Fuel 116 (2014) 151–157

Contents lists available at ScienceDirect

Fuel

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

Classification of gasoline as with or without dispersant and detergent additives using infrared spectroscopy and multivariate classification

Michelle Patrícia Ferreira da Silva [a], Lívia Rodrigues e Brito [a], Fernanda Araújo Honorato [b], Ana Paula Silveira Paim [a], Celio Pasquini [c], Maria Fernanda Pimentel [b,*]

[a] Universidade Federal de Pernambuco, Departamento de Química Fundamental, Recife, PE, Brazil
[b] Universidade Federal de Pernambuco, Departamento de Engenharia Química, Recife, PE, Brazil
[c] Universidade Estadual de Campinas, Instituto de Química, Campinas, SP, Brazil

HIGHLIGHTS

- Method based on IR spectra and classification methods to detect additives in gasoline.
- Three classification methods (LDA, PLS-DA, SVM) were compared.
- Three algorithms for selection of variables for LDA were evaluated – SW, GA, SPA.
- The best results were obtained for the MIR region using the LDA/SPA and LDA/GA.
- 100% correct classification was achieved for the test set.

ARTICLE INFO

Article history:
Received 17 April 2013
Received in revised form 25 July 2013
Accepted 26 July 2013
Available online 13 August 2013

Keywords:
Gasoline
Detergents and dispersants additives
Infrared spectroscopy
Classification

ABSTRACT

Gasoline may contain additives which can minimize the amount of pollutants emitted to the atmosphere. Detergents and dispersants added to gasoline can reduce gas emissions towards atmosphere and the formation of deposits in engines. The Brazilian Agency of Petroleum, Natural Gas and Biofuel (ANP) has established that Brazilian commercial gasoline must contain detergent and dispersant additives, thus requiring the development of methods for their identification in commercial gasoline. This work proposes a methodology which uses infrared spectra in the medium and near region (MIR and NIR) of the residue of distillation for classification of gasoline samples into two groups: with or without detergent/dispersant additives. The performances of three types of classification methods were compared: linear discriminant analysis (LDA), partial least squares discriminant analysis (PLS-DA) and Support Vector Machines (SVM). Different algorithms for selection of spectral variables for LDA were evaluated: stepwise (SW), genetic algorithm (GA) and successive projections algorithm (SPA). The best results were obtained using LDA/GA or SPA/LDA for MIR region.

1. Introduction

Gasoline is a petroleum-derived fuel which is a complex mixture of several compounds, the majority of which are classified within the group of hydrocarbons. In Brazil, the gasoline produced at refineries or petrochemical plants is called “Type A gasoline”. However, to be marketed, Brazilian laws establish that anhydrous ethanol fuel must be added to obtain “Type C gasoline” or “Common gasoline” [1]. The ethanol content in gasoline differs in various countries, in general varying from 5% to 10% (v/v) [2]. In Brazil, this content may vary from 18% (v/v) to 25% (v/v), depending on the availability and price of the ethanol [3].

Besides anhydrous ethanol, gasoline may contain additives, which are compounds added to improve fuel performance, avoid motor problems and minimize the amount of pollutants emitted to the atmosphere. There are several types of additives for gasoline, such as detergents, dispersants, antiknock agents and antioxidants. Dispersant additives prevent fuel stratification during storage at low temperatures [4]. Detergent additives reduce gas emissions to the atmosphere as well the formation of deposits in engines [4]. The chemical composition of detergent and dispersant additives is not exactly known due to the fact that it is kept secret by the manufacturing industries. However, it is well-known that these additives must exhibit good thermal stability up to 300 °C, which is the observed highest temperature of an internal combustion engine [5]. Although the literature presents several works reporting

---

* Corresponding author. Address: Universidade Federal de Pernambuco, Departamento de Engenharia Química, Laboratório de Combustíveis (LAC), CEP 50740-521 Recife, PE, Brazil. Tel./fax: +55 81 2126 7235.
E-mail addresses: mfernanda.pimentel@gmail.com, mfp@ufpe.br (M.F. Pimentel).

0016-2361/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fuel.2013.07.110

on the benefits of oxygenated additives and on they determination, no work could be found concerning detergent and dispersant additives [6–10].

Quality control of fuels has become increasingly rigorous, not only because of economic and social issues, but also due environmental concerning. The Brazilian Agency of Petroleum, Natural Gas and Biofuel (ANP) has established that after January 1st of 2014 all Brazilian commercial gasoline should contain detergent and dispersant additives [11]. The concentration depends on the formulation of the additives and its effectiveness. Currently, additives may be mixed with Type C gasoline at concentrations ranging from 100 to 5000 mg kg^{-1} and the gasoline is labeled as “Additive Type C gasoline”. The compulsory use of detergent and dispersant requires the development of analytical methods capable of identifying the presence of these additives in gasoline samples, in order to verify whether commercial gasoline accomplishes with the norms established by ANP [11].

