

STATISTICAL ANALYSIS AND CHEMOMETRIC METHODS

Discrimination and Quantification of Moroccan Gasoline Adulteration with Diesel Using Fourier Transform Infrared Spectroscopy and Chemometric Tools

ISSAM BARRA, MOHAMMED ALAOUI MANSOURI, MOHAMMED BOUSRABAT, YAHIA CHERRAH,
and ABDELAZIZ BOUKLOUZE

Mohammed V University, Faculty of Medicine and Pharmacy, Laboratory of Pharmacology and Toxicology, Pharmaceutical and Toxicological Analysis Research Team, Rabat, Morocco

MOURAD KHARBACH

Vrije Universiteit Brussel (VUB), Center for Pharmaceutical Research, Department of Analytical Chemistry and Pharmaceutical Technology, Laarbeeklaan 103, B-1090 Brussels, Belgium

In this work, transform-infrared spectroscopy (FTIR) was associated with chemometric tools, especially principal component analysis (PCA) and partial least squares regression (PLSR), to discriminate and quantify gasoline adulteration with diesel. The method is composed of a total of 100 mixtures were prepared, and then FTIR fingerprints were recorded for all samples. PCA was used to verify that mixtures can be distinguished from pure products and to check that there are no outliers. As a result of using just PC1 and PC2, more than 98% of the general variability was explained. The PLSR model based on infrared spectra has shown its capabilities to be suitable for predicting gasoline adulteration in the concentration range of 0 to 98% (w/w), with a high significant coefficient of determination ($R^2 = 99.25\%$) and an acceptable calibration and prediction errors (root mean squared error of calibration = 0.63 and root mean square of external validation and/or prediction = 0.69).

Gasoline is composed of hundreds of different natural components, such as benzene and toluene, in addition to other additives such as methyl *tert*-butyl ether (MTBE). Components of gasoline could be classified according to the following four classes: paraffins (normal and branched), cycloparaffins, olefins, and aromatics. This composition is different according to several factors, including crude oil origin, the process of obtainment (distillation, alkylation, hydrocracking, catalytic cracking, etc.), the purpose for which it is produced (automotive competitions, engine performance tests), and international legislation (contents of benzene, aromatic, sulphur, lead, etc.; 1, 2).

The adulteration is defined as blending a higher grade gasoline with other cheaper products; it may be as simple as

adding diesel to the super-grade gasoline storage tank, which leads to reduced performances of the super.

Each engine is designed and manufactured to work on specified fuel. When the fuel characteristics are changed, the emission of more pollutants can be increased. The use of adulterated fuels usually leads to engine life and performance reduction and increases the emission of harmful pollutants (3, 4).

Two main types of petroleum products are marketed in different Moroccan service stations: unleaded gasoline and diesel. The former is the most expensive one, while the latter has a cheaper price. The reason fuel pump merchants attempt to adulterate the authentic product is to gain a small profit margin per liter.

Gasoline adulteration may be highlighted by the use of gas or liquid chromatography techniques generally known by their expensiveness and complexity in use. Fourier-transform infrared spectroscopy (FTIR) has always had an important place in fuel analysis and quality control. This technique has been shown, when it is coupled with multivariate analysis, to be a powerful tool to develop quantitative models of fuel quantification because of its fastness, robustness, and analysis cheapness, and it doesn't require any sample preparation (5, 6).

Currently, the application of multivariate calibration is one of the most relevant tools that can be applied to develop a mathematical relationship between chemical variables and measured instrumental signals. They can be classified into the following three categories: data mining tools, e.g., partial least squares regression (PLS); classification tools, e.g., partial least squares (PLS) discriminant analysis; and prediction tools, e.g., partial least squares regression (PLSR; 7).

The aim of this work was to develop a chemometric model by coupling FTIR spectra to PCA and PLSR calibration for the detection and quantification of unleaded gasoline adulteration with diesel.

Materials and Methods

An FTIR spectrometer (Bruker Optics, Ettlingen, Germany) connected to an attenuated total reflectance accessory was employed to record the FTIR spectra in the range of 4000 to 400 cm; final spectra were the mean of 24 scans with a resolution of 4 cm at room temperature to improve the signal-to-noise ratio.

