

PAPER**CRIMINALISTICS**

Maria Monfreda,¹ M.Sc. and Adolfo Gregori,² M.Sc.

Differentiation of Unevaporated Gasoline Samples According to Their Brands, by SPME–GC–MS and Multivariate Statistical Analysis

ABSTRACT: One of the aims of fire investigations is to identify associations among accelerants according to their source. In this study, 50 gasoline samples—representing five brands—were analyzed using solid-phase microextraction (SPME) and gas chromatography–mass spectrometry (GC–MS). Chemometric procedures, such as principal component analysis (PCA) and discriminant analysis (DA), were applied to a data matrix obtained by the target compound chromatogram method, to discriminate samples according to their brand. PCA was successful in finding a natural grouping of samples according to their brand, suggesting that aromatic compounds were more useful than aliphatics for the purpose of this study. DA, if applied to aromatic compounds, gave both a classification ability and a prediction ability of 100%. The outstanding results obtained by this work provide the basis of a data matrix that could be used in real cases of arson to link a sample of unevaporated gasoline to its brand or refinery.

KEYWORDS: forensic science, gasoline, arson, fire debris, solid-phase microextraction, gas chromatography–mass spectrometry, multivariate statistical analysis, principal component analysis

Arson is a widespread crime, instigated by those not only intent on destroying buildings, their contents, and other property but also as a means of destroying physical evidence. Petroleum-based fuels—such as gasoline, kerosene, and diesel—are often utilized as accelerants as they increase the rate and spread of fire. However, in cases where an inflammable liquid has been used, enough unburnt accelerant may remain absorbed in the fire debris to be able to be detected and then identified.

One of the first problems encountered in forensic investigations is the physical isolation of very small quantities of accelerant from a sample of heterogeneous, charred debris taken from the fire scene. This task is typically performed by using static or dynamic headspace techniques; solid-phase microextraction (SPME) is a technique commonly used in forensic sciences, including ignitable liquid extractions, as presented in literature (1–3). The American Society of Testing and Materials (ASTM) published guidelines for the use of SPME as a preparative technique for fire debris analysis (4), a practice that is especially suitable when a very low concentration of ignitable liquid residues is found in the sample. Unlike other methods of separation and concentration, this method recovers a minimal amount of the ignitable residues contained in the evidence, leaving behind those that are suitable for subsequent resampling (4). In this regard, Sandercock published a review of

fire investigations and ignitable liquid residue analysis, which was essentially a collection of articles published between late 2001 and early 2007 (5).

Another critical area in forensic investigations is the difficulty in detecting traces of petroleum-based fuels in a fire-related sample. Gas chromatography (GC) using a mass spectrometer (MS) is a useful technique to detect traces of accelerants. Pyrolysis of synthetic materials (carpets, fabrics, foam padding, floor tiles) during a fire can produce a complex mixture of volatile compounds that interfere with low-level detection of common accelerants. Moreover, many frequently observed pyrolysis products are hydrocarbons of the type present in petroleum-based accelerants. The target compound (TC) analysis, proposed by Keto and Wineman (6,7) and developed further by Lennard (8), helps to overcome these problems. In their analysis, the total ion chromatogram (TIC) of each sample was processed by a special data analysis program that looked for GC peaks at specific retention times. It then checked the identity of each TC by considering the relative abundance of two or three characteristic m/z ions. The TCs chosen were known to be present in the accelerants under investigation and remained detectable when the accelerant was highly evaporated, diluted, and contaminated with high levels of coeluting substances. Semi-quantitative data were obtained for the TCs and used to construct a “target compound chromatogram” (TCC), that is, a bar graph of base ion peak area versus retention time for each TC identified. Finally, the TCCs were visually (6,7) or automatically (with an Excel macro; [8]) compared with those generated from standard petroleum samples. The ASTM also published guidelines for the identification of residue of ignitable liquids in extracts from fire debris

¹Agenzia delle Dogane, Direzione Regionale per il Lazio e l’Umbria, Sezione Operativa Laboratorio Chimico, via Carucci 71, 00143 Rome, Italy.

²Raggruppamento Carabinieri Investigazioni Scientifiche, Reparto di Parma, via Parco Ducale 3, 43100, Parma, Italy.

Received 29 July 2009; and in revised form 5 Feb. 2010; accepted 6 Feb. 2010.

samples by performing a GC–MS analysis and developing TCCs (9).

The sampling and classification of petroleum-based accelerants, according to their petroleum distillate class, were realized using SPME with analysis by GC–flame ionization detection in conjunction with multivariate data analysis (10). Multivariate pattern recognition techniques were also applied to GC–MS data for the same purpose (11).

In forensic investigations, it is becoming increasingly important to identify associations among accelerants according to their source. Several studies have focused on grouping accelerants using chemometric procedures, such as principal component analysis (PCA) and discriminant analysis (DA). Sandercock and Du Pasquier realized a discrimination of unevaporated gasoline samples by applying PCA and linear discriminant analysis (LDA) to the C0- to C2-naphthalene profiles obtained from GC–MS analysis (12). A successful association of evaporated gasoline samples to unevaporated gasoline samples using similar methods was also found: 35 samples made up 18 unique groups and a fair separation between the 35 samples at any given evaporation level was maintained, showing that the variation between samples was greater than that within samples (13). PCA was also applied to the C0- to C2-naphthalene profile, showing that samples tended to cluster together according to their fuel grade and the period of the year in which they were collected (i.e., in winter or in summer; [14]). A further comparison between samples collected in New Zealand and Australia showed that most samples could be grouped based on their country of origin, suggesting that the C0- to C2-naphthalene profile was dependent on the refinery where the gasoline was produced (14). PCA applied to GC–MS data was combined with artificial neural networks to classify unevaporated regular and premium gasoline samples (15). Gasoline samples were used as model mixtures to improve the classification method by data pretreatment, such as retention time alignment and analysis of variance feature selection. The following classification by LDA used optimal parameters and allowed the replicated analyses of each sample to cluster (16,17). In another study, Petraco et al. (18) discriminated 25 gasoline samples from casework by multivariate procedures from data generated by GC–MS, showing that the variability in the sample population was sufficient enough to distinguish all the samples from one another, knowing their groups in advance. Elsewhere, Hupp et al. (19) applied Pearson product moment correlation and PCA to group 25 diesel samples collected at service stations, representing eight different brands. PCA was performed on GC–MS data (after a data pretreatment) using both TICs and extracted ion chromatograms. Dimensional reduction with PCA revealed four distinct clusters for the TIC; both statistical procedures suggested that aromatic components provide the greatest discrimination among diesel samples (19).

