Macian Martinez, V.; Tormos, B.; García-Barberá, A.; Tsolakis, A. (2021). Applying chemometric procedures for correlation the FTIR spectroscopy with the new thermometric evaluation of Total Acid Number and Total Basic Number in engine oils. Chemometrics and Intelligent Laboratory Systems. 208:1-8. https://doi.org/10.1016/j.chemolab.2020.104215

Applying chemometric procedures for correlation the FTIR spectroscopy with the new thermometric evaluation of Total Acid Number and Total Basic Number in engine oils

Vicente Macián[a], Bernardo Tormos[a], Antonio García-Barberá[a,*], Athanasios Tsolakis[b]

[a] *CMT – Motores Térmicos, Universitat Politècnica de València, Camino de Vera, 46022 Valencia, Spain*
[b] *School of Mechanical Engineering, University of Birmingham, Edgbaston B15 2TT, United Kingdom*

Abstract

Currently, the main trend is to employ spectroscopic techniques for the condition assessment of used engine oils. That is a consequence of the requirement for increasing the Oil Condition Monitoring (OCM) efficiency. Usually, FTIR (Fourier Transform Infra-red) Spectroscopy is employed to achieve the ability of measuring used engine oil and uses the FTIR spectrum to generate a relation between attribute-spectrum. To allow this connection, it is necessary to develop chemometric studies. However, this procedure needs a previous validation to correlate the spectrum with the value of the attribute measured by standardised techniques. One of the potential attribute for study is the acidification of the engine oil by the Total Acid Number (TAN), specially employing the ASTM D8045 standard that measures the TAN level by thermometric titration. The new standard offers a better chemical path to quantify accurately the TAN value of used engine oil. And other basic parameter is the Total Basic Number (TBN) that also is possible to quantify by thermometric determinations. Both of these parameters are correlated, therefore this study is focused on the generation of

*Corresponding author. Tel.: +34 963 877 650,
email: angarba2@mot.upv.es

2020-11-30

a mathematical model that describes the condition of the engine oil employing the shape of its FTIR spectrum. Specifically, in this work has been shown the applicability of chemometric procedures in conjunction with thermometric titration for determining TAN and TBN values of an used engine oil.
Keywords: FTIR, TAN, TBN, Thermometric titration, Chemometric analysis

1. Introduction

Engine oil used during the real operation of an internal combustion engine is continuously changing its properties, as several reactions take place inside the engine. The main sources of changes in the engine oil are due to temperature, pressure, combustion by-products, contamination with external agents and shearing effects. As a result of all these possible degradation processes, the performance of the engine oil changes, differing from the fresh oil [1]. Therefore, it is a key step the determination of the engine oil condition, in order to assure a good engine lubrication.

Traditionally, engine oils (fresh or used) analysis consist of a combination of physical and chemical methods that allows to understand its performance. Fourier Transform Infra-red (FTIR) spectroscopy [2], has been widely studied due to its several advantages including: speed of analysis, lower cost and wide variety of information that it can be obtained. Parameters of the engine oil such has oxidation [3], nitration, sulfation, soot, water, fuel dilution [4], coolant and additives amount (ZDDP, aminic additives, phenolic additives) [5] can be determined using FTIR spectroscopy [6, 7] and as a result this techniques can replace, or rather ratify, the results obtained by traditional analysis procedures.

Total Basic Number (TBN) represents the amount of alkaline concentration that an engine oil has in its formulation. The goal of that alkaline concentration is to neutralize the acidic by-products formed as a consequence of the degradation generated by the usage of the engine oil. Therefore, TBN is a control parameter in order to know if the engine oil is being able to protect the engine against corrosive wear. This characteristic is obtained through the use of alkaline additives that

react with these acid compounds and that additives deplete during the ODI (Oil Drain Interval). In this way, when the engine oil is fresh, it has the maximum value of TBN but when it accumulates mileage that parameter decreases, and no significant filler has been produced within the ODI or additives additions that can affect the result of TBN.

Currently, due to the higher requirements to engine oils in the new engines designs, the study of the evolution of engine oil acidification is of crucial importance. This characteristic is represented by the Total Acid Number (TAN). With mileage, engine oil degrades and its acidic organic constituents (that are present in its own molecular structure) start to increase. TAN is measured by titration with a base, commonly Potassium Hydroxide (KOH), and its value is expressed in mg KOH/g of sample. The main idea behind the TAN quantification is to use the necessary amount of KOH solution (that its concentration is perfectly known) to neutralize the acidic components in the engine oil sample. Consequently, to detect the moment when the added amount of KOH is enough to neutralize the acids (equivalence point), it is necessary to employ some physical-chemical changes or signal [8]. Several degradation processes could lead to an increase of TAN, but the most important ones are: thermal oxidation and contamination with combustion by-products. As a result, if the TAN value is high (close to TBN) the engine oil becomes corrosive and it induces corrosive wear in engine lubricated components.

According to that, it is possible to choose different methodologies and some of them are standardised, for example: potentiometric determination according to ASTM D664 [9], photometric determination according to ASTM D974 [10] and thermometric determination according to ASTM D8045 [11]. The newest standardised determination is the thermometric method and therefore, it is necessary to study more deeply its performance in the tasks of the Oil Condition Monitoring (OCM) as a consequence of its novelty and reduced implementation. Studies of TAN and/or TBN oils that exist are referenced to the other methodologies: potentiometric determination and photometric determination, but not for thermometric determination. As these are different analytical methodologies,

it is not entirely correct to compare studies since the response of each of the methodologies is different, which implies the need to study thermometric determination. Unsurprisingly, in a short period of time, analytical thermometry will be more present in oil analysis laboratories and, therefore, the work derived from it will begin to develop.