Received June 8, 2018. Accepted by GL September 25, 2018.
Color images are available online at <http://aoac.publisher.ingentaconnect.com/content/aoac/jaoac>
Corresponding author's e-mail: issam.barra@gmail.com
DOI: <https://doi.org/10.5740/jaoacint.18-0179>

All the spectra of fuel samples were exported to PLS-Toolbox software (version 8.1.1, Eigenvector Research) for multivariate analysis calculations.

Samples

Gasoline and diesel samples were purchased from Total Gas Company in the city of Rabat. A group of 100 mixtures (1 mL; from 0 to 98%) were prepared by adding diesel to gasoline in small vials. The experimental design is shown in Table 1. The adulterated gasoline sample set was split into 80 samples for the training set and 20 for validation to construct a multivariate model for the quantification of adulterant.

Chemometrics

(a) *PCA*.—PCA is almost the first examination carried out on a multivariate data. It can be used for data exploration, outlier detection, and graphical clustering and reduce multidimensional data down to a few dimensions. An orthogonal transformation is applied to transform correlated variables to new perpendicular (uncorrelated) variables called principal components (PCs). As a result, PCs are plotted orthogonally to each other to obtain the maximum explained variance (8, 9).

(b) *PLSR*.—PLSR is focused on a bilinear model and usually gives good solutions, where a linear relationship exists between the experimental measured data matrix X ($n \times m$; matrix of independent variables) and the set of output variables Y ($n \times p$;

Table 1. Sample composition prepared by mixing gasoline and adulterant (diesel)

| Sample No. | Gasoline volume, μL | Diesel volume, μL | Adulterant, % | Sample No. | Gasoline volume, μL | Diesel volume, μL | Adulterant, % | Sample No. | Gasoline volume, μL | Diesel volume, μL | Adulterant, % |
|------------|--------------------------------|------------------------------|---------------|------------|--------------------------------|------------------------------|---------------|------------|--------------------------------|------------------------------|---------------|
| 1 | 1000 | 0 | 0 | 35 | 932 | 68 | 6.8 | 69 | 640 | 360 | 36 |
| 2 | 998 | 2 | 0.2 | 36 | 930 | 70 | 7 | 70 | 620 | 380 | 38 |
| 3 | 996 | 4 | 0.4 | 37 | 928 | 72 | 7.2 | 71 | 600 | 400 | 40 |
| 4 | 994 | 6 | 0.6 | 38 | 926 | 74 | 7.4 | 72 | 580 | 420 | 42 |
| 5 | 992 | 8 | 0.8 | 39 | 924 | 76 | 7.6 | 73 | 560 | 440 | 44 |
| 6 | 990 | 10 | 1 | 40 | 922 | 78 | 7.8 | 74 | 540 | 460 | 46 |
| 7 | 988 | 12 | 1.2 | 41 | 920 | 80 | 8 | 75 | 520 | 480 | 48 |
| 8 | 986 | 14 | 1.4 | 42 | 918 | 82 | 8.2 | 76 | 500 | 500 | 50 |
| 9 | 984 | 16 | 1.6 | 43 | 916 | 84 | 8.4 | 77 | 480 | 520 | 52 |
| 10 | 982 | 18 | 1.8 | 44 | 914 | 86 | 8.6 | 78 | 460 | 540 | 54 |
| 11 | 980 | 20 | 2 | 45 | 912 | 88 | 8.8 | 79 | 440 | 560 | 56 |
| 12 | 978 | 22 | 2.2 | 46 | 910 | 90 | 9 | 80 | 420 | 580 | 58 |
| 13 | 976 | 24 | 2.4 | 47 | 908 | 92 | 9.2 | 81 | 400 | 600 | 60 |
| 14 | 974 | 26 | 2.6 | 48 | 906 | 94 | 9.4 | 82 | 380 | 620 | 62 |
| 15 | 972 | 28 | 2.8 | 49 | 904 | 96 | 9.6 | 83 | 360 | 640 | 64 |
| 16 | 970 | 30 | 3 | 50 | 902 | 98 | 9.8 | 84 | 340 | 660 | 66 |
| 17 | 968 | 32 | 3.2 | 51 | 900 | 100 | 10 | 85 | 320 | 680 | 68 |
| 18 | 966 | 34 | 3.4 | 52 | 890 | 110 | 11 | 86 | 300 | 700 | 70 |
| 19 | 964 | 36 | 3.6 | 53 | 880 | 120 | 12 | 87 | 280 | 720 | 72 |
| 20 | 962 | 38 | 3.8 | 54 | 870 | 130 | 13 | 88 | 260 | 740 | 74 |
| 21 | 960 | 40 | 4 | 55 | 860 | 140 | 14 | 89 | 240 | 760 | 76 |
| 22 | 958 | 42 | 4.2 | 56 | 850 | 150 | 15 | 90 | 220 | 780 | 78 |
| 23 | 956 | 44 | 4.4 | 57 | 840 | 160 | 16 | 91 | 200 | 800 | 80 |
| 24 | 954 | 46 | 4.6 | 58 | 830 | 170 | 17 | 92 | 180 | 820 | 82 |
| 25 | 952 | 48 | 4.8 | 59 | 820 | 180 | 18 | 93 | 160 | 840 | 84 |
| 26 | 950 | 50 | 5 | 60 | 810 | 190 | 19 | 94 | 140 | 860 | 86 |
| 27 | 948 | 52 | 5.2 | 61 | 800 | 200 | 20 | 95 | 120 | 880 | 88 |
| 28 | 946 | 54 | 5.4 | 62 | 780 | 220 | 22 | 96 | 100 | 900 | 90 |
| 29 | 944 | 56 | 5.6 | 63 | 760 | 240 | 24 | 97 | 80 | 920 | 92 |
| 30 | 942 | 58 | 5.8 | 64 | 740 | 260 | 26 | 98 | 60 | 940 | 94 |
| 31 | 940 | 60 | 6 | 65 | 720 | 280 | 28 | 99 | 40 | 960 | 96 |
| 32 | 938 | 62 | 6.2 | 66 | 700 | 300 | 30 | 100 | 20 | 980 | 98 |
| 33 | 936 | 64 | 6.4 | 67 | 680 | 320 | 32 | | | | |
| 34 | 934 | 66 | 6.6 | 68 | 660 | 340 | 34 | | | | |

