Full Length Article

Rapid and sensitive method for detecting adulterants in gasoline using ultra-fast gas chromatography and Partial Least Square Discriminant Analysis

Maurílio Gustavo Nespeca[a,*], João Fernando Villarrubia Lopes Munhoz[a], Danilo Luiz Flumignan[b], José Eduardo de Oliveira[a]

[a] Center for Monitoring and Research of the Quality of Fuels, Biofuels, Crude Oil, and Derivatives (Cempeqc), Institute of Chemistry, São Paulo State University (UNESP), Prof. Francisco Degni 55, Zip Code 14800-060 Araraquara, SP, Brazil
[b] Federal Institute of Education, Science and Technology of São Paulo (IFSP), Estéfano D'avassi 625, Zip Code 1.991-502 Matão, SP, Brazil

GRAPHICAL ABSTRACT
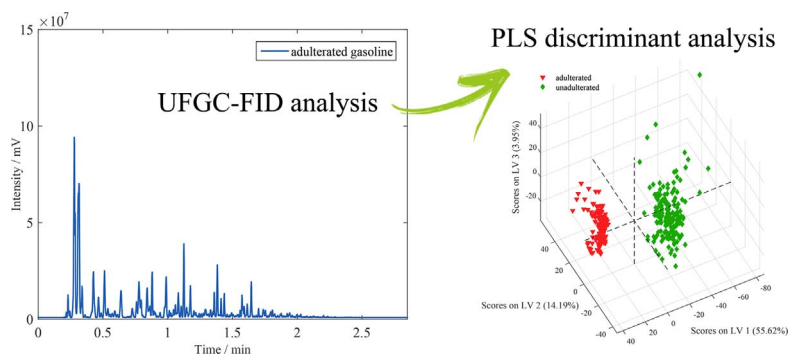
ARTICLE INFO

ABSTRACT

In the last years, the Brazilian Fuel Quality Monitoring Program drastically reduced the number of analyzed fuel samples as consequence of the current economic crisis in the country. The impoverishment of the monitoring program may lead to an increase in cases of gasoline adulteration, nonetheless, it also strengthens the search for faster and less costly methodologies for the fuel quality monitoring. Thus, this study aimed the development of a rapid analytical method to detect the adulteration of gasoline with organic solvents through ultra-fast gas chromatography with flame ionization detector (UFGC-FID) associated with the supervised pattern recognition method, Partial Least Square Discriminant Analysis (PLS-DA). The sample set consisted of 171 Brazilian common gasoline (i.e., with ethanol in its composition) and 171 adulterated gasoline prepared in laboratory using 19 different solvents in the concentration range of 2–10% (v/v). The chromatographic method required only 2.85 min and the chromatograms presented 125 peaks on average. The PLS-DA model was developed with 3 latent variables and provided correlation coefficients close to 0.99 and correct discrimination of 100% of calibration and validation samples. Therefore, the developed UFGC-FID/PLS-DA method provided a sensitive, fast and automated alternative method for the detection of adulterants in the monitoring of gasoline quality.

1. Introduction

Brazil has experienced the biggest recession in its history and several sectors of production and services were inevitably affected by the economic crisis [1]. In 2015, the Fuel Quality Monitoring Program (PMQC) was also subject to budget cuts and, as consequence, the number of fuel samples analyzed per year reduced by 80% [2]. In the same year, the number of non-conforming samples, i.e. samples with physicochemical parameters outside the specifications regulated by the National Petroleum Agency (ANP), increased 0.7%, which is the largest increase since the appearance of the PMQC [3].

The composition of gasoline has great importance in engine performance and the emission of pollutants [4]. Gasoline is a complex mixture consisting mainly of paraffinic, olefinic and aromatic hydrocarbons ranging from 4 to 12 carbon atoms and, in a lower concentration, substances containing oxygen and sulfur [5]. The final composition of the gasoline depends on several factors such as the nature of the crude oil, the process that was used to obtain it, the presence of additives (detergents, dispersants, octane improvers, etc) and the local legislation that specifies the maximum content of methanol, benzene, sulfur, aromatic and olefin hydrocarbons [6,7]. In Brazil, anhydrous ethanol is a mandatory additive to replace tetraethyl lead in gasoline composition [6,8] and its concentration varies between 20% (v/v) and 27% (v/v) according to economic and production factors [6,7,9].

Unfortunately, reducing fuel monitoring provides greater opportunities for illegal adulteration of gasoline. Due to the complexity of gasoline composition, several miscible solvents can be added to this fuel without causing major changes in its physicochemical properties [10,11]. The gasoline adulteration is commonly practiced using low-cost solvents, such as kerosene, white spirit, naphtha, thinner, rubber solvent, or lower-value fuels such as diesel and ethanol, however, the use of hexane, toluene, and xylenes has also been reported by authors [6,12,13]. Therefore, an effective quality control of commercial gasoline is highly necessary for today's economic scenario, since the addition of these compounds or mixtures to gasoline can lead to engine malfunction, increase in fuel consumption, tax evasion, and environmental damage caused by the intensification of CO and NOx emissions [12–15].

The quality of the gasoline can be verified by several routine tests. For example, distillation curves may reveal whether the gasoline has been adulterated with higher or lower boiling solvents. According to Takeshita et al. (2008) [6], the presence of only 2% (v/v) of diesel fuel in gasoline composition can be easily detected by the final boiling point (FBP) of the distillation curve. However, a sample of gasoline can present non-conform results even without the addition of adulterants, as well as being adulterated and presenting physicochemical parameters within specifications [16,17]. For this reason, the ANP invested in the strategy of adding an isotopic marker to the solvents that are sold in Brazil [12]. If the gasoline sample presented the marker in the result of the official gas chromatographic (GC) method, the adulteration of gasoline was done with some type of solvent [10]. Although the marker monitoring has been an efficient methodology, insertion of new markers periodically into the solvent market requires large financial resources and, additionally, the marker detection is a laborious process since chromatographic analysis requires about 20 min per sample [10,12,17].

