

## **Artificial Intelligence and Fourier-Transform Infrared Spectroscopy for Evaluating Water-mediated degradation of Lubricant Oils**

Christian Chimeno-Trinchet, Clarissa Murru, Marta Elena Díaz-García, Alfonso Fernández-González\*, Rosana Badía-Laiño.

Department of Physical and Analytical Chemistry, University of Oviedo, Oviedo, Asturias, 33006, Spain.

\*corresponding author: fernandezgalfonso@uniovi.es

### **Abstract**

The presence of water in lubricant oils is a parameter related to the lubricant deterioration, which can be indicative of a serious loss of tribological efficiency and, therefore, an increase in maintenance costs. Likewise, controlling the aging of the lubricant oil is a keynote issue to prevent damage on the lubricated surfaces (e.g. engine pieces). The combination of Attenuated Total Reflectance (ATR) techniques with Fourier-Transform Infrared Spectrometry (FTIR) result in an easy, simple, fast and non-destructive way for obtaining accurate information about the actual situation of a lubricant oil. The analysis of this ATR-FTIR information using Artificial Neural Networks (ANN) as well as Linear Discriminant Analysis (LDA) results in the proper classification of lubricant oils regarding the presence/absence of water, age and viscosity. The methodology proposed in this work describes procedures for identifying the deterioration degree of oils with as high as 100% success (aging week) or 97.7% (for viscosity and water presence).

### **Keywords**

FTIR, lubricant oil aging, linear discriminant analysis, artificial neural networks, LDA, ANN

### **Introduction**

Friction causes important losses of money. It is estimated that friction and related phenomena in industrialized countries cost up to €450 billion annually, despite the wide use of lubricants [1]. According to Webster dictionary, a lubricant is “a substance (such as grease) capable of reducing friction, heat, and wear when introduced as a film between solid surfaces”[2]. The idea behind a lubricant, therefore, is the creation of a film between the sliding parts of an engine (or other moving device), filling the space between the surfaces and keeping them apart. A lubricant material has to achieve some objectives: it has to improve machine-performance by reducing mechanical energy, withstand high temperatures, maintain its viscosity and clean up the impurities that are generated during the sliding of mechanical parts.

Basically, a lubricant consists in a base oil and a series of additives, which depend on the use the lubricant is designed for. Currently, three main types of base oils that differ from each other by

their origin can be considered: mineral oils, synthetic oils and semi-synthetic oils. i) Mineral oils that proceed from the fractional distillation of crude oil and are elaborated through multiple processes that generate adequate products to form base oil. ii) Synthetic lubricants (such as poly- $\alpha$ -olefins) can be manufactured using chemically modified petroleum components or synthetic esters produced from other raw materials. Synthetic oils are used in those applications with specific equipment demands that conventional mineral oil do not fulfil and when economic benefits are offered [3]. iii) The composition of a semi-synthetic lubricating oil comprises a major portion of a synthetic lubricant or a mixture of a synthetic lubricant plus a hydrocarbon mineral base oil, and a minor portion of various additive components [4]. Mineral oils are one of the main liquid lubricants used in industrial machinery and in automotive industry with an annual expenditure of approximately 9000 million liters [5].

Corrosion and degradation behavior of oil lubricants is critically related to water content. Water contamination in lubricant oils may coexist in a combination of forms: free water, dissolved in oil and emulsified. Free water is that which exists in excess of its equilibrium concentration in solution. Dissolved water is simply water in solution and its concentration in oil depends on the oil hygroscopic properties, temperature and humidity. Water out of the limits imposed by these conditions is free water. If the free water content increases the lubricant oil becomes saturated with enough water molecules to a point that water is suspended in micro/nanoscale droplets giving rise to an emulsion, often undesirable. Free water and emulsified water are the most damaging of all water-lubricant mixtures. So, free water may settle on machinery surfaces displacing any protective lubricant film, thus contributing to the corrosion of the surface [6]. On the other hand, emulsified lubricants exhibit reduced load carrying capacity due to a change in the compressibility of the oil. Besides, the emulsified water has a tendency to trap dirt and particulates creating a sludge that may grind the contacting surfaces creating more particulates. The consequent lubricant failure is then followed by permanent damage to the operating surfaces [7,8]. Besides, water contamination promotes chemical and physical changes in the lubricant oil properties. In fact, water plays a key role in the increase of lubricant aging rate, depletion of additives and destruction of base oils causing acid formation [9,10]. In some instances, there is a need to replace the lubricant oil or use demulsifiers. For those applications where the lubricant must be regularly replaced, there is a direct relationship between lubricant deterioration and consumption, along with the environmental impact of lubricant disposal [11]. Consequently, if an adequate control of the water content and lubricant stability is made before it reaches the point of causing corrosion or losing effectiveness, a better lubricant economy, a reduction in friction losses and hence decreased environmental hazards are possible [12,13].

The development of effective and efficient methods to assess the aging of lubricant is important to the lubrication industry. It is desirable that the methods do not require sample preparation and