matrix of dependent variables), e.g., the concentration (10). The X and Y matrixes are scaled and mean-centered and then are simultaneously decomposed according to a bilinear model. PLSR is the most used multivariate calibration tool, mainly for predicting a single answer (11, 12).

(c) *Figures of merit*.—The evaluation of obtained results depends on the comparison of different parameters: the coefficient of determination of model fitting (R^2), root mean squared error of calibration (RMSEC), or root mean square of external validation and/or prediction (RMSEP), which can be calculated as follows:

$$Rel.error\ in\ \% = \sqrt{\frac{\sum_{i=1}^n (\tilde{y}_i - y_i)^2}{\sum_{i=1}^n (y_i)^2}} \times 100$$

where the \tilde{y}_i are the predicted values of concentrations by the model; y_i are the measured (experimental) values; and n is the

number of samples. RMSEC is an indication of model quality to fit experimental concentrations, while RMSEP is a parameter of model prediction quality. RMSEP and RMSEC are calculated in the same way, the only exception being that the estimates are the values from external validation samples (13).

Results and Discussion

FTIR Spectra

By simple visual inspection of spectra corresponding to different ratios of adulterated gasoline samples shown in Figure 1, it can be seen that it is difficult to distinguish differences between spectra.

Also, the existence of three clusters of spectral bands can be observed. The first group appears in the spectral region between 2800 and 3000 cm, where the prominent band at

