Analytica Chimica Acta 863 (2015) 9–19

Contents lists available at ScienceDirect

Analytica Chimica Acta

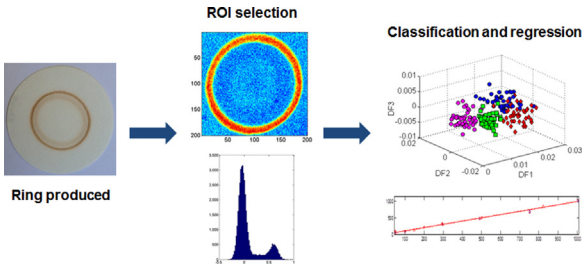
journal homepage: www.elsevier.com/locate/aca

Determination of detergent and dispensant additives in gasoline by ring-oven and near infrared hypespectral imaging

Lívia Rodrigues e Brito [a], Michelle P.F. da Silva [a], Jarbas J.R. Rohwedder [b], Celio Pasquini [b], Fernanda A. Honorato [c], Maria Fernanda Pimentel [c,*]

[a] Departamento de Química Fundamental, Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil
[b] Instituto de Química, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil
[c] Departamento de Engenharia Química, Universidade Federal de Pernambuco – UFPE, Avenida Prof. Arthur de Sá s/n, Cidade Universitária, 50740-521 Recife, PE, Brazil

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Available online 30 December 2014

Keywords:
Gasoline
Detergent dispersant additives
Ring oven
Hyperspectral imaging
Near infrared spectroscopy

ABSTRACT

A method using the ring-oven technique for pre-concentration in filter paper discs and near infrared hyperspectral imaging is proposed to identify four detergent and dispersant additives, and to determine their concentration in gasoline. Different approaches were used to select the best image data processing in order to gather the relevant spectral information. This was attained by selecting the pixels of the region of interest (ROI), using a pre-calculated threshold value of the PCA scores arranged as histograms, to select the spectra set; summing up the selected spectra to achieve representativeness; and compensating for the superimposed filter paper spectral information, also supported by scores histograms for each individual sample. The best classification model was achieved using linear discriminant analysis and genetic algorithm (LDA/GA), whose correct classification rate in the external validation set was 92%. Previous classification of the type of additive present in the gasoline is necessary to define the PLS model required for its quantitative determination. Considering that two of the additives studied present high spectral similarity, a PLS regression model was constructed to predict their content in gasoline, while two additional models were used for the remaining additives. The results for the external validation of these regression models showed a mean percentage error of prediction varying from 5 to 15%.

1. Introduction

Gasoline is a complex mixture of substances, comprising mainly hydrocarbons. In Brazil, the gasoline sold in fuel stations presently

* Corresponding author. Tel.: +55 81 21267235; fax: +55 81 21267235.
E-mail addresses: mfernanda.pimentel@ufpe.br, mfp@ufpe.br (M.F. Pimentel).

http://dx.doi.org/10.1016/j.aca.2014.12.052
0003-2670/

contains 25% (v/v) of ethanol. This percentage may vary depending on the availability and price of the alcohol [1,2].

The gasoline may also contain detergent dispersant additives, compounds that improve the engine performance by cleaning the engine (eliminating the deposits and minimizing future formation). The benefits of adding these additives in gasoline include the reduction of deposits by 60–70%, the decrease in fuel consumption by 3%, a reduction of the emissions of carbon monoxide and hydrocarbons by 50–60% and nitrogen oxides by 20% [3].

Environmental concerns about gas emissions in the atmosphere have contributed to the increase of the quality control of fuels. Following the example of some countries like USA, that in 1996 instituted the mandatory addition of detergent additives [3], the Brazilian Agency of Petroleum, Natural Gas and Biofuel (ANP) has established that from 1st July, 2015, all Brazilian commercial gasoline will contain detergent dispersant additives [4]. The concentration of the additives will depend on their effectiveness, but the maximum concentration will be 5000 mg kg^{-1} or 0.5% (w/w). The mandatory addition of the additives requires the development of analytical methodologies to determine their concentrations in gasoline in order to verify the compliance with the law.

The substances used as additives in Brazil are an industrial secrecy. However, it is known that these additives have a high boiling point, being stable up to 250 °C and decomposing at temperatures higher than 300 °C without leaving solid residues [5]. There are many works concerning the development of detergent dispersant additives reported in the literature. Some examples of the compounds used as detergent dispersant additives are: succinimide derivatives [6], quaternary ammonium salts [7] and amines [8]. Actually, there is no standard or widely accepted method for additive analysis and few works have reported on this subject. Most of them are based on size exclusion chromatography (SEC) [9–11] due to the high molecular weight of the additives compared to the molecular weight of the gasoline compounds, but some works use other techniques such as thin-layer chromatography [12]. SEC is a slow technique requiring about 30 min for one chromatographic run and needs individual optimization of the detector (evaporative light scattering) for each type of additive. As the type of additive cannot be previously known, considering gasoline samples collected in the market, the usefulness of the method is restricted to the quality control performed by manufacturers who know the type of substance added. Thin-layer chromatography is also slow and presents several limitations regarding its use as quantitative technique.

In a previous work of our group, infrared spectroscopy was employed in screening methods developed to classify gasoline samples as with or without additives [13]. In that work, the additives in the gasoline samples were pre-concentrated using the atmospheric distillation (following the ASTM D86-11b standard procedure [14]), and the medium and near infrared spectra were acquired in the distillation residues. Despite the fact that there was a 100% correct classification rate, the methods were not efficient for the quantification of the additives nor for classification of the samples according to type of additive.

For near infrared analysis, a pre-concentration stage is compulsory because the maximum concentration of the additives in gasoline is below 0.1%, the regular limit of quantification for the technique. Besides distillation, several techniques can be used to pre-concentrate the samples. For this purpose, the ring-oven technique has shown a considerable potential [15].

Herbert Weisz developed the ring-oven technique in 1954 as a pre-concentration micro-analytical technique for qualitative analysis of small sample volumes [16]. Over the years, this technique found application in different areas and has been used not only for qualitative analysis but also for semi-quantitative analysis [16]. This technique employs a simple apparatus comprised of a heated cylindrical block (the ring-oven) made of aluminum. A volume of sample (usually amounting to a few tens to hundreds of microliters) is delivered in discrete portions to a filter paper disc placed in the oven, allowing for solvent evaporation between successive additions. The remaining material can be washed by a suitable solvent, which is also added in discrete portions and evaporates after carrying the solutes to the border of the filter disc, where they accumulate to form a thin ring profile.

Nowadays, with the development of new micro-analytical techniques such as near infrared (NIR) hyperspectral imaging and laser induced breakdown spectroscopy (LIBS), it is possible to analyze small areas of solid samples [15]. Therefore, it has become possible to associate the ring oven with these modern analytical techniques in order to improve the ring-oven quantitative analysis, once made by naked eye. The ring-oven technique can be used as a pre-concentration stage, before measurement. This technique was revisited and modified by Cortez and Pasquini to pre-concentrate metals in fuel ethanol before their determination by LIBS [15].

