Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline

Leonardo S.G. Teixeira [a,*], Fábio S. Oliveira [b,c], Hilda C. dos Santos [d], Paulo W.L. Cordeiro [a], Selmo Q. Almeida [a]

[a] *Universidade Salvador, UNIFACS, Departamento de Engenharia e Arquitetura, Av. Cardeal da Silva 132, 40220-141 Salvador, Bahia, Brazil*
[b] *Instituto de Química, Universidade Federal da Bahia, Campus Universitário de Ondina, 40170-280 Salvador, Bahia, Brazil*
[c] *Universidade Federal do Recôncavo da Bahia, Centro de Ciências da Saúde, Campo do Governo, 44574-490 Santo Antônio de Jesus, Bahia, Brazil*
[d] *Serviço Nacional de Aprendizagem Industrial, Centro de Tecnologia Industrial Pedro Ribeiro, Av. Luiz Tarquínio Pontes 938, 42700-000 Lauro de Freitas, Bahia, Brazil*

Abstract

In the present work, Fourier transform infrared spectroscopy (FTIR) in association with multivariate chemometrics classification techniques was employed to identify gasoline samples adulterated with diesel oil, kerosene, turpentine spirit or thinner. Results indicated that partial least squares (PLS) models based on infrared spectra were proven suitable as practical analytical methods for predicting adulterant content in gasoline in the volume fraction range from 0% to 50%. The results obtained by PLS provided prediction errors lower than 2% (v/v) for all adulterant determined. Additionally, Soft Independent Modeling of Class Analogy (SIMCA) was performed using all spectral data (650–3700 cm^{-1}) for sample classification into adulterant classes defined by training set and the results indicated that undoubted adulteration detection was possible but identification of the adulterant was subject to misclassification errors, specially for kerosene and turpentine adulterated samples, and must be carefully examined. Quality control and police laboratories for gasoline analysis should employ the proposed methods for rapid screening analysis for qualitative monitoring purposes.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Gasoline adulteration; SIMCA model; PLS analysis

1. Introduction

Fuel, according to its type and application, requires a minimum quality so that its energy can be used in the best way possible. In Brazil, this minimum quality is established through law enforceable technical specifications, aimed at insuring that all fuels complies with a minimum quality standard. However, before consumption, changes may be introduced to the fuel through inadequate transport, handling and storage or outright adulteration, by the unlawful addition of other substances, such as solvents, kerosene or diesel. The addition of these substances to automotive gasoline is a problem that has been worrying Brazilian authorities and has prompted action from Agência Nacional do Petróleo (ANP), Brazil's petroleum regulatory agency. Fuel adulteration may lead to engine malfunction, rubber corrosion and environmental hazards, besides tax evasion, since industrial solvents and fuels have different taxations. Therefore, monitoring fuel quality is of paramount importance in the fight against such irregularities, aiming at consumer protection and the establishment of healthy market conditions.

Brazilian gasoline is usually adulterated by adding diesel oil, kerosene and petrochemical solvents such as turpentine spirit and thinner due to the large differences between the prices of gasoline and these solvents. Moreover, the easy

---

availability of these chemicals and the fact that they are miscible in gasoline make the unhealthy and unethical practice of adulteration of gasoline a very attractive proposition [1].

Gasoline adulteration detection and identification are not ordinary tasks, since solvents usually employed for this purpose are mixtures of hundreds of different compounds and some of them are found even in unadulterated gasoline [2,3]. In this context, efforts to detect and identify adulteration in a traditional fashion could demand the determination of hundreds of organic compounds in gasoline samples and further evaluation of composition profile in comparison with typical unadulterated gasolines in order to identify composition variation that could be related to adulteration. Thus, a fast and reliable method to screen large numbers of gasoline samples is required to classify and determine the adulterant amount added in the fuel.

