The Canadian Mineralogist
Vol. 41, pp. 1135-1146 (2003)

VISUAL–STATISTICAL CLASSIFICATION OF As–Fe-RICH PRODUCTS OF ALTERATION OF TAILINGS FROM THE ENGUIALÈS TUNSGTEN MINE, FRANCE

Jean-François LENAIN[§] and Alexandra COURTIN-NOMADE

Laboratoire HYDRASA (UMR 6532 CNRS), Faculté des Sciences et Techniques,
123, avenue Albert Thomas, F-87060 Limoges Cedex, France

Abstract

We have studied the products of alteration of sulfides contained in the tailings of a former tungsten mine at Enguialès, Aveyron, France. Most of these products are poorly crystalline and contain high concentrations of arsenic. A preliminary classification based on visual observations was carried out, and five products were distinguished according to their color and texture. The data obtained by electron-microprobe analysis were submitted to different statistical approaches in order to discriminate among the various materials and to test the efficiency of the visual classification. The statistics involved rely on logratios of the chemical elements because the ratio of any two elements is independent of the perturbations caused by other elements, and remains unchanged whatever the choice of subcompositional space. The statistical discrimination of the groups indicates that log(Fe/As) and log(S/As) best express the chemical differences. Both of these logratios correspond to the system Fe–As–S. In this study, the two approaches, visual observations and statistics, are complementary, and one method without the other could give only partial information or misinterpretations. A hierarchical clustering applied to the whole dataset confirms the existence of five groups, but visual and statistical classifications differ. On one hand, by using the logratio method, different subgroups with the same color were identified. On the other hand, some groups with different colors correspond to the same mineral, which was highlighted by Raman microspectrometry. These observations (concerning As-rich jarosite, for example) are explained mainly by differences in composition.

Keywords: mining environment, As-bearing products, iron oxyhydroxides, Raman microspectroscopy, compositional data, logratios, multivariate statistical analysis, Enguialès mine, France.

Sommaire

Les haldes d'une ancienne mine de tungstène à Enguialès, Aveyron, en France, présentent des produits d'altération de sulfures, pour l'essentiel faiblement cristallins, et contenant de fortes teneurs en arsenic. Ces produits d'altération ont été préalablement classés de manière visuelle, c'est-à-dire en fonction de leur couleur et de leur texture, et cinq groupes ont été ainsi définis. Les données d'analyses en microsonde électronique obtenues sur ces matériaux ont été soumises à différentes approches statistiques dans le but de les discriminer de manière plus précise, et de tester l'efficacité de la classification visuelle. Les méthodes mises en oeuvre utilisent le log des rapports entre éléments chimiques, car tout rapport de deux éléments est indépendant des perturbations causées par les autres éléments, et demeure inchangé quelle que soit la sous-composition considérée. L'analyse discriminante des groupes visuels a montré que ce sont log(Fe/As) et log(S/As) qui expriment le mieux les différences chimiques entre groupes. Ces deux logs de rapports correspondent au système Fe–As–S. Dans cette étude, nous montrons que les deux approches (visuelle et statistique) sont complémentaires, et qu'une méthode sans l'autre peut conduire à des interprétations partielles ou érronées. Une classification hiérarchique réalisée sur l'ensemble des données a aussi mis en évidence l'existence de cinq groupes; cependant, les classifications visuelle et statistique diffèrent. Différents sous-groupes de même couleur ont été différenciés de manière statistique, tandis que certains groupes de couleurs différentes correspondraient en fait à un même minéral, ce qui a été confirmé par microspectrométrie Raman. Ces observations (concernant la jarosite riche en As par exemple) s'expliquent en grande partie par des différences de composition.

Mots-clés: environnement minier, produits arséniés, oxyhydroxydes de fer, microspectroscopie Raman, données de composition, log de rapports, analyse statistique multivariée, mine d'Enguialès, France.

[§] *E-mail address*: jflenain@unilim.fr

INTRODUCTION

The characterization of minerals or phases is usually carried out *via* classical techniques, such as X-ray diffraction (XRD), scanning electron microscopy (SEM), electron microprobe (EMP) analysis. The chemical data obtained by EMP allow geoscientists to determine the precise composition and also provide information on the homogeneity (or lack thereof) of the materials. However, several problems can be encountered with EMP analytical data depending on the choice of subcompositional space, substitutions, solid solutions, and even window dimensions that exceed crystal size. Furthermore, simple consideration of the raw EMP data is difficult in the case of amorphous or poorly crystalline products and unidentified minerals. Thus there are problems in correctly interpreting EMP data.

In this study, the logratio statistical methods, developed by Aitchison (1986) for compositional data, are applied to some by-products of mining. The alteration of sulfides contained in the tailings can generate amorphous or poorly crystalline products that cannot be well identified by direct methods of mineralogical characterization. In previous studies of the alteration products resulting from the weathering of tailings, we pointed out problems concerning the precise definition of their composition (Courtin-Nomade *et al.* 2003, Courtin-Nomade 2001). These studies allowed us to distinguish products of alteration according to visual criteria (*e.g.*, color, texture). Electron-microprobe analyses indicated wide chemical variations within a single visual object. Furthermore, the analytical total generally never reaches 100% owing to a high degree of hydration. The aim of this study is (i) to show the advantages of a statistical approach applicable to compositional data of heterogeneous, amorphous or poorly crystalline products, (ii) to provide more information on a previous visual classification by testing the homogeneity of the different groups according to the chemical composition of the micrometric products of alteration, and (iii) to give an overview of the links among these materials in their evolution at the expense of the sulfide minerals.

MATERIALS AND METHODS

Characterization

Chemical analyses were carried out with a CAMECA SX50 electron microprobe equipped with four wavelength-dispersion spectrometers (WDS) and an energy-dispersion spectroscopy system (EDS; CAMPARIS, Paris). The acceleration voltage used was 15 kV, with a beam current of about 4 nA. The elements sought were: As, Fe, S, Si, Al, K, Ca, Mn, Mg, and Na. The materials were also characterized by Raman microspectroscopy with a XY 800 Dilor instrument (CRSCM, Orléans). As a source of radiation for Raman spectroscopy, we used a green laser ($\lambda = 514.5$ nm) with a low output of energy, about 5 to 10 mW, to avoid local heating of the sample. About 10% of the energy of the laser reached the sample through an Olympus microscope using a ×100 objective. The scans were usually collected from 60 to 1579 cm^{-1} and from 3000 to 3900 cm^{-1}, with a scan time of 15 minutes (300 s per window). These two methods provide complementary information: EMP data for the chemical composition, and Raman microspectroscopy for the precise mineralogical identification of the species.