Many researchers have dedicated efforts to develop alternative methods to improve the gasoline quality monitoring. Many of these methods are based on spectroscopic techniques, such as FTIR [9,14,17–19], NIR [9,20,21], HS-MS [5,20], NMR [22–24], Raman [25,26], or GC methods [11,12,16,27–33] associated with chemometric techniques. However, few studies are devoted to detecting adulterants in gasoline through alternative methods [12,14,16–18,27,34]. Tanaka et al. (2011) [34] attempted to discriminate adulterated and unadulterated samples using the gasoline physicochemical parameters,

such as atmospheric distillation temperatures, research octane number (RON) and composition, however, the SIMCA model correctly classified only 77.1% of the prediction samples. The FTIR/LDA method developed by Pereira et al. (2006) [17] was able to correctly classify 96% of adulterated samples with four different solvents (kerosene, thinner, light and heavy naphtha). The FTIR/SIMCA method of Teixeira et al. (2008) [14] obtained satisfactory results with correct discrimination of 100% of samples adulterated with diesel, kerosene, turpentine, and thinner, in the concentration range of 0–50% (v/v). Skrobot et al. (2005, 2007) [16,27] carried out two studies on the identification of adulterants (thinner, kerosene, light and heavy naphtha) in gasoline using gas chromatography with flame ionization detector (GC-FID) combined with unsupervised (HCA and PCA) and supervised (KNN and LDA) pattern recognition methods. The GC-FID analysis required 75 min for each sample and the supervised methods presented low sensitivity (89.3%), therefore, the method was unsuitable for routine quality monitoring [16,27]. Pedroso et al. (2008) [12] developed a sophisticated method for quantifying adulterants (white spirit, kerosene, and paint thinner) using comprehensive two-dimensional gas chromatography with flame ionization detection (GC × GC-FID) and second-order multivariate calibration (multi-way partial least squares regression, N-PLS). The second-order data obtained from 40 min chromatographic analyses resulted in models with accuracy (RMSEP) ranged from 3.3% (v/v) to 8.2% (v/v), depending on the adulterant. In this study, no pattern recognition models were developed to detect the presence of adulterant prior to quantification [12].

The use of fast, automated and accurate methods for the routine analysis of fuel quality is crucial for the improvement of monitoring programs since there is a reduction in the consumption of analysis time, expenditures on inputs and manual labor. In this perspective, this paper proposes an analytical method for detection of several adulterants in gasoline using ultra-fast gas chromatography with flame ionization detector (UFGC-FID) associated with a supervised pattern recognition method, Partial Least Square Discriminant Analysis (PLS-DA).

2. Material and methods

2.1. Samples

The gasoline samples used in this work were acquired from a laboratory specialized in fuel analysis, Cempeqc (Center for Monitoring and Research of the Quality of Fuels, Biofuels, Crude Oil and Derivatives), which collects fuel samples in the state of São Paulo, Brazil. From 424 analyzed samples of gasoline, we selected 171 samples to construct a representative sample set of conforming and non-conforming gasoline. Since the physicochemical parameters may be out of specification even without the adulteration of gasoline, it was important to use a substantial number of nonconforming samples to avoid false positives after modeling. The physicochemical parameters of the 171 selected samples and the number of nonconformities are shown in Table 1.

Since the detection of low concentrations of solvent in gasoline is the greatest challenge for conventional methods [15], the adulterated samples set was prepared by adding solvents to the selected gasoline samples in the concentration range of 2–10% by volume. The adulteration was carried out with hydrocarbon (7 aliphatics; 5 aromatics; 1 cyclic; and 5 mixtures of hydrocarbon groups) and oxygenated (ether and dialcohol) solvents (Table 2). Nine samples were prepared using each solvent, totaling 171 adulterated samples.

2.2. Ultra-fast gas chromatographic analysis

Chromatograms of the 342 samples were obtained using an ultra-fast gas chromatograph (UFGC), Trace GC Ultra (Thermo Scientific), equipped with a direct resistive heating module, split/splitless injector, and high-frequency FID (300 Hz). The capillary chromatographic

Table 1
Physicochemical parameters of the gasoline samples, the number of nonconformities and specifications of ANP Regulation 40/2013.

Parameter	Method	Unit	Specification	Gasoline samples set				
				min	max	average	standard deviation	nonconforming samples
Relative density	D1298	$g\,cm^{-3}$	not specified	0.7128	0.7765	0.7430	0.0075	
Distillation curve								
10% evaporated	D86	°C	65.0, max.	46.3	60	53.9	1.6	
50% evaporated		°C	80.0, max.	69.5	74.3	72.1	0.9	
90% evaporated		°C	190, max.	134.6	180.9	158.8	6.9	
Final boiling point		°C	215, max.	168.8	265.6	202.4	8.7	8
Residue		% v/v	2, max.	0	1.7	0.9	0.3	
Octane number								
Motor octane number	Correlation to D2699	–	82, min.	81.4	84.4	82.4	0.4	7
Research octane number	D2700	–	not specified	89.1	101.6	96.7	1.5	
Anti-knocking index		–	2, min.	86.2	92.1	89.6	0.8	1
Composition								
Benzene	Correlation to D1319	% v/v	1.0, max.	0	0.6	0.4	0.1	
Saturates		% v/v	not specified	37.7	76.1	48.2	5.9	
Olefins		% v/v	25, max.	0	22.6	12.5	4.7	
Aromatics		% v/v	35, max.	8.5	26.3	14.7	2.3	
Anhydrous ethanol	D5501	% v/v	25 ± 1	22	31	25.2	1.6	55

Table 2
Solvent used in adulterated gasoline sample set.