In the study presented here, unevaporated gasoline samples of several brands were analyzed by SPME–GC–MS; the same method is in fact used also when samples of fire debris are collected from the fire scene for analysis. Chromatograms were then examined using the TCC approach (commonly used to detect the presence of any accelerants in an arson-related sample) to obtain a data set for multivariate statistical analysis. Finally, PCA was applied to find a natural grouping of samples according to their brand, while DA with canonical discriminant functions was used for numerical supervised sample classification. The aim of this study was to obtain a data matrix that could be used in real cases of arson to link a sample of unevaporated gasoline (that can sometimes be found at a fire scene, e.g., in an unburned can) to its brand or refinery, with a view to providing further information to forensic investigations.

Materials and Methods

Samples

Fifty samples of gasoline were collected in Parma (northern Italy) at service stations supplied from refineries of five well-known brands, named for the sake of this study “A,” “B,” “C,” “D,” and “E.” It was possible to obtain some information about the origin of the crude oil used in two of these refineries: A bought crude oil from only one country (Libya), while D was supplied with crude oil coming from several different countries.

It is important to note that the distributor and refinery that supply gasoline to a specific service station may change depending on several factors, including their location. The composition of the gasoline will invariably depend not only on the refinery itself but also on the residual level in the tank when filled. Thus, samples analyzed for this study were accepted for analysis only if the service station in question was supplied from the same refinery or distributor for at least two consecutive times. Further information was also obtained as to which refineries the distributors received gasoline from. Finally, 10 samples for each brand, coming from five refineries located in the north of Italy, were collected over a 9-month period (May 2005–January 2006).

Sample Preparation and Analysis SPME–GC–MS

Twenty microliters of the sample was dissolved in 500 mL of water and, after shaking, 1 mL of this mixture was put in a head space vial (10 mL of capacity) and hermetically sealed. The vial was then heated to 80°C for 15 min.

A fiber coated with a polydimethylsiloxane stationary phase was exposed for 3 min to the headspace of the sample vial to extract volatile compounds. The fiber was then introduced directly in the injection port of a gas chromatograph to thermally desorb the analytes (4).

GC was performed on a 50 m × 0.2 mm i.d. × 0.3 μm film HP5-MS capillary column using an Agilent 6890 gas chromatograph (Agilent, Milano, Italy) connected to an Agilent 5979 mass selective detector (Agilent, Palo Alto, CA). The GC conditions used were as follows: split injection 20.0:1 at 260°C; temperature set to rise from 70 to 130°C at 2°C/min (held for 0.2 min), then ramped to 260°C at 10°C/min (held for 26.8 min), giving a total run time of 70 min. Helium gas carrier was held at a constant flow rate of 1 mL/min, while the detector was set at a temperature of 280°C.

An electron-impact ionization source was utilized with quadrupole mass analyzer operated in full-scan mode (m/z 40–400) with a sampling rate of 3 scans/sec; MS-Source: 280°C; MS-Quad: 150°C.

Semi-Quantitative Analysis

Several samples of unevaporated gasoline, kerosene, and diesel fuel were analyzed select characteristic components of these petroleum-based fuels, that is, the TCs. Thirty-four variables, namely 17 aromatic compounds (that include both benzene and naphthalene derivatives) and 17 aliphatic compounds, which are alkanes (both normal and branched) and cycloalkanes, were therefore chosen (listed in Table 1). These compounds used in the TCC approach are also the variables employed in multivariate statistical analysis.

From the TIC of each sample (one of these is shown in Fig. 1), a semi-quantitative report of peak areas of TCs was obtained from a special data analysis program; these values were normalized to the area of the base peak (benzene, 1,2,3-trimethyl), put to 10,000. Three independent portions for each sample of gasoline were

analyzed, and the areas of these 34 peaks were averaged. As TCs are characteristic components of gasoline, kerosene, and diesel fuels, all these accelerants (already analyzed with the same GC-MS method) were able to be processed with the identical data analysis program.

TABLE 1—Target compounds of petroleum-based fuels.

Target	Compound
1	Benzene, 1-ethyl, 2-methyl
2	Benzene, 1,2,3-trimethyl
3	Benzene, 1-methylpropyl
4	Benzene, 1-methyl-2-(1-methylethyl)
5	Benzene, 1,2,4-trimethyl
6	Cyclohexane, butyl
7	Benzene, 1-methyl-3-propyl
8	Benzene, 1-ethyl-3,5-dimethyl
9	Benzene, 1-methyl-2-propyl
10	Benzene, 1-methyl-4-(1-methylethyl)
11	Benzene, 4-ethyl-1,2-dimethyl
12	Undecane
13	Benzene, 1-ethyl-2,3-dimethyl
14	Benzene, 1,2,3,5-tetramethyl
15	Benzene, 1,2,3,4-tetramethyl
16	Cyclohexane, pentyl
17	Benzene, 1,2,4,5-tetramethyl
18	Naphthalene
19	Dodecane
20	Cyclohexane, hexyl
21	Tridecane
22	Naphthalene, 2-methyl
23	Naphthalene, 1-methyl
24	Tetradecane
25	Pentadecane
26	Hexadecane
27	Heptadecane
28	Pentadecane, 2,6,10,14-tetramethyl
29	Octadecane
30	Hexadecane, 2,6,10,14-tetramethyl
31	Nonadecane
32	Eicosane
33	Heneicosane
34	Docosane

Multivariate Statistical Analyses

A first data set relative to 50 samples and 33 variables (averaged from the analysis of three portions for each sample) was obtained from the semi-quantitative analysis. The no. 2 variable, corresponding to the base peak, was excluded from the statistical analysis because its value is equal to 10,000 for all samples, therefore does not contain any information. A second data set consisted only of aromatic compounds (which are variables 1, 3, 4, 5, 7, 8, 9, 10, 11, 13, 14, 15, 17, 18, 22, and 23); these were picked to improve the results obtained from the statistical analysis of the first data set, as will be discussed in the following section.

The Matrix condition (namely the product of the norm of a matrix multiplied by the norm of the inverse matrix) was evaluated for each data set (using MATLAB version 7.2 [20]) to check system stability; the correlation matrix was also evaluated for each data set (using SPSS, version 11.01 [21]) to check the correlation between variables. Multivariate statistical analysis was then applied to these two data sets; the chemometric methods were PCA and DA (outlined in several publications [22–24]). PCA, applied to the standardized and unstandardized variables, projected the data into a subspace where data clusters, if present in said subspace, could be visualized.

