Mathematical Geology, Vol. 32, No. 3, 2000

Supervised Mineral Classification with Semiautomatic Training and Validation Set Generation in Scanning Electron Microscope Energy Dispersive Spectroscopy Images of Thin Sections[1]

Harald Flesche,[2] Allan Aasbjerg Nielsen,[3] and Rasmus Larsen[3]

This paper addresses the problem of classifying minerals common in siliciclastic and carbonate rocks. Twelve chemical elements are mapped from thin sections by energy dispersive spectroscopy in a scanning electron microscope (SEM). Extensions to traditional multivariate statistical methods are applied to perform the classification. First, training and validation sets are grown from one or a few seed points by a method that ensures spatial and spectral closeness of observations. Spectral closeness is obtained by excluding observations that have high Mahalanobis distances to the training class mean. Spatial closeness is obtained by requesting connectivity. Second, class consistency is controlled by forcing each class into 5–10 subclasses and checking the separability of these subclasses by means of canonical discriminant analysis. Third, class separability is checked by means of the Jeffreys–Matusita distance and the posterior probability of a class mean being classified as another class. Fourth, the actual classification is carried out based on four supervised classifiers all assuming multinormal distributions: simple quadratic, a contextual quadratic, and two hierarchical quadratic classifiers. Overall weighted misclassification rates for all quadratic classifiers are very low for both the training (0.25–0.33%) and validation sets (0.65–1.13%). Finally, the number of rejected observations in routine runs is checked to control the performance of the SEM image acquisition and the classification. Although the contextual classifier performs marginally best on the validation set, the simple quadratic classifier is chosen in routine classifications because of the lower processing time required. The method is presently used as a routine petrographical analysis method at Norsk Hydro Research Centre. The data can be approximated by a Poisson distribution. Accordingly, the square root of the data has constant variance and a linear classifier can be used. Near orthogonal input data, enable the use of a minimum distance classifier. Results from both linear and quadratic minimum distance classifications are described briefly.

KEY WORDS: x-ray mapping, seed growing, feature selection, contextual and hierarchical classification.

[1]Received 18 August 1998; accepted 7 April 1999.
[2]Norsk Hydro Research Centre, N-5020 Bergen, Norway. e-mail harald.flesche@hydro.com
[3]Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

INTRODUCTION

Mineral classification and quantification is traditionally done using point counting of thin sections or by x-ray diffraction. The first of these methods is very time-consuming and requires a trained petrographer; the latter does not give any spatial information about the samples being analyzed. Point counting also has an element of subjectivity in that a more skilled petrographer is better at recognising rare minerals, separating cement from detrital grains, etc.

A third method is to do x-ray mapping or energy dispersive spectroscopy (EDS) in a scanning electron microscope (SEM). Here, an x-ray spectrum is acquired for each pixel. By means of this spectrum, the mineral present in each pixel can be identified using a manual or an automatic classification method. Spatial information about the mineral composition can thereby be obtained by an objective and reproducible method. Earlier work in this field (Minnis, 1984; Tovey and Krinsley, 1991; Clelland and Fens, 1991) used classification methods that range from lookup table to maximum likelihood classification. Earlier, long image acquisition times made the use of EDS images for mineral classification difficult. New equipment enables acquisition of a 256×256 pixels image with 12 elements mapped in 36 min. Accelerating voltage, current, dwell time, and sensor parameters are adjusted for an acceptable trade-off between data noise level and image acquisition time. The chosen configuration results in 40% dead time in the EDS detector. Typical pixel size is $2.4\ \mu\text{m} \times 2.4\ \mu\text{m}$.

This paper addresses methods for classification of EDS images. The methods applied all presume that the data (within each mineral class) are described by a multivariate normal distribution. As the data are basically photon counts from the decay of excited atoms the Poisson distribution applies. The normal distribution (ideally with the mean value equal to the variance for each variable in each class) is used as an approximation to the Poisson distribution.

We put emphasis on building a model for the classification and validating the model. The resulting classification model is used in a routine laboratory method at Norsk Hydro Research Centre for quantification of mineral composition in sedimentary rocks.

METHODS

This section describes the methods used for training and validation set generation, classification, and validation of the model.

Training Set Generation

Good supervised classification is contingent on good training sets. Stipulating a multinormal distribution for the data, statistically sound training sets are not

Figure 1. Validation area for chlorite 2. An example on training and validation areas grown with the seed algorithm.

necessarily obtained when handdrawn by a human operator. One reason for this is the human inability to obtain an overview of multidimensional spaces. Another problem with training sets drawn by humans is inconsistency. Training sets need to be extracted in a consistent way across time and classes irrespective of operator and shape of image structures. Therefore, we propose a new semiautomatic algorithm for generation of a set of training classes from a series of seed points. For each class, the operator needs only supply one or a few observations. From these points, training classes are grown in a fashion that ensures spatial and spectral closeness.

Spatial closeness is ensured by demanding that all the pixels in one training class be connected with the seed point. This is a very useful condition, because most relevant phenomena appear as connected objects. The connectivity may be defined in terms of first-order or second-order neighbors etc., which allows for small holes in the training sets. Here we apply second-order neighbors. This is useful for classes that occur as clusters of smaller objects in the image, and also in the case of classes that occur as thin strings or layers, as illustrated in Figure 1.

Spectral closeness is achieved by restricting the distance to the current mean value of the class while growing the training set. Here two distance measures are considered. One is the Euclidean spectral distance,

$$D_E^2 = \left(\mathbf{x} - \boldsymbol{\mu}_i^*\right)^T \left(\mathbf{x} - \boldsymbol{\mu}_i^*\right)$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ is the value observed in a pixel, and $\boldsymbol{\mu}_i^*$ is the current estimate of the class mean. The application of Euclidean distance to seed-growing is suggested by ERDAS (1990). The other distance measure used is the Mahalanobis distance,

$$D_M^2 = \left(\mathbf{x} - \boldsymbol{\mu}_i^*\right)^T \Sigma_i^{*-1} \left(\mathbf{x} - \boldsymbol{\mu}_i^*\right)$$

$\boldsymbol{\Sigma}_i^*$ is the current estimate of the class dispersion matrix.

For the Euclidean distance, an upper limit for the distance should be supplied by the user. The Mahalanobis distance is χ^2-distributed with n degrees of freedom. This enables us to choose a systematic threshold, defined by, e.g., the 0.99 quantile, which is a major advantage over use of the Euclidean distance.

If the seed growing begins with a single pixel, we cannot get an estimate of the dispersion matrix. Therefore, we first grow the seed point to an initial training set using the Euclidean distance method with a preset maximum distance, spectrally and spatially. From the pixels thus included, an estimate of the dispersion matrix is obtained. This estimate may first be used to exclude any outliers in the current set, and second, used to grow the initial training set further using the Mahalanobis distance method.

The application of this method gives us training data that are in good correspondence with normal distributions. Validation data are generated in the same fashion.

Consistency Check of Training Sets

Training and validation data should be checked for consistency to make sure that the multivariate data in each assumed class make up one class only. Here, a method based on a partitioning of the training and validation data for each class into five subclasses by means of an unsupervised clustering algorithm is used. First, observations called cluster seeds are selected as a first guess of the subclass means. Second, clusters are formed by assigning observations to the nearest seed as measured by Euclidean distance. After all observations are assigned, new cluster means are calculated. This step is repeated until changes in cluster means become zero (or small). This clustering is followed by a canonical discriminant analysis, which combines the original variables into new orthogonal variables called canonical discriminant functions (CDFs), which are the best possible linear discriminators between the subclasses into which the training and validation data have been clustered. If a scatter plot of the first two CDFs shows no outliers and no sign of grouping, the training and validation data are considered as being consistent.

Canonical Discriminant Analysis

Consider k groups with $m_1, \ldots, m_k$ multivariate (n-dimensional) observations represented by stochastic variables $\{\mathbf{X}_{ij}\}$, where i is the group index and j is the observation number. The group expectations are denoted $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$. Without loss of generality the overall expectation is assumed to be $\mathbf{0}$. As in a one-way analysis of variance the total sum of squares matrix $\mathbf{T}$ is split up into a sum of the "among group" matrix $\mathbf{A}$, and the "within group" matrix $\mathbf{W}$,

$$\mathbf{T} = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \mathbf{X}_{ij} \mathbf{X}_{ij}^T = \sum_{i=1}^{k} m_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T + \sum_{i=1}^{k} \sum_{j=1}^{m_i} (\mathbf{X}_{ij} - \boldsymbol{\mu}_i)(\mathbf{X}_{ij} - \boldsymbol{\mu}_i)^T,$$

i.e.,

$$\mathbf{T} = \mathbf{A} + \mathbf{W}$$

or in words: the total variation can be written as a sum of the variation of the group expectations around the overall expectation and the variation around the group expectations.