Classification		Solvent
Hydrocarbon	Aliphatic	Light aliphatic DT solvent (Carbosolv)
		Light aliphatic SB solvent (Carbosolv)
		Medium aliphatic AD solvent (Carbosolv)
		Solvent A-4070 (Carbosolv)
		n-heptane 99.5% P.A (Vetec Química Fina)
		n-hexane 95% P.A (Vetec Química Fina)
	Aromatic	Aromatic solvent AB-9 (Carbosolv)
		Benzolum crystallisabile (ECIBRAS)
		Toluene 99,9% P.A. (J.T. Baker)
		Xylol 98% P.A. (Vetec Química Fina)
		Ethylbenzene 99,8% P.A. (Acros Organic)
	Cyclic	Cyclohexane 99% P.A. (Vetec Química Fina)
	Mixture	White spirit (Thinsol Química)
		Halogen free solvent (Carbosolv)
		Rubber solvent
		Heavy naphtha
		Kerosene
Oxygenated	Ether	Ethyl ether (Isofar)
	Dialcohol	Ethylene glycol 99,5% P.A. (Carlo Erba)

column used for separation was a Thermo Scientific PH5 (5% phenyl; 95% dimethylpolysiloxane, 5 m × 0.10 mm id × 0.40 μm) surrounded by a resistor to provide high heating rates. The samples were injected by an AS3000 autosampler with a 5 μL syringe and the injection volume was 0.1 μL. The temperature of the injector and the detector were set at 270 °C. The flow of the gases used by the detector were 45 mL min^{-1} for hydrogen, 30 mL min^{-1} for nitrogen (make-up gas), and 350 mL min^{-1} for synthetic air. Hydrogen was used as carrier gas due to the high diffusion coefficient and its flow was kept constant at 0.5 mL min^{-1}. As the samples were not diluted in any solvent prior to injection, the split rate used was 1:500. The column temperature programming started at 40 °C, held for 0.60 min, and raised at a rate of 100 °C min^{-1} to 250 °C, maintained for 0.15 min. The chromatographic analysis time was 2.85 min per sample.

2.3. Chemometric analysis

The data were collected using ChromQuest software (Thermo Scientific) and then transferred to Matlab 2013a (Mathworks) with PLS toolbox 7.3.1 (Eigenvector Research Inc.), where the alignment and

chemometric analyses were performed using a Windows 8 PC running on an Intel Core i7 3632QM 2.20 GHz processor with 8.0 Gb of RAM. The obtained chromatograms were exported as vectors for the modeling and, as the data acquisition rate was about 10 points per second, each chromatogram consisted of 1711 variables. Thus, the combination of all vectors resulted in a 342 × 1711 data matrix.

The first step prior to modeling was the alignment of the chromatograms. The displacements of the peaks were corrected using the Correlation Optimized Warping (COW) algorithm [35]. Firstly, the alignment requires a reference chromatogram containing peaks which show a relatively consistent maximum. The algorithm searches for peaks in the mean chromatogram by identification of positions with a clear inflection point as a peak maximum [36]. Then, the chromatograms are divided into windows, which are aligned with the respective reference chromatogram windows. The size of the chromatograms will be extended or compressed according to the defined slack and, finally, the number of points of each window will be equalized to the reference window [35]. Different values for the slack and window size parameters were tested and evaluated visually to provide the best alignment.

After the correction of peak displacement, the samples were divided into a calibration set (230 samples) for the development of the multivariate model and validation set (112 samples) to test the model. The Onion algorithm was used to select the samples with less covariance (based on distance from the mean) for each set and, consequently, to obtain greater sample representativeness in both sets [37,38]. The 1:1 ratio of unadulterated and adulterated samples was maintained at calibration and validation.

From the premise that gasoline samples are composed mainly of hydrocarbons in the range of C6–C12 [39] and that the compounds have different response factors using FID [40], the data were autoscaled prior to modeling to equalize the impact of all variables [41].

The Partial Least Square Discriminant Analysis (PLS-DA) was used to develop the pattern recognition model. The PLS-DA is based on the PLS algorithm, however, it uses the encoded classes as dependent vector, **y**, rather than continuous numbers such as the concentration of an analyte [42]. The PLS method maximizes the relationship between the dependent variable and the scores and, consequently, the latent variables (LV) represent the directions that best discriminate the classes. When the application involves discrimination of only two classes, the PLS1 method is employed. In this paper, we labeled the classes as −1 (unadulterated) and +1 (adulterated) to center the values of **y**, making the algebraic derivations simpler, and the threshold
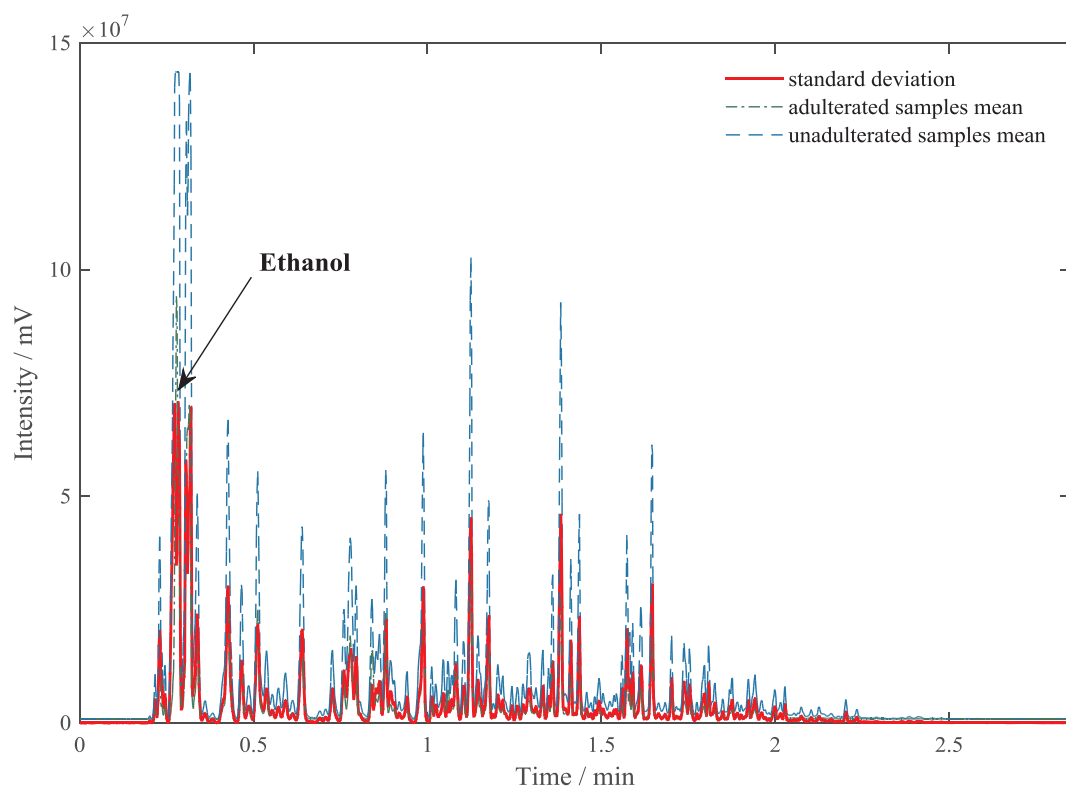
Fig. 1. Means of the chromatograms of the adulterated samples and of the unadulterated samples and the standard deviation of the two means.