The combination of infrared spectrometry and chemometric techniques provides a powerful tool for monitoring a variety of processes and, as such, is arousing increasing interest for quality control purposes. The fact that samples require virtually no treatment and the expeditiousness with which spectra can be obtained have made infrared spectrometry a preferred choice over traditional techniques involving lengthy sample conditioning procedures [4]. Chemometrics associated with infrared spectroscopy have been used to obtain accurate and simple calibration models to dictate chemical composition and physical properties of different samples [4]. In this way, these appealing features of the infrared spectrometry in association with multivariate chemometrics techniques have fostered its use in a variety of fields including food [5], pharmaceutical [6], petrochemical industries and also have been used to monitor quality of fuel [1,7–10].

In the present work, Fourier transform infrared spectroscopy (FTIR) in association with multivariate chemometrics classification techniques [1] was employed to identify gasoline samples adulterated with diesel oil, kerosene, turpentine spirit or thinner. Additionally, partial least squares (PLS) [11,12] multivariate calibration models were constructed in order to determine the adulterant content in gasoline samples which adulteration was previously detected by chemometrics classification model.

2. Experimental

2.1. Apparatus and methods

An IROX 2000 spectrometer (Grabner Instruments, Austria) was employed to obtain the FTIR spectra in the 3700–650 cm^{-1} range at a nominal resolution of 4 cm^{-1} and at 23 ± 1 °C. Final spectra were the average of 31 scans in order to improve the signal to noise ratio. To avoid cell manipulation and to increase the sample throughput, sample introduction in FTIR spectrophotometer was done using the internal equipment pump. The instrument performed automatically with 12 ml of sample. The samples were drawn in directly from the sample container through flexible tubing. To avoid out gassing, the samples were drawn carefully into an internal chamber and then pressured through the absorption cell equipped with two Zn/Se windows forming the absorption volume with a mean path length of 23 μm. After the measurement, the samples were transferred into the connected disposal container.

Distillation test for 10%, 50% and 90% evaporated as well as final boiling point was performed in an automatic distillation unit AD 865G (ISL, France), in accordance with international standard methods (ASTM D 86) [13].

The tests for determination of motor octane number (MON) and antiknock index [(RON + MON)/2] were carried out in a CFR engine (Waukesha, USA), in accordance with ASTM D2699 and D2700 [14].

Ethanol content, in the gasoline samples, was determined by the aqueous extraction method, by using 50 ml of sample and 50 ml of 10% (m/v) of NaCl aqueous solution in compliance with Brazilian NBR 13992 [15].

All the spectra of gasoline samples were exported to Unscrambler 8.0 from CAMO (A/S) and this software was employed for multivariate analysis calculations. The variables were mean centered before modeling procedures.

2.2. Samples

A group of 132 samples was prepared by adding gasoline (deriving from oil refineries), anhydrous ethanol and diesel oil, kerosene, turpentine spirit or thinner solvent. With the purpose of including variety in the gasoline composition and to construct synthetic mixtures in order to span slight variations in gasoline composition that could not be related with adulterant levels, each blend was prepared in triplicate using three different gasoline samples obtained from service stations in metropolitan Salvador, Brazil. Samples (50 ml) were prepared in volumetric flasks by mixing the gasoline with anhydrous ethanol and each adulterant. The experimental design is shown in Table 1. As can be seen in Table 1, ethanol was added to all synthetic adulterated samples in order to attain a final level of 25% (v/v) of anhydrous ethanol because in Brazil ethanol is being mixed in gasoline (gasohol, mixture of 75% gasoline and 25% anhydrous ethanol) to act as an oxygenated additive [16,17]. The FTIR spectra of all gasoline samples were acquired as previously described.

The mixture design produced 30 samples adulterated with known amounts of each of the studied solvent adulterant and three unadulterated gasoline blank samples. The synthetic adulterated gasoline sample set was randomly split into 27 samples for training set and 6 samples for validation set in order to construct multivariate models for analysis of each selected adulterant.

Gasoline samples were randomly collected from different gas stations in Salvador, Bahia State, Brazil. These samples were previously analyzed by the ANP regulated physical–chemical parameters and seven problematic samples were selected for this work. The results of physical–chemical analysis are shown in Table 2.