The literature presents many examples of application of near infrared (NIR) and middle infrared (MIR) spectroscopy associated with multivariate analysis for quality control of fuels [12–16]. Infrared (IR) spectroscopic methods show several advantages, such as the use of relatively low cost equipment that allows field analysis; minimum or no sample treatment; expeditious analysis; causing no sample destruction, and demanding no reagents.

Perhaps, the main drawback of IR based analytical methods is in its relatively low detectability. As dispersant and detergent additives are typically added to Brazilian gasoline in concentrations ranging from 200 to 500 mg kg^{-1} it is not possible to determine these species directly in gasoline using IR spectroscopy [17]. To concentrate the additives up to a concentration level capable to be detected by IR spectroscopy, gasoline samples have been submitted to distillation, following the ASTM D86-11b standard procedure [18], before analysis. As dispersant and detergent additives presents low volatility, they are concentrated in the distillation final residue by a factor of about 100 or more, allowing the use of IR spectroscopy, associated to supervised pattern recognition methods, to identify their presence in gasoline samples

Pattern recognition methods such as principal components analysis (PCA), linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA) are extensively used in classification problems [19]. PCA is an unsupervised pattern recognition method used to detect similarities among samples [19]. Different from this, LDA and PLS-DA are supervised pattern recognition methods and require a training set of known groups [18,19]. These last two methods are similar because the samples may only be classified in one of the classes.

PLS-DA is based on the PLS algorithm for calibration, but instead of using concentration information, it uses class labels as dependent y vector [12,19]. In PLS-DA, for example, a value close to one indicates that the sample belongs to the class under consideration and a value close to zero indicates that it does not. The threshold used in PLS-DA may vary, but for a two class problem it is common to use a threshold of 0.5.

LDA employs the Mahalanobis distance and seeks a linear combination function that maximizes the between-class variance relative to the within-class variance [13,20]. In LDA the number of training samples must be larger than the number of variables to be included in the LDA model [12]. When using spectral data, in which there are a great number of variables, LDA requires the use of variable selection algorithms, such as: stepwise algorithm (SW), successive projections algorithm (SPA) and genetic algorithm (GA).

The SW algorithm evaluates the individual value of each spectral variable according to its discriminability with respect to the classes under consideration [21]. The variable with the largest discriminability is selected and a leave-one-out cross-validation procedure is performed using LDA [21]. To avoid collinearity problems, the remaining variables that are highly correlated with those already selected are discarded [21]. For the decision about which variables need to be discarded, a threshold value for the coefficient of multiple correlation must be defined by the operator. The algorithm repeats this procedure until no more variables are available for selection. The set of variables that resulted in the smallest number of cross-validation errors is then adopted.

The goal of SPA is to select variables with minimum multicollinearity and maximum information [13]. In SPA the validation set is used to choose the best subset of variables by minimizing the cost function (Eq. (1)) defined as the average risk of misclassification by LDA [22].

$$G = \frac{1}{kV}\sum_{k=1}^{kV} g_k \tag{1}$$

Where g_k (risk of misclassification of the kth validation object x_k) is defined as

$$g_k = \frac{r^2(x_k, \mu_{1k})}{\min_{1j \neq 1k} r^2(x_k, \mu_{1j})} \tag{2}$$

The numerator in Eq. (2) is the squared Mahalanobis distance between object x_k (of class index Ik) and the sample mean μ_{Ik} of its true class (both row vectors). The denominator in Eq. (2) is the squared Mahalanobis distance between object x_k and the center of the closest incorrect class. The object x_k should be close to the center of its true class and distant from the centers of all other classes [22].

The GA is a stochastic algorithm invented to imitate the evolutionary process of living species [23]. Given two distinct classes, the algorithm can be used as an optimization procedure to determine the key variables that maximize the separation between the two classes [23]. As in SPA the GA uses a function to evaluate the quality of the variable selection. In this case, this function is called the fitness function, which is defined as the inverse of the cost function (Eq. (1)) [12]. For more details on GA see Refs. [23,24].

Support Vector Machine (SVM) is a machine learning technique derived from two foundations: Statistical Learning Theory and Mathematical Optimization, applied for classification, regression, and other learning tasks showing high performance in practical applications [25–28]. The support vector machine determines the limits of a decision, producing a great separation between classes by minimizing the errors. For this, SVM implements two basic mathematical operations: non-linear mapping of input vectors in a high dimensional feature space (kernels), and constructing a maximum margin hyperplane in the feature space. The construction of this hyperplane is performed in accordance with the principle of structural risk minimization (SRM). For more details on SVM see Refs. [19,25,26].