probability value was set at 50%. The number of LV was chosen based on root mean square errors of cross-validation (RMSECV) values of the two classes in order to minimize the prediction errors and avoid model overfitting [43,44]. More details on PLS-DA method can be found in the Ref. [42].

3. Results and discussion

3.1. Samples

According to the PMQC statistics, between 2007 and 2016, the average of nonconforming samples in the country was equal to 1.7% and the main nonconformities were related to the content of anhydrous ethanol (41%), distillation temperatures (31%), and octane number (11%) [3]. In this work, the sample set consisted of 38% of non-conforming samples and the nonconformities were 1% related to anti-knock index, 10% to motor octane number, 11% to final boiling point and 77% to anhydrous ethanol content. The selection of many non-conforming samples was due to the objective of this work, which is to identify the presence of adulterants regardless of the sample conformity. In addition, ethanol was not used as an adulterant, since 19% of the samples used had an ethanol content above the amount specified by the legislation. Other works found in the literature have already shown the possibility of discriminating gasoline samples according to conformity through GC-FID [30] or mid-infrared [19] combined with the chemometric techniques, therefore, it was not the focus of the current paper.

3.2. UFGC-FID analysis

3.2.1. Chromatographic separation

The main reason for the use of gas chromatography in solvent detection by the marker method is to obtain the information of individual compounds in gasoline [20,23]. However, a proper identification of the marker requires a good chromatographic resolution, which is not an

easy task since gasoline is a mixture of hundreds of compounds. A good separation of the gasoline compounds by conventional GC requires a long analysis time, thus the detection of solvents in gasoline becomes a costly process [23]. Since the UFGC system uses smaller columns with reduced internal diameter and resistive heating system, the chromatographic analyses are much faster than conventional GC and provide satisfactory resolutions [45]. Unlike the detection of solvents in gasoline by the marker method, chemometric tools allows the adulteration detection through the whole chromatogram, that is, there is no need for high-resolution [33].

The analyzed samples presented an average of 125 chromatographic peaks, i.e., many gasoline compounds were not separated by the developed UFGC method since gasoline typically has more than 230 compounds [47]. Nevertheless, the class discrimination can be performed by pattern recognition methods even when there is no high chromatographic resolution. Therefore, a satisfactory separation of the gasoline compounds for the classification model could be achieved in only 2.85 min through the UFGC-FID.

3.2.2. Peak alignment

Variations in the retention time of the analytes are very common to occur in chromatographic analyses due to oscillations of instrumental parameters such as pressure, temperature, and gas flow [46]. Thus, the application of pattern recognition or regression methods is highly dependent on peak alignment. The peak alignment was performed by the COW algorithm and the mean chromatogram was taken as reference. Different values of COW parameters were evaluated, and the best alignment was obtained using slack and window size equal to 5 and 10, respectively.

3.2.3. Comparison between class chromatograms

To evaluate the main differences between the chromatograms of adulterated and unadulterated samples, the mean chromatogram of each class was obtained, and the standard deviation of the means was calculated (Fig. 1). The standard deviation of the means of the classes

reveals the peaks with greater difference of intensity between the classes. The peak with the greatest difference was related to the retention time of the ethanol (Fig. 1). This can be justified by the dilution of the ethanol present in the mixture when other solvents are added to the gasoline.

The adulteration of gasoline using pure solvents, such as benzene, toluene, ethylbenzene and xylenes (BTEX), played an important role in verifying the sensitivity of the developed model because these compounds are important in gasoline composition and their sum may reach up to 35% by volume [7]. The addition of pure solvents to the gasoline resulted in the dilution not only of the ethanol but of the other gasoline compounds. Thus, the more intense peaks of the gasoline samples presented the highest standard deviation.

3.3. Chemometric analysis

3.3.1. Model development

The classification model was developed using the PLS-DA method after the peak alignment by the COW algorithm. Although a 1:1 ratio of adulterated and unadulterated samples was used in the model development, this ratio is not a requirement to obtain a high sensitivity model. The steps of the PLS-DA method provide an adjustment according to the class distribution. The first step of the modeling is to estimate the values of y by the PLS method. Thereafter, assuming the values of y predicted have a normal distribution, the threshold value (discriminant function) is calculated by the Bayesian method to obtain the same prediction probability for both classes. If the number of samples in each group is different, the threshold value is adjusted [42,48,49]. As the classes were coded as -1 (unadulterated) and $+1$ (adulterated) and we used the same number of samples in each class, the threshold was close to zero (0.018).

Autoscaling is a typical preprocessing of chromatographic data when FID is used as detector because different compounds usually present different response factors. In addition, autoscaling is useful to equalize compounds with different variances of concentration [41]. However, regions of the chromatogram where there is no retention of analytes, i.e. noisy variables for the model, are amplified using this preprocessing. Therefore, the initial variables of the chromatograms (between 0.00 s and 0.18 s) were excluded and the final matrix was constituted of 1602 variables.

