JOURNAL OF OPTOELECTRONICS AND ADVANCED MATERIALS Vol. 14, No. 1-2, January – February 2012, p. 1 - 7

IR-stimulated Raman spectroscopy for advanced gear-oil analysis

D. G. DORIGO*, B. R. WIESENT, A. PÉREZ GRASSI, Ö. ŞIMŞEK, A. W. KOCH
Institute for Measurement Systems and Sensor Technology, Technische Universität München,
Theresienstr. 90 / N5, D-80333, Munich

In this work, we present a new approach for gear-oil analysis, based on Raman spectroscopy and subsequent multivariate data analysis. The scattered Raman signal was collected by a Fourier Transform Infrared (FTIR) Spectrometer. The results of the multivariate data analysis applied to the collected spectra are compared to the ones obtained by analysing the IR-Spectra of the same samples. Correlations with different parameters, e.g. Total Acid Number (TAN) and element concentrations, are performed and discussed.

(Received June 10, 2011; accepted February 20, 2012)

Keywords: Oil Analysis, Raman, IR, Spectroscopy

1. Introduction

The goal set by some European governments is to increase the contingent of renewable energy up to 30 percent within 2020. Therefore, offshore wind parks will become more and more important. In order to reduce the enormous maintaining costs of such parks, online condition monitoring systems are currently being developed. Especially condition monitoring of gear-oils is a promising approach.

By now, gear-oils of such facilities are changed within fixed intervals. For reduction of costs and environmental pollution, on-demand oil changes would be much more reasonable. This latter requires in-process oil monitoring. In this paper, we propose a method for gear-oil evaluation based on Raman spectra analysis. The results are compared to those obtained from the well-known evaluation of Infrared (IR) spectra.

Raman spectroscopy [1] is well known in chemistry and well-established as a non-destructive technique for substance analysis at molecular level. Due to the Raman effect, the scattered light of a laser beam focussed on a sample shows a frequency shift with respect to the excitation light source. This is caused by the interaction of the laser beam with the molecules of the analysed sample. The frequency shift can be either positive or negative, depending on the energy level of the molecules after excitation [2].

Important oil parameters can be deduced using multivariate data analysis from different kinds of spectra, like Mid-IR and Near-IR ones [3]. These parameters provide valuable information about the quality and the properties of the oil under test and its additive depletion state. One of the most important parameters is the Total Acid Number (TAN), a measure of the analysed sample acidity [4]. The importance of the TAN value comes from the fact that it is a proxy variable for oil “age”.

Using multivariate data analysis on IR spectra, it is possible to predict the most important oil parameters which "mirror" the oil "age" [5]. In this paper, multivariate data analysis is applied on gear-oil Raman spectra. The results are then compared to those obtained from IR Spectra.

2. Multivariate data analysis

The establishment of spectroscopy techniques in chemical process monitoring has significantly increased the notoriety of multivariate data analysis as a tool for extracting relevant information from huge amount of data. Depending on the objective of the study, different multivariate analysis methods can be employed in order to extract the information of interest. This information can be used to perform a classification and to understand the interrelationships between the involved parameters [6]. Linear regression methods, like Principal Component Regression (PCR) and Partial Least Square Regression (PLS), are widely diffused in pharmaceutical industry and foodstuffs monitoring [3, 7]. In the case of PCR, the relationship between an independent and a dependent variable is estimated taking into account only the last one. Contrary to PCR, PLS estimates this relationship considering both, the independent and the dependent variable [3]. In this work, the applicability of one dimensional PLS (PLS1) regression on Raman Oil-Spectra is presented.

A major application of PLS1 is the analysis of causal relationships between a dependent variable y and an independent variable x. The relationship is modelled by linear function given by: $y = f(x)$. The generation of such model is called calibration. During the calibration, a set $S = \{(x_i, y_i), 1 < i < n\}$ of n known samples x and their corresponding y values is used to generate the model. Such a set is called training set, while the known y values are called reference values. From the known pair of values

(x_i, y_i) of the training set a linear model $\hat{y} = f(x)$ is calculated, so that from each sample x the corresponding value of y can be approximated. If the samples in S are selected correctly, the model works also for finding the value of $\hat{y}$ from samples x that are not included in S. The approximated/predicted value of y is denoted by $\hat{y}$.