Table 1
Samples composition prepared by mixing gasoline, anhydrous ethanol and adulterant (diesel oil, kerosene, turpentine spirit or thinner solvent)

Solution	Adulterant (%)	Adulterant volume (ml)	Gasoline volume (ml)	Added ethanol (ml)	Total (ml)
1	0	0.0	50.0	0.0	50.0
2	3	1.5	48.0	0.5	50.0
3	10	5.0	43.3	1.7	50.0
4	15	7.5	40.0	2.5	50.0
5	20	10.0	36.7	3.3	50.0
6	25	12.5	33.3	4.2	50.0
7	30	15.0	30.0	5.0	50.0
8	35	17.5	26.7	5.8	50.0
9	40	20.0	23.3	6.7	50.0
10	45	22.5	20.0	7.5	50.0
11	50	25.0	16.7	8.3	50.0

2.3. Chemometrics analysis

In this work, the employment of Soft Independent Modeling of Class Analogy (SIMCA) [1,12,18–20] for gasoline adulterant identification based on FTIR spectral data was proposed. The SIMCA is a well known multivariate supervised pattern recognition method based on Principal Component Analysis (PCA) [12]. SIMCA is a multivariate method based on PCA decomposition of all spectral data in few relevant principal components that explain most of the variance of spectral data. Most of the information contained in spectral data is compressed by PCA in few PCs and the boundaries of each class are defined in order to classify unknown samples. Mathematical details could be found elsewhere [1,18,20].

The association of SIMCA with FTIR spectral data can determine if an unknown gasoline sample belongs to one of the evaluated classes of solvent adulterated gasoline or none of them. The delimitation boundaries of PCA space that contain each kind of solvent are the base of SIMCA multivariate classification. The PCA region characteristic of each solvent adulterant is determined using a training sample set previously known to be adulterated with this solvent.

In this work, SIMCA was employed to identify if an unknown gasoline sample was adulterated by kerosene, diesel, thinner, turpentine spirits or none of them. Training sets containing 30 gasoline samples previously adulterated with known amounts of selected solvents (Table 1) were employed to construct SIMCA model. These training sets provide spectral information that was used to determine PCA boundaries that characterize samples adulterated by diesel, kerosene, turpentine spirit or thinner. Evaluation of SIMCA classification performance was performed by leave one out complete cross-validation [13].

Once the adulterant was identified by SIMCA using FTIR spectral data, the levels of solvent could be determined using a multivariate calibration method that relates solvent content with FTIR spectra [11,12]. Partial least squares (PLS) [11,12,20,21] regression is a powerful model for the analysis of mixtures and it is the current means of choice when prediction is the main objective. In PLS regression, the objective is to assess the degree of relationship between a set of X predictor variables (FTIR spectra) and a set of Y outcome variables (adulterant level). The PLS model attempts to derive from each source the information which is relevant to the establishment of a relationship between the two blocks of variables (X block and Y block) [21,22]. In the present work, the concentrations of diesel oil, kerosene, turpentine spirit or thinner solvent, in blends with gasohol, were determined using PLS. Individual PLS 1 models were constructed in order to predict the levels of each evaluated adulterant using 27 samples in the training set and six samples in the validation set with known amounts of adulterant, resulting in four PLS 1 models. The number of latent variables needed to construct each PLS 1 model was determined by full cross-validation. All PLS calculations were performed using Unscrambler 8.0 (CAMO) chemometrics package [20].

The proposed strategy for adulterant identification using SIMCA and further quantification by PLS are depicted in Fig. 1.

3. Results and discussion

3.1. FTIR spectra

Changes in the physical and chemical properties of gasoline are caused by solvent addition. However, visual

Table 2
Physic-chemical analysis of problematic gasoline samples collected from gas stations in Salvador, Brazil