The number of latent variables (LV) was selected through the optimal values for each class to provide the minor prediction error for both classes without generating an overfitting [44]. Fig. 2 reveals that 3 LVs is the most appropriate LV number for the modeling since the
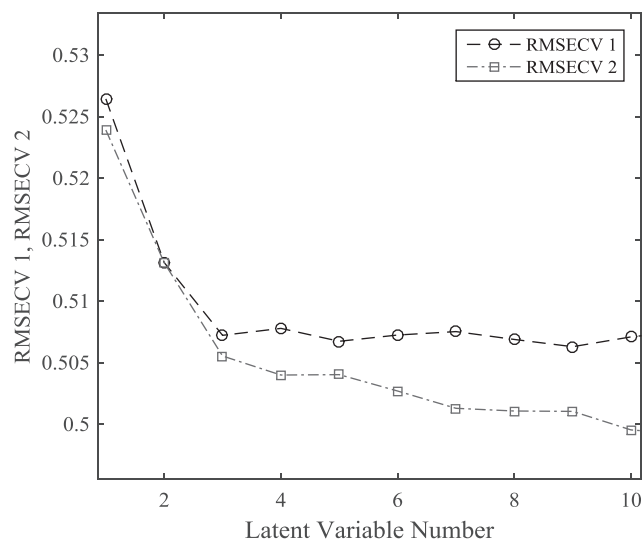

Fig. 2. Latent variable number versus RMSECV values plot.

addition of more LV does not significantly reduce the values of RMSECV1 (unadulterated samples) and RMSECV2 (adulterated samples).

The scores plot with 3 LV (Fig. 3) shows that the classes were well-separated in the 1st LV, which explained 55.62% of the X-Block variance. The addition of the 2nd and 3rd LVs increased the distance between the groups providing a better discrimination ability to the model.

3.3.2. Model validation

The application of the alternative method in the analysis of unknown samples depends on its accuracy and representativeness. The accuracy of a PLS-DA model can be evaluated by the sensitivity (true positive rate), the prediction error values and the correlation coefficients (r) [25,50]. Sensitivity equal to 100% indicates that the model correctly classified all samples, while low prediction errors and high correlation coefficients are indicative of how well fitted the model is.

The results in Table 3 were obtained by the cross-validation (venetian blinds) and by the prediction of 112 samples (validation set) that were not used in the model development. In both validations, the model correctly predicted 100% of the samples and presented correlation coefficients above 0.98. Given that the sample set consisted of 171 gasoline samples and 19 different adulterants, the developed model was accurate representative for the monitoring of gasoline adulteration. However, the constant addition of new samples to the prediction model is an important practice to make the model more robust.

3.3.3. Evaluation of multivariate filters

The exclusion of the variables between 0.00 and 0.18 s resulted in a less parsimonious modeling, but there was no significant reduction in RMSE values and increase in correlation coefficients. To reduce RMSE values and improve the model accuracy, multivariate filters were evaluated as preprocessing. Multivariate filters such as Orthogonal Signal Correction (OSC), Generalized Least Squares Weighting (GLSW) and External Parameter Orthogonalization (EPO) aim to increase model selectivity by reducing sources of unwanted variations or covariance structures. Each filter has a different mathematical approach and detailed information can be found in the Ref. [51–53]. Among the tested filters, only the EPO with two principal components improved the PLS-DA fit ($r_{CV} = 0.9878$ and $r_{pred} = 0.9886$), however, the increase in correlation coefficients was not significant. Since the application of these multivariate filters occurs in several stages, the calibration process and the prediction of new samples requires more computational processing and becomes more time-consuming. Therefore, we chose to use only autoscale preprocessing and the x variables after 0.19 s, since there was no increase in computational processing.

3.3.4. Model selectivity ratio

The selectivity ratio for Y plot is an indicative of which variables are most important for the PLS1 discrimination. Firstly, the y values (classes) are used as target to obtain a single predictive target-projected through the PLS components. Then, the selectivity ratio plot is obtained by calculating the ratio between explained and residual variance of the x variables on the target-projected component [54]. The higher the ratio, the greater the influence of the variable in the prediction. The selectivity ratio plot, in Fig. 4, showed that the most important variables were related to the retention time of the ethanol peak. The selectivity ratio also justifies the insignificant improvement when using multivariate filters (OSC, GLSW or EPO) since most of the x variables were useful in prediction. The variables after 2 min did not present high selectivity for Y, so the time of chromatographic analysis can be reduced by increasing the heating rate after 2 min without decreasing the performance of the model.

3.3.5. Model limitations

Although the sample set used in this study is representative in view of the practice of adulteration of gasoline in Brazil, the classification
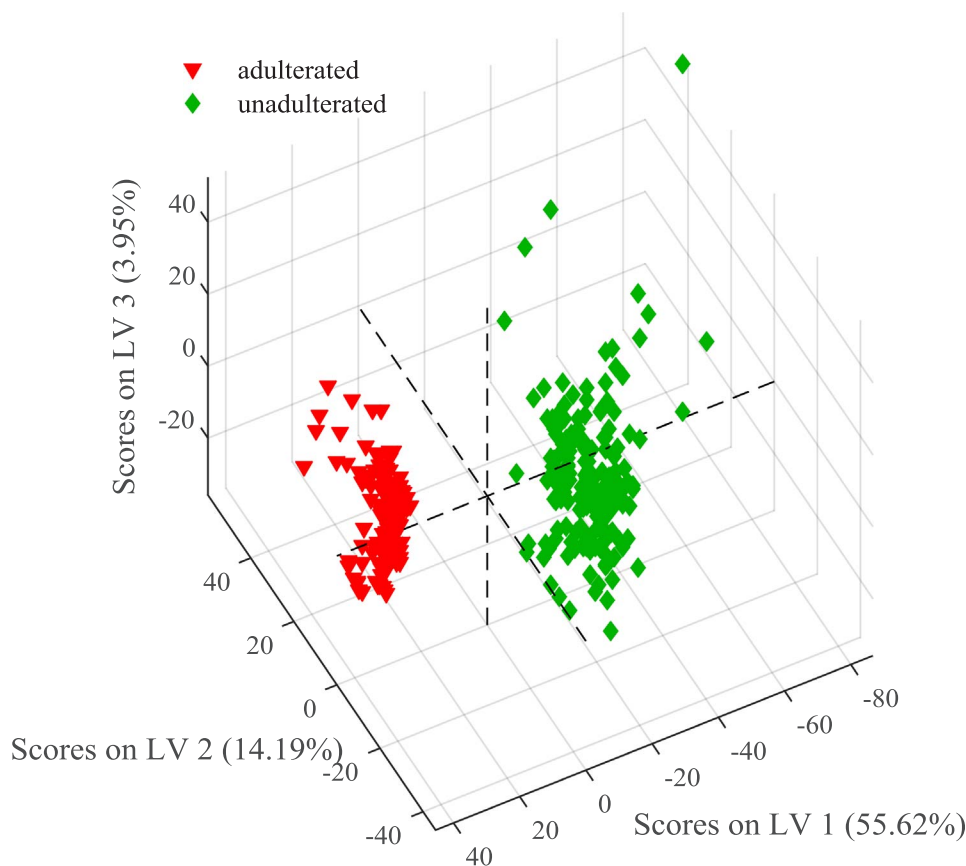
Fig. 3. Scores plot from PLS-DA model.