To conclude, an alternative procedure to all mentioned determinations methodologies [12, 13] is employing mathematical methodologies [14] to achieve a correlation, that from the mid-IR absorptions [15], provide estimated values of TAN [16, 17, 18] and TBN [19, 20]. So, if the techniques to accurately determine both parameters, as shown by the thermometric titration, are used, it is possible to generate a precise mathematical model that provides reliable results [13, 21].

2. Materials and methods

Samples. In this study, samples from real-service engine oil used in diesel and CNG (Compressed Natural Gas) engines have been analysed. Three different engine oil formulations were tested (see Table 1). This fact adds value to the study, since they are samples used under real conditions, where the engine oil has suffered degradation and contamination processes that occur under real working conditions (outside the simulated conditions that can be performed in a laboratory).

Also, the additive packages of each engine oil formulation are different. As can be seen in the Supplementary Material section 1, lubricant B and lubricant C are more similar than the lubricant A according to the analysis results.

Each additive package is designed for specific working conditions. That is why, if the conditions under which the engine oil is to be exposed are demanding, the additive package will be prepared to ensure the correct lubrication of the engine. And vice versa, if the conditions are kinder, the additive package does not requiere to be so rich. Therefore, with the exception of the base used in the engine oil formulation, according to the additives used, the IR spectrum of the lubricating oils will be different. Regarding TAN/TBN, their performance would

Table 1: Main characteristics of the engine oils tested.

Parameters	**Lub. A**	**Lub. B**	**Lub. C**
SAE grade	5W30	5W30	5W30
API category	CJ-4	CK-4	FA-4
API base oil	G-III+IV	G-III+IV	G-III+IV
KV@40 °C (cSt)	68	68	55
KV@100 °C (cSt)	11.7	12.6	10.5
HTHS@150 °C (cP)	3.58	3.57	3.10
VI (-)	<169	168	165
TBN (mg KOH/g)	10	11	12
SAPS level	Mid	Mid	Mid

be different according to the engine oil composition. However, in all cases the TBN decreases progressively, while the TAN increases slowly. At the time of the ODI, the value of TBN falls above the TAN. The specific formulation ensures a sufficient mileage window to prevent engine oil from generating corrosive attacks on the engine elements it lubricates, especially those constructed by soft metals such as copper and lead.

The protocol followed for sampling is contained in ASTM D8112 [22]. In this way, 125 mL of engine oil could be collected as follows:

- The vehicle is started and idling for a period of 10 minutes to ensure that the engine oil has recirculated throughout the engine and is tempered. This ensures that the oil content in the crankcase is homogeneous.
- The sampling kit consisting of: a vampire pump, a PP tube (one for each sample) and a 125 mL of PP bottle (one for each sample) are then prepared.
- After 10 minutes, the engine shuts down.

- Then, with the sampling kit, the PP tube is inserted through the duct where the oil level rod is accommodated to access the crankcase easily. Because of this, the length of the tube is in line with the length of the rod of each vehicle.

- Once the tube is inserted, the pump is operated so that a vacuum pressure is generated inside the sampling vessel. Thus, this pressure difference favors the suction of the lubricating oil from the crankcase, which is collected in the bottle.

Following this protocol, the different engine oil samples can be extracted: those for the model (Supplementary Material section 2) and those for validation (Supplementary Material section 3).

Thermometric titration instrumentation. Metrohm Ltd. experimental instrument was used and the experimental set-up consist of:

- 859 Titrotherm instrument: Dosino, Thermoprobe and stirrer.

- Metrohm tiamo with Thermometric Endpoint Titration (TET) command.

One of the most critical items of the Thermometric titration instrumentation is the thermistor sensor, that measures temperature changes in titration solutions as low as 10^{-5} K. In addition to this high temperature resolution, the thermometric titration is enhanced employing catalytic reagents in order to reduce the thermal signal noise and locate the endpoint at the right moment. However, the equipment uses a specific software that facilitates the interpretation of the thermal signal and the final treatment for obtaining the TAN and/or TBN value of the sample.

As a result of standardised method for TAN thermometric determination [23], according to the ASTM D8045 standard [11, 24], the software of the titration instrumentation records the data obtained from the chemical determination (see Figure 1).

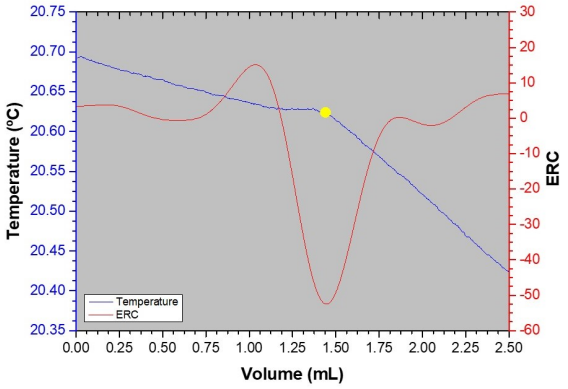

Figure 1: TAN graph, where the temperature and the ERC (Endpoint Recognition Criterion), coloured in yellow, were depicted with the added titrant volume.

$$TAN = \frac{(V_{EP} - Blank) \times c_{KOH} \times f \times M_A}{m_S} \tag{1}$$

In Equation 1 is presented the mathematical expression of TAN (Total Acid Number, in mg KOH/g, of the sample), the V_{EP} is the volume (in mL) of the titrant solution necessary to achieve the EP, whereas $Blank$ is the Blank value; that is the volume of titrant solution consumed for the used quantity of solvent. c_{KOH} is the concentration of the titrant solution expressed in Molar (mol/L). f is a correction factor (titer) and it is dimensionless. The last element in the numerator of the mathematical expression is M_A, molar mass of Potassium Hydroxide (KOH), 56.106 g/mol. To conclude, all the above elements are divided by m_S that is the sample weight in grams [25]. More information about this procedure in Supplementary Material section 4.

