Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1047 (2004) 101–109
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Detection of soil pollution by hydrocarbons using headspace–mass spectrometry and identification of compounds by headspace–fast gas chromatography–mass spectrometry

José Luis Pérez Pavón*, Armando Guerrero Peña[1], Carmelo García Pinto, Bernardo Moreno Cordero

Departamento de Química Analítica, Nutrición y Bromatología, Facultad de Ciencias Químicas, Universidad de Salamanca, 37008 Salamanca, Spain

Received 26 January 2004; received in revised form 9 April 2004; accepted 29 June 2004

Available online 27 July 2004

Abstract

The direct coupling of a headspace sampler with a mass spectrometer is proposed as a screening tool for the rapid detection of soil pollution by hydrocarbons from petroleum and derivatives. The samples are subjected to the headspace generation process, with no prior treatment, and the volatiles generated are introduced directly into the mass spectrometer, thereby obtaining a fingerprint of the sample analysed. Suitable treatment of the signal by chemometric techniques allows unequivocal characterisation of the different types of sample. The use of fast gas chromatography with a mass spectrometer detector coupled to the headspace sampler allows identification of the major hydrocarbons present in the mineral and organic polluted samples, interpretation of the results obtained, and demonstrates the analytical potential of headspace–mass spectrometry coupling.

Keywords: Soil; Headspace analysis; Mass spectrometry; Gas chromatography, fast; Chemometrics; Environmental analysis; Hydrocarbons

1. Introduction

Petroleum crude oils and derivatives are one of the main sources of the pollution by hydrocarbons of matrices of environmental interest [1]. Gas chromatographic methods [both GC–MS and GC–flame ionisation detection (FID)] have become the most popular tool for the identification and quantification of this kind of pollution [2]. Previous treatment of the sample, which may include preconcentration steps such as solid-phase microextraction or purge-and-trap [3,4], is the most costly and time-consuming part of the process and one of the most frequent sources of errors [5].

Throughput in the analytical laboratory can be increased significantly by developing non-separative methods for the resolution of different analytical problems, resulting in lower operational costs per sample. In this sense, the direct coupling of mass spectrometry with methods such as solid-phase microextraction (SPME) [6,7], or headspace (HS) generation [8–12] has been developed for the characterization of different samples. Compared to other time-consuming complex separation methods, the signal provided by these techniques can be used as a "fingerprint" that contains enough information – when suitably processed with the appropriate chemometric techniques – to make decisions about the proposed problem. In this context, direct coupling of solid-phase microextraction with Raman spectroscopy (SPME–Raman) [13] and solid-phase extraction with IR spectroscopy (SPE–IR) [14] have been proposed for the study of pollution due to hydrocarbons.

The collection of profile signals provides little information about the individual volatile compounds present in the sample. In cases in which it is necessary to obtain more specific information, chromatographic separation techniques must be

* Corresponding author. Tel.: +34 923 294483; fax: +34 923 294483.
E-mail address: jlpp@usal.es (J.L. Pérez Pavón).
[1] Present address: Laboratorio de Suelos, Plantas y Aguas, Campus Tabasco, Colegio de Postgraduados, Supera-Anuies, Mexico.

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.06.113

used. In this sense, the development of fast gas chromatographic techniques considerably reduces the analysis time required for the identification and/or quantification of the compounds [15–17]. The main development of faster GC methods has been observed in the field of environmental analysis [18–20]. The advantages of HS–MS coupled together with those of fast gas chromatography (FGC) suggest that HS–FGC–MS has an important analytical potential for the rapid detection and identification (where necessary) of pollutants in environmental matrices.

In previous works we have recently proposed the direct coupling of a headspace sampler with a mass spectrometer (HS–MS) for the rapid detection of soil contamination by crude oil and derivatives [21] and the use of multiplicative calibration transfer for the generation of multivariate calibration models valid over long periods of time, for the quantification of volatile organic compounds (VOCs) comprising the species benzene, toluene, methyl *tert*-butyl ether, ethylbenzene, *m*-xylene and mesitylene [22]. The procedure does not require any chromatographic step since it is based exclusively on the generation of volatiles and later analysis using a quadrupole mass detector.

In this work we propose the use of the HS–MS methodology as a screening tool for the rapid detection of soil pollution due to petroleum hydrocarbons. The model used to the prediction of the samples was constructed with polluted samples prepared at the laboratory with clean commercial sand. Only an internal standardization pretreatment process of the signals obtained permits the construction of a stable model valid for long periods of time with excellent results as regards prediction.

Additionally, we performed chromatographic separation of all unknown samples by headspace–gas chromatography–mass spectrometry (HS–GC–MS) and headspace–fast gas chromatography–mass spectrometry (HS–FGC–MS) to identify the volatiles of the different samples analysed. This information was used to interpret the results obtained with the chemometric techniques carried out [hierarchical cluster analysis (HCA), principal component analysis (PCA) and linear discriminant analysis (LDA)] and to show the analytical potential of the HS–MS coupling.