Materials

The materials studied are the result of the alteration of sulfide-rich tailings from an abandoned tungsten mine at Enguialès, Département de l'Aveyron, France. The main sulfide minerals are pyrite and arsenopyrite. Five different As- and Fe-rich products of alteration were identified (Courtin-Nomade *et al.* 2003) according to their color and luster: dark red (DRD), red (RED), fleecy (FLE), yellow (YEL) and resinous (RES). That study pointed out the heterogeneity of some of the species in these groups, especially the red (goethite, lepidocrocite or amorphous iron oxyhydroxides), the fleecy (jarosite or clay minerals), and the yellow materials (goethite or amorphous iron oxyhydroxides). The dark red materials correspond to hematite, and the resinous materials, to As-rich amorphous products with a well-defined chemical composition, close to $Fe_7(AsO_4)_4(OH)_9$ as a first approximation.

The data used in this study come from analyses carried out on several thin sections. In the EMP analyses, we detected no significant concentrations of Mn, Na, Mg and Ca in the materials considered, or these concentrations were below the limit of detection. Therefore, these elements were not taken into account in the following study. Concerning Si, it was very commonly detected in high concentrations. However, the presence of Si in our case is due to the presence of primary silicate minerals, and the As–Fe-rich by-products we intend to study invariably coat these silicates. Thus we consider Si as a "pollutant" element, *i.e.*, the measured concentrations do not reflect the presence of Si in the secondary products, but the beam used to analyze the products is too wide or penetrates too deeply. Thus Si measured only corresponds to the adjacent silicates. In order to avoid the influence of Si on the data during the statistical procedures, we chose to consider only the subcomposition [As, S, Al, K, Fe], according to the definition of Aitchison (1986), which represents 565 data. Among them, 21 were found to be very scattered and far from the remaining values. These outliers were discarded before carrying out the satistical analysis.

Statistical procedures

The particular nature of compositional data and the statistical methods used in this study are presented here

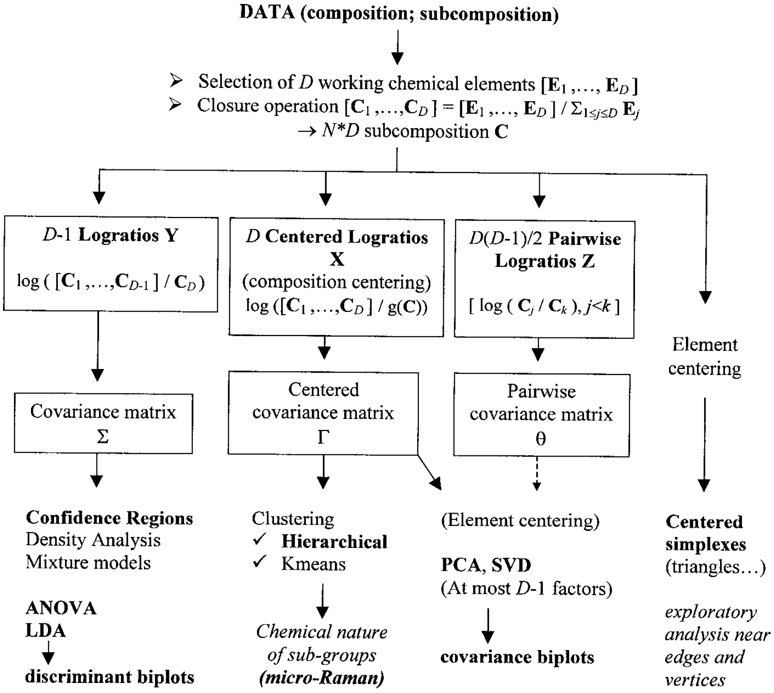

FIG. 1. Summary of the main statistical methods that can be used on the EMP data. The methods chosen for this study are in bold.

and summarized in Figure 1. The logratio transformations enable the use of standard multivariate methods that all statistical packages contain: we employed S–PLUS (Insightful Corp. 2002) for the clustering of all the data, and SYSTAT® (SPSS 2000) in the other cases. We also used some personal routines in MATLAB® language (The MathWorks 2000) largely based on the NEWCODA toolbox provided by Aitchison (2000).

Nature of the compositional data. The compositional data are closed; this mathematical dependence leads to spurious correlations and makes covariance matrices unexploitable (Aitchison 1986, Hinkle & Rayens 1995, Pawlowsky-Glahn & Egozcue 2001). In fact, the true characteristic of the compositional data is their scale invariance (Aitchison 2001, Barceló-Vidal *et al.* 2001), so that the data are still compositional even if their sum is not constant. This may happen for analytical reasons (hydration of the sample, incompatibility in modes of measurement) or because of the choice of elements (subcomposition). Normalization allows all compositions to be represented on the same plane, the simplex. Other drawbacks are the non-linear relationships in triangular diagrams, squeezed data and skewed distributions near edges, and proportional errors (Aitchison 1986, 2001, Eynatten *et al.* 2002). Moreover, Euclidian distances and linear methods are not valid in a simplex because results may lie outside this closed space (Weltje 2002).

The logratio transformations. They offer several benefits; distributions become more symmetrical, variances become independent of means, ratios become differences, and proportional errors become additive. Consequently, different scales of proportions can be jointly used, so that major, minor, and trace elements and their variances may be compared in a single analysis. The mean of logs corresponds to the geometrical mean of original proportions: it better describes the center of a group than the arithmetic mean (Philip & Watson 1988). Only Euclidian and Mahalanobis distances calculated after these transformations have the required properties for compositional data (Martín-Fernández *et al.* 1998). Three logratios matrices contain the same information: the matrix **Z** of all pairwise logratios, the matrix **Y** of additive logratios (ALR), and the matrix **X** of centered logratios (CLR), with covariance matrices θ, Σ, and Γ, respectively. For example, from a four-element composition [A, B, C, D], we get: **Z** = log([A/D, B/D, C/D, A/C, B/C, A/B]), **Y** = log([A/D, B/D, C/D]), **X** = log([A/g, B/g, C/g, D/g]), where g is the geometric mean of [A, B, C, D]. The choice of the working matrix depends on some properties like invariance, distances, and inversibility (Aitchison 1986).

Exploratory methods. With Principal Component Analysis (PCA), or Singular Value Decomposition (SVD), compositions and logratios may then be displayed on a single "biplot" (Aitchison 1986, Aitchison & Greenacre 2001). Such biplots are useful in the choice of some ways to display the data, *e.g.,* a triangular diagram representation that best shows the variability in the data. Here, **Z** and **X** give equivalent biplots, but **X** is the working matrix; the CLR biplot only displays *D* variables, and logratios are the links between the CLR apices (Aitchison & Greenacre 2001). Biplots have been applied, for example, to fluid variation in petroleum (Whelan *et al*. 2001) and marine micropaleontology (Kucera & Malmgren 1998). Hinkle & Rayens (1995) extended the method to Partial Least-Squares Regression and applied it to organic chemistry and sedimentology. Based on the example of olivine compositions, Aitchison & Thomas (1998) have also shown the importance of near-zero eigenvalues; they correspond to almost-constant eigenvectors. Returning to proportions, they can reveal typical relations between elements under the form of pseudo-equilibrium constants (Aitchison 1999).