The present paper proposes an analytical method to classify gasoline as with or without dispersant and detergent additives based on infrared spectroscopy and supervised pattern recognition methods. The spectra of the distillation residues, which contain the pre-concentrated additives, were employed to overcome the lack of detectability of the IR spectroscopy.

2. Materials and methods

2.1. Samples

A total of 125 samples of gasoline type C were acquired in fuel stations from the states of Pernambuco, Alagoas and Sergipe in the Northeast of Brazil. So as to encompass the variability of the gasoline composition, the samples were collected from different producers and distributors. From these 125 samples, 49 were commercialized containing additive and 64 without additives. In addition, 12 samples of gasoline (acquired without additives) have been spiked with different additives in the laboratory. In Brazil, five types of dispersant and detergent additives are used. However, only three were provided by ANP for this study. These three types of additives (coded as *G*, *T* and *W*), in concentrations of 100, 200, 500 and 1000 mg kg^{-1}, were employed to spike the samples. It is important to mention that the 49 samples commercialized with additives should contain the five types, since they were collected from different producers and distributors, as commented above.

According to the gasoline specification established by the ANP Resolution N.57 [1], all samples were submitted to distillation analysis, following the ASTM D86-11b method [18]. The complete distillation procedure takes about 40 min and leaves about 1 mL as the final distillation residue, which is about 1% (v/v) of the starting sample volume of 100 mL.

2.2. NIR and MIR spectra acquisition

The spectra of gasoline distillation residues were acquired using a Spectrum 400 FT-IR/FT-NIR spectrometer (Perkin Elmer), at controlled temperature (22 ± 2) °C. Each spectrum was obtained with a resolution at 4 cm^{-1} (average of 16 scans). The NIR spectra (12,000–4000 cm^{-1}) were acquired using a quartz cell with an optical path length of 1.0 mm. For MIR measurements (4000–650 cm^{-1}) a ZnSe horizontal attenuated total reflectance (HATR) accessory was employed.

2.3. Data analysis and software

NIR and MIR spectral data sets were submitted to different pre-processing techniques, such as Standard Normal Variate (SNV), Savitzky–Golay first derivatives (7, 15 and 21 point windows and second order polynomial) and range normalization.

Outlier detection was performed using PCA scores, X-residual and leverage plots. Only four samples (two gasoline residue samples without additives and two gasoline residue samples with additives) were identified as outliers and excluded from the NIR data set. Only six samples (three gasoline residue samples without additives and three gasoline residue samples with additives) were excluded from the MIR data set before building the classification models.

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 [29]. The number of samples in each class is presented in Table 1. The training and validation samples were used in the modeling procedures (including variable selection for LDA and determination of the number of factors in PLS-DA), whereas the test samples were only used in the final evaluation and comparison of predictive ability of the models.

In SW algorithm, seven threshold values (0.10, 0.30, 0.50, 0, 70, 0.80, 0.90 and 0.95) of multiple correlation coefficients were tested. The best threshold value was selected on the basis of the classification errors recorded for the validation set.

Table 1
Number of training, validation and test samples in each class.

Class	NIR data			MIR data		
	Training	Validation	Test	Training	Validation	Test
Without additives	38	12	12	37	12	12
With additives	35	12	12	34	12	12
Total	73	24	24	71	24	24

The GA algorithm was carried out over 100 generations with 200 chromosomes in each generation and the mutation and crossover probabilities were set to 10% and 60%, respectively. Moreover, the algorithm was repeated five times, starting from different random initial populations. The best solution resulting from the five runs was employed.

In PLS-DA, five threshold values (0.25, 0.375, 0.5, 0.625 and 0.75) were evaluated [19]. When a value above the selected threshold is predicted, a sample is considered as belonging to the class “gasolines with additives” while a value below the selected threshold indicates that the sample belongs to the class “gasolines without additives”. The best threshold value was selected on the basis of the classification errors recorded for the test set. For PLS-DA, the leave-one-out method of cross-validation was employed to select the number of factors. The data used to build the models included the training and validation samples.

For SVM classification a linear kernel function was used. This function was chosen based on linear trend confirmed by linear discriminant analysis (classification rate on the test set of 94%). Optimization of the *C* parameter (cost parameter) was done through cross-validation procedures to evaluate model performance and minimize the risk of overfitting [26]. The model is calculated with a training set, and an independent test set is used to test the predictive performance of the trained model.

Spectral pre-processing, PCA, PLS-DA and SVM were carried out using Unscrambler® X(CAMO S.A.) [30]. The KS, LDA-GA, LDA-SW and LDA-SPA algorithms were coded in Matlab® R2010a 7.10.0.499 [31].

3. Results and discussion

3.1. NIR and MIR spectra

Fig. 1a presents the original mean NIR spectra of the two classes of gasoline residue recorded in the range of 12,000–4000 cm^{-1}. These NIR spectra show bands assigned to the second overtones (8500–8200 cm^{-1}), first overtones (6000–5500 cm^{-1}) and the combination regions (4600–4000 cm^{-1}) of C—H bonds [32].