A common example for a *classical calibration* is given by determining the relationship between TAN values of oil samples (independent variable x) and the corresponding absorption at a specific wavelength (dependent variable y). In this case, the measured absorptions are regressed on the TAN values based on a given training set S. The parameters of the model function $\hat{y} = f(x)$ are calculated by using a least square method.

In order to obtain better prediction results, it is convenient to use an *inverse calibration* [3]. For an inverse calibration the variables x and y are exchanged. In this case, the TAN becomes a function of the absorption. The linear function $\hat{y} = f(x)$ can be described as follows:

$$\hat{y} = b_1 x + b_0, \tag{2.1}$$

where $\hat{y}$ is the modelled (or predicted) TAN value, x is the measured absorption value, b_1 the slope of the model and b_0 its intercept. For the described model, only single values x, given by the absorptions measured at a single wavelength, were used. This kind of regression is called *univariate regression*. By using spectra as an independent variable $\boldsymbol{x}$, instead of the single absorption values, a *multivariate regression* is obtained. Equation (2.1) becomes:

$$\hat{y} = \boldsymbol{x}^T \boldsymbol{b}_1 + b_0. \tag{2.2}$$

here, $\hat{y}$ is the predicted TAN value, $\boldsymbol{x}$ is the measured spectrum, $\boldsymbol{b}_1$ is the vector of the regression coefficients and b_0 is the static offset of the regression line describing the model. Both model parameters $\boldsymbol{b}_1$ and b_0 are calculated using *c* PLS-Components [3, 8, 9].

An important aspect is the model quality. In order to evaluate the model, this latter is applied to all spectra $\boldsymbol{x}$ in S, and the resulting predicted TAN $\hat{y}$ are compared with those given as reference y. By comparing these two values, a regression line can be defined as follows:

$$\widehat{y'} = b'_1 x + b'_0. \tag{2.3}$$

An ideal model is characterised by b'_1 equal to one and an intercept b'_0 equal to zero. Together with this regression line, two parameters [3] are especially relevant to measure the quality of the model: the Root Mean Square Error:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{2.4}$$

and the coefficient of determination:

$$\mathrm{r}^2 = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}}, \tag{2.5}$$

where y_i and $\hat{y}_i$ are the reference and the predicted TAN values, respectively, while $\bar{y}_i$ is the mean value of the reference TAN values. A good model is characterised by a low RMSE and r² close to one. Beside this, it is further important to keep in mind that the calculated linear relationship is only a hypothesis that has to be proved. This latter requires a good knowledge of the analysed process. For this work, inverse calibration and regressions based on the NIPALS algorithm [8, 10] were used. Data analysis was performed using the software *The Unscrambler®* (CAMO Software).

3. System description

In this work, the analysed Raman spectra were detected using the commercially available FT-Raman spectrometer Bruker MultiRAM. The setup of the spectrometer is illustrated in Fig. 1.