To obtain the value of TBN it is necessary to follow an equivalent pathway like TAN but changing the titration conditions. For the TBN titration an acid solution is required, consequently the TBN value is the acid amount expressed in milligrams of Potassium Hydroxide per gram of engine oil sample (mg KOH/g).

The TBN signal is recorded also in order to obtain the EP (Endpoint) of the titration. Figure 2 shows the signal of the temperature (plotted in blue) and the ERC (plotted in red) according the volume of Trifluoromethanesulfonic acid solution.

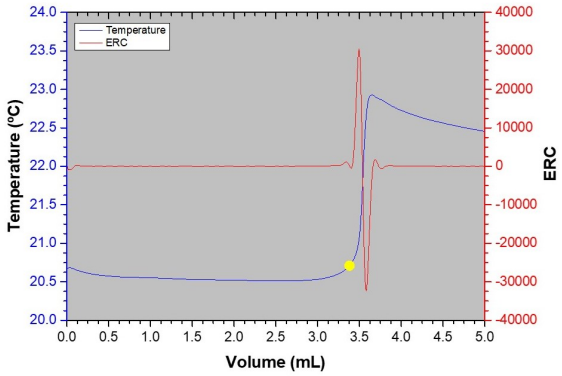

Figure 2: TBN graph, where the temperature and the ERC (Endpoint Recognition Criterion), coloured in yellow, were depicted versus the added titrant volume.

$$TBN = \frac{(V_{EP} - Blank) \times c_{TFMSA} \times f \times M_A}{m_S} \tag{2}$$

In Equation 2, each element of the mathematical expressions of the TBN (Total Base Number of the sample, in mg KOH/g) represents the next subjects: the elements of the parenthesis, V_{EP} and $Blank$ are the consumed volume (in mL) of titrant solution to reach the EP and blank value, respectively. c_{TMFSA} is the concentration of the Trifluoromethanesulfonic acid solution (titrant) in Molar units (mol/L). f is a dimensionless correction factor (titer). M_A represents 56.106 g/mol, the Molar mass of KOH. Finally, m_S is the weight, in grams, of the sample [26]. More information in Supplementary Material section 5.

This TBN thermometric determination is not standardised yet (ASTM standard method is under discussion), however the TBN value can be obtained from other standardised methods such as: potentiometric determination (ISO 3771 [27], ASTM D2896 [28]), photometric determination (ASTM D974 [10]) or conductometric determination (IP 400 [29]).

One of the first steps was to study the accuracy of the thermometric titration. Consequently, 6 different engine oil samples with differente mileage were measured several times in order to obtain the deviation for each value (Supplementary Material section 6). In all cases, the error does not surpass the 5% (in absolute value). However, the maximum difference (between average value and individual measurement) in each parameter is not relevant according to the results obtained: in TAN the maximum difference is 0.04 mg KOH/g while in TBN is only 0.21 mg KOH/g. As a result, the thermometric titration procedure has proven that lead to an accurate determination of TAN and TBN values in engine oils.

FTIR spectrophotometer. FTIR spectrophotometer employed in this study has main characteristics presented on Table 2.

Table 2: Main characteristics of the FTIR spectrophotometer.

Feature (units)	**Value or property**
Spectral range / Resolution (cm^{-1})	4700 – 590 / 4
Background Scans (-)	128
Sample Scans (-)	128
Path length (mm)	0.1
Window material (-)	Zinc Selenide (ZnSe)
Sampling interface (-)	Liquid transmission-TumblIR
Apodization (-)	Triangular

This device allows to measure engine oil samples with pH value between 4-9 unites (optimal pH range to work with ZnSe). Consequently, according to the ASTM D7946 [30], it has been controlled the pH level of the different samples to corroborate that the pH was not below 4 (as a consequence of its usage). In

that way, this corroborates that IR results are reliable. To avoid problems of erroneous FTIR spectra, each sample was analysed 3 times to minimize this potential problem due to external source effects. To conclude, each group of triplicate spectra of each sample was used to generate a global average spectra.

One of the common pre-treatment processes when using the FTIR spectrum is to subtract the spectrum of the fresh engine oil from the spectrum of the sample in order to remove the contribution of the fresh engine oil to the sample. However, the equipment used on this study is not advisable as a consequence of the imprecision when the FTIR spectrophotometer tries to fix the background level. That zero-line (no IR-absorption bands) is not completely stable and possess a fluctuation that in some cases could generate an error if it is worked with spectral subtraction (Supplementary Material section 7 shows this problem). Other possibility is to apply the 1^{st} or 2^{nd} derivatives to subtract the background effect. However, this procedure is also affected by the problem of fixing the background line (Supplemetary Material section 8). The procedure to use the FTIR spectrum is influenced by the capability of the employed FTIR spectrophotometer. It has been decided not to pretreat the signal, due to the operating limitations of the IR spectrophotometer (mainly by sampling interface that it does not control the dispersive phenomena that occurs inside the sample: IR light pathway and its vicinities). So, to avoid any loss or error of information/data when performing any alteration of the spectrum.

Therefore, to ensure the correct development of the study, the different samples were correlated with their respective new engine oil. In this way, a starting signal (reference) is recorded and analyzed as the spectrum varies as the engine oil accumulates use: during the test, all the spectrum of the engine oils a different mileage were collected. In that way, the evolution of the FTIR spectrum due to the usage of the engine oil is plotted in Figure 3.