DA was applied thereafter to classify the gasoline samples into five separate groups, according to their brand. The samples were divided into a training set, to develop a rule for the classification of the unknown samples, and a test set on which the model could be assessed. The percentage of the training set and test set members correctly classified by the model—called classification ability and prediction ability respectively—were evaluated for the first and the second data sets. In this study, classification and prediction ability were estimated using the “leave one out” method, which repeated the model calculations n times, each time leaving out a different sample (test set), with the other $n - 1$ samples forming the training set. Classification ability was calculated from the training set, excluding each time the sample that formed the test set, while prediction ability was estimated by predicting the classification of the test set sample from a model applied to the other $n - 1$ samples.

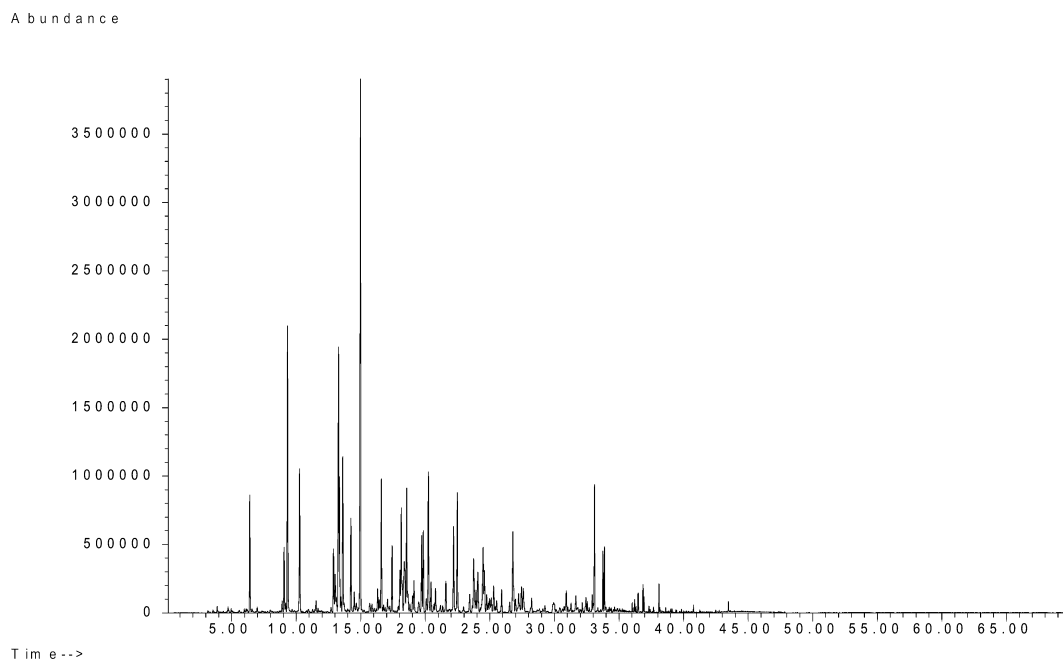


FIG. 1—Typical TIC of a gasoline sample.

Because DA provides reliable results if the ratio between the number of samples and variables is >3 (25), for the first data set (50 samples and 33 variables) this statistical analysis was applied to the first three principal components (PCs), while for the second data set DA was applied to the original variables.

DA was performed calculating canonical discriminant functions; this method is an extension of LDA, which finds a number of variables that reflect as much as possible the difference between the groups. The results are not dependent on scale, so no pretreatment of the data is necessary. PCA and DA were performed using SPSS (version 11.01 [21]).

Results and Discussion

PCA: 1st Data Set, Standardized Variables

Six PCs were extracted, with eigenvalues >1 ; the first three PCs explain 71.6% of the total variance; these results are presented in Table 2.

With regard to the score plot of PC2 versus PC1 and PC3 versus PC1, these are shown in Figs 2 and 3, respectively. A separation of samples in accordance with their brand was obtained for the "E" and "A" samples: "E" gasoline has the highest positive scores on PC1 (>1) while "A" samples have high positive scores on PC3 (>1) and negative scores on PC1 and PC2. "B," "C," and "D" samples are not as well separated as in the cases of "A" and "E," but show a trend in their differentiation. PCA was then used to identify the variables that contribute most to the variance in the gasoline data set, in other words the variables with the highest

(positive or negative) loadings on the PC1, PC2, and PC3; these can be seen in Table 3.

The results obtained can be related to the information on the origin of the crude oil (where available): A samples, tightly clustered in the scatterplots, were obtained from crude oil coming from only one country. This oil might have consistent chemical properties, which would help to explain why A samples appear to be very similar to each other. On the other hand, D samples, which are fairly well spread out in the same scatterplots, were obtained from crude oil

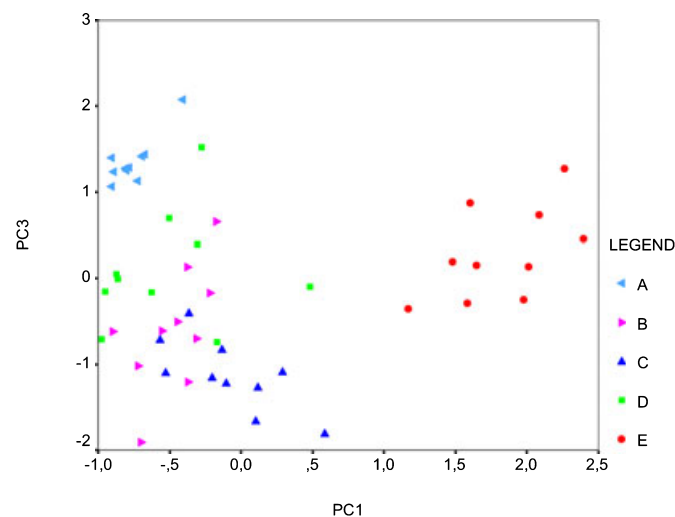


FIG. 3—Score plot of PC3 versus PC1, obtained from the 1st data set, standardized variables.

TABLE 2—PCs with eigenvalues >1 , extracted applying PCA to the 1st data set, standardized variables.

PC	Variance %	Cumulative %
1	44.08	44.08
2	21.03	65.11
3	6.44	71.55
4	5.97	77.52
5	4.41	81.93
6	3.43	85.37

TABLE 3—Variables with the highest loadings on the PC1, PC2, and PC3 (1st data set, standardized variables).

