Determination of Gasoline Adulteration by Principal Components Analysis-Linear Discriminant Analysis Applied to FTIR Spectra

Rita C. C. Pereira,[†] Vinicius L. Skrobot,[†] Eustáquio V. R. Castro,[‡] Isabel C. P. Fortes,[†] and Vânya M. D. Pasa*,[†]

Laboratório de Ensaios de Combustíveis, Departamento de Química, ICEx Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, Minas Gerais, Brasil CEP: 31.270-901, and Departamento de Química, Centro de Ciências Exatas, Universidade Federal do Espírito Santo, Av. Fernando Ferrari, s/n°, Vitória, Espírito Santo, Brasil CEP: 29060-900

Brazilian automotive gasoline has been the target of constant adulterations in an attempt to raise profit margins as a result of the market flexibility and the ensuing increased competition. Gasoline adulteration with solvents is common because solvents present lower taxation in comparison to gasoline. In the face of this, the importance of developing low-cost analytical techniques such as spectroscopy to certify the quality and authenticity of fuels used in Brazil is conspicuous. This work used IR spectroscopy and multivariate techniques (principal components analysis-linear discriminant analysis, PCA-LDA) to determine gasoline adulteration by solvent and to identify the solvent added. The results show that FTIR associated with PCA-LDA is a powerful technique in the quality control of automotive gasoline. The method sensitivity was 8% v/v with 96% efficiency in the classification of adulterated and unadulterated gasoline and 93% efficiency in the identification of the type of solvent added.

1. Introduction

Adulteration of gasoline, a criminal practice, has been observed in Brazil since market reform with and the end of the fuel distribution monopoly. The National Agency of Petroleum (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, ANP), in association with other Brazilian inspection agencies, has made every possible effort to minimize this problem as it may cause damage to vehicles and consequently to consumers. Additionally, it can affect public coffers through tax embezzlement because solvents, petrochemical fractions, and fuels are levied at different rates. Distributors and fuel station owners are not allowed to mix solvents or petrochemical fractions to formulate gasoline, even if to improve properties. Gasoline can only be produced in refineries or by petrochemical formulators with the express authorization of ANP. Nevertheless, tax evasion due to unauthorized mixture of solvents to gasoline amounts to millions of dollars per year. Solvents such as mineral spirits, kerosene, rubber solvent, petrochemical naphtha, diesel and thinner, and their mixtures have been used in gasoline adulteration. The main factors that contribute to the use of solvents in gasoline adulteration are their similar chemical composition, low cost, and lower tax rates.

Recently, ANP begun to introduce tracers in all solvents used in Brazil, national or imported.[1] Marked solvents added to gasoline can be detected by a specific analytical technique. This technology has been shown to be efficient, but it consumes some thousands of dollars every year in both marking and/or tracer detection. It is very common to find in the Brazilian market gasoline with all physicochemical proprieties in accordance with ANP specification but with the presence of solvent traces.

Therefore, inexpensive analytical methods to certify the quality and authenticity of the fuels used in Brazil are important. Spectroscopic methods are potential suitable candidates. The use of near- or mid-IR absorption spectroscopy associated with multivariate techniques such as linear discriminant analysis (LDA) for the analysis of adulterated foods and drugs has been well publicized.[2−4]

LDA is used to classify a group observation when the quadratic distance of the average of the observation of the group is minimized.[5−7] Each group has a linear discriminant function. However, LDA cannot be used for group samples if the number of variables of the data matrix is higher than the number of samples. This is due to the linear discriminant functions obtained to classify the samples having a covariance matrix that cannot have an inverse.[7−9] This type of problem is very common when dealing with spectral date matrix. To curb this problem, the matrix is downsized and analysis is carried out through the principal components analysis (PCA) in the original date matrix.[9−16]

* Corresponding author. Phone: 55-31-3499-6650. Fax: 55-31-3499-5700. E-mail: vanya@ufmg.br.
† Universidade Federal de Minas Gerais.
‡ Universidade Federal do Espírito Santo.

(1) Portaria Agência Nacional do Petróleo, 274 de 01/11/2001. http://www.anp.gov.br.

(2) Al-Jowder, O.; Kemsley, E. K.; Wilson, R. H. *J. Agric. Food Chem.* **2004,** *50,* 1325−1329.

