

# Automated ore microscopy based on multispectral measurements of specular reflectance. I – A comparative study of some supervised classification techniques



Alfredo López-Benito<sup>a,\*</sup>, Juan Carlos Catalina<sup>b</sup>, David Alarcón<sup>b</sup>, Úrsula Grunwald<sup>b</sup>, Paulo Romero<sup>b</sup>, Ricardo Castroviejo<sup>b</sup>

<sup>a</sup> Universidad Politécnica de Madrid, Center for Computational Simulation, Department of Geological and Mining Engineering, Ríos Rosas, 21 28003 Madrid, Spain

<sup>b</sup> Universidad Politécnica de Madrid, Laboratory for Applied Microscopy, Department of Geological and Mining Engineering, Ríos Rosas, 21 28003 Madrid, Spain

## ARTICLE INFO

### Keywords:

Automated microscopy  
Specular reflectance  
VNIR  
AMCO System  
Image analysis  
Discriminant analysis

## ABSTRACT

Automated mineralogy, including quantitative compositional and textural information, is a requirement for an efficient ore processing, and is comprised as an important input for geometallurgical planning. Classical ore microscopy is seen by many potential users as an outdated, time consuming, tool. Thus, SEM-based systems are often the choice for those who can afford them, in spite of their evident limitations for some minerals (e.g. for iron oxide ores). However, automated and quantitative mineral characterisation of metallic ores is also possible with optical (reflected light) microscopes, attaining similar performance at a much lower price than SEM systems, as shown by the AMCO (Automated Microscopic Characterisation of Ores) prototype. This system relies on the measurement of multispectral specular reflectance,  $R$ , on polished ore sections, to achieve automated identification of the ore species (reflectance is computed from grey levels of digital images acquired by automated scanning of the sample). The performance of this approach, supported by a multispectral reflectance database covering the VNIR range (370–1000 nm) built with the AMCO System, is analysed in this paper, comparing the reliability of different classification methods to achieve ore identification.

The work outlined in this article focuses on checking the actual behaviour of four classification techniques, based respectively on spectral angle mapper, euclidean distance, Mahalanobis distance and linear discriminant analysis. The tests carried out reveal that the last two techniques are powerful tools to determine to which mineral corresponds a pixel based on its reflectance spectrum.

The obtained results prove that automated multispectral optical microscopy is a reliable tool for mineral characterisation of common/industrial ores, with few exceptions (distinction of cassiterite and chromite, or sphalerite and wolframite). For optimisation of its performance, the multispectral information may be complemented with some additional criteria, such as paragenesis or type of deposit (e.g. to differentiate cassiterite from chromite). The reflectance database can also be enriched with measures of local ores, to account for existing reflectance variations due to local conditions or to learn new ores. The performance is particularly good for iron oxide ores (hematite, magnetite, goethite), a very important commodity. In the tests carried out, they are successfully identified in more than 99% of the cases.

## 1. Introduction. Aims of this study

Ore microscopy, as a tool for the study of ore deposits, became a mature technique about a century ago (or at least by the 1930s, e.g. [Schneiderhöhn and Ramdohr \(1931\)](#)). Reflected light microscopy has been widely applied to support ore geology, mineral exploration, and ore processing all along the second half of the twentieth century. In recent years, however, its use is less common and the academic earth-

science and mining curricula in universities tend to neglect or even forget it. This may in part be due to the time, means, and attention needed for its effective learning. Also to the effort and expertise required to apply such devices as the traditional point counter to support ore processing with quantitative data. And last but not least, to the development of automated techniques based on SEM-EDS (Scanning Electron Microscope with Energy Dispersive Spectroscopy) ([Wightman et al., 2016](#)): these may be expensive, but they are faster and far more

\* Corresponding author.

E-mail address: [alfredo.lopez@upm.es](mailto:alfredo.lopez@upm.es) (A. López-Benito).

<https://doi.org/10.1016/j.mineng.2019.106136>

Received 16 April 2019; Received in revised form 20 November 2019; Accepted 22 November 2019

0892-6875/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

performant than point counter devices, so they are becoming part of the ancillary plant equipment of large mining or metallurgical operations.

This development is enhanced by the raising awareness of the need of geological and mineralogical information to optimize plant work, leading to the present-day concept of Geometallurgy, which refers to an interdisciplinary activity combining Geology, Mining Engineering, Geotechnics, Metallurgy, Mineral Economics and Environmental Sciences in order to maximize a mining project's economic value, reduce risk, build resilience and demonstrate the good management of the resource (Dominy et al., 2018).

Ore mineral characterisation, a basic specific input for an efficient geometallurgical planning, requires automation to deal with the huge amount of information to be reported in world-class deposits. Smaller mines also need this support, but cannot always afford SEM microscopes. Facing this need (as well the instrumental limitations of SEM-EDS), it is reasonable to look at alternative systems based on an affordable technique, such as automated optical microscopy. This is the option of the AMCO System (Automated Characterisation of Ores, EU/EIT Project No. 15039, H2020), which provides automated ore identification relying on the multispectral measurement of specular reflectance in polished ore sections in the near-UV, visible and near-IR (VNIR) ranges (Catalina et al., 2019). Short-wave infrared (SWIR) data have also been acquired, but their discussion is beyond the scope of this work.

The aim of the present work is to check the performance of this option, analyzing the reliability of the different spectral classification methods to achieve ore identification. Preliminary works published so far by various authors (e.g., Kühnel et al. (1980), Bernhardt (1987), Gerlitz et al. (1989), Pirard (2004), Pirard et al. (2008), Pirard (2016), Castroviejo et al. (2009), Castroviejo et al. (2010), Castroviejo et al. (2014), Catalina and Castroviejo (2017), among others) suggest this should be feasible, and a selection of ores will therefore be investigated in a systematic way, using the VNIR spectra measured with the AMCO System to test automated classification. Recently, López-Benito et al. (2015), López-Benito et al. (2017) have used successfully this kind of methods to deduce the chemical composition and typology of chromites, while Chopard et al. (2019) use principal component analysis and supervised k-means clustering on multispectral reflectance measures to classify some sulfide minerals.

## 2. Materials and methods

### 2.1. Data acquisition. The AMCO System

The measurement of multispectral specular reflectance of ores was carried out with the AMCO System developed at the LMA (Laboratory of Applied Mineralogy - Universidad Politécnica de Madrid). The AMCO System is based on a fully motorized research-grade reflected-light optical microscope, which has been specifically modified to allow the acquisition of multispectral images of specular reflectance beyond the visible range.

Two proprietary software applications complete the System:

- *amcoCapture* controls the different components of the instrument and applies patented image-processing procedures to calibrate the system and to capture and correct each of the bands of a field, composing a multispectral image that guarantees the accuracy and repeatability of the specular reflectance measurements and the perfect registration between the image bands. Images can be acquired either manually by the system operator, or automatically by the unattended scanning of a region of the polished section.
- *amcoAnalysis* analyses the multispectral images, recognizing the different minerals by applying advanced classification techniques that compare the multispectral specular reflectance values in the images with those of the database. In this way, a map of the distribution of mineral phases is generated for each field, and the

modal analysis of the sample is computed. These maps are the base for grain size, liberation and texture analyses that provide a complete geometallurgical characterization of the sample, invaluable for industrial applications.

The basic version of the AMCO System is able to acquire a multi-spectral image covering the VNIR (visible and near infrared) range of the spectrum (between 350 and 1000 nm). The current prototype acquires multispectral images composed of 20 reflectance bands, with the following central wavelengths: 370, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 750, 800, 850, 900, 950 and 1000 nm.

Automation bound to image analysis has some important implications, in this case, for the methodology and for the microscope setup:

- The automated scanning of the ore polished section requires a motorized stage with x-y-z displacement. The stage is therefore not rotatable and, given the random orientation of the minerals in the section, polarisation must be avoided (i.e. no polars are introduced in the light path), to ensure that anisotropic ore particles have a similar and reproducible specular reflectance value for any orientation. This means that a specific VNIR database had to be produced with own measures, and that the IMA/COM values of QDF3 (Criddle and Stanley, 1993) cannot be applied directly, since they are related to anisotropy and, moreover, they are bound to the visible range (no IR or UV values). However, QDF3 spectra provide a very useful tool to check the AMCO measures, which in the visible range must be coincident for isotropic minerals or intermediate between  $R_1$  and  $R_2$  (or  $R_a$ ,  $R_b$  and  $R_c$ ) for anisotropic ores.
- For the same reason, work under oil immersion is avoided to prevent measuring artifacts (e.g. air bubbles), and only dry objectives are used for the automated scanning.
- AMCO works with digital images as required for image analysis, while most of the commonly used microscope spectrophotometers acquire no images but just measure  $R$  at a point located in the center of the field. This means that  $R$  is measured by AMCO at every point of the acquired image, and this image can be stored, and then retrieved and processed mathematically to quantify any desired parameter, any time. This capability does not exist either with classical photometers, or with manual point-counter devices.

### 2.2. Multispectral reflectance data source

The objective of the study is the selection of a suitable classification technique for the discrimination of the most common metallic ores, based on their specular reflectance in the visible and near infrared ranges.

The source of information on ore reflectance is the database generated in the LMA for the Project AMCO. This database contains over seventy million measurements of multispectral specular reflectance from seventy-four metallic ores and associated minerals of economic interest. A measurement consists of a set of specular reflectance values covering the wavelength range between 370 and 1000 nm, each one corresponding to the gray level that a pixel of a given mineral has in each of the bands of a multispectral reflectance image. To build this database, numerous images of several samples of each mineral were acquired in an AMCO prototype, and mineral regions were manually outlined on them by an ore microscopy specialist using the *amcoAnalysis* software.

### 2.3. Selection of ores

The distribution and abundance of the known ore mineral species in nature is very uneven. Most of them are scarce or just rarities. Actually, not even 10% of the known ores are common. According to Criddle (1998), "the number of minerals known to absorb light strongly enough that they can be considered as opaque is about 660" (and they are

surely many more, if the new minerals which were defined in the twenty years of research elapsed since then are added), but “the fact that [Bowie and Simpson \(1980\)](#) selected just thirty-seven of them as representative of ore minerals that a student might encounter gives an idea of how few are common to most types of ore deposits”.

