Article

Multivariate Quality Control of Lubricating Oils Using Fourier Transform Infrared Spectroscopy

Alessandra Borin and Ronei Jesus Poppi

Instituto de Química, Universidade Estadual de Campinas, C P 6154, 13084-971 Campinas - SP, Brazil

Controle de qualidade multivariado, em conjunto com espectroscopia no infravermelho por transformada de Fourier (FTIR), foi usado para detectar qualitativamente o tipo e a condição de óleos lubrificantes. O procedimento multivariado baseou-se na análise de componentes principais (PCA), primeiro para classificar o tipo do lubrificante (mineral, sintético e semi-sintético) e depois para desenvolver duas cartas de controle: uma carta T^2 usando os componentes principais mais significativos e uma carta Q com os componentes principais não utilizados na primeira carta. A partir destas duas cartas, foi possível identificar amostras de óleo, baseado no parâmetro da viscosidade, que se encontravam fora do padrão normalmente presente em lubrificantes em condições de uso.

Multivariate quality control in conjunction with Fourier transform infrared Spectroscopy (FTIR) were used to qualitatively detect the class and the condition of lubricating oils. The multivariate approach was based on principal component analysis (PCA), first to classify the lubricating oil type (mineral, synthetic and semi-synthetic) and then to develop two control charts: a T^2 chart using the most significant principal components and a Q chart with the PC not used in the first chart. From these two charts it was possible to identify oil samples, based on a viscosity parameter, which were out of the pattern normally present in utilizable lubricating oils.

Keywords: Fourier Transform Infrared Spectroscopy, lubricant control quality, Principal Component Analysis, multivariate control charts

Introduction

Lubricating oils consist of complex mixtures of hydrocarbons with molar masses in the range 250-1000 g mol[-1]. For use, they are supplied with varying amounts of different additives intended to reduce friction, wear and heat, save power and prevent corrosion. While the major components of oils are hydrocarbons, their most important characteristics are provided by the additives. These are mostly salts of organic acids and such metallic ions as zinc, barium, magnesium or calcium.

During lubricant use, contamination, loss of additive performance and an increase in oxidation products can occur.[1,2] The lubricants could have an unlimited lifetime if they are not contaminated by any kind of agents. The quality control of lubricating oil is essential for preservation of the longevity and the performance of industrial machines, automotives, and equipment which depends on hydraulic fluids. Based on this, changes in oil quality need to be detected and potential problems fixed before they become serious. Analysis of oils during use can also help prevent unnecessary replacement of oil and premature engine overhauls.

The analyses normally employed in lubricant quality control are viscosity, total base number (TBN), heptane-insolubles, water and fuel contamination levels, among others.[3] The viscosity is the most important parameter to be monitored in quality control, because an increase in viscosity can indicate the presence of insolubles, oxidation products, replacement by degraded oil or water. On the other hand, a decrease can indicate the presence of fuel, replacement by a different oil and additive breakdown.

In recent years, research has been carried out to apply FTIR for the quality control of lubricant oils. Methods for the determination of TBN,[4] water,[5] oxidation and nitration products have been developed. FTIR is a suitable tool to perform these determinations because it is generally rapid, can be automated and can reduce the need for solvents and toxic reagents associated with wet chemical methods for lubricating oil analysis.

* e-mail: ronei@iqm.unicamp.br

In this work, approaches were developed based on FTIR in conjunction with multivariate data analysis based on PCA to classify the lubricating oil type (mineral, synthetic or semi-synthetic) and to develop two control charts: a T^2 chart using the most significant principal components and a Q chart with the remaining PC.

From these two charts it was possible to identify the condition of oil samples, based on a viscosity parameter, which were out of the pattern normally present in utilizable lubricating oils. No quantitative analysis was performed as no specific quality parameter was monitored and only the viscosity was employed as basis for a priori knowledge of the oil state. In this case, an additional advantage was reached, where the arduous and time consuming viscosity measurement was changed by a simple, fast and possibly automated spectra measurement.