(3) Paradkar, M. M.; Sivakesava, S.; Irudayaraj, J. *J. Sci. Food Agric.* **2003,** *83,* 714−721.

(4) Seregély, Z.; Deák, T.; Bisztray, G. D. *Chemom. Intell. Lab. Syst.* **2004,** *72,* 195−203.

(5) Yang, J.; Yang, J. *Pattern Recognit.* **2003,** *36,* 563−566.

(6) Klecka, W. R. *Discriminant Analysis*; Sage Publications: Beverly Hills, CA, 1980.

(7) Huberty, C. J. *Applied Discriminant Analysis*; Wiley & Sons: New York, 1994.

(8) Brereton, R. G. Chemometrics: Data Analysis for the Laboratory and Chemical Plant, 1st ed.; Wiley & Sons: Hoboken, NJ, 2003.

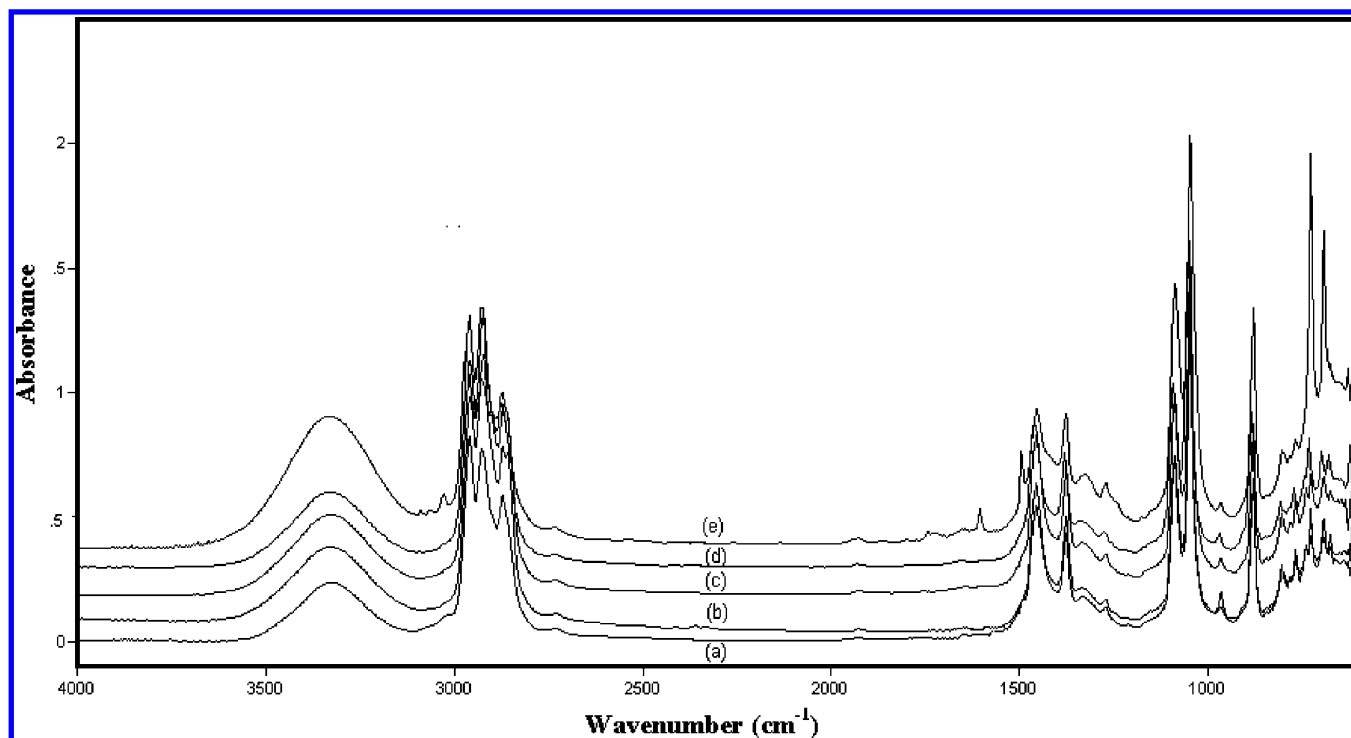(9) Keemley, E. K. *Chemom. Intell. Lab. Syst.* **1996,** *33,* 47−61.

(10) Kim, H.; Kim, D.; Bang, S. Y. *Pattern Recognit.* **2003,** *36,* 1095−1105.