| Parameter	Gasoline sample no.							Official specification
	I	II	III	IV	V	VI	VII	
Ethanol (%, v/v)	24	25	26	26	25	25	25	25 ± 1
Distillation temperatures (°C)								
10% evaporated	54.3	53.4	59.3	57.9	53.9	54.6	53.4	65, max
50% evaporated	72.2	73.0	72.2	71.9	71.9	73.8	73.4	80, max
90% evaporated	168.1	166.2	178.1	181.2	165.0	168.7	164.5	190, max
Final boiling point	249.1	234.7	231.8	247.8	226.1	239.7	244.7	220, max
Residue (%, v/v)	0.8	2.3	1.5	1.5	2.0	1.6	1.4	2.0, max
MON (min)	84.2	83.6	78.0	77.8	83.6	84.5	84.4	82.0, min
Antiknock index (min)	89.6	90.5	83.3	83.0	90.5	90.2	90.2	87.0, min
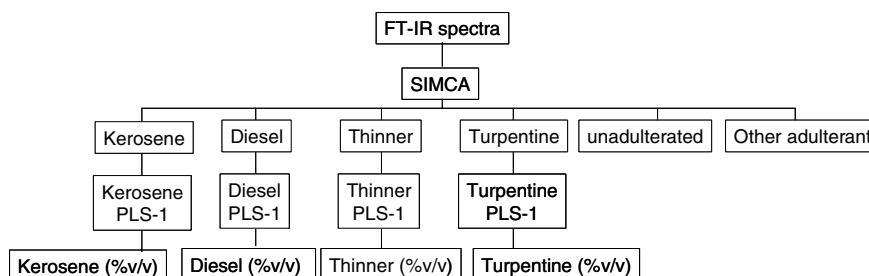
Fig. 1. Proposed strategy for adulterant identification and quantification of gasoline adulterants with multivariate data analysis based of FTIR spectra.

inspection of sample infrared spectrum was not efficient in discriminating corrupted gasoline samples. The spectra corresponding to gasoline samples are shown in Fig. 2. As can be seen, differences between the spectra are not easy to perceive with a visual inspection, so multivariate statistical treatment is essential.

3.2. Adulterant identification by SIMCA

Chemometrics classification methods, such as SIMCA, associated with information provided by FTIR spectral data can be an alternative to identify corrupted gasoline samples. The SIMCA method was chosen due to the possibility of sample classification into one of the classes defined by the training set, more than one classes or even none of them. In this work, SIMCA analysis was performed using all spectral data (650–3700 cm^{-1}) for sample classification into adulterant classes defined by training set.

In order to carry out SIMCA method, 30 samples representing one of the evaluated adulterants were used to compose the training set and seven samples collected from different gas stations and suspected of adulteration were used in the test set.

In SIMCA classification, one distinct PCA model was built for each adulterant group (kerosene, diesel, thinner and turpentine spirits) using all spectral data (650–

Table 3
Performance of PCA models by SIMCA for identification of adulterant in gasoline samples

PCA model	No. PCs	Explained variance (%)
Diesel	2	99.3
Kerosene	2	99.1
Turpentine	2	99.4
Thinner	3	99.5

Number of retained principal components (no. PCs) and percentile of explained variance in the spectral data.

700 cm^{-1}). Baseline adjustment was not carried out in this work. First derivate, baseline adjustment and multiple scatter correction preprocessing techniques were tested in preliminary studies and better results were obtained by using crude spectral data after mean centering. The parameters that describe the performance of each of these PCA models are listed in Table 3. The results obtained by full cross-validation for SIMCA classification can be visualized by sample to model distance *versus* leverage graph (Fig. 3) [1,18,20]. In this figure, the data points contained into the lower left quadrant were pointed by SIMCA to belong to the adulterant class in evaluation at 95% confidence level. This figure shows fair performance of SIMCA associated with FTIR spectral data to identify the solvent employed for adulteration of gasoline samples, since some errors in classification were located.

The SIMCA errors in classification could be of two types: type I (object not included in its own class) and type II (object included in a wrong class). Table 4 summarizes the error in classification pointed out by full cross-validation of SIMCA. As can be seen in Table 4, while type I error (object not included in its own class) was not observed, a significant number of type II (sample classified in the wrong class) was obtained in the analysis of kerosene and turpentine spirits adulterated gasoline samples.