Principal Component Analysis

Chemometric multivariate analysis, based on PCA, was first applied to the FTIR spectra of lubricants, to evaluate its ability to differentiate the oil type (mineral base, synthetic or semi-synthetic). The PCA employed in this work is the fundamental basis of most methods in multivariate analysis.[6] It is used to investigate the correlation between variables and to explore the structure of large data sets; it organizes the data *via* condensation into a reduced number of variables, which are more easily comprehended.

PCA has been used in pattern knowledge problems,[7] where similar spectra are classified in the same class (or group) not using any a *priori* information about the data set. It was applied for evaluation of synchronous scan fluorescence spectroscopy as means of determining characteristics of mineral insulating oils used in electrical apparatus.[8]

In brief, PCA finds combinations of variables, or factors, that describe major trends in the data. It is a method used to decompose a matrix $\mathbf{X}$ of rank r into a sum of r matrices of rank 1:

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \ldots + \mathbf{M}_r \qquad (1)$$

Rank is a number expressing the true underlying dimensionally of a matrix and $\mathbf{X}$ is formed by the spectrum of different lubricants, where in rows are disposed the samples (lubricants) and in the columns the absorbance values at pre-established wavelengths. The rank 1 matrices, $\mathbf{M}_r$, can be written either as outer products of two vectors, a score $\mathbf{t}_r$ and a loading $\mathbf{p}_r$:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}'_1 + \mathbf{t}_2\mathbf{p}'_2 + \mathbf{t}_3\mathbf{p}'_3 + \ldots + \mathbf{t}_r\mathbf{p}'_r \qquad (2)$$

or in the matrix form: $\mathbf{X} = \mathbf{TP}'$, where $\mathbf{P}'$ is made up with $\mathbf{p}'$ as rows and $\mathbf{T}$ with $\mathbf{t}$ as columns.

Multivariate control charts

Normally in quality control, Shewart charts[9] are used to monitor a small number of variables to detect an event having a 'special cause' that makes the process out-of-control. This special cause can be found and eliminated, helping to improve the process and the product quality.

A univariate control chart is constructed based on a plot of sample number or time on the abscissa and three important values on the ordinate: mean value $\bar{x}$ and limits derived from the standard deviation s: warning limits ($\pm 2\omega/\sqrt{n}$) and control limits ($\pm 3\omega/\sqrt{n}$). The out-of-control process is characterized when there are unusual or nonrandom patterns in the data. The limits are usually determined by analyzing the variability in a reference set of process data collected when variabilities from 'normal' or 'commom' causes are present. Different procedures can be used to distinguish between out-of-control and in-control situations: a) one or more observations outside of the control limits; b) at least 8 observations, with values up or down the center line defined by $\bar{x}$ value; c) two or more consecutive observations outside the warning limits but still inside the control limits; d) an unusual or nonrandom pattern in the data.

A disadvantage in the use of univariate control charts appears when there are a great number of variables to be monitored, because each variable is monitored with its own control chart. In this case, many times it is impractical to control a process with univariate control charts when many variables are involved. It demands much work to check the process with these univariate charts and the probability to make mistakes is larger when several control charts need to be checked simultaneously. Also, 'out of control' samples can be missed due to correlations in the data set. To overcome this problem multivariate control charts have been used.[10-12]

In the multivariate approach, two control charts, which are based on PCA are used. The use of PCA eliminates the correlation problems, because, by definition, the new variables used in the multivariate control charts are not correlated.

The first step in building the multivariate control chart is the determination of the number of principal components, that will be used in each chart. The method adopted was based on the evaluation of the eigenvalues, that is, related to the quantity of variance explained by each principal component of the $\mathbf{X}$ matrix. The number of

principal components were chosen to explain practically all the system variance.

The significant principal components are used to construct the T^2 chart and the remaining principal components are used to construct the Q chart.

The T^2 has its origin in the work of Hotelling's and is estimated from the scores of PCA in a space formed by the most significant principal components. The T^2 is a measure of the variation in each sample within the PCA model and is defined as:[13]

$$\mathbf{T}_i^2 = \mathbf{t}_i \lambda^{-1} \mathbf{t}_i^T = \mathbf{x}_i \mathbf{P}_K \lambda^{-1} \mathbf{P}_K^T \mathbf{x}_i^T \qquad (3)$$

where $\mathbf{t}_i$ refers to the ith row of $\mathbf{T}_k$ (the matrix of K scores vectors from the PCA model), λ is the diagonal matrix containing the eigenvalues associated with the eigenvectors included in the PCA model, $\mathbf{x}_i$ is the ith row in $\mathbf{X}$ and $\mathbf{P}_k$ is the matrix of the K loadings vectors retained in the PCA model.