A priority list of forty ore species has thus been established applying practical criteria, and taking into account not only the abundance or frequency of natural occurrences of the ore, but also its industrial interest. Therefore, the ores selected for this study comprise the most common species, such as pyrite or magnetite, as well as some uncommon but economic ores, such as gold or platinum. The forty ores chosen are all comprised in the AMCO database, and they span a large range of specular reflectance values, from low reflectance oxides such as cassiterite (luminance  $Y = 10.7 - 12.1\%$ , QDF3) to high reflectance species such as native silver (luminance  $Y = 92.9\%$ , QDF3). The current AMCO database (comprising over seventy mineral species) is easily extensible, so new species can be added to be identified by the same methodology.

Gangue minerals and epoxy resin, which have very low but similar reflectance values ( $R \approx 5\%$ ) are not identified as mineral species, so they are not comprised among the forty ores analysed here; nevertheless they have also been measured and a procedure to distinguish between gangue and resin has been developed by [Grunwald et al. \(2019\)](#).

The metallic ores involved in this study are listed alphabetically in [Table 1](#), with indication of their composition and of the abbreviations used in this article.

**Table 1**  
Minerals used in the study.

Mineral	Abbrv.	Formula
Acanthite	ac	Ag <sub>2</sub> S
Arsenopyrite	asp	FeAsS
Bismuth	Bi	Bi
Bismuthinite	bm	Bi <sub>2</sub> S <sub>3</sub>
Bornite	bn	Cu <sub>5</sub> FeS <sub>4</sub>
Cassiterite	cs	SnO
Chalcocite	cc	Cu <sub>2</sub> S
Chalcopyrite	ccp	CuFeS <sub>2</sub>
Chromite	chr	FeCr <sub>2</sub> O <sub>4</sub>
Cinnabar	cin	HgS
Cobaltite	cob	CoAsS
Copper	Cu	Cu
Covellite	cv	CuS
Cubanite	cbn	CuFe <sub>2</sub> S <sub>3</sub>
Cuprite	cup	Cu <sub>2</sub> O
Enargite	en	Cu <sub>3</sub> As <sub>4</sub>
Galena	gn	PbS
Goethite	gth	FeO(OH)
Gold	Au	Au
Graphite	gra	C
Hematite	ht	Fe <sub>2</sub> O <sub>3</sub>
Ilmenite	il	FeTiO <sub>3</sub>
Magnetite	mt	Fe <sup>2+</sup> Fe <sup>3+</sup> O <sub>4</sub>
Manganite	mng	MnO(OH)
Marcasite	mc	FeS <sub>2</sub>
Molybdenite	mol	MoS <sub>2</sub>
Nickeline	nl	NiAs
Pentlandite	pn	(Ni,Fe) <sub>9</sub> S <sub>8</sub>
Platinum	Pt	Pt
Pyrite	py	FeS <sub>2</sub>
Pyrolusite	prl	MnO <sub>2</sub>
Pyrrhotite	po	Fe <sub>0.8-1</sub> S
Rutile	rt	TiO <sub>2</sub>
Silver	Ag	Ag
Sphalerite	sp	SZn
Stibnite	sbt	Sb <sub>2</sub> S <sub>3</sub>
Tetrahedrite	td	(Cu,Fe) <sub>12</sub> Sb <sub>4</sub> S <sub>13</sub>
Titanite	tit	CaTiSiO <sub>5</sub>
Uraninite	urn	UO <sub>2</sub>
Wolframite	wf	(Fe,Mn) WO <sub>4</sub>

#### 2.4. Mathematical tools. The Problem of Classification

Once the multispectral specular reflectance of a pixel of an image has been measured, the next task is to determine to which mineral it corresponds. This is the problem that we will refer to as the *Problem of Classification*. Basically there are two types of techniques to solve it, called respectively *Supervised Classification* and *Non-Supervised Classification*. The first one is applied when there is prior knowledge about the properties that the characteristic spectrum of each mineral possesses and the second when this information is not available. In the research described in this article we have used supervised methods because the spectral reflectance characteristics of each candidate mineral (patterns) have been previously determined and each new multispectral reflectance measurement is compared with them, being assigned to that mineral whose pattern is the most similar. Therefore, classification requires a precise definition of *similarity*.

From a mathematical point of view, a spectrum built with reflectance measurements at  $m$  different wavelengths can be considered as a point  $x$  of the  $\mathbb{R}^m$  space. From this standpoint, a great variety of measures of similarity can be defined. We have tested four of them: Spectral Angle Mapper (**SAM**), Normalized Euclidean Distance (**NED**), Mahalanobis Distance (**MD**) and Linear Discriminant Analysis (**LDA**), which will be described in next paragraphs.

In the design of a supervised classifier, four different stages can be considered: *data acquisition stage*, in which data of individuals belonging to each of the classes are obtained; *training stage*, in which the classification patterns are determined using a part of the acquired data; *validation stage*, in which the behaviour of the classifier is evaluated, applying it on individuals from another part of the acquired data; and, finally, *application stage*, in which the classifier is applied to classify unknown samples.

Validation of a classifier is assessed mainly by means of the cross-validation confusion matrix. The element  $A_{i,j}$  of this matrix is the number of individuals from the class  $i$  that are classified as class  $j$ . A confusion matrix can be read in two ways: reading by rows gives an idea of how accurate is the classification of a certain mineral, while reading by columns gives an estimation of the uncertainty in the classification of an individual as a particular mineral. Therefore, in what follows, we will say that a pixel is *misclassified* when, being of a certain mineral, the classifier assigns it to another mineral. In this context, the ability of a classifier to identify a mineral must be expressed in terms of both the percentage of misclassified pixels and the percentage of pixels assigned by the classifier to a mineral when they are actually of another mineral. The first of these percentages is known as *error of omission* (EO) and the second as *error of commission* (see, for example, [Campbell and Wynne \(2011\)](#)). Formally, if  $c$  is the number of groups, these errors can be expressed as:

$$EO_i = 100 \left( 1 - \frac{A_{i,i}}{\sum_{j=1}^c A_{i,j}} \right), \quad EC_j = 100 \left( 1 - \frac{A_{i,i}}{\sum_{i=1}^c A_{i,j}} \right)$$

Other important measure of the accuracy of a classifier is the so called *coefficient of agreement or  $\kappa$ -index* ([Cohen, 1960](#)) which evaluates how much better the classification is if it is compared with a random classification. It is ranked from  $-1$  to  $1$ . If  $N = \sum_{i=1}^m \sum_{j=1}^c A_{i,j}$  is the total number of individuals on which the classifier is validated, the  $\kappa$ -index has the form:

$$\kappa = \frac{N \sum_{i=1}^c A_{i,i} - \sum_{i=1}^c \left( \sum_{j=1}^c A_{i,j} \cdot \sum_{j=1}^c A_{j,i} \right)}{N^2 - \sum_{i=1}^c \left( \sum_{j=1}^c A_{i,j} \cdot \sum_{j=1}^c A_{j,i} \right)}$$

A value of  $\kappa$  close to zero indicates that the classification is not better than a random classification while  $\kappa = 1$  means that all the non diagonal entries of the confusion matrix are zero, i.e. the classifier leads to a perfect classification. This index should not be confused with the overall accuracy (OA) defined by:

$$OA = \frac{\sum_{i=1}^c A_{i,i}}{N}$$

It is noteworthy that, although usually  $\kappa$  and OA have similar values, they express different concepts.

#### 2.4.1. Spectral Angle Mapper (SAM)

This is one of the simplest methods. It is widely used in Remote Sensing. SAM uses as similarity measure the angle formed by the vectors that connect the point  $\mathbf{O} = (0, 0, \dots, 0)$  with  $\mathbf{X} = (X_1, X_2, \dots, X_i)$  and  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_i)$  (Kruse et al., 1993). This angle is given by:

$$\alpha = \cos^{-1} \left( \frac{\sum_{i=1}^m X_i Y_i}{\sqrt{\sum_{i=1}^m X_i^2} \sqrt{\sum_{i=1}^m Y_i^2}} \right)$$

This similarity measure has the disadvantage that it does not take into account the length of vectors (equivalent to the spectral albedo when working with diffuse reflectance), so it only measures differences in the spectral shape, being insensitive to variations in the total reflectance of model spectra (Dennison et al., 2004).

#### 2.4.2. Normalized Euclidean Distance (NED)

Euclidean distance between two points  $\mathbf{X} = (X_1, X_2, \dots, X_i)$  and  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_i)$  of  $\mathbb{R}^m$  is defined by:

$$d_E(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{i=1}^m (X_i - Y_i)^2}$$

The use of this distance for classification purposes has two serious drawbacks. The first one is that differences between point coordinates with high values contribute much more than differences between point coordinates with low values. In other words, this distance is sensitive to changes of scale. The second drawback is derived from the fact that point coordinates can be correlated. If this occurs, the differences between the values of the point coordinates provide redundant information. As a consequence, the dissimilarity between points may be over-emphasized.

To avoid the first cited drawback, individuals can be normalized to zero mean and unit standard deviation by subtracting the mean and dividing by the standard deviation. Even so, in spite of normalization, correlation may continue being a problem.

#### 2.4.3. Mahalanobis Distance (MD)

Mahalanobis distance between two points  $\mathbf{X}$  and  $\mathbf{Y}$  of  $\mathbb{R}^m$  is defined by:

$$d_M(\mathbf{X}, \mathbf{Y}) = \sqrt{(\mathbf{X} - \mathbf{Y})^T \mathbf{V}^{-1} (\mathbf{X} - \mathbf{Y})}$$

where  $\mathbf{V}$  is the covariance matrix. This distance has advantageous properties that avoid the two above mentioned drawbacks of Euclidean distance: it is invariant to changes of scale and it takes into account the correlations between the point coordinates and corrects the effect of the redundancy. It is interesting to remark here that when the point coordinates are uncorrelated, Mahalanobis distance coincides with normalized Euclidean distance.

#### 2.4.4. Linear Discriminant Analysis (LDA)

Discriminant Analysis is a set of multivariate statistical tools widely

used in many applications related to Data Mining, Pattern Recognition, Machine Learning and many others. Its purpose is, given a population stratified in classes or groups, to find a rule that allows classifying each individual of the population into its group of origin. This analysis is based on a Fisher's idea (Fisher, 1936), although later authors have made numerous contributions to this technique: Mika et al. (1999), McLachlan (2004), Tharwat et al. (2017) and many others.

