

# Multivariate statistical methods for the analysis of microscope image series: applications in materials science

N. BONNET

INSERM Unit 314 (IFR 53) and University of Reims (LERI), 21, rue Clément Ader, BP 138,  
51685 REIMS Cedex, France

**Key words.** Automatic classification, dimensionality reduction, electron microscopy, image series, microanalysis, multivariate image segmentation, multivariate statistical analysis, neural networks, watersheds.

## Summary

Multivariate data sets are now produced in several types of microscopy. Multivariate statistical methods are necessary in order to extract the useful information contained in such (image or spectrum) series. In this review, linear and nonlinear multivariate methods are described and illustrated with examples related both to the segmentation of microanalytical maps and to the study of variability in the images of unit cells in high-resolution transmission electron microscopy. Concerning linear multivariate statistical analysis, emphasis is put on the need to go beyond the classical orthogonal decomposition already routinely performed through principal components analysis or correspondence analysis. It is shown that oblique analysis is often necessary when quantitative results are expected. Concerning nonlinear multivariate analysis, several methods are first described for performing the mapping of data from a high-dimensional space to a space of lower dimensionality. Then, automatic classification methods are described. These methods, which range from classical methods (hard and fuzzy C-means) to neural networks through clustering methods which do not make assumptions concerning the shape of classes, can be used for multivariate image segmentation and image classification and averaging.

## Introduction

Physical and chemical sensors provide data concerning the physical and chemical state and composition of specimens. At the microscopic level, even at very high resolution, several different signals (or different attributes of the same signal) can be recorded simultaneously, thus providing multivariate data sets. Such multivariate data sets are recorded more and more often, because of the intuitive idea that multivariate measurements can provide information well beyond the limits achievable with individual measurements. However, this intuitive idea becomes a reality only in

the cases where data analysis methods are available for extracting useful compact information from the enormous amount of data recorded. From this point of view, one could say that, although several groups of methods are already in use, there is still a lot to do before the full recorded information can be optimally extracted.

Since we are dealing with multivariate data sets, the generic term for these data analysis methods (which often rely on some kind of statistics) should be multivariate statistical analysis. However, from the point of view of terminology, the situation is slightly confused, because multivariate statistical analysis (MSA) is the name generally used for representing a specific group of methods dealing with the analysis of data sets by linear methods. However, linear methods (such as principal components analysis (PCA), correspondence analysis (CA), Karhunen–Loëve Analysis (KLA), ...) represent only a small part of all available methods; nonlinear methods are also very promising, and will probably constitute the largest part of further developments.

In this paper, I will attempt to cover nonlinear methods as well as linear ones. On the other hand, analysis of data sets (which, as we will see, can be considered as the mapping onto a subspace) is not the only piece of information extraction; another aspect consists of the interpretation of the projections. Although several facets can again be recognized, I will concentrate on the question of data (or object) classification. Obviously, this aspect is largely connected to the topics of pattern recognition and, in some way, of artificial intelligence. Supervised classification methods, which require a preliminary training phase (the classifier is trained with examples constituting the training set), can be used. Alternatively, unsupervised classification methods can be attempted, where the data are gathered into several classes without the help of an 'expert', on the basis of their information content (signature) only.



It should be stressed that some of the concepts and methods described in this paper (linear MSA, classification of images in the representation space) were introduced in electron microscopy by researchers working in the field of three-dimensional reconstruction of macromolecules (van Heel & Frank, 1981; Frank & van Heel, 1982; van Heel, 1984, 1989; Frank, 1990; Borland & van Heel, 1990). Although these methodological contributions and the great success they have had in recent years in elucidating the 3D structure of important biological macromolecules are not described in this paper, their importance in the diffusion of related (although different) techniques for materials science applications should be recognized.

The outline of the paper is the following. In the next section, I give some examples of multivariate physical data sets and I introduce those that I will use for illustration in the rest of the paper. The following section is devoted to linear multivariate statistical analysis. Since these methods are described in many textbooks (Lebart *et al.*, 1984) and are already in use in several laboratories, only a brief description of them will be given. Emphasis will be put on the extension of orthogonal MSA to oblique MSA. The section thereafter will be devoted to nonlinear mapping, an extension of linear MSA. Several approaches will be discussed and illustrated, ranging from the minimization of a criterion (cost function) to neural networks approaches. The last section will be devoted to automatic classification. I will concentrate on unsupervised classification, which does not mean that supervised classification does not deserve attention. After briefly describing some classical statistical classification techniques (which make assumptions concerning the shape of clusters in the parameter space), I will put emphasis on new methods which do not make assumptions concerning the shapes of classes.

### Some examples of multivariate data sets

Multivariate data sets produced in the domain of physical sciences (as well as in other scientific domains) are very diverse in nature. They can be 'simple data', series of spectra, series of two- or three-dimensional images, spectrum-images, etc.

Examples of 'simple' data are:

- the results of measurements (concentrations of different elements for instance) made at different positions on a specimen. Examples are described in Quintana (1991) and Quintana & Bonnet (1994a,b).
- different preparation conditions related to some characteristics of the specimens obtained (see for instance the paper by Simeonova *et al.* (1996), which concerns the conditions of preparation of high-temperature superconducting thin films).

Examples of multivariate spectra are:

- sets of spectra recorded as a function of time (time-

resolved spectroscopy) (Ellis *et al.*, 1985). Examples of multivariate statistical analysis of such data sets can be found in Bonnet *et al.* (1991) and Jbara *et al.* (1995).

- sets of spectra recorded as a function of position, through an interface for instance (Tencé *et al.*, 1995). Examples of multivariate statistical analysis of such data sets can be found in Gatts *et al.* (1995), Müllejans & Bruley (1995), Brun *et al.* (1996) and Titchmarsh & Dumbill (1996).

Multivariate two-dimensional (2D) image sets include (Bonnet, 1995a):

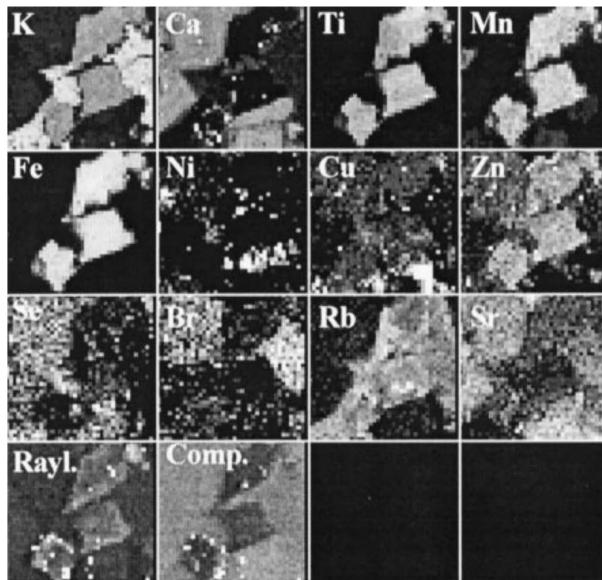
- sets of different elemental (or chemical) maps of a specimen recorded in different microanalytical modes (Auger, EELS, X-ray emission, X-ray fluorescence, X-ray differential absorption). Examples of the processing of this type of data can be found in Bonnet *et al.* (1992), Prutton *et al.* (1990, 1996), Cazaux (1993), Quintana & Bonnet (1994a,b), Collieux *et al.* (1994) and Trebbia *et al.* (1995). As an illustration of this type of data, I have selected a series of 14 X-ray fluorescence maps of a specimen of granite (courtesy of K. Janssens and collaborators, Department of Chemistry, University of Antwerp: Wekemans *et al.*, 1997). The series is displayed in Fig. 1.

- sets of images of unit cells recorded by high-resolution transmission electron microscopy (HRTEM) of interfaces between two crystals. Such data sets have been analysed (with the purpose of visualizing the gradual change of composition across the interface) by pattern recognition techniques (Ourmazd *et al.*, 1990; De Jong & Van Dyck, 1990; Kisielowski *et al.*, 1995). Their analysis by multivariate statistical methods has also begun (Rouvière & Bonnet, 1993; Aebersold *et al.*, 1996). I will also show some possibilities of this technique. Such a data set is displayed in Fig. 2.

Multivariate three-dimensional (3D) images can also be obtained with some microanalytical techniques (secondary ion mass spectroscopy (SIMS) (Van Espen *et al.*, 1992) or fluorescence confocal laser microscopy, for instance). Techniques working with 2D images can be extended to 3D images relatively easily, thanks to the increasing capabilities of computers.

Four-dimensional (4D) image data sets are, for instance, 3D images recorded as a function of time, a mode which begins to be feasible in fluorescence (confocal or not) videomicroscopy.

Spectrum-images (or a variant of them: image-spectra) will be the multivariate data sets of choice at the beginning of the 21st century (Jeanguillaume & Collieux, 1989). Combining spatial and full spectral information, they will mix the advantages of spectroscopy and microscopy. Although several acquisition procedures are already in use, in different fields of physics, chemistry, biology and teledetection, the procedures for analysing the data sets are still in their infancy and I will not address them in this paper.

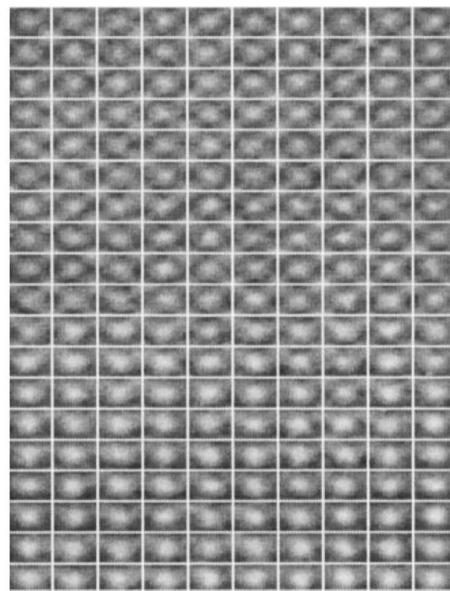


**Fig. 1.** Example of multivariate image set. The image series consists of 14 X-ray fluorescence maps of a specimen of granite (courtesy of K. Janssens and collaborators, University of Antwerp). The aim of the analysis is to segment the specimen area into regions of homogeneous composition, which means labelling pixels according to their content in the different images (pixels are represented by vectors in a 14-dimensional space). If this labelling can be performed successfully, further quantification and characterization of the specimen can take place: the percentage of area occupied by the different phases can be computed, for instance.