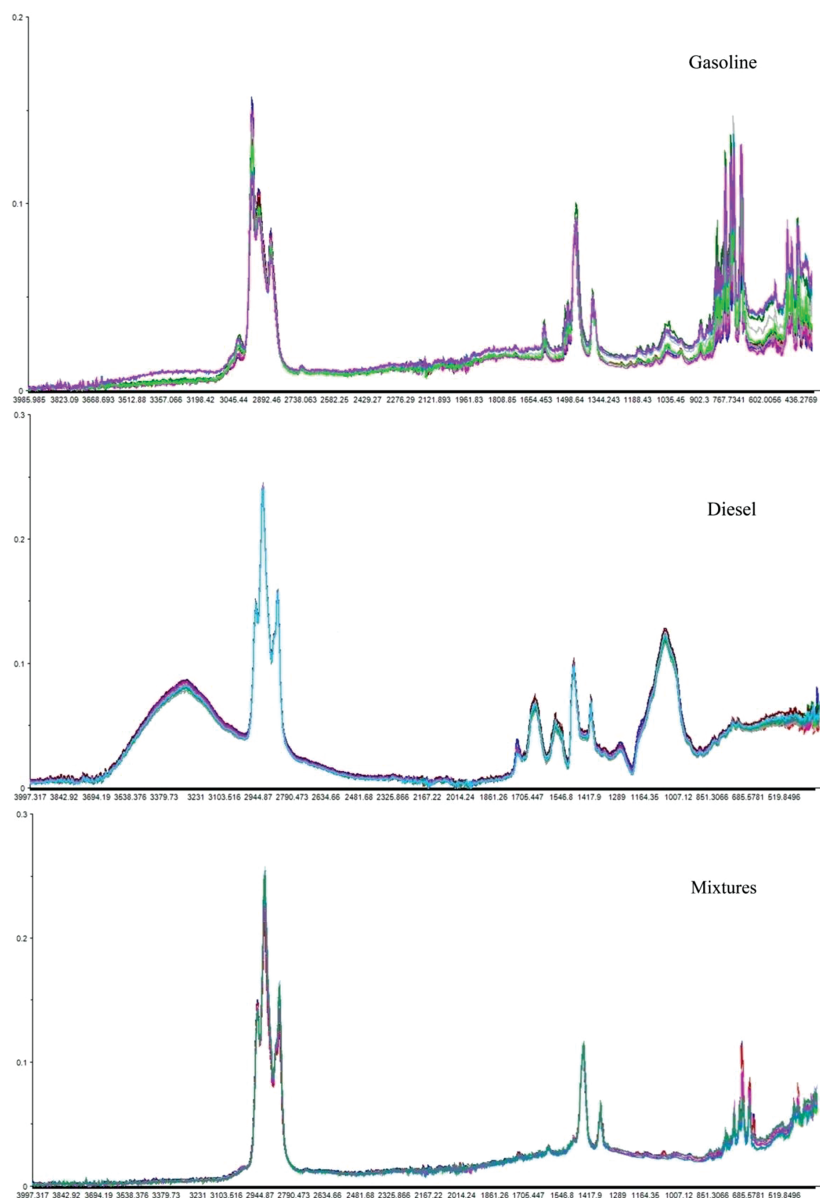


Figure 1. Fourier transform infrared spectra of three types of samples: gasoline, diesel, and adulterated gasoline samples with diesel.

2920 cm is attributed to symmetric CH_3 stretch. The band at 2850 cm attached to symmetric CH_2 stretch and the band at 2950 cm corresponds to asymmetric CH_2 stretch. The second cluster of bands is in the spectral region between 1200 and 1500 cm, where the most important band at 1450 cm can be assigned to the CH_3 asymmetric deformation. The signal at 1400 was attributed to the CH_2 scissor bending vibration and the 1380 cm assigned to the CH_3 symmetric deformation. The last spectral bands group is localized at the 900–700 cm region because of $\text{C}_1\text{--C}_2$ stretching (14, 15).

PCA

The first part of this study treats the visualization of data and verification of the possibility of distinguishing between the three types of products: two pure products of gasoline and diesel and, the third, adulterated samples by PCA. This is why PCA with full cross validation was applied to the data set of pure and adulterated samples exploring the full acquired spectral range (4000–400 cm).

The score graph obtained (Figure 2) shows that the two first PCs, PC1 and PC2, explained more than 98% of the original variability and display the formation of three groups, one formed by unadulterated gasoline samples, the second by diesel samples, and the third by adulterated samples.

Quantitative Determination of Gasoline Adulteration with Diesel Amount by PLSR Model

After detecting the adulteration of a gasoline sample by PCA, the level of adulterant was predicted by a PLSR model. Calibration model development was performed using preprocessed spectra to eliminate noise and scattering by applying two types of data pretreatments, Mean Center

Table 2. Performance of PLS model for determination of adulterant levels in gasoline samples

| | No. of latent variables | R^2 | RMSEC ^a | RMSEP ^b |
|------|-------------------------|-------|--------------------|--------------------|
| PLSR | 5 | 99.25 | 0.63 | 0.69 |

^a RMSEC = Root mean square error of calibration.

^b RMSEP = Root mean square of external validation and/or prediction.

Algorithms and Standard Normal Variate transformation. A set of 20 validation samples with known content of diesel was used to evaluate the model's performance by studying four figures of merit (number of retained latent variables, RMSEC, RMSEP, and percentile of explained variance) presented in Table 2.