The entries of the original data matrix,  $\mathbf{X}$ , are the measures on each individual of the population of a  $m$ -dimensional variable  $\mathbf{x}$ . In other words, each individual is a point of  $\mathbb{R}^m$ . The key idea is to find a subspace of  $\mathbb{R}^m$  in which the distances among the projections of the group means are maximum. This can be done by splitting the variance matrix of  $\mathbf{x}$ ,  $\mathbf{S}$ , into two different matrices: the between class variance matrix,  $\mathbf{S}_b$ , and the within class variance matrix,  $\mathbf{S}_w$ . The first of them represents the distances between the means of the groups ( $\mu_i$ ) and the total mean ( $\mu$ ), while the second represents the distances between individuals and the mean of the group which each of them belongs to.

Suppose now a vector  $\mathbf{a} \in \mathbb{R}^m$  and the multivariate variable  $\mathbf{y} = \mathbf{a}^T \mathbf{x}$ . Variance of  $\mathbf{y}$  is:

$$\mathbf{W} = \mathbf{a}^T \mathbf{S} \mathbf{a} = \mathbf{a}^T \mathbf{S}_b \mathbf{a} + \mathbf{a}^T \mathbf{S}_w \mathbf{a}$$

The goal of LDA is finding the vector  $\mathbf{a}$  that maximizes the variance between groups. Usually this is done maximizing

$$J(\mathbf{a}) = \frac{\mathbf{a}^T \mathbf{S}_b \mathbf{a}}{\mathbf{a}^T \mathbf{S}_w \mathbf{a}}$$

which is known as the Fisher's criterion. Maximizing  $J(\mathbf{a})$  is equivalent to finding the direction along which the distances between the projected groups means are maximum (numerator) and the variance within the groups is minimum (denominator) (Mika et al., 1999). It can be proved that this problem is equivalent to finding a vector  $\mathbf{a}$  and a scalar  $\lambda$  verifying:

$$(\mathbf{S}^{-1} \mathbf{S}_b) \mathbf{a} = \lambda \mathbf{a}$$

which means that  $\mathbf{a}$  is the eigenvector of  $\mathbf{S}^{-1} \mathbf{S}_b$  corresponding to the eigenvalue  $\lambda$  (Tharwat et al., 2017). It can be proved also that  $\lambda_i / \sum_{j=1}^m \lambda_j$  is the proportion of total variance explained in the direction of vector  $\mathbf{a}$ .

Therefore, parametric equations of the projection subspace are  $\mathbf{y} = \mathbf{A} \mathbf{x}$ , where the rows of the  $n \times m$  matrix  $\mathbf{A}$  are the eigenvectors of  $\mathbf{S}^{-1} \mathbf{S}_b$  related with its  $n$  ( $n \leq m$ ) larger eigenvalues. Variables  $\mathbf{y}$  are called discriminant variables or canonical discriminant variables if variables  $\mathbf{x}$  are normalized to zero mean and unit standard deviation.

Once the discriminant variables have been extracted, the problem is how to classify a new individual into its corresponding class. This can be accomplished in several ways. One of the most widely used is to assign the individual to the most likely group by means of the Bayes theorem. Let  $\mathbf{d}_i$  be the vector of scores of an individual  $i$  on the discriminant variables  $\mathbf{y}$  and let  $P(\mathbf{d}_i | g_k)$  be the probability that an individual of group  $g_k$  obtains the score  $\mathbf{d}_i$  (prior probability). Let  $P(g_k)$  the probability that individual  $i$  belongs to group  $g_k$ . By Bayes theorem, the probability that an individual with score  $\mathbf{d}_i$  belongs to group  $g_k$  (posterior probability) is:

$$P(g_k | \mathbf{d}_i) = \frac{P(\mathbf{d}_i | g_k) P(g_k)}{\sum_{j=1}^c P(\mathbf{d}_i | g_j) P(g_j)}$$

where  $c$  is the number of groups or classes. Calculation of  $P(\mathbf{d}_i | g_j)$  is made under the assumption of multi-normality of discriminant variables and that of  $P(g_j)$  is based on the previous knowledge. By default, they are assigned according to the size of the groups or considering all groups with the same probability.

As the multivariate normal probability density function of a  $n$ -dimensional variable  $\mathbf{X}$  is:

$$f(\mathbf{x}) = K \exp\left(-\frac{d_M^2(\mathbf{x}, \mu)}{2}\right)$$

where  $K$  is a constant depending on the determinant of the variance matrix, there is an obvious correspondence between the probability of an individual  $\mathbf{x}$  of belonging to group  $g$  and the Mahalanobis distance from  $\mathbf{x}$  to the group mean of  $g$ : the greater this distance, the lower the probability density of  $\mathbf{x}$ .

### 2.5. Preparation of databases for training and validation

A subset of the original multispectral specular reflectance database compiled in project AMCO was obtained, containing the measurements of the 40 metallic ores previously selected for the study (see Table 1). This database contained all the measurements of 3674 manually-outlined mineral regions, for a total of 59.6 million pixels.

From this database, different sets of training and validation data files were extracted, in order to carry out the classification tests. For every set, each of the mineral regions was randomly assigned to either the training or the validation file, and then 5000 pixels of each mineral were selected at random from the regions assigned to each file. In this way, it was ensured that a given mineral region only contributed pixels to either the training or the validation data file of a set. Five pairs of sets of data files (for training and for validation) were extracted in this way in order to check the stability of the classification results, not finding significant differences between the classification results obtained with each of them.

Obviously, during the training stage of the development of the classifier it is necessary the collaboration of an expert microscopist, able to decide unambiguously what mineral is being measured. For the processing of a set of data files, the first action that was performed was the elimination of those pixels whose spectra were clearly different from the average spectrum of the mineral they belong to. The presence of these anomalous pixels may have several causes: dirt or dust particles, polishing defects, voids, pits, etc., i.e. features that do not correspond to an actual change of mineralogy but to a physical artifact that must be discarded.

This cleaning process can be carried out once a certain zone (containing several hundred pixels) has been measured in a field. If a normal distribution of reflectance values is assumed, it can be supposed that 95% of them are at less than two standard deviations from the mean and that 99% of these values are at less than three standard deviations from the mean. Therefore, it can be considered that pixels whose reflectance values differ by more than a given number (two or three) of standard deviations from the mean have a low probability of corresponding to the majority mineral in the measured zone. This cleaning process has an important impact on the classification, especially in the training stage, since it allows to obtain more precise patterns. During the validation and application stages, it is not necessary to eliminate any pixels because if their measurements differ greatly from the patterns, they simply will not be classified as the main mineral of the measured zone.

The average reflectance values obtained for each studied mineral are displayed in Table 2. These average spectra were taken as the patterns with which to compare the spectra from unknown minerals. Fig. 1 shows some of the most typical spectrum shapes and in Fig. 2 the patterns of the forty ore minerals used in this study are displayed.

## 3. Experimental process and results

As one of the goals of this study was to compare the performance of some classification methods, the following four techniques were evaluated: Spectral Angle Mapper (SAM), Normalized Euclidean Distance (NED), Mahalanobis Distance (MD) and Linear Discriminant Analysis (LDA).

### 3.1. Spectral Angle Mapper based classifier (SAM)

Globally, the performance of this classifier is moderate, yielding a correct classification for only 86.35% of the pixels used for validation ( $\kappa = 0.8600$ ). It seems to work properly (less than 1% of misclassified pixels and error of commission less than 1%) for chalcopyrite, copper, cubanite, gold and pentlandite pixels. Bornite, cobaltite, cuprite, hematite and pyrite pixels are also usually well classified but there is a significant error of commission (1.30 – 6.09%) when a pixel is classified as one of these minerals.

### 3.2. Normalized Euclidean Distance based classifier (NED)

The performance of the NED classifier is not very good, as only 83.17% of pixels ( $\kappa = 0.8316$ ) used for validation were correctly classified (even less than the 86.35% attained by the SAM classifier). However, its mean error of commission is slightly better (12.60% for SAM and 8.46% for NED).

This classifier gives good results for chalcopyrite, gold, nickeline and rutile. Arsenopyrite, covellite, pentlandite and silver are also usually well classified but there is a significant error of commission (1.52 – 10.94%) when a pixel is classified as one of these minerals.

### 3.3. Mahalanobis Distance based classifier (MD)

The Mahalanobis Distance based classifier identifies correctly most of the minerals involved in this study, with 99.37% of pixels well classified ( $\kappa = 0.9935$ ). Exceptions to this general good performance are cassiterite (91.92%), platinum (98.04%), pyrolusite (97.78%), sphalerite (98.48%) and wolframite (96.58%). Errors of commission are also moderate (less than 1%) except for acanthite (2.84%), bismuth (2.84%), chromite (7.32%), gold (1.11%), manganite (5.04%) and sphalerite (2.84%). More details about these results can be seen in Tables 3 and 5. From these tables, we can realize that MD has some difficulties to discriminate between cassiterite and chromite (8.08% of swapped pixels), platinum and bismuth (1.96% of swapped pixels), sphalerite and manganite (1.52% of swapped pixels), and wolframite and sphalerite and manganite (3.42% of swapped pixels). Pyrolusite is sometimes confused with acanthite, bismuth and gold (2, 22% of swapped pixels). Larger error of commission is found when a pixel is classified as chromite because a good few pixels classified as chromite are in fact cassiterite (7.32%). To a lesser extent, this also occurs with pixels classified as manganite, which sometimes are really sphalerite or wolframite (5.04%). Acanthite, bismuth, gold and sphalerite also have error of commission greater than 1%. Values of EO, EC, OA and  $\kappa$  can not be obtained from Table 3 because it does not show the full confusion matrix.

Although results of this classifier are much better than those of SAM or NED, it has still some problems with the classification of cassiterite, platinum, pyrolusite, sphalerite and wolframite pixels. Furthermore, classification as acanthite, chromite or manganite has a significant error of commission over 2%.

### 3.4. Linear Discriminant Analysis based classifier (LDA)

Results from linear discriminant analysis were twenty discriminant variables, linear combination of the original variables (the values of reflectance measured for each wavelength). Of these, the two first explain 83.58% of total variance. As a consequence of this high value, perhaps a reduced model including only the two first discriminant variables, could be accurate enough. However, in this study we have preferred to use the full model with the twenty discriminant variables. The value of the Wilks' lambda statistic for this model is very close to zero ( $4.84 \cdot 10^{-13}$ ) indicating that the total variance can be explained almost exclusively by the deviations between the groups. In order to perform the classifier validation, all the prior probabilities were set to

**Table 2**  
Average specular reflectance values (370 – 1000 nm) measured at the mineral species selected for the study.