#### Linear multivariate statistical analysis (LMSA)

The purpose of LMSA is to reduce the number of components of the objects studied. This is useful (and sometimes necessary) because a multivariate data set always contains redundant information: the  $N$  measurements are never completely independent and some correlation (or anticorrelation) is always present. LMSA can help both to reduce redundancy and to define a new representation space (onto which the components of the objects are less correlated). This step is performed on the basis of the variance–covariance matrix (classically, the concept of variance is supposed to be one of the concepts connected to the information. However, other information descriptors can also be used (Bonnet, in preparation)).

Consider  $X$  as the data set, arranged as a matrix: rows are objects (or individuals) and columns are descriptors (or features, or variables) of these objects. For instance, if one wants to classify a set of images, images are individuals and pixel intensities are features. But if one wants to classify pixels (image segmentation), rows are composed of pixels and columns represent the different image contents for every pixel.



**Fig. 2.** Another example of a multivariate image set. The image series is composed of 190 subimages (unit cells) extracted from a high-resolution transmission electron microscope image of a GaAs–GaAlAs interface. Here, the aim is to analyse the differences between subimages and to deduce from them the variation of composition across the interface. Subimages are digitized as  $25 \times 25$  pixels and can thus be described by a vector in a 625-dimensional space.

$$Y = X \cdot X^t \quad (1)$$

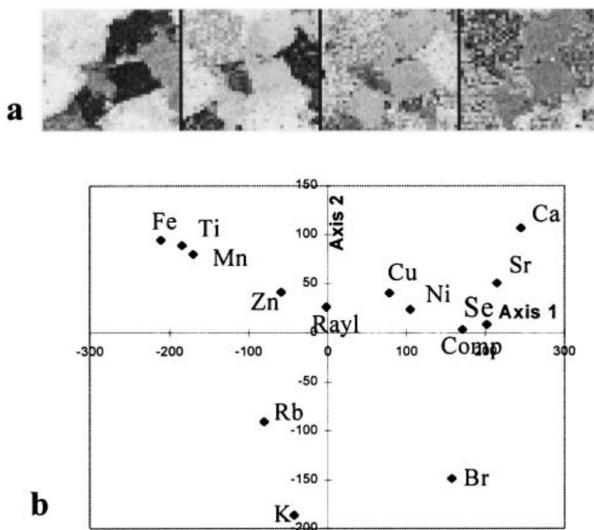
where  $X^t$  indicates the transposed matrix,  $Y$  is the variance–covariance matrix (variances of the images are along the diagonal, and covariances, which represent the exchange of information between pairs of images, are off the diagonal).

The next step consists of computing the eigenvalues (and eigenvectors) of the variance–covariance matrix. The eigenvectors then correspond to the new representation space. The associated eigenvalues are proportional to the strength of the corresponding eigenvectors in the variance–covariance matrix, that is to say the amount of information carried by the new direction of representation (thus, eigenvalues are sorted in descending order). Note that the nature of an eigenvector is analogous to that of an original individual (spectrum, image, ...). They can thus be displayed as spectra (eigen-spectra) or images (eigen-images), which can help in their interpretation.

Now, an original individual can be described as a linear combination of the eigenvectors:

$$X_i = \sum_j \alpha_{ij} A_j \quad (2)$$

where  $\alpha_{ij}$  represents the weight (or score) of the object  $i$  on the axis (eigenvector) number  $j$ . The scores of the different



**Fig. 3.** Results of applying correspondence analysis to the series of 14 microanalytical images (Fig. 1): (a) first four factorial images; (b) scores of the 14 images onto the first two factorial axes. See text for the interpretation of these results.

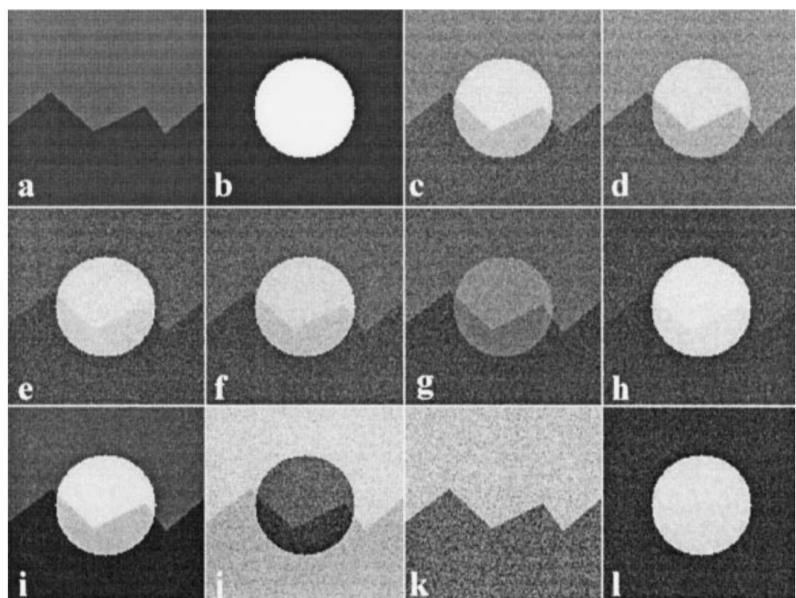
objects can thus also be displayed (for display purposes, two scores,  $\alpha_{ij}$  and  $\alpha_{ij'}$ , are often represented simultaneously, for all objects  $i=1 \dots N$ ). The two display possibilities (of objects and of objects' descriptors) often help to interpret the complete data set in terms of sources of information: how many sources are present? what do they represent?

It should be stressed that the true number of sources of information ( $M$ ) is often smaller than the number of components ( $N$ ) in the experimental set. Thus, a large

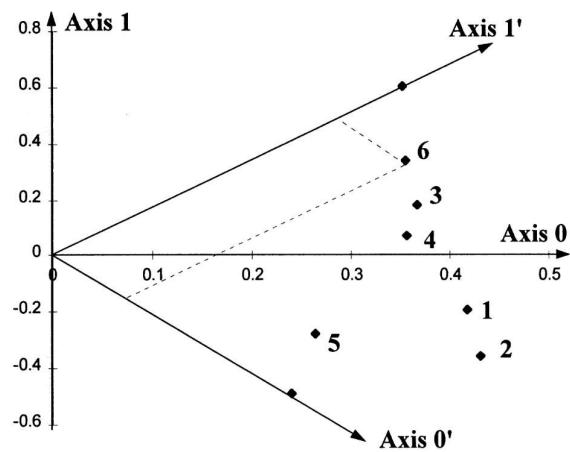
compression of information can be performed. Since the eigenvectors are orthogonal, noise uncorrelated with the useful signal is rejected into specific components, which can thus be eliminated after proper inspection. The same is true for experimental artefacts (Hannequin & Bonnet, 1988; Trebbia & Mory, 1990). Therefore, the next step may be to reconstitute the data set after selecting some 'useful' components and discarding some 'useless' ones. Since the decomposition is linear, there is no difficulty in following the reverse path (Bretaudiere & Frank, 1986) for reconstituting a 'filtered' data set.

As an example, I will consider the application of LMSA to sets of images. After its introduction in electron microscopy by the groups around Frank and van Heel, this technique has also been applied successfully in the domain of materials science and physics (Trebbia & Mory, 1990; Van Espen *et al.*, 1992; Geladi, 1992; Rouvière & Bonnet, 1993; Quintana & Bonnet, 1994a,b; Aebersold *et al.*, 1996; Trebbia, 1996). Thus, I will just describe briefly how MSA can be applied to an example such as the one represented in Fig. 1 (series of microanalytical maps). Then, I will introduce the discussion concerning the need to go further than the orthogonal LMSA described above.

The results of applying the orthogonal MSA to the 14 images of Fig. 1 are displayed in Fig. 3: Fig. 3(a) represents the first four factorial images obtained after applying correspondence analysis (one variant of LMSA: see Trebbia & Bonnet, 1990, for an extended description) to this data set. Figure 3(b) represents the scores of the 14 images on the first two factorial axes, which account for 71% and 20% of the total variance, respectively. Altogether, these two figures allow us to understand the content of the data set.



**Fig. 4.** An illustration of the need to go from orthogonal MSA to oblique MSA when quantitative results are expected. (a,b) Two basic (reference) images. (c–h) Six images obtained from a linear combination of the two basic images. These images are supposed to represent an experimental multivariate data set (Poisson noise was added to each image independently). (i,j) The first two orthogonal factorial images were obtained after principal components analysis. The corresponding scores are displayed in Fig. 5. (k,l) The two oblique factorial images were obtained after oblique analysis. They compare very well with the original basic images (a,b).



**Fig. 5.** Scores of the six images (Fig. 4c–h), numbered 1–6, on the first two orthogonal factorial axes (0 and 1) obtained after principal components analysis. These scores (coordinates on orthogonal axes) do not correspond to the weighting factors used for the simulation. Thus, PCA is not a quantitative method in this situation (because the two sources of information (Fig. 4a,b) are partly correlated). After oblique analysis, the two axes  $0'$  and  $1'$  are obtained. At which point, the scores (coordinates) on these two axes correspond perfectly to the weighting factors. Thus, oblique analysis is a quantitative method (provided, of course, the experimental images satisfy the underlying assumption of linear combination).

The first source of information, represented by the factorial axis number 1, opposes Ca and Sr (right of Fig. 3b) to Fe, Ti and Mn (left of Fig. 3b). This corresponds to the spatial localization displayed in the first factorial image in Fig. 3(a) (left). The second source of information, represented by the factorial axis 2, opposes Ca, Fe, Ti, Mn (top of Fig. 3b) to K (bottom of Fig. 3b). This corresponds to the spatial localization displayed in the second factorial image in Fig. 3(a). Altogether, we can anticipate that there are finally three groups of different regions within the analysed area. This will become more evident when other tools will be used for analysing this data set (see the following sections).