Table 3

Parameters of the PLS-DA model for adulterated and unadulterated gasoline discrimination.

Parameter	Value	
Variables	1602	
X block variance captured	73.77%	
Y block variance captured	48.93%	
Latent variables	3	
	unadulterated	adulterated
Sensitivity (cal)	1.000	1.000
Sensitivity (CV)	1.000	1.000
Sensitivity (pred)	1.000	1.000
RMSEC	0.5096	0.5010
RMSECV	0.5072	0.5055
RMSEP	0.4990	0.5128
r (cal)	0.9893	0.9893
r (CV)	0.9871	0.9871
r (pred)	0.9883	0.9883

Fig. 4. Selectivity ratio for each variable from PLS-DA model.

model is limited to the physicochemical parameters of the calibration samples, to the solvents used in the adulteration and to the Brazilian gasoline composition, which has anhydrous ethanol. Changes in the specifications of gasoline physicochemical parameters will require a revalidation of the model. In addition, new samples should be periodically added to the model to make it more robust and up-to-date with the fuel trade.

The limitations mentioned above will be present in prediction models independently of the instrumental technique used. In this perspective, the use of UFGC-FID associated to the PLS-DA method has advantages such as speed, automation and sensitivity for the detection of adulterants in commercial gasoline. Therefore, the method proposed in this work is suitable for the routine monitoring of gasoline quality.
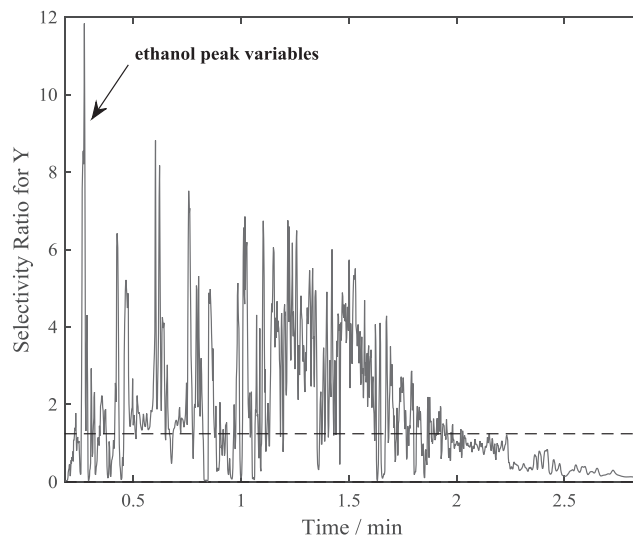
4. Conclusion

Ultra-fast gas chromatography associated with Partial Least Square Discriminant Analysis provided a rapid and sensitive analytical method for detecting adulterants in gasoline. Through chromatographic analyses of only 2.85 min, it was possible to correctly discriminate 100% of the adulterated samples from the unadulterated samples. The method was sensitive to 19 different solvents in concentrations between 2 and 10% (v/v) and the discrimination ability was independent of the nonconformities of the gasoline samples. According to the selectivity ratio of the x variables for the prediction of Y, the time of chromatographic analysis can be further reduced since the x variables after 2 min are not

significantly useful for class discrimination. In view of the current need to implement a fast and automated method for the detection of adulterants in gasoline, the method proposed here is fully adequate for the monitoring of Brazilian gasoline quality.

Acknowledgements

References

[1] Irredeemable? Econ 2016:1–9. https://www.economist.com/news/briefing/21684778-former-star-emerging-world-faces-lost-decade-irredeemable.

[2] ANP. Boletim de Monitoramento da Qualidade dos Combustíveis 2017. http://www.anp.gov.br/wwwanp/publicacoes/boletins-anp/2388-pmqc-edicoes-anteriores.

[3] ANP. Anuário estatístico brasileiro do petróleo, gás natural e biocombustíveis: 2016. Rio de Janeiro: 2016.

[4] Wang Y, Rong Z, Qin Y, Peng J, Li M, Lei J, et al. The impact of fuel compositions on the particulate emissions of direct injection gasoline engine. Fuel 2016;166:543–52. http://dx.doi.org/10.1016/j.fuel.2015.11.019.

[5] Ferreiro-González M, Ayuso J, Álvarez JA, Palma MG, Barroso C. New headspace-mass spectrometry method for the discrimination of commercial gasoline samples with different research octane numbers. Energy Fuels 2014;28:6249–54. http://dx.doi.org/10.1021/ef5013775.

[6] Takeshita EV, Rezende RVP, de Souza SMAGU, de Souza AAU. Influence of solvent addition on the physicochemical properties of Brazilian gasoline. Fuel 2008;87:2168–77. http://dx.doi.org/10.1016/j.fuel.2007.11.003.

[7] ANP. Resolução ANP No 40 DE 25/10/2013 2013.

[8] Da Silva R, Cataluña R, Menezes EW De, Samios D, Piatnicki CMS. Effect of additives on the antiknock properties and Reid vapor pressure of gasoline. Fuel 2005;84:951–9. http://dx.doi.org/10.1016/j.fuel.2005.01.008.