Mineral	Wavelength (nm)																			
	370	400	425	450	475	500	525	550	575	600	625	650	675	700	750	800	850	900	950	1000
Acanthite	28.82	29.35	29.65	30.15	30.59	30.12	29.50	28.35	27.73	27.23	26.96	26.80	26.70	26.57	26.43	26.24	26.08	25.53	25.39	25.14
Arsenopyrite	50.98	51.40	50.91	51.45	51.57	52.09	52.01	52.41	52.31	52.27	52.16	51.99	51.97	51.83	51.49	50.81	50.96	49.69	49.33	48.60
Bismuth	48.32	51.57	53.04	54.83	55.79	57.67	59.25	61.25	63.09	64.79	66.05	67.11	67.88	68.35	68.86	68.60	68.74	69.24	69.39	69.47
Bismuthinite	40.46	42.40	42.87	43.79	43.42	44.08	43.74	43.90	43.46	43.12	42.72	42.34	41.98	41.41	40.62	39.59	38.82	37.61	36.25	35.32
Bornite	22.43	21.40	19.58	17.95	17.28	16.91	17.38	18.26	19.53	21.06	22.41	24.05	25.58	27.11	29.76	31.79	33.45	34.69	35.21	35.84
Cassiterite	12.96	12.80	12.60	12.42	12.23	12.11	12.06	11.88	11.86	11.77	11.73	11.65	11.67	11.60	11.53	11.40	11.45	11.24	11.33	11.25
Chalcocite	31.68	33.02	33.29	33.42	32.96	32.09	31.11	30.12	29.18	28.40	27.67	27.02	26.40	25.81	24.93	24.29	23.84	23.48	23.54	23.70
Chalcopyrite	12.28	15.52	20.50	27.44	32.98	38.85	42.17	44.85	46.26	47.09	47.41	47.57	47.59	47.41	47.01	46.32	45.91	44.94	44.36	44.50
Chromite	12.88	12.71	12.38	12.22	11.99	11.92	11.80	11.65	11.55	11.45	11.42	11.30	11.29	11.21	11.15	11.08	11.08	10.92	10.94	10.88
Cinnabar	32.16	31.13	29.00	27.85	26.90	26.51	26.08	25.46	25.11	24.82	25.70	26.28	26.19	26.15	25.90	25.63	25.43	25.48	25.28	25.36
Cobaltite	46.28	47.73	46.83	47.50	46.99	48.48	48.91	50.28	51.33	52.51	53.29	53.96	54.35	54.37	54.29	53.66	53.32	52.04	50.53	49.96
Copper	42.24	46.15	49.39	52.52	54.69	56.55	57.98	62.35	76.17	87.67	90.53	92.13	92.88	93.67	94.66	95.42	96.60	97.32	98.21	98.32
Covellite	19.62	20.45	20.11	19.59	18.19	16.46	14.45	12.46	10.63	9.46	9.40	11.01	14.76	21.29	33.08	39.42	43.58	46.91	48.62	50.45
Cubanite	16.84	19.55	23.22	26.99	30.62	33.65	35.95	37.69	39.03	40.12	41.03	41.93	42.74	43.44	44.41	44.87	45.38	45.07	44.78	44.56
Cuprite	34.83	33.38	32.53	31.78	30.89	30.02	29.23	27.44	26.19	25.22	24.50	24.23	23.90	23.49	22.90	22.34	22.12	21.17	20.99	20.61
Enargite	27.66	27.48	27.10	26.98	26.85	26.57	26.34	26.10	26.08	26.20	26.40	26.69	26.89	26.96	26.76	26.84	27.19	27.19	26.14	25.38
Galena	47.88	47.05	45.29	44.19	42.99	42.06	41.15	40.73	40.35	40.30	40.39	40.59	40.65	40.59	40.12	39.59	39.24	38.34	37.80	37.38
Goethite	19.79	19.19	18.19	16.94	16.09	15.62	15.23	14.61	14.21	13.85	13.63	13.39	13.27	13.10	12.88	12.65	12.56	12.32	12.30	12.23
Gold	29.89	31.34	31.97	32.87	35.27	43.73	57.72	68.36	75.22	79.34	81.88	84.00	85.36	86.36	87.79	88.19	89.38	89.87	90.43	90.19
Graphite	11.14	11.44	11.64	11.86	12.04	12.35	12.73	12.90	13.24	13.53	13.73	14.12	14.26	14.58	15.12	15.61	16.19	16.81	17.29	17.86
Hematite	27.56	29.65	30.38	30.45	29.81	29.43	28.71	28.64	28.24	27.04	25.88	24.84	24.18	23.59	22.80	22.11	21.86	21.21	21.03	20.67
Ilmenite	20.79	20.43	19.12	18.49	17.89	17.86	17.70	18.07	18.16	18.36	18.53	18.55	18.63	18.77	18.79	18.54	18.43	17.91	17.51	17.32
Magnetite	20.67	20.83	20.24	19.85	19.53	19.57	19.50	19.65	19.70	19.82	19.89	19.88	19.81	19.63	18.98	18.18	17.55	16.72	16.35	16.22
Manganite	19.29	19.84	20.06	20.09	19.86	19.66	19.31	19.02	18.67	18.36	18.18	18.01	17.94	17.76	17.64	17.60	17.70	17.39	17.38	17.27
Marcasite	38.77	40.34	41.28	43.55	44.86	46.84	47.78	48.25	48.32	48.26	48.01	47.73	47.44	46.96	46.23	45.63	45.54	44.81	44.73	44.61
Molybdenite	26.85	28.45	29.38	30.89	30.77	29.66	28.42	27.69	27.22	27.55	28.15	27.76	28.60	28.36	27.20	26.79	26.66	26.28	26.13	26.13
Nickeline	43.77	42.99	41.00	40.96	41.26	43.82	46.54	50.22	53.42	56.23	58.30	60.26	61.80	63.12	65.46	67.17	69.57	70.15	70.74	71.12
Pentlandite	21.90	25.75	29.88	34.15	37.44	40.96	43.40	45.78	47.52	49.12	50.27	51.36	52.30	53.03	54.29	55.19	56.38	56.67	56.72	57.29
Platinum	55.50	57.12	58.45	60.13	61.50	62.95	64.22	65.01	65.82	66.61	67.28	67.97	68.74	69.02	70.22	71.10	72.46	71.36	71.50	71.29
Pyrite	36.76	38.04	39.90	43.08	46.01	49.09	50.80	52.15	52.97	53.53	53.94	54.36	54.72	54.76	54.08	52.55	51.48	49.53	48.23	47.04
Pyrolusite	28.09	29.09	29.94	30.60	31.01	31.15	31.26	30.99	30.75	30.48	30.27	29.92	29.74	29.44	29.09	28.83	28.78	28.17	27.82	27.34
Pyrrhotite	28.08	28.86	29.76	31.21	32.52	34.24	35.55	37.02	38.16	39.30	40.20	41.14	42.01	42.81	44.20	45.30	46.72	47.38	47.93	48.69
Rutile	26.59	25.05	23.82	23.04	22.50	22.03	21.69	21.43	21.11	20.93	20.89	20.77	20.78	20.62	20.57	20.58	20.76	20.38	20.39	20.13
Silver	58.89	64.07	68.45	70.96	74.44	75.95	78.69	79.55	81.37	82.43	83.32	84.22	84.95	85.64	86.71	87.30	88.53	89.05	89.80	89.73
Sphalerite	20.00	19.26	18.42	17.87	17.54	17.26	17.05	16.96	16.72	16.62	16.47	16.34	16.28	16.20	16.12	15.98	16.01	15.75	15.62	15.58
Stibnite	38.86	38.75	38.33	38.25	38.02	37.75	37.44	36.83	36.37	35.90	35.51	35.43	35.37	35.00	34.15	33.50	33.78	33.48	33.39	32.89
Tetrahedrite	29.49	29.90	29.89	29.87	29.92	30.06	30.33	30.69	30.67	30.43	30.02	29.45	28.93	28.45	27.58	26.72	26.24	25.40	24.95	24.69
Titanite	11.06	10.75	10.59	10.36	10.37	10.25	10.19	10.30	10.12	10.06	10.11	10.00	10.07	10.02	10.12	10.11	10.12	9.99	10.05	9.93
Uraninite	16.42	15.75	14.99	14.62	14.23	14.14	14.01	13.90	13.89	13.89	13.92	13.95	14.04	14.12	14.22	14.21	14.35	14.31	14.40	14.37
Wolframite	19.07	17.92	17.30	16.59	16.78	16.64	16.85	16.94	16.88	16.85	16.86	16.68	16.64	16.42	16.25	16.05	15.99	15.60	15.41	15.06

1/40 = 0.025, that is, all pixels have the same probability a priori to be classified as any mineral.

Regarding the classification, all the minerals studied are classified correctly (>99% of well classified pixels) except cassiterite (93.08%), pyrolusite (97.86%), sphalerite (97.94% and wolframite (98.46%).

Globally, there is only 0, 04% ( $\kappa = 0.9949$ ) of misclassified pixels. Errors of commission are also small (less than 1%), except for acanthite (2.18%), bismuth (1.77%), chromite (6.35%), gold (1.09%) manganite (2.38%) and wolframite (1.40%). In Tables 4 and 5 a summary of the classification results can be seen. As in the MD case, values of EO, EC,

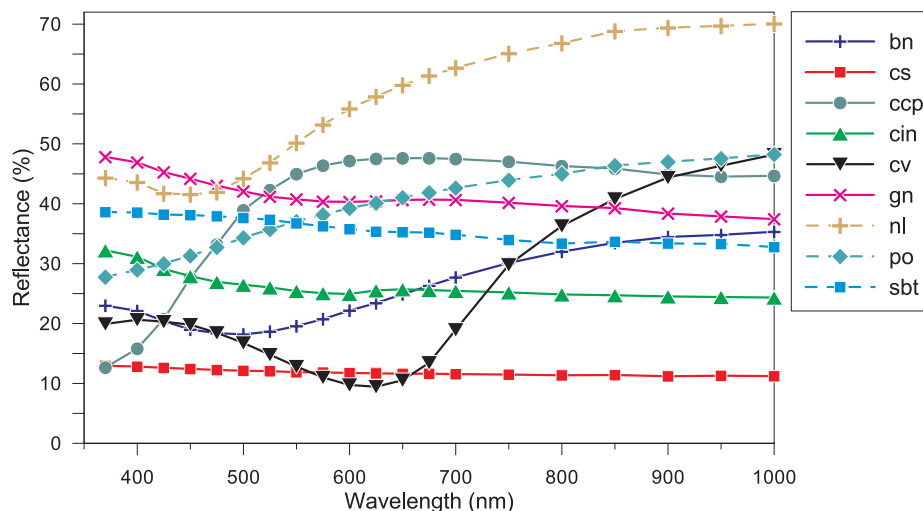


Fig. 1. Characteristic reflectance spectra of some minerals showing the most typical spectrum shapes.