Variable	Compound	Loading on PC1
10	Benzene, 1-methyl-4-(1-methylethyl)	0.067
11	Benzene, 4-ethyl-1,2-dimethyl	0.067
9	Benzene, 1-methyl-2-propyl	0.066
13	Benzene, 1-ethyl-2,3-dimethyl	0.066
7	Benzene, 1-methyl-3-propyl	0.065
8	Benzene, 1-ethyl-3,5-dimethyl	0.065
14	Benzene, 1,2,3,5-tetramethyl	0.064
15	Benzene, 1,2,3,4-tetramethyl	0.062
18	Naphthalene	0.062
4	Benzene, 1-methyl-2-(1-methylethyl)	0.062
16	Cyclohexane, pentyl	0.061
3	Benzene, 1-methylpropyl	0.060
22	Naphthalene, 2-methyl	0.059
23	Naphthalene, 1-methyl	0.059
		Loading on PC2
26	Hexadecane	0.133
27	Heptadecane	0.128
28	Pentadecane, 2,6,10,14-tetramethyl	0.120
25	Pentadecane	0.117
31	Nonadecane	0.115
29	Octadecane	0.099
24	Tetradecane	0.098
34	Docosane	0.095
17	Benzene, 1,2,4,5-tetramethyl	-0.051
5	Benzene, 1,2,4-trimethyl	-0.074
		Loading on PC3
7	Benzene, 1-methyl-3-propyl	-0.125
1	Benzene, 1-ethyl, 2-methyl	-0.406

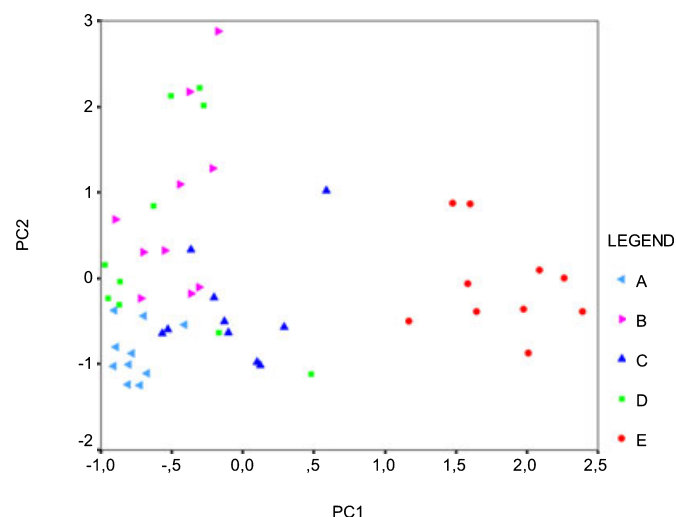


FIG. 2—Score plot of PC2 versus PC1, obtained from the 1st data set, standardized variables.

TABLE 4—PCs with eigenvalues >1, extracted applying PCA to the 1st data set, unstandardized variables.

PC	Variance %	Cumulative %
1	73.24	73.24
2	16.92	90.16
3	7.58	97.74

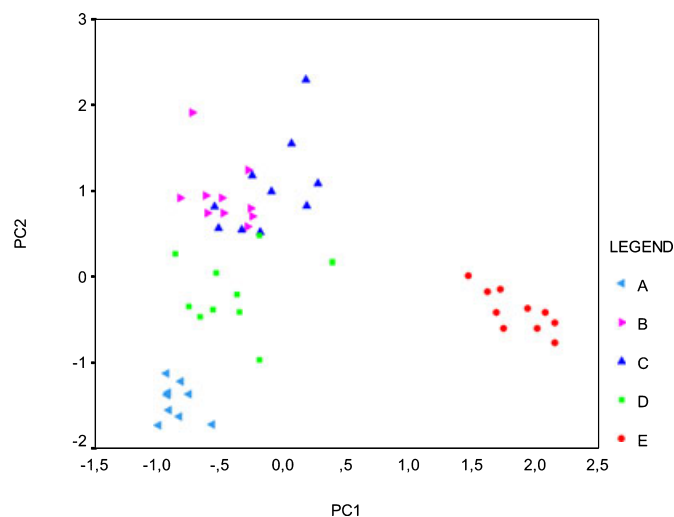


FIG. 4—Score plot of PC2 versus PC1, obtained from the 1st data set, unstandardized variables.

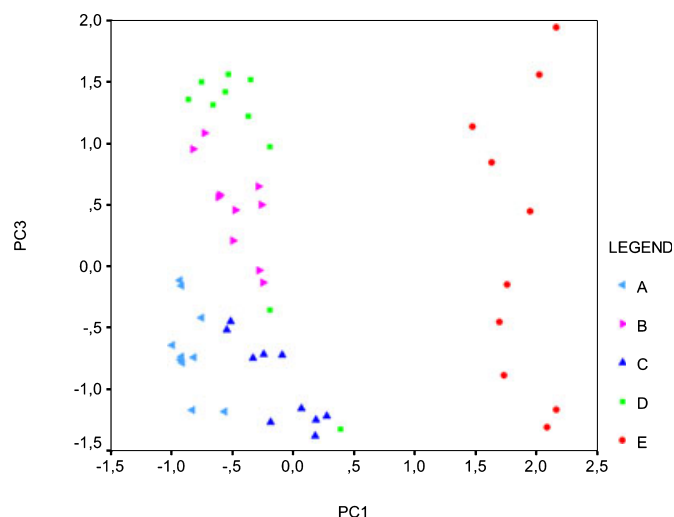


FIG. 5—Score plot of PC3 versus PC1, obtained from the 1st data set, unstandardized variables.

coming from several countries that may well have varying chemical properties. This would explain why D samples look very different from each other and why it is therefore more difficult to group them, compared to the more easily distinguishable A samples.

PCA: 1st Data Set, Unstandardized Variables

Three PCs were extracted, with eigenvalues >1; they account for 97.8% of the variance in the data, as shown in Table 4. The score

TABLE 5—Variables with the highest loadings on the PC1, PC2, and PC3 (1st data set, unstandardized variables).