For introducing the need to go towards oblique analysis, I will first choose a simple example. I would like to consider the set of simulated images displayed in Fig. 4. These images were simulated according to the following protocol: first, two basic images (representing two sources of information) are created (Fig. 4a,b); then, these basic images are linearly combined to produce six new images, which are supposed to constitute the experimental data set (Poisson noise was added in order to produce a more realistic data set) (Fig. 4c–h). These six images were submitted to principal components analysis (PCA), which means that no normalization was applied to the data set before the variance–covariance analysis. The first two eigen-images obtained are displayed in Fig. 4(i,j) while the scores of the different images on the principal axes 1 and 2 are displayed in Fig. 5 (numbers 1–6

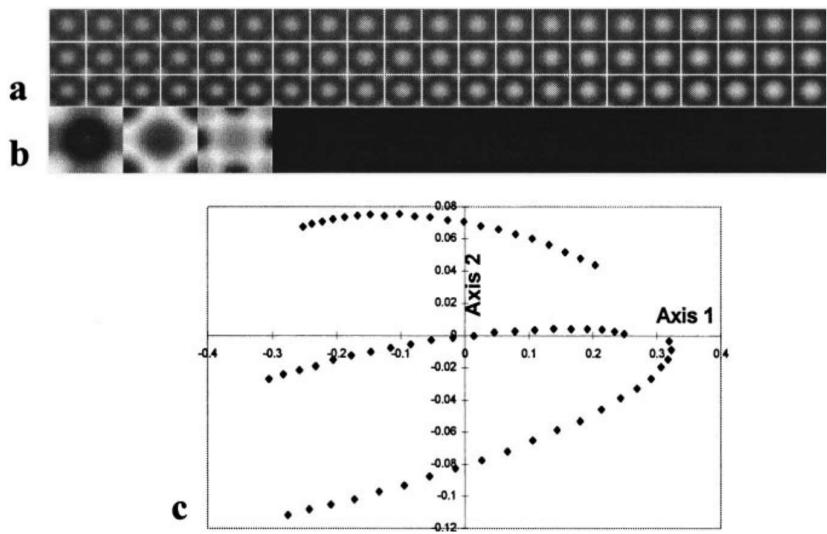
represent the six images 4c–h). One can notice that, although displaying some similarity with the two basic images, the two eigen-images are not strictly equivalent (the eigen-images are still a mixture of the two sources). Thus, the scores on the two principal axes cannot be used as estimates of the weighting factors. One can reformulate this by saying that, in general, orthogonal LMSA is not a quantitative method (see also, for instance, the comment of T. Walker in the discussion of Trebbia, 1996).

The reason for this drawback is that LMSA decomposes the data set into orthogonal components. On the other hand, the real sources of information have very little chance to be orthogonal. Thus, the eigenvectors do not, in general, represent the basic sources of information faithfully and the scores on the principal components do not correctly represent the weighting factors.

If one wants to access the elementary sources of information, one must perform an additional step after the orthogonal decomposition. This step is often called oblique analysis or factor analysis (Malinowski & Howery, 1980) because it consists of rotating the significant orthogonal axes until they are consistent with the nature of the basic information sources. Contrary to the orthogonal analysis which is completely assumption-free (except for the choice of one of the variants of LMSA), oblique analysis implies that some additional information is injected by the end-user. This extra information (see Trebbia & Bonnet (1990) for a discussion on the role of extra information in image processing) can take many different forms, indicating that many variants of oblique analysis have been suggested, ranging from completely interactive forms to completely automatic variants. In completely interactive variants, the end-users must supply the basic components (or, equivalently, they must specify the scores of  $M$  images onto the  $M$  different basic components). Examples of this way of proceeding can be found in Garenström (1986) and Sarkar *et al.* (1993). In chemometrics, this kind of approach is called partial least-squares (PLS) (Geladi & Kowalski, 1986) or principal component regression (PCR) (Esbensen *et al.*, 1992).

For performing completely automatic oblique analyses, some additional knowledge concerning the basic components and the scores of the individuals on them must be assumed. For instance, a frequent assumption is that they both verify a positivity constraint (Di Paola *et al.*, 1982; Benali *et al.*, 1994). With this assumption, several algorithms can be used for performing the rotation of axes automatically (oblimax method, varimax method, etc.).

Returning to the example illustrated in Figs. 4 and 5, we can improve the results if we can provide some extra knowledge. For instance, if we know that image 1 is a linear combination of the two basic images with weighting factors 1 and 0.5, and image 2 is a linear combination with weighting factors 1.2 and 0.4, the solution to the problem



**Fig. 6.** Multivariate analysis of a series of simulated HRTEM images of the GaAlAs system. (a) Series of  $21 \times 3$  images simulated at a constant defocus value ( $z = 10$  nm): the three rows correspond to three different thicknesses ( $t = 5, 6$  and  $7$  nm), the  $21$  columns correspond to  $21$  different compositions (content in Al varying from  $0\%$  to  $100\%$ ). (b) First three factorial images (in descending order). (c) Scores of the  $21 \times 3$  images on the first two factorial axes. Each original image is a linear combination of the first factorial images displayed in (b) with the weighting coefficients displayed in (c).

is trivial. The rotation matrix  $T$  can be obtained easily by linear regression. Each of its columns is given by

$$T_l = (\psi^t \psi)^{-1} \psi^t \psi_l \quad (3)$$

where  $\psi$  is the matrix of (orthogonal) eigenvectors and  $\psi_l$  is a column vector representing the weights of the different images on the oblique axis  $l$ .

The new (rotated) axes are also displayed in Fig. 5 (axes  $0'$ ,  $1'$  ...). The basic images obtained after rotation are displayed in Fig. 4(k,l) and, as expected, are now very close to the basic images (Fig. 3a,b). The new scores (coordinates of the experimental images onto the oblique axes) can be obtained ( $0.3$  and  $0.8$  for image 6, for instance) and are in fact very close to the true ones, which means that oblique analysis is a quantitative method.

It could be argued that the extra information provided by the end-user in this case is, in fact, a large part of the solution, and that this solution could have been obtained directly by linear regression, without the intermediate step of orthogonal LMSA. This is partly true, but the role of the orthogonal MSA step is to prepare a proper application of the linear regression (cleaning the data set by rejecting part of the noise in specific components, determining the number of components, etc.). Also, as mentioned above, less stringent extra information can sometimes be used. In the previous example, the same result can be obtained by using the non-negativity constraint only.

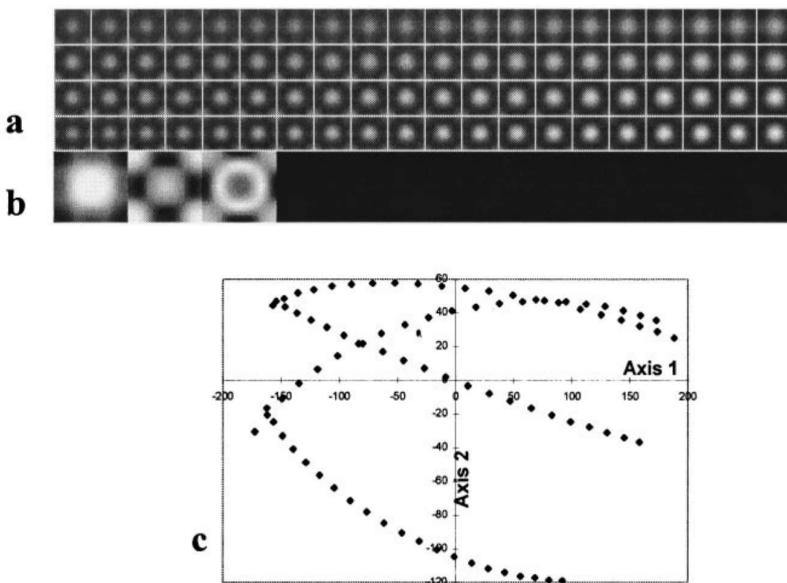
Of course, the procedure is not limited to two sources of information. In fact, one important step in the procedure consists of determining the 'true' number of sources of information (or degrees of freedom) present in the original data set. New tools are presently being studied for getting this piece of information.

Examples of this way of proceeding can be found in

Hannequin & Bonnet (1988), Bonnet & Trebbia (1992), Geladi (1992) and Trebbia *et al.* (1995).

To conclude this section, I will come back to the possible application of MSA to HREM. In Rouvière & Bonnet (1993), we performed a preliminary investigation of this possibility on experimental and simulated images of the GaAs/AlGaAs system. Concerning the experimental interface, we showed that LMSA was able to perform as well as the pattern recognition technique developed by Ourmazd *et al.* (1990) for the determination of the averaged chemical profile across an interface. In addition, it offers significant advantages which come mainly from the fact that intermediate results can be obtained and displayed (namely, the different eigen-images). These intermediate results may be very useful for a detailed interpretation of the data set content. For instance, we were able to depict a very small drift of some cells relative to others, simply because the second eigen-image was not centrosymmetric. These intermediate results provide a clear advantage over methods which are based only on the 'blind' computation of a single parameter.

However, it should be stressed that the chemical profile across the interface has been easily obtained (after careful registration) only because the entire information contained in the data set was concentrated on one factorial component only (axis 1 represents the variation in Al concentration). In this case, the results are directly quantitative. If this were not the case (if a source of information other than Al concentration were present), none of the scores on the different orthogonal factorial axes would represent the Al concentration. In order to obtain this concentration, it would be necessary to perform oblique analysis, that is to rotate the axes in such a way that one oblique axis represents the concentration and the other



**Fig. 7.** Multivariate analysis of a series of simulated HRTEM images of the GaAlAs system. (a) Series of  $21 \times 4$  images simulated at constant thickness ( $t = 6$  nm): the four rows correspond to four different defocus values ( $z = 5, 10, 15$  and  $20$  nm), the 21 columns correspond to 21 different compositions (content in Al varying from 0% to 100%). (b) First three factorial images (in descending order). Note: these eigen-images look similar to the negatives of those in Fig. 5(b). The sign of eigen-images does not have any meaning by itself: it must be combined with the sign of the scores for a proper interpretation. (c) Scores of the  $21 \times 4$  images on the first two factorial axes. Each original image is a linear combination of the first factorial images displayed in (b) with the weighting coefficients displayed in (c).

oblique axis represents the other source of information (thickness effect, for instance).