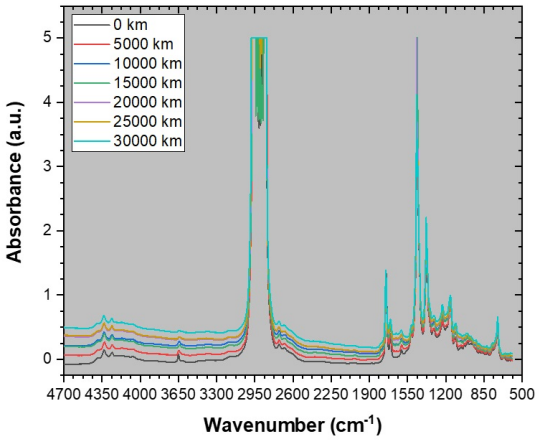

Figure 3: IR spectrum of the same engine oil with different mileages. In that case, they are samples from EEV diesel engine that employs the lubricant C.

Figure 3 is an example of the evolution of the absorption IR bands due to the degradation/contamination of the engine oil as a consequence of the progressive accumulation of mileage. As seen in Figure 3, the fresh engine oil (0 km) spectrum has a slight deviation of its background line. The other samples suffer the same problems due to the proper functioning of the sampling interface.

Nevertheless, employing this FTIR spectrophotometer, it is possible to develop a chemometric model. The procedure followed in this study is therefore, based on observing the evolution of the IR spectrum (direct trending). The FTIR spectrum employed to develop this chemometric study is the result of the average of 3 different FTIR spectrums. For this reason, it is necessary to apply a smoothing process to reduce noise of the average calculation.

According to the experience in the OCM, it is possible to remove some regions of complete IR spectrum that are not relevant for monitoring some parameters correlated with the TAN and TBN such as: oxidation, nitration, additives depletion, etc [3, 5, 16, 31, 32, 33]. Consequently, the problem can be simplified.

The total wavenumbers amount of 2208 (before the smooth treatment) can be reduced to 339 (smooth treatment included). This data reduction process is due to the necessity to balance the amount of data (wavenumbers) with the TAN and TBN measured values. Consequently, the stages of this process are:

- 2208 original wavenumbers obtained from the FTIR spectrophotometer.
- Smooth process over the 2208 wavenumbers reduces the database to 552 wavenumbers. With this treatment, it is possible to remove noise from a data set. However, it may lead to certain data points being ignored. Everytime the smooth process was applied, the curve fitting was checked for it to adjust to the original IR spectrum.
- Finally, select just the IR zones more useful in order to describe the changes in the IR spectrum as a consequence of the degradation processes that affects TAN and TBN values: 4500-4000 cm^{-1}, 3800-3100 cm^{-1} and 1900-590 cm^{-1}. In these three regions appear the different parameters that are controlled by FTIR in OCM following the ASTM E2412 standard [34]: water (3500 to 3150 cm^{-1}), oxidation (1800 to 1670 cm^{-1}), nitration (1650 to 1600 cm^{-1}), antiwear additives (1025 to 960 cm^{-1}), aminic antioxidant additives (1550 to 1490 cm^{-1}), diesel (815 to 805 cm^{-1}), sulfate by-products (1180 to 1120 cm^{-1}) and ETG coolant (1100 to 1030 cm^{-1}).

Process described before can be observed in the Supplementary Material section 9. Following these steps, the final database is composed by 339 different wavenumbers. Employing just that reduced dataset it will be possible to generate an accurate mathematical model that allows monitoring the evolution of TAN and TBN value.

Chemometric study and treatment. TAN [16, 17, 35] and TBN are complex parameters to be calculated by the IR spectrum. In that case it is necessary to employ chemometrics models [36] that uses a database of IR spectrum and the value of both parameters determined by direct analysis (in this study by

thermometric titration). According to that, it is generated a database of IR spectrums and TAN/TBN values: 182 samples (each sample was analysed 3 times). In order to generate a robust model, firstly is necessary to calibrate the proposed model and then proceed with the model validation using engine oil samples measured by direct titration.

The amount of data could further be reduced by employing some statistical procedures, in this study PCA (Principal Component Analysis) has been chosen [37]. The main reason of this selection is due to the ability of PCA, using an orthogonal transformation, to reduce all data (variables) in a simple group of a set of components called Principal Components (PC) that are linearity uncorrelated. Next, employing the reduced amount of data it is possible to obtain a correlation between the PCs and the TAN/TBN values by different regression models: Partial Least Squares (PLS) regression [17, 33, 38, 39], SVM (Supporting Vector Machines) regression models and GPR (Gaussian Process Regression) models. Finally, the regression obtained was checked using 20 different samples which, in the same way as the previous group, were analysed 3 times each.

(1) **The calibration of the model** requires a selection of samples, in this study 546 different data of TAN and TBN values are used that were determined by thermometric titration.

    The average value of the 3 set measurements of each engine oil sample is plotted in Figure 4. TBN values are progressively decreasing along the ODI while the performance of the TAN is increasing as expected.

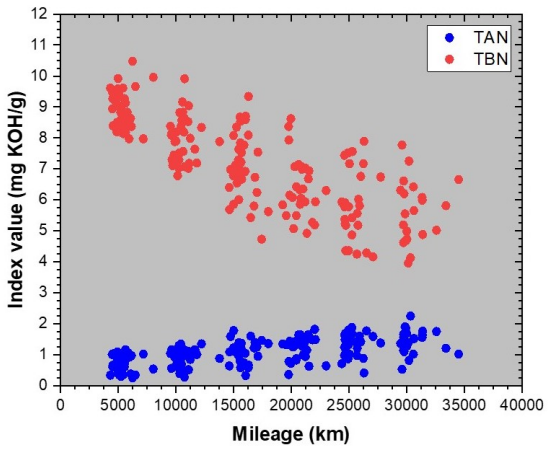

Figure 4: TAN and TBN average values of 182 different engine oil samples.

(2) **Mathematical model validation** 20 engine oil samples were used to validate the model. These 20 samples were distributed along the ODI: from the beginning (5,000 km) to the end at 30,000 km (in stages of 5,000 km). With this set of samples, it was tried to obtain a homogeneous distribution along all of these different mileages in order to obtain the maximum variety of mileage (for more information, see the Supplementary Material section 3). The following graph, Figure 5, represents that selected group of values.