We are looking for new variables $Y = \mathbf{e}^T\mathbf{X}$ that maximise the ratio between variation among groups and variation within groups; the latter can be considered as the natural level of variance of the variables $\mathbf{X}$. The idea of maximizing this ratio is due to Fisher (1936). This ratio equals the Rayleigh coefficient $\mathbf{e}^T\mathbf{Ae}/\mathbf{e}^T\mathbf{We}$, i.e., the transformation is defined by the conjugate eigenvectors $\mathbf{e}_l$ of $\mathbf{A}$ with respect to $\mathbf{W}$

$$\mathbf{Ae}_l = \lambda_l \mathbf{We}_l$$

We define the canonical correlation coefficients R_l by their squares $R_l^2 = \mathbf{e}_l^T\mathbf{Ae}_l/\mathbf{e}_l^T\mathbf{Te}_l$. This gives the relation $R_l^2 = \lambda_l/(\lambda_l+1)$. The new variables $Y_l = \mathbf{e}_l^T\mathbf{X}$ are the CDFs. The first CDF defined by $\mathbf{e}_1$ is the linear transformation of the original variables that gives the best discrimination between the k groups. A higher order CDF is the linear combination of the original variables that gives the best discrimination between the k groups subject to the constraint that it is orthogonal (with respect to $\mathbf{A}$ and $\mathbf{W}$) to the lower order CDFs. Note, that the number of CDFs is given by rank considerations for $\mathbf{A}$ and $\mathbf{W}$. If $\mathbf{A}$ and $\mathbf{W}$ have full rank, this number equals $\min(k-1, n)$.

Scatterplots of the first few CDFs give a good visual impression of the separability of the groups.

Classification

We apply a series of classifiers. All of these are parametric supervised classifiers. For general references on classifiers, see Swain and Davis (1978) or Richards (1993). In accordance with the training set generation method, all classifiers presume that the data may be described by a multivariate Gaussian distribution. The Gaussian distribution is by far the most commonly adopted density model for continuous image features. This is partly supported by the central limit theorem, and partly due to the resulting simple analytical expressions obtained in various kinds of analyses.

Suppose that a pixel is an observation from one of the populations $\pi_1, \pi_2, \ldots, \pi_k$. The classification of the observation depends on its feature vector, which we will denote $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$. The Gaussian class conditional density function of class π_i is

$$f_i(\mathbf{x}) = P(\mathbf{X} = \mathbf{x} \mid C = \pi_i) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sqrt{\det \Sigma_i}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right)$$

for $i = 1, \ldots, k$, where C is the class variable. Furthermore, let us assume

knowledge of the prior distribution of the classes, i.e., the prior probabilities, $P(C = \pi_i) = p_i, i = 1, \ldots, k$. This distribution determines the probability with which an arbitrary feature vector from a particular class has been generated. The posterior probability for the class variable becomes

$$k(\pi_i \mid \mathbf{x}) = \frac{p_i f_i(\mathbf{x})}{\sum_{i=1}^{k} p_i f_i(\mathbf{x})}$$

The Bayes solution to a classification problem chooses the action that minimizes the posterior expected loss. In the case of equal losses, i.e., the same loss is assigned to all misclassifications, the Bayes solution consists of choosing the class that has the highest posterior probability. For computational efficiency, we may simplify the decision rule by taking the logarithm of the posterior probability and excluding terms that are common to all classes. The decision function then becomes

$$S_i = \log p_i - \frac{1}{2}\log(\det \Sigma_i) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)$$

If this function is maximum for $i = v$ we assign the pixel to population π_v.

In the case of class-dependent dispersion matrices, the logarithm of the posterior probability is a quadratic function of $\mathbf{x}$. This is therefore denoted quadratic classification. In the case of equal dispersion matrices, $\Sigma_i = \Sigma$, we only need to consider a linear function of $\mathbf{x}$ because some terms in the logarithm of the posterior probability are common to all classes and can be neglected.

In minimum distance classification, off-diagonal elements in Σ are set to zero (i.e., the input variables are noncorrelated), which makes inversion and determinant calculation easy and fast.

Contextual Classifiers

When applying classical classification schemes in image analysis, the spatial structure of the data is neglected. This is not satisfactory because more information obviously can be drawn from the spatial arrangement of pixels; e.g., neighboring pixels tend to be of the same class. We will refer to this type of information as contextual information.

Contextual information can be taken into account in a number of ways when performing classification. One way is to include derived features that hold information of the neighborhood of a given pixel, i.e., contextual features. Another way to take the spatial nature into account is in the classification itself. Here we will consider a technique for including spatial information in the classifier that was first published by Owen (1984), Hjort (1985), and Hjort and Mohn (1984).

An alternative algorithm has been proposed by Welch and Salter (1971) and by Haslett (1985).

We will denote the feature vector of the neighboring pixels $\mathbf{X}_N$, $\mathbf{X}_S$, $\mathbf{X}_E$, and $\mathbf{X}_W$ for the north, south, east, and west pixel, respectively. The augmented feature vector consisting of the feature vectors for the neighbors of a pixel will be denoted $\mathbf{D}_\Delta = (\mathbf{X}_N^T, \mathbf{X}_S^T, \mathbf{X}_E^T, \mathbf{X}_W^T)^T$. The augmented feature vector consisting of the feature vector of a pixel and those of its neighbors will be denoted $\mathbf{D} = (\mathbf{X}^T, \mathbf{D}_\Delta^T)^T$.

The posterior distribution for the class variable becomes

$$k(\pi_v \mid \mathbf{d}) = P(C = \pi_v \mid \mathbf{D} = \mathbf{d}) = \frac{P(C = \pi_v)P(\mathbf{D} = \mathbf{d} \mid C = \pi_v)}{\sum_{i=1}^{k} P(C = \pi_i)P(\mathbf{D} = \mathbf{d} \mid C = \pi_i)}$$
$$= \frac{\sum_{a,b,c,d} p_v P(\mathbf{D} = \mathbf{d} \mid \mathbf{C} = (\pi_v, \pi_a, \pi_b, \pi_c, \pi_d))g(\pi_a, \pi_b, \pi_c, \pi_d \mid \pi_v)}{h(\mathbf{d})}$$

where $h(\mathbf{d})$ is the unconditional density of the augmented feature vector, (a, b, c, d) is one of the possible k^4 configurations of the class variables of the neighboring pixels, $\mathbf{C}$ is the class configuration corresponding to the augmented feature vector $\mathbf{D}$, and $g(\pi_a, \pi_b, \pi_c, \pi_d \mid \pi_v)$ is the probability of the configuration of the class variables of the neighboring pixels given that the centre pixel has class π_v. In the numerator the prior probability p_v is assumed known as before. Also as before the class conditional density, $P(\cdots \mid \cdots)$, is assumed multidimensional Gaussian. In order to reduce the huge number of terms in the summation, $g(\cdots)$ is nonzero only for very few configurations of the class variable. The nonzero configurations are shown in Figure 2. In this figure, pixels with the same greytone are the same class. Contextual information enters into the model in two ways, first in the spatial dependence of the feature vectors (specification of the conditional distribution of the augmented feature vector), and second in the specification of prior distribution of the class configurations g.

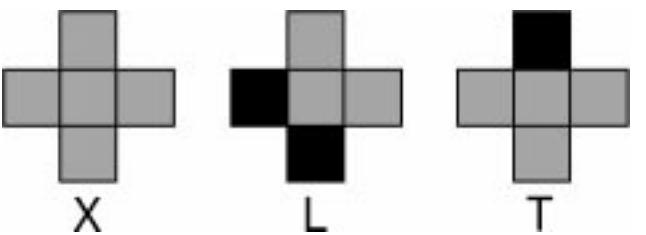

Figure 2. Nonzero probability configuration of pixels for contextual classification.