Figure 1. Infrared absorption spectra of (a) unadulterated gasoline, (b) gasoline + heavy naphtha 40% v/v, (c) gasoline + kerosene and 40% v/v, (d) gasoline + light naphtha 40% v/v, and (e) gasoline + thinner 40% v/v.

Table 1. Average Physical Chemical Properties of the Brazilian Gasoline Studied

characteristic/method	ANP specification	unadulterated gasoline	gasoline + heavy naphtha	gasoline + kerosene	gasoline + light naphtha	gasoline + thinner
fuel anhydride ethyl alcohol (%v/v)/NBR 13992	25.00	25.10	25.23−25.90	24.80−25.20	24.90−25.80	25.40−**33.70**
MON, min/ASTM D 2700	82.00	83.40	82.00−84.67	**81.10**−83.60	**81.10**−83.60	83.90−113.40
IAD, min/ASTM D 2699	87.00	90.80	87.40−92.60	87.40−90.03	**86.83**−90.80	**73.10**−91.90
benzene %v/v, max/ASTM D3606, D5443, D6277	1.00	0.60	0.09−0.59	0.21−0.58	0.50−0.67	0.60−**1.59**
density (g/cm^3)/ASTM D 4052	not specified	0.76	0.76−0.79	0.76−0.80	0.74−0.77	0.76−0.82
distillation, °C/ASTM D 86						
10% evaporated, max	65.00	54.70	53.30−**75.20**	55.80−**73.60**	52.70−64.80	52.40−**70.30**
50% evaporated, max	80.00	72.60	72.90−**154.00**	72.60−**159.00**	68.40−75.50	71.80−76.50
90% evaporated, max	190.00	172.10	173.00−185.30	174.30−**215.90**	71.50−124.70	129.00−171.50
final boiling point, max	220.00	208.90	198.20−209.40	212.80−238.80	194.60−210.90	189.80−208.30
residue, % v/v, max	2.00	1.00	0.90−1.50	1.00−1.60	0.70−1.30	0.90−1.60

The objective of this research is to investigate the use of IR spectroscopy data and multivariate techniques (PCA-LDA) to determine if gasoline adulteration has occurred and the type of solvent used. We believe that this analytical method can be applied to other countries aiming to solve fuel quality problems.

2. Experimental Section

2.1. Samples Used in the Model Construction and Validation. Samples of commercial gasoline were collected between October 2003 and May 2004 at gas stations in Minas Gerais state, in Brazil. Theses samples were submitted to solvent tracer test. The samples were doped with one of the four solvents (heavy naphtha, kerosene,

light naphtha, and thinner) within the range of 1−50% v/v with concentration variation of 1% v/v (200 samples). Seventeen unadulterated gasoline samples purchased from gas stations were also submitted to a solvent tracer test to confirm their good quality and were used in the construction of the model, making up 217 samples in all (200 samples adulterated and 17 samples unadulterated). The samples used in the model construction and validation were divided into two sets: modeling set, approximately 70% of the samples, and validation set, approximately 30% of the samples.

Table 1 gives the average physicochemical properties of unadulterated Brazilian gasoline and of a range of gasoline adulterated with the solvents studied.[17]

2.2. Samples Used in Model Testing. Other gasoline samples were collected at gas stations in Minas Gerais state and tested five months after the construction of the model. This procedure was carried out to evaluate the performance of the model with the possible variations of the fuel (petroleum composition and refine process). At this stage, eight unadulterated gasoline samples (without solvent tracer) were collected and a further 20 gasoline samples were doped with one of the four solvents (heavy naphtha,

(11) Klemenc, S. *Forensic Sci. Int.* **2001**, *115*, 43−52.

(12) Otto, M. *Chemometrics: statistics and computer application in analytical chemistry*; Wiley & Sons: New York, 1999.

(13) Sandercock, P. M. L.; Du Pasquier, E. *Forensic Sci. Int.* **2003**, *134*, 1−10.

(14) Sandercock, P. M. L.; Du Pasquier, E. *Forensic Sci. Int.* **2004**, *140*, 71−77.

(15) *Minitab-14*, Statistical Software for Windows; Minitab, Inc: State College, PA, 2003.

(16) Kim, H. C.; Kim, D.; Bang, S. Y. *Pattern Recognit.* **2003**, *36*, 1095−1105.