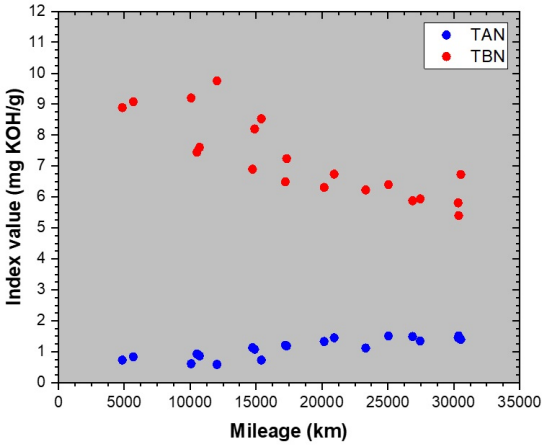

Figure 5: Engine oil samples employed to validate the model.

With these values, extended along the ODI, it is possible to corroborate the accuracy of the response of the proposed model.

The different sets of samples, calibration and prediction, were selected trying to maintain the ratios of mileage-formulations-vehicles as much as possible (for more information, Supplementary Material section 7).

3. Results and discussion

In the calibration step, it was necessary that the parameters (TAN and TBN) of interest had been determined accurately. Thermometric titration is a properly direct measurement method to determine that parameters. Next, to generate the chemometric model, a value of TAN and TBN was assigned to each IR spectrum. In order to reduce the number of variables, multivariate methods: PCA and PLS were applied in this study. Thus, it was possible to obtain a simple method that allow to generate correlations between the IR spectrum and the TAN and TBN.

First of all, employing PCA it is possible to obtain a new set of variables, called Principal Components (PC), which describes all the set of IR data. Next, according to the relevance of each PC, they could be selected or not, to be used in the chemometric model. The purpose of the PCA analysis is to obtain a small number of linear combinations of the 339 variables (wavenumbers) that explain the greatest variability in the data. For this reason, according to Figure 6, if 10 components are selected they will explain more than 99.5% of the variability in the original data (see Table 3).

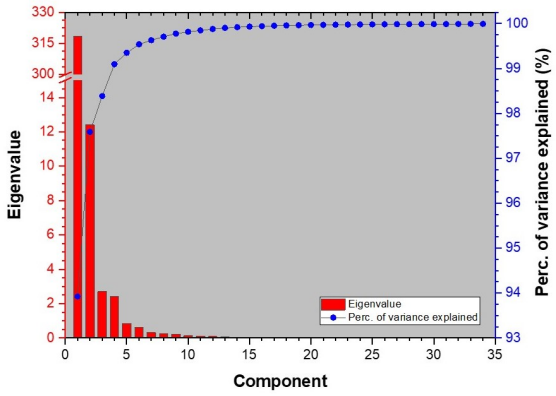

Figure 6: Percentage of variance explained per each PC.

Table 3: PCA analysis and the main characteristics of the four selected PCs.

PCi	**Eigenvalue**	**Perc. of variance (%)**	**Perc. vari. accumulated (%)**
1	318.406	93.925	93.925
2	12.414	3.662	97.587
3	2.707	0.799	98.385
4	2.406	0.710	99.095
5	0.860	0.254	99.349
6	0.635	0.187	99.536
7	0.314	0.093	99.629
8	0.260	0.077	99.706
9	0.237	0.070	99.776
10	0.135	0.040	99.815

According of the eigenvalue (value that determines the magnitude of the PC) of each PC, it is only interesting the PCs whose eigenvalue is more than 1: PC1, PC2, PC3 and PC4. These results can be considered a checkpoint that the correct number of PCs have been selected. Besides, the first two components (PC1 and PC2) gather most of the entire variance in the data. In this way, using the score plot it could be detected some group or families of data (called clusters) that follow some trends, and identify the presence of outliers. Next plot, Figure 7, presents the scores of the principal component two versus the scores of the principal component one (which they gather almost the 98% of the variance), allows to distinguish between the 3 different engine oil formulations (shown in a different color) employed and detect data outliers.

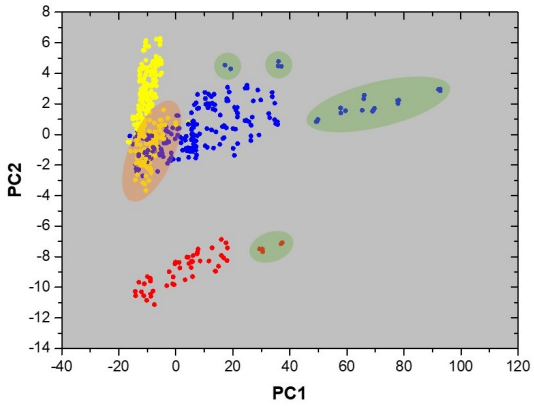

Figure 7: PC1 vs PC2 score plot where it is coloured each different engine oil formulation employed in that study.

In concordance to Figure 7, PC1 and PC2 can explain the differences between the engine oil formulations: the red dots are the formulation of lubricant A while the others are lubricant B (blue dots) and lubricant C (yellow dots).

In this study, one engine oil considered is completely different to the others; this engine oil is coloured by red and consequently it appears quite separated from the other two. However, the other two formulations are similar, only the balance of the additives package changes. Consequently, both engine formulations share some IR bands and has particular values, resulting in both families being closer (see the orange highlighted zone). Lubricant A has in its formulations a higher amount of calcium while lubricant B and C have a significant variances in their boron, calcium and molybdenum content (see Supplementary Material section 1). This graph also allows to detect outlier data, that does not follow the distribution (see green highlighted zones).

In the following Figure 8 are plotted the scores of each PC according to the wavenumbers. This could be usefully to get an idea of these regions of the IR spectrum were minimum variations could generate high deviations.