Methods of classification. No difficulty arises in using rules of classification found in standard statistical packages; the Aitchison distance between compositions is the Euclidian distance between CLR. The Ward method of Hierarchical Clustering (HC) gives compact clusters and clear hierarchies (Martín-Fernández *et al.* 1999, Martín-Fernández 2001). Pawlowsky-Glahn & Buccianti (2002) applied cluster analysis to the chemical composition of a fumarole on the Island of Vulcano, Italy. Mixture models also give excellent results, with overlapping groups (Barceló-Vidal *et al*. 1999, Martín-Fernández *et al*. 1997).

Analysis of Variance (*ANOVA*) *and Linear Discriminant Analysis* (*LDA*). **Y** is the working matrix and produces identical results whatever the choice of the reference element (Aitchison 1986). In most cases, a major element or oxide is chosen as the reference, *e.g.*, SiO_2 (Clarke *et al*. 1989, Verrucchi & Minissale 1995). Ridenour (1999) applied these methods to hydrology, and Guitart *et al*. (1999) made biochemical comparisons of two marine vertebrates. Only multivariate tests must be considered, because individual tests only concern the corresponding binary subcompositions. As a consequence, the selection of variables must be replaced by a choice of subcomposition. This choice was done using discriminant biplots. The correlations between the variables (here the CLR) and the discriminant factors are superimposed on the compositional plot. The best represented CLR on the correlation circle are candidates for formation of a discriminant subcomposition.

Confidence regions (*CR*). A recent review covers this subject extensively (Weltje 2002). CR can be drawn from **Y** by using multivariate normal models (Weltje 2002) or kernel estimates (Aitchison & Lauder 1985, Aitchison 1986). For a three-element subcomposition, the CR can be displayed in a triangular diagram using the inverse ALR transformation (Weltje 2002). Pawlowsky-Glahn & Buccianti (2002) used this representation with data on fumaroles.

RESULTS

Bulk compositions

The first attempt was to compare visual and statistical classifications based on electron-microprobe data. Unlike Linear Discriminant Analysis (LDA), Hierarchical Clustering (HC) is not based on a predefined classification. It is an adequate method of evaluation if some "natural" populations can be distinguished, without preconceived ideas. Euclidian distances between compositions were calculated following the Ward method in S–PLUS from the five centered logratios **X** = CLR(S, As, Al, K, Fe), and used to construct the dendrogram in Figure 2. A distance equal to 17 enabled the creation of five clusters that were compared with the five predefined visual groups by cross-tabulation (Table 1a). It seemed first that the three visually most important groups, FLE, RED and RES, were split into several clusters. This indicates that the one color may correspond to many chemical entities, and that the homogeneity of visually recognized groups has to be verified. Cluster 1, an iron-rich composition, groups together DRD, the greatest part of RED and a small part of FLE. Cluster 4 groups together another major part of RED and FLE, which seems to correspond to As-rich compositions.

The second stage was to determine the most discriminant ratios as well as the most discriminant triangular diagram with the help of LDA and of a discriminant biplot. The variables involved for LDA are the four logratios **Y** = ALR(S, As, Al, K, Fe), whereas the correlations between the CLR **X** and the discriminant factors (scores) are calculated and superimposed to form the biplot (Fig. 3a). The links S–As and As–Fe are nearly

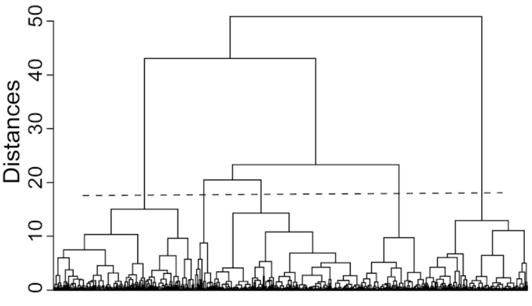

FIG. 2. Hierarchical clustering of the 544 EMP data (using Ward method and Euclidian distance on the five centered logratios). The dotted line defines five clusters.

orthogonal in As and best express the variance between the groups. In this way, we have considered the orthogonal logratios log(S/As) and log(Fe/As) (Fig. 3b), and the corresponding triangular diagram Fe–As–S (Fig. 3c). The subcomposition [S, As, Fe] best explains the visual classification. The Ward method applied to the three centered logratios **X** = CLR(S, As, Fe) shows that the RES group is highly distinct from the others (Table 1b).

This visual classification, which seemed satisfactory on the whole as a first approach, remains too imprecise. The presence of very different groups having the same color is now clearly observed. This is the case for two of the five groups: the YEL and the RED groups. Two sets could be identified for each of them: YEL_1 and YEL_2, and RED_1 and RED_2, respectively. A clustering carried out on these sets confirmed this first observation (Fig. 4). Furthermore, with the five visually defined groups, this global approach allows one to identify two main end-members, an "iron" and an "arsenic" one. According to these previous results the RED_2, YEL_2 and DRD can already be classified in the "iron" end-member, whereas the RED_1, YEL_1, FLE and RES belong to the "arsenic" end-member.

Group-by-group analysis

The "iron" end-member. ANOVA and tests use the ALR matrix **Y**, as does LDA (Table 2a). Considering all the variables, the chemical identity of the three subgroups RED_2, YEL_2 and DRD was confirmed: all differences were found to be significant. However, RED_2 and YEL_2 are not significantly different, considering the subcomposition [As, S, Fe] (Table 2b). The LDA of the three groups (not shown here) indicates that the link As–Fe is the most discriminant variable, and that the second factor corresponds to the variation of K/S and Al/S logratios in YEL_2. A second test was performed, but this time only with the binary composition [As, Fe], *i.e.,* with log(Fe/As) variable. Tukey pairwise comparisons were used in this univariate case (Table 2c), allowing a recovery between the RED_2 and YEL_2 groups to be observed; RED_2 and YEL_2 show no difference, whereas the test is significant in comparison to DRD. This recovery can be explained in mineralogical terms. Analyses by Raman microspectroscopy show that RED_2 and YEL_2 both correspond to goethite (α-FeOOH) (Fig. 5a), whereas the DRD group corresponds to hematite (α-Fe_2O_3; spectrum not shown). These results indicate that two compounds can have the same mineralogy, a close chemical composition, but a different color. The main chemical difference between the RED_2 and YEL_2 groups is due to K/S and Al/S variations, whereas the main difference among the three groups of the "iron" end-member is explained by variations in Fe/As.