Hyperspectral imaging associates spatial and spectral information and has been used in many different areas such as pharmaceutical [17], forensics [18] and medical [19]. With the development of new and faster imaging instruments (cameras), computers and software, hyperspectral imaging techniques have been expanded to obtain qualitative and quantitative information [20]. The images contain large volumes of data. In order to extract relevant information from this data, it is necessary to use chemometric techniques. Some classical multivariate techniques have been adapted to work with hyperspectral imaging, such as principal component analysis (PCA) and multivariate curve resolution (MCR) [20].

PCA has found several applications in image analysis. This multivariate data analysis technique can be used as an exploratory tool, to detect similarities between the pixels spectra, and to compress the images (reducing the data dimensionality). MCR is a resolution technique used to obtain the pure spectra and the concentration maps of a mixture of sample components. The MCR algorithm used in this work was the multivariate curve resolution-alternating least squares (MCR-ALS) which calculates the concentration matrix $\mathbf{C}$ and the pure spectra $\mathbf{S^T}$ iteratively by the ALS algorithm [21]. The ALS optimally calculates the $\mathbf{C}$ and $\mathbf{S^T}$ matrices during the iterative cycles, using different constraints such as non-negativity of the spectra and concentration [21].

Before applying these techniques, it is necessary, however, to pre-process the images [20]. At this stage it is possible to compress the images, reducing the storage space; to select the region of interest (ROI), eliminating the undesired parts of the images such as the background; and to pre-process spectral data to eliminate instrumental noise and unwanted effects, such as light scattering.

The selection of the ROI can be done manually by visual inspection of a PCA scores image, by selecting a threshold value using the scores of a PCA model, and by using histograms showing the distribution of the PCA score values [20]. Sometimes, the selection of ROI is not straightforward because there may not be a clear threshold to separate the different parts of the image, making the process very time consuming [20]. Furthermore, when score histograms are employed, it may be necessary to adopt a robust criterion to avoid selecting any of the background, even though part of the ROI can be missed. Alternatively, it may be necessary to ensure that all the ROI is selected, even though some parts of the background are included in the selection.

After the selection of the ROI, it may be necessary to have a representative spectrum in order to build a multivariate classification or regression model. A partial least squares (PLS) model, for example, can be built using the representative spectra of the ROI of the images in order to predict the bulk concentration of the components present in the sample. After the validation, the model can be used to predict the bulk concentrations of unknown samples using data of new images generated by those samples [22].

Linear discriminant analysis (LDA) can be applied to images, also using representative spectra, when it is necessary to classify samples based on their components. In LDA the number or variables needs to be lower than the number of training samples; so when using spectral data it is necessary to perform a variable selection previously. Variable selection can be done manually or by using a variable selection algorithm such as successive projections algorithm (SPA) [23,24] and genetic algorithm (GA) [25].

In the present work, an analytical method is proposed based on pre-concentration by the ring-oven technique and NIR hyperspectral imaging, first to identify the additive present in gasoline samples and then submit its representative spectrum extracted from the ROI to be evaluated by the proper model in order to determine its concentration.

2. Materials and methods

2.1. Samples

The Brazilian National Petroleum Agency (ANP) provided four detergent dispersant additives (coded as G, T, W and Y) which should be added in Brazilian gasoline. Due to the high viscosity of the additives, standard solutions containing 100,000 mg kg^{-1} of the compounds were prepared in toluene. Two sample sets were prepared; the first one was used in the preliminary studies (to select the best ring oven operational parameters, the ROI, and the best strategy to extract the representative spectra) and the other one to build the final classification and regression models.

For the preliminary studies, a gasoline without additives was spiked, in the laboratory, with the solutions of the four additives, resulting in 32 gasoline samples, 8 for each additive, whose concentrations (ranging from 50 to 1,000 mg kg^{-1}) are as shown in Table 1 (with an "x" symbol). Triplicate rings were prepared from each of these gasoline samples (named here type 1 gasoline samples) on filter paper discs (Whatman 42). Twenty-four rings from each additive (96 in total) were then obtained.

To build the final classification and regression models, the samples were prepared using 10 different gasoline samples without additives, collected from different distributors, to encompass the variability of the gasoline composition. The concentrations of the additives in these samples (named here as type 2 gasoline samples) are marked with an "o" in Table 1. Using these type 2 samples, triplicate rings were prepared for each additive, resulting in 108 rings. A total of 204 rings were then prepared (considering the sum of type 1 and type 2 gasoline samples). It is important to mention that the preparation of the rings from type 1 samples was carried out over a short period of time in cold and wet weather as compared with the rings from the type 2 samples which were prepared six months later (during the dry and hot season).

2.2. Ring oven pre-concentration

The system used in this work for the ring preparation was similar to the one described by Cortez and Pasquini [15]. The main differences are: a syringe (Microlab® PSD/4, Hamilton) was employed for sampling and sample delivering, and a cylindrical heating element was inserted directly to the cylindrical base of the oven instead of using a heating tape.

The ring preparation comprises the following steps: (1) the heating system is turned on and reaches thermal equilibrium at the desired temperature (150 °C); (2) the filter paper (Whatman 42) is placed in the oven; (3) the syringe and tubes are prefilled with washing solvent (toluene); (4) an air bubble is aspirated through the capillary followed by the selected sample volume; (5) the sample volume is delivered in programmed discrete volumes to the paper. Enough time is given between additions for the volatile components of the gasoline sample to evaporate; (6) a certain volume of the washing solvent (toluene) is delivered also in discrete volumes to the paper until all the chosen volume has been completed. Time is given to evaporate the toluene between additions; (7) the paper is kept in the heated oven for additional 5 min after the sample transference and washing operations are finished.

Toluene was employed in the preparation of the samples and as washing solvent because previous tests showed that the additives have high solubility in this solvent, it evaporates at the oven temperature and it is a hydrocarbon usually present in gasoline. The oven temperature was kept at (150.0 ± 0.5) °C, considering the temperature of evaporation of the toluene and of the most of volatile compounds present in gasoline. In addition, this temperature causes negligible damage to the filter paper.

Different protocols for ring preparation were tested, varying: (1) the total volume of sample, (2) the total volume of toluene (washing solvent), (3) the volume of each discrete addition and (4) the time interval between additions. With several protocols tested, the rings that were not well defined were immediately discarded and their images were not acquired. The four conditions for which the ring images were acquired are presented in Table 2.

The third condition (Table 2) evaluated for the ring preparation was chosen (1,800 μL of sample and 240 μL of toluene) because the rings formed were more clearly defined in the score images than the rings attained by using the first and second condition. Under the third condition, it takes about 35 min to prepare a ring. The additives are pre-concentrated about 78 times in the ring region. Though the time demanded for the ring preparation can be considered high, the process is mechanized and precludes the presence of an operator. Besides, the pre-concentration operation can be considered less aggressive to the environment as low volumes of non-toxic solvent are employed.

Table 1
Concentrations of the additives in the gasoline samples prepared from one gasoline sample without additive (type 1, marked with an "x") and from 10 different gasoline samples without additives (type 2, marked with an "o").