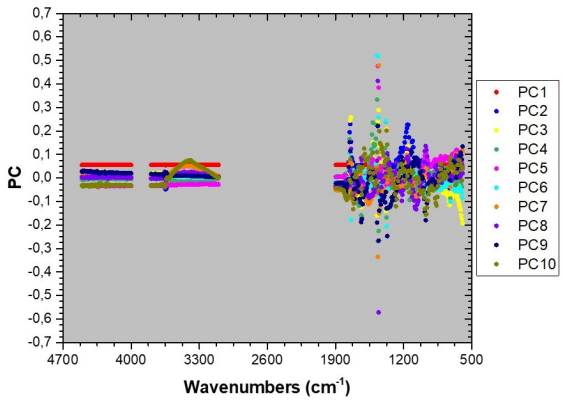

Figure 8: PC scores according to IR wavenumber.

The sensitive regions are located in the right zone of the spectrum, around 2000-500 cm^{-1}. This fact is quite logical, because in this spectral range is where the additives and degradation bands appear more clearly. The other regions, selected in the study, have significance as indicated by the responses of the different PCs versus the wavenumber.

Once the PCA is completed, different analysis can be performed: PLS (Partial Least Squares) [39], Linear Regressions (LR), Support Vector Machine (SVM) and Gaussian Process Regression (GPR). In Table 4 are written the results of each regression focusign the interest in the Corrletaion Coefficnet (R^2) and the Root Mean Square Error (RMSE).

Table 4: Regression models studied.

Regresion model	**Parameter**	**R^2 cal.**	**RMSE cal.**	**R^2 val.**	**RMSE val.**
PLS	TAN	0.57595	0.21596	0.75452	0.16225
PLS	TBN	0.75680	0.68391	0.77265	0.71879
Linear	TAN	0.96264	0.21889	0.97884	0.15878
Linear	TBN	0.99257	0.63210	0.99074	0.70632
Linear SVM	TAN	0.96177	0.22178	0.97770	0.16206
Linear SVM	TBN	0.99195	0.65669	0.99376	0.58421
Quadratic SVM	TAN	0.90772	0.34463	0.79540	0.52046
Quadratic SVM	TBN	0.99166	0.66978	0.99203	0.67552
Cubic SVM	TAN	0.88141	0.41101	0.85284	0.40351
Cubic SVM	TBN	0.83586	3.30178	0.30041	14.04915
Fine Gaussian SVM	TAN	0.97241	0.18657	0.95277	0.22883
Fine Gaussian SVM	TBN	0.99011	0.72714	0.98564	0.87735
Medium Gaussian SVM	TAN	0.96764	0.20083	0.96592	0.20055
Medium Gaussian SVM	TBN	0.99386	0.57197	0.99278	0.62953
Coarse Gaussian SVM	TAN	0.94579	0.25599	0.96394	0.20694
Coarse Gaussian SVM	TBN	0.98846	0.78215	0.98882	0.77609
Squared Exponential GPR	TAN	0.97962	0.16271	0.94731	0.24517
Squared Exponential GPR	TBN	0.99572	0.47982	0.99518	0.51920
Matern 5/2 GPR	TAN	0.98072	0.15834	0.94926	0.24040
Matern 5/2 GPR	TBN	0.99603	0.46161	0.99525	0.51529
Exponential GPR	TAN	0.98311	0.14851	0.96347	0.20801
Exponential GPR	TBN	0.99671	0.42011	0.99346	0.59821
Rotational Quadratic GPR	TAN	0.98308	0.14875	0.98308	0.14875
Rotational Quadratic GPR	TBN	0.99607	0.46005	0.99430	0.56306

Next, the detection and elimination of outliers is necessary because these can influence the accuracy of the model. Consequently, in each regression model, the response is increased (R^2 closer to 1 and RMSE lower than 1) according the results of the Table 5.

Table 5: Regression models without the outliers.

Regresion model	**Parameter**	**R^2 cal.**	**RMSE cal.**	**R^2 val.**	**RMSE val.**
PLS	TAN	0.88169	0.10629	0.73764	0.16235
PLS	TBN	0.84601	0.52074	0.83215	0.61004
Linear	TAN	0.99143	0.11132	0.97898	0.16326
Linear	TBN	0.99515	0.51716	0.99357	0.59544
Linear SVM	TAN	0.99109	0.11233	0.97708	0.16707
Linear SVM	TBN	0.99501	0.52101	0.99329	0.60964
Quadratic SVM	TAN	0.98879	0.12494	0.39800	1.19896
Quadratic SVM	TBN	0.99659	0.42904	0.87487	2.95411
Cubic SVM	TAN	0.96691	0.21847	0.5633	0.84993
Cubic SVM	TBN	0.99432	0.56046	0.55425	7.71572
Fine Gaussian SVM	TAN	0.98842	0.12678	0.94267	0.26041
Fine Gaussian SVM	TBN	0.99286	0.62583	0.98079	1.03102
Medium Gaussian SVM	TAN	0.98968	0.12038	0.96444	0.20753
Medium Gaussian SVM	TBN	0.99492	0.52297	0.99269	0.63405
Coarse Gaussian SVM	TAN	0.97821	0.16897	0.93733	0.28245
Coarse Gaussian SVM	TBN	0.98974	0.75465	0.98699	0.84822
Squared Exponential GPR	TAN	0.99129	0.11295	0.94956	0.24803
Squared Exponential GPR	TBN	0.99620	0.45272	0.99568	0.49450
Matern 5/2 GPR	TAN	0.99054	0.11554	0.94801	0.25181
Matern 5/2 GPR	TBN	0.99656	0.43094	0.99564	0.49689
Exponential GPR	TAN	0.99046	0.11779	0.96243	0.21658
Exponential GPR	TBN	0.99718	0.38956	0.99422	0.56370
Rotational Quadratic GPR	TAN	0.99237	0.10498	0.95829	0.22791
Rotational Quadratic GPR	TBN	0.99637	0.44259	0.99499	0.52925

Finally, the best option to generate a model to obtain TAN or TBN values from the IR response of the engine oil is:

- Linear Regression has the most accurate response to calculate TAN value. In Figure 9 is plotted the fitting results of the samples employed to generate the model (A) and the validation samples (B).