A close examination of SIMCA results showed a complete overlap between kerosene and turpentine adulterated gasoline samples in the PCA space and distinction of these solvents could not be attained by SIMCA associated with FTIR spectral data. This behavior can be attributed to a very close similarity in the composition of kerosene and turpentine adulterated gasoline samples and consequent
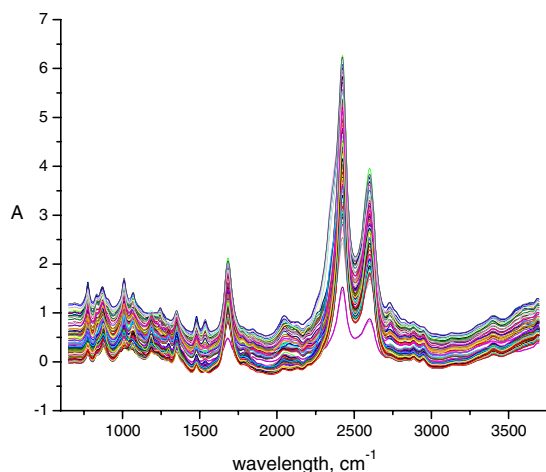

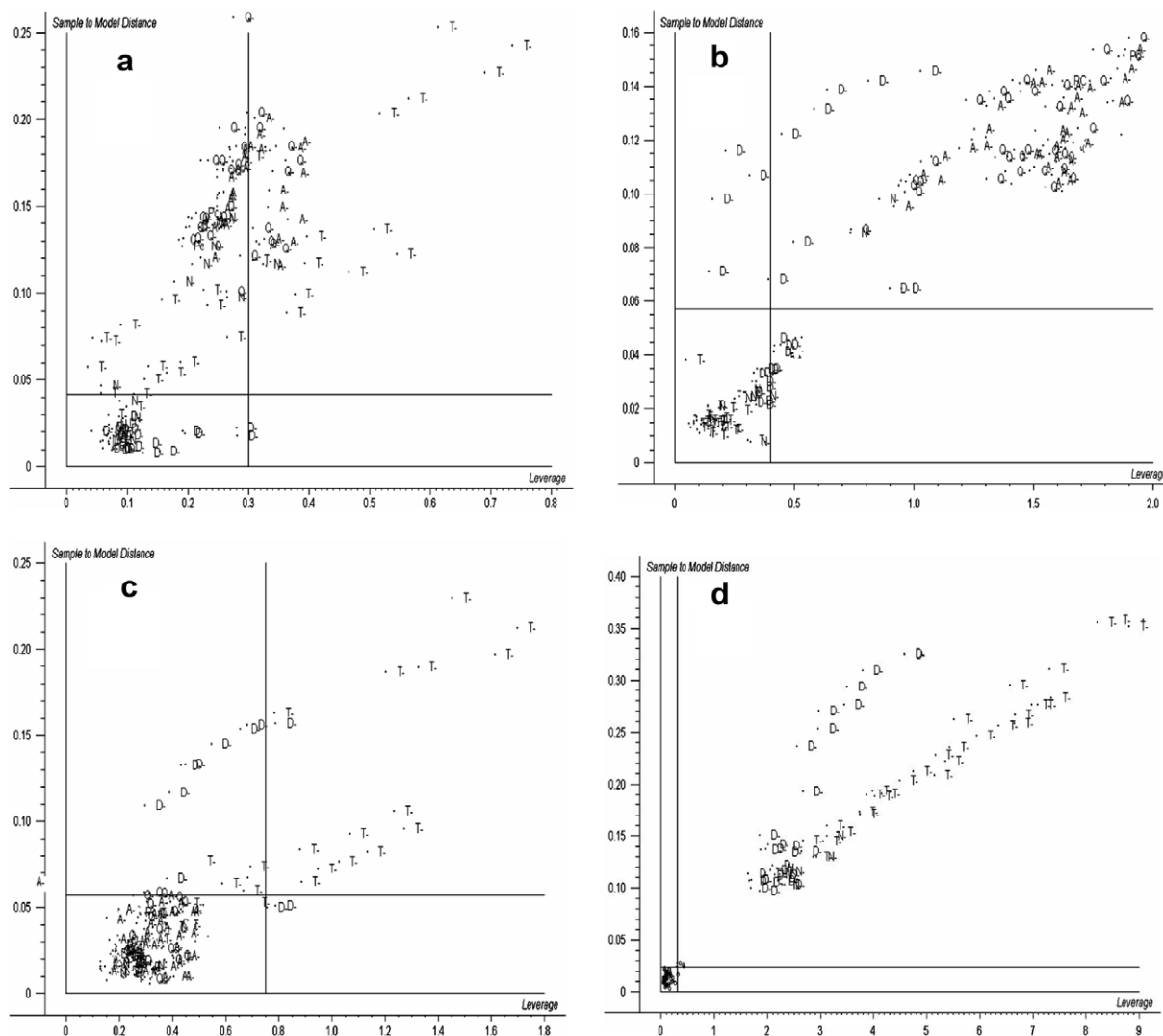Fig. 2. Fourier transform infrared spectra of gasoline samples.