Additive	Concentration (mg kg^{-1})										
	50	100	150	200	300	400	500	600	750	800	1,000
G	x o	x o	x	x o	x o	o	x o	o	o	x	x o
T	x o	x o	x	x o	x o	o	x o	o	–	x o	x o
W	x o	x o	x o	x o	x	o	x o	o	x o	–	x o
Y	x o	x o	x o	x o	x o	o	x o	o	x	o	x

Table 2
Ring preparation conditions.

Condition	Sample			Toluene		
	Volume (μL)		Time (s)	Volume (μL)		Time (s)
	Total	Discrete		Total	Discrete	
1	600	10	10	150	25	10
2	1,200	10	10	150	25	10
3	1,800	12	10	200	25	10
4	2,400	10	10	240	40	20

2.3. Image acquisition

The near infrared hypespectral images (HI-NIR) of the filter papers containing the rings were acquired using a SisuCHEMA SWIR hyperspectral camera (Specim, Finland). The spectral range employed was 928–2,524 nm, the spectral sampling was 6.3 nm, and the spectral resolution was 10 nm. Images were acquired with pixel sizes 150 μm × 150 μm. The filter papers were fixed on a Teflon board before image acquisition in order to keep them uniformly flat.

2.4. Image processing

The original size of the images was initially reduced to contain only the ring and its adjacent region, resulting in files containing 200 × 200 pixels and 256 wavelengths per pixel. Noisy regions of the spectra (from 928 to 1073 nm and from 2493 to 2524 nm) were eliminated and the spectra were then submitted to standard normal variate (SNV).

Three approaches were evaluated for the selection of the region of interest (ROI): (A) manual selection of the limits in the MCR-ALS concentration maps, (B) visual selection in the score images of the first principal component and (C) automatic selection using a pre-defined threshold value from the histograms of the distribution of the score values in the first principal component. These approaches were compared using preliminary PLS models built using "type 1 gasoline samples".

In the MCR-ALS the only constraint used was the non-negativity of concentration values. The average spectrum of the ROI (selected visually in the PC1 score image) of the rings prepared from the solutions containing 1000 mg kg^{-1} of each additive in toluene was used to estimate the spectra of pure additives. The average spectrum of the ROI (selected visually in the PC1 score image) of the rings prepared using a gasoline sample (without additives) was used as initial estimate of the resulted "pure" gasoline spectrum. Finally, average spectrum of a filter paper image was used to obtain the paper spectrum.

After the ROI selection, the representative spectra of the selected pixels were extracted in two ways: (i) by averaging the spectra set of the ROI pixels, and (ii) by summing up the same spectra set. After extraction, the spectra were submitted to different pre-processing techniques, such as Savitzky–Golay smoothing (3 and 7 point windows and second order polynomial), Savitzky–Golay first and second derivatives (7 and 11 point windows and second order polynomial) and range normalization. Outlier detection was performed visually (comparing the spectra of one ring with the spectra of the replicate rings and with the spectra of the other samples containing the same additive) and using the PCA plots. Outlier detection was performed visually (comparing the spectra of one ring with the spectra of replicate rings and with the spectra of rings of other samples containing the same additive) and by using PCA scores plots. The spectra that were eliminated were just the ones visually different from the others and, consequently, showing unexpected values of scores when compared with the scores of total spectra set.

Classification models were built to identify the type of additive present in gasoline samples. Different regression models (according to the type of the additive) were built to quantify the concentration of additives in the samples.

2.4.1. Classification

Classification models were built using LDA. All individual rings of all samples were employed. The selection of samples to compose the training (60% of samples), validation (20% of samples) and test (20% of samples) sets was carried out using the Kennard–Stone (KS) algorithm. The number of samples of each additive (Table 3) may be different from each other because some outliers were excluded.

Table 3
Number of training, validation and test samples of each additive.

Additive	Sets		
	Training	Validation	Test
G	28	10	10
T	29	10	10
W	27	9	9
Y	30	10	10
Total	114	39	39

The variable selection was carried out using the successive projection algorithm (SPA) and the genetic algorithm (GA). Three approaches were used: (1) the variables were selected from the full spectra, (2) the variables were selected only from the regions in which the additives had characteristic absorption bands (identified by the MCR-ALS optimized spectra) and (3) as the number of variables in those regions (53) was lower than the number of samples in the training set, the use of all of these variables in the model building was also evaluated without further selection. For the GA algorithm, 100 generations with 200 chromosomes, mutation probability of 10% and crossover probability of 60% were used. The algorithm was repeated five times and the best solution was the one with the lowest number of misclassification errors in the test set.

2.4.2. Regression

The regression models were built using the PLS algorithm. The models were built using (1) the full spectra, (2) the variables selected by the Jack-knife (JK) algorithm and (3) only the regions in which the additives had characteristic absorption bands.

The preliminary PLS models were built using type 1 gasoline samples and considering each triplicate ring as a different sample, resulting in a total of 24 sample (rings) for each type of additive (8 different concentrations, as shown in Table 1 with an "x"). The final PLS models were built using type 1 and type 2 gasoline samples. The representative spectrum obtained for each ring were averaged, resulting in 17 mean representative spectra for each additive, as shown in Table 1 (with "x" and "o").

In this part of the work, the models were validated for prediction using 5 gasoline samples purchased from different distributors, not employed in modeling stage.

The original image format conversion to the Matlab format was carried out using Evince 2.7.0 (Umbio). Image cutting, application of SNV, ROI selection, PCA and LDA were carried out using Matlab® R2010a 7.10.0.499 (MathWorks). Spectral pre-processing and PLS were carried out using Unscrambler® X10.3 (CAMO S.A.).

3. Results and discussion

Fig. 1a shows the original spectra of an image acquired from a paper disc containing a ring prepared by pre-concentrating a gasoline sample containing 50 mg kg^{-1} of additive G. Fig. 1b shows the spectra of the same image but without the noisy spectral regions and pre-processed using SNV. The SNV was applied to correct for light scattering effects. The spectra of the image are very similar to the paper spectrum (Fig. 1c) because most of the image is composed of pixels from the ring substrate, and even after the pre-concentration step, the concentration of the additive is low. The compensation of the paper spectral contribution previously to all the analysis has also been evaluated in the initial stage of this work. However, the extraction of the additives spectra from the images was compromised because their concentration in each pixel is low and the spectra resulting after paper subtraction did not allow