The R^2 (99.25%) value explains the good linearity of the model built (Figure 3) and indicates a good agreement between predicted and real levels of adulterants in gasoline for calibration and validation data sets. The reduced RMSEC and RMSEP values (0.63 and 0.69) obtained with five Latent Variables indicate a high accuracy of the model in terms of calibration and prediction.

Conclusions

This work showed that the combination of FTIR with chemometric tools has a high usefulness to discriminate and quantify gasoline adulteration with diesel. The detection of adulterated samples was easily highlighted by the use of PCA. On the other hand, the PLSR model has shown a good accuracy to predict the amount of adulterant in the range 0 to 98% (w/w).

This approach based on coupling FTIR and PLSR can be useful in routine quality control tests of adulteration problems in the field of petroleum products.

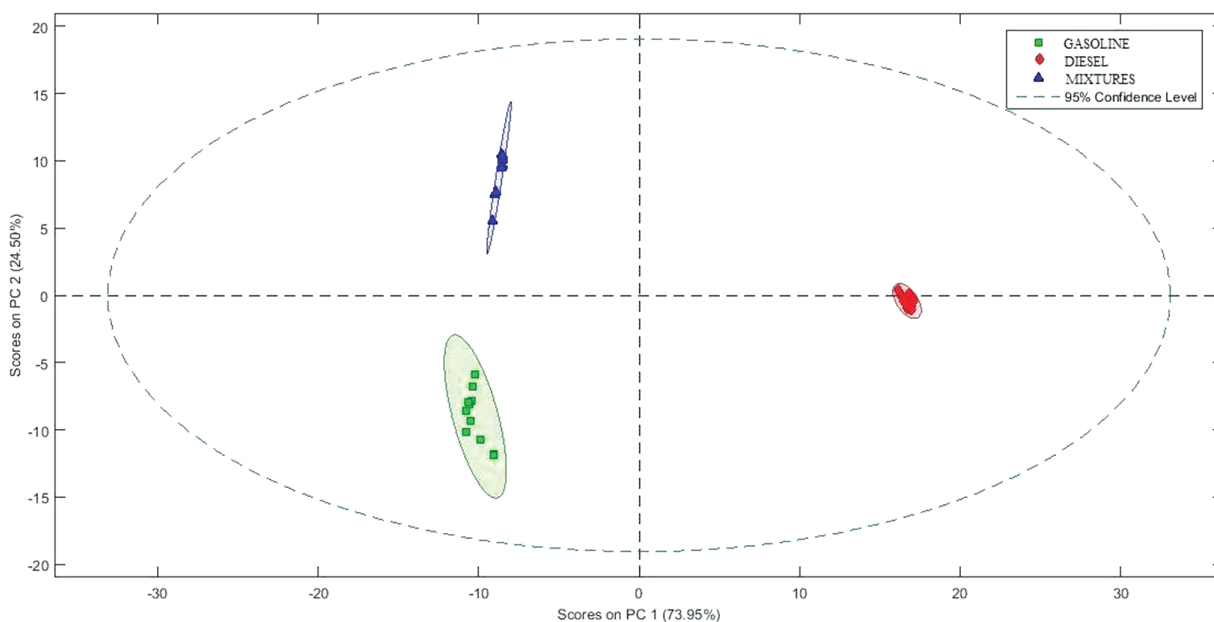


Figure 2. Score graph for the first two principal components obtained for the discrimination of three classes of samples: gasoline, diesel, and adulterated samples.