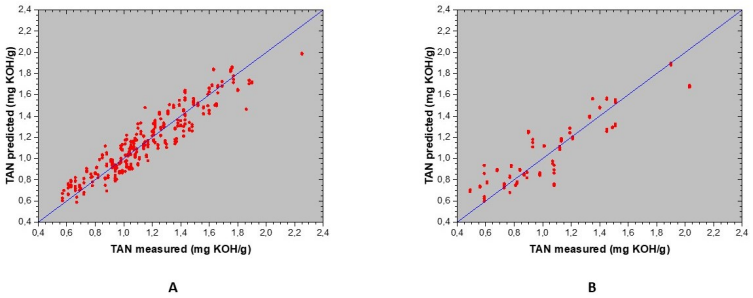

Figure 9: TAN model fitting is represetned by the plot A while the selected samples for validating the model is plotted in the plot B.

- While for TBN, the Squared Exponential GPR describes better this parameter. Figure 10 is the fitting results of the calibration (A) and validation (B) of the model to obtain the TBN value from the IR.

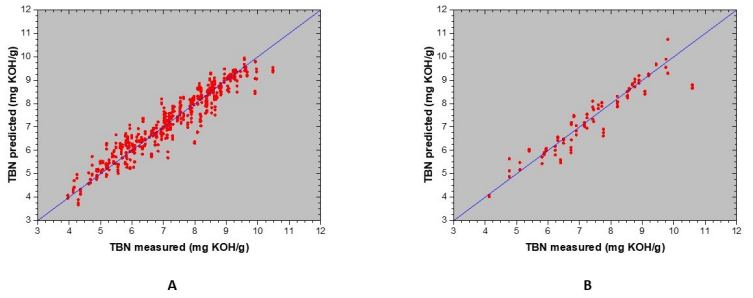

Figure 10: TBN model generates a good fitting, plot A, as same as the validation samples, plot B.

4. Conclusion

Thermometric titration is a good option to determine the TAN and TBN values. However, that technique requires reagents and approximately 3-5 minutes of sample preparation time to be able to obtain a result. The FTIR spectroscopy, on the other hand, has some advantages: no reactive agents are required and its measuring time is faster (∼2-3 minutes per sample). The report obtained from

the FTIR can also offer much more information to the user. Consequently, the use of FTIR spectroscopy, combined with chemometric techniques, allows to know the TAN/TBN level of the engine oil, both fresh and used. The chemometric model shows that it could be a powerful tool to detect some anomaly trend in the performance of the engine oil along its ODI. The correlations obtained in the proposed method for both parameters, TAN and TBN, are good enough for the model to be implemented in a routine oil monitoring program for engine oils.

Acknowledgments

Antonio García-Barberá is supported through the Programa Nacional de Formación de Recursos Humanos de Investigación of Spanish Ministerio de Ciencia e Innovación [grant number BES-2016-078073].

References

[1] A. Kupareva, P. Mäki-Arvela, H. Grénman, K. Eränen, R. Sjöholm, M. Reunanen, D. Y. Murzin, Chemical characterization of lube oils, Energy & fuels 27 (1) (2013) 27–34.

[2] F. R. V. D. Voort, J. Sedman, R. A. Cocciardi, D. Pinchuk, FTIR Condition Monitoring of In–Service Lubricants: Ongoing developments and future perspectives, Tribology Transactions 49 (3) (2006) 410–418.

[3] N. Gracia, S. Thomas, P. Bazin, L. Duponchel, F. Thibault-Starzyk, O. Lerasle, Combination of mid-infrared spectroscopy and chemometric factorization tools to study the oxidation of lubricating base oils, Catalysis Today 155 (3-4) (2010) 255–260.

[4] M. Sejkorová, Application of FTIR spectrometry using multivariate analysis for prediction fuel in engine oil, 2017.

[5] S. Zzeyani, M. Mikou, J. Naja, A. Elachhab, Spectroscopic analysis of synthetic lubricating oil, Tribology International 114 (2017) 27–32.

[6] A. Borin, R. J. Poppi, Application of mid infrared spectroscopy and iPLS for the quantification of contaminants in lubricating oil, Vibrational Spectroscopy 37 (1) (2005) 27–32.

[7] J. Stodola, A. Breznicka, A. Chovanec, D. Rakúsová, P. Lipták, P. Stodola, Diagnosis engine oils utilizing infrared spectroscopy, in: 2017 International Conference on Military Technologies (ICMT), IEEE, 2017, pp. 93–96.

[8] A. Wolak, Changes in lubricant properties of used synthetic oils based on the total acid number, Measurement and Control 51 (3-4) (2018) 65–72.

[9] ASTM D664-18e2 Standard test method for acid number of petroleum products by potentiometric titration, Tech. rep., ASTM International (2018). URL `https://doi.org/10.1520/D0664-18E02`

[10] ASTM D974-14e2 Standard test method for acid and base number by color-indicator titration, Tech. rep., ASTM International (2014).
URL https://doi.org/10.1520/D0974-14E02

[11] ASTM D8045-17e1 Standard test method for acid number of crude oils and petroleum products by catalytic thermometric titration, Tech. rep., ASTM International (2017).
URL https://doi.org/10.1520/D8045-17E01

[12] Spectro Scientific, Guide to measuring TAN and TBN in oil, Tech. rep. (2017).