There are only one confidence limit for T^2, it is calculated by the F distribution and is given as:

$$T^2 = \frac{(m-1)k}{m-K} F_{k,m-k,\alpha} \qquad (4)$$

where m is the number of samples utilized to develop the PCA model, K is the number of significant principal components and α corresponds to a confidence limit region. The T^2 limit presents the multivariate distance to the data center normally found for the modeled samples in the reduced PCA space.

The Q chart corresponds to a lack of fit statistic for PCA models and is simply the sum of squares of the residual matrix ($\mathbf{E}$) of each row (sample) and it is defined as:

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T = \mathbf{x}_i (I - \mathbf{P}_k \mathbf{P}_k^T) \mathbf{x}_i^T \qquad (5)$$

where $\mathbf{e}_i$ is the ith row of $\mathbf{E}$ and $\mathbf{I}$ is the identity matrix.

Q is a scalar that measures the amount of variation not accounted for the PCA model. If the process is controlled, the Q values must be small. The confidence limit can be calculated according to:

$$Q\alpha = \Theta \left[\frac{c\alpha \sqrt{2\Theta_2 h_o^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_o (h_o - 1)}{\Theta_1^2} \right]^{\frac{1}{h_o}} \qquad (6)$$
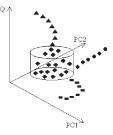
$$\Theta_i = \sum_{j=k+1}^{A} \lambda_j^i \quad \text{for } i = 1, 2, 3 \qquad (7)$$

$$h_o = 1 - \frac{2\Theta_1 \Theta_3}{3\Theta_2^2} \qquad (8)$$

where $c\alpha$ in equation 6 is the normal standard deviation corresponding to the upper $(1-\alpha)$ percentile. In equation 7, K is the number of principal components retained in the model, A is the total number of principal components and λ_j is the eigenvalue associated with the jth principal component.

Based on these two charts, different types of 'out of control' situations can be observed, as illustrated in Figure 1. In this Figure, Q is plotted against PC1 and PC2 (in this case, is supposed that 2 principal components explain the main system variance) and the cylinder includes 'in control' samples. It can be observed 'out of control' samples only in T^2 chart (blocks in Figure 1) if the process is disturbed in one or more modeled variables, without changing the model relation. For example, in spectroscopy, suppose that there is a strong increase (not normal) in a band related to a specific functional group and this band has a large correlation with another bands related to another functional groups. Due to this correlation, also the another bands have increased and the correlation structure is maintained. Secondly, it can observed 'out of control' samples only in Q chart (triangles in Figure 1) if a new process occurs, which is not covered by the model. Using the spectroscopic example, it will be the case where one band increases while another one decreases, but in normal conditions both bands had to increase, so this process is not longer covered by the model. Combinations of first and second situations described above can also be observed (circles in Figure 1) and it causes 'out of control' situation in both T^2 and Q charts.

Figure 1. Multivariate control chart. (◆) 'in control' samples; (■) 'out of control' in the T^2 chart; (▲) 'out of control' in Q chart; (●) 'out of control' in both T^2 and Q charts.

After the development of multivariate control charts from samples under control, new multivariate observations of new samples can be projected onto a hyperspace defined by the PCA loading vectors to obtain their new scores, T^2 and Q. If these new values of T^2 and Q are inside the confidence limit in the multivariate control charts the sample is 'in control', otherwise the sample is 'out of control'.

Experimental

Lubricants for gasoline motors

In the classification of the lubricating oil type, it was used new AGIP motor oils of four different types: SJ50, HD40 (both mineral based), semi-synthetic and synthetic. The semi-synthetic oil has a preponderance of mineral base (parafinic) and the synthetic base is a ester. The synthetic oil has two synthetic bases: polyolefin and ester. The aim to choice oils from the same manufacturer was to observe the distinctions between the different products. Oils from different bottles and batches were sampled to capture the variation due to oils blended/refine, feedstock and/or process. The samples were collected in triplicate and in randomized order.