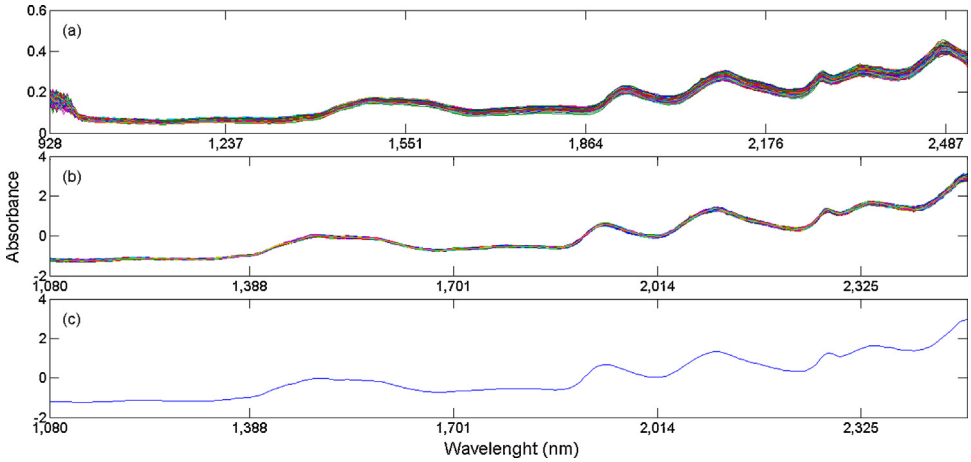


Fig. 1. (a) Original spectra and (b) spectra without the noisy spectral regions and pre-processed with SNV of an image acquired from a paper containing a ring prepared by using a gasoline sample containing 50 mg kg^{-1} the additive G and (c) mean paper spectrum.

identifying the ring spectra. Thus, the compensation of the paper spectral contribution was carried out only after the extraction of the ROI spectra and just in the case of "the summed spectra approach", for representative spectra extraction (Section 3.2).

3.1. ROI selection and representative spectra extraction: preliminary tests

As mentioned, PLS calibration models were used preliminary to evaluate the best approach for ROI selection and spectra extraction. In these tests, the models were built using the rings of the samples prepared during a short period of time (one week) using only one gasoline in order to avoid matrix effects (type 1 gasoline samples).

As most of the image is composed by the pixels of the paper, it was necessary to select the pixels of the ring, which constitute the ROI, before building the models.

Three approaches were evaluated to guide the ROI pixel selection, the first (A) was based on MCR-ALS concentration maps and the other two on the PCA (B) score images and (C) on the corresponding histograms of PCA score values distribution.

Regarding to the representative spectra extraction, the average spectra and the summed spectra were calculated. It was observed that the rings obtained by pre-concentrating samples containing higher concentrations of additive present a larger diameter and thickness, as shown in Fig. 2, for MCR results. Therefore, the averaging may not be the best alternative to obtain the representative spectrum because the spectral intensities of larger rings (higher concentrations) could be lower than expected, due to the higher number of pixels selected. The sum of spectra from the ROI, then, might be a better alternative because it accounts for all the information related to the additives. The problem is that this procedure adds the information from the paper as well. Therefore, to use the summed spectra approach, it is necessary to subtract the same number of summed paper spectra (pixels) in order to enhance the information from the additives. Therefore, for each ROI spectra set (pixels) that is summed, an equivalent number of paper spectra (pixels) needs to be subtracted. In these preliminary tests, the paper spectra employed for compensation was the mean spectrum of an image obtained from one filter paper of the same batch employed for ring production. To compare these two strategies, PLS models were built using both representative spectra (average and summed).

3.1.1. Approach (A): MCR-ALS – concentration maps

This first approach was based on the visual determination of limit values in the concentration maps. Only the pixels whose concentration values were higher than the limit were selected. In order to have better resolution results, the augmented MCR-ALS was used, grouping the rings obtained with samples in different concentrations (Fig. 2). The optimized spectra in Fig. 3 show that there are two regions (from 1,670 to 1,801 nm, and from 2,238 to 2,425 nm) where the spectra of the additives show a significant

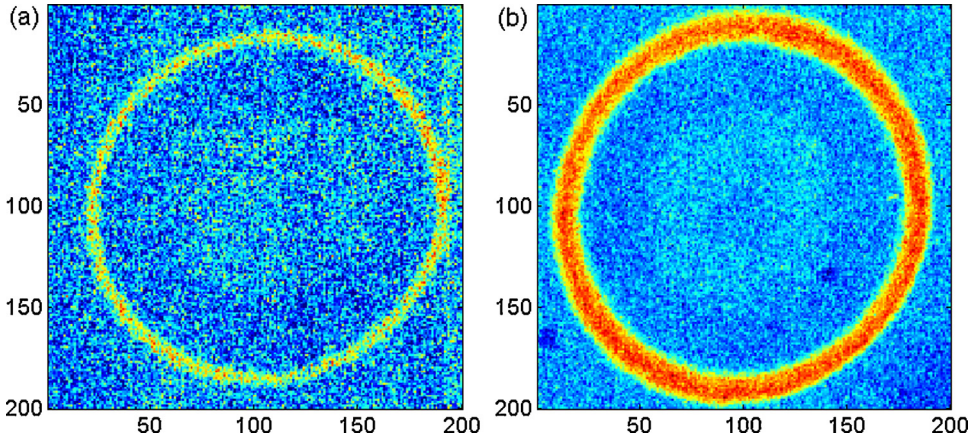


Fig. 2. Approach (A) to select the ROI based on MCR-ALS and concentration limit. The concentration distribution map of a sample with (a) 50 mg kg^{-1} and (b) 1,000 mg kg^{-1} of additive G.

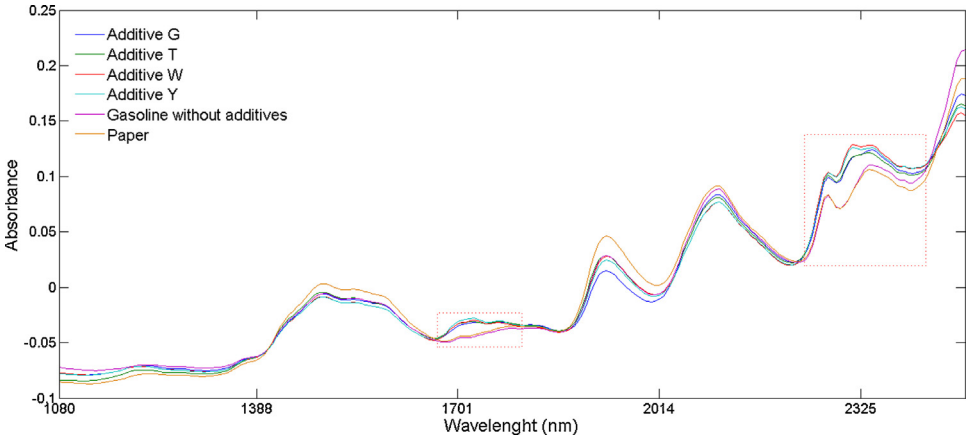


Fig. 3. MCR-ALS optimized spectra of the additive, paper and the gasoline without additives.

Table 4
Best PLS results for the approaches evaluated for ROI selection, and representative spectra extraction. LV is the number of latent variables.

Approach	Additive	Representative spectra	LV	RMSECV (mg kg^{-1})
(A)	G	Summed[a]	1	133
	T	Summed	1	64
	W	Summed	2	70
	Y	Summed[b]	1	43
(B)	G	Summed[b]	1	102
	T	Summed[b]	1	52
	W	Summed[b]	2	64
	Y	Summed[b]	2	29
(C1) (grouped images)	G	Summed[a]	1	71
	T	Summed[a]	1	58
	W	Summed	2	63
	Y	Average[b]	1	46
(C2) (individual images)	G	Summed[a]	1	79
	T	Summed[b]	1	46
	W	Summed[b]	2	41
	Y	Summed[b]	2	40

RMSECV, root mean square error of cross validation.
[a] Variables selected with the Jack-knife algorithm.
[b] Variables from 1670 to 1801 nm and from 2238 to 2425 nm.

spectral difference from the spectra of gasoline without additive. These two regions are assigned to the first overtones of C—H stretching and to the combination regions of C—H bonds [26]. The best results of the regression models for each additive are shown in Table 4. The drawback of this approach is that it is subjective and time consuming.