2. Experimental

2.1. Samples

Twenty-five soil samples from Mexico were studied. Three of them were reference samples known not to contain remains of hydrocarbons from petroleum [Histosol (H), Vertisol (V) and Arenosol (A)]. The other 22 samples (samples 1–22) were soils collected from different zones of the province of Tabasco, concerning which their possible contamination was a priori unknown. These samples included both mineral and organic soils. The first ones (samples 1, 5, 6, 8, 10, 13, 15, 17, 18, 21, and 22) were from horizon type Ap, with an organic matter content between 0.1 and 14%, a clay content between 50 and 480 g/kg, pH between 4.6 and 8.3, and a cation-exchange capacity (CEC) between 0.8 and 51.1 cmol/kg. The second samples (2, 3, 4, 7, 9, 11, 12, 14, 16, 19, and 20) were from horizon type Ai, with an organic matter content between 27 and 95%, pH between 4.0 and 6.8, and a CEC between 39 and 130.2 cmol/kg. All samples were collected in metallic containers and stored at −5 °C until analysis. All of them were analysed by HS–MS, HS–GC–MS and HS–FGC–MS.

Table 1
Set of samples used in this work

Class	Samples	Total[a]
Unpolluted	Clean sand	10
	Beach sands (from Spain)	3
	Arenosol, Histosol and Vertisol (from Mexico)	3
Polluted[b]	Iran light crude oil (1.4–431 mg/kg)	10
	Brass river light crude oil (1.3–409 mg/kg)	10
	Diesel fuel (1.2–371 mg/kg)	10
Unknown	Soils 1–22 (from Mexico)	22

[a] Beach sand samples and polluted samples were run in triplicate.
[b] All samples in the polluted class have been prepared in the laboratory by spiking the commercial clean sand with the different pollutants at 10 uniformly distributed concentration levels.

Additionally, a set of samples prepared at the laboratory and measured two and a half years previously [21] was used for the generation of the PCA and LDA models. This set of samples comprised a subset of 90 polluted samples (commercial clean sand spiked at ten uniformly distributed concentration levels with Iran light crude oil, Brass River light crude oil, and diesel fuel, respectively and analysed in triplicate) and a subset of 19 unpolluted samples (10 commercial clean sand, and three beach sand samples from Spain, the last ones analysed in triplicate). Table 1 shows the set of samples used in this work both for constructing the different models and for the prediction of unknown samples.

2.2. Apparatus

Sample analyses were performed with an Agilent 6890/5793 GC/MSD system coupled to a headspace sampler. This headspace sampler (HP 7694) is equipped with a tray for 44 consecutive samples, an oven with positions for six sample vials, where the headspace is generated, and a sampling system comprising a stainless steel needle, a 316-SS six-port valve with a nickel loop, and two solenoid valves (for pressurization and venting). Data collection was performed with Pirouette v3.0 [23] software from Infometrix on a Hewlett-Packard PC computer that also controlled the MS detector parameters. Two different columns were used for the chromatographic separations: an HP-5MS (5%)-diphenyl-(95%)-dimethylsiloxane capillary column (30 m × 0.25 mm; film thickness, 0.25 μm) purchased from Supelco (Bellefonte, PA, USA), and a poly(siloxane) phase DB-VRX capillary column (20 m × 0.18 mm; film thickness, 1.00 μm) purchased

from J&W Scientific (Folsom, CA, USA). The composition of DB-VRX is considered proprietary information.

2.3. Procedure

2.3.1. HS–GC–MS

For the analysis of volatile compounds, aliquots of 2 g of each soil sample available – polluted, unpolluted, and unknown – were placed in 10-mL vials and sealed hermetically with silicone septum caps. These vials were introduced in the oven of the headspace sampler at a temperature of 95 °C for 45 min where the headspace is generated, whereas the temperature of the nickel loop was 120 °C. The volatiles generated were injected in the chromatographic system through a thermostatted transfer line heated to 130 °C.

To perform the gas chromatographic measurements, the column (HP-5MS) was initially maintained at 35 °C for 2 min; then, temperature was increased to 250 °C at a rate of 10 °C/min, which was then held for an additional 3 min. For the fast gas chromatographic analysis, the column (DB-VRX) was initially maintained at 50 °C for 0.1 min; after this time, the temperature was first increased to 175 °C at a rate of 60 °C/min and then increased to 240 at 45 °C/min, which was then held for an additional 4 min.

The carrier gas was helium N50 (99.995% pure, from Air Liquide). The *m*/*z* range was 49–160 amu and the compounds were identified by comparison of their experimental spectra with those of the NIST'98 data bank (NIST/EPA/NIH Mass Spectral Library, version 1.6, USA).

2.3.2. HS–MS

In order to measure the patterns of volatiles of the soil samples without chromatographic separation (HS–MS methodology), the oven temperature was maintained high enough (240 °C) to prevent retention of the injected volatile compounds. The total ion current signal was obtained in the *m*/*z* range considered (49–160 amu; threshold: 150; scans/s: 6.48).

2.4. Data analysis

Internal normalization involves expressing each mass fragment of each individual spectrum as a percentage of the sum of the mass fragments and was accomplished using Pirouette software.

These normalized data were subjected to analysis with the different pattern recognition techniques to evaluate the discriminating power of the HS–MS methodology. HCA and PCA were performed with Pirouette v3.0 software, while the PARVUS statistical package (Geneva, Italy) [24] was used to perform the LDA.

The hierarchical cluster analysis was performed with the 25 samples of Mexican soils. PCA was carried out in a first stage with the same 25 samples. Later, PCA was used for classification purposes, constructing a model of principal components with the artificially polluted samples (90), and the unknown samples were predicted.

For analysis with supervised pattern recognition techniques (LDA) the model generation step was carried out with a set of samples (training set) formed by the spiked clean sand and clean sand samples, the three beach sand samples, and the reference samples Histosol, Vertisol, and Arenosol (90 polluted samples and 22 unpolluted samples). The 22 unknown samples of Mexican soil (not included in the models) were used to study the prediction capacity of the model.