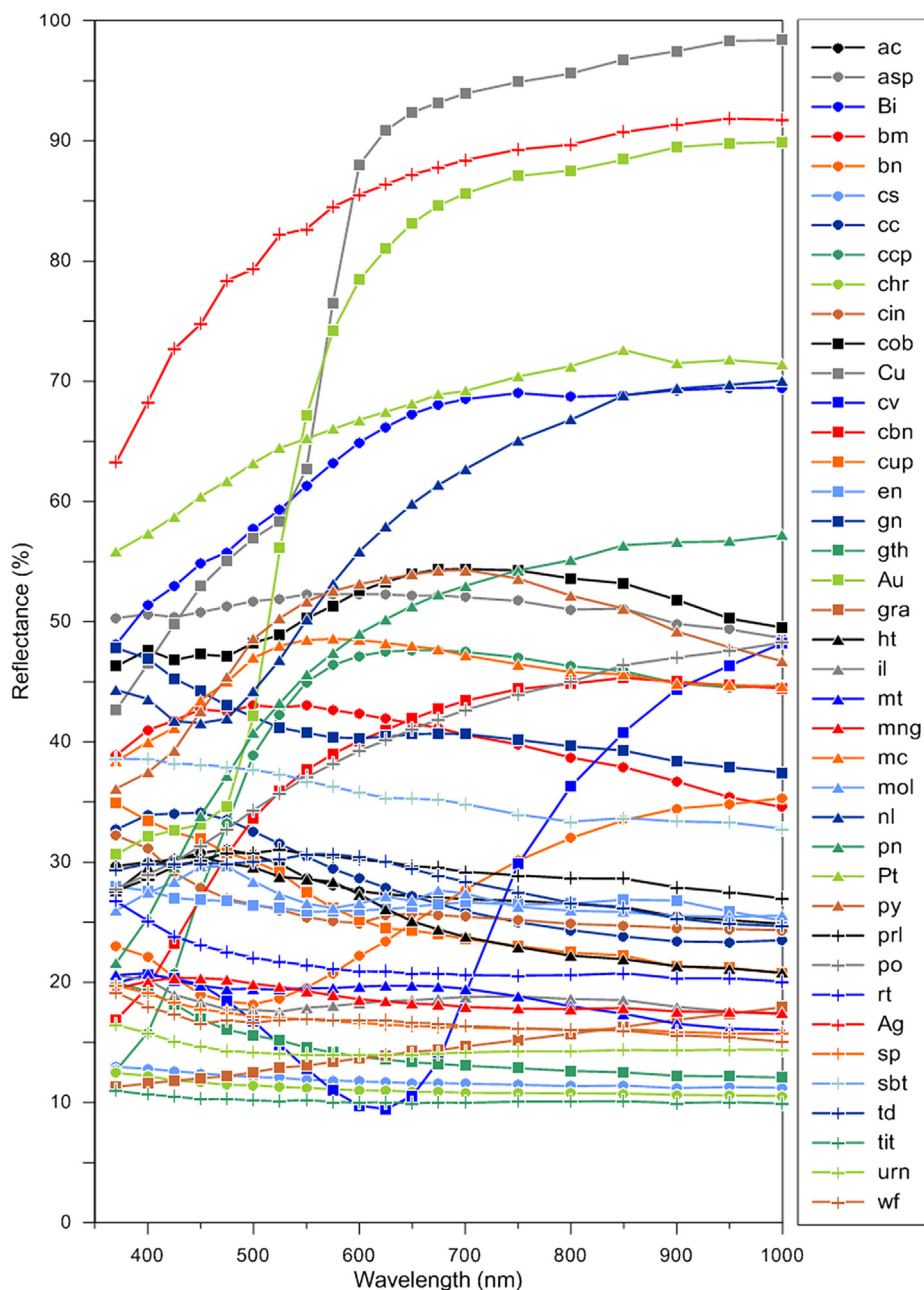


Fig. 2. Average specular reflectance spectra of the forty minerals used in this study.

Table 3

Cross-validation for the Mahalanobis Distance based classifier (only minerals with more than ten misclassified pixels are represented).

	Classified as										EO
	ac	Bi	cs	chr	Au	mng	Pt	prl	sp	wf	
ac	4988	0	0	0	0	0	0	0	0	0	0.24
Bi	0	5000	0	0	0	0	0	0	0	0	0.00
cs	0	0	4596	391	0	0	0	0	0	0	8.08
chr	0	0	14	4964	0	0	0	0	0	0	0.72
Au	0	0	0	0	5000	0	0	0	0	0	0.00
mng	0	0	0	0	0	4998	0	0	0	0	0.04
Pt	0	98	0	0	0	0	4902	0	0	0	1.96
prl	32	24	0	0	51	0	0	4889	0	0	2.22
sp	0	0	0	0	0	61	0	0	4924	0	1.52
wf	0	0	0	0	0	85	0	0	85	4829	3.42
EC	2.84	2.84	0.30	7.32	1.11	5.04	0.00	0.87	1.7	0.19	

EO: error of omission (%). EC: error of commission (%).

**Table 4**

Cross-validation for Linear Discriminant Analysis (only minerals with more than ten misclassified pixels are represented).

	Classified as										EO
	ac	Bi	cs	cc	chr	Au	mng	prl	sp	wf	
ac	4985	0	0	0	0	0	0	0	0	0	0.30
Bi	0	5000	0	0	0	0	0	0	0	0	0.00
cs	0	0	4654	0	337	0	0	0	0	0	6.92
cc	25	0	0	4971	0	0	0	0	0	0	0.58
chr	0	0	19	0	4969	0	0	0	0	0	0.62
Au	0	0	0	0	0	5000	0	0	0	0	0.00
mng	0	0	0	0	0	0	4996	0	0	0	0.08
prl	27	24	0	0	0	51	0	4893	0	0	2.14
sp	0	0	0	0	0	0	29	0	4897	70	2.06
wf	0	0	0	0	0	0	26	0	49	4923	1.54
EC	2.18	1.77	0.41	0.04	6.35	1.09	1.38	0.81	0.99	1.4	

EO: error of omission (%). EC: error of commission (%).

**Table 5**

Classification performance for each classifier: percentages of misclassified pixels (or errors of omission) and errors of commission).

	Misclassified				Error of commission			
	SAM	NED	MD	LDA	SAM	NED	MD	LDA
Acanthite	22.58	82.96	0.24	0.30	34.70	0.23	2.84	2.18
Arsenopyrite	11.92	0.20	0.22	0.18	19.18	1.52	0.06	0.04
Bismuth	16.78	3.78	0.00	0.00	6.85	28.56	2.84	1.77
Bismuthinite	2.84	4.60	0.42	0.34	8.10	5.66	0.00	0.00
Bornite	0.38	6.14	0.00	0.00	6.09	0.02	0.04	0.06
Cassiterite	38.12	65.40	8.08	6.92	52.55	0.00	0.30	0.41
Chalcocite	15.54	48.46	0.72	0.58	1.15	1.90	0.02	0.04
Chalcopyrite	0.18	0.68	0.00	0.00	0.00	0.00	0.06	0.06
Chromite	52.30	2.04	0.72	0.62	53.65	42.32	7.32	6.35
Cinnabar	22.76	37.42	0.42	0.36	4.41	0.00	0.02	0.04
Cobaltite	0.14	2.40	0.28	0.22	4.53	0.00	0.00	0.00
Copper	0.00	1.46	0.00	0.00	0.08	0.00	0.00	0.00
Covellite	6.24	0.00	0.02	0.02	0.02	10.94	0.00	0.00
Cubanite	0.10	13.52	0.10	0.10	0.48	0.00	0.00	0.00
Cuprite	0.04	19.00	0.10	0.06	3.64	0.00	0.00	0.00
Enargite	5.92	67.28	0.66	0.54	13.10	0.00	0.00	0.00
Galena	13.02	9.24	0.00	0.00	10.29	0.00	0.04	0.04
Goethite	1.52	2.72	0.12	0.10	0.77	0.02	0.68	0.40
Gold	0.08	0.00	0.00	0.00	0.16	0.02	1.11	1.09
Graphite	61.76	1.04	0.80	0.50	2.35	27.60	0.16	0.14
Hematite	0.04	32.36	0.16	0.10	4.47	0.50	0.00	0.00
Ilmenite	3.24	47.80	0.32	0.10	7.14	0.11	0.00	0.00
Magnetite	6.08	16.76	0.72	0.42	0.30	0.02	0.00	0.00
Manganite	48.20	12.06	0.04	0.08	38.29	42.43	5.04	2.38
Marcasite	1.38	1.36	0.48	0.48	13.69	0.78	0.10	0.12
Molybdenite	44.30	34.76	0.02	0.02	1.69	84.19	0.30	0.30
Nickeline	2.82	0.00	0.02	0.00	9.38	0.04	0.00	0.00
Pentlandite	0.50	0.04	0.00	0.00	0.10	2.38	0.00	0.00
Platinum	5.74	37.82	1.96	0.96	34.18	5.67	0.00	0.00
Pyrite	0.06	1.04	0.12	0.10	1.30	0.72	0.28	0.28
Pyrolusite	31.86	38.20	2.22	2.14	17.59	34.58	0.87	0.81
Pyrrhotite	11.00	2.36	0.00	0.00	31.76	13.56	0.22	0.22
Rutile	11.68	0.10	0.00	0.00	16.80	0.36	0.00	0.00
Silver	26.06	0.76	0.02	0.02	19.53	1.55	0.00	0.00
Sphalerite	21.52	10.90	1.52	2.06	13.63	7.26	1.70	0.99
Stibnite	39.98	9.30	0.08	0.08	30.60	16.19	0.12	0.12
Tetrahedrite	2.40	23.88	0.48	0.42	0.16	0.00	0.00	0.00
Titanite	11.86	8.40	0.22	0.16	24.50	0.52	0.00	0.04
Uraninite	1.68	23.38	0.54	0.28	11.98	0.00	0.02	0.04
Wolframite	3.26	3.68	3.42	1.54	4.93	8.77	0.19	1.40
Average	13.65	16.83	0.63	0.50	12.60	8.46	0.61	0.48

OA and  $\kappa$  can not be obtained from Table 4 because it does not show the full confusion matrix.