Hierarchical Classifiers

When applying a classification scheme it is desired to use as few features as possible. This will minimize the computational load as well as optimize the estimation accuracy (Hughes, 1968). The problem of choosing few variables with sufficient discriminating power may basically be addressed in two ways. In many cases, we see high correlations between original features, which allows us to concentrate the relevant information in a lower dimensional space. Traditionally linear transformations such as principal components (Anderson, 1984) or maximum autocorrelation factors (Switzer and Green, 1984; Green and others, 1988; Nielsen, 1994; Nielsen, Conradsen, and Simpson, 1998) are used. Another possibility is to perform a selection among the original features. This may be done using (stepwise) selection schemes based on different measures of class separability, e.g., the Mahalanobis distance, the divergence or the Jeffreys–Matusita (JM) distance, JM_{ij}, (or equivalently, the Bhattacharyya distance α_{ij}). The JM measure of separability between classes i and j is given by

$$JM_{ij}^2 = \int (\sqrt{f_i(\mathbf{x})} - \sqrt{f_j(\mathbf{x})})^2 d\mathbf{x} = 2(1 - e^{-\alpha_{ij}}) = 2\left(1 - \int \sqrt{f_i(\mathbf{x})\, f_j(\mathbf{x})}\, d\mathbf{x}\right)$$

where f_i and f_j are the density functions for classes i and j, respectively (see Matusita, 1966; Ersbøll, 1989).

When applying feature selection schemes to classification problems the feature selection will normally be based on a weighted sum of the separability among all class pairs. This approach may be regarded as suboptimal in the sense that separation of certain class pairs may be given little weight. Also, in general more features are required to separate more classes. We use the following definition (in which we assume that the prior probabilities of all classes are equal),

$$JM_{\mathrm{AVE}} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} JM_{ij},$$

whereas in the literature, average JM value is calculated from the entire JM matrix, including the diagonal elements which are zero. Therefore, the maximum value of the average JM is a function of the number of classes. With the definition above, the maximum value for the average JM value is $\sqrt{2}$.

In order to reduce the number of features needed, we may consider hierarchical classifiers. An overview of hierarchical classifiers is given by Safarian and Landgrebe (1991). A particular implementation is described by Jia and Richards (1996). However, the technique described by Jia and Richards (1996) suffers from a potential drawback, namely that the classes are considered in a given (preselected) order. The selection of the ordering may influence the classification result. The classification scheme is sketched in Figure 3. In the first layer, all pixels pass

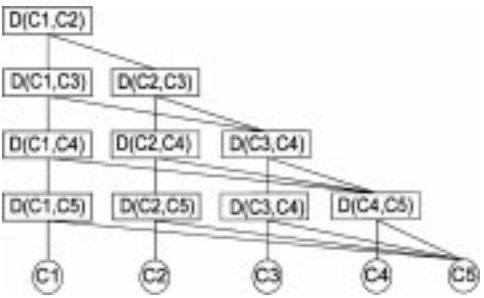

Figure 3. Schematic chart for the hierarchical classification. D(Ci,Cj) is the classification rule used for discriminating between classes i and j.

through the classification rule that discriminates between classes 1 and 2. All pixels that have been assigned class label 1 are then passed to the classification rule that discriminates between classes 1 and 3, whereas the pixels that received the label 2 in the first layer are passed to a classification rule that discriminates between classes 2 and 3. In this way, we progress down through the tree until a final classification has been reached in the last layer. Note that at each node a feature selection for that particular class pair is performed.

We propose an extension to this algorithm that eliminates the problem of using a preselected ordering of the classes by considering all pairs of classes for all pixels. This will greatly increase the computational load, because all pixels have to pass through all classification rules. Furthermore, an ambiguity is introduced: what should be done if for instance class 1 has higher posterior probability than class 2 (using one feature set), class 2 has higher posterior probability than class 3 (using another feature set), and class 3 has higher posterior probability than class 1 (using yet another feature set). We solve this problem by introducing a majority-voting scheme. In a k class classification problem a single pixel will be considered by $k(k-1)/2$ classification rules corresponding to all possible pairs of classes. We will then assign this pixel to the class that has been selected most often by these classification rules.

Reject Class

We may introduce a reject class to the classification schemes described above. By this we mean a null class consisting of pixels that in some sense are too far from the known populations to be classified as any of these.

If we restrict each class to a certain Mahalanobis distance from the class mean we obtain a consistent way of doing this. The Mahalanobis distance from a feature

vector $\mathbf{x}$ to the ith class mean is given by

$$D_M^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$$

As $D_M^2 \in \chi^2(n)$, a convenient way of specifying the reject class is by a quantile in the χ^2-distribution.

Distance Between Class Centers

Having generated a series of training sets, it is useful to establish a measure of class uniqueness. The purpose of this is to determine which classes we can differentiate between and which classes should be merged. The elements of the posterior probability matrix for classification of class centers are given by

$$A_{ij} = k(\pi_i \mid \mathbf{x} = \boldsymbol{\mu}_j)$$

i.e., in the jth column we find the posterior probabilities for all classes given that the class centre for the jth class has been observed. Thus, each column adds to 1. This matrix of posterior probabilities gives important information on the uniqueness of a class. If the center of a particular class has a high probability of belonging to another class, then this is an indication for overlap between the classes. This should result in the classes being merged after classification. On the other hand, if the posterior probability of the center of a particular class belonging to that particular class is close to 1, then this class may be considered as unique. This procedure may also be used to evaluate a validation set.

DATA: MINERALS AND ELEMENTS

As the aim is to use the method for standard studies of sedimentary rocks, it is important to cover the most frequently occurring minerals. Table 1 shows all

Table 1. Mineral Classes in the Model

Albite	Chlorite 2	Gypsum	Quartz
Ankerite	Chlorite 3	Illite/Muscovite	Rutile
Apatite	Dolomite	Ilmenite	Siderite 1
Barite	Fe-calcite	Kaolin	Siderite 2
Biotite 1	Garnet 1	K-feldspar	Titanite
Biotite 2	Garnet 2	Monazite	Tourmaline
Calcite	Garnet 3	Porosity	Zincblende
Chlorite 1	Glauconite	Pyrite	Zircon

Table 2. Parameter Setting on the SEM

Pulse processing time	19 μs
Energy range	10 keV
Channel width	10 eV
Dwell time	0.020 s
Map resolution	256 $\times$ 256 pixels
Magnification (typical)	100 x
Pixel size (100 $\times$ magnification)	2.4 μm $\times$ 2.4 μm
Accelerating voltage	12 V
Current	1 nA
Deadtime	40%

minerals in the model. In some cases, more than one class is needed to describe a mineral. This can be due to natural variation in the chemical composition of the mineral, such as in the biotites, the chlorites, the garnets, and the siderites. There is also a case, for illite and muscovite, where it is known in advance that different minerals have the same chemical composition. They will therefore be overlapping in the EDS measurements. In addition to minerals, it is also important to have a porosity class.

The data are counts from the EDS detector. The data are stored as 16 bits per pixel, as the range of the data goes beyond the value 255 with the current setting of the microscope parameters, which is shown in Table 2.

The mapped elements reflect the major components in the minerals. It is normally the K_α line that is mapped, but in some instances this is superimposed by another element's L_α line. This is the case for P and Zr, Ti and Ba, and Na and Zn. There have not been any problems with this duality of the data; it has rather increased the possibility of discriminating between more minerals. An image with all elements is shown in Figure 4. The set of mapped elements is given in Table 3. The range of data in the different bands is shown in Table 4. To give an impression of correlation between the variables, Table 5 and Figure 5 show eigenvalues of correlation matrices for the 32 training classes and simple statistics. We see that most eigenvalues for most classes are close to 1 indicating very little correlation between the variables. Chlorite 1, chlorite 2, illite/muscovite, kaolin, zincblende, and siderite 2 have both higher and lower eigenvalues than most other classes, indicating slightly higher correlations between variables for these classes.

During the course of this work, the data quality has been improved in several ways. The parameter settings in the SEM have been varied in order in obtain well-separated peaks in the EDS spectra, to obtain an intensity level that clearly separates the peaks from the background noise level and to obtain a balance in the EDS sensor sensitivity of light vs. heavy elements. Dwell time and dead time in the EDS sensor affect the acquisition time for an image and thereby the number

