ELSEVIER

Geoderma 76 (1997) 253–262

GEODERMA

An application of discriminant analysis to pattern recognition of selected contaminated soil features in thin sections

Alexandra B. Ribeiro [a,*], Allan A. Nielsen [b]

[a] *Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, P-2825, Monte da Caparica, Portugal*

[b] *Department of Mathematical Modelling, Technical University of Denmark, DK-2800, Lyngby, Denmark*

Received 7 June 1996; accepted 28 November 1996

Abstract

A soil contaminated with copper, chromium and arsenic from an industrial site was characterized with respect to heavy metal (Cu, Cr) and metalloid (As) 'total' content, and for metal speciation by sequential chemical extraction. The contaminated soil contained a negligible proportion of Cu, As and Cr in the 'soluble and exchangeable' phase, these elements being associated primarily with 'amorphous-crystalline Fe-oxides', 'organic matter' and/or 'resistant' phases. The results obtained with sequential extraction were the prerequisite to the attempt to identify the Cr and As distribution in the solid phase. If high concentrations of contaminants are indicated by chemical wet analysis, these contaminants must occur directly in the solid phase. Thin sections of soil aggregates were scanned for Cu, Cr and As using an electron microprobe, and qualitative analysis was made on selected areas. Microphotographs of thin sections of domains (or parts of them), obtained with plane polarized light, and which the electron microprobe showed to be of interest, were saved on a Kodak photo CD. These relevant identified soil features were shown to be iron-aluminium-silicon, which always had a yellowish colour and showed the common qualitative microprobe results: 'present elements' Al, Si, Cr, Fe, As (associated with others). Selected groups of calibrated images (same light conditions and magnification) were submitted to discriminant analysis, in order to find a pattern of recognition in the soil features corresponding to contamination already identified in the thin sections. The authors present a procedure to study the spatial distribution of contaminants in images of contaminated soil features taken from thin sections.

Keywords: contaminated soil; thin section; image processing; Mahalanobis distance

* Corresponding author. Fax: +351 1 2948554. E-mail: abr@mail.fct.unl.pt

PII S0016-7061(96)00102-4

1. Introduction

The use of some image analysis techniques largely drawn from remote sensing has been applied successfully to soil science. However, image analysis of soil thin sections is still in an early stage of development. Until now only relatively simple soil features such as quartz and clay coatings have been identified, and problems such as quantification and shape analysis have been tackled (Terribile and Fitzpatrick, 1995). Tovey et al. (1992a,b) presented several techniques to map automatically the shape and size of domains, as well as to describe soil microfabric rather than give detailed descriptions of the fabric. Although, few workers have used image analysis methods to study the distribution and shape and orientation of the solid particles (Tovey et al., 1992b).

Pollution by heavy metals is of great concern because, whatever their source, metals tend to be immobilized in the top layer of soils, except under extremely acid conditions (McGrath, 1993). There is much need of studying the spatial distribution of heavy metals in soils and their relationship with other soil features. Image analysis techniques seems a promising approach for this purpose, when previously combined with other methods. If high concentrations of contaminants are obtained with chemical wet analysis, these contaminants have to be expressed directly in the solid phase.

The authors present a procedure to study the spatial distribution of what is expected to be contaminated in images of contaminated soil features taken from thin sections. The method is based on the recognition of a pattern of contaminated soil features ('fingerprints'), directly from the solid phase of the contaminated soil which first was submitted to sequential chemical extraction and to electron microprobe analysis. These 'fingerprints', after being determined, can be used directly for recognition of other similar features in the soil, just by applying the same quick procedure to other images taken from the thin sections, without the need to resort to the expensive and time consuming electron microprobe. This study reports results from an application of discriminant analysis to the investigation of pattern recognition of Cr and As contaminated soil features in thin sections of the contaminated soil.

Probably, in future, if we succeed in compiling several patterns of contaminated soil features in a kind of data base acting as a 'fingerprints catalogue', we may reach the point that one can directly identify contaminated soil features in the solid phase, by overlapping their patterns.

2. Materials and methods

The procedure employed is outlined in Fig. 1.