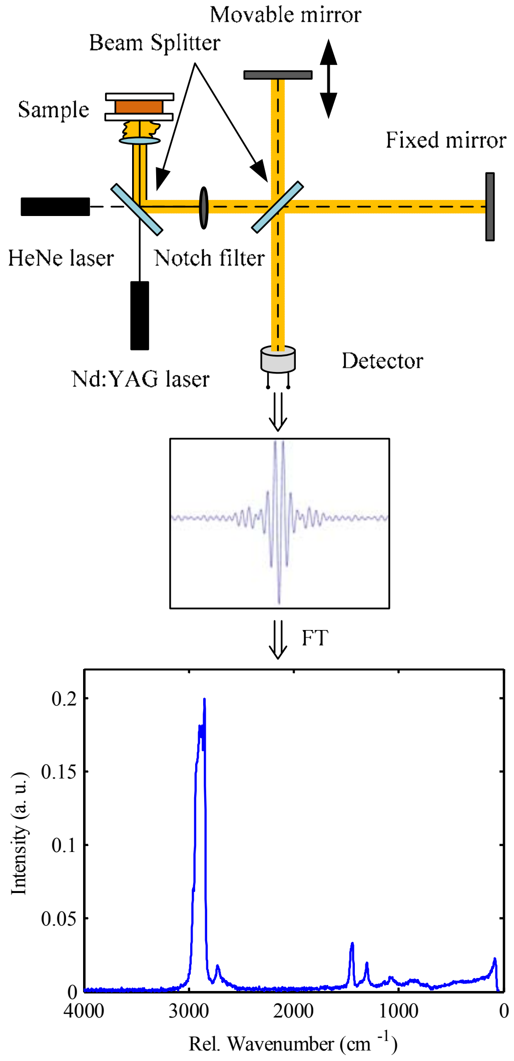


Fig. 1. The measurement setup of the used FT-Raman-Spectrometer is shown. The sample was excited by a Nd:YAG laser (1064 nm) with an output optical power of 800 mW. A Raman spectrum of an oil sample is shown on the bottom.

The spectrometer has been selected because of the source wavelength of 1064 nm. This wavelength is necessary in order to reduce fluorescence [11]. By exciting the samples with wavelengths in the visual spectral range, e.g. 532 nm, strong fluorescence overlap the weak Raman signal making an analysis of Raman spectra extreme difficult [12]. Choosing an excitation wavelength in the Near Infrared (NIR) region, e.g. 1064 nm, the fluorescence is reduced approximately to zero because of the lower energy of the exciting light quanta [12]. On the other hand, also the Raman signal intensity decreases drastically [13] and high detector sensitivity is required. For this reason, the spectrometer is equipped with a liquid nitrogen cooled germanium detector. In order to maximize the signal intensity without overheating the sample, the optical output power of the laser was set to 800 ± 20 mW. The power regulation was performed using polarisation filters.

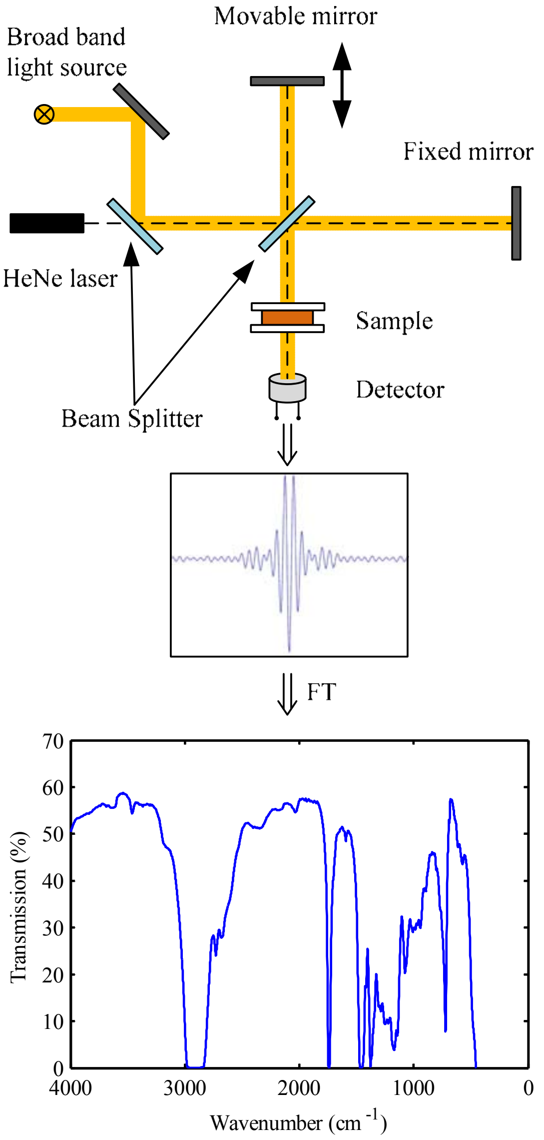


Fig. 2. The measurement setup of the used FTIR spectrometer is shown. The broad band light source has a similar spectral characteristics to the one of a blackbody. An IR spectrum of the oil sample illustrated in Fig. 1 is shown on the bottom.