Fig. 1b shows the original mean MIR spectra of the two classes of gasoline residue recorded in the range of 4000–650 cm^{-1}. It is possible to observe peaks corresponding to C—H stretching of alkanes and alkenes (2800–3100 cm^{-1}), C—H stretching of aromatic rings (1475–1600 cm^{-1}), C—H bend of alkanes (1300–1500 cm^{-1}) and C—H out-of-plane bend of alkenes and aromatic rings (650–1000 cm^{-1}) [33].

Fig. 1 shows that it is very difficult to distinguish between the two types of gasoline based only on their NIR or MIR spectra of distillation residues, thus demonstrating the need for using chemometric techniques for classification.

3.2. LDA and PLS-DA classification

Principal component analysis (PCA) was performed as an exploratory tool and for outlier detection. Due to the fact that the additives are minor constituents and the gasoline composition varies widely according to origin and manufacturing process, it was already expected that the first components would not show a separation related to the presence of the additives, but associated with the variation of the major components of the gasoline. Hence, PCA results did not show any clear separation of classes, but only a tendency of separation, with substantial dispersion and overlapping of classes, observed for higher order PCs.

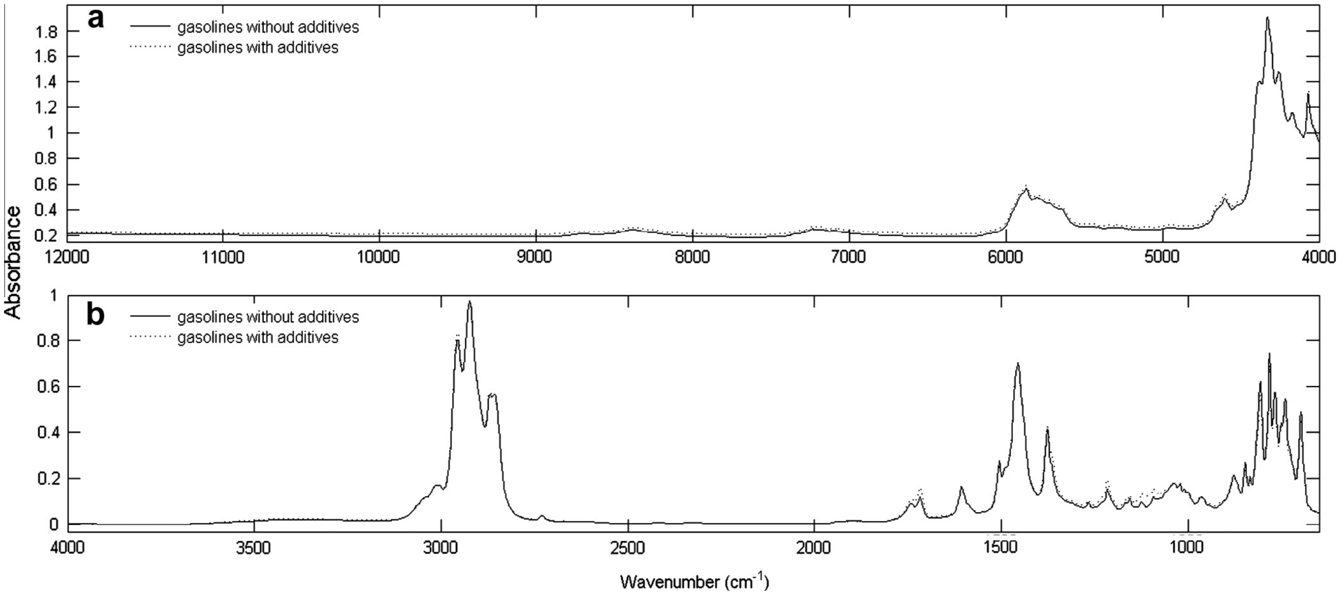

Fig. 1. Original mean spectra (gasoline residue) of the two gasoline classes recorded in the region of (a) NIR and (b) MIR.

The LDA models obtained with the variables selected by SW, SPA and GA algorithms and the PLS-DA models were applied to the classification of the test set (external validation set). The best results for the models in the NIR region were achieved by using different preprocessing techniques. For the LDA/SPA, LDA/GA and PLS-DA, SNV was the best preprocessing. For LDA/SW, the best result was obtained using a combination of Savitzky–Golay first derivative (21 point windows and second order polynomial) followed by range normalization.