2.1. Soil

Thin sections were prepared from several aggregates of the A_1 horizon (0–10 cm depth) of a contaminated soil from a Portuguese wood preservation site, at Famalicão, in the north of the country (step 2 in Fig. 1). The aggregates were air dried, vacuum

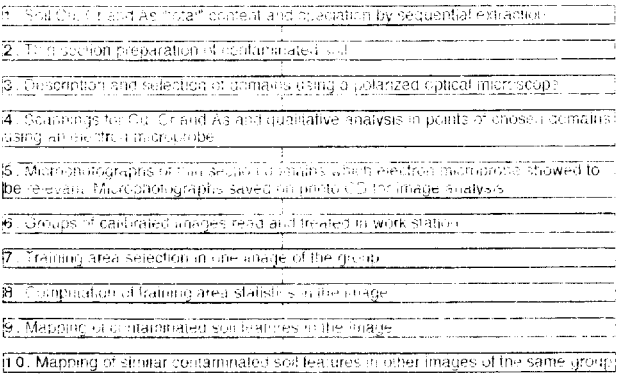

Fig. 1. Flowchart showing the steps involved in the procedure.

checked, impregnated with epoxy resin and thin sections polished for further analysis in an electron microprobe and polarized optical microscope (steps 3 and 4).

The soil is classified as Urbic Anthrosol (FAO-UNESCO, 1990), derived from a Cambisol, and it was sampled from a storage area for wood freshly treated with chromate copper arsenate (CCA) as the wood preservative. This soil is heavily polluted with copper, chromium and arsenic. The 'total' metal/metalloid content, extracted by HNO_3-$HClO_4$-HF, as well as other soil characteristics are presented in Table 1 (step 1 in Fig. 1). A six phase sequential chemical extraction scheme divided the 'total' contaminant content into six different soil fractions (step 1 in Fig. 1): (a) 'soluble and exchangeable', (b) 'Mn-oxides', (c) 'organic matter', (d) 'amorphous Fe-oxides', (e) 'crystalline Fe-oxides', (f) 'resistant' (Table 1, according to the description in Ribeiro et al., 1997).

2.2. Equipment

The study of the thin sections was conducted using a polarized optical microscope Orthoplan-Pol (step 3 in Fig. 1). After selection of the study areas (domains) and their description, the thin sections were vacuum-coated with carbon. Using a wavelength dispersive X-ray system (WDS) attached to the Jeol JCXA Superprobe 733, scannings for Cu, Cr and As were performed, as well as qualitative analysis (QLA) in points of the previously chosen domains (accelerating potential = 25 kV; electron probe beam diameter = 5 μm) (step 4 in Fig. 1). The QLA software used was also able to search for Cu, Cr and As. The detection limit of the microprobe was estimated at 0.05%.

Plane polarized light microphotographs were taken of the thin sections of the domains (or parts of them) which electron microprobe showed relevant (step 5 in Fig. 1). For this purpose, different types of illumination and magnification were used with the microscope Orthoplan-Pol, a systematized Leitz camera, and a Microsix-L photometer. The microphotographs were saved on a Kodak photo CD (step 5).

Working on images with size 512 by 768 corresponding to a pixel size of 0.42 by 0.42 μm, the highest resolution rate of the photo CD, previously chosen groups of

Table 1
Characteristics of the soil used in the study (adapted from Ribeiro et al., 1997)
Depth = 0–10 cm.
Texture = loamy sand (coarse sand = 49.4%, fine sand = 24.4%, silt = 21.7%, clay = 4.5%).
$pH_{(H2O)}$ = 6.99, $pH_{(KCl)}$ = 6.10; organic matter = 70.9 g kg^{-1}.
Exchangeable bases ($cmol_c$ kg^{-1}): Ca = 10.0; Mg = 1.0; K = 0.4; Na = 0.2.
Cation exchange capacity = 12.8 $cmol_c$ kg^{-1}; base saturation = 92%.