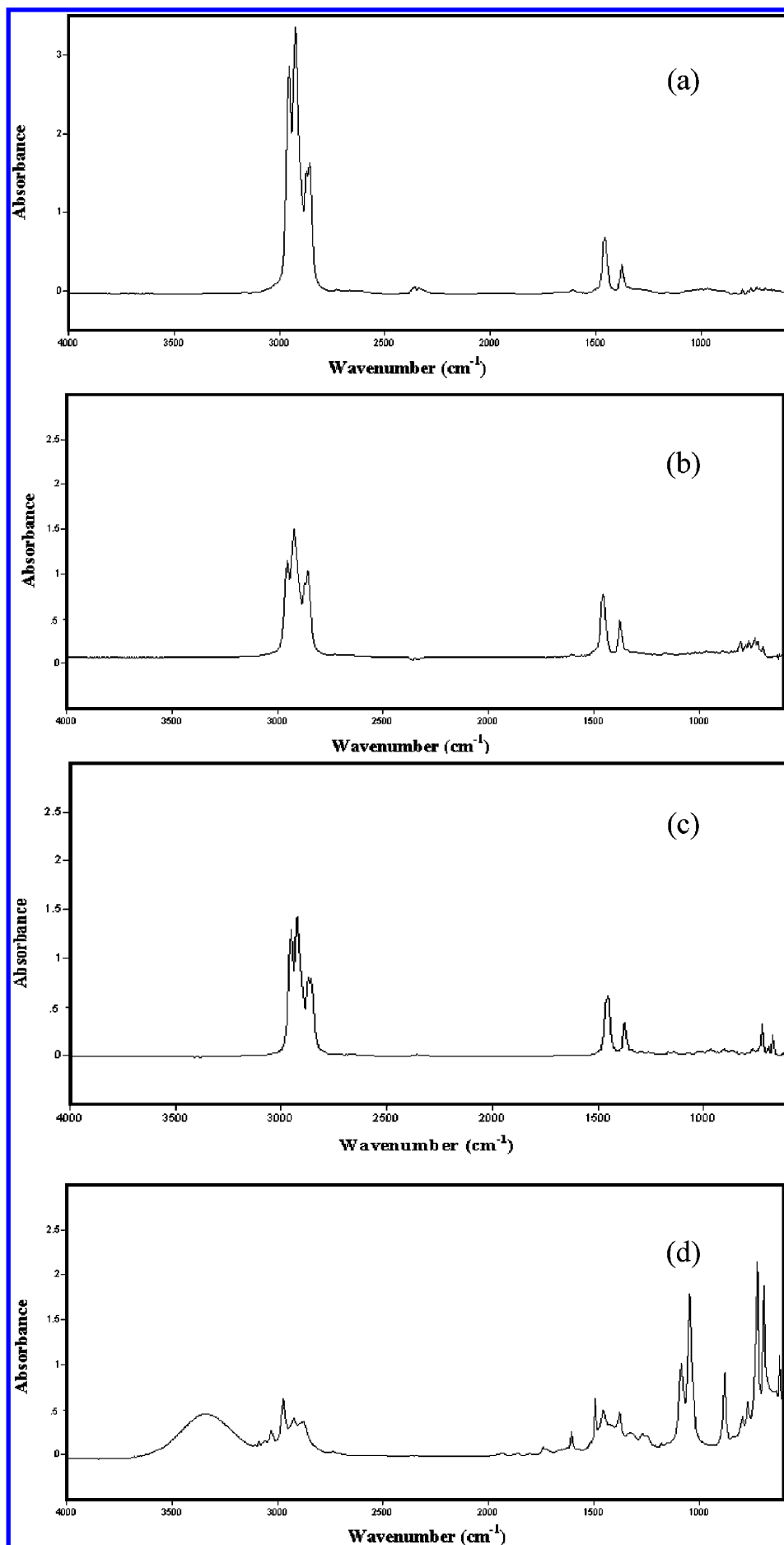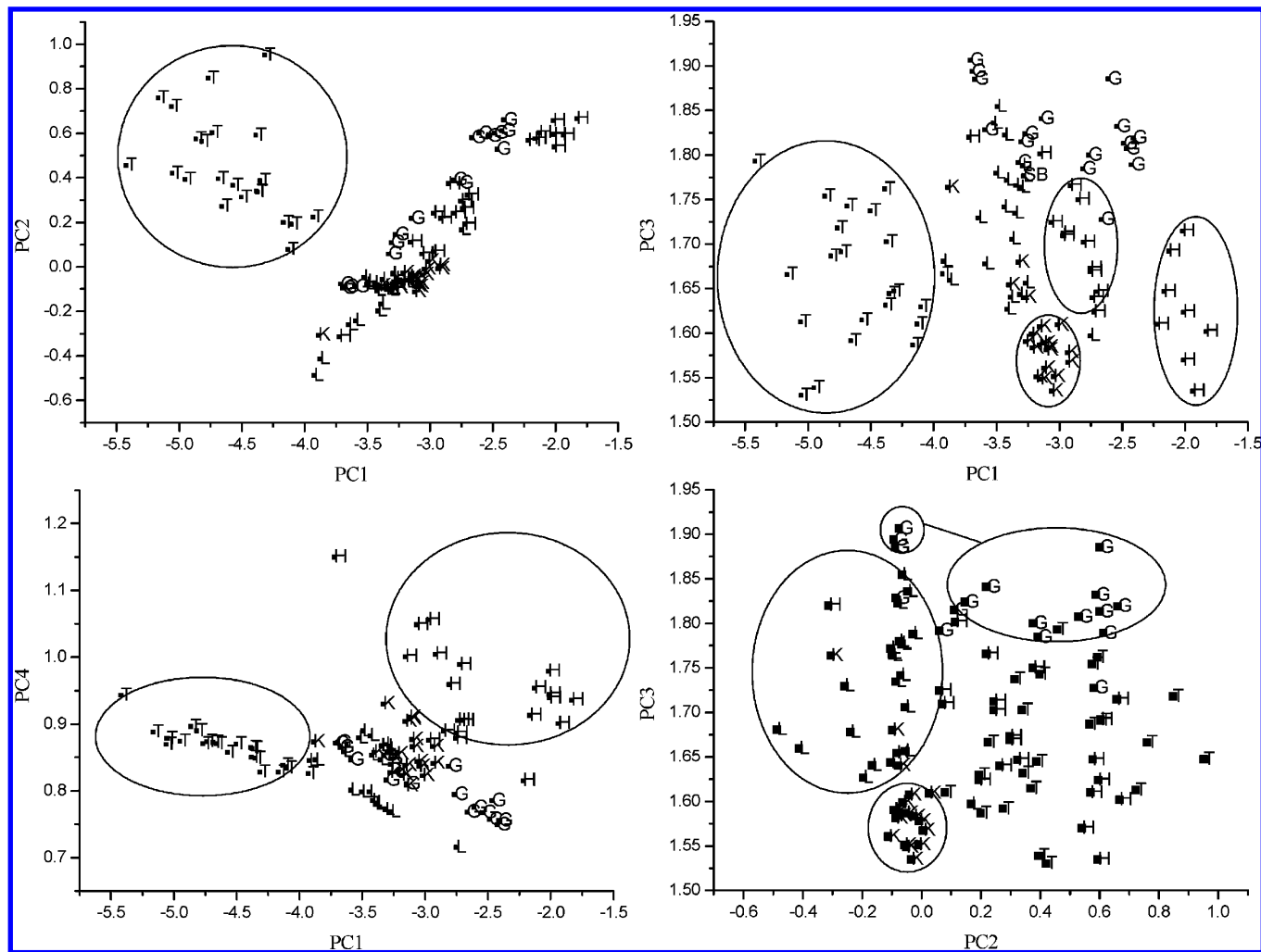(17) Portaria da Agência Nacional do Petróleo, 309 de 27/12/2001. http://www.anp.gov.br.