The "arsenic" end-member. This end-member comprises distinct populations, each of which contains various amounts of arsenic. RES and YEL_1 groups seem homogeneous and well defined if we consider the subcomposition [As, S, Fe] (Fig. 3a). These groups are also the richest in arsenic. However, YEL_1 and also RED_1 are superimposed on the remaining and widely dispersed group FLE. Thus the major questions involve the homogeneity of FLE and its relation to RED_1 and YEL_1. Mineralogical studies indicate that both RED_1 and YEL_1 correspond to Fe-amorphous products enriched in As (am-FeOOH ± As). However, they are chemically very different (Table 3). An average of their composition in terms of iron and arsenic content, the main components of these two groups, can be estimated by their Fe/As values, which indicate, respectively, for RED_1 and YEL_1: Fe/As ≈ 10.43 (geometric mean $As_{(g)} \approx 7.24\%$ in the composition [Al, K, S, As, Fe]) and Fe/As ≈ 4.08 ($As_{(g)} \approx 18.46\%$) (Table 4). Thus the main difference between RED_1 and YEL_1 is chemical and not mineralogical, which can be explained by two different

TABLE 1. VISUAL GROUPS BY CLUSTER FREQUENCIES

Visual group	Color	Cluster 1	2	3	4	5	Total
		(a) **with subcomposition [As, S, Al, K, Fe]**					
FLE	fleecy	18	38	8	37	1	102
YEL	yellow	10	7	23	0	8	48
RED	red	93	8	14	37	6	158
DRD	dark red	40	1	0	0	7	48
RES	resinous	0	115	73	0	0	188
Total		161	169	118	74	22	544
		(b) **with subcomposition [As, S, Fe]**					
FLE	fleecy	15	2	50	2	33	102
YEL	yellow	9	10	29	0	0	48
RED	red	56	46	24	0	32	158
DRD	dark red	30	18	0	0	7	48
RES	resinous	2	0	7	179	0	188
Total		112	76	110	181	65	544

TABLE 2. MULTIVARIATE TESTS OF BETWEEN-GROUP EQUALITY HYPOTHESES

Hypothesis	Wilks' Lambda	F-Statistic	df	probability
(a) **with subcomposition [As, S, Al, K, Fe]**				
$RED_2 = YEL_2$	0.778	11.608	4, 163	**0.000**
$RED_2 = DRD$	0.866	6.297	4, 163	**0.000**
$DRD = YEL_2$	0.722	15.671	4, 163	**0.000**
(b) **with subcomposition [As, S, Fe]**				
$RED_2 = YEL_2$	0.992	0.657	2, 165	0.520
$RED_2 = DRD$	0.884	10.818	2, 165	**0.000**
$DRD = YEL_2$	0.930	6.192	2, 165	**0.003**

(c) **with subcomposition [As, Fe]; matrix of pairwise comparison probabilities**

	YEL_2	RED_2	DRD
YEL_2	1		
RED_2	0.709	1	
DRD	**0.001**	**0.000**	1

Highly significant probabilities are shown in bold.

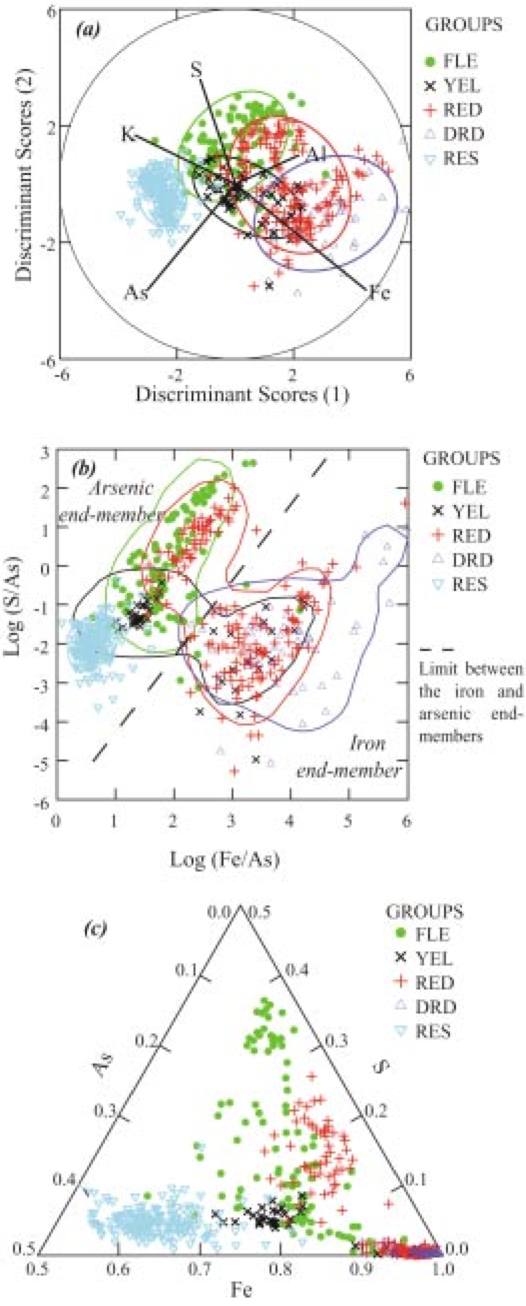

FIG. 3. (a) Linear Discriminant Analysis (LDA) biplot of the five visual groups FLE (fleecy), YEL (yellow), RED (red), DRD (dark red), and RES (resinous), with 80% normal confidence region. LDA is based on additive logratios, whereas elements are displayed as correlations between centered logratios and discriminant factors. (b) Representation of the logratios log(S/As) / log(Fe/As), with 80% kernel confidence region. (c) Corresponding triangular diagram Fe–As–S.

components from distinct sources. RED_1 is probably a result of the alteration of pyrite, whereas YEL_1 comes from the alteration of arsenopyrite.

The dispersion of the FLE group was analyzed by clustering. The dendrogram obtained for this group clearly shows four subgroups (Fig. 6), and the PCA enables the main axes of variation (Fig. 7) to be determined. The two major subgroups FLE_{1a} and FLE_{1b} are noted, whereas the less important are FLE_2 and FLE_3.

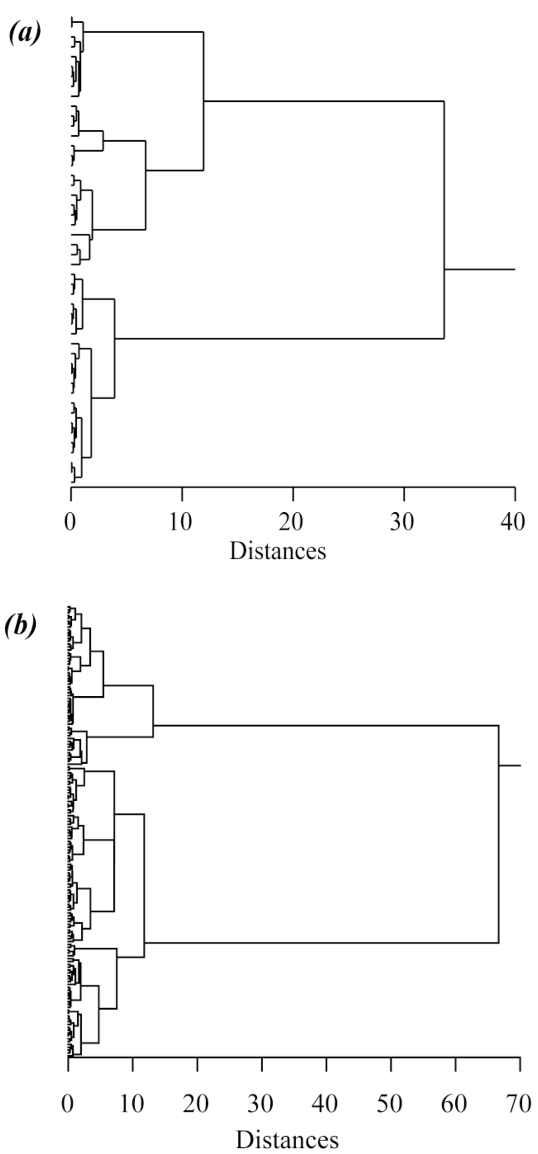

FIG. 4. Cluster tree of (a) the YEL (yellow) group indicating the presence of two subgroups, and (b) the RED (red) group showing also a division into two subgroups.