The Raman spectra were detected with a resolution of 4 cm^{-1} over a spectral range between 50 cm^{-1} and 4000 cm^{-1}. The mirror position in the Michelson interferometer was adjusted using a HeNe laser.

In order to allow a good comparison between the results obtained by analysing the Raman spectra and those obtained by analysing the IR spectra, the same samples were investigated using both systems. Fig. 2 shows the setup of the used FTIR spectrometer. A similar Michelson disposition can be seen. The IR spectra were detected with a resolution of 1 cm^{-1} over a spectral range between 500 cm^{-1} and 4000 cm^{-1}.

Two commercially available synthetic gear-oils have been analysed. These will be named Oil-1 and Oil-2. Oil-1 and Oil-2 are both poly-alpha-olefin (PAO) oils with Oil-1 having a higher ester component than Oil-2.

4. Partial least square regression applied on raman oil spectra

The detected Raman spectra were analysed by means of multivariate data analysis without previous signal processing. For model generation two training sets, one for each oil type (Oil-1 and Oil-2), were used. Each set has $n = 26$ samples with a uniform distribution of the TAN values. In order to determine the TAN reference values y, potassium hydroxide (KOH) is added to the oil samples until its pH values turns to neutral. The amount of added KOH gives quantitative information about the oil acidity level [4].

The TAN values for Oil-1 are uniformly distributed between 0.3678 mgKOH/g (milligram potassium hydroxide per oil gram) and 1.3956 mgKOH/g. The TAN values of Oil-2, however, are uniformly distributed between 2.6208 mgKOH/g and 3.6435 mgKOH/g. The use of a uniformly distributed training set is very important for the model quality. This is because the error between measured and predicted values increases with increasing distance from the model centre point [3]. Both training sets were also used to generate models for other parameters like viscosity and element concentrations (iron, copper, etc.). The results in these cases are not optimal, because the training set was selected with a uniform distribution of TAN values. The latter results in a reduced model quality. However, the prediction feasibility of the other studied parameters could be examined.

In order to increase the model quality, multivariate data analysis has not been applied to the entire measured spectrum but only to those regions containing useful information in terms of peaks. For both oil types, these regions are defined between 3110 cm^{-1} - 2555 cm^{-1} and 1765 cm^{-1} - 698 cm^{-1}.

In order to generate the training set S, all samples of Oil-1 and Oil-2 are measured in order to obtain their Raman spectra and the reference values y of the parameters under study. The resulting training sets are used to generate a model for each parameter according to equation (2.2). Once the model is found the regression line $\widehat{y'}$ is calculated according to equation (2.3) together with the

RMSE and r^2 values. The number of PLS components used for the PLS1 analysis was chosen in order to obtain the best model quality according to quality parameters. This procedure was applied on both oil types, Oil-1 and Oil-2. Different parameters were tested, but not all suited for a regression.

5. Partial least square regression applied on IR oil spectra

The transmittance IR spectra (T) of the same sample set as used for the Raman models were detected using the setup shown in Fig. 2. These have been converted to absorbance spectra (A) as follows [14]:

$$\mathrm{A} = \log_{10} 1/\mathrm{T}. \tag{5.1}$$

For quantitative analysis of absorbance spectra, the Bouguer–Beer–Lambert law [14], also known as Beer's law, has to be respected. This law states that the component's concentrations in a homogeneous solution is linearly proportional to the intensities of the corresponding absorption peaks until the absorbance level is below 2.5 [5, 14]. In order to keep the relationship linear, the following three regions were excluded for Oil-1: 3067 cm^{-1} - 2750 cm^{-1}, 1780 cm^{-1} - 1680 cm^{-1} and 1500 cm^{-1} - 1350 cm^{-1}. For Oil-2, the excluded regions were 3067 cm^{-1} - 2750 cm^{-1} and 1500 cm^{-1} - 1350 cm^{-1}.

The IR spectra and the corresponding reference values were then used to generate linear models, as explained in the section above.