			Copper mg Cu kg^{-1} ± SD	Arsenic mg As kg^{-1} ± SD	Chromium mg Cr kg^{-1} ± SD
	'Total' element by HNO_3-HClO_4-HF extraction		346 ± 1	966 ± 23	855 ± 21
	Element distribution by the 6 phase sequential extraction	Operational condition			
E	(a): 0.5 M $Ca(NO_3)_2$	16 h contact	1.3 ± 0.5	1.60 ± 1.60	0.08 ± 0.18
X	(b): 0.1 M $NH_2OH \cdot HCl$ acidified to pH 2 with 0.1 M HNO_3	30 min continuous agitation	51 ± 4	76 ± 22	33 ± 2
T					
R	(c): 0.1 M $K_4P_2O_7$	24 h contact	85 ± 2	176 ± 28	95 ± 4
A	(d): 0.25 M $NH_2OH \cdot HCl$ + 0.25 M HCl	40 min in a water-bath at 50°C	88 ± 8	239 ± 48	145 ± 15
C					
T	(e): 0.25 M $NH_2OH \cdot HCl$ + 0.25 M HCl	40 min in a water-bath at 90°C	62 ± 5	215 ± 9	159 ± 15
A					
N	(f): 4 M HNO_3	16 h in a water-bath at 80°C	30 ± 4	101 ± 30	207 ± 21
T					
	Recovery ratio (quantitative mass balance) of the sequential extraction		92% Cu	84% As	75% Cr

The extractant sequence from (a) to (f) represents increasing vigour of extraction.

calibrated images (same light conditions and magnification) were read and treated in a UNIX work station, using software written by the Section for Image Analysis of the Department of Mathematical Modelling, Technical University of Denmark (step 6 in Fig. 1). The images were submitted to texture analysis, principal component analysis, and discriminant analysis, in order to find a pattern of recognition in the soil features to be associated with contamination previously identified in the thin sections: a simple texture analysis was performed to see if the local variance in a 3 by 3 window would give a good discrimination between soil components, as well as a 'one class classification' involving application of the Mahalanobis distance to a region of interest ('training area') was used. Theories and algorithms of the employed techniques are considered to be outside the scope of this paper. The reader is referred to literature for details (e.g., Anderson, 1958; Cooley and Lohnes, 1971; Davis, 1973; Kendall and Stuart, 1977; Clausen and Harpøth, 1980).

3. Results and discussion

The contaminated soil studied shows a negligible proportion of Cu, As and Cr in the 'soluble and exchangeable' phase (a) (Table 1 and step 1 in Fig. 1). Considering (b) 'Mn-oxides', (c) 'organic matter', (d) 'amorphous Fe-oxides', (e) 'crystalline Fe-oxides', (f) 'resistant' (the operationally defined phases of the applied sequential chemical extraction scheme) the contaminants show the following orders of association (Table 1):

copper (d) > (c) > (e) > (b) > (f),
arsenic (d) > (e) > (c) > (f) > (b) and
chromium (f) > (e) > (d) > (c) > (b).

The reasoning that the high concentrations of contaminants obtained with the wet chemical analysis would be found in the solid phase proved to be fruitful. Five selected thin sections examined and submitted to scanning for Cu, As and Cr in the electron microprobe revealed the presence of As and Cr in some of the soil features, but never the presence of Cu (step 4 in Fig. 1). The qualitative results obtained with the electron microprobe, in some points of previously chosen domains, also show the 'presence' of

Table 2
Some characteristics of the soil features microphotographs chosen for image analysis

Image No.	Magnification	Domain No. (Point No.) submitted to electron microprobe beam	Electron microprobe qualitative results as 'present elements'
7	250×	2 (6)	Al, Si, P, Ca, Cr, Fe, As
13		3 (1)	Al, Si, Cl, Cr, Fe, As
21	400×	1 (6)	Mg, Al, Si, K, Ca, Cr, Fe, As
23		1 (4)	F, Mg, Al, Si, Cl, K, Ca, Cr, Fe, As
25		4 (1)	Mg, Al, Si, K, Ca, Cr, Mn, Fe, As