The subgroup FLE_1 shows a slight dispersion, which essentially reflects a chemical variation due to As concentrations. FLE_{1a} and FLE_{1b} correspond to two end-members, one rich in As, with the As geometric mean $As_{(g)} \approx 8.3\%$ (FLE_{1b}), and an other with smaller amounts of As, with $As_{(g)} \approx 6.0\%$ (FLE_{1a}) (Fig. 7, Table 4). Mineralogically, according to Raman microspectroscopy, we determined that FLE_{1a} and FLE_{1b} both correspond to jarosite (Fig. 5b) that contains adsorbed or coprecipitated As (substitution of AsO_4 for SO_4 can easily occur in the structure of jarosite: Paktunc & Dutrizac 2003). No Raman spectra were obtained for the second subgroup, FLE_2, involving poorly crystalline (*e.g.,* clay minerals) or non-crystalline (amorphous) products. According to a previous study (*e.g.,* XRD), the FLE_2 subgroup corresponds to clay minerals; they are known to have an affinity for As, as is observed here with an average concentration of $As_{(g)} \approx 9.9\%$ (Table 4). Finally, the FLE_3 subgroup corresponds to misclassified data and could preferentially be classified with the YEL_1 group. The test indicates no significant differences between these two groups (Table 3).

Statistical observations carried out on the FLE group allow three different subgroups to be clearly identified, whereas the visual classification does not distinguish them. Furthermore, the cluster analysis performed on the RED_1, YEL_1 and FLE (1, 2 and 3) groups indicates that they are well classified except for the RED_1, which could easily be confused with the FLE_1 group (Table 5). However, a visual classification enabled discrimination between these two groups. Thus visual classification appears essential in this case. The study of these three groups shows the complementarity of the two approaches (statistical and visual); one method without the other would likely give partial interpretations or misinterpretations.

Contrary to the two previous groups, RES is, after YEL_1, the least scattered group in terms of the [As, S,

TABLE 3. MULTIVARIATE TESTS OF THE HYPOTHESIS $YEL_1 = FLE_3$

Subcomposition	Wilks' Lambda	F-Statistic	df	probability
[As, S, Al, K, Fe]	0.785	2.185	4, 32	0.093
[As, S, Fe]	0.985	0.263	2, 34	0.771

This hypothesis cannot be rejected.

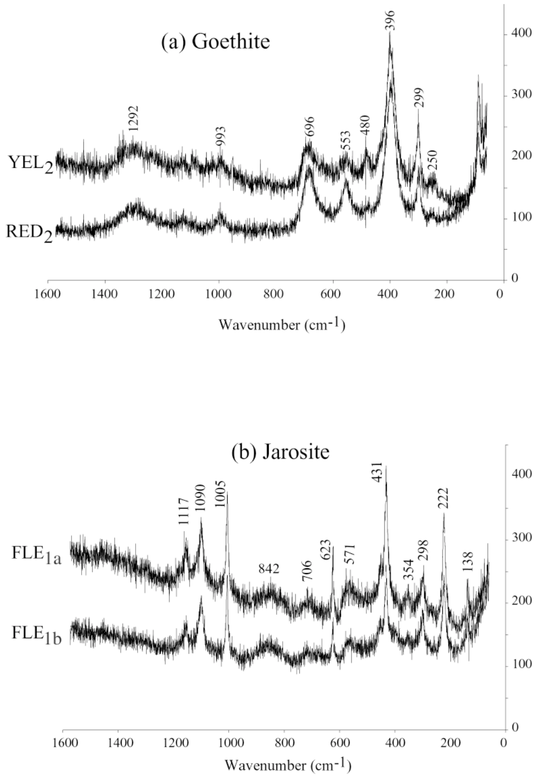

FIG. 5. Raman spectrum of (a) the YEL_2 (yellow subgroup 2) and the RED_2 (red subgroup 2) products, both corresponding to goethite; (b) the $FLE_{1a,b}$ (fleecy subgroup 1) product identified as jarosite, more or less enriched in arsenic.

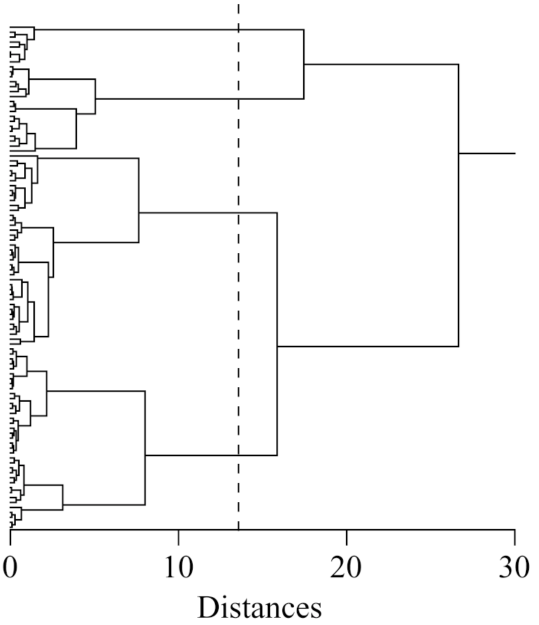

FIG. 6. Cluster tree on the FLE (fleecy) group. The subdivision defines four subgroups.

Fe] subcomposition (Fig. 3). This fact is confirmed by PCA (or SVD) analysis of the group. There is no evidence for a subgroup structure in spite of the chemical variations due to Al and K.

A principal component analysis reveals a near-zero eigenvalue (the last one), which corresponds to an almost constant eigenvector with equal contributions of As and Fe (Table 5) and a negligible contribution of S. Following Aitchison (1999), we deduce that the ratio Fe/As is the fundamental characteristic of this group. RES is the richest in arsenic, with an average Fe/As ratio of 1.9 (Table 4); RES is also the only group of the "arsenic" end-member where S varies independently of As (and Fe). This characteristic is better displayed with a centered triangular diagram, which prevents a large part of the data squeeze near the triangle edges (Fig. 8). This transformation retains straight lines (Eynatten *et al.* 2002).

Evolutionary model

The precise definition of all of these groups enables us to identify their source components, pyrite or arsenopyrite. It is well known that these minerals can generate secondary products such as jarosite, goethite and poorly crystalline materials, such as ferrihydrite (Blowes *et al.* 1994, Jambor & Dutrizac 1998, Savage *et al.* 2000). All of these products have greater or lesser affinities with arsenic and can trap it (adsorption or coprecipitation: Manceau 1995, Waychunas *et al.* 1995, Lumsdon *et al.* 2001). A statistical analysis coupled with a microspectroscopy technique allow us to evaluate the alteration of the primary minerals and to specify what type of secondary products are generated.