From these tables it can be concluded that LDA tends to confuse frequently cassiterite and chromite. A 6.74% of cassiterite pixels were

assigned to chromite and a 6.35% of pixels classified as chromite were actually cassiterite. Other somewhat problematic minerals are pyrolusite (2.35% of misclassified pixels), sphalerite (2.06% of misclassified pixels) and wolframite (1.54% of misclassified pixels). Errors of commission are moderate (less than 1%) except for acanthite (2.18%), bismuth (1, 77%), chromite (6.35%), gold (1.09%) and manganite (2.38%).

The central part of Fig. 3 shows the projection of the poorly classified minerals on the plane defined by the two first discriminant variables. It can be seen that projections of sphalerite and wolframite are very close, like those of cassiterite and chromite, with confidence intervals (at 99% of confidence) almost concentric. This fact can be confirmed in the lateral parts of the same figure, where the marginal density probability functions of these discriminant variables have been depicted. It is noteworthy the wide overlapping areas between the probability density functions of cassiterite and chromite and between wolframite and sphalerite, which leads to a significant uncertainty in the classification of these minerals.

Compared with other classification methods, LDA has a slightly better overall accuracy (99.51% well classified pixels) than the Mahalanobis distance based method (99.37%), and much better than methods based on NED (83.17%) and on SAM (86.35%). More specifically, and attending to the percentage of misclassified pixels, classification results for cassiterite, pyrolusite and wolframite are slightly better using LDA than using MD. Errors of commission are also somewhat better for acanthite, bismuth, chromite, manganite and sphalerite. However, for wolframite, LDA shows larger error of commission than MD.

### 3.5. Discussion

The first noteworthy observation is that chalcopyrite is always well classified (more than 99% of pixels) and has very small error of commission (less than 1%), regardless of the classification method used. The four classification methods give also good results (less than 5.00% of misclassified pixels) for cobaltite, copper, goethite, gold, pentlandite and pyrite, with error of commission lower than 5.00% (see Table 5). Cassiterite is rather poorly classified, none of the studied methods being able to classify well at least 95% of the pixels. Pixels of chromite are usually detected by NED, MD and LDA, but with a significant uncertainty in the classification. These two minerals provide good examples for a critical appraisal of the methodology, so they are further discussed below.

Of the four methods tested, NED and SAM have much worse overall behaviour (less than 87% of well-classified pixels) than MD and LDA, which have similar performance (more than 99% of well-classified pixels). This observation applies also to the errors of commission: 12.60% for SAM, 8.46% for NED, 0.61% for MD and 0.48% for LDA.

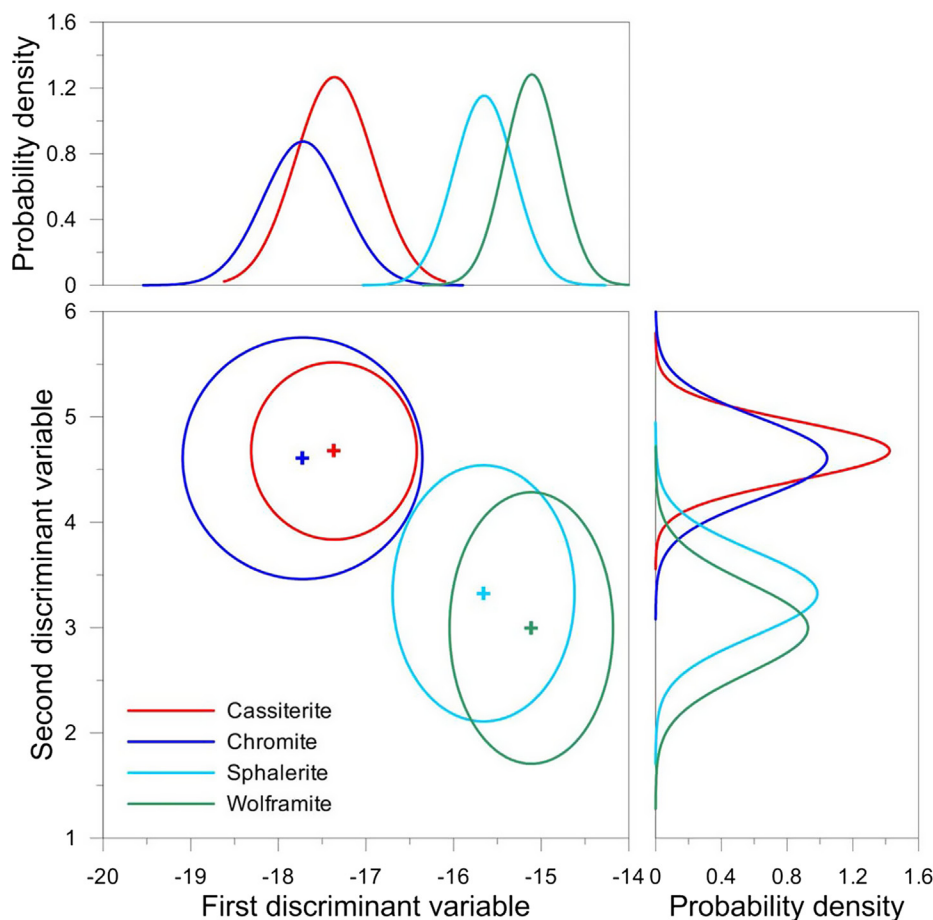


Fig. 3. Projection of the poorly classified minerals (cassiterite, chromite, sphalerite and wolframite) on the plane defined by the two first discriminant variables. Ellipses represent the 99% confidence interval for each mineral. Figures on top and on the right of the central figure display the marginal probability density functions of the first two discriminant variables for these minerals. Multinormality of discriminant variables for each mineral is assumed. Crosses represent the group centroids.

As the main difference between MD and LDA is that LDA works on discriminant variables instead of original variables, it is logical that there is not much difference in the performance of both classifiers. LDA gives slightly better results for cassiterite, platinum and wolframite and reduces the error of commission of bismuth, chromite and manganite. In contrast, it shows somewhat poorer performance for sphalerite and larger error of commission for wolframite.

As shown by the results compiled in Table 5, most ores are correctly classified with MD and LDA, so these methods are consistent for a reliable automated identification. This is particularly the case for iron oxide ores (goethite, hematite, magnetite), which are not easily classified with SEM-EDS due to their close chemical composition. Iron oxides and ferro-alloys belong, by the amounts invested and produced and by the volume of ore processed, to the most important metallic raw materials. So they represent a most significant industrial application for the automated optical microscopy.

There are however some challenging results which are to be analysed and understood from a mineralogical point of view, keeping always in mind the constraints of the experimental method chosen, which avoids polarisation. This is a reasonable choice, to account for the unavoidable random orientation of the mineral particles, but the implication is that the properties bound to anisotropy, e.g. bi-reflectance, cannot be measured. As a consequence, there is a risk for flat spectra to be confused, particularly those corresponding to low reflectance values. So, first risk to be faced is the confusion of gangue minerals and the epoxy resin used for polished block embedding, both having flat spectra, with roughly similar  $R$  values (about 5%). This important problem, a traditional barrier for automated optical methods, has been

solved in the LMA using combinations of resin dyes and particular ratios of selected spectral values in polished thin sections (Grunwald et al., 2019). The goal of this procedure is to detect and mask the epoxy resin, independently from the gangue species present. This is safer than undertaking the identification of the individual gangue species, because their  $R$  spectra will be often overlapping.

This question, critical for industrial applications, will be briefly discussed. Actually, the reflectance value of low absorbing (transparent) minerals, a class comprising most of the gangue minerals in ore deposits, is dictated by their refractive index ( $n$ ), according to the Fresnel's equation:

$$R = 100 \frac{(n - 1)^2 + k^2}{(n + 1)^2 + k^2}$$

where  $R$  is the reflectance (%),  $n$  is the refractive index and  $k$  is the absorption coefficient (transparent minerals have  $k \approx 0$ ). Further explanations on this subject can be found in Cervelle and Moëlo (1990), Criddle (1998). Fig. 4 is a plot of some of the transparent minerals on the  $n$ - $R$  plane according to the Fresnel's equation. Common gangue minerals and epoxy resin share similar reflectance values ( $R \approx 4.5 - 5\%$ ), very low if compared with most ores. The assumed  $R$  value refers to the most common gangue species, such as quartz, albite, adularia, sericite, chlorite, clays and many others, whose refractive index slightly exceeds 1.5, and therefore their reflectance according to the Fresnel's equation and to experimental results corresponds roughly to the  $R$  value of the resin; the same holds for fluorite ( $n \approx 1.45$ ), yielding a slightly lower  $R$  value (<4%).

The behavior of other possible gangue minerals, such as barite,

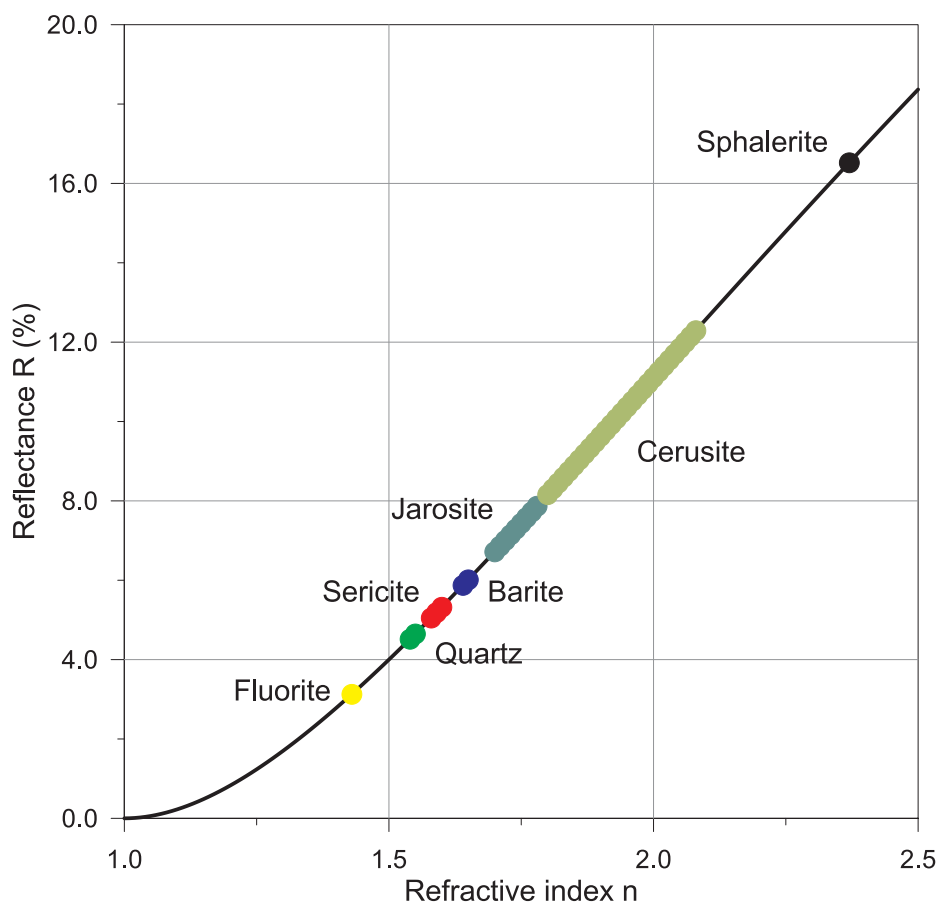


Fig. 4. Reflectance of some transparent minerals as a function of the refractive index.