3.1.2. Approach (B): PCA – score images

PCA was performed using only the regions of the spectra in which the spectra of the additives show significant differences from the spectrum of paper and from the gasoline without additives. The two approaches based on PCA were carried out only on PC1 results because this PC accounted for most of the information related to the ring.

The approach based on PCA score images, approach (B), was performed using the rings produced with additives in different concentrations grouped together to improve the score image definition of the rings obtained by the lower concentration of the additives. PC1 score images were then used for visual definition of the limits (Fig. 4). The best results of the regression models for each additive are shown in Table 4. The main issue with this approach is the same as in approach (A).

3.1.3. Approach (C): PCA – histograms

In order to make the selection of the ROI pixels more objective, a third approach was investigated based on the determination of a PCA score threshold value capable to better split the image pixels belonging to the ring (ROI) and those belonging to the paper (background). Typical histograms of the PCA score values from

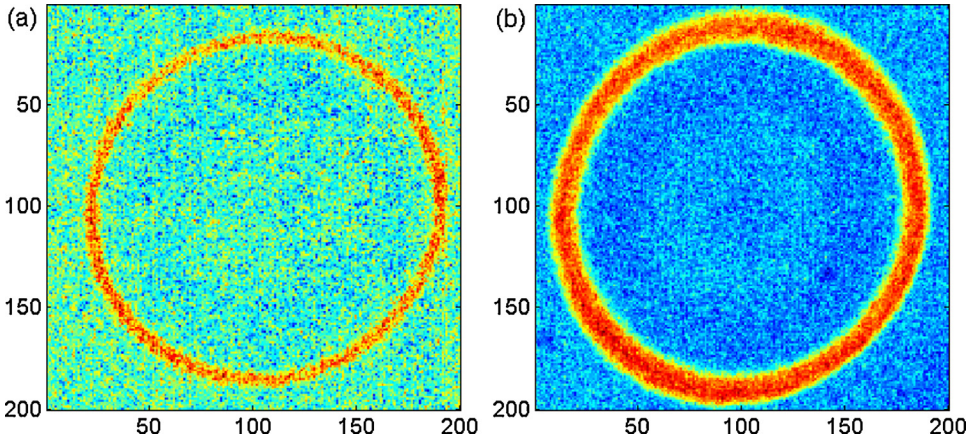


Fig. 4. PC1 score images obtained from samples with (a) 50 mg kg^{-1}and (b) 1,000 mg kg^{-1} of additive G.

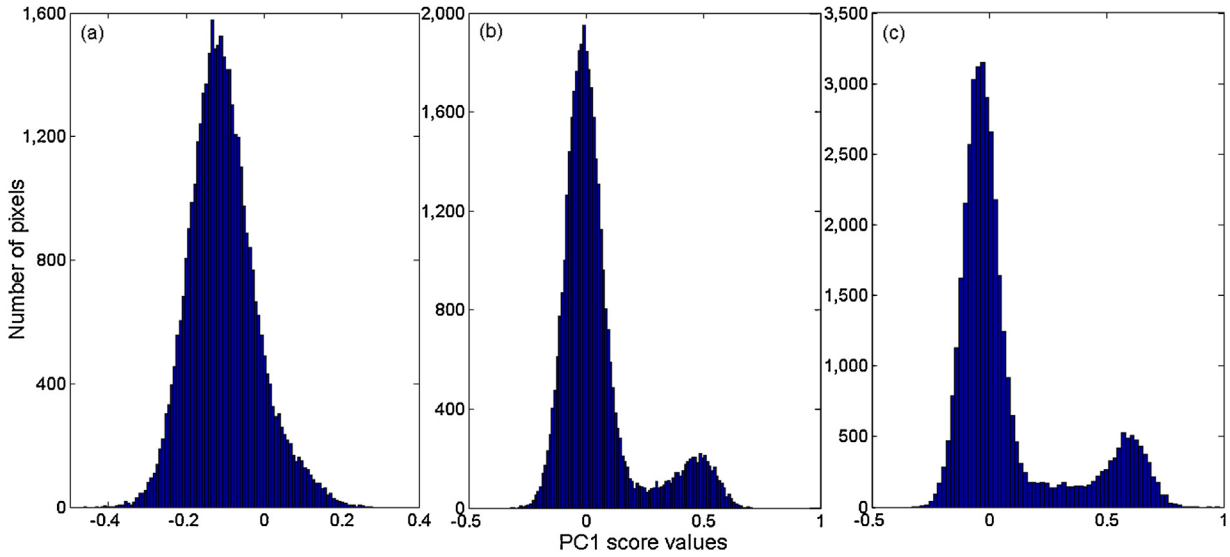


Fig. 5. Histograms of the PC1 score values for samples with (a) 50 mg kg^{-1}, (b) 500 mg kg^{-1} and (c) 1,000 mg kg^{-1} of additive G.

images are shown in Fig. 5 for gasoline samples containing additive G in different concentrations. The spectra of the pixels containing this additive are the most difficult to distinguish from the paper background, and a clear distinction (a second mode) is only observed in Fig. 5b and c for higher concentrations. Nevertheless, the histogram of a sample with the lowest additive concentration (50 mg L^{-1}), Fig. 5a, shows a slight asymmetry on the right side caused by the superposition of pixels containing spectral information from the ring and from paper. This histogram was employed to find the best value for a common threshold to be used in the ROI selection.

The procedure to determine the threshold is based on the determination of the standard deviation of the score values that can be unequivocally attributed to pixels belonging to the paper, found at the left side of the more intense mode in the histogram. Then, black-and-white images are produced by attributing to the ROI the pixels (black) beyond one, two, three and four standard deviation values from the maximum toward the right side of the histogram, as shown in Fig. 6a–d. The criteria to select the threshold, expressed in units of standard deviation, was based on the number of pixels correctly attributed to the ROI (black). The number of pixels correctly attributed to the ROI, visually identified,