From the previous results, we contend that the FLE_{1a} and RED_1 subgroups correspond to the first products of alteration of pyrite, whereas the FLE_{1b}, YEL_1 and RES groups ("arsenic" end-member) result from the primary stage of alteration of arsenopyrite (Fig. 9). Furthermore, these products have themselves been altered and have generated more "evolved" materials, which represent the "iron" end-member corresponding to crystalline end-products. Thus this evolution indicates a subsequent release of As with time and degree of alteration.

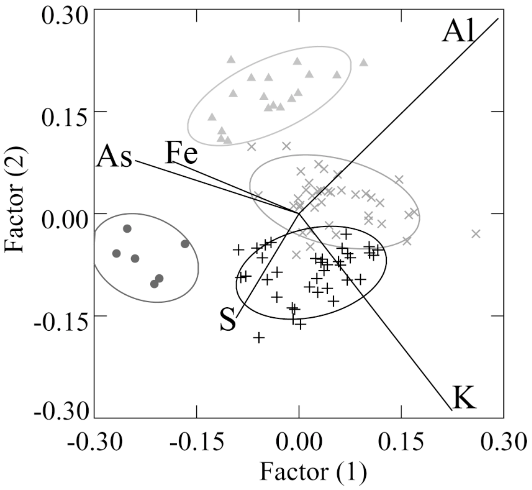

FIG. 7. Principal Component Analysis (PCA) covariance biplot of the FLE (fleecy) group, with representation of the four subgroups defined by Hierarchical Clustering and their corresponding 80% normal confidence regions. The four subgroups are therefore well identified: $FLE_{1a,b}$, FLE_2 and FLE_3. Clusters: • FLE_3, × FLE_{1b}, + FLE_{1a}, ▲ FLE_2.

TABLE 4. GEOMETRIC CENTERS IN % OF SUBCOMPOSITION [As, S, Al, K, Fe]

Subgroup	**As** center	σ_{CLR}	**Al** center	σ_{CLR}	**S** center	σ_{CLR}	**K** center	σ_{CLR}	**Fe** center	σ_{CLR}	Fe:As ratio
DRD	1.73	1.07	3.27	1.33	0.26	0.72	0.12	1.00	94.61	0.70	54.66
RED_1	7.24	0.67	2.18	0.99	13.24	0.53	1.75	0.75	75.57	0.47	10.43
RED_2	2.94	0.69	4.11	0.82	0.39	0.66	0.16	0.85	92.40	0.48	31.44
YEL_1	18.46	0.66	0.43	1.23	5.40	0.58	0.34	0.85	75.36	0.56	4.08
YEL_2	3.08	1.04	11.88	0.51	0.30	0.71	0.93	1.61	83.80	0.56	27.20
FLE_{1a}	5.91	0.63	6.51	0.67	22.51	0.29	8.06	0.42	57.01	0.30	9.65
FLE_{1b}	10.56	0.65	22.35	0.55	6.13	0.53	5.68	0.64	55.27	0.46	5.23
FLE_2	7.35	0.60	17.58	0.88	3.02	0.94	0.17	0.71	71.87	0.53	9.77
FLE_3	19.44	0.40	0.13	0.71	5.75	0.44	0.37	0.51	74.29	0.43	3.82
RES	30.76	0.63	3.13	1.71	4.20	0.72	3.33	0.52	58.56	0.57	1.90

Centers are calculated as $100^*[g(As),\ldots,g(Fe)]/\sum_j g_j$, where g is the geometric mean concentration of each element. σ_{CLR} values are the standard deviations of the centered logratios.

CONCLUSIONS

In this study, we have shown that a preliminary visual classification of the materials is necessary to separate the major products of alteration. Nevertheless, a statistical analysis enables this classification to be refined in order to obtain more homogeneous chemical and mineralogical groups. This approach is validated by confirming some of the groups identified using specialized techniques of characterization such as Raman microspectroscopy. The complementarity of the two approaches is remarkable: only the visual classification allows one to define the YEL_1 and the RED_1 groups (both corresponding to am-FeOOH but with different concentrations of As), which are chemically very close to some other minerals groups. A statistical analysis easily allows us to differentiate some groups with the same color and to detect the presence of misclassified individuals.

As already observed in some other studies, Aitchison's logratio method appears to be very efficient. This model allows most of the usual methods of statistical analysis to be used, provided that all kinds of linear relationships between proportions are forgotten. It allows a general model of the different As-bearing products of evolution to be established and a discrimination of the precursor minerals (arsenopyrite and pyrite).

TABLE 5. PRINCIPAL COMPONENT ANALYSIS OF THE COVARIANCE MATRIX OF RES (RESINOUS GROUP)

(a) **subcomposition [As, S, Al, K, Fe]**

	Principal factors (eigenvectors)			
	1	2	3	4
	Percent of total variance explained			
	89.614	7.196	2.776	0.415
Centered logratios	Component loadings			
CLR (As)	0.647	0.128	0.146	-0.094
CLR (Al)	-1.704	0.158	-0.014	-0.005
CLR (S)	0.653	0.089	-0.289	0.001
CLR (K)	-0.118	-0.507	0.018	0.000
CLR (Fe)	0.522	0.132	0.140	0.099

(b) **subcomposition [As, S, Fe]**

	Principal factors (eigenvectors)	
	1	2
	Percent of total variance explained	
	83.081	16.919
Centered logratios	Component loadings	
CLR (As)	0.137	-0.117
CLR (S)	-0.293	0.005
CLR (Fe)	0.156	0.112

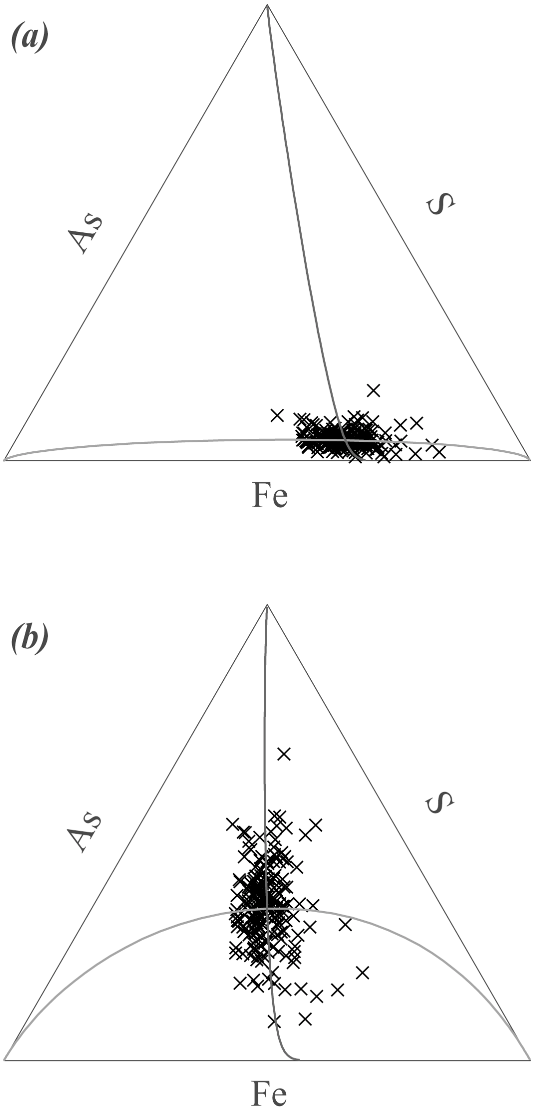

FIG. 8. Original (a) and centered (b) triangular diagrams of the RES (resinous) group. The centering moves the geometric center $[g(As), g(S), g(Fe)] / \Sigma_j g_j$ of the group to the geometric center of the triangle (g = geometric mean). The principal factors defined in the logratio space are also displayed by inverse transformation. The independent variation of S is clearly revealed by centering.