[9] Da Silva MPF, Brito LRE, Honorato FA, Paim APS, Pasquini C, Pimentel MF. Classification of gasoline as with or without dispersant and detergent additives using infrared spectroscopy and multivariate classification. Fuel 2014;116:151–7. http://dx.doi.org/10.1016/j.fuel.2013.07.110.

[10] de Paulo JM, Mendes G, Barros JEM, Barbeira PJS. A study of adulteration in gasoline samples using flame emission spectroscopy and chemometrics tools. Analyst 2012;137:5919–24. http://dx.doi.org/10.1039/c2an35441a.

[11] Ugena L, Moncayo S, Manzoor S, Rosales D, Cáceres JO. Identification and discrimination of brands of fuels by gas chromatography and neural networks algorithm in forensic research. J Anal Methods Chem 2016;2016. http://dx.doi.org/10.1155/2016/6758281.

[12] Pedroso MP, de Godoy LAF, Ferreira EC, Poppi RJ, Augusto F. Identification of gasoline adulteration using comprehensive two-dimensional gas chromatography combined to multivariate data processing. J Chromatogr A 2008;1201:176–82. http://dx.doi.org/10.1016/j.chroma.2008.05.092.

[13] Ré-Poppi N, Almeida FFP, Cardoso CAL, Raposo Jr. JL, Viana LH, Silva TQ, et al. Screening analysis of type C Brazilian gasoline by gas chromatography – flame ionization detector. Fuel 2009;88:418–23. http://dx.doi.org/10.1016/j.fuel.2008.10.014.

[14] Teixeira LSG, Oliveira FS, dos Santos HC, Cordeiro PWL, Almeida SQ. Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline. Fuel 2008;87:346–52. http://dx.doi.org/10.1016/j.fuel.2007.05.016.

[15] Mabood F, Gilani SA, Albroumi M, Alameri S, Al Nabhani MMO, Jabeen F, et al. Detection and estimation of Super premium 95 gasoline adulteration with Premium 91 gasoline using new NIR spectroscopy combined with multivariate methods. Fuel 2017;197:388–96. http://dx.doi.org/10.1016/j.fuel.2017.02.041.

[16] Skrobot VL, Castro EVR, Pereira RCC, Pasa VMD, Fortes ICP. Identification of adulteration of gasoline applying multivariate data analysis techniques HCA and KNN in chromatographic data. Energy Fuels 2005;19:2350–6. http://dx.doi.org/10.1021/ef050031l.

[17] Pereira RCC, Skrobot VL, Castro EVR, Fortes ICP, Pasa VMD. Determination of gasoline adulteration by principal components analysis-linear discriminant analysis applied to FTIR spectra. Energy Fuels 2006;20:1097–102. http://dx.doi.org/10.1021/ef050203e.

[18] Al-Ghouti MA, Al-Degs YS, Amer M. Determination of motor gasoline adulteration using FTIR spectroscopy and multivariate calibration. Talanta 2008;76:1105–12. http://dx.doi.org/10.1016/j.talanta.2008.05.024.

[19] Khanmohammadi M, Garmarudi AB, Ghasemi K, De La Guardia M. Quality based classification of gasoline samples by ATR-FTIR spectrometry using spectral feature selection with quadratic discriminant analysis. Fuel 2013;111:96–102. http://dx.doi.org/10.1016/j.fuel.2013.04.001.

[20] Ferreiro-González M, Ayuso J, Álvarez JA, Palma M, Barroso CG. Gasoline analysis by headspace mass spectrometry and near infrared spectroscopy. Fuel 2015;153:402–7. http://dx.doi.org/10.1016/j.fuel.2015.03.019.

[21] Balabin RM, Safieva RZ. Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. Fuel 2008;87:1096–101. http://dx.doi.org/10.1016/j.fuel.2007.07.018.

[22] Monteiro MR, Ambrozin ARP, Liao LM, Boffo EF, Tavares LA, Ferreira MMC, et al. Study of Brazilian gasoline quality using hydrogen nuclear magnetic resonance (^{1}H NMR) spectroscopy and chemometrics. Energy Fuels 2009;23. http://dx.doi.org/10.1021/ef800436p.

[23] Flumignan DL, Boralle N, De Oliveira JE. Screening Brazilian commercial gasoline quality by hydrogen nuclear magnetic resonance spectroscopic fingerprintings and pattern-recognition multivariate chemometric analysis. Talanta 2010;82:99–105. http://dx.doi.org/10.1016/j.talanta.2010.04.002.

[24] Kaiser CR, Borges JL, dos Santos AR, Azevedo DA, DAvila LA. Quality control of gasoline by 1H NMR: aromatics, olefinics, paraffinics, and oxygenated and benzene contents. Fuel 2010;89:99–104. http://dx.doi.org/10.1016/j.fuel.2009.06.023.

[25] Li S, Dai LK. Classification of gasoline brand and origin by Raman spectroscopy and a novel R-weighted LSSVM algorithm. Fuel 2012;96:146–52. http://dx.doi.org/10.1016/j.fuel.2012.01.001.

[26] Flecher PE, Welch WT, Albin S, Cooper JB. Determination of octane numbers and Reid vapor pressure in commercial gasoline using dispersive fiber-optic Raman spectroscopy. Spectrochim Acta Part A Mol Biomol Spectrosc 1997;53A:199–206. http://dx.doi.org/10.1016/S1386-1425(97)83026-0.

[27] Skrobot VL, Castro EVR, Pereira RCC, Pasa VMD, Fortes ICP. Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration. Energy Fuels 2007;21:3394–400. http://dx.doi.org/10.1021/ef0701337.

[28] Pierce KM, Hope JL, Johnson KJ, Wright BW, Synovec RE. Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis. J Chromatogr A 2005;1096:101–10. http://dx.doi.org/10.1016/j.chroma.2005.04.078.