carbonates, amphiboles, etc., is also dependent on their  $n$  value. Barite  $R$  value (0.6%) exceeds slightly that of quartz, as corresponds to its  $n$  value ( $\approx 1.63$ ). Carbonates are characterized by their strong birefringence, with at least one of their  $n$  values over 1.5, therefore corresponding in this orientation to a reflectance higher than epoxy. The carbonate group as a whole shows a large span of  $R$  values depending on composition, e.g. much larger for cerussite ( $n \approx 1.9 - 2$ ) than for calcite ( $n \approx 1.5 - 1.6$ ): this means for the group a span of  $R$  values from 5 to 12%. However, the effect of birefringence is significantly reduced in the AMCO System, which measures reflectance values with unpolarised light and provides an intermediate value between the extreme theoretical values, i.e. maximum and minimum dictated by  $n_g$  and  $n_p$  resp. Therefore, the AMCO carbonate  $R$  values may approach those of  $Fe - Mg$  silicates, such as pyroxenes, amphiboles, epidote or olivine ( $R \approx 8 - 9\%$ ), which can also appear as gangue in some ores.

This means that some of these higher  $R$  gangue minerals might be distinguished from resin by the multispectral measurements of the AMCO system, without resorting to the resin dyeing explained by (Grunwald et al., 2019). However, since this method allows the direct discrimination of resin, the classification of gangue minerals as such becomes redundant: if the goal is gangue-resin discrimination (and quantification, e.g. for characterisation of milled ore concentrates), it is more straightforward and safer to identify, mask, and cut off the resin from the image as a first step. In this way, a further source of uncertainty is avoided: the known but unpredictable variations of  $R$  measures on the polished surface of gangue or any transparent mineral, as a consequence of internal reflections, which may at random increase the brightness, i.e. the  $R$  measure on the mineral surface. Thus, this procedure has been chosen as the standard procedure for the characterisation of milled ore samples.

A similar problem is found in the case of cassiterite and chromite,

whose distinction is not always reliable (Tables 3–5). This is to be expected, given the similarity of their almost flat spectra (Table 2, Fig. 2) and thus their overlapping in the first and second discriminant variables space (Fig. 3). Ore dyeing is not possible, since it would damage the polished surface to be measured, so other kind of solutions have to be used. First, a detailed analysis of the spectral values from AMCO and other sources shows that the reflectance values of chromite are significantly composition-dependent and can be used for a chemical and typological characterisation of the chrome ore (López-Benito et al., 2017). This is related to the large compositional variations of spinels, including chromite, which determine a relatively large span for chromite and Cr-spinel spectral values (e.g. for the typical value of  $R_{550}$ : from 10.0% to 15.1%, but spanning down to 7.8% for Cr-poor chromian spinel and up to 17.6% for  $Fe^{3+}$ -rich ferrian chromite, respectively, after López-Benito et al. (2017) and QDF3, resp.). Therefore, as the chromite values in Fig. 2 show, some types of chromite ores show multispectral signatures which depart from those of cassiterite ( $R_{550} = 11.8\%$ , Table 2) and would thus not be misclassified. However, a reliable method aiming at a general application should not be limited to particular ore compositions. This means that some additional specific criteria may be needed in some cases to avoid mistakes, and Geology can provide useful tools. In this case, the paragenesis and the geological environment provide practical discrimination criteria, since both minerals occur in very different types of deposits: cassiterite is typical of felsic rocks or occurs more rarely associated to volcanogenic massive sulphides (as in Neves-Corvo, Portugal), while chromite is typically related to ultramafic rocks. These criteria can be added to the identification scheme, so that confusion is avoided by a preliminary screening.

Sphalerite (sp) and wolframite (wf) can also be confused (Tables 3 and 4), even if their colour is not exactly the same ( $\lambda_d = 475$  for sphalerite and  $\lambda_d = 468 - 472$  for wolframite, after QDF3) and their

**Table 6**  
Recommended methods for each mineral.

	SAM	NED	MD	LDA
Acanthite			✓	✓
Arsenopyrite			✓	✓
Bismuth			✓	✓
Bismuthinite			✓	✓
Bornite			✓	✓
Cassiterite	*	*	*	*
Chalcocite			✓	✓
Chalcopyrite	✓	✓	✓	✓
Chromite	*	*	*	*
Cinnabar			✓	✓
Cobaltite			✓	✓
Copper	✓		✓	✓
Covellite			✓	✓
Cubanite	✓		✓	✓
Cuprite			✓	✓
Enargite			✓	✓
Galena			✓	✓
Goethite			✓	✓
Gold	✓	✓	✓	✓
Graphite			✓	✓
Hematite			✓	✓
Ilmenite			✓	✓
Magnetite			✓	✓
Manganite			✓	✓
Marcasite			✓	✓
Molybdenite			✓	✓
Nickeline		✓	✓	✓
Pendlandite	✓		✓	✓
Platinum			✓	✓
Pyrite			✓	✓
Pyrolusite	*	*	*	*
Pyrrhotite			✓	✓
Rutile		✓	✓	✓
Silver			✓	✓
Sphalerite	*	*	*	*
Stibnite			✓	✓
Tetrahedrite			✓	✓
Titanite			✓	✓
Uraninite			✓	✓
Wolframite	*	*	*	*

\*: minerals requiring specific criteria (see discussion on text).

level of misclassification is relatively low: 1.52% (sp), 3.42% (wf), and 2.06% (sp), 1.54% (wf) for MD and LDA, respectively (Tables 3 and 4). To understand this small deviation, the samples measured for the AMCO database have been re-examined and checked. Again the compositional variations of both minerals can contribute to the confusion. Sphalerite luminance from different QDF3 samples spans from  $Y = 16.3\%$  to  $Y = 17.4\%$ , even if it is a cubic (therefore optically isotropic) mineral, and the abundance and color of internal reflections (IR) are strongly dependent on composition (Fe, Mn, Cd... contents). The appearance of wolframite ( $Y = 15.3 - 16.4\%$ , QDF3) shows a noteworthy change with composition. Fe and mixed (Fe, Mn) wolframite members show the typical optical properties of wolframite (brown-grey color, with few internal reflections), but the Mn member (hübnerite) is only very weakly light-absorbent, has abundant reddish-brown internal reflections, and looks like sphalerite under the microscope (with no polars). The wolframite samples selected for the database comprise a variety of compositions, including hübnerite from Pasto Bueno (Peru), occurring together with sphalerite. Both minerals show abundant IR, and these contribute to the measured reflectance value, as they come out of the polished surface of the crystal under study, adding to the reflected light and increasing the apparent  $R$ . Such particular cases are uncommon, but they show the convenience of a preliminary check of the local mineralogy and chemistry, as a first step of the study. For this reason, the database should be open to allow the addition of new measures by the user. A further addition whose implementation and testing is underway at the LMA-UPM is the extension of the AMCO database with

measurements in the SWIR spectral range, so as to improve the discriminant capacity in some particular cases, as e.g. the so called *white arsenides*.

Other small deviations have no mineralogical explanation and are probably due to artifacts. The slight misclassification ( $EO = 2.22\%$  for MD, 2.14% for LDA, Tables 3 and 4 resp.) of pyrolusite (prl) as native bismuth (Bi) or gold (Au), which have very different spectra, may be likely due to the typical contraction cracks of pyrolusite (secondary after manganite), which scavenge abrasive particles, dirt, and oil during the polishing process and make it difficult to get a perfectly clean surface, and to avoid any interference produced by reflection from these pollutants inside the cracks.

The inclusion of new minerals into a pre-existing classification model is trivial in the case of the MD classifier (as in the case of SAM or NED), since it is sufficient to add the mean and the covariance matrix of the reflectance measurements made on the new species to the set of means and covariance matrices of the model. This process is a bit more complicated for the LDA classifier because each new mineral added to the model affects all discriminant variables, that must be calculated again. This fact represents a clear disadvantage for the practical implementation of the LDA method compared to other techniques. In contrast, LDA has the advantage over other classifiers that it is possible in a very easy way to assign prior probabilities to each mineral on the basis of the previous knowledge of the studied samples. For example, if we are facing the study of a certain class of mineral deposit, we know in advance that the occurrence of some minerals can be ruled out, so these minerals come into the model with zero prior probability; or, if we know the mineralogical composition, we can assign prior probabilities based on the relative abundance of each mineral.

For the practical implementation of any of the classifiers described in this article, it should be taken into account that the spectral values measured are sometimes dependent to a certain extent on local deposit conditions, ruling such properties as chemical composition, solid solutions, trace elements content, crystallinity and crystallization history, annealing, alteration, colloidal adsorption, etc. These properties can sometimes determine the specular reflectance of a particular mineral phase as various researches have shown, e.g. (Beran, 1978; Bowles et al., 1983; Cerville, 1967; Cerville et al., 1971; Charlat and Lévy, 1971; Chikhaoui and Lévy, 1982; Hall et al., 1974; Leonard et al., 1969; Lévy, 1967). So, the reliability of the classification can be enhanced if databases are open to include the particular values of the local ores to be studied. These can be checked directly as the first step of the survey, so that any significant peculiarity in the samples from the deposit that is being studied is timely detected.