ACKNOWLEDGEMENTS

We gratefully acknowledge the PDZR program for financial support. Special thanks to Dr. J.M. Bény of the CRSCM–CNRS, ISTO/Orléans, for the Raman

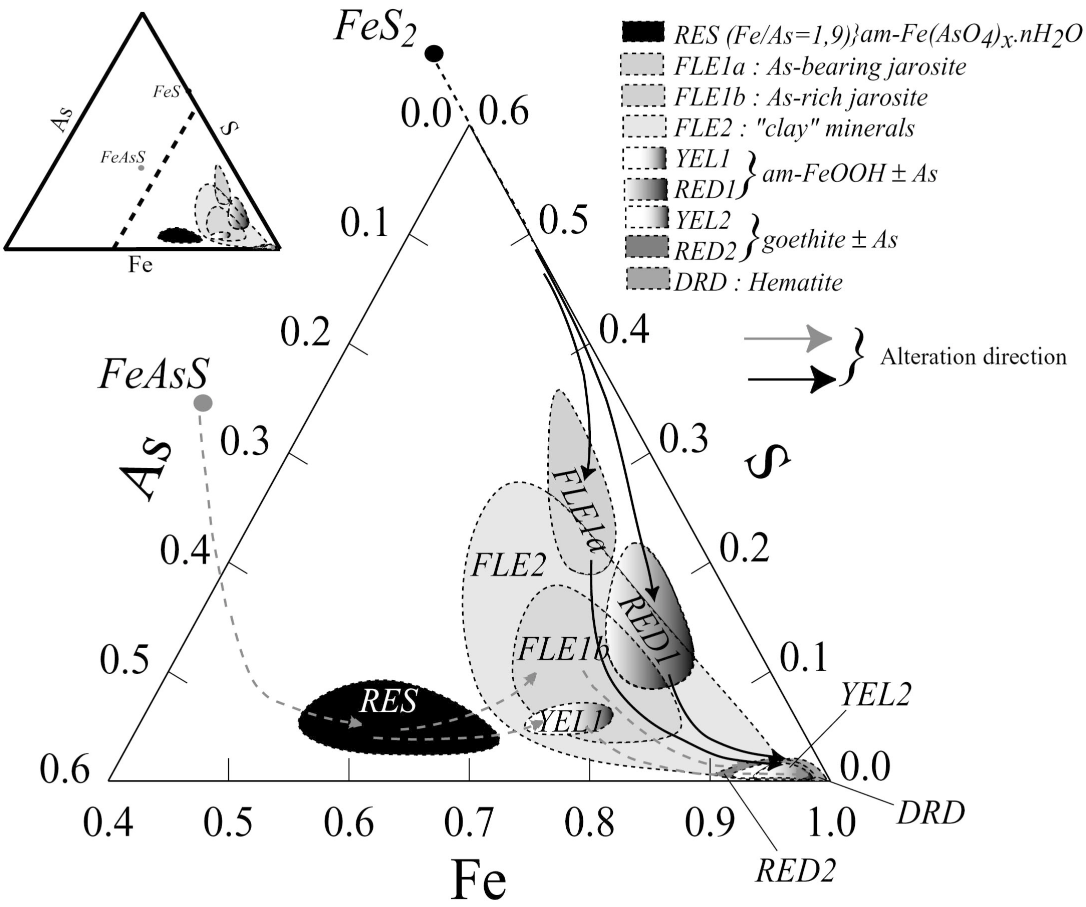

FIG. 9. Final triangular diagram Fe–As–S indicating the evolution of the newly formed materials from the precursors arsenopyrite (FeAsS) and pyrite (FeS$_2$) to the more evolved products. Each group is represented by its 80% normal confidence region in the logratio space, back-transformed in the triangle.

microspectroscopy analyses, and to Dr. J. Aitchison, who kindly provided us the NewCoda toolbox and preprints at the beginning of this study. Professor H. Bril and Dr. K. Ludwig also provided very helpful comments on an early draft of this paper. Constructive reviews and numerous comments from Drs. R.F. Martin and M. Raudsepp were gratefully appreciated, as well as those from Dr. Chang-jo Chung and an anonymous reviewer.

REFERENCES

AITCHISON, J. (1986): *The Statistical Analysis of Compositional Data*. Chapman and Hall, New York, N.Y.

________ (1999): Logratios and natural laws in compositional data analysis. *Math. Geol.* **31**, 563-580.

________ (2000): NEWCODA, a software package for compositional data analysis. (a Matlab® toolbox available from the author).

________ (2001): Simplicial inference. *In* Algebraic Methods in Statistics, Contempory Mathematics Series (M. Viana & D. Richard, eds.). American Mathematical Society, New York, N.Y.

________ & GREENACRE, M. (2001): Biplots of compositional data. *Economics Working Papers* **557**. Department of Economics and Business, Universitat Pompeu Fabra, Spain.

________ & LAUDER, I.J. (1985): Kernel density estimation for compositional data. *Applied Statistics* **34**, 129-137.

________ & THOMAS, C.W. (1998): Differential perturbation processes: a tool for the study of compositional processes. *In* Proc. of IAMG'98, the Fourth Annual Conference of the Int. Association for Mathematical Geology II (A. Buccianti, G. Nardi & R. Potenza, eds.). De Frede Editore, Naples, Italy (499-504).

BARCELÓ-VIDAL, C., MARTÍN-FERNÁNDEZ, J.A. & PAWLOWSKY-GLAHN, V. (2001): Mathematical foundations of compositional data analysis. *In* Proc. of IAMG'01, the Seventh Annual Conference of the Int. Association for Mathematical Geology. Cancún, México (CD-ROM).

________, PAWLOWSKY-GLAHN, V. & BOHLING, G. (1999): Classification of compositional data using mixture models: a case study using granulometric data. *In* Computerized Modeling of Sedimentary Systems (J. Harff, W. Lemke & K. Stattegger, eds.). Springer-Verlag, Heidelberg, Germany (389-400).