Variable	Compound	Loading on PC1
18	Naphthalene	0.286
22	Naphthalene, 2-methyl	0.195
11	Benzene, 4-ethyl-1,2-dimethyl	0.149
8	Benzene, 1-ethyl-3,5-dimethyl	0.108
7	Benzene, 1-methyl-3-propyl	0.092
		Loading on PC2
1	Benzene, 1-ethyl, 2-methyl	0.715
5	Benzene, 1,2,4-trimethyl	-0.170
		Loading on PC3
23	Naphthalene, 1-methyl	0.129
10	Benzene, 1-methyl-4-(1-methylethyl)	-0.119
14	Benzene, 1,2,3,5-tetramethyl	-0.148
15	Benzene, 1,2,3,4-tetramethyl	-0.333

plots obtained without standardizing variables (Figs 4 and 5) indicate a better differentiation of samples according to their brand than the one obtained through the previous analysis.

The first PC clusters all “E” gasoline samples as they have the highest positive scores on this component (>1, as in the previous analysis), while all “A” samples cluster on PC2, having the highest negative scores on this component (<-1). “B,” “C,” and “D” samples tend to display a starker differentiation: “D” gasoline shows lower scores on PC2 and tends to shift toward higher scores on PC3 (except for two samples) than “B” and “C” samples. A separation between “B” and “C” samples is achieved on PC3, where the samples of the former brand show higher scores than the latter brand gasoline.

Variables with the highest loadings on PC1, PC2, and PC3 (shown in Table 5) include only aromatic compounds, whereas in the previous analysis some aliphatic compounds also had the highest loadings on the first three PCs (Table 3). This fact can be explained by considering the fact that in gasoline samples alkanes and cycloalkanes have very low chromatographic signals, so their loadings can be compared with aromatic ones only if the variables are standardized.

PCA: 2nd Data Set, Standardized Variables

The results obtained from the statistical analysis of the first data set demonstrated that standardized variables were characterized by a greater loss in discrimination power than unstandardized variables. As another difference between these two previous analyses was the respective presence and lack of aliphatic compounds between variables with the highest loadings on the PCs, it can be deduced that alkanes and cycloalkanes were not very useful in discriminating automotive gasoline samples in relation to their brand. Therefore, a further PCA was performed with a second data set that consisted only of aromatic compounds. Sixteen variables were used, that is, all the aromatic compounds listed in Table 1. PCA was first applied to the standardized variables. Three PCs were extracted, with eigenvalues >1; they account for 95.6% of the variance in the data, as shown in Table 6.

The score plots of PC2 versus PC1, and PC3 versus PC1 are shown in Figs 6 and 7, respectively, while variables with the highest loadings on PC1, PC2, and PC3 are shown in Table 7. From this table, it can be seen that the top contributors to PC1 and PC2 are benzene derivatives, while the top contributors to PC3 are naphthalene and its derivatives.

TABLE 6—PCs with eigenvalues >1, extracted applying PCA to the 2nd data set, standardized variables.

PC	Variance %	Cumulative %
1	73.50	73.50
2	14.82	88.32
3	7.29	95.61

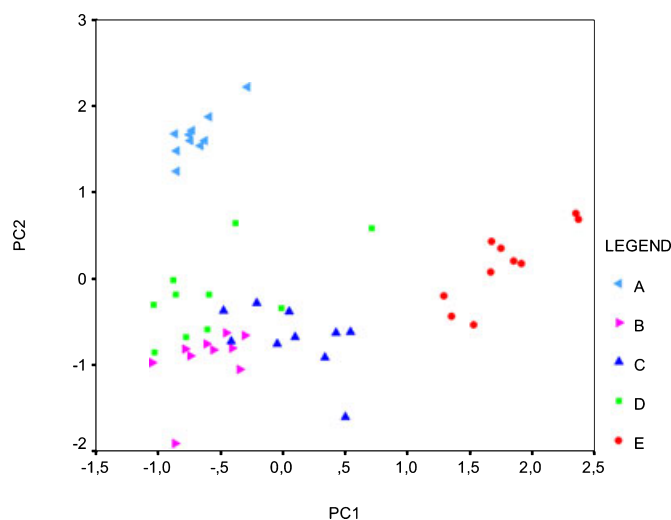


FIG. 6—Score plot of PC2 versus PC1, obtained from the 2nd data set, standardized variables.

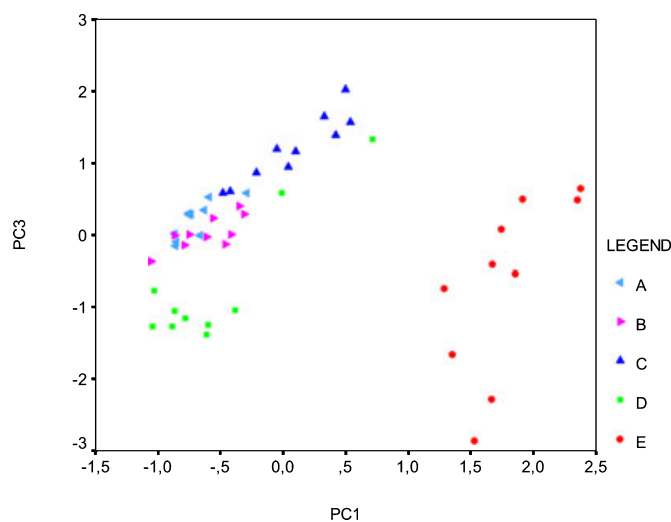


FIG. 7—Score plot of PC3 versus PC1, obtained from the 2nd data set, standardized variables.

Gasoline samples are better grouped according to their brand than results obtained in the two previous analyses. It can be noticed that “E” gasoline samples, even though they are relatively spread out, are separated from all other samples as they have the highest positive scores on PC1 (>1, as in the previous analyses). Therefore, they contained higher concentrations of the components that are positive in PC1 (Table 7), compared to the other gasoline samples.

“A” gasoline samples, with the highest positive scores on PC2 (>1), form the best cluster as they are the least spread samples. They contained higher concentrations of the components that are positive in PC2 and lower concentrations of the components that

TABLE 7—Variables with the highest loadings on the PC1, PC2, and PC3 (2nd data set, standardized variables).