3. Results and discussion

3.1. HS–MS methodology

3.1.1. Cluster analysis

Once the signals of the unknown and reference samples had been obtained, a cluster analysis was performed. To apply this technique of exploratory analysis, the signals were normalised internally in such a way that the intensity of each mass–charge relationship was divided by the sum of the intensities of all the fragments in the interval recorded (49–160).

When the Euclidean distance was used as a measure of similarity and complete linkage as a way to generate clusters, the dendrogram shown in Fig. 1 was obtained. In that figure, two main clusters can be seen ("a" and "b"), each subdivided into two subclusters (a1–a2 and b1–b2, respectively). The "a" group included the clean soil samples used as reference.

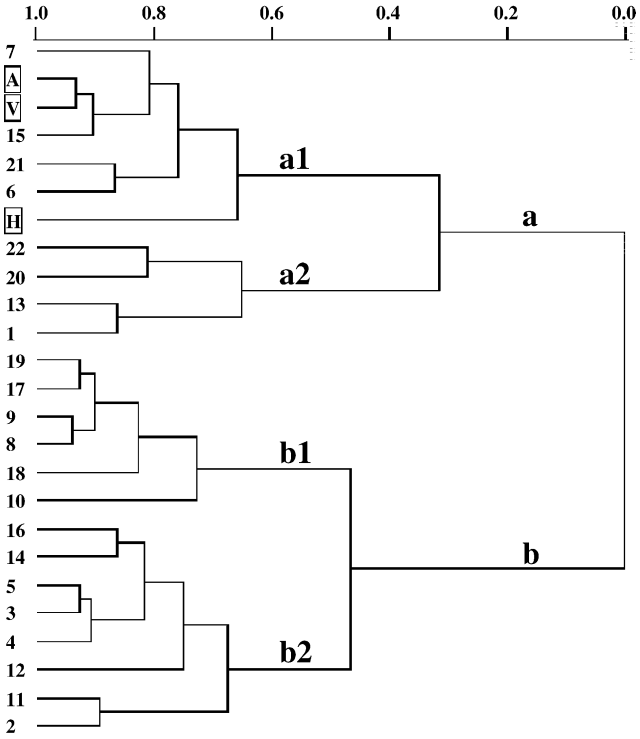

Fig. 1. Complete linkage dendrogram obtained with hierarchical cluster analysis for the samples from Mexico including the reference Arenosol (A), Histosol (H), and Vertisol (V) samples and the unknown samples.

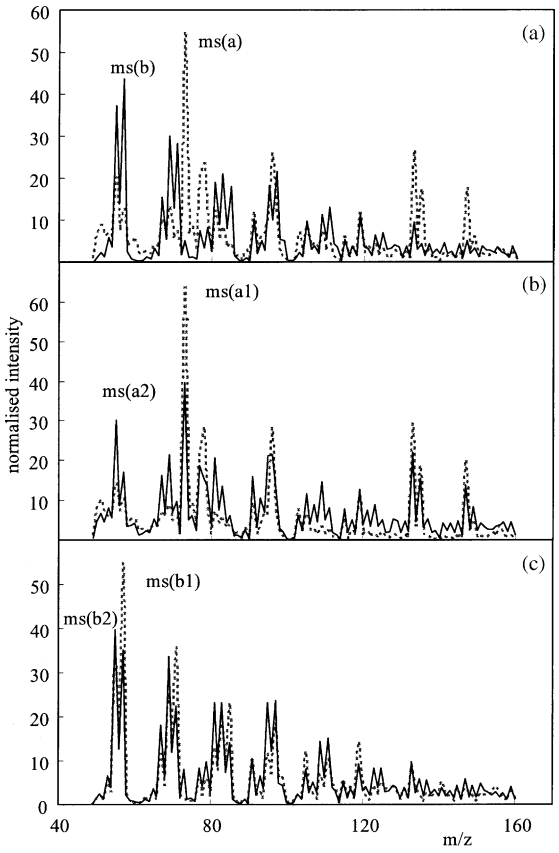

Fig. 2. (a) Mean mass spectra of the samples of groups "a" and "b" (ms(a) and ms(b), respectively) of the dendrogram of the Fig. 1, (b) of the subclusters"a1" and "a2" (ms(a1) and ms(a2), respectively) and (c) of the subclusters"b1" and "b2" (ms(b1) and ms(b2), respectively).

The differences in the signals produced by this separation into two groups were seen by obtaining the mean spectra of the two sample groups (Fig. 2). In the mean spectrum of the "b" group [ms(b) in Fig. 2a], it is possible to note the presence of signals corresponding to the characteristic patterns of alkanes, such as the intensity ratio of fragments 57–71–85. The mean spectrum of the "a" (ms(a)) group does not show similarity with any other group of compounds present in petroleum fractions.

It therefore seems that the two groups generated correspond to samples polluted by hydrocarbons from petroleum fractions (group "b") and to samples unpolluted with this type of compound (group "a").

The division of group "a" into two subclusters – "a1" and "a2" – can also be attributed to certain differences in the mass patterns generated by the volatiles contained in them, and these can be appreciated by comparison of the mean spectra of each sample subcluster (ms(a1) and ms(a2) in Fig. 2b), although with this type of signal it is not possible to attribute such differences to particular compounds.

From a comparison of the mean spectra of the "b1" and "b2" subclusters (ms(b1) and ms(b2) in Fig. 2c), it is possible to observe more specific differences, such as those concerning the intensity ratio of fragments 55 and 57, which could be attributed to patterns of different alkanes, with a higher content in cyclic alkanes in one of the two subclusters.