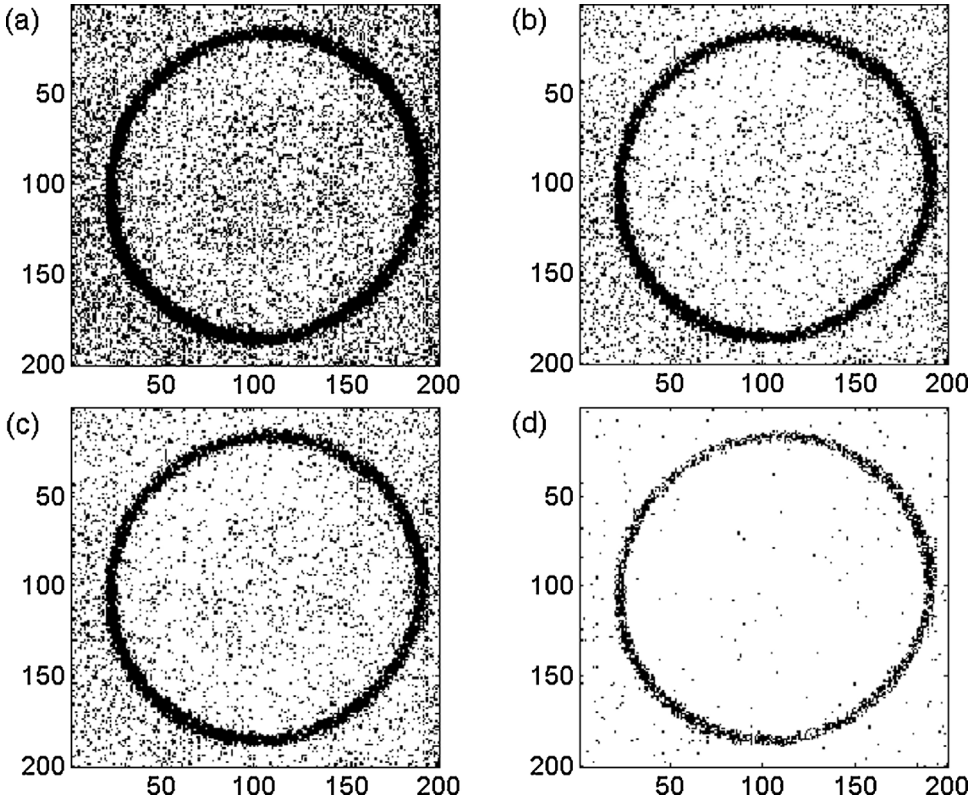


Fig. 6. Images of a ring from a sample with 50 mg kg^{-1} of additive G with ROI selected by using histograms of the score images. (a) With one standard deviation, (b) with two standard deviation, (c) with three standard deviation, (d) with four standard deviation.

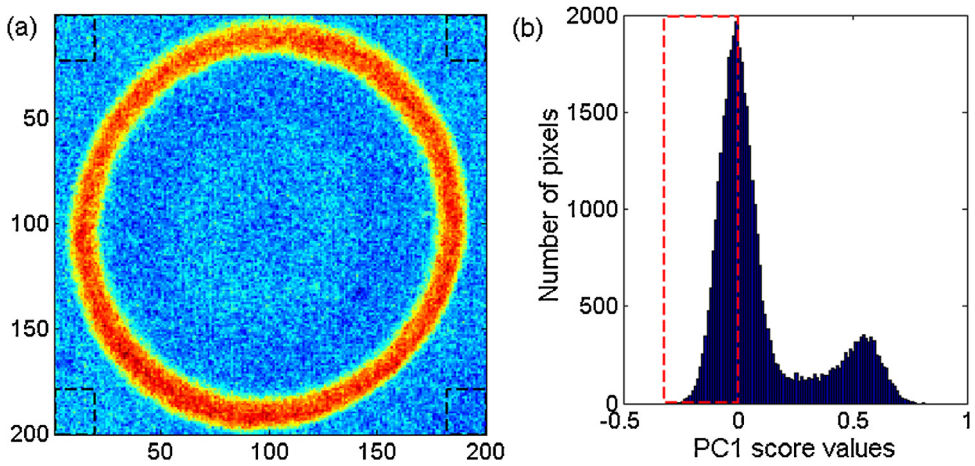


Fig. 7. The two method to extract the paper spectra: (a) approach (D) – pixels on the edges of the images and (b) approach (E) – pixels whose score values in PC1 were lower than the histogram maximum.

must be maximized, so more ROI pixels and less paper pixels are selected.

Observing Fig. 6a–d, it was possible to conclude that by taking the threshold equal to three (3) standard deviations more pixels of the ROI (ring) were selected and less pixels of the paper background were wrongly attributed to the ROI. Therefore, a threshold value of three standard deviation was selected to maximize the number of pixels correctly attributed to the ROI.

Once the threshold has been determined, it was applied for each image independently as an objective procedure to select the ROI. The score values located at the left side from the maximum of the more intense mode of the score values histogram for each image are used to calculate the standard deviation. The pixels with score values exceeding 3 standard deviations on the right of the same histogram are taken as belonging to the ROI. The procedure is independent of the paper batch and of its physical–chemical properties, such as moisture.

In approaches A and B, MCR-ALS and PCA models were built with all the rings together to improve the visualization of the rings produced from samples with low additive concentrations. The third approach is not based on a visual inspection and the PCA models were built (C1) grouping all the rings (as in approaches A and B) and also a PCA model was built with only one ring at a time (C2). The best results of the regression models for each additive are shown in Table 4.

Comparing the results presented in Table 4, it can be seen that the full RMSECV values for the models based on the histograms of a PCA performed with the rings individually (C2) were lower for the additives T and W and were statistically equivalent, at a 95% confidence level, to the lower values found for the additives G and Y. Therefore, the C2 approach was considered the best. Regarding the determination of the representative spectra, the sum of the spectra of the ROI pixels was chosen because it results in better models.

3.2. Influence of the paper in the spectra extraction process

For all calculations described before, the rings (produced by type 1 gasoline samples) and images were obtained within a short period of time (one week). When the spectra of these rings were compared with the rings obtained from the type 2 gasoline samples, significant differences were observed in regions from 1,413 to 1,657 nm and from 1,864 to 2,057 nm, which are assigned to the combination regions of O—H bonds of water molecules [26]. As mentioned in Section 2.1 these rings were obtained in different periods of the year (cold and wet weather and hot and dry weather) and this is reflected in the paper spectra (variability in humidity). In order to avoid this kind of problem, the compensation of the paper information should be carried out individually. So the paper spectra region around the ring of each image must be considered in the calculation.

Two methods (Fig. 7) were evaluated to decide the best way to extract the paper spectra. The first was (D) to select the pixels on the edges of the images and the second was (E) to select the pixels whose score values in PC1 were lower than the histogram maximum. In order to define the best approach, regression models for T and W additives were built using the mean spectra of the rings prepared in triplicates, excluding the outliers. For these models, 12 mean spectra were used to calibrate and 5 mean spectra (obtained from different type 2 gasoline samples) were used as the external validation set. The relative errors of the test samples were calculated using Eq. (1) where c_i is the real concentration and $\hat{c}_i$ is the concentration predicted by the model. The mean relative error was calculated using Eq. (2) where n is the number of samples in the external validation set. As

Table 5
Comparison of the results obtained by using the two paper compensation approaches.

Approach	Additive	Pre-processing	LV	Mean percentage error (%)
(D)	T	2nd order Savitzky–Golay derivative (11 points window)[a]	3	14
	W	2nd order Savitzky–Golay derivative (11 points window)	1	23
(E)	T	1st order Savitzky–Golay derivative (7 points window)[b]	2	9
	W	1st order Savitzky–Golay derivative (11 points window)	2	19

[a]Variables from 1670 to 1801 nm and from 2238 to 2425 nm.
[b]Variables selected with the Jack-knife algorithm.