The SW algorithm selected wavenumbers in the region between 6000 cm^{-1} and 5500 cm^{-1}, which may be associated with the first overtone of C—H stretching [31]. Most of the wavenumbers selected by SPA algorithm are located in the region between 4600 cm^{-1} and 4000 cm^{-1}, which can be attributed to combination bands of C—H bond [31]. The GA algorithm selected wavenumbers in the region close to the second (8500–8200 cm^{-1}) and first overtones (6000–5500 cm^{-1}) of C—H stretching [31].

The best results for the NIR models are presented in Table 2. The values express both correct classifications (predicted class index equal to correct class index) and incorrect classifications (predicted class index different from correct class index). For PLS-DA, it is important to mention that, as leave-one-out cross, validation was employed, the results shown for the training set were those obtained in the calibration and the results shown for the validation set correspond to those obtained in the cross-validation procedure.

Overall, the best results obtained using the NIR region were achieved with GA and SPA, which showed similar performance for the test samples.

The best classification results for the models based on the MIR region were also achieved by using different preprocessing techniques, depending on the algorithm employed. For the LDA/SW models, the best preprocessing was SNV. For the LDA/SPA models, the best results were achieved using the Savitzky–Golay first derivative (21 point windows and second order polynomial). For the LDA/GA models, the combination of Savitzky–Golay first derivative (15 point windows and second degree polynomial) followed by range normalization presented the best results. Finally, for the PLS-DA, the best pre-processing was a combination of Savitzky–Golay first derivative (21 point windows and second degree polynomial) with range normalization.

Most of the wavenumbers selected by the SW and SPA algorithms are located in the region between 1000 cm^{-1} and 650 cm^{-1} (C—H out-of-plane bend of alkenes and aromatic rings) [32]. The GA algorithm also selected wavenumbers in this region.

Table 2
Classification results obtained with LDA/SW, LDA/SPA, LDA/GA and PLS-DA models in the NIR region. (1): Gasoline without additives and (2) Gasoline with additives. The number of wavenumbers or latent variables employed in each model is indicated in parentheses. *N* indicates the number of samples employed in the study.

Training set		LDA/SW (16)[a] Predicted index		LDA/SPA (25) Predicted index		LDA/GA (23) Predicted index			PLS-DA (8)[b] Predicted index	
True class index	*N*	1	2	1	2	1	2	*N*	1	2
NIR data										
1	38	34	4	36	2	37	1	50	48	2
2	35	8	27	1	34	1	34	47	7	40
Validation set		LDA/SW (23)[a]		LDA/SPA (25)		LDA/GA (23)			PLS-DA (8)[b]	
1	12	12	0	12	0	11	1	50	46	4
2	12	3	9	2	10	1	11	47	10	37
Test set										
1	12	11	1	12	0	12	0	12	12	0
2	12	0	12	1	11	1	11	12	2	10

[a] Threshold 0.7.
[b] Threshold 0.5.

The three algorithms also selected absorption peaks between 3100 cm^{-1} and 2800 cm^{-1} (C—H stretching of alkanes and alkenes) and between 1600 cm^{-1} and 1300 cm^{-1} (C—Hbend of alkanes and C—H stretching of aromatic ring) [33]. The best results for the MIR models are presented in Table 3.

The best results in the MIR region were achieved with LDA/SPA and LDA/GA which correctly classified 100% of the test samples. These result are better than those obtained for the NIR region, which showed at most 92% of correct classification.

Superimposing the MIR spectra of the gasoline residue samples containing and not containing the additives *G*, *W* and *T* (1000 mg kg^{-1}), it can be observed that both algorithms SPA and GA have selected variables corresponding to bands characteristic of the additives. To illustrate, Figs. 2 and 3 show some variables selected by GA and SPA.

Fig. 2a shows that the variable selected by GA at 870 cm^{-1} can be associated to spectral differences between the residue of a gasoline sample with additive "W" (1000 mg kg^{-1}) from the other gasoline residues (with *G* and *T* additives and without additives). Fig. 2b shows the variable selected by GA at 720 cm^{-1}, which distinguishes spectral differences between the residue of gasoline samples with additives *G*, *T* and *W* from the gasoline residue without additives.

The variables selected by SPA at 1733 cm^{-1}, 1725 cm^{-1} and 1718 cm^{-1} are shown in Fig. 3a. As can be seen, these variables evidence spectral differences between the residue of the gasoline sample without additives from those samples with additives *G*, *W* and *T* (1000 mg kg^{-1}). Fig. 3b shows segments of MIR spectra containing the selected variables at 1097 cm^{-1}, 1060 cm^{-1} and 1050 cm^{-1}, as selected by SPA, which suggest spectral differences between the residue of a gasoline sample with additive "W" (1000 mg kg^{-1}) from the others gasoline residues samples (with *G* and *T* additives and without additives).