3.1.2. Principal component analysis

PCA is a technique used for reducing dimensionality that is usually used in exploratory analysis as a method for visualization. Here it was carried out using the same set of samples as that used in the cluster analysis.

The first two principal components explained 95.6% of the variance of the data (75.84% and 19.76%, respectively). Plotting the scores of the samples in the second principal component showed the division of the samples into two groups: those with positive scores and those with negative scores (Fig. 3a). It may be seen that all the samples with a positive score (among which were the reference samples of clean soils) coincided with those from group "a" (Fig. 1) and the samples with negative scores coincided with those belonging to group "b".

Fig. 3b shows that the loadings of principal component 2 were negative for variables corresponding to fragments typical of linear (57, 71, 85) and cyclic (55, 69, 83, 97) alkanes, such that it is likely that the samples showing negative scores would correspond to soils in which compounds from petroleum fractions are found.

PCA is the basis of the SIMCA technique, which is useful for modelling classes and the later prediction of unknown

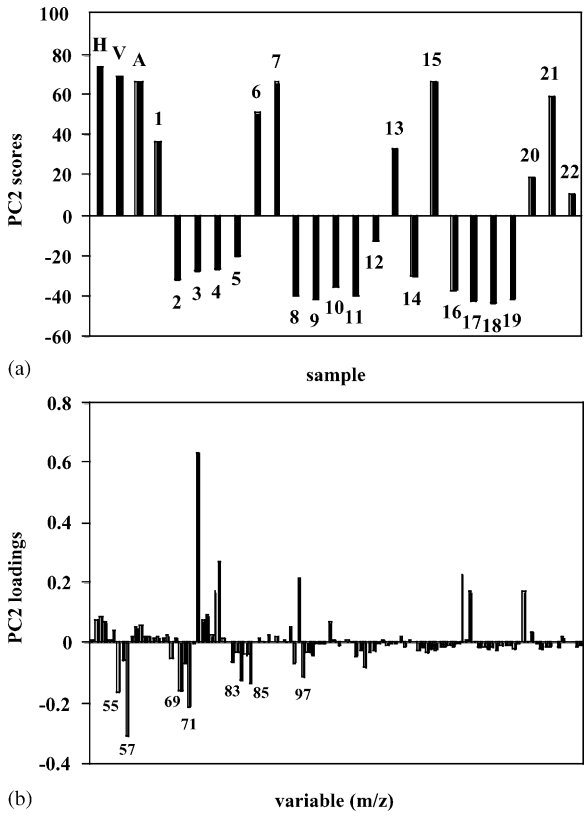

Fig. 3. (a) Plot of the scores of the samples with respect to the second principal component; (b) plot of the loads of the second principal component against the *m*/*z* ratio.

samples. In this case, our interest lay not in deciding which two classes or more a sample belonged to but, instead, in deciding whether a sample belonged to a class (polluted with compounds from petroleum fractions) or not. Whereas the characteristics of the polluted samples are relatively well defined, those of the unpolluted soil samples may vary considerably.

To generate the model, we used the data available from a set of samples obtained by spiking clean commercial sand with different amounts two petroleum crude oils and a diesel fuel. These samples had been measured two and a half years before the unknown samples to be analysed, although as has been reported earlier [21], the process of internal normalization used as a technique for data pretreatment allows the use of signals generated at very different times since variations in the sensitivity of the apparatus are compensated.

With this set of samples, the first two principal components explained 97.35% of the total variance and hence a model with two components was fixed (Fig. 4a).

In PCA, predictions are carried out by projecting the unknown samples onto the space defined by the principal components of the training set. The decision as regards whether a given sample differs significantly from those of the training set is mainly based on the magnitude of the residuals when that sample is projected onto the space of the model. From these residuals it is possible to calculate a probability which can then be compared with a limit value.

Fig. 4b shows the values of the residuals against the Mahalanobis distance for each of the unknown samples and the three reference samples when they were projected in the principal component model of the samples spiked at the laboratory with the two petroleum crude oils and the diesel fuel.

The samples located in the region that is clearly outside the limits marked by both magnitudes (99% probability) should be considered as samples that are significantly different from those of the model; that is, samples free of pollution. These samples (H, V, A, 1, 6, 7, 13, 15, 20, 21 and 22) are the same as those that formed group “a” in the cluster analysis. Within the limits for the model are all the remaining samples except two (samples 4 and 5), which – although they are within the limit for the Mahalanobis distance – slightly surpass the limit values for the residuals. Using this classification system, samples 4 and 5 could be considered as doubtful.

3.1.3. Linear discriminant analysis

Study of the signals using the LDA method was carried out in two steps. In the first, a classification model was built