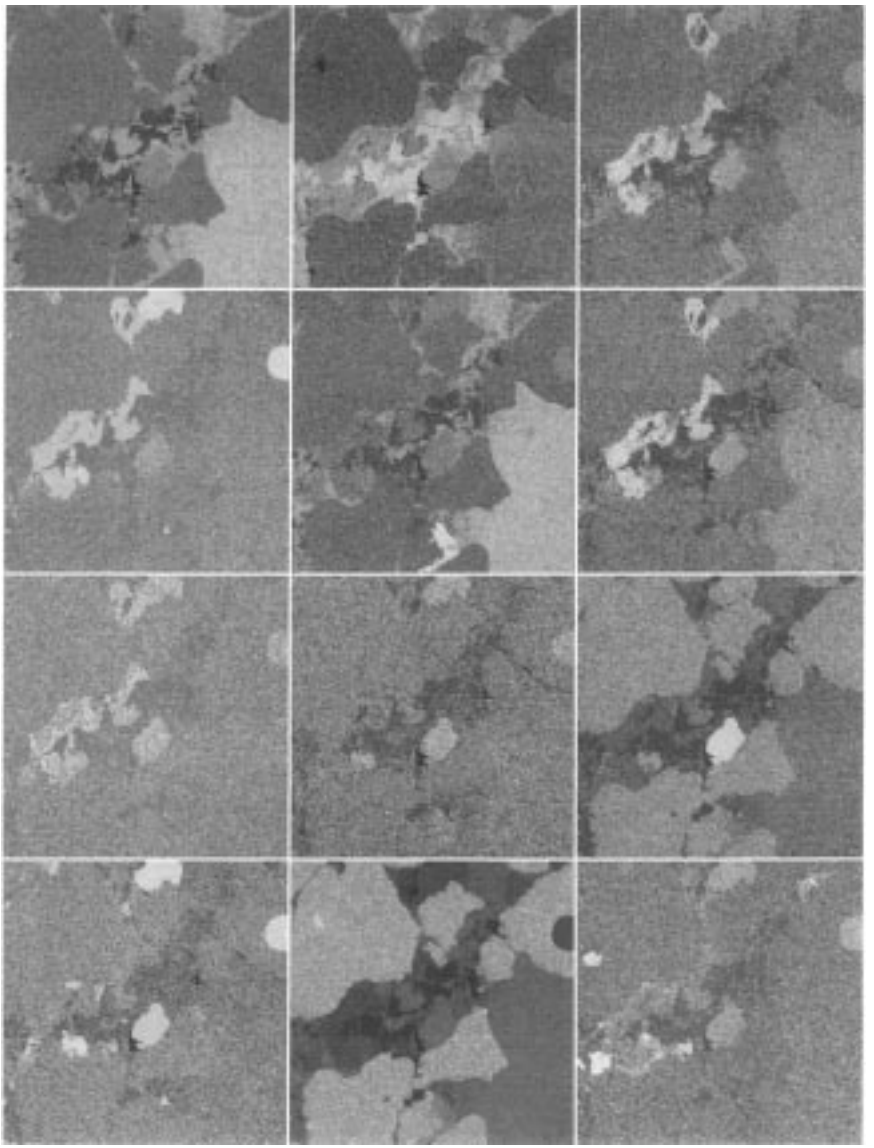

Figure 4. Example on SEM EDS image. The elements mapped are (rowwise) Al, C, Ca, Fe, K, Mg, Mn, Na (+Zn), P (+Zr), S, Si, and Ti (+Ba).

Table 3. Mapped Elements

Al	C	Ca
Fe	K	Mg
Mn	Na (+Zn)	P (+Zr)
S	Si	Ti (+Ba)

Table 4. Range of Values of the Mapped Elements for All Training Areas[a]

	Al	C	Ca	Fe	K	Mg	Mn	Na	P	S	Si	Ti
Minimum	0	1	0	0	0	0	0	0	0	0	0	0
Maximum	149	165	124	44	60	99	16	228	171	282	331	110
Mean	34.00	44.48	28.48	6.18	9.26	24.30	1.91	10.89	12.38	21.10	73.55	5.29
Median	16	44	8	6	6	17	2	9	9	7	57	3
SD	30.58	20.39	32.65	6.54	9.61	19.03	1.50	13.85	14.66	43.53	80.81	10.22

[a] The total number of pixels in the training areas is 59207.

Table 5. Eigenvalues for the Correlation Matrices for the Individual Classes

Albite	1.25	1.17	1.12	1.11	1.08	1.03	0.96	0.94	0.92	0.90	0.78	0.72
Ankerite	1.26	1.14	1.06	1.05	1.03	1.00	0.97	0.95	0.94	0.92	0.85	0.82
Apatite	1.65	1.50	1.40	1.21	1.14	0.99	0.90	0.80	0.67	0.63	0.58	0.53
Barite	1.61	1.40	1.15	1.10	1.04	0.96	0.92	0.90	0.82	0.75	0.72	0.63
Biotite	1.50	1.24	1.10	1.04	1.03	1.00	0.96	0.94	0.86	0.85	0.79	0.70
Biotite 2	1.21	1.15	1.12	1.10	1.09	1.05	1.01	0.99	0.94	0.90	0.82	0.61
Calcite	1.15	1.10	1.10	1.05	1.03	1.01	0.99	0.96	0.95	0.93	0.91	0.82
Chlorite 1	2.42	1.16	1.14	1.04	0.97	0.94	0.91	0.85	0.76	0.68	0.59	0.54
Chlorite 2	2.37	1.26	1.11	1.07	1.01	0.96	0.91	0.85	0.78	0.66	0.55	0.48
Chlorite 3	1.28	1.15	1.08	1.07	1.05	0.99	0.98	0.97	0.91	0.89	0.84	0.81
Dolomite	1.21	1.08	1.08	1.07	1.05	1.02	0.99	0.97	0.95	0.89	0.85	0.83
Fe-calcite	1.15	1.10	1.07	1.06	1.03	1.03	1.00	0.96	0.95	0.92	0.89	0.85
Garnet 1	1.40	1.30	1.16	1.13	1.09	1.02	0.99	0.96	0.80	0.74	0.71	0.70
Garnet 2	1.18	1.15	1.13	1.10	1.03	0.99	0.98	0.95	0.93	0.89	0.87	0.79
Garnet 3	1.22	1.17	1.15	1.10	1.09	1.02	0.98	0.96	0.86	0.85	0.82	0.79
Glauconite	1.21	1.18	1.13	1.10	1.06	1.00	0.95	0.94	0.93	0.92	0.83	0.74
Gypsum	1.33	1.24	1.04	1.04	1.01	0.99	0.98	0.94	0.92	0.87	0.85	0.79
Ill/Musc	1.99	1.21	1.18	1.13	1.07	1.00	0.99	0.87	0.82	0.71	0.60	0.44
Ilmenite	1.29	1.22	1.19	1.09	1.07	1.01	0.99	0.97	0.93	0.85	0.74	0.65
Kaolin	2.25	1.14	1.10	1.08	1.02	0.97	0.94	0.92	0.88	0.83	0.44	0.42
K-feldspar	1.21	1.15	1.08	1.07	1.03	1.01	0.98	0.95	0.92	0.91	0.87	0.82
Monazite	1.40	1.31	1.25	1.15	1.09	1.03	0.96	0.90	0.86	0.75	0.69	0.61
Porosity	1.33	1.13	1.11	1.07	1.04	1.02	1.00	0.94	0.92	0.88	0.84	0.72
Pyrite	1.29	1.20	1.15	1.11	1.06	1.02	0.99	0.96	0.92	0.84	0.78	0.66
Quartz	1.10	1.09	1.08	1.05	1.03	1.01	1.00	0.97	0.93	0.92	0.91	0.89
Rutile	1.56	1.36	1.25	1.14	1.08	1.02	0.95	0.86	0.83	0.71	0.62	0.61
Siderite	1.32	1.18	1.16	1.07	1.06	1.03	0.95	0.94	0.90	0.83	0.81	0.75
Siderite 2	2.22	1.20	1.10	1.04	1.00	0.91	0.89	0.84	0.80	0.74	0.66	0.60
Titanite	1.29	1.16	1.14	1.11	1.08	1.01	0.98	0.93	0.87	0.86	0.81	0.77
Tourmaline	1.37	1.14	1.12	1.09	1.04	0.99	0.98	0.97	0.90	0.89	0.87	0.64
Zincblende	2.13	1.43	1.23	1.07	1.01	0.94	0.87	0.84	0.76	0.68	0.64	0.40
Zircon	1.43	1.18	1.16	1.12	1.07	0.99	0.96	0.93	0.86	0.84	0.74	0.73
Mean	1.49	1.21	1.14	1.09	1.05	1.00	0.96	0.93	0.88	0.83	0.76	0.68
SD	0.39	0.10	0.07	0.04	0.03	0.03	0.04	0.05	0.07	0.09	0.12	0.13
Min	1.10	1.08	1.04	1.04	0.97	0.91	0.87	0.80	0.67	0.63	0.44	0.40
Max	2.42	1.50	1.40	1.21	1.14	1.05	1.01	0.99	0.95	0.93	0.91	0.89