Variable	Compound	Loading on PC1
10	Benzene, 1-methyl-4-(1-methylethyl)	0.084
11	Benzene, 4-ethyl-1,2-dimethyl	0.084
9	Benzene, 1-methyl-2-propyl	0.084
13	Benzene, 1-ethyl-2,3-dimethyl	0.083
7	Benzene, 1-methyl-3-propyl	0.082
14	Benzene, 1,2,3,5-tetramethyl	0.082
8	Benzene, 1-ethyl-3,5-dimethyl	0.081
4	Benzene, 1-methyl-2-(1-methylethyl)	0.080
15	Benzene, 1,2,3,4-tetramethyl	0.080
		Loading on PC2
5	Benzene, 1,2,4-trimethyl	0.399
17	Benzene, 1,2,4,5-tetramethyl	0.228
3	Benzene, 1-methylpropyl	-0.121
1	Benzene, 1-ethyl, 2-methyl	-0.383
		Loading on PC3
18	Naphthalene	-0.382
22	Naphthalene, 2-methyl	-0.464
23	Naphthalene, 1-methyl	-0.474

are negative in PC2 (Table 7), compared to the other gasoline samples. As “A” samples also show negative scores on PC1, they contained small relative quantities of the components that are positive in PC1 (Table 7). Scores on PC3 tend toward positive values and are approximately halfway between the negative scores of “D” samples and positive scores of “C” gasoline. Therefore, naphthalene and its derivative contents in “A” gasoline were halfway between the higher contents of “D” gasoline and lower contents of “C” gasoline.

“D” samples are spread and have high negative scores on PC3 (except for two samples); therefore, their trend was toward higher concentrations of naphthalene and its derivatives (Table 7) compared to the other gasoline samples (except for some “E” samples, which are spread on this component). “D” samples also show negative scores on PC1 (except for two samples) and thus contained small relative quantities of the components that are positive in PC1 (Table 7). Scores on PC2 are approximately halfway between negative scores of “B” and “C” samples and positive scores of “A” gasoline.

“C” samples, with the highest positive scores on PC3 (except for the two “D” samples named before), contained lower concentrations of the naphthalene and its derivatives (Table 7) compared to the other gasoline samples. They also have negative scores on PC2 and therefore contained small relative quantities of the components that are positive in PC2 and high relative quantities of the components that are negative in PC2 (Table 7).

“B” samples have negative scores on PC1, like “A” and “D” samples, and thus contained small relative quantities of the components that are positive in PC1 (Table 7). “B” samples also have negative scores on PC2, as “C” samples, while scores on PC3 are similar to those of “A” gasoline.

PCA: 2nd Data Set, Unstandardized Variables

Three PCs were extracted, with eigenvalues >1; they account for 98.1% of the variance in the data, as shown in Table 8.

From the score plots (Figs 8–10), it can be seen that the grouping of samples is very similar to that obtained with standardized variables. However, there are some improvements: “D” samples

TABLE 8—PCs with eigenvalues >1, extracted applying PCA to the 2nd data set, unstandardized variables.

PC	Variance %	Cumulative %
1	73.50	73.50
2	17.00	90.51
3	7.58	98.08

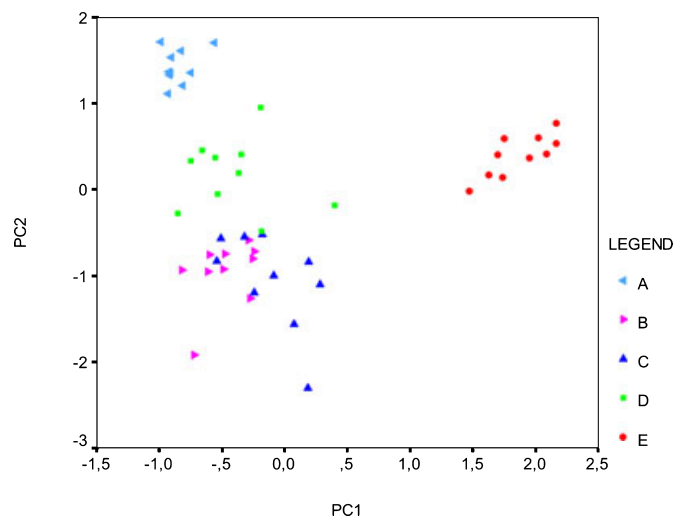


FIG. 8—Score plot of PC2 versus PC1, obtained from the 2nd data set, unstandardized variables.

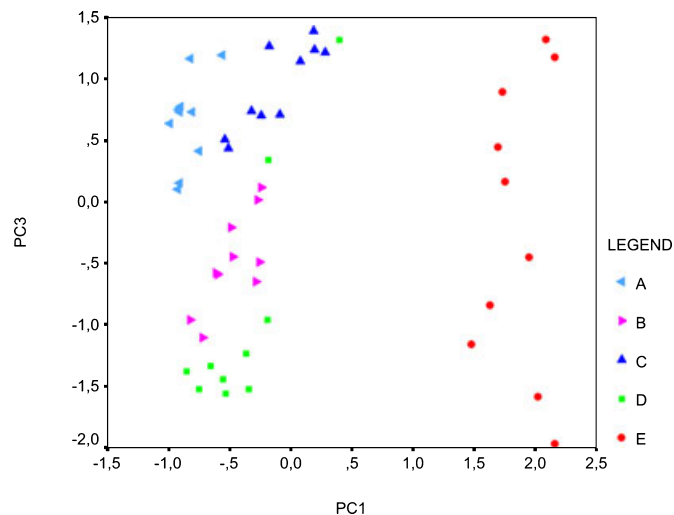


FIG. 9—Score plot of PC3 versus PC1, obtained from the 2nd data set, unstandardized variables.

are better separated on PC2 between positive scores of “A” gasoline and negative scores of “B” and “C” samples. Finally, the separation between “B” and “C” samples increases on PC3. Variables with the highest loadings on PC1, PC2, and PC3 (shown in Table 9) are, however, the same as those obtained with the first data set and unstandardized variables.

PCA was also applied to other data sets obtained from examining the correlation matrix, thus eliminating the variables that were most correlated to each other. The results obtained from these analyses do not show significant improvements compared to those

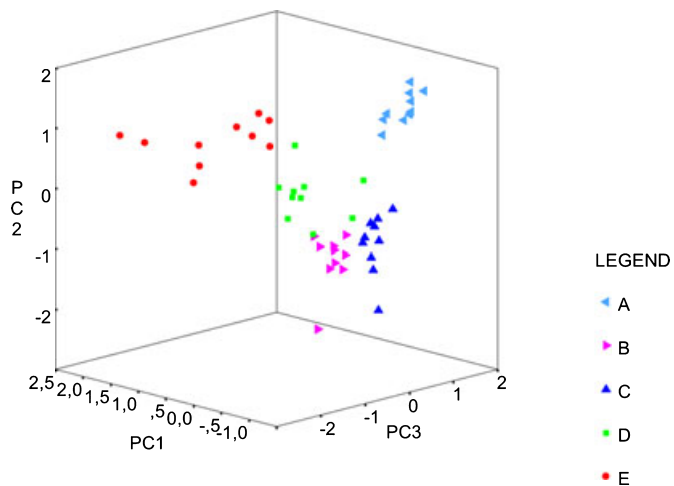


FIG. 10—Score plot of PC3 versus PC2 versus PC1, obtained from the 2nd data set, unstandardized variables.