The used lubricants for gasoline motors were collected using two test-cars: a GM-Kadett from the year 1990 and a GM-Astra from 1999, both made in Brazil. The same four types of oils from AGIP, described above, were used: SJ50, HD40 and synthetic collected in the Kadett and the semi-synthetic oil collected in the Astra. The samples were obtained from the motor compartment, using a hose and a syringe. The samples had been used for the following distances traveled: 11 samples of SJ50 from 375 to 5600 km, 7 samples of HD40 from 481 to 5453 km, 4 samples of semi-synthetic oil from 904 to 4000 km and 3 samples of synthetic oil from 770 to 7326 km. Also 26 samples of new oils were used in this study: 9 of SJ50, 5 of HD40, 5 of semi-synthetic and 7 of synthetic.

Lubricants for diesel motors

In the multivariate control charts application, oils for diesel motors were employed, available from the garage of a local bus company - Rapido Luxo Campinas – Brazil, which used a Shell – Rimula Plus oil for bus motors. Seven buses were monitored that travel on urban routes, from which 43 samples with 522 to 45896 km of traveling were obtained. Another 33 samples of oils, considered degraded by the bus company, were also collected and analyzed.

The criteria adopted by the bus company to discard the oils is a distance traveled of 25,000 km. However, in practice, as bus company have a big number of busses to be monitored, the garage make agendas to change the oil every 15 days. In this way, some busses travel longer distances than others and, in consequence, they exceed the limit, reaching values higher than 45,000 km.

An unused oil employed in this study, presents a viscosity at 100 °C of 14.06 cSt and a TBN of 10.1 mg KOH/g of sample.

Instrumentation

The instrument used in this study was a BOMEM MB100 FTIR. To minimize water and CO_2 interference, the instrument was purged with nitrogen. A horizontal ATR sampling accessory equipped with a ZnSe crystal was used to obtain the spectra of the oil samples. For the ATR data acquisition, 1 mL of oil sample was pipetted onto the crystal and its spectra was recorded. An air spectrum was used as reference in absorbance calculations and they were collected in the range of 650 to 4000 cm^{-1}, using 128 scans at 4 cm^{-1} resolution.

The PCA and the Control Charts were made using the Matlab 6.1 for Windows and the PLS Toolbox for use with Matlab,[14] with mean centered data.

Results and Discussions

Lubricant classification

First, a study was accomplished using only the 26 new motor oils (SJ50, HD40, semi-synthetic and synthetic) to verify possible group separations. Figure 2 shows the mid-infrared spectra of the 26 new oils, where it is possible to verify that the region in which these sets of spectra have accentuated differences is between 650 to 1300 cm^{-1}, mainly due to the absorption bands related to the additives and to hydrocarbons.

PCA modeling was developed using the spectra of the 26 oils, where three principal components described 99.1% of data variance. Based on scores plots of the first two principal components, the oils can be classified in 4 groups (SJ50, HD40, semi-synthetic and synthetic), as shown in Figure 3. In Figure 4, the loadings plot for each variable of
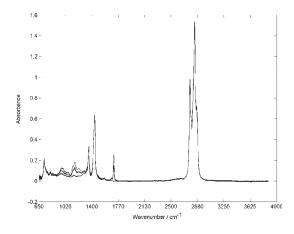

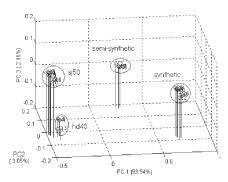**Figure 2.** FTIR spectra of 26 gasoline motor lubricant samples.

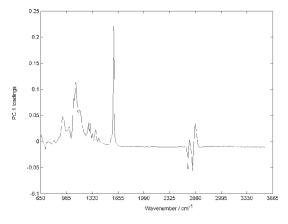**Figure 3.** Scores plot of gasoline motor lubricant samples.

Figure 4. Loadings plot of the first principal component of gasoline motor lubricant samples.