6. Results

Table 1 shows the quality parameters b'_1, b'_0, RMSE, r^2 and the number of PLS components c for the modelled parameters of Oil-1. Models for TAN, viscosity and different element concentrations, like iron, copper etc., were calculated. The TAN values are measured in mgKOH/g, the element concentrations in mg/kg (mg element per kg of oil), and finally the viscosities (V40 and V100) in mm^2/s. V40 is the viscosity measured at 40 °C and V100 is the one measured at 100 °C.

Table 1. Quality parameters for the calculated models of Oil-1. The models were obtained by multivariate data analysis applied on Raman spectra.

y	b'_1	b'_0	RMSE	r^2	c
TAN	0.663	0.297	0.177	0.964	4
Fe	0.669	8.456	11.866	0.870	4
Cu	0.187	1.533	1.627	0.612	4
Si	0.127	-8.881	7.004	0.692	3
P	0.710	102.96	42.645	0.986	4
Zn	0.230	7.411	6.893	0.692	3
S	0.318	2524.6	218.33	0.997	3
V40	0.793	66.538	6.985	0.999	3
V100	0.751	9.044	0.832	0.999	3

The analysis shows that for Oil-1 the concentrations of copper, silicon, zinc and sulphur do not suit for prediction by multivariate data analysis (see Table 1). The parameter which suits best is the viscosity measured at 40 °C. It has to be considered that the training set was chosen in order to guarantee a uniformly distributed TAN. Due to this, only for the TAN a minimal model error could be calculated. Using training sets especially selected for the other parameters, the model quality could be improved.

Table 2. Quality parameters for the calculated models of Oil-1. The models were obtained by multivariate data analysis applied on IR spectra.

y	b'_1	b'_0	RMSE	r^2	c
TAN	0.925	0.0661	0.086	0.991	4
Zn	0.032	6.485	6.462	0.729	3
V40	0.696	97.680	5.351	0.999	4

For a better evaluation of the obtained results, a comparison to multivariate data analysis applied on IR spectra was performed. The IR spectra are obtained from the same sample set of Oil-1 used for the Raman spectra analysis. Table 2 shows quality parameters for TAN, zinc concentration and viscosity measured at 40 °C. It can be seen that also in this case zinc concentration does not suit for regression. For TAN, a better model than the one obtained for the Raman spectra is achieved. For V40, however, the model based on Raman spectra fits better.

An absolute comparison between the RMSE values of the different models cannot be performed because of the different scales of the analysed parameters (see Figure 3). In this case, the RMSE values should be normalised on the input parameter values.

Table 3. Quality parameters for the calculated models of Oil-2. The models were obtained by multivariate data analysis applied on Raman spectra.

y	b'_1	b'_0	RMSE	r^2	c
TAN	0.797	0.636	0.128	0.998	4
Ca	0.328	13.341	3.100	0.976	4
Mo	0.775	261.64	142.45	0.986	6
Mg	0.464	826.55	139.95	0.992	5
P	0.805	276.83	58.162	0.998	6
Zn	0.804	309.89	71.986	0.998	6
S	0.407	3396.4	315.07	0.998	3
V40	0.546	148.45	9.728	0.999	4
V100	0.528	15.82	1.030	0.999	4

In Fig. 3 the regression line $\widehat{y'}$ together with the $(y, \hat{y})$ pairs for TAN, zinc concentration and viscosity of Oil-1 are plotted. In each plot, the model calculated by multivariate data analysis applied on Raman spectra is compared with the one obtained from IR spectra. In all cases, the same samples were used. It can be seen that the model quality for TAN is better when calculated on IR spectra.