Table 6
Classification results obtained with LDA. The number of wavelengths employed in each model is indicated in parentheses. *N* indicates the number of samples employed in the test set.

True class	*N*	LDA/SPA[a] (60)				LDA/GA[b] (28)				LDA[c] (53[d])			
		Predicted class				Predicted class				Predicted class			
		G	T	W	Y	G	T	W	Y	G	T	W	Y
G	48	43	–	4	1	43	–	4	1	43	–	4	1
T	49	–	49	–	–	–	49	–	–	–	47	–	2
W	45	7	–	38	–	6	–	39	–	4	–	41	–
Y	50	–	–	–	50	–	–	–	50	–	1	–	49

[a] Savitzky–Golay smoothing (7 points window) and range normalization.
[b] Savitzky–Golay smoothing (3 points window) and range normalization with variables selected in the regions in which the additives have different absorption bands.
[c] Savitzky–Golay smoothing (3 points window) and range normalization.
[d] Variables from 1670 to 1801 nm and from 2238 to 2425 nm.

can be seen in Table 5, the errors are lower for the second approach (E). Therefore, method E was chosen.

$$e(\%) = \frac{(c_i - \hat{c}_i)}{c_i} \times 100\% \quad (1)$$

$$E(\%) = \frac{\Sigma \text{absolute}(e(\%))}{n} \quad (2)$$

In summary, the best conditions used in the image processing were: selection of the pixels of the ROI using the histograms of the score values of PCA models individually built for each ring image (approach C2), taking the representative spectra as the sum the spectra of ROI, and applying a compensation for the paper spectra with score values in PC1 lower than the histogram maximum (approach E).

3.3. Classification

Once the best approaches to select the ROI and the best representative spectra were chosen, classification models were built. The results of the best models built with the variables selected by SPA and GA and with all the 53 variables of the regions in which the additives have characteristic absorption bands (from 1,670 to 1,801 nm and from 2,238 to 2,425 nm) are shown in Table 6. The models were built with the spectra pre-processed with Savitzky–Golay smoothing and 1st derivative, varying the number of window points and the order of the derivative, and range normalization. As can be seen in Table 6, the best result was achieved with LDA/GA that correctly classified 92% of the test samples and 94.3% of all the samples. Most of the errors in this model (10 of 11) occurred due to misclassifications between additives G and W. These errors may indicate that the additives are chemically similar and that a single PLS regression model could be used to quantify them. The discriminant functions (DF) of the model (Fig. 8) show the separation of additives T and Y in the DF2. Regarding additives G and W there is a tendency towards separation, with dispersion and some overlapping of the two classes.

3.4. Regression

The results of all the best regression models built are shown in Table 7. For strengthening the robustness of the results, a number of training–validation–testing splits were performed. Five different models were built using 5 different calibration-test sets. In Table 7, the results reported are the average results of these five different combinations of calibration-testing samples. It is important to mention that in all the five test sets, different gasoline samples were employed (type 2 gasoline samples) to assure variability in the chemical matrix. In addition, as the classification model showed that several samples containing the G and W additives are very similar to each other, a single regression model

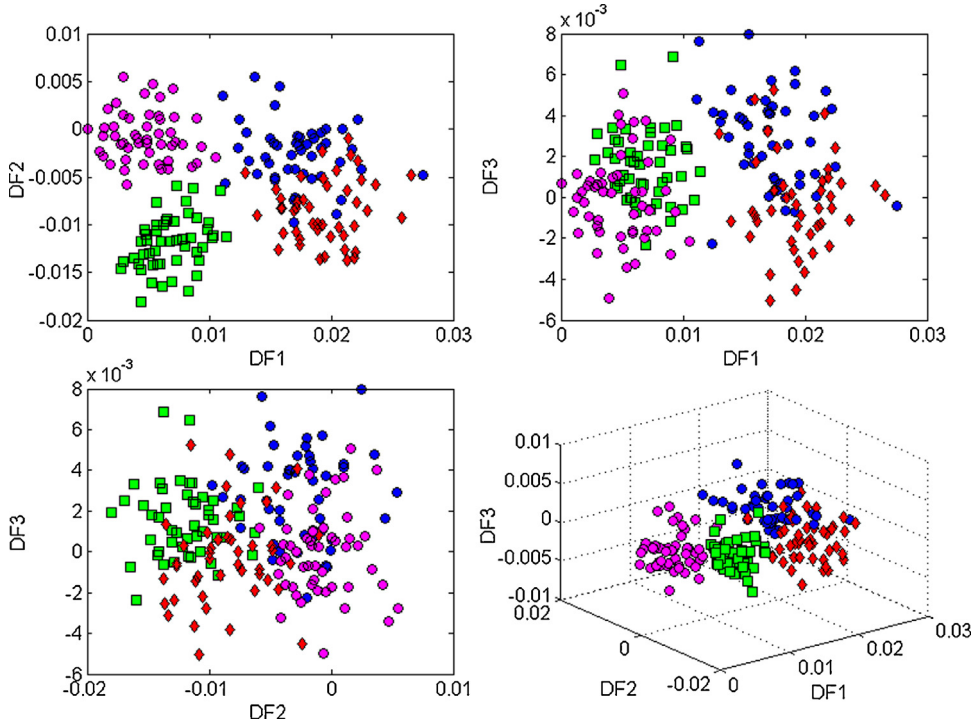


Fig. 8. Discriminant functions of the LDA/GA model. (●, Additive G; ■, Additive T; ◆, Additive W; ●, Additive Y).

Table 7
Mean regression results obtained with PLS models.

Additive	Pre-processing	LV	RMSECV (mg kg^{-1})	Mean relative error (%)
G	2nd order Savitzky–Golay derivative (11 points window)[a]	1	64	15
T	1st order Savitzky–Golay derivative (7 points window)[a]	2	34	10
W	1st order Savitzky–Golay derivative (11 points window)	2	43	14
Y	2nd order Savitzky–Golay derivative (11 points window)	2	34	5
G and W	2nd order Savitzky–Golay derivative (7 points window)	2	48	12

[a] Variables selected with the Jack-knife algorithm.

Table 8
Mean prediction results obtained for the test sets using different additive models.

Additive	Model	Mean relative error (%)
G	G	15
	T	26
	W	15
	Y	29
T	G	21
	T	10
	W	24
	Y	17
W	G	9
	T	30
	W	14
	Y	32
Y	G	28
	T	18
	W	45
	Y	5

was built to determine these two additives. As can be seen in Table 7, this model shows an intermediate value for the RMSECV and a lower value for the mean relative error, when compared to the individual models of these two additives. Therefore, this model can be used to quantify both additives. The mean relative errors varied from 5 to 15% which is adequate for this kind of application.