Fig. 3. Sample to models distance *versus* leverage graph plot obtained by SIMCA to classify adulteration by (a) diesel, (b) thinner, (c) turpentine spirit and (d) kerosene.

Table 4
Percentile errors in classification obtained by SIMCA

Adulterant	Error (%)	
	Type I	Type II
Diesel	0	2
Kerosene	0	33
Turpentine spirits	0	33
Thinner	0	20

of highly overlapped FTIR spectra in the selected wavelength range. In this way, kerosene adulterated samples were correctly classified in its own class and misclassified as turpentine adulterate while turpentine adulterated samples were correctly classified in its own class but misclassified as kerosene by SIMCA.

Type II error occurrences were related to few gasoline samples adulterated with low levels of diesel that were cor-

rectly pointed out in its own class but misclassified by SIMCA as adulterated with thinner.

The results obtained by SIMCA using FTIR spectra indicated that undoubted adulteration detection was possible but identification of the adulterant was subject to misclassification errors (type II), specially for kerosene and turpentine adulterated samples, and must be carefully examined.

Application of SIMCA to a test set containing seven gasoline samples with anomalous physical chemistry proprieties (Table 2) resulted in three of these samples that were classified as adulterated by kerosene. This fact reinforced the supposition of heavy solvent addition indicated by distillation test presented in Table 2. All other gasoline samples of this test set were not classified as adulterated by solvents in evaluation or as unadulterated. This fact suggests the possibility of contamination or intentional addition of a different solvent from those ones studied in this work.

3.3. PLS models

In order to predict adulterant content in blends with gasohol, multivariate calibration models were built, using different pre processing of the spectra data, by partial least squares regression (PLS 1). Since the identification of an adulterant was pointed out by SIMCA, the level of this adulterant was predicted by a PLS 1 multivariate calibration model constructed for prediction of this solvent content in gasoline. The performance of four PLS 1 models constructed to determine the adulterant levels in gasoline samples was evaluated using six samples of validation with previously known content for each adulterant. Root mean square error of prediction (RMSEP), root mean square error of the calibration (RMSEC), number of retained latent variables (no. LVs) and percentile of explained variance are listed in Table 5. The results obtained by PLS provided prediction errors lower than 2% (v/v) for all adulterant determined and high dimensionality reduction using few latent variables.

Validation of PLS 1 models was carried out using an independent validation set of six samples of blends of gasoline with a specific adulterant in evaluation. The results obtained by PLS 1 models are illustrated in Fig. 4, showing a good agreement between predicted and real levels of adulterants in gasoline for calibration and validation data sets. The slopes and intercepts of curves depicted in

Table 5
Performance of PLS 1 models for determination of adulterant levels in gasoline samples

PLS model	No. LVs	RMSEP (%, v/v)	RMSEC (%, v/v)	Explained variance (%)
Diesel	5	1.8	1.6	98.4
Kerosene	4	1.9	1.7	98.3
Turpentine	3	1.8	1.4	93.9
Thinner	2	1.9	1.7	96.9

Number of retained latent variables (no. LVs), root mean square error for prediction (RMSEP), root mean square error of the calibration (RMSEC) and percentile of explained variance of prediction.

Fig. 4 are close to unity and zero, respectively, indicating low bias and systematic regression errors.

The classification error could be overcome if PLS 1 resulted in the prediction of non-significant levels of adulterant that were not added to the gasoline samples and this adulteration was wrongly pointed out by SIMCA. In this context, the application of PLS 1 models for determination of adulterant previously identified by SIMCA was carried out. The PLS regression resulted that diesel and thinner predicted levels were close to the real ones, even for samples misclassified as containing this solvent.