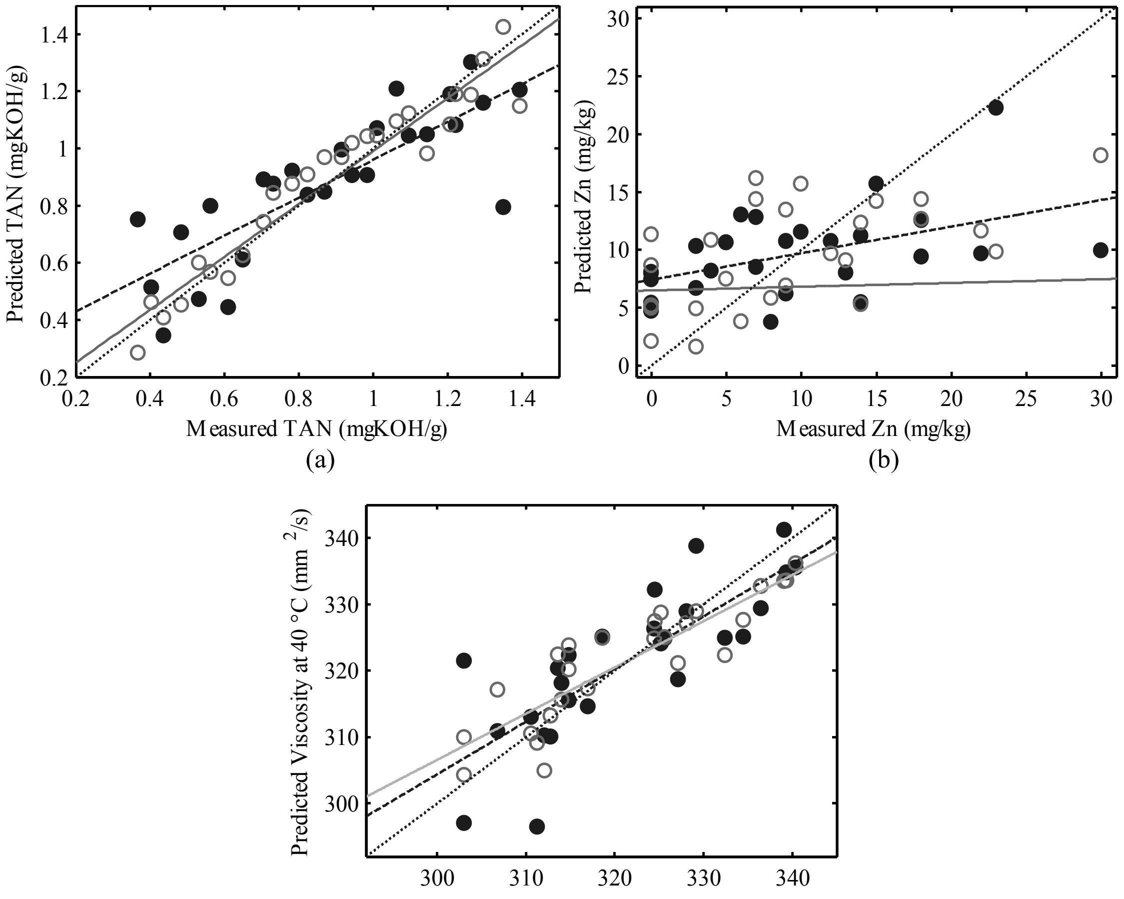
Zinc concentration does not suit for multivariate data analysis in both cases. For the viscosity measured at 40 °C, the prediction model calculated using Raman spectra fits better. The quality parameters concerning the plots in Fig. 1 are displayed in Table 1 and Table 2. It is evident that the prediction error increases with increasing distance from the model centre point.

Similar to Table 1, Table 3 shows different model quality parameters for Oil-2. The models were obtained by analysing the Raman spectra. The results show that in this case calcium, magnesium and sulphur concentration do not suit for multivariate data analysis. The results for the viscosities are worse than for Oil-1. The parameters which suited best were phosphor and zinc concentration. Also for Oil-2, the samples were chosen in order to achieve a uniformly distributed TAN.

Table 4. Quality parameters for the calculated models of Oil-2. The models were gained by multivariate data analysis applied on IR spectra.

y	b'_1	b'_0	RMSE	r²	c
TAN	0.889	0.348	0.102	0.999	4
Zn	0.978	34.235	44.556	0.999	2
V40	0.494	165.45	9.013	0.999	2

Like for Oil-1, the three selected Oil-2 parameters were analysed by multivariate data analysis on IR spectra. The tested samples are the same as used for Raman analysis. The results are displayed in Table 4. Here, the model quality for TAN and silicon concentration are better than those obtained from Raman spectra. However, the model for viscosity measured at 40 °C based on Raman spectra is better than the one based in IR spectra.