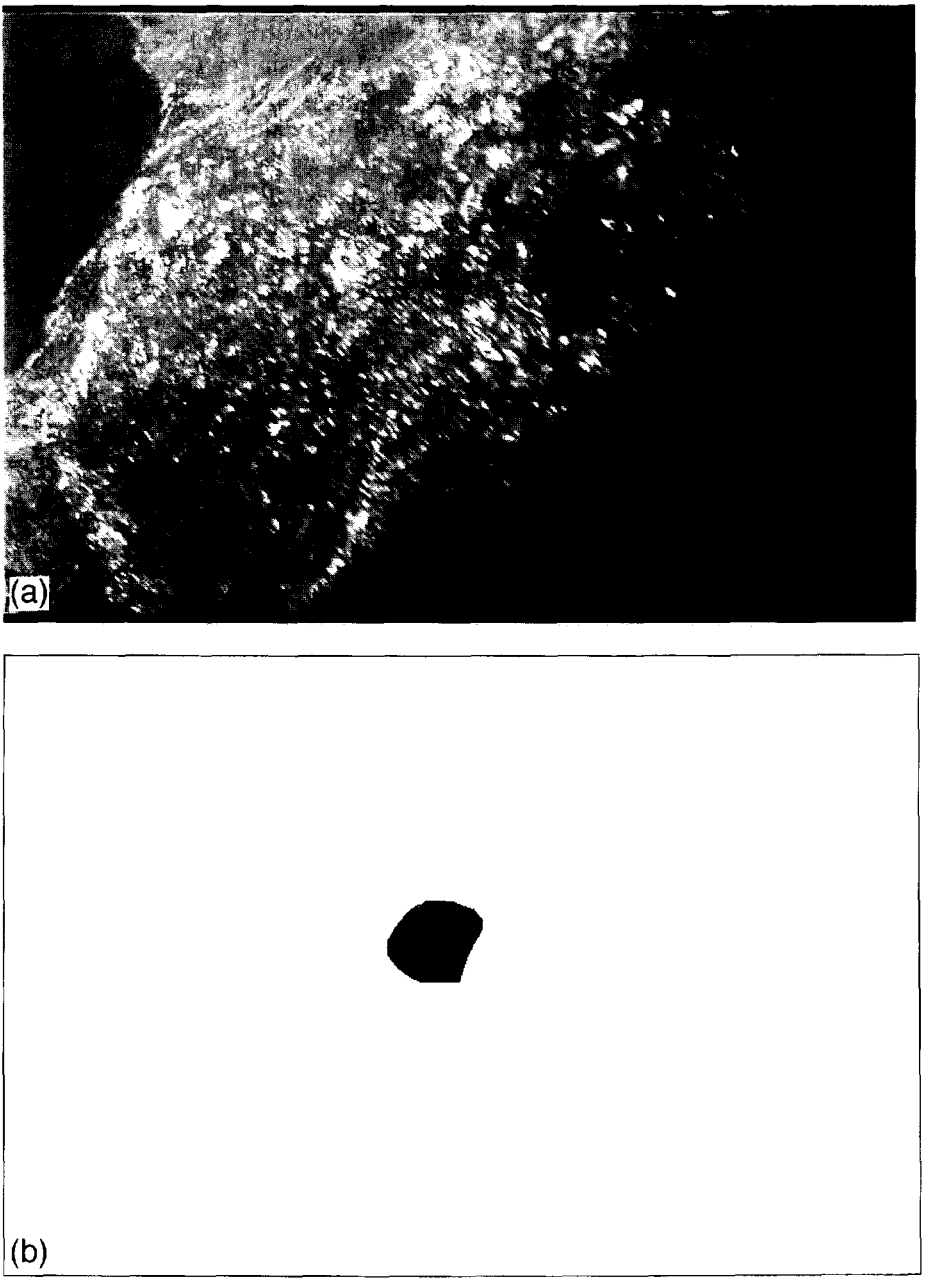

Fig. 2. (a) Red channel of image No. 7 (to locate the chosen 'training area'); 1 cm corresponds to 25 ca. μm. (b) 'Training area' chosen in image No. 7; (c) Mahalanobis distance from each pixel to the center of the distribution estimated in the 'training area' (the darker the tone the nearer to the center of the distribution of the observations in the 'training area'; the lighter the tone the more unlikely it is that the pixel belongs to the same distribution of the 'training area').