TABLE 9—Variables with the highest loadings on the PC1, PC2, and PC3 (2nd data set, unstandardized variables).

Variable	Compound	Loading on PC1
18	Naphthalene	0.286
22	Naphthalene, 2-methyl	0.196
11	Benzene, 4-ethyl-1,2-dimethyl	0.150
8	Benzene, 1-ethyl-3,5-dimethyl	0.108
7	Benzene, 1-methyl-3-propyl	0.093
		Loading on PC2
5	Benzene, 1,2,4-trimethyl	0.170
1	Benzene, 1-ethyl, 2-methyl	-0.715
		Loading on PC3
15	Benzene, 1,2,3,4-tetramethyl	0.335
14	Benzene, 1,2,3,5-tetramethyl	0.149
10	Benzene, 1-methyl-4-(1-methylethyl)	0.119
23	Naphthalene, 1-methyl	-0.130

derived from the second set of data, suggesting that it is more useful to only choose between aromatic compounds and all the variables for grouping samples, than to choose variables simply in relation to their correlation.

Discriminant Analysis

The classification and prediction ability obtained performing DA on the first data set (the first three PCs from unstandardized variables, which account for 97.8% of the variance in the data, were preferred to corresponding PCs from standardized variables, which explain only 71.6% of the variance in the data) and on the second data set (16 aromatic compounds) are shown in Tables 10 and 11, respectively. The corresponding plots of the first two canonical discriminant functions are shown in Figs 11 and 12, respectively.

The best classification ability and prediction ability (100% in both) are obtained by performing DA on the second data set (Tables 10 and 11). When DA is applied to PCs of the first data set (from unstandardized variables), there are some misclassified samples, both by applying cross-validation and the algorithm that creates canonical discriminant functions (i.e., classification ability).

This result is confirmed in the plots of the first two canonical discriminant functions (Fig. 12), where the five groups of gasoline

TABLE 10—Classification ability and prediction ability obtained performing DA on the first three PCs coming from the 1st data set, unstandardized variables.

	Predicted Group Membership (%)				
	1	2	3	4	5
Original					
1	100	0	0	0	0
2	0	100	0	0	0
3	0	0	80	0	20
4	0	0	0	100	0
5	0	0	0	0	100
Cross-validated					
1	100	0	0	0	0
2	0	100	0	0	0
3	0	0	70	10	20
4	0	0	0	90	10
5	0	0	0	20	80

TABLE 11—Classification ability and prediction ability obtained performing DA on the 2nd data set.

	Predicted Group Membership (%)				
	1	2	3	4	5
Original					
1	100	0	0	0	0
2	0	100	0	0	0
3	0	0	100	0	0
4	0	0	0	100	0
5	0	0	0	0	100
Cross-validated					
1	100	0	0	0	0
2	0	100	0	0	0
3	0	0	100	0	0
4	0	0	0	100	0
5	0	0	0	0	100

brands are well separated only if discriminant functions are obtained from the second data set. In Fig. 11, “D,” “B,” and “C” groups are not as well separated as “E” and “A” clusters, in line with results obtained by performing the PCA.

Conclusions

Multivariate statistical analysis applied to the peak areas of TCs obtained by SPME–GC–MS analysis was successful in grouping 50 unevaporated gasoline samples according to their brands. Both PCA and DA procedures provided the same results.

PCA was first applied to 33 variables that are characteristic components of gasoline, kerosene and diesel fuel (first data set). The results showed a fair differentiation of “E” and “A” brands and also suggested that aromatic compounds were more useful than aliphatics in grouping gasoline brands. Even better results were obtained by applying the PCA to the second data set, consisting of only 16 aromatic compounds: “B,” “C,” and “D” samples improved their grouping according to their brand. For both data sets, the best results were obtained with unstandardized variables. For two brands, where information had been obtained about the origin of the crude oil, it was possible to link the chemical characteristics of samples to the crude oil used. DA confirmed the results given by PCA because a classification ability and a prediction ability of 100% in both were obtained only by performing the analysis on the second data set.

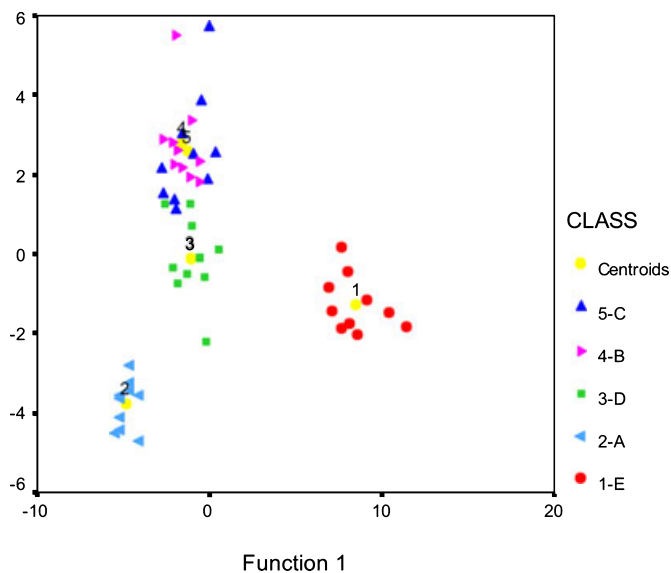


FIG. 11—Plot of the first two canonical discriminant functions obtained by performing DA on the first three PCs coming from the 1st data set, unstandardized variables.