Figure 2. Infrared absorption spectra of (a) heavy naphtha, (b) kerosene, (c) light naphtha, and (d) thinner.

Figure 3. Separation of the adulterated gasoline samples with heavy naphta (H), kerosene (K), light naphta (L), thinner (T), and unadulterated gasoline samples, in the first four principal components.

kerosene, light naphta, and thinner) at five different concentrations (8, 18, 28, 32, and 42% v/v).

2.3. ATR-FTIR Analysis. Spectra were obtained by ATR-FTIR. An ABB Bomen IR spectrometer model MB 102 equipped with a deuterated triglycerin sulfate detector (DGTS) was used. Spectra were read in the $4000-600$ cm^{-1} range with 16 scans, resolution 4 cm^{-1}, zinc selenete crystal, and incidence angle of 45°.

2.4. GC/MS Analysis. Unadulterated gasoline and solvents were analyzed with a gas chromatographer coupled to a mass spectrometer (GC/MS) model GC/MS-QP5050 using a PONA column (50 m × 0.20 nm; df = 0.20 μm) with dimethylsiloxane stationary phase. Helium was used as carrier gas, flow of 0.40 mL/min. Sample aliquots of 0.5 μL were injected in the split mode (1:24). The column was kept at 34 °C for 5 min, and then the temperature was raised to 60 °C at 2 °C/min, and then to 185 °C at 3 °C/min. Finally, the oven was heated to 250 °C at a rate of 10 °C/min and kept at 250 °C for 10 min. The mass spectrometer was operated at 70 eV. The ionization source was kept at 250 °C, and the quadrupole at 100 °C.

2.5. Software. Data were treated with two computer programs. The program GRAMS 32/AI (Galactic Industries, version 6.01) supplied by the spectrometer ABB Bomen manufacturer was used to collect and treat spectra, and MINITAB (version 14 for Windows) was used in statistical treatment.

3. Results and Discussion

3.1. Model Construction and Validation. Table 1 showed that many gasoline samples, even adulterated, corresponded to

the ANP specification. Non-corresponding properties are presented in bold. The solvent kerosene presented the highest deviation of the values established by law, for contents higher than 5% v/v whereas light naphta presented the lowest deviation. It was observed that many adulterated samples (up to 30% v/v) were within ANP specification.

Figure 1 presents a typical infrared spectrum of adulterated and unadulterated gasoline. It can be seen that they are coincident, and it is not possible to distinguish by visual inspection whether gasoline is adulterated or not, except for adulterated samples with thinner. Figure 2 shows spectra of the solvents studied that may be used in the adulteration of automotive gasoline. It can be observed that only solvent thinner presents a profile distinct from those of the other solvents. For this solvent, the band at 3400 cm^{-1} is assigned to the axial deformation of alcohol OH;[18] the bands at 1600, 1490, and 1450 cm^{-1} are due to the axial deformation of the C=C aromatic ring; the CH$_3$ angular deformation at 1375 cm^{-1}; the alcohol C−O angular deformation occurs at 1079 and 1048 cm^{-1}; and the angular deformation out of the ring C−H bond plane is observed at 874, 719, and 688 cm^{-1}.[18] This indicates that samples adulterated with solvent thinner can be easily identified. However, gasoline adulterated with solvents heavy naphtha, kerosene, or light naphtha could not be discriminated with ease

---

(18) Silverstein, R. M.; Webster, F. X. *Spectrometric Identification of Organic Compounds*, 6th ed.; Wiley: New York, 1998.
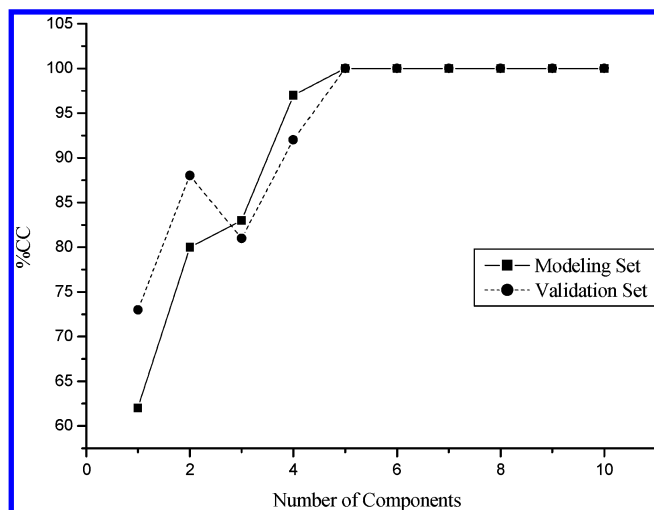
Figure 4. Percentage of correctly classified of LDA as a function of the number of principal components for the modeling and validation sets.