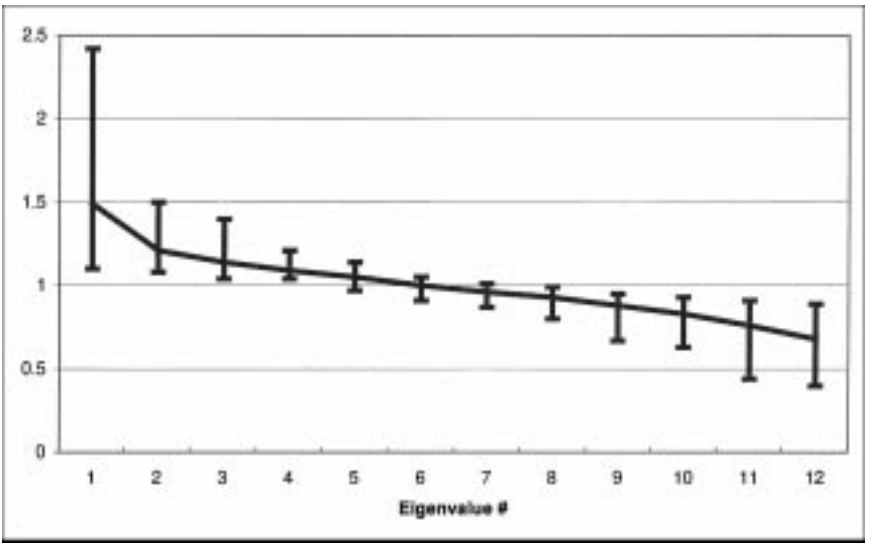

Figure 5. Mean value and range of eigenvalues from correlation matrices of all training classes. Values close to 1 indicate noncorrelated input variables. The values are given in Table 5.

of samples that it is possible to analyze. Also, the suite of samples selected for training and validation of the model are chosen so that they represent the most common minerals in both siliciclastic and carbonate rocks.

The total data set used in the study comprise 36 images each with 256×256 pixels. This gives 2,359,296 observations in total. The pixel size is 2.4 μm $\times$ 2.4 μm.

Training Areas

Training areas are made with the seed algorithm as described above. Only one pixel is chosen as the starting point for each training area. In the first step, a radius of 5 pixels is normally sufficient to get an initial estimate of the mean value and the dispersion matrix of the class. For more spatially dispersed minerals, such as chlorite 2 (which is a chlorite coating on grains) and illite/muscovite it is necessary to increase the radius to 10. The Euclidean distance that is used as an acceptance limit is normally set to 30. Barite, K-feldspar, zincblende, and zircon need a Euclidean distance of 50 in order to include a sufficient number of pixels. The final training areas are grown based on the Mahalanobis distance method with a 0.99 quantile in the χ^2-distribution with 12 degrees of freedom. The resulting training area for chlorite 2 is shown in Figure 1.

Scatter plots of the two first CDFs from most training areas show a consistent set, with no subgroups and with little scatter around the main cluster. This

is contrary to results from training areas created by painting an area in the image, which is what most classification packages encourage. Poor definition of the structure of the real class and inclusion of much noise is often the result from such an approach. Figure 6 shows the scatter plots as described above of a hand drawn training area compared to a seed-grown training area for kaolin.

Validation Areas

Validation areas are defined in exactly the same manner as the training areas. Validation samples are preferably chosen from other wells (or fields) than the samples for the training areas to assure independence. For some of the rarer minerals, it has not been possible to find independent samples and no validation therefore exists. The results of the validation are discussed below.

Selection of Features—Jeffreys–Matusita Distance

It is relatively easy to figure out which elements to map in order to cover the most important components in the minerals of the model. However, we cannot know in advance that all of these elements are needed to discriminate between the classes. Nor do we know in advance that it will be possible to separate all classes that we have included. The JM distance measure estimates the separation ability for all input variables (features), single and in groups, for all classes. The classes are assumed to have multinormal distributions, and the mean vectors and dispersion matrices that are estimated from the training areas are used as input.

Table 6 shows the results of the average JM value JM_{AVE}. It is shown in the table that all features contribute to the overall separation ability of the model, but the last included features have little influence on the classification. The average JM value for the training set is 1.4102 against the maximum value of 1.4142, or 99.72% of the maximum. This result is very good, and it must be ascribed to the way the training areas are constructed. Standard ways of making training areas would have included more outliers that would increase the tolerance of the classes, and thereby decrease the separability between the classes.

The detailed result of the JM analysis is a 32×32 matrix, which is too large to reproduce here. Instead, the pairs of classes that have the highest degree of overlap are shown in Table 7. It is clear that calcite and Fe-calcite cannot be separated and that they should be combined into one class after classification. The accelerating voltage in the SEM favors good resolution in the lightest elements at the expense of resolution in the heavy elements. This may be a reason for the poor separation between calcite the Fe-calcite. There are no other overlaps between the calcites and other classes. The three garnet classes also have a degree of overlap that is too high and they should therefore be combined into one class after classification. Other classes overlap as well, but not to such an extent that it is necessary to combine them. All class pairs that are not mentioned in Table 7 have a JM value

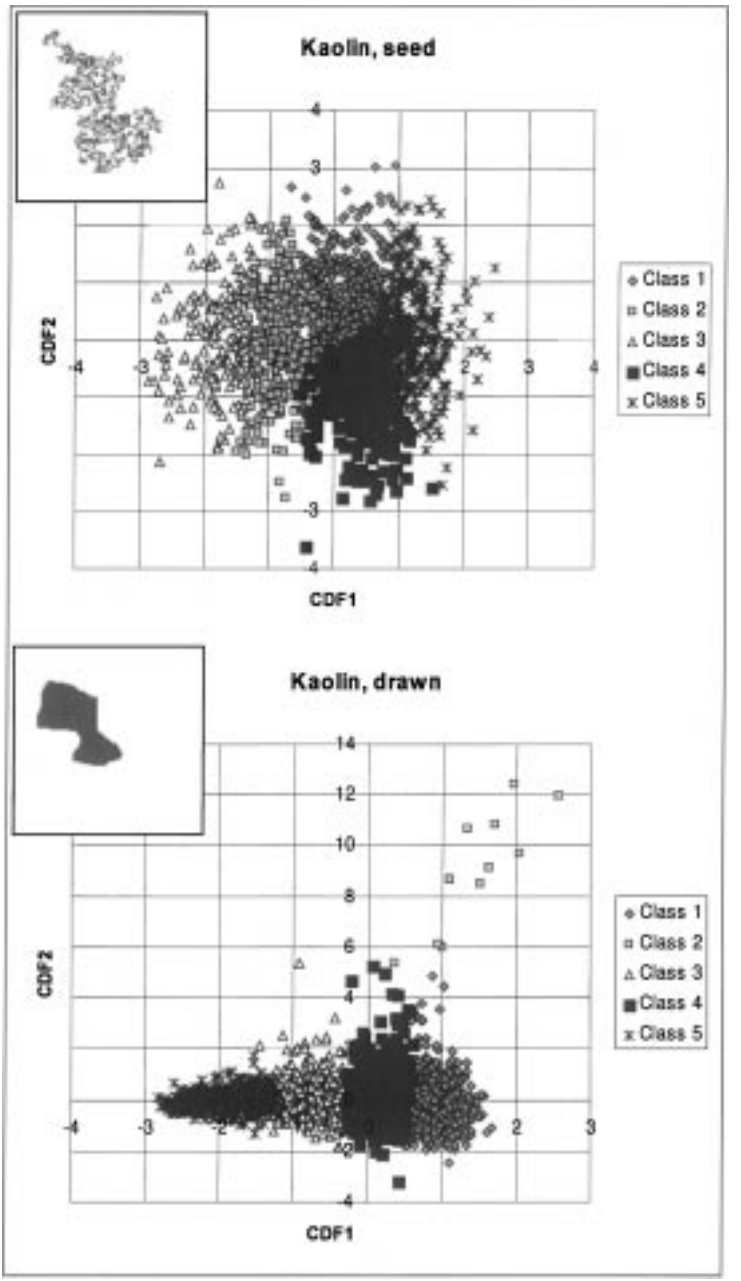

Figure 6. Scatter plots of training areas for kaolin, comparison between the results of a training area grown with seed algorithm and a hand drawn training area. The training areas (in the same scale) are shown in the plots.

Table 6. Optimal Selection of Features According to the JM Method

No. of features	Included features	Average JM value
1	Si	1.1257
2	Ca, Si	1.3081
3	C, Ca, Si	1.3679
4	C, Ca, S, Si	1.3876
5	Al, C, Ca, Mg, S	1.3992
6	Al, C, Ca, Mg, S, Si	1.4035
7	Al, C, Ca, K, Mg, P, S	1.4060
8	Al, C, Ca, Fe, K, Mg, P, S	1.4083
9	Al, C, Ca, Fe, K, Mg, Mn, P, S	1.4093
10	Al, C, Ca, Fe, K, Mg, Mn, P, S, Ti	1.4099
11	Al, C, Ca, Fe, K, Mg, Mn, P, S, Si, Ti	1.4101
12	Al, C, Ca, Fe, K, Mg, Mn, Na, P, S, Si, Ti	1.4102

Table 7. Pairs of Overlapping Classes with Their JM Value

JM value	Mineral 1	Mineral 2
0.6336	Fe-calcite	Calcite
0.8715	Garnet 3	Garnet 1
1.1504	Garnet 3	Garnet 2
1.2139	Garnet 2	Garnet 1
1.3665	Kaolin	Ill/Musc
1.3725	Siderite 2	Siderite 1
1.3768	Chlorite 3	Chlorite 2
1.3768	Chlorite 2	Chlorite 1

that is higher than 1.400 or 98.99% of maximum value, and the absolute majority of those have perfect separation to 4 decimal places.