produce rapid results and information on multiple parameters simultaneously. Among the techniques that provide functional-group information Fourier Transform Infrared Spectroscopy (FTIR) allows to ascertain their presence in a sample, as each group absorb in defined wavelength regions. So, chemical changes taking place in lubricant oils over time are associated with the formation of new or loss of particular functional groups due lubricant oil aging. FTIR is rapid and non-destructive technique. However, FTIR is not sufficient to provide a plausible differentiation criterion if chemical mapping is applied to large data sets that include spectra from many lubricant oils. In the last ten years, research has been carried out to apply FTIR combined with chemometrics to address this shortcoming by analyzing large amounts of spectral information. The viscosity index (VI) and the base number (BN) of motor oils were successfully determined using the FTIR analyses along with the PLS-1 calibration. Gracia et al. applied FTIR spectroscopy combined to Principal Component Analysis (PCA) in order to extract chemical information during oil oxidation process in the presence of iron as a catalyst. Results demonstrated that the presence of iron did not lead to significant change in the global chemical composition as revealed by the first principal component. However, the second principal component indicated clearly that the initial formation of alcohols and esters was favored by the presence of iron. The influence of iron was highlighted by PCA analysis of on-line FTIR data [14]. In other study, Nguele et al. [15] studied the depletion of additives in lubricating oils due to their degradation through a series of chemical reactions which resulted in loss of their primary functions. To tackle the subject, authors used FTIR-ATR spectral data combined with a curve fitting technique and mathematical models, that describe the behavior of additives within the engine. Results showed that depletion of additives followed an exponential regression rather than polynomial and that the chemical breakpoint (the initiation of deterioration of additives) depended on the composition of the base stock. The breakpoint was found to be two times higher in a fully synthetic model lubricating oil than a semi synthetic model one. In 2018, Hossain et al. [16], applied FTIR in connection with Artificial Neural Network (ANN), Principal Component Regression (PCR) and Partial Least-Square Regression (PLSR) for determination of VI of motor oils. Results showed that among the calibration techniques studied, PLSR provided the best prediction results with Savitzky-Golay smoothed FTIR spectral data, the method requiring shorter turnaround times and lower expenses than conventional approaches.

In this work, we compared the performance of non-linear (ANN) and linear calibration techniques (Linear Discrimination Analysis, LDA) for prediction of the aging degree of lubricant oils from Attenuated Total Reflectance-Fourier-Transform Infrared Spectroscopy (ATR-FTIR) spectra. The model was created for four important lubricant oils properties: water content, viscosity, oxidation and time of the experiment. ANN demonstrated to be a good chance to identify base oils according to their viscosity, as well as to detect the presence/absence of water with a fast, simple and non-destructive ATR-FTIR measurements. Also, ANNs have an outstanding

performance for detecting the aging time of the base oil. ANN classify without error samples in periods of one week and show a good trend to group samples in shorter periods of times. Results thrown by LDA were compatible and coherent with those obtained by ANN. In our knowledge, ATR-FTIR aided by ANN has never been used to predict the contamination by water and the aging time of lubricant base oils.

## **Material and methods**

### *Oils samples and instrumentation*

Base oils with different viscosities were kindly provided by REPSOL S.A. A Varian 670-IR spectrometer equipped with a DLaTGS detector and a diamond-based Golden Gate ATR device, with an internal reflection (crystal area 1 x 1 mm), was employed for all the measurements. The spectrometer was completely software-controlled by the Varian Resolutions Pro software provided by Varian Inc. Mathematical data processing and calculations were performed with MatLab©. Linear Discriminant Analysis and statistical analyses were carried out with Origin by Origin Lab Corporation.

### *Oil samples aging*

The study of degradation of lubricant oils is usually carried out on artificially aged oils which are subjected to experimental conditions simulating the working circumstances of the lubricant. These simulations can be performed at small or large scale, and have proven to be a reliable tool for the study of aging according to different parameters like oxidation, viscosity and water content. Good correlations between the values obtained in the simulations and those collected from reference lubricants in use can be obtained [17]. In this work, the accelerated aging of the oils was carried out employing a small-scale simulation of the thermal degradation of three base oils with different viscosity: Base 68, Base 46 and Base 32, the higher the number, the higher the viscosity. Each oil was aged from two different perspectives, a dry aging, where the oil was heated in absence of water, and a wet aging, where the oil was saturated with water previously to the heating.

Dry aging was performed by keeping the oil under constant stirring 30 mL of oil at 100°C in a spherical flask for three weeks; wet aging was carried out by adding 1 mL of ultrapure water to 30 mL base oil and keeping the emulsion under constant stirring at 100°C in a spherical flask for three weeks.

### *Measurement protocol*

Daily, several aliquots of each base oil were collected and their infrared spectra were acquired in triplicate in the ATR-FTIR device, which allowed the oils to be analyzed without any previous treatment of the sample. Each spectrum (as average of 32 scans) was taken from 600 cm<sup>-1</sup> to 4000

$\text{cm}^{-1}$  with a resolution of  $2 \text{ cm}^{-1}$ . A total of 487 spectra were obtained to perform the study, according to the sampling scheme shown in Supplementary Information.

#### *Selection of input variables*

The variables used for the ANN and MA calculations were selected according to the Mid-Infrared (Mid-IR) absorption maxima shown in Figure 1, which are also listed in Table 1.

Figure 1. Average absorption spectra of the three lubricating oil employed in this study (Base 68, 46 y 32)