On the other hand, several variables selected by the different algorithms could not be directly associated to spectral features of the additives *G*, *W* and *T*. This fact can be explained considering that the sample set employed for variables selection contains commercial gasoline samples with additives, besides the surrogate samples prepared to contain *G*, *W* and *T* additives. For these commercial samples it was not possible to know the type of additive and, of course, its spectral features. In addition, sometimes an algorithm may select variables that do not add to the classification power of the model, because it is not associated with any spectral feature of the analyte. At the same time, those variables do not jeopardize the classification power.

Table 3
Classification results obtained with LDA/SW, LDA/SPA, LDA/GA and PLS-DA models in the MIR region. (1) Gasoline without additives and (2) Gasoline with additives. The number of wavenumbers employed in each model is indicated in parentheses. *N* indicates the number of samples employed in the study.

Training set		LDA/SW (20)[a] Predicted index		LDA/SPA (26) Predicted index		LDA/GA (27) Predicted index			PLS-DA (6)[b] Predicted index	
True class index	*N*	1	2	1	2	1	2	*N*	1	2
MIR data										
1	37	37	0	35	2	36	1	49	43	6
2	34	2	32	2	32	0	34	46	8	38
Validation set										
1	12	11	1	9	3	12	0	49	43	6
2	12	0	12	0	12	1	11	46	11	35
Test set										
1	12	10	2	12	0	12	0	12	12	0
2	12	1	11	0	12	0	12	12	1	11

[a] Threshold 0.95.
[b] Threshold 0.5.

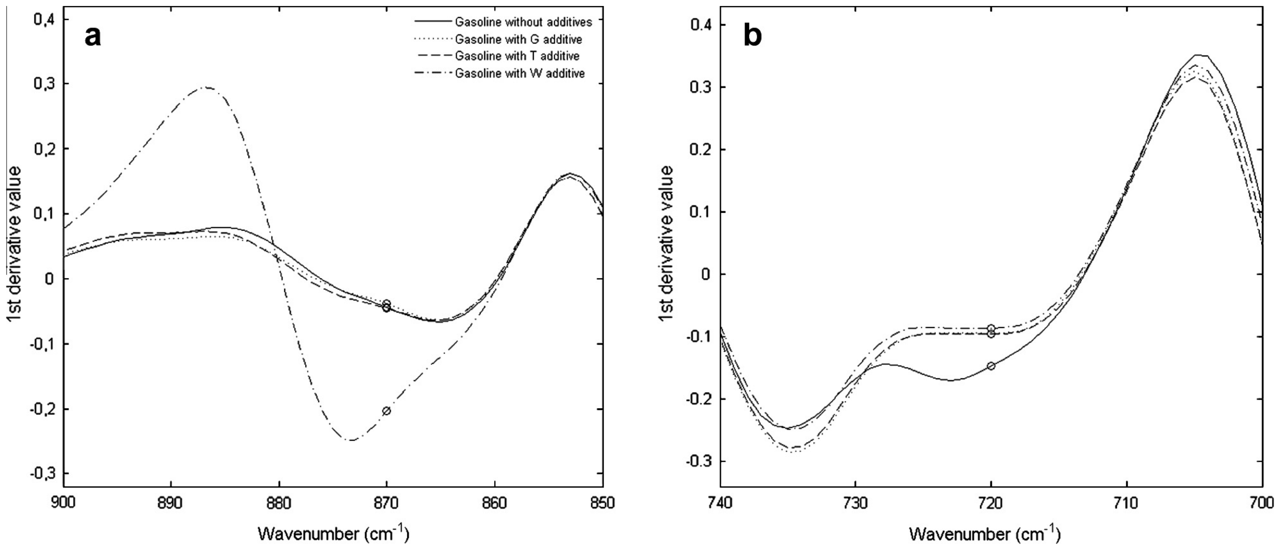

Fig. 2. Derivative MIR spectra segment of the gasoline residues (with and without additives), emphasizing the variables selected by GA (a) at 870 cm^{-1} and (b) at 720 cm^{-1}.