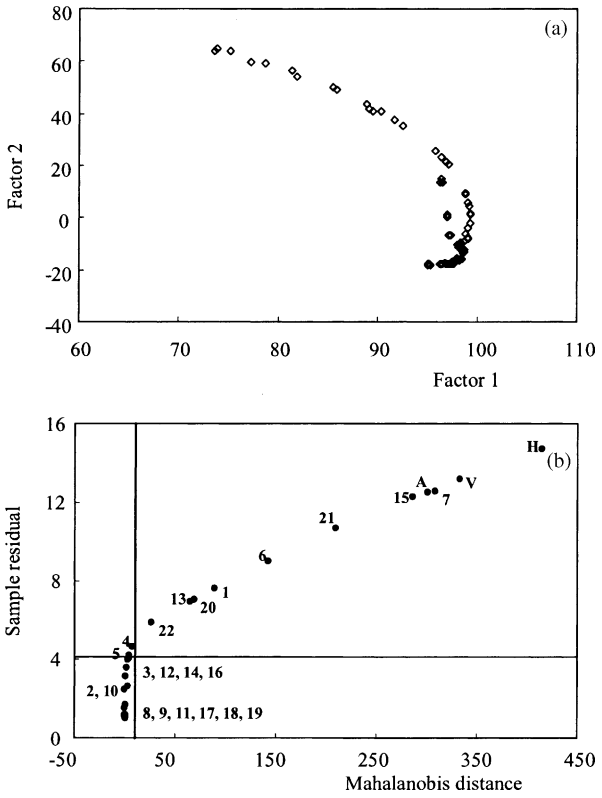

Fig. 4. (a) Model obtained with the first two principal components corresponding to samples at the laboratory from clean commercial sand; (b) values obtained upon projecting residuals against the Mahalanobis distance of the unknown samples including the references samples from Mexico.

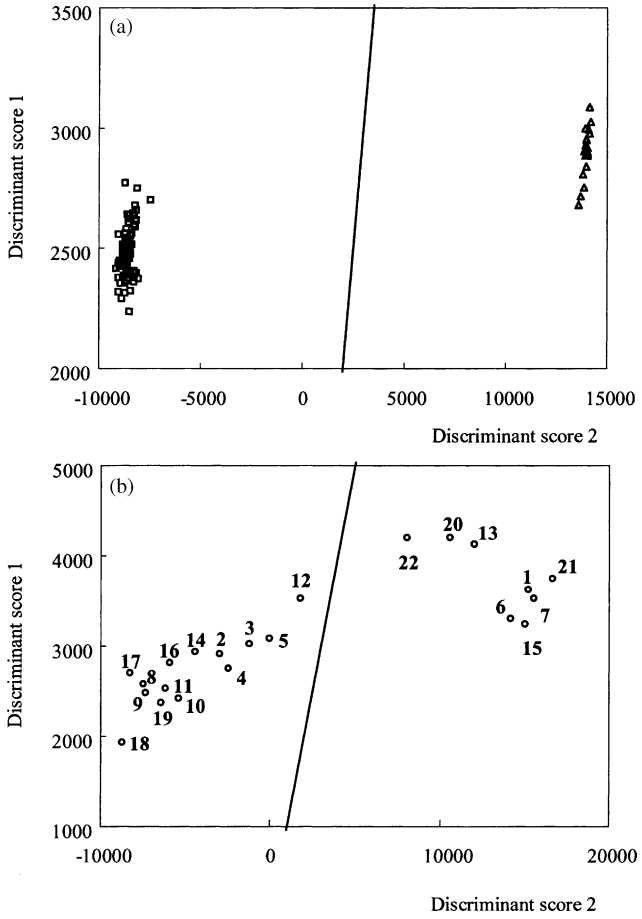

Fig. 5. (a) Plots of the discriminant scores for the model (40 variables) obtained with clean sand samples (including the reference samples) and (b) for the prediction of the external validation set formed by unknown samples from Mexico.

in which the training set comprised spiked sand samples to define the polluted class, and clean sand samples, the three beach sand samples, and the reference samples – Histosol, Vertisol, and Arenosol – to define the non-polluted class.

A StepLDA (stepwise linear discriminant analysis) variable selection process was carried out in order to obtain the greatest Mahalanobis distances between the closest classes. To validate the model, a cross-validation process was used [21].

The model was constructed with 20, 30, 40 and 50 variables and in all cases 100% hit rates were obtained both in classification and in prediction. The increase in the number of variables produced an increase in the separation between classes, although dispersion among samples within the same group also increased. For 40 variables (Fig. 5a) a suitable distance between classes was obtained, maintaining acceptable dispersion within the group.

With this model we carried out the prediction of the unknown set, in this case formed by the Mexican soil samples. Fig. 5b shows the plot of discriminant scores for the prediction set. All the samples grouped in the "b" cluster in the hierarchical cluster analysis were recognised by the model as polluted samples, just as the samples of the "a" cluster were categorised within the non-polluted sample class. The classification of samples achieved with the supervised pattern recognition technique is thus in concordance with the results obtained previously using the unsupervised pattern recognition techniques. The fact that most of the samples used to generate the model (except Histosol, Vertisol and Arenosol) were measured two and a half years before the samples used in the prediction points to the stability of the model and the capacity of the methodology proposed for the detection of pollution by hydrocarbons from petroleum.

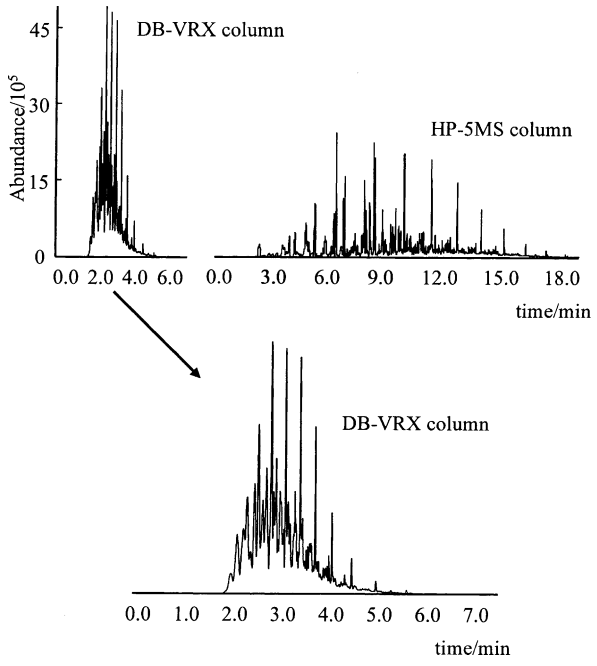

Fig. 6. Chromatograms obtained with the two columns used in this work.