Figure 5. Scores plot of new and used diesel motor lubricant samples.

the first principal component is presented, which describes 95.4% of data variance. It is possible to note that the spectral range between 650 to 1300 cm^{-1} has the highest loadings (and the most important variables for group separation) and it is correlated with different lubricant formulations. In this spectral region we can observe absorptions due to hydrocarbons and from Zn, Ca and Mg salts of organic acids, such as alkylthiophosphate, sulphonate, phenolate.

The next step in the study was to combine the new and used oils, building a set with 51 samples. PCA was applied to this new data set and three principal components explained 94.1% of data variance. Figure 5 shows the scores plot of the first two principal components, where it is possible to observe that the first principal component (80.0% of variance explained) is responsible for separation into mineral (SJ50 and HD40 – low scores in PC1) , semi-synthetic (medium scores in PC1) and synthetic (high scores in PC1) oils. Analyzing the loadings of the first principal component, indicates that the region between
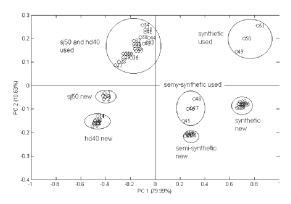
900 to 1300 cm^{-1} is the most important, and it is related mainly with the absorption bands of the additives. In this case, no separation between SJ50 and HD40 oils was verified, because the spectra are more complex, presenting absorption bands due to oxidation products in the same region (650-1300 cm^{-1}) used before to differentiate them.

Also a separation into new and used oils for mineral, semi-synthetic and synthetic oils was observed. In this case, the second principal component (10.0% of variance explained) was responsible, where low values of scores in this principal component are related to new oils, and high values to used oils. The loadings for the second principal component indicate that the most important variables (higher loadings values) are related to the absorption bands of C-O bonds of alcohols, carboxylic acids and esters, that are increased due to oxidation products and other sub-products from lubricant use.[15,16]

Multivariate control charts

The T^2 chart and the Q chart were developed based on data obtained from spectra of 43 samples collected from bus motors (considered as the control group) that have viscosities inside the range 13.28 to 15.76 cSt at 100 ºC, that is accepted as normal by Shell (manufacturer of Rimula Plus). In these samples, we can consider that only common causes of variations were present, such as distance traveled and motor conditions, but all had acceptable viscosity indices and, as consequence, they were able to use. The first step in building multivariate control charts is the correct determination of the number of significant principal components. This number was determined by examination of the variance described by each principal component. Three principal components were chosen, because 90.8% of data variance is described and little or no significant variation is observed after this number.

The multivariate control charts were then developed based on 3 principal components. The T^2 chart used the most significant principal components (three in this case) and the remainder were used to build the Q chart. In the calculations of T^2 and Q, equations 3 and 5 were used and the control limits for these multivariate control charts were found using equations 4 and 6.

The control limit for the T^2 chart was 8.4 and, for the Q chart, it was 0.0021 with 95% confidence. All samples used in the development of the model fell within the control limits, denoting no abnormal variations among the oils.

The behavior of the other 33 samples (considered as the test group) that had been discarded by bus company, could be referenced against the samples of the control group. The samples of the test group were projected onto loadings vectors found with the model developed from the samples of the control group and their scores, T^2 and Q were obtained.

The T^2 and Q values could be compared with control limits (95% confidence) established in the multivariate control charts and, therefore, 'in control' and 'out of control' samples could be visualized in these plots. These results are shown in Figures 6 and 7, where the control group samples used for model development are numbered 1 to 43 and the test group from 43 to 76. A sample is considered 'out of control' when its T^2 or Q values were found outside the control limit. In Figure 6, the T^2 chart shows that samples 48, 49, 51, 60, 62 and 63 are outside of the control limit and the Q chart (Figure 7) shows that samples 49, 51 and 69 have high residuals and they also are outside of the control limit.

By using the information of both control charts, only 7 samples (samples number 48, 49, 51, 60, 62, 63 and 69) had abnormal behaviors, based on the their FTIR spectra and considering as prior information the viscosity parameter. For those samples considered 'out of control', the sample number 49 has viscosity value of 18.86 cSt, the sample 51 has 18.77 cSt, the sample 60 has 15.78 and the other four samples have values around 15.00 cSt. All samples considered 'in control' have viscosity values below the specified limit by manufacturer.