This point is illustrated below. [001] HREM images of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  crystals were simulated by J. L. Rouvière (CEA Grenoble, France) for different compositions and thicknesses ( $8 \text{ nm} \leq t \leq 12 \text{ nm}$ , around the optimum thickness defocus condition for chemical analysis) (see fig. 4 in Rouvière & Bonnet, 1993). Figure 6(a) shows simulated images obtained for three different thicknesses (rows) and 21 values of the Al concentration (columns). After correspondence analysis, it appears that the whole information is spread onto two orthogonal eigenvectors, representing 91% and 9% of the total variance. The first three factorial images are displayed in Fig. 6(b) and the scores onto the first two axes are displayed in Fig. 6(c). As expected, one can see that the main information source (chemical composition) is conveyed by axis 1, and the second source of information (thickness effect) is mostly conveyed by axis 2. Unfortunately, these two sources of information are not completely orthogonal (otherwise, the three curves should be parallel) and thus the coordinate on axis 1 is not directly related to the chemical composition, whatever the thickness. The same is true when a focus variation is added to the variation in composition. This is illustrated in Fig. 7. Figure 7(a) represents images of cells simulated for four defocus values (rows) and 21 composition values (columns). The first three factorial images are displayed in Fig. 7(b) and the scores on the first two axes (which represent 63% and 18% of the total variance, respectively) are displayed in Fig. 7(c). Again, one can see that the effects of composition and defocus are not completely independent. As expected, the effect of defocus is even more severe than the effect of thickness. Although this fact can be considered as a

drawback, it could also be used in a positive manner when comparing experimental images to simulated ones: instead of using one single experimental image for this comparison, it would probably be better to consider focus series (for experimental and for simulated images). Then, performing multivariate statistical analysis on the mixed series (one experimental focus series and several simulated focus series), one can expect to be able to identify the correct structure (or composition) according to its trace in the factorial space spanned by axes 1 and 2.

#### Nonlinear multivariate statistical analysis (NLMSA)

As described above, the aim of MSA is to reduce (or concentrate) the information contained in a multidimensional data set into a set of reduced dimension. Besides linear methods, nonlinear methods can also be envisaged. Besides projection (also called mapping) methods, other tools are also available for analysing these data sets. Classification methods, for instance, may play an important role in pattern recognition and/or segmentation. I will describe some methods belonging to these two groups (mapping and classification).

#### Nonlinear mapping methods

I have described above how LMSA allows us to concentrate a multidimensional set (which can be represented in a parameter space  $R^N$ ) into a data set with lower dimensionality ( $R^M$ ,  $M < N$ ). This was done by rotating the axes of representation so that the variance along the new low order axes is maximal.

The same result can be obtained (sometimes with more efficiency, which means with a lower final dimensionality) by nonlinear mapping, a transformation which does not involve only linear combinations of the original coordinates. Several nonlinear mapping methods have already been suggested: heuristic approaches, minimization of a distortion measure and neural networks approaches.

#### (a) A heuristic approach

In Bonnet *et al.* (1995), we have suggested a heuristic approach in order to generalize the concept of scatterplot often used when the data set is composed of two or three images only. The idea is first to reduce the  $N$ -dimensional data set to two synthesized images. These images are obtained by coding the information 'seen by observers looking at the  $N$ -dimensional parameter space'. This means that the grey levels of the transformed images are the distances from the  $N$ -dimensional pixels of the original data set to the observers. The scatterplots can be built by combining two of these synthesized images. The choice of the coded distance depends on the nature of the observers chosen, which is of a heuristic nature. For instance, they can be the corners (or the diagonals) of the  $N$ -dimensional hyper-rectangle. This method is easy to implement and very fast, but its drawback is that there are many possibilities for choosing the observers. Some of them lead to interesting results (displaying a large part of the information contained in the original data set), but others are not as interesting (part of the information is hidden). Since our aim was mainly exploratory data analysis (more specifically, discovering the number of clusters in the scatterplot built from two mappings of the data set), our choice between 'good' pairs of maps and 'bad' pairs was essentially made on a visual basis. However, this may prove insufficient and more objective methods may be needed.

#### (b) Nonlinear mapping based on the minimization of a distortion criterion

A long time ago, American psychologists developed methods for reducing the dimensionality of 'simple' multidimensional data sets (such as the results of enquiries, soundings, etc.). These methods include multidimensional scaling (MDS) (Kruskal, 1964), Sammon's mapping (Sammon, 1969) and other variants.

They start from the simple idea that when projecting  $N$ -dimensional objects onto a space of reduced dimension, the distances between these objects must be preserved as far as possible. Thus, a criterion  $E$  describing this preservation may be used to evaluate the quality of the mapping. It can also be used as an intermediate for designing the mapping algorithm itself. The different variants of this approach rely on different definitions of the preservation criterion. For

instance, MDS is based on the criterion

$$E_{\text{MDS}} = \sum_{i < j} (D_{ij} - d_{ij})^2 \quad (4)$$

where  $D_{ij}$  is the distance between objects  $i$  and  $j$  in the original space and  $d_{ij}$  is the distance between the same objects in the projection space.

Sammon's mapping is based on the criterion\*

$$E_S = \sum_{i < j} (D_{ij} - d_{ij})^2 / D_{ij}. \quad (5)$$

As an example, the Sammon's mapping algorithm consists of moving each object projection (in the  $M$ -dimensional space) according to the rule deduced from a minimization of  $E_S$  according to the Newton's approach:

$$\vec{x}_{i+1} = \vec{x}_i + \alpha \vec{A} / B \quad (6)$$

where

$$A = \frac{\partial E}{\partial x_{il}} = - \sum_j \left( \frac{D_{ij} - d_{ij}}{D_{ij} d_{ij}} \right) (x_{il} - x_{jl})$$

and

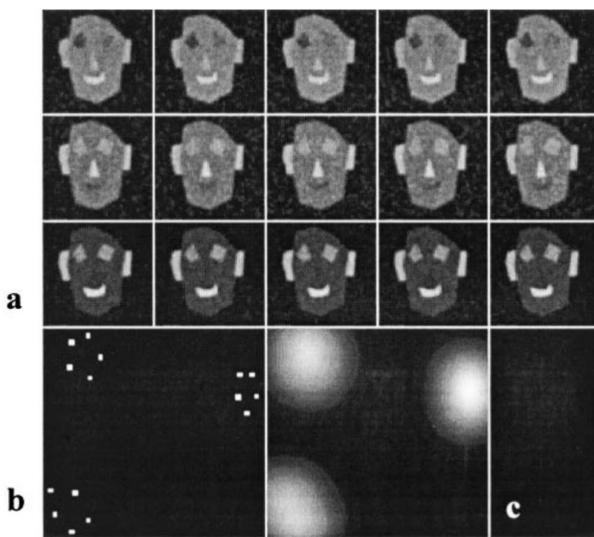
$$B = \frac{\partial^2 E}{\partial x_{il}^2} = - \sum_j \frac{1}{D_{ij} d_{ij}} \left[ D_{ij} - d_{ij} - \frac{(x_{il} - x_{jl})^2}{d_{ij}} \left( 1 + \frac{D_{ij} - d_{ij}}{d_{ij}} \right) \right].$$

This approach can be used for mapping the different types of multidimensional data discussed above. In the case of spectra or images, there are two possibilities: projection of spectra/images or projection of energy channels/pixels. Below, we illustrate these two possibilities.

*Projection of images.* The aim is to identify several groups of images within an image set. This approach is often used in high-resolution electron microscopy of isolated macromolecules, where it is necessary to compute an average of similar images (views) before performing three-dimensional reconstruction. Although only linear mapping is usually performed in this context, one application of nonlinear mapping (Radermacher & Frank, 1985) can be noticed. The approach could also probably take place in materials science problems (classification of unit cells in HRTEM of crystals, for instance).

Figure 8(a) displays a set of 15 noisy simulated images. These images are composed of  $64 \times 64$  pixels, which means that each of them can be represented in  $R^{4096}$ . Figure 8(b) represents the result of the mapping onto a two-dimensional space ( $R^{4096} \rightarrow R^2$ ). Each point represents an image. From this, it is easy (a) to recognize that there are three groups of images, (b) to know which image belongs to which group

\*Denominators usually appearing in the expressions of the MDS and Sammon's criteria have been suppressed, since they are constant and thus are not involved in the minimization process.



**Fig. 8.** Illustration of Sammon's mapping procedure applied to images (or subimages). (a) Fifteen simulated images belonging to three different classes. (b) Mapping of the 15 images (size  $64 \times 64$ : dimension 4096) onto a two-dimensional parameter space according to the Sammon's algorithm ( $R^{4096} \rightarrow R^2$ ). (c) Probability density function estimated from the result of mapping, according to the Parzen window technique. From this estimated pdf, it can be recognized that three classes of images are present. Other tools allow us to determine which images belong to the different classes (see below).

(automatic classification procedures will be described in the next section) and (c) to obtain an averaged view of the images within each class.

(Figure 12(a) is the result of mapping the 190 subimages of Fig. 2 onto a two-dimensional space ( $R^{625} \rightarrow R^2$ ). The number of classes is more difficult to estimate in this situation.)

*Projection of pixels.* The aim is now to identify several groups of pixels within a multivariate image. This approach can be used to segment such an image (which represents different signals recorded from the same area of a specimen, in any type of microanalysis for instance), in order to visualize the different regions with homogeneous composition, and perform quantification in terms of area or spatial distribution of the different phases.

(Figure 11(a) represents the result of mapping the 1560 pixels of images in Fig. 1 onto a two-dimensional space ( $R^2$ ) according to the Sammon's algorithm. One can easily see that there are three main clouds (one of them could still be split into two subclasses). From this information, a classification of the different pixels into three (or four) classes can be performed, either interactively, by selecting the boundaries of classes in the two-dimensional map with a

graphic mouse (this approach is often named 'interactive correlation partitioning': ICP), or automatically (see the next section: I suggested naming this approach 'automatic correlation partitioning': ACP) (Bonnet, 1995b).)