In order to verify the behavior of the mean relative error of prediction (Eq. (2)) if a sample of one type of additive is predicted using a model built for the other type, the test samples were quantified using the models constructed for other additives. The results are shown in Table 8. As in Table 7, the results represent also the average of five different models, built using 5 different calibration-test sets. As can be seen, if the concentration of G and W additives are predicted using W and G models, respectively, the errors are similar to the ones observed when the same samples are predicted using their own models. If these samples, however, were predicted using the T and Y models, the errors increase significantly. Therefore, G and W additives seem to be chemically similar, and even the wrong classification of gasoline containing these additives would not result in significant prediction errors of their content. Regarding the T and Y samples, they cannot be predicted using different models because the error increases significantly.

4. Conclusions

In this work, different approaches were used to select the pixels of the ROI, to extract representative spectra from ring-oven images, and to compensate for the paper spectrum (background) contribution. These different approaches were evaluated in order to identify the simpler, faster and more objective way to implement the technique on a routine basis in fuel analysis laboratories.

Regarding to the classification results, 10 of the 11 misclassification errors obtained by the LDA/GA model resulted in G and W additives being wrongly classified in each other's class. It was assumed, then, that these two additives are chemically similar and a single regression model could be used to quantify them. Regarding the regression results, the mean relative error of prediction varied from 5 to 15% and the models were built with 2 latent variables at most. The classification step is a necessary pre-stage in the analysis to identify correctly the true class of the samples before quantification, made by using the specific regression model constructed for the assigned class. However, due to the similarity previously noted between additives G and W, a single regression model constructed for both G and W additives can be used in the quantification stage.

The results obtained indicate that NIR hyperspectral imaging, associated with the ring-oven pre-concentration technique and multivariate techniques, is promising for quantifying additives in gasoline samples. The objective method developed to select the ROI and pre-treat the ring-oven images can be easily automated to facilitate the use of the proposed method in routine analysis. The same methodology proposed in this work can be adapted to other additives used worldwide.

As the ring oven system has a low cost (U$ 2,000 at most) it is possible to have several systems working in parallel, minimizing the disadvantage related to the relative long time necessary to obtain the rings (35 min). Because the rings once formed on the paper substrate are very stable, they can be stored for long time and transported over long distances (even by regular mail) to be measured in the place the HI facility is allocated. Therefore, the procedure described in this study is potentially applicable for routine analysis in fuel laboratories.

Acknowledgments

This work is a contribution of the National Institute for Advanced Analytical Science and Technology (INCTAA) processes FAPESP 2008/57808-1, CNPq 573894/2008-6, CNPq and FACEPE. The authors are also grateful to ANP for providing the additives and for supporting this work. The English version of this text has been revised by Sidney Pratt, Canadian, BA, MAT (The Johns Hopkins University), RSAdip (TEFLA, Cambridge University).

References

[1] Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP), Resolution No. 40 of 25.10.2013. www.anp.gov.br (accessed 27.08.14).
[2] Ministry of Agriculture, Livestock and Food Supply (MAPA), Ordinance No. 105 of 28.02.2013. www.anp.gov.br (accessed 27.08.14).
[3] E.A. Nikitina, V.E. Emel'yanov, I.F. Krylov, A.V. Fedorova, Detergent additives to automotive gasolines, Chem. Technol. Fuels Oils 42 (1) (2006) 30–34.
[4] Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP), Resolution No. 1 of 06.01.2014. www.anp.gov.br (accessed 27.08.14).
[5] J.-J. Lin, J.-J. Wu, S.-M. Shau, Y.-S. Ho, Thermal stability and combustion behaviors of poly(oxybutylene) amides, Polym. J. 34 (2002) 72–80.
[6] S. Herbstman, T.E. Hayden, T.E. Nalesnik, N. Benfaremo, Gasoline detergent additive. Patent US 5030,249.9 July (1991).
[7] P.R. Stevenson, D. Thetford, J.S. Vilardo, Quaternary ammonium salt of a Mannich compound. Patent US 7906,470 B2. 15 March (2011).
[8] O. Graupner, M. Mundt, A. Schutze, J.J.J. Louis, D.R. Kendall, N.P. Tait, Gasoline additives. Patent US 2005/0172545 A1. 11 Aug. (2005).

[9] S.J. Brauer, T.-C.T. Miin, Method for quantitatively determining detergent fuel additives in fuel samples. Patent CA 2132,806.25 March (1995).
[10] S. Colaiocco, M. Lattanzio, Determination of additives in gasoline by SEC coupled to a light-scattering detector–chemometrics application, J. High Resolut. Chromatogr. 18 (1995) 387–388.
[11] G. Federova, Analytical method for the detection and quantification of fuel additives. Patent US 2004/0214341 A1. 28 Oct. (2004).
[12] L. Guoqing,W. Lili, F. Minxing, Method for separating and measuring cleaning agent in gasoline. Patent CN 1,673,737 A. 28 Sept. (2005).
[13] M.P.F. da Silva, L.R. e Brito, F.A. Honorato, A.P.S. Paim, C. Pasquini, M.F. Pimentel, Classification of gasoline as with or without dispersant and detergent additives using infrared spectroscopy and multivariate classification, Fuel 116 (2014) 151–157.
[14] Standard test method for distillation of petroleum products at atmospheric pressure. ASTM D863-11b. Am. Soc. Test Mater. (2011).
[15] J. Cortez, C. Pasquini, Ring-oven based preconcentration technique for microanalysis: simultaneous determination of Na, Fe, and Cu in fuel ethanol by laser induced breakdown spectroscopy, Anal. Chem. 85 (2013) 1547–1554.
[16] H. Weisz, Microanalysis by the Ring-Oven Technique, second ed., Pergamon Press, New York, 1970.
[17] J.M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco, Study of pharmaceutical samples by NIR chemical-image and multivariate analysis, TrAC Trends Anal. Chem. 27 (2008) 696–713.
[18] M.A.F. Ossa, J.M. Amigo, C. García-Ruiz, Detection of residues from explosive manipulation by near infrared hyperspectral imaging: a promising forensic tool, Forensic Sci. Int. 242 (2014) 228–235.
[19] S. Piqueras, L. Duponchel, R. Tauler, A. de Juan, Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares, Anal. Chim. Acta 705 (2011) 182–192.
[20] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images: essential steps before image analysis, Chemom. Intell. Lab. Syst. 117 (2012) 138–148.
[21] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemom. Intell. Lab. Syst. 76 (2005) 101–110.
[22] J.M. Prats-Montalbán, A. de Juan, A. Ferrer, Multivariate image analysis: a review with applications, Chemom. Intell. Lab. Syst. 107 (2011) 1–23.
[23] E.D.T. Moreira, M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, Near infrared spectrometry classification of cigarettes using the successive projections algorithm for variable selection, Talanta 79 (2009) 1260–1264.
[24] M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, P.N.T. Moreira, O.D. Pessoa Neto, G.E. José, T.C.B. Saldanha, The successive projections algorithm for spectral variable selection in classification problems, Chemom. Intell. Lab. Syst. 78 (2005) 11.
[25] L.H. Chiang, R.J. Pell, Genetic algorithms combined with discriminant analysis for key variable selection, J. Process Control 14 (2004) 143–155.
[26] J. Workman Jr, L. Weyer, Practical Guide and Spectral Atlas for Interpretative Near-Infrared Spectroscopy, second ed., CRC Press, New York, 2012.