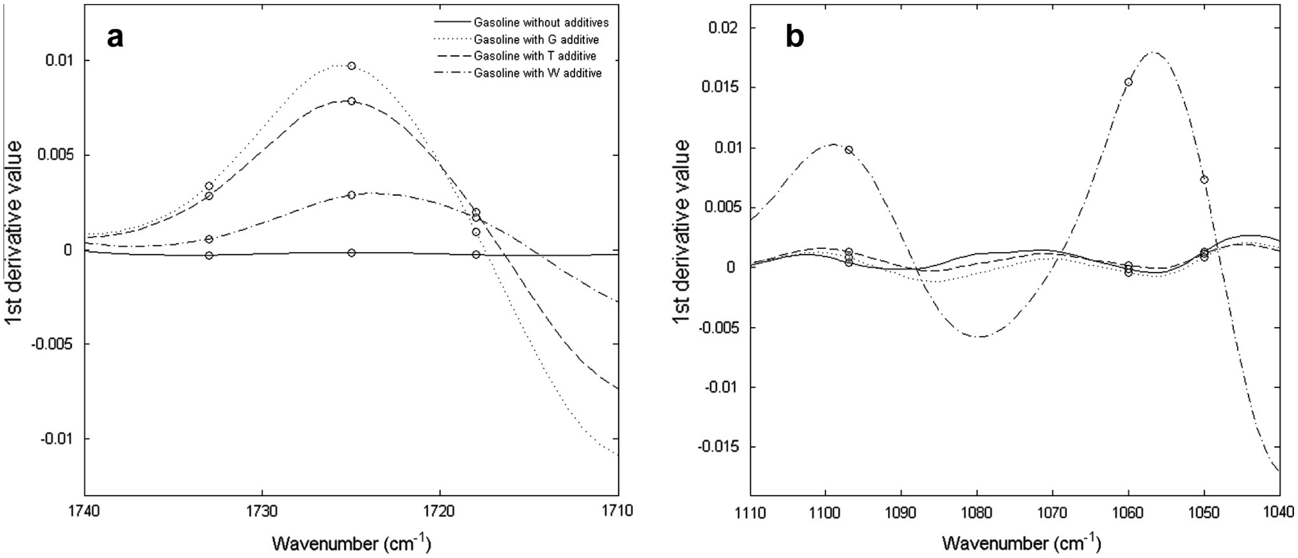

Fig. 3. Derivative MIR spectra segment of gasoline residues (with and without additives), emphasizing the variables selected by SPA (a) at 1733 cm^{-1}, 1725 cm^{-1} and 1718 cm^{-1} and (b) at 1097 cm^{-1}, 1060 cm^{-1} and 1050 cm^{-1}.

3.3. SVM classification

The linear kernel function was adopted and optimization of the *C* parameter used threshold values of 0.1, 1.0, 10, 100, 1000 in the raw and pre-processed spectra in both regions, NIR and MIR.

It was observed that for higher *C* values (100 and 1000) the number of support vectors decreases and the rate of misclassified samples increases. According Devos et al. [26], high values of *C* minimize the margin, leading to more complex boundary limits; and smaller values of *C* (1 and 10) lead to large margins, smooth boundaries and the models are more parsimonious and easier to interpret.

SVM models, constructed with NIR spectra, presented the best result when the full spectrum, without pre-processing, was employed, optimized with a *C* value of 10, and a total of 52 support vectors on the margin border were employed. The SVM model correctly classified 11 of the 12 test samples. Considering the MIR spectral region, the best SVM model, produced a correct classification for 11 of the 12 test samples using the raw spectra (*C* value of 10 and 33 support vectors of the margin borders).

4. Conclusions

In this study, different spectral regions (NIR and MIR), preprocessing techniques (SNV, Savitzky–Golay first derivatives with 7, 15 and 21 point windows and second degree polynomial and range normalization), supervised pattern recognition methods (PLS-DA, SVM and LDA) associated with variable selection algorithms (GA, SW and SPA) were used to distinguish pre-concentrated gasoline (residue obtained after distillation) containing additives or not. The classification techniques PLS-DA, LDA and SVM produced comparable results. However, 100% correct classification for the external validation set was obtained using LDA with variables selected by SPA and GA, for the MIR region. The results obtained indicate that infrared spectroscopy is a promising technique for classification of gasoline with and without additives when associated with supervised pattern recognition methods and a pre-concentration step (distillation). Moreover, because distillation is a customary assay in the quality control of gasoline, the classification procedure described is potentially applicable in the daily routine of fuel analysis 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, FACEPE, FINEP, CAPES and CNPq. The authors are also grateful to ANP for providing the additives and for supporting this work.

