6560	0,3440
KV	0,1309	0,8800	8,5931	0,0047	0,7965	0,2035
POK	0,1215	0,9477	3,4752	0,0670	0,5287	0,4713
TBN	0,1194	0,9649	2,2950	0,1348	0,8190	0,1810

Table 2 shows that discriminators *FP* and *TCI* are not included in it, i.e. they were excluded due to failure to meet parameters of the crucial criteria of the set of indicators.

Another process was related to the examination of discriminant functions, so as to determine each discriminator's importance for classification.

The traditional approach to interpreting discriminant functions examines the sign and size of standardized discriminant coefficients $a^T = [a_0, a_1, \dots, a_p]$. The discriminators with relatively high coefficients contribute to the discrimination strength of the discriminant function more than the discriminators with lower coefficients. The sign indicates either the positive or negative contribution of the discriminator. The program quantified the coefficient of discriminant functions for classification of objects in the "good oil" and "bad oil" classes.

Below are the discriminant functions for classification:

$$Z_{good} = 2.120x_{Cu} + 6.775x_{TAN} + 19.463x_{CCR} - 0.423x_{Pb} + 70.460x_{KV} + 0.001x_{POK} + 0.413x_{TBN} - 448.841$$

$$Z_{bad} = 4.088x_{Cu} + 17.549x_{TAN} + 46.278x_{CCR} - 0,871x_{Pb} + 64.513x_{KV} + 0x_{POK} + 1.260x_{TBN} - 447.281$$

The oil sample is classified in the class in which it assumes the highest values of the discriminant score.

The predictive ability of the proposed model is determined by the number of objects classified in correct classes. Owing to the fact that the analyzed selection of samples of engine oils was quite narrow, the discriminant functions were created from the entire selection and then the entire selection was used for classification. Table 3 proves that the achieved classification of oils into classes was 100 %.

Table 3: Classification matrix of oils

	Classification Matrix		
	Percent correct	Good p=0.67606	Bad p=0.32394
Good	100.000	48	0
Bad	100.000	0	23
Total	100.000	48	23

4. Conclusion

This paper presents the possibilities of using classification methods of multidimensional statistical analysis on specific results of the analysis of engine oil OA-M6 ADS II, which was used in TATRA 815 heavy off-road lorries.

The discriminant analysis method quantified the value of the discriminant function for the class of engine oils, which were known to comply with the required qualitative parameters, and for the class of engine oils, where the values of the monitored qualitative parameters went beyond the range ensuring serviceable condition of the engine with the particular oil fill.

The effectiveness of the performed discrimination was verified by means of re-substitution, i.e. application of the discrimination classification on the sample selection. All the classified oils were classified correctly.

5. References

1. LIPPAY, J.: Kandidátská disertační práce. Vojenská akademie v Brně. Brno 1991.
2. STODOLA, J.: Možnosti modelování a vyhodnocování degradačních procesů spalovacích motorů. In Sborník přednášek 7. ročníku semináře REOTRIB 2001 – zjišťování kvality paliv a maziv. Velké Losiny, 22. – 25. května 2001. Ostrava: ReoTrade; Praha: VŠCHT, Ústav technologie ropy a petrochemie, 2001, str. 77 – 86. ISBN 80-7080-426-2.
3. MELOUN, M., MILITKÝ, J., HILL, M.: Počítačová analýza vícerozměrných dat v příkladech. Academia, Praha 2005. ISBN 80-200-1335-0.