### (c) Neural network approach to nonlinear mapping

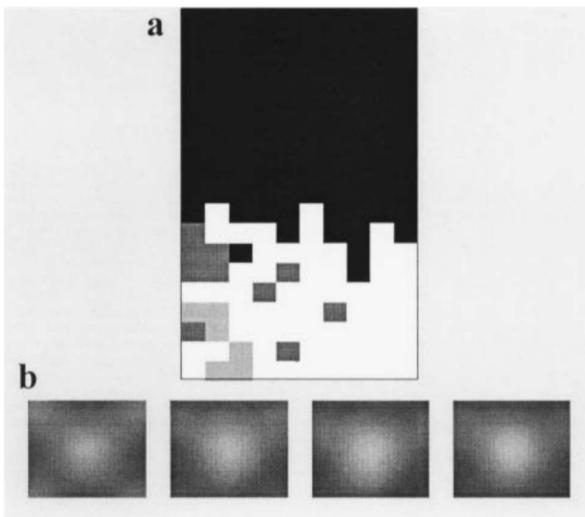
Artificial neural networks (ANNs) have been developed as an alternative to classical statistical analysis for solving pattern recognition problems. One of the advantages they are assumed to possess is their ability to cope with nonlinear problems as well as with linear ones. Several ANNs have been devised for solving the mapping problem. The self-organizing map (SOM) (Kohonen, 1984) and the Diabolo network (Kramer, 1991) are examples of such nets. I will describe and illustrate the SOM technique only. It consists of defining a set of neurones, arranged on a regular grid (a square grid is often used, and corresponding to a projection onto  $R^2$ , although one-dimensional or three-dimensional grids have also been used). An  $N$ -dimensional vector is attached to each neurone of the grid composed of  $P^M$  neurones ( $M=2$  in the case of the square grid). The aim of the first phase (often called the training phase, although there is no learning in the sense of supervised procedures) is to dispatch all the objects (pixels, images, spectra or energy channels) available among the neurones in such a way that the vector associated with every neurone is similar to all the vectors describing objects associated with this neurone. For this purpose, a 'competitive' learning procedure is used: each original object is compared to all the prototypes (neurones) and a 'winner' is selected according to a comparison rule (distance, scalar product ...). Then the neurones (prototypes) associated with the winner and its neighbours are updated according to their similarity with the object being processed:

$$\vec{p}_{t+1} = \vec{p}_t + \alpha(t+1) \cdot (\vec{o} - \vec{p}_t) \quad (7)$$

where  $\vec{p}_t$  and  $\vec{p}_{t+1}$  are the prototype vectors at iteration  $t$  and  $t+1$ ,  $\vec{o}$  is the processed object and  $\alpha(t+1)$  is the learning rate at iteration  $t+1$ .

This process is repeated until convergence, with empirical rules for decreasing the size of the neighbourhood and the updating coefficient  $\alpha$ . At the end of the process, all objects are mapped in such a way that similar objects are represented by neurones close to each other, thus preserving the topology of the original data set, hence the name 'topological map'.

An example is given in Fig. 9, which represents the mapping of the 190 unit cells (Fig. 2) onto a one-dimensional space ( $R^1$ ) composed of four neurones. Figure 9(b) represents the four prototypes obtained after the training phase. Figure 9(a) represents the labelling. Each unit cell is represented by the number of its associated neurone.



**Fig. 9.** Results of the classification of the 190 subimages (see Fig. 2) by the self-organizing map procedure: (a) results of classification into four classes (the grey levels correspond to the labels of the different subimages after classification); (b) prototypes associated with the four neurones of the one-dimensional map.

This approach was used by Marabini & Carazo (1994) for classifying sets of macromolecules. The advantages of this method as compared to the linear MSA approach are discussed in their papers. Further studies are necessary for comparing this mapping technique with other nonlinear ones in different situations.

#### Automatic clustering

Besides dimensionality reduction, another task consists of classifying the different objects which constitute the

multivariate data set. The purpose of this task is often to compute an average of the different objects which belong to the same class (Van Heel & Frank, 1981). In microanalysis, it may also be to segment multivariate images into regions of homogeneous composition, and subsequently to characterize them (Bonnet *et al.*, 1997).

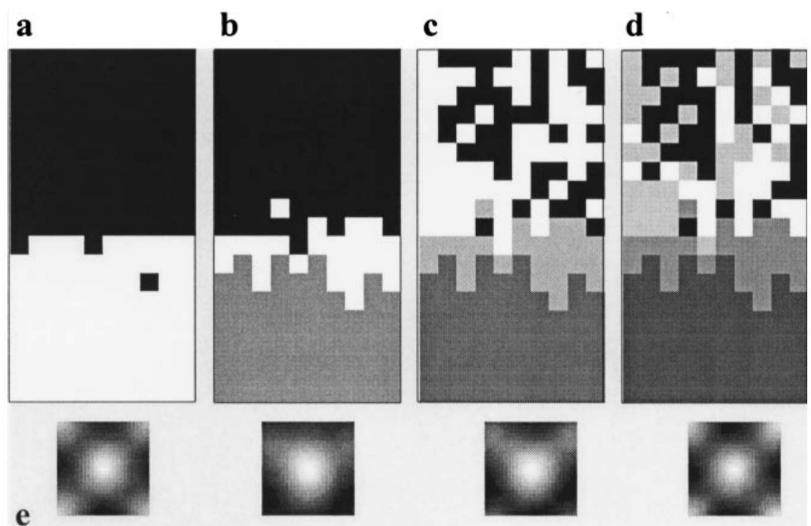
Some of the methods for automatic clustering work in the original multidimensional space ( $R^N$ ) while others can only work in a space of reduced dimensionality, which implies a preliminary mapping ( $R^N \rightarrow R^M$ ).

I will describe some of the methods available for performing automatic clustering. The name 'clustering' refers to unsupervised classification. I will not address the task of supervised classification in this paper. Some examples of the applications of supervised classification for microscope images can be found in Tovey *et al.* (1992), Livens *et al.* (1996) and Aebersold *et al.* (1996). These methods range from classical statistical methods to new methods working directly in a reduced parameter space and to neural networks approaches. All of them aim to classify all the objects available into a limited number of classes on the basis of their similarity/dissimilarity. (Note that contrary to supervised methods, the number of classes is unknown and finding this number is also a part of the clustering procedure.)

#### (a) A classical statistical method: the C-means techniques

The aim of this technique is to dispatch objects into a selected number of classes, and to modify the statistical characteristics of the classes accordingly, until the sum of the within-class variances is minimized.

The practical algorithm is described in many textbooks (Duda & Hart, 1973) and in Bonnet (1995b). Thus, I will not reproduce it here.



**Fig. 10.** Results of the classification of the 190 subimages (see Fig. 2) by the K-means procedure: (a) results of classification into two, three, four and five classes (the grey levels correspond to the labels of the different subimages after classification). (b) Averaged subimages after classification into four classes.

Finding the number of classes consists of performing the classification for a variable number of classes and trying to optimize a criterion describing the quality of the partition. Unfortunately, the many different criteria suggested do not often lead to the same result, except for the trivial case of very compact and well-separated classes. A partial list of the criteria suggested can be found in Bonnet *et al.* (1997). A criterion which often gives good results when the number of classes is greater than three is the second derivative of the total within-class variance.

Note that other graphical tools can also help in deciding the number of classes, depending on the purpose of the classification process. For instance, it is possible to display the correlation coefficient between any pixel vector and the vector representing its class centre (see Bonnet *et al.*, 1997, for more details). Once classification is made, averaging can also be performed: the pixel vector is replaced by the pixel vectors averaged over all pixels belonging to the same class. This procedure leads to a dramatic improvement of the signal-to-noise ratio, without the degradation of the image resolution resulting from the application of low-pass filters.

Of course, this approach can also be used for classifying images (or spectra) instead of pixels. This approach was recently attempted by Aebersold *et al.* in the context of HRTEM images of the  $(\gamma, \gamma')$ -interface of an Ni base superalloy, but this attempt was made within the framework of supervised classification (which implies training the classifier with examples). I illustrate the procedure within the framework of unsupervised classification (clustering). Figure 10(a)–(d) displays the result of classifying the 190 subimages (Fig. 2) into a varying number of classes ( $C = 2, 3, 4$  and 5). Figure 10(e) displays the result of averaging the unit cells which belong to the same class, for a classification into four classes (Fig. 10c).

*(b) A variant of the C-means technique: the fuzzy C-means technique*

Although it works well for easy classification problems, the C-means technique becomes rather sensitive to some pitfalls when the problem is more complicated (overlapping classes...). A variant of the technique was proposed within the framework of fuzzy logic. (In order to avoid confusion, the classical technique is now called the hard C-means (HCM) technique.) The idea in the fuzzy C-means algorithm (FCM) is that, during the iterative process, objects to be classified should not be attached to a single class, but attached to all the  $C$  classes, with different degrees of membership ( $0 \leq \mu_{ic} \leq 1$ , with  $\sum_c \mu_{ic} = 1$ , where  $i$  is the object number and  $c$  is the class number).

It can be shown that this approach is often less sensitive to the initialization and less prone to becoming stuck in the local minima of the objective function to be minimized. More details concerning this variant and the criteria which

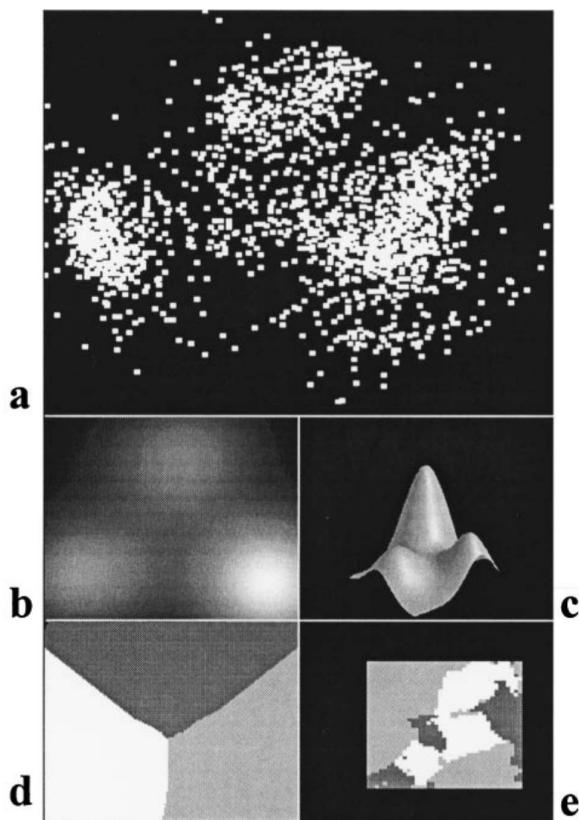
can be used for determining the number of classes can be found in Bonnet *et al.* (1997).