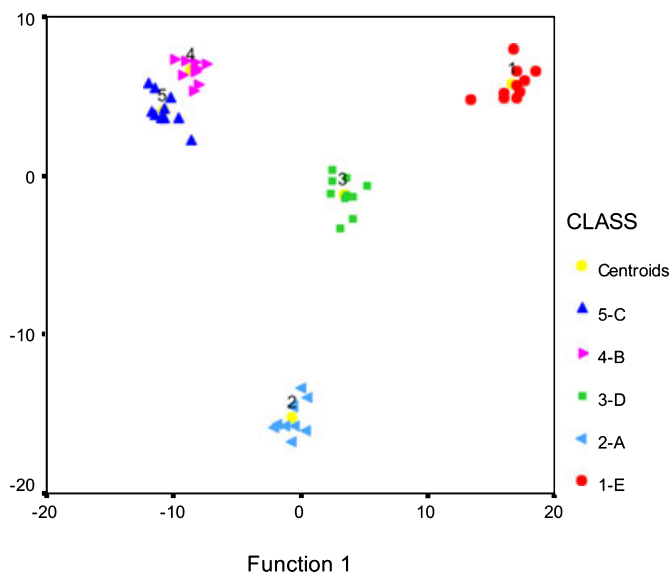


FIG. 12—Plot of the first two canonical discriminant functions obtained by performing DA on the 2nd data set.

This study provides the basis of a data matrix that could be used in real cases of arson to link a sample of unevaporated gasoline to its brand or refinery. This might help to answer questions posed by military, civil, and legal bodies about the origin of unevaporated gasoline samples taken from a suspected arsonist. It is possible to update the data matrix, according to variations that may occur in the composition of gasoline for each brand, to compare a gasoline sample taken from a fire scene with gasoline refined, however, no more than a year earlier.

Further developments of this project might provide an analogous data set for gasoline samples at different levels of evaporation, which in turn could be used for samples collected from fire debris where only traces of evaporated gasoline are found. The whole procedure could be applied to diesel fuel, commonly used as an accelerant, by using the same SPME–GC–MS analysis method and the identical data analysis program employed for gasoline.

References

1. Snow NH, Slack JC. Head-space analysis in modern gas chromatography. *Trends Anal Chem* 2002;21(9/10):608–17.
2. Almirall JR, Furton KJ. New developments in sampling and sample preparation for forensic analysis. In: Pawliszyn J, editor. *Comprehensive Analytical Chemistry*, Vol. 37. Elsevier Science, 2002:919–38.
3. Pert AD, Baron MG, Birkett JW. Review of analytical techniques for arson residues. *J Forensic Sci* 2006;51(5):1033–49.
4. American Society of Testing and Materials. ASTM Method E 2154–01 Standard practice for separation and concentration of ignitable liquid residues from fire debris samples by passive headspace concentration with solid phase microextraction (SPME). In: *Annual book of ASTM standards*. ASTM 2005;14(02):860–2.
5. Sandercock PML. Fire investigation and ignitable liquid residue analysis—a review: 2001–2007. *Forensic Sci Int* 2008;176:93–110.
6. Keto RO, Wineman PL. Target-compound method for the analysis of accelerant residues in fire debris. *Anal Chim Acta* 1994;288:97–110.
7. Keto RO, Wineman PL. Detection of petroleum-based accelerants in fire debris by target compound gas chromatography/mass spectrometry. *Anal Chem* 1991;63:1964–71.
8. Lennard CJ, Tristan Rochaix V, Margot P, Huber K. A GC–MS database of target compound chromatograms for the identification of arson accelerants. *Sci Justice* 1995;35(1):19–30.
9. American Society of Testing and Materials. ASTM Method E 1618–01 Standard test method for ignitable liquid residues in extracts from fire debris samples by gas chromatography–mass spectrometry. In: *Annual book of ASTM standards*. ASTM 2005;14(02):539–49.
10. Bodle ES, Hardy JK. Multivariate pattern recognition of petroleum-based accelerants by solid-phase microextraction gas chromatography with flame ionization detection. *Anal Chim Acta* 2007;589:247–54.
11. Tan B, Hardy JK, Snavely RE. Accelerant classification by gas chromatography/mass spectrometry and multivariate pattern recognition. *Anal Chim Acta* 2000;422:37–46.
12. Sandercock PML, Du Pasquier E. Chemical fingerprinting of unevaporated automotive gasoline samples. *Forensic Sci Int* 2003;134(1):1–10.
13. Sandercock PML, Du Pasquier E. Chemical fingerprinting of gasoline 2. Comparison of unevaporated and evaporated automotive gasoline samples. *Forensic Sci Int* 2004;140(1):43–59.
14. Sandercock PML, Du Pasquier E. Chemical fingerprinting of gasoline Part 3. Comparison of unevaporated automotive gasoline samples from Australia and New Zealand. *Forensic Sci Int* 2004;140:71–7.
15. Doble P, Sandercock PML, Du Pasquier E, Petocz P, Roux C, Dawson M. Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. *Forensic Sci Int* 2003;132:26–39.
16. Pierce KM, Hope JL, Jonson 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.
17. Watson NE, VanWingerden MM, Pierce KM, Wright BW, Synovec RE. Classification of high-speed gas chromatography-mass spectrometry by principal component analysis coupled with piecewise alignment and feature selection. *J Chromatogr A* 2006;1129:111–8.
18. Petracco NDK, Gil M, Pizzola PA, Kubic TA. Statistical discrimination of liquid gasoline samples from casework. *J Forensic Sci* 2008;53(5):1092–101.
19. Hupp AM, Marshall LJ, Campbell DI, Smith RW, McGuffin VL. Chemometric analysis of diesel fuel for forensic and environmental applications. *Anal Chim Acta* 2008;606:159–71.
20. MATLAB [computer program version 7.2], <http://www.mathworks.com/help/techdoc/rn/bqsqsom.html>.
21. SPSS, Inc. SPSS [computer program, version 11.01], <http://support.spss.com/Student/Patches/SPSS/Windows/Pre12/SPSS-1101ReadMe.html>.
22. Gemperline PJ. *Practical guide to chemometrics*, 2nd edn. Boca Raton FL: CRC Press, 2006.
23. Massart DL, Vandeginste BGM, Deming SN, Minchotte Y, Kaufman L. *Chemometrics: a textbook*. New York, NY: Elsevier, 1988.
24. Jolliffe IT. *Principal component analysis*, 2nd edn. New York, NY: Springer, 2004.
25. Kaufmann A. Multivariate statistics as a classification tool in the food laboratory. *J AOAC Int* 1997;80:665–75.

Additional information and reprint requests:
 Maria Monfreda, M.Sc.
 Agenzia delle Dogane
 Direzione Regionale per il Lazio e l'Umbria
 Sezione Operativa Laboratorio Chimico
 Via Carucci 71
 00143 Rome
 Italy
 E-mail: maria.monfreda@gmail.com