3.2. HS–GC–MS measurements

As seen, use of the HS methodology permits simple and rapid discrimination between samples polluted by hydrocarbons from petroleum and samples "free" of such pollution. Accordingly, the technique can be employed as a screening tool. Despite this, the profiles of the signal obtained do not allow appropriate identification of the compounds present in the different samples.

Table 2
List of compounds identified by HS–GC–MS in the soil number 10 with the two columns used in this work

Compound	HP-5MS column		DB-VRX column	
	t_R[a]	Match quality	t_R[a]	Match quality
Linear alkenes				
n-Heptane	3.919	85		
n-Octane	5.323	70		
n-Nonane	6.979	68	2.669	79
n-Decane	8.660	77	2.950	96
n-Undecane	10.266	77	3.243	94
n-Dodecane	11.779	78	3.545	94
n-Tridecane	13.200	82	3.855	94
n-Tetradecane	14.536	89	4.204	93
n-Pentadecane	15.798	88	4.612	92
n-Hexadecane	16.993	88	5.113	92
n-Heptadecane	18.128	90	5.741	90
n-Octadecane	19.208	88		
n-Nonadecane	20.240	81		
Branched alkenes				
2,6-Dimethylundecane	11.975	74		
2,6,11-Trimethylundecane	12.441	77		
2,6,10-Trimethyldodecane	14.228	89		
Cyclic alkenes				
Methylcyclohexane	4.221	86	2.349	83
1,3-Dimethylcyclohexane	4.997	84		
1,1,3-Trimethylcyclohexane	5.931	87		
1,3,5-Trimethylcyclohexane	6.172	87		
Benzene derivatives				
Toluene	4.843	94		
C2-Benzene	6.354	97		
C2-Benzene	6.488	97	2.691	92
C2-Benzene	6.884	96	2.764	94
C3-Benzene	7.420	95		
C3-Benzene	8.051	95	2.933	88
C3-Benzene	8.162	94		
C3-Benzene	8.359	96	2.999	93
C3-Benzene	8.587	95	3.050	96
C3-Benzene	9.071	91	3.152	89
C4-Benzene	9.528	90	3.257	88
C4-Benzene	9.585	84	3.451	78
C4-Benzene	9.703	92	3.576	89
C4-Benzene	9.790	92		
C5-Benzene	11.345	81	3.563	88
Naphthalene derivatives				
C1-Naphthalene	13.251	80	4.144	91
C1-Naphthalene	13.496	89	4.217	87
C2-Naphthalene	14.939	94		
C2-Naphthalene	14.985	90		

[a] t_R = retention time (min).

With a view to identifying the major hydrocarbons and in order to unequivocally check the results obtained with the chemometric techniques used previously, we performed a chromatographic analysis of all the Mexican soil samples. Conventionally, this analysis is carried out with capillary columns (HP-5MS in this work), which prolongs the analysis time to a considerable extent (18 min). Therefore, in this work we also carried out a chromatographic analysis using a capillary column for fast chromatography (DB-VRX), with which it was possible to reduce the time of analysis to 6 min, such that even this second confirmation step would be rapid.

Fig. 6 shows the chromatograms corresponding to one of the polluted soils (sample 10) recorded using the two chromatographic modes, obtained with the total ion current (TIC) mode. As well as a reduction in the time of analysis, it may be seen that both chromatograms have a similar shape, typical of samples polluted with hydrocarbons. Identification of the different hydrocarbons present in the sample was accomplished with the extracted ion chromatograms. Thus, linear and branched (m/z = 57, 71, 85) hydrocarbons, cyclic alkanes (m/z = 55, 69, 83), and hydrocarbons derived from benzene (m/z = 91, 105, 119) and from naphthalene (m/z = 128, 142, 156) were identified. Table 2 shows the compounds identified in each sample, using both the HP-5MS and the DB-VRX columns, together with the retention times and the match quality between the experimental spectrum and that of the database used for identification.

Table 3
List of compounds identified by HS–fast GC–MS in the polluted soils studied

Compound	t_R[a]	Samples									
		8	9	10	11	12	14	16	17	18	19
Linear alkenes											
n-Nonane	2.669	X		X						X	X
n-Decane	2.950	X		X	X				X	X	X
n-Undecane	3.243	X		X	X				X	X	X
n-Dodecane	3.545	X	X	X	X				X	X	X
n-Tridecane	3.855	X	X	X	X				X	X	X
n-Tetradecane	4.204	X	X	X	X				X	X	X
n-Pentadecane	4.612	X		X	X				X	X	X
n-Hexadecane	5.113	X		X	X				X	X	X
n-Heptadecane	5.741	X		X	X				X	X	X
n-Octadecane	6.553										X
Branched alkenes											
3-Methylnonane	3.781					X	X			X	X
2,6,10-Trimethyldodecane	4.139		X		X	X	X	X			X
2,8-Dimethylundecane	4.466		X			X	X	X			X
2,6-Dimethylundecane	5.428		X				X			X	
2,6,11-Trimethylundecane	5.823		X		X		X	X			X
Cyclic alkenes											
Methylcyclohexane	2.349			X							
1,1,3-Trimethylcyclohexane	2.604					X	X	X			
Benzene and derivatives											
Benzene	2.260					X					
Toluene	2.445									X	X
C2-Benzene	2.688					X					
C2-Benzene	2.691	X		X	X	X				X	
C2-Benzene	2.764			X					X	X	X
C3-Benzene	2.933			X							X
C3-benzene	2.999	X		X					X	X	X
C3-Benzene	3.050	X		X						X	X
C3-Benzene	3.152	X		X						X	X
C4-Benzene	3.257	X		X					X	X	X
C4-Benzene	3.451			X						X	X
C4-Benzene	3.576			X						X	X
C5-Benzene	3.563			X						X	X
C5-Benzene	3.667									X	
Naphthalene and derivatives											
Naphthalene	3.746									X	X
C1-Naphthalene	4.144	X		X					X	X	X
C1-Naphthalene	4.217	X		X				X	X	X	X
C2-Naphthalene	4.529									X	