Fig. 2 (continued).

As and Cr in some soil features. Copper rarely appears and only as 'probably present'. The difficulties in detecting Cu may be due to the even dispersion of this element in the matrix, as indicated by the 'speciation' obtained by the chemical fractionation, where Cu is much less present in phases (c), (d), (e) and (f) than Cr and As (Table 1). Also, in the case of Cu, it has to be remembered that this element is easily complexed by the organic matter, and that the organic sections are quickly destroyed by the electron beam conditions used.

The identified soil features were shown to be iron-aluminium-silicon. They always have a yellowish colour and common qualitative electron microprobe results 'present elements': Al, Si, Cr, Fe, As (sometimes associated with P, Mg, Ca, Cl, K, Mn among others).

From the microphotographs taken of these soil features (step 5 in Fig. 1), two groups were chosen to be submitted to image analysis: a group of two (No. 7 and 13), and a group of three (No. 21, 23 and 25). All these images were taken with plane polarized light and had yellow as a predominant colour; Table 2 summarizes details obtained for them.

In order to find a pattern of recognition in the soil features that were previously shown to be contaminated in the thin sections, a simple texture analysis and a principal component analysis were performed with the referred groups of images. As good discrimination between the soil components was not obtained, the results from these analysis were not encouraging enough to continue with these techniques.

An application of discriminant analysis was then performed, in order to look for a