References

[1] Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP), Resolution No. 57 of 20.10.2011. <www.anp.gov.br> [accessed 17.07.13].
[2] Opekar F, Cabala R, Kadlecová T. A simple contactless impedance probe for determination of ethanol in gasoline. Anal Chim Acta 2011;694:57–60.
[3] Presidency of the Brazilian Republic. Law No. 12,490 of 16.09.2011. <www.planalto.gov.br/> [accessed 17.07.13].
[4] Danilov AM. Development and use of fuel additives during 2006–2010. Chem Tech Fuels Oil+ 2012;47:470–84.
[5] Alemán-Vázquez LO, Vellagómez-Ibarra JR. Polyisobutenylsuccinimides as detergents and dispersants in fuel: infrared spectroscopy application. Fuel 2001;80:965–8.
[6] Dabbagh HA, Ghobadi F, Ehsani MR, Moradmand M. The influence of ester additives on the properties of gasoline. Fuel 2013;104:216–23.
[7] Song CL, Zhang WM, Pei YQ, Fan GL, Chu GP. Comparative effects of MTBE and ethanol additions into gasoline on exhaust emissions. Atmos Environ 2006;40:1957–70.
[8] Kaminski M, Kartanowicz R, Przyjazny A. Application of high-performance liquid chromatography with ultraviolet diode array detection and refractive index detection to the determination of class composition and to the analysis of gasoline. J Chromatogr A 2004;1029:77–85.
[9] Pavón JLP, Sánchez MN, Pinto CG, Laespada MEF, Cordero BM. Determination of methyl tert-butyl ether in gasoline: a comparison of three fast methods based on mass spectrometry. J Chromatogr A 2004;1048:133–9.
[10] Apicella B, Palma TM, Wang X, Vellota R, Armenate M, Spinelli N. A mass spectrometric study of gasoline anti-knocking additives: Part I: Electron impact ionization and fragmentation processes of MTBE. Int J Mass Spectrom 2007;262:105–13.
[11] Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP), Resolution No. 38 of 09.12.2009. <www.anp.gov.br> [accessed 17.07.13].
[12] Silva AC, Pontes LFBL, Pimentel MF, Pontes MJC. Detection of adulteration in hydrated ethyl alcohol fuel using infrared spectroscopy and supervised pattern recognition methods. Talanta 2012;93:129–34.
[13] Pontes MJC, Pereira CF, Pimentel MF, Vasconcelos FVC, Silva AGB. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. Talanta 2011;85:2159–65.
[14] Honorato FA, Neto BB, Pimentel MF, Stragevitch L, Galvão RKH. Using principal component analysis to find the best calibration settings for simultaneous spectroscopic determination of several gasoline properties. Fuel 2008;87:3706–9.
[15] Kardamakis AA, Pasadakis N. Autoregressive modeling of near-IR spectra and MLR to predict RON values of gasolines. Fuel 2010;89:158–61.

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

[17] Medeiros ARB. Uso de ATR/FTIR and FTNIR associado a técnicas quimiométricas para quantificação de aditivos em gasolina automotiva. Disssertação de Mestrado. Brasília, DF: Universidade de Brasília, Instituto de Química; 2009.

[18] Standard test method for distillation of petroleum products at atmospheric pressure. ASTM D86-11b. Am Soc Test Mater 2011.

[19] Brereton RG. Chemometrics for pattern recognition. 1st ed. Chichester: Wiley; 2009.

[20] Moreira EDT, Pontes MJC, Galvão RKH, Araújo MCU. Near infrared spectrometry classification of cigarettes using the successive projections algorithm for variable selection. Talanta 2009;79:1260–4.

[21] Caneca AR, Pimentel MF, Galvão RKH, Matta CE, Carvalho FR, Raimundo Jr IM, et al. Assessment of infrared spectroscopy and multivariate techniques for monitoring the service condition of diesel-engine lubricating oils. Talanta 2006;70:344–52.

[22] Pontes MJC, Galvão RKH, Araújo MCU, Moreira PNT, Pessoa Neto OD, José GE, et al. The successive projections algorithm for spectral variable selection in classification problems. Chemometr Intell Lab 2005;78:11–8.

[23] Chiang LH, Pell RJ. Genetic algorithms combined with discriminant analysis for key variable selection. J Process Contr 2004;14:143–55.

[24] Leardi R. Genetic algorithms in chemometrics and chemistry: a review. J Chemometr 2001;15:559–69.

[25] Li H, Liang Y, Xu Q. Support vector machines and its applications in chemistry. Chemometr Intell Lab 2009;95:188–98.

[26] Devos O, Ruckeebusch C, Durand A, Duponchel L, Huvenne JP. Support vector machines (SVM) in near infrared (NIR) spectroscopy: focus on parameters optimization and model interpretation. Chemometr Intell Lab 2009;96:27–33.

[27] Balabin RM, Safieva RZ, Lomakina EI. Near-infrared (NIR) spectroscopy for motor oil classification: from discriminant analysis to support vector machines. Microchem J 2011;98:121–8.

[28] Yu HY, Niu XY, Lin HJ, Ying YB, Li BB, Pan XX. A feasibility study on on-line determination of rice wine composition by Vis–NIR spectroscopy and least-squares support vector machines. Food Chem 2009;113:291–6.

[29] Kennard RW, Stone LA. Computer aided design of experiments. Technometrics 1969;11:137–48.

[30] The Unscrambler® X.1. CAMO Software AS. NedreVollgate 8, N-0158, OSLO, Norway.

[31] Matlab® R2010a 7.10.0.499. The Mathworks. Natick, MA, USA.

[32] Workman Jr J, Weyer L. Practical guide to near-infrared spectroscopy. 1st Ed. New York: CRC Press; 2007.

[33] Pavia D, Lampman GM, Kriz GS, Vyvyan JR. Introduction to spectroscopy. 4th ed. California: Brooks/Cole; 2008.