Conclusions

In this paper the feasibility of the use of Fourier transform infrared spectroscopy in conjunction with multivariate statistics, based on PCA, is demonstrated to develop a quality control strategy for classification of lubricant type and usage conditions. The procedure is simple, fast and adaptable to oil process monitoring and requires only infrared spectra acquisition of a set of patterns or normal samples to develop the control charts. In a bus company, that has a lot of busses to be monitored, instead
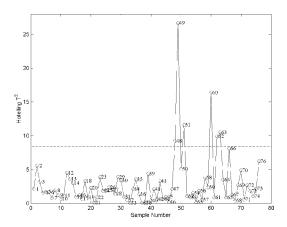

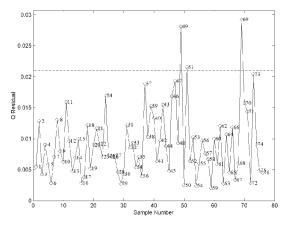**Figure 6.** T^2 chart of diesel motor lubricant samples.

Figure 7. Q chart of diesel motor lubricant samples.

of change the oil after a specific time or distance traveled, they could be tested by the proposed methodology and used longer, getting extra life and saving money.

By using PCA, it was possible to carry out a clear separation of the motor lubricant types into mineral, semi-synthetic and synthetic, as well as, into new and used oils.

The multivariate control charts were capable of identifying 7 'out of control' samples, of the 33 initially considered degraded by bus the company. These samples considered 'out of control' really presented viscosity values near or above the limits established by the manufacturer.

Acknowledgments

References

1. Wiseman, M.; Ah-Sue, A.; *Lubric. Eng.* **1992**, *48*, 236.

2. Dong, J.; Voort, F.R.V.; Ismail, A.A.; Koblé, E. A.; Pinchuk, D.; *Lubric. Eng.* **200**, *56*, 12.

3. Stuart, A. D.; Trotman, S.M.; Doolan, K.J.; Fredericks, P.M.; A*ppl. Spectrosc.* **1989**, *43*, 55.

4. Kauffman, R. E.; *Lubric. Eng.* **1998**, *54*, 39.

5. Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; González, R.; *Mikrochim. Acta* **1998**, *128***,** 235.

6. Seasholtz, M. B.; *Chemom. Intell. Lab. Syst.* **1999**, *45*, 55.

7. Massart, D. L.; Vandeginste, B. G. M.; Buydens; L. M.; de Jong, C.S.; Lewi, P. J.; Smeyers-Verbeke, J.; *Handbook of Chemometrics and Qualimetrics: Part B,* Elsevier: Amsterdam,1998.

8. Bessant, C.; Ritchie, L.; Saini, S.; Vanpour, B. P.;Wilson, G.; *Appl. Spectrosc.* **2001**, 55, 840.

9. Montgomery, D. C.; *Statistical Quality Control: An Introduction*, 2nd ed., Wiley: New York, 1991.

10. Barthus, R. C.; Poppi, R. J.; S*pectrosc. Lett.* **2002**, *35***,** 729.

11. Nijhuis, A.; Jong, S; Vandeginste,B.G.M.; *Chemom. Intell. Lab. Syst.* **1997**, *38*, 51.

12. Nijhuis, A.; Jong, S; Vandeginste,B.G.M.; *Chemom. Intell. Lab. Syst.* **1999**, *47*, 107.

13. Johnson,A.R.; Wichem,W.D.; *Applied Multivariate Statistical Analysis*, 4th ed., Prentice Hall: Upper Saddle River, 1998.

14. Wise, B. M.; Gallagher, N. B.; *PLS Toolbox for Use with Matlab,* ver. 2.1*,* Eigenvector Technologies: Manson, 1998.

15. Nakanishi, H.; Onodera, K.; *Lubric. Eng.* **1997**, *53*, 29.

16. Rizvi, S. Q. A.; *Lubric. Eng.* **1999**, *55*, 33.