[29] Watson NE, VanWingerden MM, Pierce KM, Wright BW, Synovec RE. Classification of high-speed gas chromatography-mass spectrometry data by principal component analysis coupled with piecewise alignment and feature selection. J Chromatogr A 2006;1129:111–8. http://dx.doi.org/10.1016/j.chroma.2006.06.087.

[30] Flumignan DL, Tininis AG, Ferreira F de O, de Oliveira JE. Screening Brazilian C gasoline quality: application of the SIMCA chemometric method to gas chromatographic data. Anal Chim Acta 2007;595:128–35. http://dx.doi.org/10.1016/j.aca.2007.02.049.

[31] Flumignan DL, de Oliveira Ferreira F, Tininis AG, de Oliveira JE. Multivariate calibrations in gas chromatographic profiles for prediction of several physicochemical parameters of Brazilian commercial gasoline. Chemom Intell Lab Syst 2008;92:53–60. http://dx.doi.org/10.1016/j.chemolab.2007.12.003.

[32] Rudnev VA, Boichenko AP, Karnozhytskiy PV. Classification of gasoline by octane number and light gas condensate fractions by origin with using dielectric or gas-chromatographic data and chemometrics tools. Talanta 2011;84:963–70. http://dx.doi.org/10.1016/j.talanta.2011.02.049.

[33] Parastar H, Mostafapour S, Azimi G. Quality assessment of gasoline using comprehensive two-dimensional gas chromatography combined with unfolded partial least squares: a reliable approach for the detection of gasoline adulteration. J Sep Sci 2016;39:367–74. http://dx.doi.org/10.1002/jssc.201500720.

[34] Tanaka GT, De Oliveira Ferreira F, Ferreira da Silva CE, Flumignan DL, De Oliveira JE. Chemometrics in fuel science: demonstration of the feasibility of chemometrics analyses applied to physicochemical parameters to screen solvent tracers in Brazilian commercial gasoline. J Chemom 2011;25:487–95. http://dx.doi.org/10.1002/cem.1394.

[35] Nielsen NPV, Carstensen JM, Smedsgaard J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. J Chromatogr A 1998;805:17–35. http://dx.doi.org/10.1016/S0021-9673(98)00021-1.

[36] Eigenvector Research. Registerspec 2009.

[37] Sousa AG, Ahl LI, Pedersen HL, Fangel JU, Sørensen SO, Willats WGT. A multivariate approach for high throughput pectin profiling by combining glycan microarrays with monoclonal antibodies. Carbohydr Res 2015;409:41–7. http://dx.doi.org/10.1016/j.carres.2015.03.015.

[38] Shrestha S, Deleuran L, Gislum R. Classification of different tomato seed cultivars by multispectral visible-near infrared spectroscopy and chemometrics. J Spectr Imaging 2016;5:a1. http://dx.doi.org/10.1255/jsi.2016.a1.

[39] Wiedemann LSM, D'Avila LA, Azevedo DA. Adulteration detection of Brazilian gasoline samples by statistical analysis. Fuel 2005;84:467–73. http://dx.doi.org/10.1016/j.fuel.2004.09.013.

[40] Scanlon JT, Willis DE. Calculation of flame ionization detector relative response factors using the effective carbon number concept. J Chromatogr Sci 1985;23:333–40. http://dx.doi.org/10.1093/chromsci/23.8.333.

[41] Gemperline P. Practical Guide to Chemometrics, second ed., 2006. doi: 10.1201/9781420018301.

[42] Brereton RG, Lloyd GR. Partial least squares discriminant analysis: taking the magic away. J Chemom 2014;28:213–25. http://dx.doi.org/10.1002/cem.2609.

[43] Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004;44:1–12. http://dx.doi.org/10.1021/ci0342472.

[44] Di Anibal CV, Callao MP, Ruisánchez I. 1H NMR variable selection approaches for classification. A case study: the determination of adulterated foodstuffs. Talanta 2011;86:316–23.

[45] Dorman FL, Overton EB, Whiting JJ, Cochran JW, Gardea-torresdey J. Gas Chromatography. Micro 2008;80:4487–97. http://dx.doi.org/10.1021/ac800714x.

[46] Johnson KJ, Wright BW, Jarman KH, Synovec RE. High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. J Chromatogr A 2003;996:141–55. http://dx.doi.org/10.1016/S0021-9673(03)00616-2.

[47] O'Shay TA, Hoddinott KB. Analysis of soils contaminated with petroleum constituents. Philafrlphia: ASTM; 1994.

[48] Wong KH, Razmovski-Naumovski V, Li KM, Li GQ, Chan K. Differentiation of Pueraria lobata and Pueraria thomsonii using partial least square discriminant analysis (PLS-DA). J Pharm Biomed Anal 2013;84:5–13. http://dx.doi.org/10.1016/j.jpba.2013.05.040.

[49] Eigenvector Research Incorporated. How is the prediction probability and threshold calculated for PLSDA? FAQ n.d.:1. http://www.eigenvector.com/faq/index.php?id=38%7C (accessed January 1, 2017).

[50] Yin M, Tang S, Tong M. Identification of edible oils using terahertz spectroscopy combined with genetic algorithm and partial least squares discriminant analysis. Anal Methods 2016;8:2794–8. http://dx.doi.org/10.1039/C6AY00259E.

[51] Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. Chemom Intell Lab Syst 1998;44:175–85. http://dx.doi.org/10.1016/S0169-7439(98)00109-9.

[52] Zorzetti BM, Shaver JM, Harynuk JJ. Estimation of the age of a weathered mixture of volatile organic compounds. Anal Chim Acta 2011;694:31–7. http://dx.doi.org/10.1016/j.aca.2011.03.021.

[53] Roger JM, Chauchard F, Bellon-Maurel V. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemom Intell Lab Syst 2003;66:191–204. http://dx.doi.org/10.1016/S0169-7439(03)00051-0.

[54] Rajalahti T, Arneberg R, Berven FS, Myhr KM, Ulvik RJ, Kvalheim OM. Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. Chemom Intell Lab Syst 2009;95:35–48. http://dx.doi.org/10.1016/j.chemolab.2008.08.004.