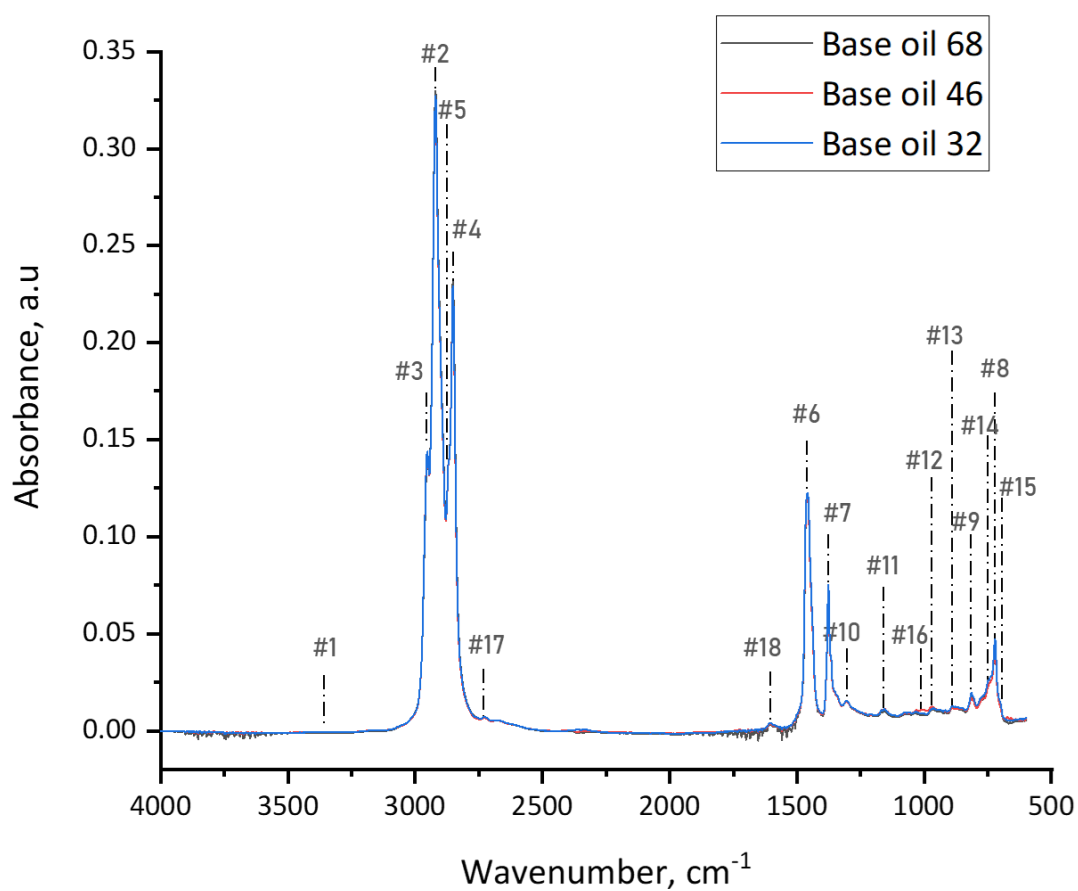


Table 1. Identification of FTIR peaks and their vibrational intensity.

Wave number, cm <sup>-1</sup>	Functional groups	Vibrational intensity
2952	C-H stretching of alkyl chains (CH <sub>2</sub> )	Strong
2852	C-H stretching of alkyl chains (CH <sub>2</sub> )	Strong
1458	bending C-H	Strong
1378	bending C-H	Strong
723	C-H asymmetry bending of alkyl chains (CH <sub>2</sub> and CH <sub>3</sub> )	Medium
3410	O-H stretching of water	Varied
2920	C-H stretching of CH <sub>3</sub>	Strong
2871	C-H stretching of CH <sub>3</sub>	Medium
2729	C-H bending of aldehyde groups	Medium
1610	C=C bending of aromatic groups	Weak/ medium
1305	C=C stretching	Weak/ medium
1157	C=C stretching	Weak/ medium
1032	C=C bending	Weak
967	C=C bending	Weak
889	C-H bending	Medium
812	C-H bending	Medium
752	C-H bending	Medium
701	Out-of-plane aromatic ring C – H flextions	Medium

#### *Artificial Neural Network (ANN) training*

The 85% of the data were used to train the neural network and 15% to test datasets (spectra of every dataset chosen with *dividerand* function of MatLab®). The performance was checked with cross-validation (15%) of training dataset. The selected ANN for this work is a three-layer-feed-forward network with a simple perceptron with sigmoidal activation [18], and the network was trained with a scaled conjugate gradient backpropagation. Several neural networks were prepared for different sample identifications, the main ones were: classification of the base oils by their viscosity grade (vs-ANN) and the identification of the degree of degradation according to the degradation time (number of weeks undergoing heating), (aw-ANN). Each of the networks consisted of an input layer of 18 neurons, a hidden layer with 17 neurons [19] and an output layer with 6 neurons for the degree of viscosity and with 3 neurons for classification per week of degradation. The 18 neurons in the input layer corresponded to the areas of the 18 peaks selected from the FTIR spectra. Samples were not centered, although every peak was normalized with the MatLab *mapminmax* function, so the input data were in the range [-1, 1]. Every spectrum was

taken as the average of 32 scans, providing a good signal-to-noise ratio. Furthermore, the use of peak areas instead of heights contributed to minimize the effect of the noise in the signal.

## Results and discussion

Mid-IR spectra provide information about the functional groups present in the different lubricant oils. The assignation of the most intense bands in the Mid-IR spectra of the base oils to the different functional groups is listed in Table 1. These data demonstrate the aliphatic nature of the hydrocarbon chains in the base oils (bands at 2952, 2920, 2871, 2852, 1458, 1378, 723  $\text{cm}^{-1}$ ), as well as the presence of water (3410  $\text{cm}^{-1}$ ), aldehyde groups (2729  $\text{cm}^{-1}$ ), double bonds C=C (1610, 1305, 1157, 1032, 967  $\text{cm}^{-1}$ ) and aromatic groups (889, 812, 752, 701  $\text{cm}^{-1}$ ).

Since the chemical composition of the tested base oils is very similar, we did not expect many differences in the Mid-IR absorption spectra. In order to check the similarity of the different oils the correlation degree (Equation 1) was used, which corresponds to Pearson's product moment correlation coefficient.

$$\rho_{a,b} = \frac{\text{cov}(A,B)}{\sigma_a \sigma_b} = \frac{E((A - E(A))(B - E(B)))}{\sigma_a \sigma_b}$$

Equation 1

Where  $E(x)$  and  $\text{cov}(x)$  indicate the expected value of the variable and means covariance, respectively. We used the number of wavenumbers scanned (3528) as degrees of freedom. Using Li et al.'s approximation [20], the null hypothesis ( $H_0 =$  'the spectra have no correlation') is discarded whenever p value falls below 0.05. The value of  $\mathbf{p}$  was calculated from Student's  $\mathbf{t}$  related to the correlation coefficient determined (Table 2). The probability associated for the t-values resulted to be lower than 0.001 in all cases and, consequently, the spectra are correlated.