Table 2. Prediction of Adulterated Gasoline Samples with Heavy Naphtha (1−5), Kerosene (6−10), Light Naphtha (11−15), Thinner (16−20), and Unadulterated Gasoline G (21−28)

sample	adulterant	content (%v/v)	result
1		8	correct
2		18	correct
3	heavy naphtha	28	correct
4		**32**	**incorrect**
5		42	correct
6		8	correct
7		18	correct
8	kerosene	28	correct
9		32	correct
10		42	correct
11		**8**	**incorrect**
12		18	correct
13	light naphtha	28	correct
14		32	correct
15		42	correct
16		8	correct
17		18	correct
18	thinner	28	correct
19		32	correct
20		42	correct
21		−	correct
22		−	correct
23		−	correct
24		−	correct
25	none[a]	−	correct
26		−	correct
27		−	correct
28		−	correct

[a] G = unadulterated gasoline.

using the infrared technique due to their chemical similarity, therefore lending itself to the use of multivariate techniques to discriminate the sample.

Spectra of 200 gasoline samples adulterated with solvents and of 17 unadulterated gasoline samples were obtained. Data of the 217 spectra were converted into an X data matrix with the samples in the rows and absorbance values in the columns. The preprocessing used to treat the X data matrix was mean-centered. Analyzing the data by PCA, we observed that the samples with solvent concentrations lower than 8% v/v clustered randomly. These samples were withdrawn from the data matrix. The method did not present sensitivities for concentrations lower than 8% v/v. To make visualization of the results easier, only unadulterated samples and adulterated samples with one of the four solvents (8 to 30%v/v) were used. Figure 3 showed that
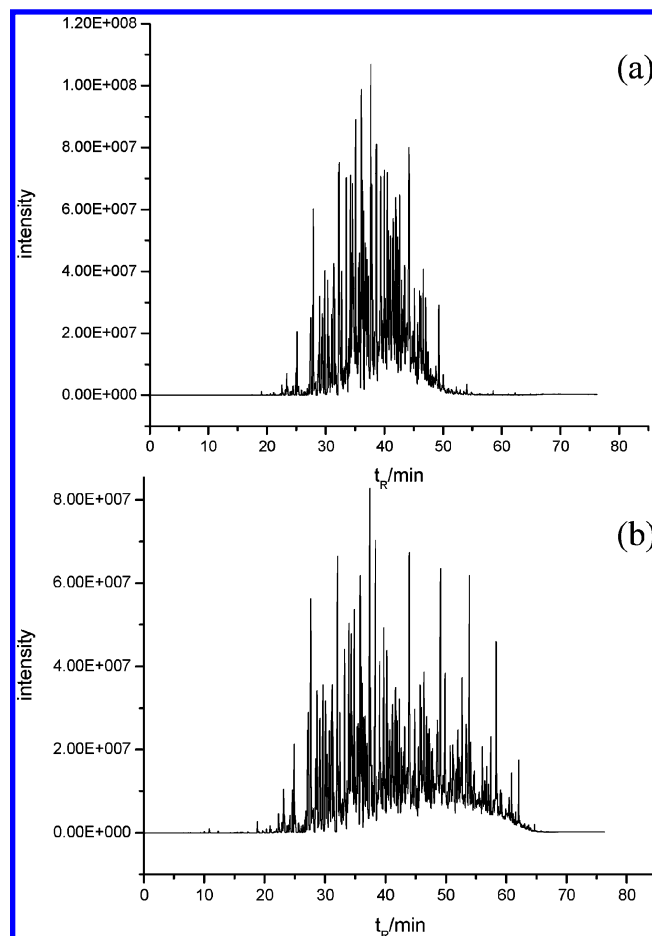

Figure 5. (a) Chromatogram of heavy naphtha. (b) Chromatogram of kerosene.

the first four principal components (which explain 99.4% of the total variance) allow the separation of the samples adulterated with thinner (T), heavy naphtha (H), kerosene (K), and light naphtha (L) of the unadulterated samples (G).

Once the modeling set matrix (70% of the samples) presented a number of variables (579) larger than the number of samples per class (∼40), covariance PCA was performed to reduce the dimension of the original data and visualize the most important information with a smaller number of principal components. Next, the validation set matrix (30% of the samples) was rearranged into this new axis system (principal components) to generate a score matrix for its use in the model validation.

LDA revealed that when five components were used (which explain 99.7% of the total variance), classification of both modeling set and validation set was 100% correct, as shown in Figure 4.