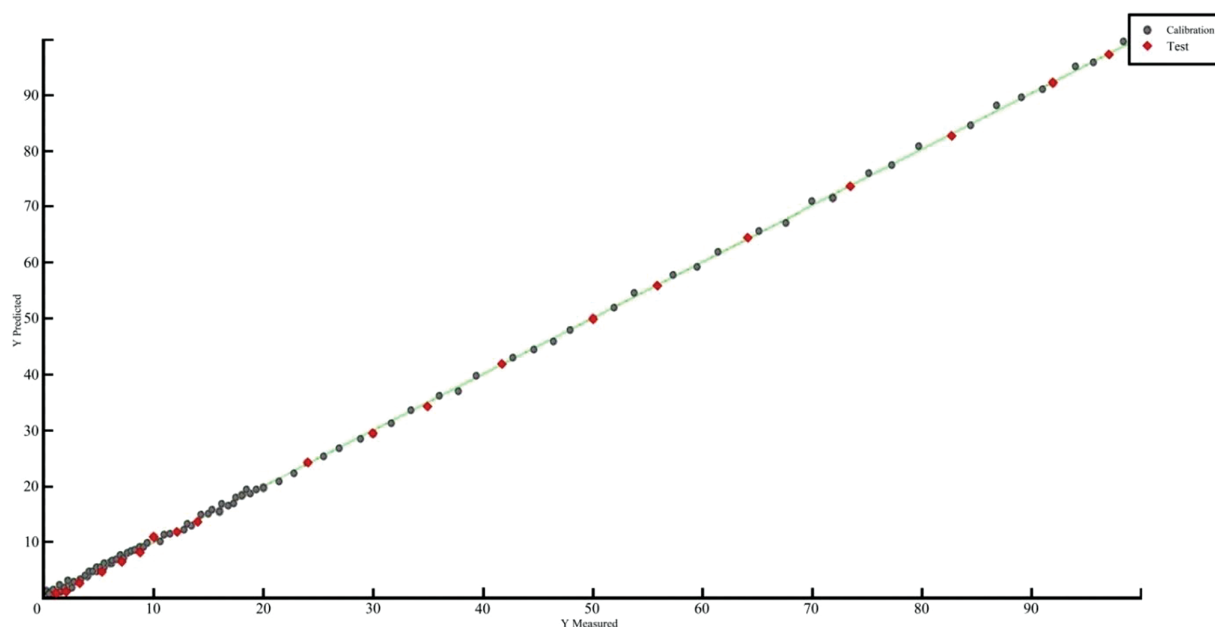


Figure 3. Predicted versus real adulterant contents graph obtained by PLS model; (●): calibration set, (◆): test set.

References

- (1) Irion, W.W., & Neuwirth, O.S. (2005) *Oil Refining* 7–12. doi:10.1002/14356007.a18_051
- (2) Takeshita, E.V., Rezende, R.V.P., de Souza, S.M.A.G.U., & de Souza, A.A.U. (2008) *Fuel* **87**, 2168–2177. doi:10.1016/j.fuel.2007.11.003
- (3) Ale, B.B. (2003) *J. Inst. Eng.* **3**, 12–16
- (4) Pasadakis, N., & Kardamakis, A.A. (2006) *Anal. Chim. Acta.* **578**, 250–255. doi:10.1016/j.aca.2006.06.072
- (5) Al-Ghouti, M.A., Al-Degs, Y.S., & Amer, M. (2010) *Talanta* **81**, 1096–1101. doi:10.1016/j.talanta.2010.02.003
- (6) De Luca, M., Terouzi, W., Ioele, G., Kzaiber, F., Oussama, A., Oliviero, F., Tauler, R., & Ragno, G. (2011) *Food Chem.* **124**, 1113–1118. doi:10.1016/j.foodchem.2010.07.010
- (7) Kalivas, J.H. (2010) *Compr. Chemom.* **3**, 1–32
- (8) Esbensen, K.H., & Geladi, P. (2010) *Compr. Chemom.* **2**, 211–226
- (9) Bro, R. (2014) *Anal. Methods* **6**, 433–459. doi:10.1039/c3ay41907j
- (10) dos Santos, V.H.J.M., Bruzza, E.D.C., de Lima, J.E., Lourega, R.V., & Rodrigues, L.F. (2016) *Energy & Fuels* **30**, 4905–4915. doi:10.1021/acs.energyfuels.6b00447
- (11) Kim, M., Han, I., & Han, C. (2003) *Comput.-Aided Chem. Eng.* **15**, 876–881
- (12) Geladi, P., & Kowalski, B.R. (1986) *Anal. Chim. Acta.* **185**, 1–17
- (13) Bassbasi, M., Hafid, A., Platikanov, S., Tauler, R., & Oussama, A. (2013) *Fuel* **104**, 798–804. doi:10.1016/j.fuel.2012.05.058
- (14) Mazivila, S., De Santana, F.B., Mitsutake, H., Gontijo, L.C., Santos, D.Q., & Neto, W.B. (2015) *Fuel* **142**, 222–226. doi:10.1016/j.fuel.2014.11.014
- (15) de Miranda, A.M., Castilho-Almeida, E.W., Ferreira, E.H.M., Moriera, G.F., Achete, C.A., Silva, R.A., Dos Santos, H.F., & Jorio, A. (2014) *Fuel* **115**, 118–125. doi:10.1016/j.fuel.2013.06.038