Table 2. Similitude of IR absorbance spectra of the different lubricant oil according to COR and Student's t. BO means Base Oil.

<b>COR t</b>	<b>BO46</b>	<b>BO32</b>	<b>Wet BO68</b>	<b>Wet BO46</b>	<b>Wet BO32</b>
BO68	0.99975 21.56	0.9998 37.66	0.99975 23.35	0.99994 77.64	0.99992 42.96
BO46	---	0.99988 14.72	0.99993 101.62	0.99983 12.36	0.99983 68.73
BO32	---	---	0.99983 75.34	0.9999 3.18	0.9999 105.24
Wet BO68	---	---	---	0.99977 56.28	0.99988 7.00
Wet BO46	---	---	---	---	0.99991 87.39

In order to know whether the Mid-IR spectrum of a given oil was statistically undistinguishable from that of the others, the spectra were compared using a Kruskal-Wallis test, assuming as null hypothesis  $H_0$  'the spectra come from the same distribution'. The comparison was carried out using the mean of all the spectra of the same category (B68, B46, B32, B68w, B46w and B32w) as a representative of the category. Furthermore, spectra were normalized to mean 0 and standard deviation 1 before calculations. When comparing the six different categories (dry and wet oils different physical and chemical properties) the null-hypothesis was rejected with  $p < 10^{-6}$ , what suggested that the spectra were different. The same kind of test was applied independently to the spectra set of wet oils (B68 wet, B46 wet and B32 wet) and to those of the dry oils (B68, B46 and B32), obtaining in both cases  $p < 10^{-3}$ . A less significant dissimilitude was obtained when using Kruskal-Wallis to compare the dry and wet condition of each oil, although in any case  $p$  was lower than  $10^{-2}$  (see Table 3). These results seem to indicate that the spectra were indeed different.

**Table 3.** p-values for the Kruskal-Wallis test performed on different sets of spectra

<b>Spectra sets</b>	<b>p</b>
The six different categories (B68, B46, B32, B68w, B46w and B32w)	$1.02 \cdot 10^{-7}$
Dry base oils (B68, B46 and B32)	$10^{-4}$
Wet base oils (B68w, B46w and B32w)	$1.6 \cdot 10^{-5}$
Base oil 68 wet and dry (B68, B68w)	$1.8 \cdot 10^{-3}$



Base oil 46 wet and dry (B46, B46w)	$8.4 \cdot 10^{-3}$
Base oil 32 wet and dry (B32, B32w)	$2 \cdot 10^{-3}$

As ANN have been successfully exploited in comparable scenarios [21,22], it could be a plausible choice in the present study. However, taking into account that the similarities of the Mid-IR spectra of the oils with different viscosity could somehow mask the capabilities of the ANN to identify the aging process of the oils, we first checked if ANN were able of identifying the type of the base oil and the presence of water using the FTIR absorbance at different wavenumbers as input variables. To do that, the ANN (vs-ANN) was trained and results presented in the confusion matrix (Table 4). The data state the successes (percentage of base oil's classes correctly classified) and reliability (percentage of base oils classified in the right category, or 100 - percentage of false positives) of the vs-ANN, obtaining a percentage of total accuracy of 97.7%. On the other hand, from the data collected in Table 4, vs-ANN was unable to recognize the presence of water in the oil samples with the lowest viscosity, misclassifying the moisture in eight of them. Moreover, there is a certain degree of confusion between the dry base oil 46 and the wet oil 68, as well as between the wet oil 68 and the dry oil 46. Notwithstanding, the total number of erroneously classified samples is not significant.

Table 4. Confusion matrix considering training, validation and test data for vs-ANN

vs-ANN		Target viscosity						Precision
		487 samples	B68	B46	B32	B68 Wet	B46 Wet	
Output viscosity	B68	81	0	0	0	0	0	100%
	B46	0	79	0	1	0	0	98.8%
	B32	0	0	79	0	0	5	94.0%
	B68 Wet	0	2	0	80	0	0	97.6%
	B46 Wet	0	0	0	0	81	0	100%
	B32 Wet	0	0	3	0	0	76	96.2%
	Sensitivity	100%	97.5%	96.3%	98.8%	100%	93.8%	97.7%

Consequently, we decided to generate a model for the determination of the degree of degradation by grouping the samples every four days, creating a neural network for the base oil 68 with 18 inputs, 17 neurons in the hidden layer and 6 output categories (fd<sub>68</sub>-ANN). Although the training of this network showed interesting results, the limited number of data for the validation and the test of the generated model caused that areas under the ROC curve had values equal to zero. In order to have a higher number of spectra which allowed both the validation and the test of the model, a new network (aw-ANN) was trained, classifying the oils according to the aging week (week in which the samples were collected) using all the spectra taken. Table 5 shows the confusion matrix corresponding to aw-ANN displaying a good capacity to classify the data groups according to 3 different weeks with a total accuracy percentage of 97.3%, a percentage similar to that obtained for vs-ANN. Taking into account that the error rate was similar to that in vs-ANN, we considered that the misclassifications appearing in vs-ANN could be affecting the results of aw-ANN. Since the type of the oil (32, 46 or 68) is known in actual situations, we trained six different ANN in order to identify the aging week, knowing the oils viscosity and if they were dry or wet. So we got aw<sub>32</sub>-ANN, aw<sub>46</sub>-ANN, aw<sub>68</sub>-ANN, aw<sub>32w</sub>-ANN, aw<sub>46w</sub>-ANN and aw<sub>68w</sub>-ANN, all them showing a 100% accuracy and 100% reliability. Table 5 records the confusion matrixes for aw-ANN, aw<sub>32</sub>-ANN, aw<sub>46</sub>-ANN and aw<sub>68</sub>-ANN.