[a] t_R = retention time (min).

The information gained from both columns was similar, although the fast GC mode afforded less resolution power and hence the number of compounds appropriately identified was lower. However, from the data shown in the table it may be deduced that fast chromatography provides sufficient information concerning the major hydrocarbons present in the samples of polluted soils.

With these results, fast GC was used for the identification of hydrocarbons in all the soil samples from Mexico. Table 3 shows the major compounds found in the samples classified as polluted using the HS–MS methodology, with the exception of samples 2, 3, 4 and 5. It may be seen that in most samples linear, branched and cyclic alkanes and some derivatives of benzene and naphthalene were identified as major compounds. Nevertheless, in some of the samples analysed (samples 12, 14 and 16) the main source of pollution was related to branched hydrocarbons. Fig. 7 shows two chromatograms, corresponding to samples 16 and 18, representative of the two types of sample. It may be seen that the shapes of the chromatograms are clearly different. In the case of sample 18, it is patent that the major peaks are those corresponding to linear hydrocarbons, while for sample 16 the chromatogram has less well defined peaks, among which almost exclusively branched hydrocarbons are identified.

The remaining samples classified as polluted by LDA (samples 2, 3, 4 and 5) are characterised by the low intensity of the chromatographic signals, indicating a relatively low degree of pollution. The shape of the chromatograms corresponding to these samples (Fig. 8a shows the chromatogram of sample 4) is characteristic of unresolved species and it was thus impossible to perform the identification of individual hydrocarbons. Despite this, for all of them it was possible

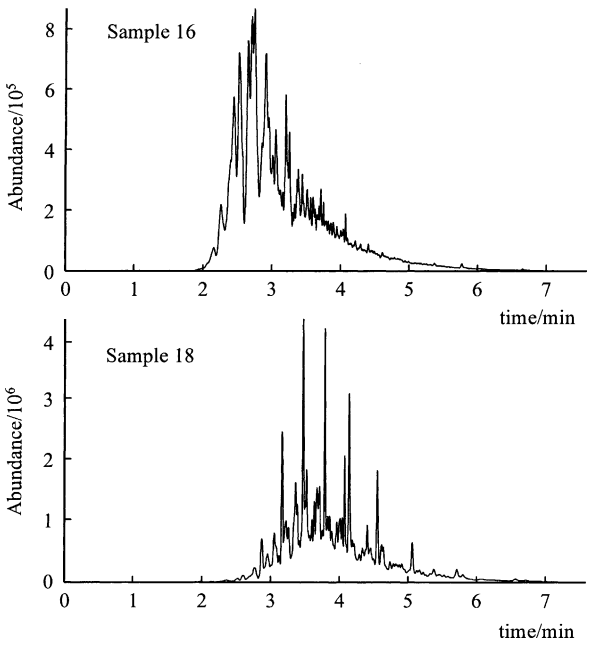

Fig. 7. Chromatograms obtained with a DB-VRX column for two samples with different shapes.

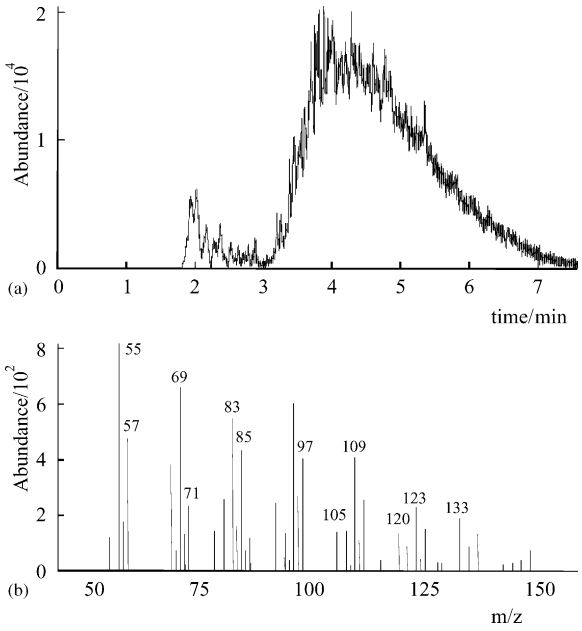

Fig. 8. (a) Chromatogram corresponding to soil 4 (unresolved hydrocarbons) and (b) average spectrum (from 1.8 to 7.5 min) of the same sample.

to obtain the mean spectrum (Fig. 8b), in which it is possible to observe the characteristic profile of samples polluted with hydrocarbons with *m/z* ratios of 57, 71, 85, 55, 69, 83, 97, 105, etc. These profiles clearly justify the inclusion of these samples within the polluted class.