*(c) An approach combining mapping and clustering, which do not make assumptions concerning the shape of the classes*

The two approaches (hard C-means and fuzzy C-means) described above make assumptions concerning the shape of the classes (these assumptions are made implicitly through the use of the concepts of class centres and of distances to these centres). These methods were developed in the context of classical statistics, starting with the idea of Gaussian mixtures. Standard implementations assume that classes have hyperspherical shapes (when the Euclidean distance is used) or hyperelliptical shapes (when the Mahalanobis distance is used). Variants are able to cope with linear or hyperplane shapes (Krishnapuram & Freg, 1992). However, it is very difficult, within this framework, to cope with a mixture of different class shapes.

If one wants to do so, one has to move towards other kinds of techniques. One such approach consists of estimating the probability density function (pdf) of the underlying distribution and of partitioning the parameter space according to the estimated pdf (Zhang & Postaire, 1994; Herbin *et al.*, 1996). The first step can be accomplished by the Parzen technique: the pdf is estimated as the convolution of the point distribution (one object is represented by one point in the parameter space) with a smoothing kernel (Gaussian kernel or Epanechnikov kernel). This result in a smoothed distribution, where several modes can be identified as representing several classes (see Fig. 11b,c).

The second step can be performed according to several approaches. We have recently suggested using tools available within the framework of mathematical morphology (but which work on the estimated pdf instead of on the original images). In Herbin *et al.* (1996), we have suggested performing iterative thresholding of the estimated pdf and applying the skeleton by influence zone (SKIZ) to the thresholded pdf, in order to define the zones of influence of the modes and thus partitioning the parameter space. In Bonnet *et al.* (1997) we have suggested replacing the binary SKIZ with its grey-level equivalent: the watershed. The watershed technique is a powerful tool (Beucher & Meyer, 1992) for partitioning a space described by a pseudocontinuous function. Here, the seeds for the different regions (one region in the parameter space represents one class) are given by the modes of the estimated pdf. Then, a flooding process is used for attaching the points of the parameter space to one of the classes (waiting lists are used, as described by Vincent, 1990). When all points are attached, the partition of the parameter space is completed. The remaining step is trivial: the real objects are classified according to their position in the parameter space.



**Fig. 11.** Results of automatic multivariate image segmentation by the watershed-based method. (a) Result of Sammon mapping: the 1560 pixels (Fig. 1) in a space of dimension 14 are projected onto a two-dimensional space: each white point represents a pixel. One can identify three main clouds of points, representing three classes of pixels. (b) Smoothing of the parameter space is performed according to the Parzen technique. The result is an estimation of the probability density function. The three modes of the estimated pdf are apparent. (c) Same result as in (b), but represented as a three-dimensional plot. (d) Partition of the parameter space (a, b and c) with the watershed technique. Each position (of the parameter space) is labelled (i.e. associated to one of the three modes) according to the estimated values of the probability density function. (e) Partition of the image space: each pixel is labelled (the grey levels represent the labels) according to the position of its projection (see a) in the labelled parameter space (d).

It must be stressed that this method does not make any assumption concerning the shape of the classes (represented by clusters in the parameter space): the distance of a point (in the parameter space) to the mode of the class is not used in the partitioning process. Only the estimated pdf value at this point is used to insert the point in the waiting list.

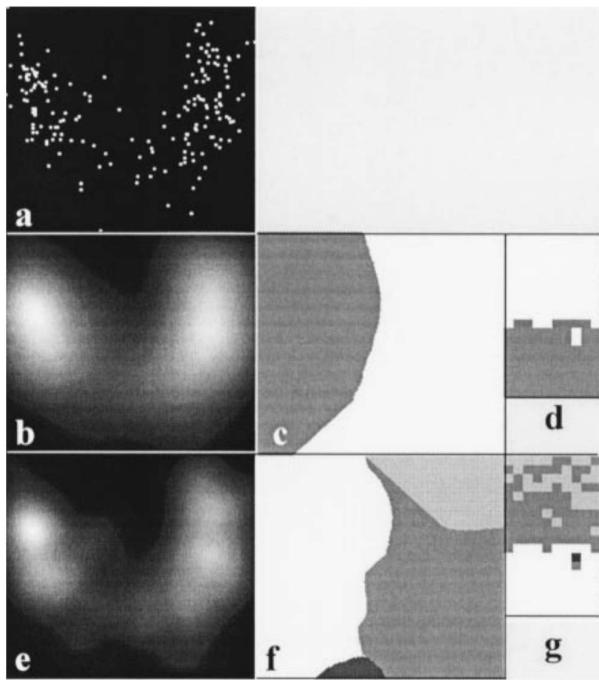
It must also be stressed that this method cannot be applied directly to data in a high-dimensional space (neither the Parzen window technique nor the watershed technique can

be applied in  $R^{1024}$  for instance, when images with  $32 \times 32$  pixels are to be classified). Thus, in general, this technique must be combined with one of the (linear or nonlinear) mapping techniques described in the first part of this paper.

One key point of this method (as of any unsupervised classification methods) is the determination of the number of classes. This results from the choice of the parameter of the smoothing kernel. At the present state of development of this method, a curve is drawn displaying the number of modes of the estimated pdf as a function of this parameter (the standard deviation of the Gaussian kernel). Then, values of the parameter corresponding to extended plateaus of the curve can be selected. In general, they correspond to several possibilities for the number of classes (or subclasses) present in the data set. In the future, we hope to be able to select the value of the smoothing parameter automatically.

*Illustrations of the method:* (i) *classification of pixels.* Figure 11 shows the result obtained from the segmentation of the multivariate image (Fig. 1). Figure 11(a) is the result of mapping the pixels onto a two-dimensional space (from  $R^{14}$  to  $R^2$ ). Figure 11(b) displays the pdf estimated from Fig. 11(a) (by the Parzen method). The pdf is displayed in three dimensions (the third dimension represents the amplitude of the estimated pdf) in Fig. 11(c). Figure 11(d) represents the labelled parameter space: a label is associated with each mode of the estimated pdf (Fig. 11b). Then each label is propagated (by the watershed technique) so that each position of the quantified parameter space is attached to one of the classes. Figure 11(e) represents the final result: each pixel of the multivariate image is associated with a label (and represented by a grey level) according to the coordinates of its projection onto the parameter space. As such, the homogeneous regions are represented by the same label. Further quantification can take place: for instance, the percentage of the area occupied by the different regions can be computed.

(ii) *Classification of images.* Figure 12 shows the result of applying the watershed-based classification technique to the 190 subimages (Fig. 2). Figure 12(a) is the scatterplot built with the scores of these subimages onto the two first factorial axes (after applying correspondence analysis). Alternatively, a nonlinear mapping algorithm can also be used for this first step. Figure 12(b) (left) is the probability density function estimated from this point distribution according to the Parzen technique. One can see that two modes are apparent, which means that the data set can be split into two classes. The result of applying the watershed technique to divide the parameter space into two regions is displayed in the middle of the row (Fig. 12c). Finally, the different subimages may be classified according to their projection onto the parameter space (right of the row: Fig. 12d). Figure 12(e)–(g) displays the same kind of results, but with a smaller parameter of the Parzen kernel. In this



**Fig. 12.** Results of the classification of the 190 subimages (see Fig. 2) by the watersheds technique applied to the results of correspondence analysis. (a) Scatterplot built with the scores of the subimages (each point represents a subimage) on the first two factorial axes obtained after correspondence analysis. (b) Probability density function (pdf) obtained after application of the Parzen technique to the point distribution in (a). Two modes are obtained. (c) Labelling of the parameter space obtained by the watershed technique applied to the pdf. The two classes correspond to the two modes of the estimated pdf. (d) Labelling of the subimages according to the labelling of the parameter space. (e,g) Same as (b,d) but with a smaller standard deviation of the smoothing kernel. Four modes are observed in the estimated pdf, which results in four classes of subimages.

situation, four modes of the estimated pdf are obtained (left of the row), and four classes are displayed in the parameter space (middle of the row) and in the image space (right of the row).

#### (d) Neural network approaches to classification

One of the main applications of ANN is also automatic classification. Multilayers feedforward networks are essentially dedicated to supervised classification (a training phase uses prototypes for adjusting the weights of interconnections and a second phase exploits the results of training, in order to perform the classification of new data). Concerning unsupervised classification, several neural networks architectures have been suggested, which will only be mentioned

briefly here, but would justify a deeper study for comparison with classical approaches.

*ART-based neural networks.* The adaptive resonance theory (ART) is based on the classical concept of correlation, enriched by the neural network concepts of plasticity–stability (Carpenter & Grossberg, 1987). Simply, an ART-based neural network consists of defining as many neurones as are necessary to split an object set into several classes (one neurone represents one class) according to a predefined vigilance parameter  $\rho$ . Each time a new object is presented (each object is presented several times, until convergence), it is compared to all existing neurones. A winner is defined as the closest neurone to the object presented. If a similarity criterion with the winner is verified (resonance), the object is attached to the corresponding class and the neurone descriptor (vector) is updated. If not, a new neurone is created, whose descriptor is initialized with the object vector. Of course, the final number of classes is related to the value of the vigilance parameter  $\rho$ .