On the other hand, even PLS 1 multivariate calibration model was not able to distinct kerosene and turpentine adulteration. In other words, PLS 1 model built for kerosene
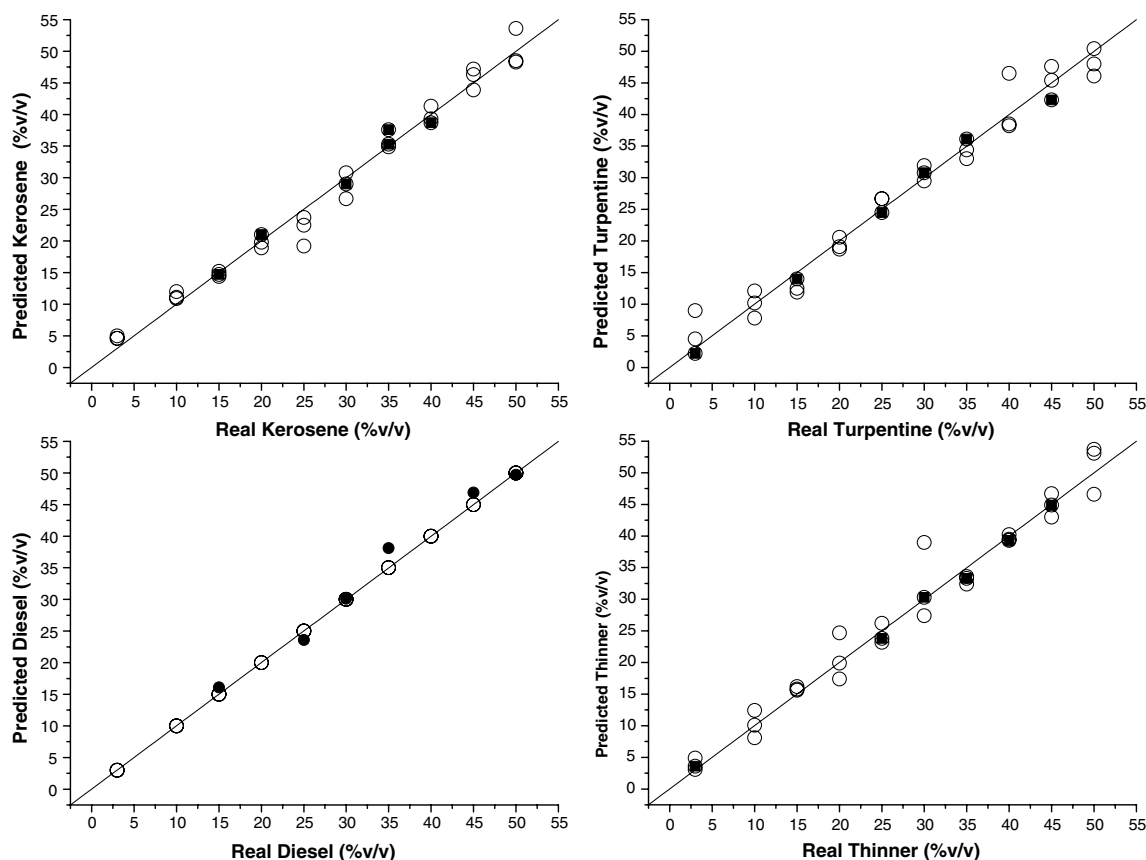

Fig. 4. Predicted *versus* real adulterant contents graph obtained by PLS 1 models for (○) calibration and (●) validation set.

predicted the solvent levels of turpentine adulterated samples with RMSEP = 6% (v/v) while PLS 1 model for turpentine predicted the levels of kerosene adulterated samples with RMSEP = 3% (v/v).

The samples in the test set pointed out as kerosene adulterated by SIMCA were subjected to PLS 1 kerosene model and presented predicted adulterant levels in the range of 11.8–13.4% (v/v), reinforcing evidences of this kind of solvent adulteration.