BLOWES, D.W., PTACEK, C.J. & JAMBOR, J.L. (1994): Mineralogy of sulfide-rich tailings and their oxidation products. *In* The Environmental Geochemistry of Sulfide Mine-Wastes (J.L. Jambor & D.W. Blowes, eds.). *Mineral. Assoc. Can., Short Course Handbook* xxx, 59-102.

CLARKE, D.B., REARDON, N.C., CHATTERJEE, A.K. & GREGOIRE, D.C. (1989): Tourmaline composition as a guide to mineral exploration: a reconnaissance study from Nova Scotia using discriminant function analysis. *Econ. Geol.* **84**, 1921-1935.

COURTIN-NOMADE, A. (2001): *Mobilité de l'arsenic, liaisons arsenic–fer et spéciation de l'arsenic dans les haldes d'anciennes mines du Massif Central français*. Thèse de doctorat, Univ. Limoges, Limoges, France.

________, BRIL, H., NÉEL, C. & LENAIN, J.F. (2003): Arsenic in iron cements developed within tailings of a former metalliferous mine – Enguialès, Aveyron, France. *Appl. Geochem.* **18**, 395-408.

EYNATTEN, H., PAWLOWSKY-GLAHN, V. & EGOZCUE, J.J. (2002): Understanding perturbation on the simplex: a simple method to better visualise and interpret compositional data in ternary diagrams. *Math. Geol*. **34**, 249-257.

GUITART, R., MARTINEZ SILVESTRE, A., GUERRERO, X. & MATEO, R. (1999): Comparative study on the fatty acid composition of two marine vertebrates: striped dolphins and loggerhead turtles. *Comp. Biochem. Physiol. B*. **124**, 439-443.

HINKLE, J. & RAYENS, W. (1995): Partial least squares and compositional data: problems and alternatives. *Chemometrics and Intelligent Laboratory Systems* **30**, 159-172.

INSIGHTFUL CORP. (2002): S-PLUS® 6.1 (http://www.insightful.com).

JAMBOR, J.L. & DUTRIZAC, J.E. (1998): Occurrence and constitution of natural and synthetic ferrihydrite, a widespread iron oxyhydroxide. *Chem. Rev*. **7**, 2549-2585.

KUCERA, M. & MALMGREN, B.A. (1998): Logratio transformation of compositional data; a resolution of the constant sum constraint. *Marine Micropaleontology* **34**, 117-120.

LUMSDON, D.G., MEEUSSEN, J.C.L., PATERSON, E., GARDEN, L.M. & ANDERSON, P. (2001): Use of solid phase characterisation and chemical modelling for assessing the behaviour of arsenic in contaminated soils. *Appl. Geochem*. **16**, 571-581.

MANCEAU, A. (1995): The mechanism of anion adsorption on iron oxides: evidence for the bonding of arsenate tetrahedra on free $Fe(O,OH)_6$ edges. *Geochim. Cosmochim. Acta* **95**, 3647-3653.

MARTÍN-FERNÁNDEZ, J.A. (2001): *Measures of difference and non-parametric cluster analysis of compositional data*. Ph.D. thesis, Univ. Politècnica de Catalunnya, Barcelona, Spain (in Spanish).

________, BARCELÓ-VIDAL, C. & PAWLOWSKY-GLAHN, V. (1997): Different classifications of the Darss Sill data set based on mixture models for compositional data. *In* Proc. of IAMG'97, the Third Annual Conference of the Int. Association for Mathematical Geology, vol. I (V. Pawlowsky-Glahn, ed.). CINME, Barcelona, Spain. (151-156).

________, ________ & ________ (1998): Measures of difference for compositional data and hierarchical clustering methods. *In* Proc. IAMG'98, Fourth Annual Conf. of the Int. Association for Mathematical Geology II (A. Buccianti, G. Nardi & R. Potenza, eds.). De Frede Editore, Naples, Italy (526-531).

________, BREN, M., BARCELÓ-VIDAL, C. & PAWLOWSKY-GLAHN, V. (1999): A measure of difference for compositional data based on measures of divergence. *In* Proc. IAMG'99, Fifth Annual Conf. of the Int. Association for Mathematical Geology I (S. Lippar, A. Naess & R. Sinding-Larsen, eds.). Tapir, Trondheim, Norway (211-216).

PAKTUNC, D. & DUTRIZAC, J.E. (2003): Characterization of arsenate-for-sulfate substitution in synthetic jarosite using X-ray diffraction and X-ray absorption spectroscopy. *Can. Mineral.* **41**, 905-919.

PAWLOWSKY-GLAHN, V. & BUCCIANTI, A. (2002): Visualisation and modelling of sub-populations of compositional data: statistical methods illustrated by means of geochemical data from fumarolic fluids. *Int. J. Earth Sci.* **91**, 357-368.

________ & EGOZCUE, J.J. (2001): About covariance and correlation on the simplex. *In* Proc. IAMG'01, Seventh Annual Conf. of the Int. Association for Mathematical Geology. Cancún, México (CD-ROM).

PHILIP, G.M. & WATSON, D.F. (1988): Determining the representative composition of a set of sandstone samples. *Geol. Mag*. **125**, 267-272.

RIDENOUR, G.S. (1999): Compositional data analysis of the influence of drainage area and stream order on hydraulic geometry. *In* Proc. Fourth Int. Conf. on GeoComputation, Mary Washington College, Virginia, U.S.A.

SAVAGE, K.S., TINGLE, T.N., O'DAY, P., WAYCHUNAS, G.A. & BIRD, D.K. (2000): Arsenic speciation in pyrite and secondary weathering phases, Mother Lode gold district, Tuolumne County, California. *Appl. Geochem.* **15**, 1219-1244.

SPSS INC. (2000): SYSTAT® 10 statistical package. Chicago, IL (http://www.systat.com).

THE MATHWORKS INC. (2000): MATLAB® 6, language of technical computing (http://www.mathworks.com).

VERRUCCHI, C. & MINISSALE, A. (1995): Multivariate statistical comparison of northern Apennines Paleozoic sequences: a case study for the formations of Monti Romani (southern Tuscany – northern Latium, Italy). *Appl. Geochem.* **10**, 581-598.

WAYCHUNAS, G.A., DAVIS, J.A & FULLER, C.C. (1995): Geometry of sorbed arsenate on ferrihydrite and crystalline FeOOH: re-evaluation of EXAFS results and topological factors in predicting sorbate geometry, and evidence for monodentate complexes. *Geochim. Cosmochim. Acta* **59**, 3655-3661.

WELTJE, G.J. (2002): Quantitative analysis of detrital modes: statistically rigorous confidence regions in ternary diagrams and their use in sedimentary petrology. *Earth-Science Reviews* **57**, 211-253.

WHELAN, J.K., EGLINTON, L., KENNICUTT, M.C. & QIAN, YAORONG (2001): Short-time-scale (year) variations of petroleum fluids from the U.S. Gulf Coast. *Geochim. Cosmochim. Acta* **65**, 3529-3555.

Received January 7, 2003, revised manuscript accepted August 12, 2003.