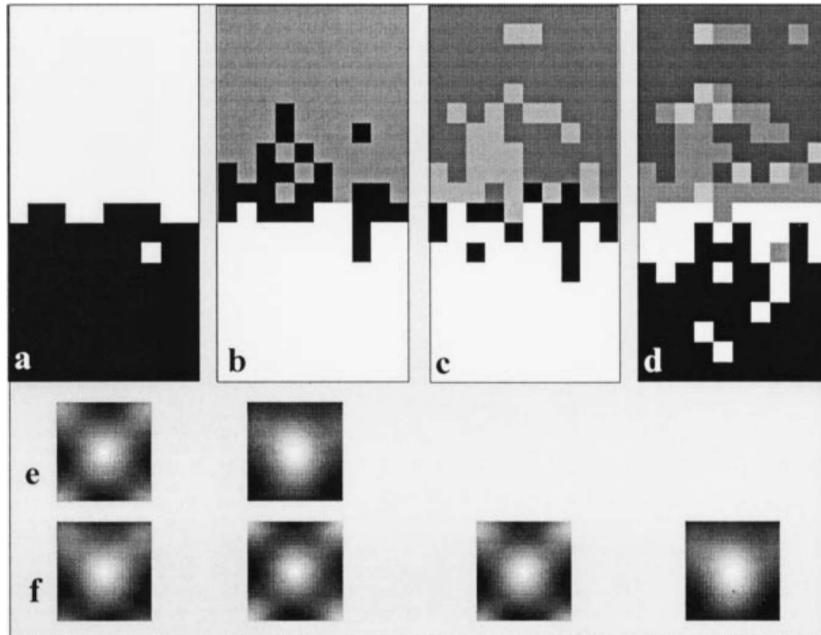
Examples of the applications of this neural network technique in chemical and physical sciences can be found in Wienke & Buydens (1995) and Gatts *et al.* (1995).

Figure 13 is an illustration of the possibilities of this approach for the classification of cells in HRTEM: Fig. 13(a)–(d) are the results of classification obtained with the ART algorithm for four different values of the vigilance parameter  $\rho$ , which led to classifications into two, three, four and five classes. Figure 13(e,f) display the averaged cell images obtained in the case of two classes (Fig. 13a) and four classes (Fig. 13c), respectively.

*Learning vector quantization techniques.* At the early stages, learning vector quantization (LVQ) techniques were developed in the framework of supervised techniques, hence the word 'learning' (see Kohonen, 1984, for instance). The aim was to define several prototypes for each class, on the basis of the training set, and then to classify unknown objects using an algorithm like the K-nearest neighbours (KNN) for instance.

Afterwards, the concept of prototypes was incorporated within the framework of unsupervised classification (clustering) through methods like Fuzzy-LVQ (FLVQ) (Bezdek & Pal, 1995) or generalized-LVQ (GLVQ) (Pal *et al.*, 1993). Briefly, these methods try to characterize each class with a prototype and to divide the whole data set into classes represented by this prototype.

From my point of view, all neural network approaches developed to this point for automatic clustering tasks suffer from one serious drawback: they all rely on the concept of class centre (or single prototype), which restricts their efficiency to hyperspherical or hyperelliptical clusters. Classes with arbitrary shapes cannot be handled properly.



**Fig. 13.** Results of the classification of the 190 subimages (see Fig. 2) by the neuromimetic ART-based procedure: (a–d) results of classification into two, three, four and five classes (the grey levels correspond to the labels of the different subimages after classification); (e) averaged subimages after classification into two classes; (f) averaged subimages after classification into four classes.

A useful jump will be accomplished when the concept of single prototype is replaced by the concept of multi-prototypes. Work in this direction is in progress.

### Conclusion

In this paper, I have reviewed several aspects of multivariate statistical methods applied to physical data analysis. I started from linear methods, which have now been in use for more than 10 years. I have shown that linear multivariate statistical methods allow us to compress high-dimensional data sets into a limited set of orthogonal components. This can be used for a qualitative interpretation of the sources of information present in the multivariable data set. It also allows us to detect experimental artefacts and to filter out these artefacts and a large part of the noise. In certain circumstances, it is also possible to use this decomposition for performing data interpolation or extrapolation (Bonnet *et al.*, 1992).

However, this orthogonal decomposition is often insufficient, because the real sources of information are rarely orthogonal. Thus, an extension to oblique analysis is often necessary when one wants to go from qualitative to quantitative interpretation. New developments are still necessary before oblique analysis can be used routinely in physical data analysis, especially when the number of factorial components is greater than two.

Besides linear multivariate statistical methods, I have shown that other tools, such as nonlinear methods, may also prove very useful. I have described two groups of such

tools: tools for mapping high-dimensional data sets onto low-dimensional data spaces and tools for performing automatic classification.

Tools for mapping ( $R^N \rightarrow R^M$ ) include statistical methods which attempt to minimize the distortion criterion (distances between objects in the destination space should resemble the distances between the same objects in the original space) as well as neural network methods, such as self-organizing mapping. These mapping tools can be useful for exploratory data analysis (finding structures in the data set), for display purposes or for preparing the subsequent classification step.

Automatic classification may be performed following a supervised or unsupervised strategy. I did not consider the former in this paper. Unsupervised classification (clustering) can be performed through different approaches: classical aggregation of objects around centres of classes, a variant of this aggregation method according to the general principles of fuzzy logic, clustering methods based on the estimation of the probability density function followed by the partition of the parameter space according to this pdf, and neural network approaches. There is still work to be done before complete confidence in the results of these automatic methods can be ascertained (see, for instance, the discrepancy obtained when trying to classify the 190 unit cell images into more than two classes). However, I would say that these methods can now be used to interpret qualitatively (and even on a quantitative basis) the multivariate data sets produced by many sensors in the field of physical sciences.

Some points which, from my point of view, need further developments, are:

- the determination of the intrinsic dimensionality of a data set: before attempting to perform dimensionality reduction (mapping), it would be useful to know the dimensionality of the data set, that is to say the dimension of the subspace spanned by the data. Some ideas have been formulated (the fractal dimension of the data set may be a candidate) but to my knowledge, they have not been studied in detail.
- oblique analysis (an extension of orthogonal MSA) should be compared with the source separation method, which provides independent components of a mixture instead of only orthogonal components (Jutten & Heraut, 1991; Comon *et al.*, 1991).
- also, nonlinear MSA methods (nonlinear PCA for instance: Karhunen & Joutsensalo, 1995) should be compared with the other nonlinear methods described in this paper.
- in this paper, I have neglected the problem of robustness, that is to say the sensitivity of the methods to the presence of outliers and noise. Some references dealing with this problem in the context of MSA should be consulted (Walczak & Massart, 1995; Liang & Kvalheim, 1996). Further studies of this problem must be conducted.
- MSA methods mostly consider the variance (and covariance) as the support of information in a data set. It is well known from information theory that this is not the only way to characterize information. Entropy and mutual entropy constitute other supports. Entropy-based MSA is the subject of a forthcoming paper.

### Acknowledgments

I am pleased to thank my colleagues M. Herbin and P. Vautrot who contributed to the development of the Parzen watershed classification technique. I would also like to thank K. Janssens and B. Wekemans for allowing me to use their X-ray fluorescence data set and to J. L. Rouvière for allowing me to use the experimental and simulated images of the GaAlAs system.

### References

- Aebersold, J.F., Stadelman, P.A. & Rouvière, J.L. (1996) Quantitative interpretation of HRTEM images using multivariate statistics: the case of the  $(\gamma,\gamma')$ -interface in a Ni base superalloy. *J. Microsc.* **62**, 171–189.
- Benali, H., Buvat, I., Frouin, F., Bazin, J.P. & Di Paola, R. (1994) Foundations of factor analysis of medical image sequences: a unified approach and some practical implications. *Image Vision Comput.* **12**, 375–385.
- Beucher, S. & Meyer, F. (1992) The morphological approach to segmentation: the watershed transformation. *Mathematical Morphology in Image Processing* (ed. by E. R. Dougherty), pp. 433–481. Dekker, Inc., New York.
- Bezdek, J.C. & Pal, N.R. (1995) Two soft relatives of learning vector quantization. *Neural Networks*, **8**, 729–743.
- Bonnet, N. (1995a) Processing of images and image series: a tutorial review for chemical microanalysis. *Mikrochim. Acta*, **120**, 195–210.
- Bonnet, N. (1995b) Preliminary investigation of two methods for the automatic handling of multivariate maps in microanalysis. *Ultramicroscopy*, **57**, 17–27.
- Bonnet, N., Herbin, M. & Vautrot, P. (1995) Extension of the scatterplot approach to multiple images. *Ultramicroscopy*, **60**, 349–355.
- Bonnet, N., Herbin, M. & Vautrot, P. (1997) Multivariate image analysis and segmentation in microanalysis. *Scanning Microsc. Suppl.* **11**, in press.
- Bonnet, N., Simova, E., Lebonvallet, S. & Kaplan, H. (1992) New applications of multivariate statistical analysis in spectroscopy and microscopy. *Ultramicroscopy*, **40**, 1–11.
- Bonnet, N., Simova, E. & Thomas, X. (1991) Application of multivariate statistical analysis to time dependent spectroscopy. *Microsc. Microanal. Microstruct.* **2**, 129–142.
- Bonnet, N. & Trebbia, P. (1992) Multi-dimensional data analysis and processing in electron-induced microanalysis. *Scanning Microsc. Suppl.* **6**, 163–177.
- Borland, L. & van Heel, M. (1990) Classification of image data in conjugate representation spaces. *J. Opt. Soc. Am. A*, **7**, 601–610.
- Bretaudiere, J.P. & Frank, J. (1986) Reconstitution of molecule images analysed by correspondence analysis: a tool for structural interpretation. *J. Microsc.* **144**, 1–14.
- Brun, N., Bonnet, N., Colliex, C. & Tencé, M. (1996) Automatic data processing procedures applied to electron energy loss spectroscopy spectrum-line profiles. *Proc. EUREM Dublin*.
- Carpenter, G.A. & Grossberg, S. (1987) ART2: self-organization of stable category recognition codes for analog input patterns. *Appl. Opt.* **26**, 4919–4930.
- Cazaux, J. (1993) A new quantification procedure for elemental mapping by X-ray (absorption) microscopy. *Microsc. Microanal. Microstruct.* **4**, 513–537.
- Colliex, C., Tencé, M., Lefèvre, E., Mory, C., Gu, H., Bouchet, D. & Jeanguillaume, C. (1994) Electron energy loss spectrometry mapping. *Mikrochim. Acta*, **114/115**, 71–87.
- Comon, P., Jutten, C. & Heraut, J. (1991) Blind separation of sources. Part II: Problems statements. *Signal Proc.* **24**, 11–20.
- De Jong, A.F. & Van Dyck, D. (1990) Composition determination from HREM images of substitutional alloys. *Ultramicroscopy*, **33**, 269–279.
- Di Paola, R., Bazin, J.P., Aubry, F., Aurengo, A., Cavailloles, F., Herry, J.Y. & Kahn, E. (1982) Handling of dynamic sequences in nuclear medicine. *IEEE Trans. Nucl. Sci.* **NS29**, 1310–1321.
- Duda, R.O. & Hart, P.E. (1973) *Pattern Classification and Scene Analysis*. Wiley Interscience, New York, pp. 189–259.
- Ellis, T.H., Dubois, L.H., Kevan, S.D. & Cardillo, M.J. (1985) Time-resolved electron energy loss spectroscopy. *Science*, **230**, 256–261.
- Esbensen, K.H., Geladi, P.L. & Grahn, H.F. (1992) Strategies for multivariate image regression. *Chem. Int. Lab. Syst.* **14**, 357–374.