[13] D. P. F. R. Voort, J. Pinchuk, FTIR Acid and Base Number analyses: Their potential to replace ASTM methods, 2009.

[14] A. Hirri, M. Bassbasi, A. Oussama, Classification and quality control of lubricating oils by Infrared Spectroscopy and Chemometric, 2013.

[15] S. Marinović, A. Jukić, D. Doležal, B. Špehar, M. Krištović, Prediction of used lubricating oils properties by infrared spectroscopy using multivariate analysis, Goriva i maziva 51 (3) (2012) 205–215.

[16] B. L. De Rivas, J.-L. Vivancos, J. Ordieres-Meré, S. F. Capuz-Rizo, Determination of the total acid number (TAN) of used mineral oils in aviation engines by FTIR using regression models, Chemometrics and Intelligent Laboratory Systems 160 (2017) 32–39.

[17] Y. Felkel, N. Dörr, F. Glatz, K. Varmuza, Determination of the total acid number (TAN) of used gas engine oils by IR and chemometrics applying a combined strategy for variable selection, Chemometrics and Intelligent Laboratory Systems 101 (1) (2010) 14–22.

[18] D. Li, J. Sedman, D. García-González, F. van De Voort, Automated acid content determination in lubricants by FTIR spectroscopy as an alternative to acid number determination, Journal of ASTM International 6 (6) (2009) 1–12.

[19] C. L. DiCologero, T. G. Barraclough, P. F. Henning, Spectro, Using Infrared Spectroscopy for the determination of TAN and TBN in machinery, 2011.

[20] J. Dong, F. R. van de Voort, V. Yaylayan, A. A. Ismail, D. Pinchuk, A. Taghizadeh, Determination of total base number (TBN) in lubricating oils by mid–FTIR spectroscopy, Lubrication engineering 57 (11).

[21] C. Winterfield, F. Van de Voort, Automated acid and base number determination of mineral–based lubricants by Fourier transform infrared spectroscopy: commercial laboratory evaluation, Journal of laboratory automation 19 (6) (2014) 577–586.

[22] ASTM D8112-17 Standard guide for obtaining in-service samples of turbine operation related lubricating fluid, Tech. rep., ASTM International (2017). URL https://doi.org/10.1520/D8112-17

[23] D. Muray, TAN thermometric method evaluation, Canadian Crude Quality Technical Association Technical Report.

[24] Competence Center Titration Metrohm International Headquarters, Acidity in crude oils and petroleum products by thermometric titration according to ASTM D8045, Metrohm Application Bulletin 427/2 e Edition.

[25] Competence Center Titration Metrohm International Headquarters, Determination of total acid number in petroleum products, Metrohm Application Bulletin AB–404/2 e Edition.

[26] Competence Center Titration Metrohm International Headquarters, Determination of the total base number in petroleum products, Metrohm Application Bulletin 405/2 e Edition.

[27] D. ISO, 3771: Petroleum products-determination of base number-perchloric acid potentiometric titration method, International Organization for Standardization, Geneva.

[28] ASTM D2896-15 Standard test method for base number of petroleum products by potentiometric perchloric acid titration, Tech. rep., ASTM International (2015).
URL https://doi.org/10.1520/D2896-15

[29] IP 400: Determination of base number of petroleum products – conductimetric titration method, Tech. rep., Energy Institute (2001).
URL https://publishing.energyinst.org/ip-test-methods/full-list-of-ip-test-methods-publications/ip-400-determination-of-base-number-of-petroleum-products-conductimetric-titration-meth

[30] ASTM D7946-19 Standard test method for initial ph (i-ph)-value of petroleum products, Tech. rep., ASTM International (2019).
URL https://doi.org/10.1520/D7946-19

[31] A. M. Abdul-Munaim, T. Holland, P. Sivakumar, D. G. Watson, Absorption wavebands for discriminating oxidation time of engine oil as detected by FT–IR spectroscopy, Lubricants 7 (3) (2019) 24.

[32] M. S. Rauscher, A. J. Tremmel, M. Schardt, A. W. Koch, Non–dispersive infrared sensor for online condition monitoring of gearbox oil, Sensors 17 (2) (2017) 399.

[33] D. Rivera-Barrera, H. Rueda-Chacón, D. Molina, Prediction of the total acid number (TAN) of colombian crude oils via ATR–FTIR spectroscopy and chemometric methods, Talanta 206 (2020) 120186.

[34] ASTM E2412-10 Standard practice for condition monitoring of in-service lubricants by trend analysis using fourier transform infrared (FT-IR) spectrometry, Tech. rep., ASTM International (2018).
URL https://doi.org/10.1520/E2412-10R18

[35] M. J. Adams, M. J. Romeo, P. Rawson, FTIR analysis and monitoring of synthetic aviation engine oils, Talanta 73 (4) (2007) 629–634.

[36] M. Khanmohammadi, A. B. Garmarudi, M. de la Guardia, Characterization of petroleum–based products by infrared spectroscopy and chemometrics, TrAC Trends in Analytical Chemistry 35 (2012) 135–149.

[37] M. Bassbasi, A. Hafid, S. Platikanov, R. Tauler, A. Oussama, Study of motor oil adulteration by infrared spectroscopy and chemometrics methods, Fuel 104 (2013) 798–804.

[38] G. Parisotto, M. F. Ferrao, A. L. Muller, E. I. Müller, M. F. Santos, R. C. Guimaraes, J. C. Dias, É. M. Flores, Total acid number determination in residues of crude oil distillation using ATR–FTIR and variable selection by chemometric methods, Energy & Fuels 24 (10) (2010) 5474–5478.

[39] Ö. Yeniay, A. Göktaş, A comparison of partial least squares regression with other prediction methods, Hacettepe Journal of Mathematics and Statistics 31 (2002) 99–111.