(a)

(b)

(c)

Fig 3. Comparison between models calculated on Raman and IR spectra for Oil-1. (a) displays the TAN models, (b) the zinc ones and finally (c) shows the V40 (viscosity measured at 40 °C) models. The solid circles represent the measured versus predicted data for the Raman based models. The gray circles represent the data based on the IR models. The dashed black lines are the regression lines for the Raman models, the gray lines are the regression lines for the IR models and the dotted black lines are the ideal regression lines. See Table 1 and 2 for the corresponding quality parameters.

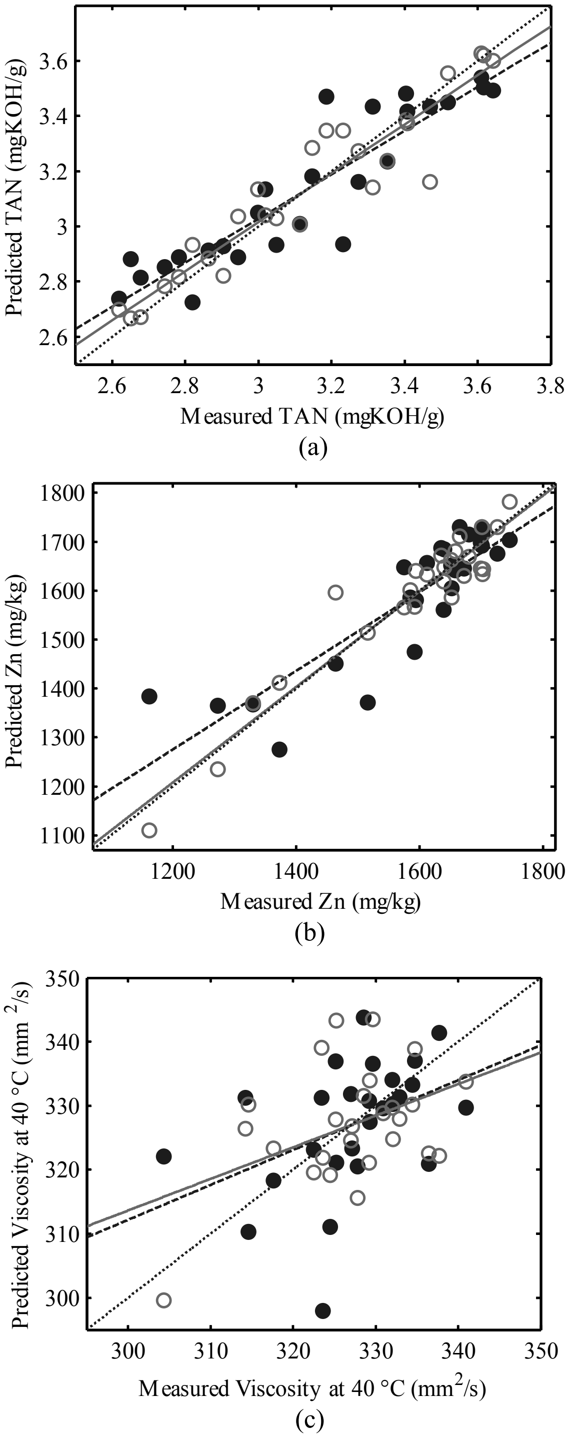


Fig. 4. Comparison between models calculated on Raman and IR spectra for Oil-2. (a) displays the TAN models, (b) the zinc ones and finally (c) shows the V40 (viscosity measured at 40 °C) models. The solid circles represent the measured versus predicted data for the Raman based models. The gray cilcles represent the data based on the IR models. The dashed black lines are the regression lines for the Raman models, the gray lines are the regression lines for the IR models and the dotted black lines are the ideal regression lines. See Table 3 and Table 4 for the corresponding quality parameters.

In Fig. 4, the regression line $\widehat{y'}$ together with the $(y, \hat{y})$ pairs for TAN, zinc concentration and viscosity of Oil-2 are plotted. A comparison between models gained by Raman and models gained by IR spectra is given. Also in these plots it is clear that the prediction error increases with increasing distance from the model centre point.