- Frank, J. (1990) Classification of macromolecular assemblies studied as 'single particles'. *Quart. Rev. Biophys.* **23**, 281–329.
- Frank, J. & van Heel, M. (1982) Correspondence analysis of aligned images of biological particles. *J. Mol. Biol.* **161**, 134–137.
- Garenström, S. (1986) Application of factor analysis to elemental detection limits in sputter depth profiling. *Appl. Surf. Sci.* **26**, 561–574.
- Gatts, C., Duscher, G., Müllejans, H. & Röhle, M. (1995) Analysing line scan EELS data with neural pattern recognition. *Ultramicroscopy* **59**, 229–239.
- Geladi, P. (1992) Some special topics in multivariate image analysis. *Chem. Intell. Lab. Syst.* **14**, 375–390.
- Geladi, P. & Kowalski, B.R. (1986) Partial least-squares regression: a tutorial. *Anal. Chem. Acta* **185**, 1–17.
- Hannequin, P. & Bonnet, N. (1988) Application of multivariate statistical analysis to energetic image series. *Optik* **81**, 6–11.
- Herbin, M., Bonnet, N. & Vautrot, P. (1996) A clustering method based on the estimation of the probability density function and on the skeleton by influence zones. Application to image processing. *Pattern Rec. Lett.* **17**, 1141–1150.
- Jbara, O., Cazaux, J. & Trebbia, P. (1995) Sodium diffusion in glasses during electron irradiation. *J. Appl. Phys.* **78**, 868–875.
- Jeanguillaume, C. & Colliex, C. (1989) Spectrum-image: the next step in EELS digital acquisition and processing. *Ultramicroscopy* **28**, 252–257.
- Jutten, C. & Herault, J. (1991) Blind separation of sources. Part I: An adaptive algorithm based on a neuromimetic architecture. *Signal Proc.* **24**, 1–10.
- Karhunen, J. & Joutsensalo, J. (1995) Generalizations of principal components analysis, optimization problems, and neural networks. *Neural Networks* **8**, 549–562.
- Kisielowski, C., Schwander, P., Baumann, F.H., Seibt, M., Kim, Y. & Ourmazd, A. (1995) An approach to quantitative high-resolution transmission electron microscopy of crystalline materials. *Ultramicroscopy* **58**, 131–155.
- Kohonen, T. (1984) *Self-Organization and Associative Memory*. Springer, Berlin, pp. 119–157.
- Kramer, M.A. (1991) Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* **37**, 233–243.
- Krishnapuram, R. & Freg, C. (1992) Fitting an unknown number of lines and planes to image data through compatible cluster merging. *Pattern Rec.* **25**, 385–400.
- Kruskal, J.B. (1964) Multidimensional scaling by optimizing goodness-of-fit to a non-metric hypothesis. *Psychometrika* **29**, 1–27.
- Lebart, L., Morineau, A. & Warwick, K.M. (1984) *Multivariate Descriptive Statistical Analysis*. Wiley, New York.
- Liang, Y.-Z. & Kvalheim, O.M. (1996) Robust methods for multivariate analysis- a tutorial review. *Chem. Intell. Lab. Syst.* **32**, 1–10.
- Livens, S., Scheunders, P., Van de Wouwer, G., Van Dyck, D., Smets, H., Winkelmann, J. & Bogaerts, W. (1996) A texture analysis approach to corrosion image classification. *Microsc. Microanal. Microstruct.* **7**, 1–10.
- Malinowski, E. & Howery, D. (1980) *Factor Analysis in Chemistry*. Wiley-Interscience, New York.
- Marabini, R. & Carazo, J.M. (1994) Pattern recognition and classification of images of biological macromolecules using artificial neural networks. *Biophys. J.* **66**, 1804–1814.
- Müllejans, H. & Bruley, J. (1995) Electron energy-loss near-edge structure of internal interfaces by spatial difference spectroscopy. *J. Microsc.* **180**, 12–21.
- Ourmazd, A., Baumann, F.H., Bode, M. & Kim, Y. (1990) Quantitative chemical lattice imaging. *Ultramicroscopy* **34**, 237–255.
- Pal, N.R., Bezdek, J.C. & Tsao, E. (1993) Generalised clustering networks and kohonen's self-organising scheme. *IEEE Trans. Neural Nets* **4**, 549–557.
- Prutton, M., Barkshire, I.R., Kenny, P.G., Roberts, R.H. & Wenham, M. (1996) Multi-imaging and multivariate statistics used for 3D characterization of surfaces. *Philos. Trans. R. Soc. London A* **354**, 2683–2695.
- Prutton, M., El Gomati, M.M. & Kenny, P.G. (1990) Scatter diagrams and Hotelling transforms: application to surface analytical microscopy. *J. Electron Spectrosc. Rel. Phenom.* **52**, 197–219.
- Quintana, C. (1991) X-ray analysis of cell nuclei. *J. Electron Microsc. Tech.* **18**, 411–423.
- Quintana, C. & Bonnet, N. (1994a) Multivariate statistical analysis applied to X-ray spectra and X-ray mapping of liver cell nuclei. *Scanning Microsc.* **8**, 563–586.
- Quintana, C. & Bonnet, N. (1994b) Improvements in biological X-ray microanalysis: cryoembedding for specimen preparation and multivariate statistical analysis for data interpretation. *Scanning Microsc. Suppl.* **8**, 83–99.
- Radermacher, M. & Frank, J. (1985) Use of nonlinear mapping in multivariate image analysis of molecule projections. *Ultramicroscopy* **17**, 117–126.
- Rouvière, J.L. & Bonnet, N. (1993) Towards a systematic pattern analysis in high resolution electron microscopy: application to quantitative chemical analysis. *Inst. Phys. Conf. Ser.* **134**, 11–14.
- Sammon, J.W. (1969) A non-linear mapping for data structure analysis. *IEEE Trans. Comput.* **C18**, 401–409.
- Sarkar, M., Calliari, L., Gonzo, L. & Marchetti, F. (1993) Auger depth profiling of silicon dioxide on silicon: a factor analysis study. *Surf. Interface Anal.* **20**, 60–68.
- Simeonova, P., Tsakovski, S., Manjukov, P., Simeonov, V. & Lovchinov, V. (1996) High-temperature superconducting thin films: a multivariate statistical approach. *Mikrochim. Acta* **124**, 151–165.
- Tencé, M., Quartuccio, M. & Colliex, C. (1995) PEELS compositional profiling and mapping at nanometer resolution. *Ultramicroscopy* **58**, 42–54.
- Titchmarsh, J.M. & Dumbill, S. (1996) Multivariate statistical analysis of FEG-STEM EDX spectra. *J. Microsc.* **184**, 195–207.
- Tovey, N.K., Dent, D.L., Corbett, W.M. & Krinsley, D.H. (1992) Processing multispectral scanning electron microscopy images for quantitative microfabric analysis. *Scanning Microsc. Suppl.* **6**, 269–282.
- Trebbia, P. (1996) Maxwell's demon and data analysis. *Philos. Trans. R. Soc. London* **A452**, 2697–2711.
- Trebbia, P. & Bonnet, N. (1990) EELS elemental mapping with unconventional methods. I. Theoretical basis: image analysis with multivariate and entropy concepts. *Ultramicroscopy* **34**, 165–178.

- Trebbia, P. & Mory, C. (1990) EELS elemental mapping with unconventional methods. II. Applications to biological specimens. *Ultramicroscopy*, **34**, 179–203.
- Trebbia, P., Wulveryck, J.M. & Bonnet, N. (1995) Progress in quantitative elemental mapping by X-ray imaging. *Microbeam Anal.*, **4**, 85–102.
- Van Espen, P., Janssens, G., Vanhoolst, W. & Geladi, P. (1992) Imaging and image processing in analytical chemistry. *Analisis*, **20**, 81–90.
- Van Heel, M. (1984) Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy*, **13**, 165–183.
- Van Heel, M. (1989) Classification of very large electron microscopical image data sets. *Optik*, **82**, 114–126.
- Van Heel, M. & Frank, J. (1981) Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, **6**, 187–194.
- Vincent, L. (1990) *Algorithmes morphologiques à base de files d'attente et de lacets. Extension aux graphes*. PhD thesis, Ecole des Mines de Paris.
- Walczak, B. & Massart, D.L. (1995) Robust principal components regression as a detection tool for outliers. *Chem. Intell. Lab. Syst.*, **27**, 41–54.
- Wekemans, B., Janssens, K., Vincze, L., Aerts, A., Adams, F. & Hertogen, J. (1997) Automated segmentation of  $\mu$ -XRF image sets. *X-ray Spectrom.*, **26**, 333–346.
- Wienke, D. & Buydens, L. (1995) Adaptive Resonance Theory based Neural Networks – the 'ART' of real-time pattern recognition in chemical process monitoring? *Trends Anal. Chem.*, **14**, 398–406.
- Zhang, R.D. & Postaire, J.-G. (1994) Convexity dependent morphological transformations for mode detection in cluster analysis. *Pattern Rec.*, **27**, 135–148.