CLASSIFICATION

Due to the expected Poisson nature of the data (which means that the mean equals the variance for all variables in each class), we find it necessary to have separate dispersion matrices for each class in order to define the natural variation in the variables. Therefore, it is not advisable to use linear classification. The hierarchical method described by Jia and Richards (1996) is found to introduce a small bias with respect to the ordering of the classes and is therefore considered not fully satisfactory. We will go through the results of four different classification methods—namely, simple quadratic classification, contextual quadratic classification, and hierarchical and extended hierarchical quadratic classification—with the main emphasis on the simple quadratic classification method. The effect of choosing different classification methods with regard to error rates is covered in more detail in the next sections. Here we will discuss some aspects that are not clear from the confusion matrices and reject class.

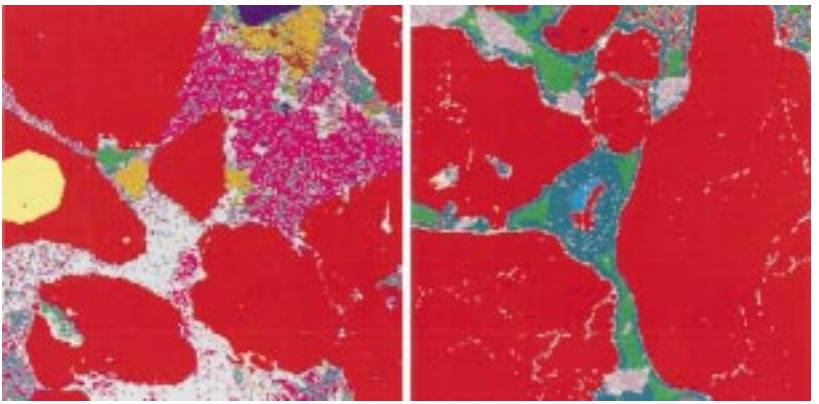

Figure 7. Example on classified images. In the left image, we see how quartz cement engulfs a pyrite grain. The pore space between the grains is mostly filled with clays, such as illite and kaolin. In the right image, we see chlorite coating the pores. There is also some siderite cement. See Figure 8 for color legend.

Figure 8. Color legend for classified images.

The different classification methods have different complexity and thus different computation time. For many purposes, the fastest method, i.e., the hierarchical quadratic method, may be preferred. However, there are some advantages of the other methods that should be noted.

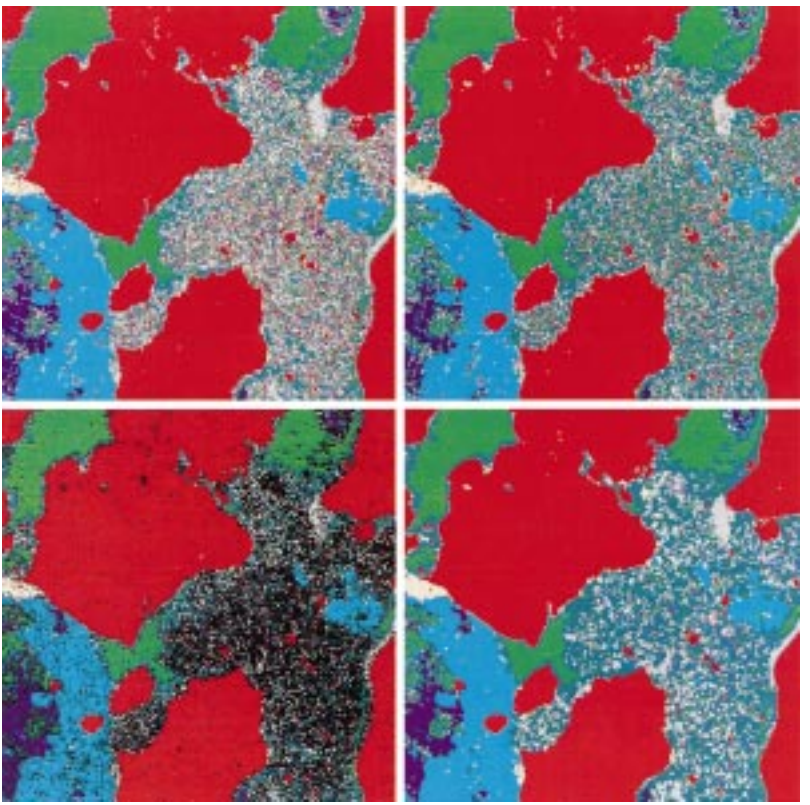

Figure 9. Result of different classification methods and reject class. The upper left image is classified with the quadratic classification method, the upper right with the extended hierarchical classification method, and the lower right with the contextual classification method. The lower left image is the same as the upper left, but with reject class added in black. The large area with mixed color is a clast of detrital clay. The pores are coated with chlorite. See Figure 8 for color legend.

An illustration of classification results is seen in Figure 7. In the left part of the image, we see how quartz cement engulfs a pyrite grain. In the right part, we see how chlorite coats the quartz grains, thereby inhibiting quartz cementation. A color legend for the classified images is shown in Figure 8.

The contextual classification estimates the posterior probability based on a neighborhood around each pixel. The result is that the classes have a more continuous distribution in space. The images classified with this method look less noisy, compare upper left and lower right sections of Figure 9.

The visual appearance of the simple quadratic classification and the extended hierarchical classification may be similar, see upper left and upper right sections of

Table 8. Processing Time for Different Classification Methods

Method	Total time (s)	Time per pixel (ms)	Relative time
Min. dist. linear	85.72	0.04	0.19
Min. dist. quad.	216.19	0.09	0.48
Quadratic	452.39	0.19	1.00
Context-Q	2497.83	1.06	5.52
Hierarchical	319.58	0.14	0.71
Ext. hierarchical	1353.96	0.57	2.99

Figure 9, but the results differ, especially in areas with detrital clays. The strength of the hierarchical method is its possibility to optimize the selection of variables for two classes at a time as opposed to all classes simultaneously. This will ensure a robust classification, and it is found that the JM selection of variables is similar to the intuitive choice. For example, to separate quartz and albite, only the Al feature is used, which is very satisfactory. The only cases where several variables are included are the ones where we know that the classes overlap. Even with the reduced number of variables, this method requires more CPU time than simple quadratic classification, as the number of classifications is far higher. Table 8 shows the processing time on an HP 9000/782 system with PA-2 8200 236 MHz processors for a 1536×1536 pixels image and processing time per pixel for the different classification methods.

Based on the observations above, we see that the classification methods all have their pros and cons. The choice of method is governed by the application where demand for system throughput or accuracy may be the most important factors. The conclusions of the sections on training and validation set based confusion matrices (below) must also be taken into account.