7. Conclusion

In this work two commercially available synthetic gear-oils on PAO basis have been analysed by means of multivariate data analysis. The prediction models have been calculated using Raman spectra. TAN, element concentrations and viscosity were modelled and predicted. The used training sets were chosen in order to guarantee a uniformly distributed TAN. Due to this fact, only for TAN an optimal model could be achieved. However, the prediction feasibility of the other studied parameters could be examined. Furthermore, a comparison to multivariate data analysis models built on IR spectra was performed.

We showed that oil parameter prediction by means of multivariate data analysis applied on Raman spectra is possible. The analysis on Raman spectra validated the results achieved by analysis of IR spectra. Moreover, for some parameters, the results were better than those obtained from the IR spectra. The obtained results based on Raman spectra can be improved for a given parameter by generating the model from a training set, which presents a normal distribution of such a parameter. This work shows that multivariate data analysis on Raman spectra not only validate the results of such analysis on IR spectra; but also improve the prediction accuracy for certain parameters. This motivates further work in this area in order to determine which other parameters could be predicted through Raman spectra significantly better as with IR spectra.

Acknowledgments

The authors would like to thank Oelcheck GmbH for providing the oil samples and the correspondent analysis reports. Furthermore, special thanks go to Dr. Dave Wooton (Wooton Consulting) and Dr. Siegfried Lochschmidt (Klüber Lubrication München KG) for the valuable discussions. We would also like to thank the Department of Inorganic Chemistry – Prof. Kornath Research Group – Ludwig Maximilians Universität München, for allowing us to perform measurements with their Raman spectrometer. Funding by “Bayerisches Statsministerium für Wirtschaft, Verkehr und Technologie” and administrative support of VDI/VDE are gratefully acknowledged. This work was also supported by the TUM Graduate School.

References

[1] N. B. Colthup, L. H. Daly, S. E. Wiberley, Introduction to infrared and Raman spectroscopy (3rd ed.). Boston: Academic Press (1990).

[2] D. A. Long, The Raman effect: a unified treatment of the theory of Raman scattering by molecules. John Wiley & Sons Inc. (2002).

[3] W. Kessler, Multivariate Datenanalyse in der Bio-und Prozessanalytik. Weinheim: Wiley-VCH (2006).

[4] D. M. Pirro, A. A. Wessol, J. G. Wills, Lubrication Fundamentals. World Wide Web Internet And Web Information Systems (Second Edi.). CRC. (2001).

[5] B. R. Wiesent, D. G. Dorigo, M.Schardt, A. W. Koch, Gear oil condition monitoring for offshore wind turbines using band limited low resolution spectra. Proceedings of Oildoc. (2011).

[6] K. Backhaus, B. Erichson, W. Plinke, R. Weiber, Multivariate Analysemethoden (12th ed.). Springer. (2008).

[7] P. Geladi, Journal of Chemometrics, **2**(4), 231 (1988).

[8] P. Geladi, B. R. Kowalski, Analytica Chimica Acta, **185**(1), 1 (1986).

[9] S. Wold, Chemometrics and Intelligent Laboratory Systems **58**(2) 109 (2001).

[10] Y. Miyashita, T. Itozawa, H. Katsumi, S. I. Sasaki, Journal of chemometrics, **4**(1), 97 (1990). Wiley Online Library. Retrieved May 23, 2011.

[11] J. R. Lakowicz, (Ed.). (2006). Principles of fluorescence spectroscopy (3rd ed.). Boston: Springer Science+Business Media, LLC.

[12] B. Schrader, Raman Spectroscopy of Mineral Oil Products. Part I: NIR/FT-Raman Spectra of Polycyclic Aromatic Hydrocarbons. Applied spectroscopy **45(**8), 1230 (1991).

[13] M. Fujiwara, H. Hamaguchi, M. Tasumi, Applied Spectroscopy, **40**(2), 137 (1986).

[14] F. A., Settle, ed. 1997. Infrared Spectroscopy. In Handbook of Instrumental Techniques for Analytical Chemistry, 247-284. Prentice Hall PTR, May 11.

---

*Corresponding author: d.dorigo@tum.de