Table 5. Confusion matrices considering training, validation and test data for aw-ANN, aw<sub>32</sub>-ANN, aw<sub>46</sub>-ANN and aw<sub>68</sub>-ANN

aw-ANN		Target week			Precision	aw <sub>68</sub> -ANN		Target week			Precision
Output week	487 samples	Week 1	Week 2	Week 3		Output week	81 samples	Week 1	Week 2	Week 3	
	Week 1	194	3	0	98.5%		Week 1	33	0	0	100%
	Week 2	4	119	2	95.2%		Week 2	0	21	0	100%
	Week 3	0	4	161	97.6%		Week 3	0	0	27	100%
Sensitivity		98%	94.4%	98.8%	97.3%	Sensitivity		100%	100%	100%	100%

aw <sub>46</sub> -ANN		Target week			Precision	aw <sub>32</sub> -ANN		Target week			Precision
Output week	81 samples	Week 1	Week 2	Week 3		Output week	82 samples	Week 1	Week 2	Week 3	
	Week 1	33	0	0	100%		Week 1	33	0	0	100%
	Week 2	0	21	0	100%		Week 2	0	21	0	100%
	Week 3	0	0	27	100%		Week 3	0	0	28	100%
Sensitivity		100%	100%	100%	100%	Sensitivity		100%	100%	100%	100%

With these promising results, we tried to combine both ideas in a single network classifying the samples according both to the aging week and viscosity (w1-B68, w2-B68, w3-B68, [...], w2-B32W, w3-B32W), aw&vs-ANN. The percentage of total accuracy of the aw&vs-ANN was 92.4% with values of the area under the ROC curve (AUC) above 0.97 in all cases, indicating a good classification performance. The base oil 32 of the first week was the most affected by false positives, as can be seen in Table 6 with a success rate of only 63.6%. This result is consistent with that found previously in the case of vs-ANN where the identification of wet samples in the case of base oil 32 provided the worst results. The capacity of aw&vs-ANN to classify the samples was great, although the percentage of success and reliability in the classification of some categories were less than 90%.

Table 6. AUC values considering training, validation and test data for aw&vs ANN and their corresponding success and reliability percentages

<b>Dry aged oils</b>			
<b>Aw&amp;vs ANN</b>	<b>AUC</b>	<b>Success</b>	<b>Reliability</b>
w1-B68	0.997 ± 0.006	97.0 %	91.4 %
w2-B68	0.998 ± 0.003	90.5 %	100 %
w3-B68	1 ± 0	100 %	100 %
w1-B46	0.999 ± 0.002	93.9 %	93.9 %
w2-B46	0.998 ± 0.003	90.5 %	95.0 %
w3-B46	0.999 ± 0.002	96.3 %	92.9 %
w1-B32	0.97 ± 0.02	63.6 %	95.5 %
w2-B32	0.995 ± 0.006	90.5 %	79.2 %
w3-B32	0.998 ± 0.005	96.4 %	90.0 %
<b>Water aged oils</b>			
<b>Aw&amp;vs ANN</b>	<b>AUC</b>	<b>Success</b>	<b>Reliability</b>
w1-B68W	0.999 ± 0.002	90.9 %	90.9 %
w2-B68W	0.998 ± 0.002	95.2 %	95.2 %
w3-B68W	1 ± 0	100.0 %	100.0 %
w1-B46W	0.9995 ± 0.0006	97.0 %	94.1 %
w2-B46W	0.999 ± 0.001	95.2 %	100.0 %
w3-B46W	1 ± 0	100.0 %	100.0 %
w1-B32W	0.99 ± 0.01	90.9 %	83.3 %
w2-B32W	0.99 ± 0.01	81.0 %	77.3 %
w3-B32W	0.997 ± 0.009	96.3 %	89.7 %

Based on the results obtained a new approach was carried out to evaluate which ones of the selected peaks of the Mid-IR spectra influenced the most each one of the trained ANN. Several algorithms with this purpose have been described being the algorithm of the weight connection proposed by Olden et al. [23] the most accurate. Table 7 (up) collects the five most influencing variables to identify a category with vs-ANN. In the classification of the base oils with this ANN, the variables with the highest weight were 1157 cm<sup>-1</sup> and 2852 cm<sup>-1</sup> for the identification of dry oils while 967 cm<sup>-1</sup>, 812 cm<sup>-1</sup> and 752 cm<sup>-1</sup> were for the identification of wet oils. Taking the results as a whole, the most weight fell on 1157 cm<sup>-1</sup>, 967 cm<sup>-1</sup>, 2852 cm<sup>-1</sup> and 812 cm<sup>-1</sup> (Table 7, up). The main errors in Table 6 appeared for the misclassification of base oil 32 dry as it were wet and vice-versa, thus suggesting that the confusion arose from the band at 967 cm<sup>-1</sup>, with a strong influence in both categories. It was evident that for dry oils the main variable was 1157 cm<sup>-1</sup>, corresponding to C = C stretching vibrations whereas for wet oils the 967 cm<sup>-1</sup> band was the most important one, assigned to C = C bending vibrations, which have been assigned to out-of-plane bending vibrations of the trans -HC=CH- group present in disubstituted olefins [24,25]. The higher influence of the olefin-related absorptions in the identification of the viscosity is in