The chromatograms of the unpolluted samples show very weak signals of chromatographic peaks and in no case do their mean spectra correspond to profiles typical of pollution by hydrocarbons. This allows us to affirm that all samples in which no remains of hydrocarbons were detected by HS–GC–MS were classified as unpolluted by HS–MS. The minimum concentration in the training set used to build the model for prediction by HS–MS was 1.2 mg/kg and all samples were correctly classified in the cross-validation step, so the limit for false negatives has to be $\leq$1.2 mg/kg with this technique.

4. Conclusions

The use of a headspace sampler coupled to a mass spectrometer offers a simple and effective tool for the rapid detection of contamination of soils by hydrocarbons from petroleum and derivatives. The mathematical treatment of the signals generated by the system, with a previous data normalization process, by chemometric techniques allowed the complete discrimination of polluted and non-polluted samples. The ability of the models built with laboratory-prepared samples, and measured two and a half years previously, to predict real samples points to the stability of the models generated.

The results obtained with the chemometric methods applied to the profile signal obtained by HS–MS coupling were

confirmed by fast GC mass spectrometry, which allows suitable identification of the major hydrocarbons present in the polluted samples in less than 6 min.

The results obtained allow us to affirm that the HS–MS methodology offers a powerful screening tool for the detection of pollution by hydrocarbons in soils with very different characteristics (organic and mineral soils). While PCA and HCA techniques are useful as preliminary visualization tools, LDA can be used to classify unknown samples as polluted or unpolluted. In cases in which the specific identification of individual pollutants is required, coupling with GC affords satisfactory results without excessively lengthening the time of analysis.

Acknowledgement

We acknowledge the DGICYT (Project BQU2001-1858) and the Consejería de Educación y Cultura of the Junta de Castilla y León y la Unión Europea (Fondo Social Europeo, project SA079/02) for financial support of this research.

References

[1] G. Xie, M.J. Barcelona, J. Fang, Anal. Chem. 71 (1999) 1899.
[2] Z. Wang, M. Fingas, L. Sigouin, J. Chromatogr. A 909 (2001) 155.
[3] W.J. Havenga, E.R. Rohwer, J. Chromatogr. A 848 (1999) 279.
[4] M. Llompart, K. Li, M. Fingas, Talanta 48 (1999) 451.
[5] J.C. Flórez Menéndez, M.L. Fernández Sánchez, J.E. Sánchez Uría, E. Fernández Martínez, A. Sanz-Medel, Anal. Chim. Acta 415 (2000) 9.
[6] R.T. Marsili, J. Agric. Food Chem. 47 (1999) 648.
[7] C. Pérès, C. Viallon, J.L. Berdagué, Anal. Chem. 73 (2001) 1030.
[8] V. Shiers, A.D. Squibb, Proceedings of the 5th Symposium on Olfaction and Electronic Nose, Hunt Valley, Baltimore, MD, 1998.
[9] J.L. Berdagué, C. Viallon, N. Kondjoyan, C. Denoyer, C. Thonat, Viandes Prod. Carnés 19 (1998) 78.
[10] I. Marcos Lorenzo, J.L. Pérez Pavón, M.E. Fernández Laespada, C. García Pinto, B. Moreno Cordero, J. Chromatogr. A 945 (2002) 221.
[11] C. Pérès, F. Begnaud, J.L. Berdagué, Sens. Actuators 87 (2002) 491.
[12] I. Marcos Lorenzo, J.L. Pérez Pavón, M.E. Fernández Laespada, C. García Pinto, B. Moreno Cordero, L.R. Henriques, M.F. Peres, M.P. Simões, P.S. Lopes, Anal. Bioanal. Chem. 374 (2002) 1205.
[13] M.J. Jager, D.P. McClintic, D.C. Tilotta, Appl. Spectrosc. 54 (2000) 1617.
[14] R.W. Current, D.C. Tilotta, J. Chromatogr. A 785 (1997) 269.
[15] C. Ching Chang, G. Rong Her, J. Chromatogr. A 893 (2000) 169.
[16] P. Korytár, H.-G. Janssen, E. Matisová, U.A.Th. Brinkman, Trends Anal. Chem. 21 (2002) 558.
[17] E. Matisová, M. Dömötörová, J. Chromatogr. A 1000 (2003) 199.
[18] T. Veritti, R. Sacks, Anal. Chem. 73 (2002) 3045.
[19] J.W. Cochran, J. Chromatogr. Sci. 40 (2002) 248.
[20] E. Matisová, M. Simeková, S. Hrouzková, P. Koytár, M. Dömötörová, J. Sep. Sci. 25 (2002) 1325.
[21] J.L. Pérez Pavón, M. del Nogal Sánchez, C. García Pinto, M.E. Fernández Laespada, B. Moreno Cordero, A. Guerrero Peña, Anal. Chem. 75 (2003) 2034.
[22] J.L. Pérez Pavón, M. del Nogal Sánchez, C. García Pinto, M.E. Fernández Laespada, B. Moreno Cordero, Anal. Chem. 75 (2003) 6361.
[23] Pirouette: Comprehensive Chemometrics Modeling Software, Infometrix, Ver. 3.0, Woodinville, WA, 2000.
[24] M. Forina, R. Leardi, C. Armanino, S. Lanteri, PARVUS: An Extendable Package of Programs for Data Exploration, Classification and Correlation, Ver. 1.1, Elsevier Scientific Software, 1990.