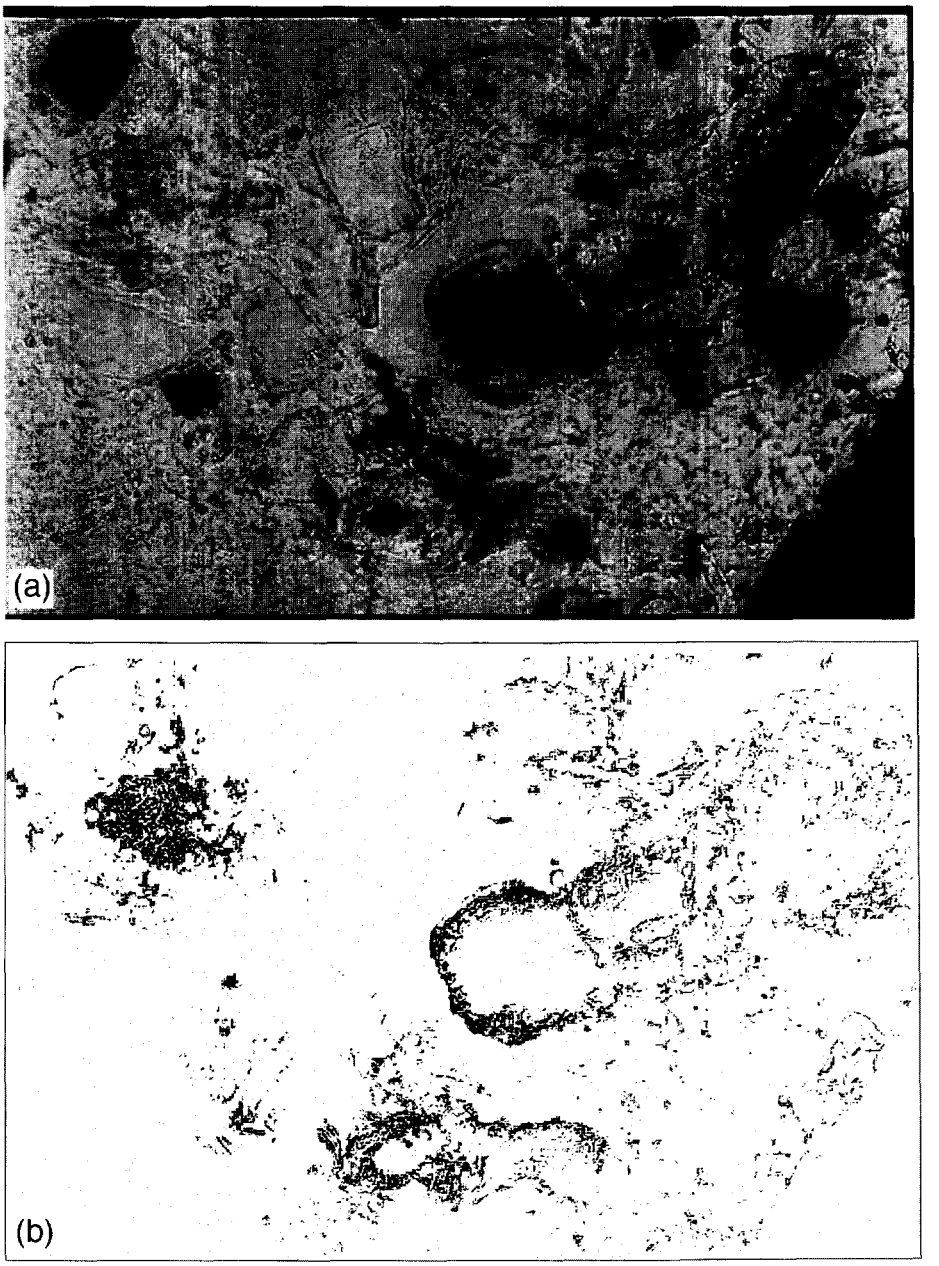

Fig. 3. (a) Red channel of image No. 13; 1 cm corresponds to ca. 25 μm. (b) Mahalanobis distance from each pixel in image No. 13 to the center of the distribution estimated in the 'training area' in image No. 7. The scale is the same as in Fig. 2c.

pattern of recognition in the selected contaminated soil features. A 'one class classification' involving application of the Mahalanobis distance (Pillai, 1985) to a region of interest ('training area') was used (steps 7, 8 and 9 in Fig. 1).

Fig. 2a–c and Fig. 3a,b illustrate the procedure used with images No. 7 and 13. The starting (and very critical) point for this analysis involves the choice of a 'training area' (step 7 in Fig. 1). Fig. 2a shows the red channel of image No. 7, to allow positioning of the chosen 'training area'. Fig. 2b shows a chosen 'training area' of image No. 7, for which the Mahalanobis distance was calculated (statistics in the training area are shown in Table 3, corresponding to step 8 in Fig. 1). Fig. 2c (step 9 in Fig. 1) shows the Mahalanobis distance from each pixel to the center of the distribution estimated in the 'training area' (Table 3): the darker the tone the nearer to the center of the distribution of the observations in the 'training area'; the lighter the tone the more unlikely it is that the pixel belongs to the same distribution [Mahalanobis distance $> \chi^2_{0.99}$ (3) is saturated in $\chi^2_{0.99}$ (3)].

A representation of image No. 13 was performed based on the statistical results for image No. 7: what was black in image No. 13 was statistically similar to what was black in image No. 7, although on a relative scale. After 'stretching' the scales used or, in other words, after calculating the minimum Mahalanobis distance differences in images 7 and 13 (for image 7 = 0.01684 and for image 13 = 0.02640), new images for No. 7 and 13 were produced, with the blacks corresponding on the same scale. Fig. 3a shows the red channel of image No. 13. Fig. 3b shows the Mahalanobis distance from each pixel in image No. 13 to the center of the distribution estimated in the 'training area' (Table 3) in image No. 7 (step 10 in Fig. 1). The scale is the same as in Fig. 2c.

The results (Fig. 2c and Fig. 3b) obtained allow an immediate visualization of what is expected to be contaminated in the images submitted to analysis. In a way these are 'fingerprints' directly identifiable from the solid phase of the studied contaminated soil. These 'fingerprints' are known to be related to Cr and As contaminated soil features, which were also shown to be iron-aluminium silicon. This pattern of contaminated soil features can now be used directly for recognition of other similar features in this soil, just by applying the same quick procedure to other images taken from the thin sections, without the need to resort to the expensive and time consuming electron microprobe.