agreement with the fact that unsaturated content is related to the viscosity [26]: the higher the olefin content, the more viscous the base oil is [27]. Taking this in consideration, less viscous base oils were expected to have a lower unsaturated content and, therefore, a poorer interaction with water, thus explaining the difficulties of the ANN to identify wet or dry B32 samples. On the other hand, it is remarkable the similitude in oils B46 and B68, which share four variables (Table 7, up) but were not misclassified by vs-ANN, indicating a strong discriminant power in the absorbance at  $701\text{ cm}^{-1}$  and at  $2920\text{ cm}^{-1}$ , ascribed to C-H aromatic and C-H stretching, respectively. The importance of these two variables can also explain why B68w is misclassified with B46 but not with B68: three out of the five most important variables according to Olden's procedure for B68w are related to C=C ( $1032\text{ cm}^{-1}$ ,  $967\text{ cm}^{-1}$  and  $1157\text{ cm}^{-1}$ ). Thus, since the aromatic moieties (which contain C=C bonds) are capital for identifying B46, but not B68, it is more likely that the algorithm gets confused more easily between B46 and B68w rather than between B68 and B68w. The four most important variables for the identification of the week for each one of the  $a_{w\text{oil}}$ -ANN are summarized in Table 7 (down). The two most repeated variables were  $889\text{ cm}^{-1}$  and  $2920\text{ cm}^{-1}$  which, in turn, have the strongest relationship with the aging week.  $2920\text{ cm}^{-1}$  along with  $812\text{ cm}^{-1}$  were the main variables for classifying in the first and the third aging week, while the  $889\text{ cm}^{-1}$  variable was ranked highly for recognizing the second aging week. These vibrations are ascribed to C-H bonds in the alkyl chain (Table 1), which confirmed that oil aging results in chemical modifications of the carbon chains.

Table 7. Variable importance for vs-ANN to evaluate factors related to water presence and absence (up) and evolution of variable importance for  $a_{w\text{oil}}$ -ANN to evaluate factors related to aging week (down)

		Wavenumbers / $\text{cm}^{-1}$				
		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>
vs-ANN	B32	1157	967	812	701	1375
	B46	1157	2852	701	2871	2952
	B68	1157	2852	2952	2920	2871
	B32w	752	889	967	1032	2920
	B46w	2852	1375	812	1305	967
	B68w	812	1032	752	967	1157

		Wavenumbers / cm <sup>-1</sup>				
		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	
ANNs trained for aging week classification	aw <sub>B68</sub> -ANN	1st week	1157	752	812	2920
		2nd week	1305	752	889	1375
		3rd week	1305	1157	1032	2920
	aw <sub>B46</sub> -ANN	1st week	2920	2729	2952	1375
		2nd week	2920	701	723	1610
		3rd week	1305	1157	1032	2920
	aw <sub>B32</sub> -ANN	1st week	1157	889	752	1032
		2nd week	889	1157	1610	1375
		3rd week	701	2729	967	2952
	aw <sub>B68W</sub> -ANN	1st week	701	967	2852	1032
		2nd week	2952	2852	701	889
		3rd week	2729	2852	889	723
	aw <sub>B46W</sub> -ANN	1st week	812	2920	2952	2871
		2nd week	2952	1305	2920	3410
		3rd week	2952	2871	701	2852
	aw <sub>B32W</sub> -ANN	1st week	889	812	723	1610
		2nd week	1157	889	1375	2952
		3rd week	889	1375	723	2920

### Linear Discriminant Analysis (LDA)

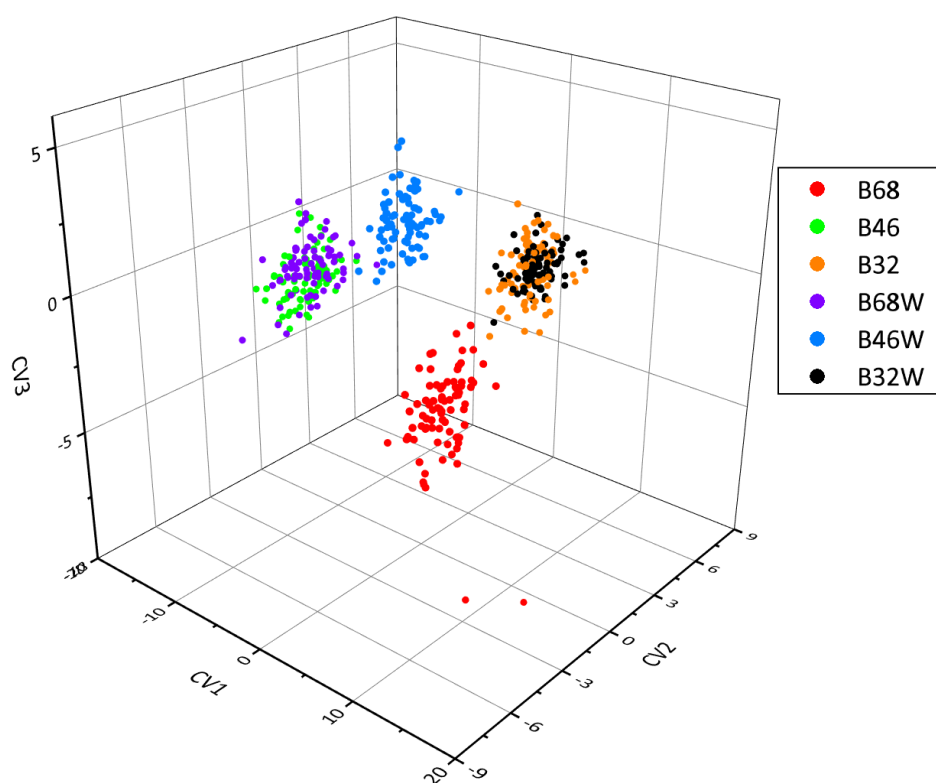
LDA is a widely used tool for data classification and dimensional reduction [28]. This method groups the samples into categories by reducing the variance within groups while maximizing the variance among different groups. The classification of the base oils was performed planning two LDA workouts using the same 18 areas corresponding to the peaks selected in the Mid-IR spectra. The first one, was designed for the classification of the base oils according to their viscosity (dry and wet), vs-LDA, and the second model was devoted to the categorization of the base oils according to their viscosity and degradation week, aw&vs-LDA. In both models, it was possible to calculate the canonical variables that best separated the different categories with a percentage of success lower than that obtained by artificial neural networks.