3.2. Model Evaluation. The model test, five months after its construction with a new set of 20 laboratory-adulterated samples and eight unadulterated gasoline samples, showed that samples 4 and 11 were classified incorrectly, as shown in Table 2. Sample 4 was adulterated with heavy naphtha (concentrations of 32% v/v), but it was classified as adulterated with kerosene. The incorrect classification occurred because heavy naphtha and kerosene have similar chemical compositions. Figure 5 presents the chromatograms of heavy naphtha and kerosene. Approximately 70% of the peaks present in the chromatogram of heavy naphtha are also present in that of kerosene. This shows that heavy naphtha is a significant fraction of kerosene, which makes their unequivocal identification difficult.
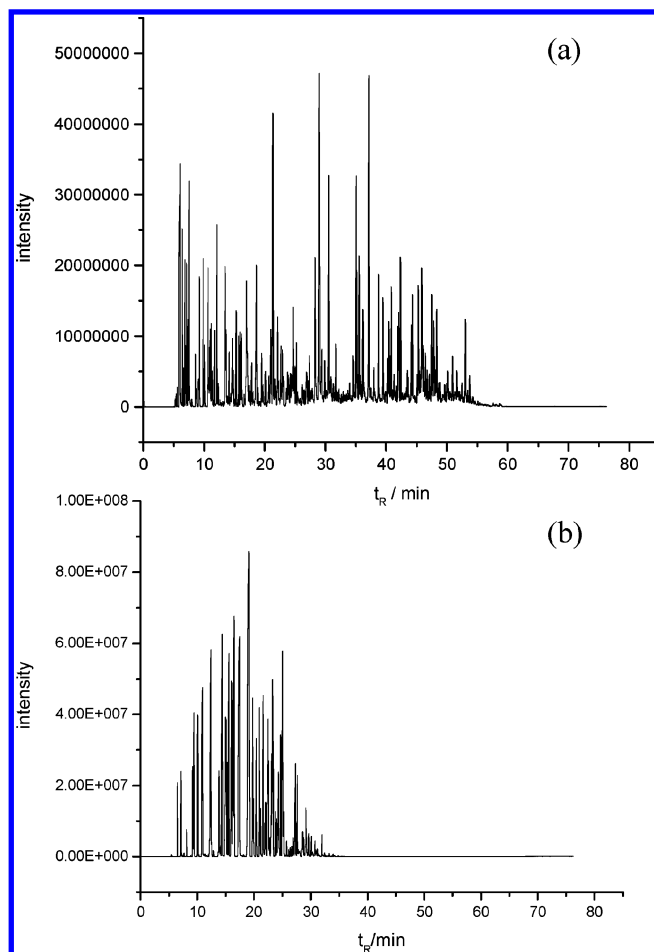
Figure 6. (a) Chromatogram of an unadulterated gasoline sample. (b) Chromatogram of light naphtha.

Sample 11, doped with light naphtha at a concentration of 8% v/v, was classified as an unadulterated gasoline sample (G). This incorrect classification occurred because most components of light naphtha (82%) are also gasoline components, which can be confirmed by comparison of their chromatograms (Figure 6). The fact that light naphtha is practically a gasoline fraction can explain why the sample adulterated with a small amount of this solvent was classified as an unadulterated sample. All unadulterated gasoline samples (G) were correctly classified (Table 2).

The model had an efficiency of 96.4% in the classification of adulterated and unadulterated gasoline samples and approximately 93% in the identification of the solvents (misclassification of 2 samples out of 28).

4. Conclusion

This study show that the methodology developed, the use of infrared absorption spectroscopy associated with PCA-LDA, is valuable in the quality control of automotive gasoline, particularly due to its speed, low cost, easy access, and efficiency. The problem of fuel adulteration is so complex due to the similarity between doped agents and gasoline because both are petrochemical fractions.

It was also shown that it is possible to discriminate samples of unadulterated gasoline and those adulterated with diverse petrochemical solvents, many with compositions similar to that of gasoline, up to 8% (v/v) of contaminant. This method is a fast and inexpensive alternative to using tracer assays, a method that can detect adulterated gasoline with 100% accuracy but at a high cost and using technology that is not readily available. The method outlined here achieved 96% efficiency in the classification of adulterated and unadulterated gasoline and 93% efficiency in the identification of the type of solvent added to gasoline.

It can be concluded that, using the method described here, it is possible to identify whether gasoline has been adulterated or not and the type of adulterant used. This information is relevant in inspections and police investigations to map adulteration origin.

EF050203E