4. Conclusion

In this work, the combination of infrared spectrometry and chemometric techniques was used for rapid screening analysis to identify gasoline samples adulterated with diesel oil, kerosene, turpentine spirit or thinner. The SIMCA classification model associated with FTIR spectral data was successfully applied for gasoline adulteration detection and identification of added solvent was fair due to kerosene and turpentine complete overlap. In this way, careful examinations of classification results must be taken in order to correctly identify kerosene and turpentine gasoline adulterations.

The partial least squares multivariate model based on FTIR spectra, developed in this work, was proven suitable as a practical analytical method to predict adulterant content in gasohol in the range from 0% to 50% (v/v). Even for misclassified samples, the prediction results presented good agreement with real ones. So, quality control and police laboratories for gasoline analysis should employ the proposed methods for rapid screening analysis for qualitative monitoring purposes.

A drawback of the present work is related with the efforts for synthetic adulterated gasoline samples set preparation, since a representative number of calibration and validation solutions must be used for multivariate models construction. This problem could be circumvented using automated systems for standard adulterated gasoline solutions preparation.

References

[1] Oliveira FS, Teixeira LSG, Araujo MCU, Korn M. Fuel 2004;83:917.
[2] Wiedemann LSM, d'Avila LA, Azevedo DA. Fuel 2005;84:467.
[3] Skrobot VL, Castro EVR, Pereira RCC, Pasa VMD, Fortes ICP. Energy Fuel 2005;19:2350.
[4] Blanco M, Pagès J. Anal Chim Acta 2002;463:295.
[5] Goodacre R, Hammond D, Kell DB. J Anal Appl Pyrolysis 1997;40–41:135.
[6] Yoon WL, Jee RD, Charvill A, Lee G, Moffat AC. J Pharm Biomed Anal 2004;34:933.
[7] Garrigues S, Andrade JM, de la Guardia M, Prada D. Anal Chim Acta 1995;317:95.
[8] Guachardi R, Costa Filho PA, Poppi RJ, Pasquini C. J Near Infrared Spectrosc 1998;6:333.
[9] Faber NM, Duewer DL, Choquette SJ, Green TL, Chesler SN. Anal Chem 1998;70:2972.
[10] Pereira RCC, Skrobot VL, Castro EVR, Fortes ICP, Pasa VMD. Energy Fuel 2006;20:1097.
[11] Oliveira FS, Teixeira LSG, Korn M. Lab Robot Automat 2000;12:305.
[12] Brereton RG. Chemometrics: data analysis for the laboratory and chemical plant. Chichester: Wiley; 2003.
[13] ASTM D86. Standard test method for distillation of petroleum products at atmospheric pressure. Washington, DC: American Society for Testing and Materials; 1987.
[14] ASTM D2700. Standard test method for motor octane number of spark ignition engine fuel. Washington, DC: American Society for Testing and Materials; 1999.
[15] NBR 13992. Gasolina Automotiva: Determinação de Álcool Etílico Anidro Combustivel (AEAC). São Paulo, Brazil: Associação Brasileira de Normas Técnicas; 1997.
[16] Teixeira LSG, Leão ES, Dantas AF, Pinheiro HLC, Costa ACS, de Andrade JB. Talanta 2004;64:711.
[17] Teixeira LSG, Souza JC, dos Santos HC, Pontes LAM, Guimarães PRB, Sobrinho EV, et al. Fuel Process Technol 2007;88:73.
[18] Wold S, Sjostrom M, Kowaslki BR. Chemometrics: theory and application. Washington: American Chemical Society; 1973. p. 243.
[19] De Maesschalck R, Candolfi A, Massart DL, Heuerding S. Chemolab 1999;47:65.
[20] Esbensen KH. Multivariate data analysis in practice: an introduction to multivariate data analysis and experimental designing. Norway: Camo Process AS; 1994.
[21] Garrigues S, Andrade JM, de la Guardia M, Prada D. Anal Chim Acta 1995;317:95.
[22] Pimentel MF, Ribeiro GMGS, da Cruz RS, Stragevitch L, Pacheco Filho JGA, Teixeira LSG. Microchem J 2006;82:201.