The three-dimensional representation of the data according to the first three canonical variables showed a true separation in four groups (Figure 2): a) one mainly composed by B68 samples, b) a second one was mainly made up of wet B46, c) B46 and wet B68 samples, which are very close

but slightly separated were part of a third group and d) a fourth group composed of samples dry and wet B32. Projections of the 3D-space into planes CV1-CV2, CV1-CV3 and CV2-CV3 are shown in the Supplementary Information. These results are coherent with those obtained previously using ANN, as the mixed categories in LDA are also the ones showing misclassification in ANN.

For the sake of comparison, a LDA classification including both the viscosity, wetness and aging week was also evaluated. As expected, taking into account that the original data were the same, still four different zones appeared in the canonical variable space with a similar distribution than that previously shown in Figure 2. Misclassification in mixed zones were more frequent in B68 or in wet B46.

Figure 2. Graphical representation of the Canonical Variables for viscosity classification of the base oils B68, B46, B32, B68Wet, B46Wet and B32Wet. A color version of this figure is available online.



## Conclusions

Artificial Neural Networks are a good chance to identify base oils according to their physical and chemical properties, as well as to detect the presence/absence of water. This is an interesting tool to detect the oil contamination by water with a fast, simple and non-destructive FTIR measurements. Also, ANNs have an outstanding performance for detecting the aging time of the oils. ANN classified without error samples in periods of one week and showed a good trend to group samples in shorter periods of times. The analysis of the most influencing variables of ANN showed that the presence of olefins in the samples played a capital role in the misclassification of oils B32 – 32w and 46 – 68w. Results thrown by Linear Discriminant Analysis were compatible and coherent with those obtained by ANN, finding four different zones in the Canonical Variables space. Two of these zones were formed only by a single oil (B68, red color in Figure 2 and B46W, blue color in Figure 2). These categories were also those classified with 100% success and reliability in ANN. On the other hand, LDA mixed B32 and B32W in a single spatial zone (orange and black respectively in Figure 2) and B68W and B46 (purple and green respectively in Figure 2) which is consistent with the confusion matrix of ANN (Table 4).

### **Acknowledgements**

Authors are gratefully acknowledged to Ministerio de Economía y Competitividad, the European Community and the Foundation for the Promotion in Asturias of the Applied Scientific Research and Technology) for the economic resources. We want also to thank the (Photo)electronic and Vibrational Spectroscopy Unit from the Scientific and Technological Resources of the Universidad de Oviedo for the FTIR measurements.



## Funding

This work was supported by the Spanish Ministry for Economy and Competitiveness (Ministerio de Economía y Competitividad) and the European Regional Development Fund [MINECO/FEDER, projects # MAT2015-66747-R and # RTI2018-099756-B-I00]; and the Foundation for the Promotion in Asturias of the Applied Scientific Research and Technology [FICYT project # GRUPIN14-023]

## References

- [1] Holmberg, K.; Erdemir, A., Influence of tribology on global energy consumption, costs and emissions. *Friction* 2017, 5, 263-284 DOI: 10.1007/s40544-017-0183-5
- [2] Lubricant in Merriam-Webster. On-line Access at <https://www.merriam-webster.com/dictionary/lubricant?src=search-dict-box>
- [3] 32United States Patent (19) Hammond et al. 54 SEMI-SYNTHETICLUBRICATING OIL COMPOSITION Kenneth G. Hammond, Poughkeepsie; Joseph C. Sendra, Fishkill, both of N.Y.; Lewis W. Watts, Jr., Austin, Tex.; Edward T. Marquis, Austin, Tex.; John M. Larkin, Austin, Tex.
- [4] Cen, H.; Morina, A.; Neville, A., Effect of lubricant ageing on lubricants' physical and chemical properties and tribological performance; Part I: effect of lubricant chemistry. *Industrial Lubrication and Tribology* 2018, 70 (2), 385-392.
- [5] Erhan, S. Z.; Sharma, B. K.; Perez, J. M., Oxidation and low temperature stability of vegetable oil-based lubricants. *Industrial Crops and Products* 2006, 24 (3), 292-299, DOI: 10.1016/j.indcrop.2006.06.008
- [6] H.P. Bloch, Ed. *Practical Lubrication for Industrial Facilities*, 2nd Ed. Chapter 17, 2009, Taylor&Francis Ltd. Boca Raton. FL. USA
- [7] Singh, A. K.; Mohapatra, S.; Pani, B., Corrosion inhibition effect of Aloe Vera gel: Gravimetric and electrochemical study. *Journal of Industrial and Engineering Chemistry* 2016, 33, 288-297 DOI: 10.1016/j.jiec.2015.10.014
- [8] Viesca, J. L.; Mallada, M. T.; Blanco, D.; Fernández-González, A.; Espina-Casado, J.; González, R.; Hernández Battez, A., Lubrication performance of an ammonium cation-based ionic liquid used as an additive in a polar oil. *Tribology International* 2017, 116, 422-430 DOI: 10.1016/j.triboint.2017.08.004
- [9] Du, Y.; Wu, T.; Gong, R., Properties of water-contaminated lubricating oil: variation with temperature and small water content. *Tribology - Materials, Surfaces & Interfaces* 2017, 11(1), 1-6. DOI: 10.1080/17515831.2017.1279845
- [10] Eachus, A.C., The trouble with water. *Tribology and Lubrication Technology* 2005, 61, 32-38. WOS: 000232615200009
- [11] Nowak, P.; Kucharska, K.; Kamiński, M., Ecological and Health Effects of Lubricant Oils Emitted into the Environment. *International Journal of Environmental Research and Public Health* 2019, 16(16), 3002. DOI: 10.3390/ijerph16163002
- [12] Tung, S.C.; McMillan, M.L., Automotive tribology overview of current advances and challenges for the future. *Tribology International* 2004, 37, 517-536 DOI: 10.1016/j.triboint.2004.01.013
- [13] Souza de Carvalho, M. J.; Rudolf Seidl, P.; Pereira Belchior, C. R.; Ricardo Sodr e, J., Lubricant viscosity and viscosity improver additive effects on diesel fuel economy. *Tribology International* 2010, 43(12), 2298-2302 DOI: 10.1016/j.triboint.2010.07.014
- [14] Gracia, N.; Thomas, S.; Bazin, P.; Duponchel, L.; Thibault-Starzyk, F.; Lerasle, O., Combination of mid-infrared spectroscopy and chemometric factorization tools to study the oxidation of lubricating base oils. *Catalysis Today* 2010, 155(3-4), 255-260. DOI: 10.1016/j.cattod.2009.11.012