Poisson Distributed and Near Orthogonal Data

To support the remarks on the expected Poisson nature of the data mentioned in the introduction and in previous section, a plot of element variance vs. element mean for all elements and all mineral classes is shown in Figure 10. The data used are based on results from the simple quadratic classification with a 99% reject quantile. A weighted regression analysis shows that Variance $= 1.61 \times$ Mean with $R^2 = 0.94$ (the intercept is not significant), which indicates some overdispersion. As a (standard) remedy for variance stabilization of Poisson-distributed data, square roots of all quantities are taken. Ideally this causes all variances to become constant (equal to $\frac{1}{4}$). In this case the results of taking the square root are shown in Figure 11 (with the same training and validation areas, i.e., grown on the original data prior to taking the square root). An alternative to the classification methods described above is thereby possible. With constant variance a linear classification can be performed. Considering the near orthogonality of the features measured

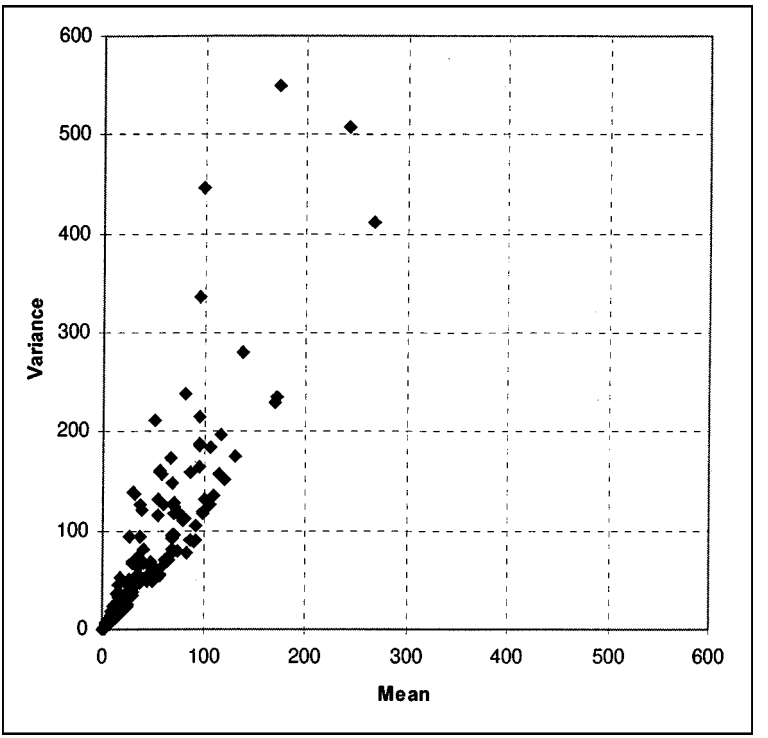

Figure 10. Scatter plot of element variance and mean for all elements and all mineral classes.

linear and quadratic minimum distance classifications are carried out (with the same training and validation data as above). Summary of the results are shown in Table 10 and Table 12. Processing times are shown in Table 8. These results are not discussed further.

POSTCLASSIFICATION ANALYSIS—QUALITY CONTROL

Training Set Based Confusion Matrix

We assume that this analysis will show much of the same results as the JM analysis did. From Table 9 we see that the highest error rates occur for the same minerals, which showed class overlap in the JM analysis. Regardless of classification method, the garnets and calcites have the highest error rates. Apart from these, chlorite 2 and illite/muscovite have the highest error rates, but these are very low. When comparing the total error rates, we find that the contextual

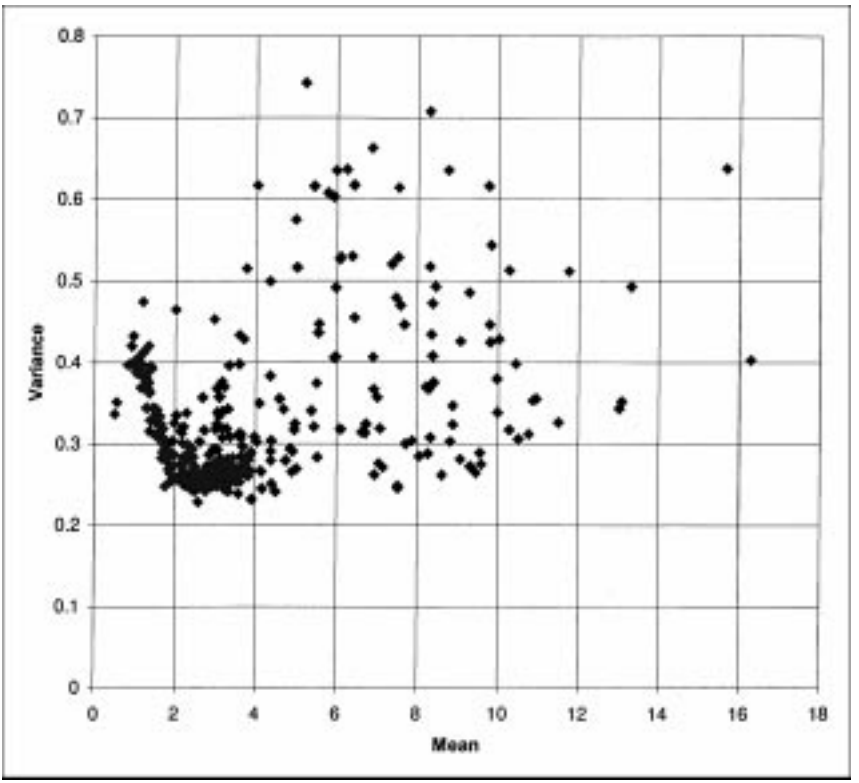

Figure 11. Scatter plot of element variances and mean for all elements and classes after taking the square root of all data.

quadratic classification method is the best. However, because of the simpler method and therefore much lower computing time, the simple quadratic method may be preferred.

If we combine the calcites into one class and the garnets into one class, we find that the rates of misclassification drop from 1.8–4.2% to 0.25–0.33% (see Table 9 and Table 10).

Validation Set Based Confusion Matrix

Validation of a classification model is more trustworthy if an independent validation set, not used in building the model, is used. In this case, we have validation samples for all minerals except the following four: biotite 2, chlorite 3, siderite 2, and zincblende. The error rates from classification of the validation set are given in Table 11.

Table 9. Summary of Training Set Based Confusion Matrix: Fractions of Misclassification

	Quadratic	Contextual quadratic	Hierarc.	Ext. hierarc.
Albite	0.000	0.000	0.000	0.000
Ankerite	0.001	0.000	0.005	0.005
Apatite	0.000	0.000	0.000	0.000
Barite	0.000	0.000	0.000	0.000
Biotite	0.003	0.001	0.002	0.002
Biotite 2	0.000	0.000	0.003	0.005
Calcite	0.274	0.138	0.290	0.290
Chlorite 1	0.021	0.036	0.027	0.016
Chlorite 2	0.008	0.006	0.014	0.012
Chlorite 3	0.009	0.001	0.012	0.012
Dolomite	0.001	0.002	0.003	0.003
Fe-calcite	0.238	0.097	0.239	0.239
Garnet 1	0.250	0.047	0.282	0.280
Garnet 2	0.088	0.010	0.092	0.091
Garnet 3	0.200	0.022	0.206	0.205
Glauconite	0.000	0.000	0.001	0.001
Gypsum	0.000	0.000	0.000	0.000
Illite/Musc	0.021	0.087	0.036	0.036
Ilmenite	0.000	0.000	0.000	0.000
Kaolin	0.008	0.010	0.018	0.018
K-feldspar	0.000	0.000	0.001	0.002
Monazite	0.000	0.000	0.000	0.000
Porosity	0.000	0.000	0.000	0.000
Pyrite	0.000	0.000	0.000	0.000
Quartz	0.000	0.000	0.000	0.000
Rutile	0.000	0.008	0.000	0.000
Siderite	0.008	0.017	0.007	0.007
Siderite 2	0.007	0.002	0.020	0.020
Titanite	0.000	0.000	0.000	0.000
Tourmaline	0.002	0.001	0.004	0.003
Zincblende	0.000	0.031	0.000	0.000
Zircon	0.000	0.000	0.000	0.000
Total	0.038	0.018	0.042	0.042

Table 10. Overall Misclassification Rates with Calcite and Fe-Calcite Combined, the Three Garnet Classes Combined and Biotite Removed from the Validation Set

	Quadratic	Contextual quadratic	Hierarc.	Ext. hierarc.	Min. distance linear	Min. distance quadratic
Training set	0.0025	0.0033	0.0025	0.0025	0.0046	0.0034
Validation set	0.0106	0.0065	0.0109	0.0113	0.0188	0.0141

We expect to find the same errors here as reported in the section above, and this is the case. In addition, there is one clear error in that biotite is totally misclassified. The reason for this turns out to be that the sample picked for validation of biotite is altered to such a degree that it is closer to the chlorites in its chemical composition. Biotite is a mineral that is easily altered and therefore hard to validate with deeply

Table 11. Summary of Validation Set Based Confusion Matrix: Fractions of Misclassification

	Quadratic	Contextual quadratic	Hierarc.	Ext. hierarc.
Albite	0.000	0.000	0.000	0.000
Ankerite	0.067	0.006	0.037	0.037
Apatite	0.000	0.000	0.000	0.000
Barite	0.000	0.000	0.000	0.000
Biotite	1.000	1.000	1.000	1.000
Biotite 2				
Calcite	0.325	0.196	0.318	0.316
Chlorite 1	0.070	0.002	0.128	0.078
Chlorite 2	0.115	0.072	0.131	0.124
Chlorite 3				
Dolomite	0.000	0.005	0.005	0.004
Fe-calcite	0.859	0.948	0.924	0.928
Garnet 1	0.119	0.009	0.178	0.119
Garnet 2	0.059	0.004	0.078	0.076
Garnet 3	0.964	0.996	0.960	0.960
Glauconite	0.000	0.000	0.014	0.005
Gypsum	0.000	0.000	0.000	0.000
Illite/Musc	0.003	0.000	0.013	0.011
Ilmenite	0.068	0.000	0.111	0.111
Kaolin	0.042	0.092	0.111	0.131
K-feldspar	0.000	0.001	0.000	0.000
Monazite	0.000	0.000	0.000	0.000
Porosity	0.000	0.000	0.000	0.000
Pyrite	0.000	0.000	0.000	0.000
Quartz	0.000	0.000	0.000	0.000
Rutile	0.000	0.027	0.000	0.000
Siderite	0.002	0.011	0.004	0.004
Siderite 2				
Titanite	0.000	0.000	0.000	0.021
Tourmaline	0.003	0.000	0.010	0.013
Zincblende				
Zircon	0.000	0.000	0.000	0.000
Total	0.169	0.165	0.185	0.185

buried samples. Apart from these obvious errors, the highest error rate is found for chlorite 2—namely, 11.5% for quadratic classification—but most of this is against chlorite 1. The rest of the error rates are less than 7%, which must be characterised as low.