#### 4. Concluding statement

The automated recognition of metallic ores based on the multi-spectral measure of their specular reflectance on polished sections, comprised in the AMCO database, has been tested using four methods: Spectral Angle Mapper (SAM), Normalized Euclidean Distance (NED), Mahalanobis Distance (MD), and Linear Discriminant Analysis (LDA). Of these, MD and LDA work very well for almost all the minerals studied (performance for 36 of the 40 minerals tested: success ratio over 99%, or <1% misclassified pixels), while the SAM and NED performances are poor. Yet, there are a few minerals for which all the methods seem to work properly, while for the rest some specific methods are preferable.

A summary report on the performance of each method for automated ore identification is presented in Table 6, showing the recommended method for each mineral. It also shows the four ore minerals (cassiterite, chromite, sphalerite and wolframite, marked “\*” in this table) which with more than 1% misclassified pixels challenge the routine automation procedure. The correct classification of these minerals can be reached introducing additional criteria, such as paragenesis or type of deposit, or local ore mineralogical information, as

shown in the discussion.

The recommended approach for automation is therefore to work with a flexible software, able to include additional ad hoc or local information.

Automated systems cannot replace mineralogists, but can enhance enormously their performance, especially for quantified ore characterization, a most important tool to support Geometallurgy.

As shown by the results of Table 6, iron oxide ores (goethite, hematite, magnetite) are reliably classified by their multispectral specular reflectance, thus avoiding the typical confusion by current methods based on SEM-EDS, due to their close chemical composition. The method discussed, applying the AMCO System, can therefore be a most useful tool to enhance ore processing for this important field of mining.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This paper presents some results from the work done in UPM for the EIT (European Institute of Innovation and Technology) Project no 15039: AMCO (Automated Microscopic Characterization of Ores, 2016–2018). The support and encouragement of EIT-Raw Materials (H2020 Programme of the European Commission) and of the colleagues of the AMCO Consortium (Université de Liège, Belgium; ThinSectionLab, France; and the mining companies KGHM Polska Miedz, Poland, and Cobre Las Cruces, Spain) are thankfully acknowledged.



EIT RawMaterials is supported by the EIT,  
a body of the European Union

### References

- Beran, A., 1978. Die bestimmung der zusammensetzung von dolomit-ankerit-und magnetit-siderit-mischkristallen mit hilfe von reflexionsmessungen. *Neues Jahrbuch für Mineralogie - Monatshefte* 559–565.
- Bernhardt, H.J., 1987. A simple, fully-automated system for ore mineral identification. *Mineral. Petrol.* 36, 241–245.
- Bowie, S., Simpson, P., 1980. The Bowie-Simpson System for the Microscopic Determination of Ore Minerals. *McCrone Research Associates, London.*
- Bowles, J., Atkin, D., Lambert, J., Deans, T., Phillips, R., 1983. The chemistry, reflectance, and cell size of the erlichmanite (oss 2)-laurite (rus 2) series. *Mineral. Mag.* 47, 465–471.
- Campbell, J., Wynne, R., 2011. *Introduction to Remote Sensing*. Guilford Press.
- Castroviejo, R., Brea, C., Pérez-Barnuevo, L., Catalina, J.C., Segundo, F., Bernhardt, H.J., Pirard, E., 2009. Using computer vision for microscopic identification of ores with reflected light: preliminary results. In: Williams, P.J, et al. (Ed.), *Smart Science for Exploration and Mining (Proceedings of the Tenth Biennial SGA Meeting)*. Townsville, Australia (17–20 August, 2009), vol. 2, pp. 682–684.
- Castroviejo, R., Catalina, J.C., Bernhardt, H.J., Pirard, E., Brea, C., Pérez-Barnuevo, L., Segundo, F., Espí, J., 2014. Multispectral (visible and near infra-red, 400–1000 nm range) reflectance data file from common ore minerals. Data Base at IMA-COM web site: [http://projects.gtk.fi/com/results/reflectance\\_data.html](http://projects.gtk.fi/com/results/reflectance_data.html).
- Castroviejo, R., Catalina, J.C., Bernhardt, H.J., Pirard, E., Segundo, F., Brea, C., Pérez Barnuevo, L., 2010. A fully automated system for multispectral ore microscopy. In: Pál-Molnár, E. (Ed.), *Acta Mineralogica-Petrographica (Proceedings of the IMA2010 Meeting)*. Budapest, Hungary (21–27 August, 2010), p. 281.
- Catalina, J.C., Alarcón, D., Grunwald, U., López-Benito, A., Castroviejo, R., 2019. Automated ore microscopy based on multispectral measurement of specular reflectance. II - Methodology. AMCO Syst. (in progress).
- Catalina, J.C., Castroviejo, R., 2017. Microscopía de reflectancia multispectral: Aplicación al reconocimiento automatizado de menas metálicas. *Revista de Metalurgia* 53, 1–20.
- Cervelle, B., 1967. Contribution à l'étude de la série ilménite-geikéilite. *Bull. BRGM* 5, 1–26.
- Cervelle, B., Lévy, C., Caye, R., 1971. Dosage rapide du magnésium dans les ilménites par microreflectométrie. *Miner. Deposita* 6, 34–40.
- Cervelle, B., Moëlo, Y., 1990. Reflected-light optics. In: Jambor, J., Vaughan, D. (Eds.), *Advanced Microscopic Studies of Ore Minerals. Short Course Handbook 17*. Mineralogical Association of Canada, Ottawa, Canada, pp. 87–108.
- Charlat, M., Lévy, C., 1971. Influence des principales substitutions sur les propriétés optiques dans la série tennantite-tétraédrite. *Bull. Soci. Francaise Mineral. Cristallograp.* 99, 29–37.
- Chikhaoui, S., Lévy, C., 1982. Expression quantitative de la couleur dans la série tennantite-tétraédrite: influence des substitutions. *Can Mineralog.* 20, 101–109.
- Chopard, A., Marion, P., Royer, J.J., Taza, R., Bouzahzah, H., Benzazoua, M., 2019. Automated sulfides quantification by multispectral optical microscopy. *Miner. Eng.* 131, 38–50.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Measur.* 20, 37–46.
- Criddle, A., 1998. Ore microscopy and photometry. In: Cabri, L.J., Vaughan, D.J. (Eds.), *Modern Approaches to Ore and Environmental Mineralogy. volume 27 of MAC Short Course Series*, pp. 1–74.
- Criddle, A.J., Stanley, C.J. (Eds.), 1993. *Quantitative Data File for Ore Minerals, 3rd ed.* Chapman & Hall, London.
- Dennison, P., Halligan, K., Roberts, D., 2004. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sens. Environ.* 93, 359–367.
- Dominy, S., O'Connor, L., Parbhakar-Fox, A., Glass, H., Purevgerel, S., 2018. *Geometallurgy—A route to more resilient mine operations*. Minerals 8.
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Ann. Hum. Genet.* 7, 179–188.
- Gerlitz, C., Leonard, B., Criddle, A., 1989. QDF database system, version 1.0: reflectance of ore minerals; a search-and-match identification system for IBM and compatible microcomputers using the IMA/COM quantitative data file for ore minerals, second issue, Part A-Program documentation. Technical Report.
- Grunwald, U., Catalina, J., Alarcón, D., López-Benito, A., Castroviejo, R., 2019. A reliable method for the automated distinction of quartz gangue and epoxy resin with reflected light microscopy and its application to digital image analysis. In: *Proceedings of the 15th SGA Biennial Meeting*. Glasgow, Scotland (27–30 August 2019), pp. 1528–1531.
- Hall, A., Cervelle, B., Lévy, C., 1974. Effect of substitution of Cu by Zn, Fe and Ag on optical properties of synthetic tetraedrite  $Cu_{12}Sb_4S_{13}$ . *Bull. Soci. Francaise Mineralog. Cristallograp.* 97, 18–26.
- Kruse, F., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., Goetz, A., 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* 44, 145–163.
- Kühnel, R., Prins, J., Roorda, H., 1980. The 'Delft' system for mineral identification. Vol. 1: 'Opaque minerals'. Delft University Press.
- Leonard, B., Desborough, G., Page, N., 1969. Ore microscopy and chemical composition of some laurites. *Am. Mineral.: J. Earth Planet. Mater.* 54, 1330–1346.
- Lévy, C., 1967. Contribution à la minéralogie des sulfures de cuivre du type  $Cu_3XS_4$ . volume 54. Éditions BRGM.
- López-Benito, A., Gervilla, F., Catalina, J.C., Castroviejo, R., 2015. Determination of chromite composition from multispectral reflectance measurements (400–1000 nm). In: André-Mayer, A., Cathelineau, M., Mucchez, P., Pirard, E., Sindern, S. (Eds.), *Mineral Resources in a Sustainable World (Proceedings of the 13th SGA Biennial Meeting)*. Nancy, France (24–27 August 2015), vol. 4, pp. 1435–1438.
- López-Benito, A., Gervilla, F., Catalina, J.C., Castroviejo, R., 2017. Chromite typology and composition characterized through multispectral specular reflectance. *Ore Geol. Rev.* 89, 132–142.
- McLachlan, G., 2004. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, Hoboken, New Jersey.
- Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R., 1999. Fisher discriminant analysis with kernels. In: *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, Ieee. pp. 41–48.
- Pirard, E., 2004. Multispectral imaging of ore minerals in optical microscopy. *Mineral. Mag.* 68, 323–333.
- Pirard, E., 2016. Optical microscopy. In: Becker, M., Wightman, E., Evans, C. (Eds.), *Process Mineralogy. Julius Kruttschnitt Mineral Research Centre. Indooroopilly, Queensland, Australia*, pp. 51–66.
- Pirard, E., Bernhardt, H.J., Catalina, J.C., Brea, C., Segundo, F., Castroviejo, R., 2008. From spectrophotometry to multispectral imaging of ore minerals in visible and near infrared (VNIR) microscopy. In: *Proceedings of the Ninth International Congress for Applied Mineralogy*. Brisbane, Australia (8–10 September, 2008), pp. 1–6.
- Schneiderhöhn, H., Ramdohr, P., 1931. *Lehrbuch der Erzmikroskopie, 2 vol.* Borntraeger Verlag, Berlin.
- Tharwat, A., Gaber, T., Ibrahim, A., Hassanien, A.E., 2017. Linear discriminant analysis: A detailed tutorial. *Ai Commun.* 30, 169–190.
- Wightman, E., Evans, C., Becker, M., Gu, Y., 2016. Automated scanning electron microscopy with energy dispersive spectrometry. In: Becker, M., Wightman, E., Evans, C. (Eds.), *Process Mineralogy. Julius Kruttschnitt Mineral Research Centre. Indooroopilly, Queensland, Australia*, pp. 97–107.