- 
- [15] Nguele, R.; Al-Salim, H.; Mohammad, K., Modeling and Forecasting of Depletion of Additives in Car Engine Oils Using Attenuated Total Reflectance Fast Transform Infrared Spectroscopy. *Lubricants* 2014, 2(4), 206–222. DOI: 10.3390/lubricants2040206
- [16] Hossain, M. N.; Uddin, M. N.; Karim, M. M.; Das, A. K.; Rana, A. A.; Jahan, R. A., Prediction of viscosity index of motor oils using FTIR spectral data and chemometrics. *Journal of Scientific and Innovative Research* 2018 7, 43-49.
- [17] Besser, C.; Agocs, A.; Ronai, B.; Ristic, A.; Repka, M.; Jankes, E.; McAleese, C.; Dörr, N. Generation of engine oils with defined degree of degradation by means of a large scale artificial alteration method. *Tribology International* 2019, 132, 39-49 DOI: 10.1016/j.triboint.2018.12.003
- [18] Anderson J.A., *An Introduction to Neural Networks*, The MIT press, 1997
- [19] Tamura S.; Tateishi M., Capabilities of a Four-Layered Feedforward Neural Network: Four Layers Versus Three. *IEEE Transactions on Neural Networks* 1997, 8(2), 251-255 doi: 10.1109/72.557662
- [20] Li, J.; Hibbert, D.B.; Fuller, S.; Vaughn, G. A comparative study of point-to-point algorithms for matching spectra. *Chemometrics and intelligent laboratory systems* 2006, 82 (1-2), 50-58 DOI: 10.1016/j.chemolab.2005.05.015
- [21] Murru, C.; Chimeno-Trinchet, C.; Díaz-García, M.E.; Badía-Laíño, R.; Fernández-González, A. Artificial Neural Network and Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy to identify the chemical variables related to ripeness and variety classification of grapes for Protected Designation of Origin wine production. *Computers and Electronics in Agriculture* 2019, 164, 104922 DOI: 10.1016/j.compag.2019.104922
- [22] Fernández-González, A.; Montejo-Bernardo, J.M.; Rodríguez-Prieto, H.; Castaño-Monllor, C.; Badía-Laíño, R.; Díaz-García, M.E. Easy-to-use analytical approach based on ATR-FTIR and chemometrics to identify apple varieties under Protected Designation of Origin (PDO), *Computers and Electronics in Agriculture* 2014, 108, 166–172 DOI: 10.1016/j.compag.2014.07.009
- [23] Olden, J. D.; Joy, M. K.; Death, R. G. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* 2004, 178, (3-4), 389-397 DOI: 10.1016/j.ecolmodel.2004.03.013
- [24] Poiana M.A.; Alexa E.; Munteanu M.F.; Gligor R.; Moigradean D.; Mateescu C., Use of ATR-FTIR spectroscopy to detect the changes in extra virgin olive oil by adulteration with soybean oil and high temperature heat treatment, *Open Chem.* 2015, 13, 689-698, DOI: 10.1515/chem-2015-0110
- [25] Socrates, G., *Infrared and Raman Characteristic Group Frequencies: Tables and Charts*. 3 ed.; John Wiley & Sons: Chichester, 2001
- [26] George E. Totten, R.J. Shah, *Hydrocarbon base oil chemistry in Fuels and Lubricants Handbook: Technology, properties, performance and testing*, ASTM International, George E. Totten, R.J. Shah, S.R. Westbrook, 2003, ISBN of 978-0-8031-4551-1
- [27] Storch H.H.; Anderson R.B.; Hofer L.J.E.; Hawk C.O.; Anderson H.C.; Golumbic N., *Synthetic liquids fuel from hydrogenation of carbon monoxide*, U.S. Government Printing Office: Washington, 1948.  
<https://books.google.es/books?id=0gXyAAAAMAAJ&lpg=PA64&hl=es&pg=PA64#v=onepage&q&f=false>
- [28] MacFarland J.W., Gains D.J., *Linear Discriminant Analysis and Cluster Significance Analysis*. In Ramsden CA (Editor), *Comprehensive Medicinal Chemistry*, 4, New York, 1990, Pergamon Press.