A sensitive point of the proposed approach consists in the imperative need of calibrated images (e.g., same light conditions and magnification), in order to make reliable comparisons between the statistics for images (e.g., using statistical distance).

Table 3
Statistics calculated in training area chosen in image No. 7

Statistical parameters	Channel		
	Red	Green	Blue
Mean	211.256	122.573	45.119
Dispersion	288.778	154.258	−49.896
(Covariance matrix)	154.258	144.984	−48.259
	−49.896	−48.259	68.892

$\chi^2_{0.99}$ (3) = 11.342

A drawback at present in the proposed approach involves the choice of the 'training area'. This decision is user based and is very critical for the analysis. A promising improvement remains in the definition of the 'training area' by graphical computation: the user will just have to point to a starting point with the cursor (give the coordinates of a starting point) which is thought to be relevant for the pattern recognition and, according to its spatial and spectral contrast, the computer will progressively advance and will define itself the area for further study (calculation of the Mahalanobis distance). This software is expected to be written by the Section for Image Analysis of the Department of Mathematical Modelling, Technical University of Denmark.

Acknowledgements

A.B. Ribeiro would like to express her gratitude to Dr. J.M. Vieira e Silva for his valuable help in performing the impregnations, describing the thin sections and taking the photographs, to Mr. R. Raposo for preparing the thin sections, to Mr. O. Chaveiro for the microprobe analysis, to Prof. A. Réfega and Prof. A. Villumsen for their encouragement and comments, and Prof. G. Bech-Nielsen, Prof. J. C. Reeve and Prof. D. Bridgwater for careful reading of the manuscript.

References

Anderson, T.W., 1958. Introduction to Multivariate Statistical Analysis. Wiley, New York, 374 pp.

Clausen, F.L. and Harpøth, O., 1980. The use of discriminant analysis techniques to reveal the structural patterns in data from panned heavy-mineral concentrates-Central East Greenland. Working Report 3. Department of Mineral Industry, Technical University of Denmark, Lyngby, Denmark, 60 pp.

Cooley, W.W. and Lohnes, P.R., 1971. Multivariate Data Analysis. Wiley, New York, 325 pp.

Davis, J.C., 1973. Statistics and Data Analysis in Geology. Wiley, New York, 550 pp.

FAO-UNESCO, 1990. Soil Map of the World, revised legend. FAO, Rome.

Kendall, M. and Stuart, A., 1977. The Advanced Theory of Statistics, Vol. 3. C. Griffin and Co. Ltd., London, pp. 176–180.

McGrath, S.P., 1993. Soil quality in relation to agricultural uses. In: H.J.P. Eijsackers and T. Hamers (Editors), Integrated Soil and Sediment Research: A Basis for Proper Protection. Kluwer, Dordrecht, pp. 187–200.

Pillai, K.C.S., 1985. Mahalanobis D^2. In: Encyclopedia of Statistical Sciences, Vol. 5. Wiley-Interscience, Wiley, New York, pp. 176–181.

Ribeiro, A., Villumsen, A., Jensen, B., Réfega, A. and Vieira e Silva, J.M., 1997. Redistribution of Cu, Cr and As in a Portuguese polluted soil during electrokinetic remediation. Proc. of Contaminated Soils, Third Int. Conf. on the Biogeochemistry of Trace Elements, Paris, France, 15–19 May 1995, INRA-Versailles (accepted for publication).

Terribile, F. and Fitzpatrick, E.A., 1995. The application of some image-analysis techniques to recognition of soil micromorphological features. Eur. J. Soil Sci., 46(1): 29–45.

Tovey, N.K., Smart, P., Hounslow, M.W. and Leng, X.L., 1992a. Automatic orientation mapping of some types of soil fabric. Geoderma, 53(3/4): 179–200.

Tovey, N.K., Krinsley, D.H., Dent, D.L. and Corbett, W.M., 1992b. Techniques to quantitatively study the microfabric of soils. Geoderma, 53(3/4): 217–235.