With combined classes as above and disregarding biotite and the minerals that lack validation, the misclassification rates drop from 16.5–18.5% to 0.65–1.13% (see Table 10 and Table 11).

We can therefore conclude that the classification model is highly successful, after combination of identified overlapping classes. This combination of classes makes sense from a mineralogical point of view.

Table 12. Fraction of Rejects

	Quadratic	Contextual quadratic	Hierarc.	Ext. hierarc.	Min. distance linear	Min. distance quadratic
Rejects in training areas	0.00	0.00	0.00	0.00	0.02	0.00
Rejects in validation areas	0.12	0.14	0.07	0.07	0.10	0.12

Reject Class

Whereas the confusion matrices can be used in the model building phase only, reject class is used in all future classifications. The numbers given in Table 12 are estimated from the images used for training and validating the model. The reject class contains data points with a Mahalanobis distance to the nearest class center beyond the 0.99 quantile in the χ^2-distribution. Choosing a smaller quantile in the estimation of the dispersion matrices would produce more rejects. We have a trade-off between being able to separate classes and rejecting natural variation in the chemical composition of the minerals.

The rejects represent three different cases:

- minerals not included in the classification model,
- variations in the chemical composition of the minerals that are not reflected in the samples used for training,
- variations in the image acquisition.

Whereas the first two of these conditions result in isolated grains, the third condition results in large areas with rejects, especially in the quartz grains. This is used as an indication that the image acquisition must be repeated for the affected samples. Precautions are taken to keep the conditions in the SEM as steady as possible, but it is still worthwhile to have the reject class as a quality control. Table 12 shows that the amount of rejects for quadratic and contextual classifications is approximately the same and that the hierarchical classifications give a smaller amount of rejects.

The lower left part of Figure 9 shows that most rejects are found in an area with detrital clays, but that we also find some scattered rejected pixels within the grains.

Routine analyses of more than 2000 images from 28 different wells based on the classification model developed, show a constant rate of rejects. Closer inspection of the rejects shows that most of them come from minerals not included in the classification model. Variation in the image acquisition has not given rise to rejects except in extreme cases such as filament breakage during image acquisition.

We can therefore conclude that the reject class serves its purpose as a quality control, and that the image acquisition is stable.

Distance to Nearest Class

The estimates of distances between class centers show the effect of removing a class from the model. If a class were removed from the model, the data originally belonging to that class would most likely be classified as the nearest class. In this manner, we get an idea of the behavior of the reduced model without having to build and validate the model once more.

The distance to the nearest class center is defined as the minimum posterior probability of a class center belonging to all other classes. Thus, we can construct a matrix that shows distances to all other classes. If the probability of belonging to the nearest class is comparable to that of belonging to the correct class, we have once more an indication of partly overlapping classes. If the probability is clearly below that of the correct class, but still within measurable range, we identify the nearest class and see whether the choice makes sense from a mineralogical point of view. In some cases, it is found that the probability of belonging to another class is extremely low. This is the case for several of the heavy minerals that have chemical compositions that are far from those of the other classes. In one extreme case, for zincblende, the posterior probability of the nearest class is so small that it is zero to the precision of 8 bytes floating-point numbers (see Figure 12).

In most cases, the choice of nearest class corresponds to mineral grouping. It seems right that albite is the nearest class to K-feldspar (it should be noted that the opposite is not the case). However, for the heavy minerals, the chemical composition can be so different from the other classes that the choice of nearest class becomes more arbitrary.

Sensitivity Study of the Seed-Growing Algorithm

In a separate sensitivity study (Larsen, Nielsen, and Flesche, 1999) it is found that the seed-growing algorithm is very robust with respect to the setting of its parameters. The sensitivity is evaluated with respect to the misclassification rate of the validation set.

The Mahalanobis seed growing requires one parameter setting—namely, the choice of quantile for the threshold. The algorithm is insensitive to the choice of this quantile.

For the initial Euclidean seed growing, two parameter settings are required—in particular, a maximum spatial range and a threshold for the spectral distance. Sensitivity to the setting of these two parameters is restricted to the spatially dispersed classes.

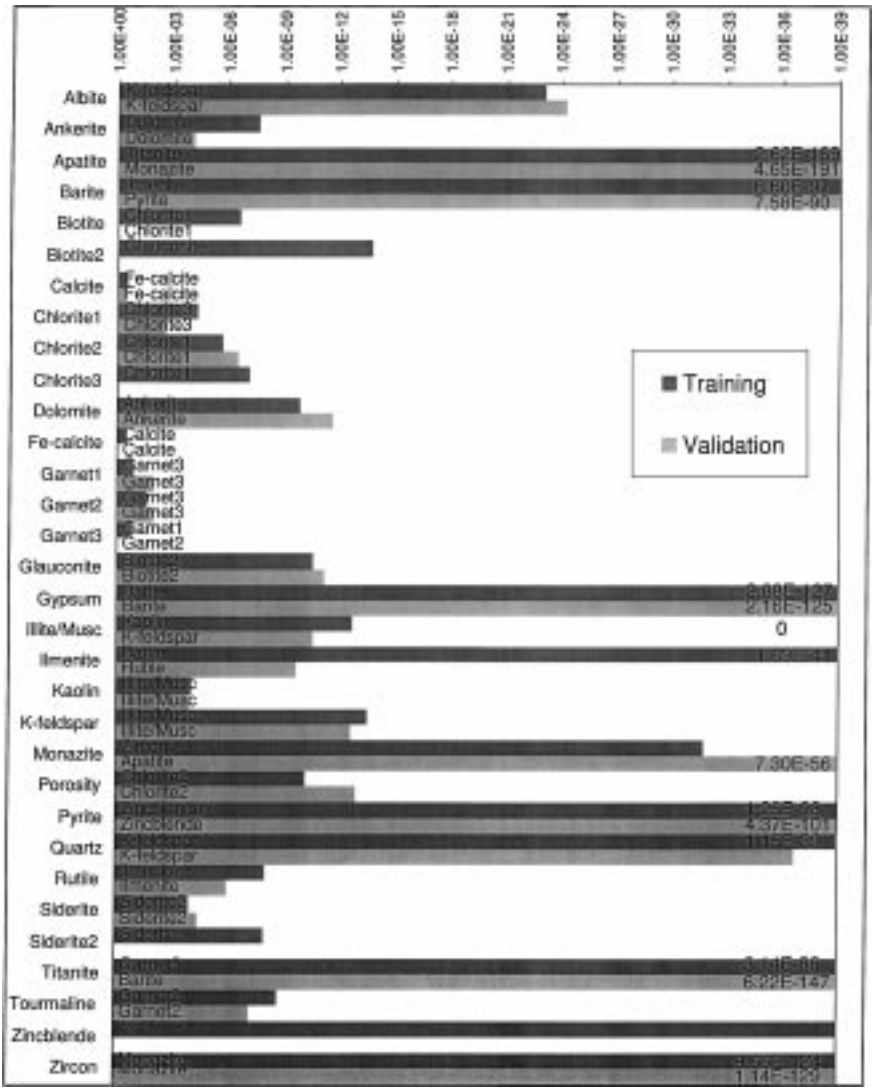

Figure 12. Distance to nearest class.

Furthermore, the initial choice of seeding point is not critical, except for the spatially dispersed classes.

The conclusion of the study is that the seed growing is very robust—only in the case of spatially dispersed classes tuning of the parameters for the initial Euclidean growing is necessary. The study also shows that updating parameters during the Mahalanobis distance growing should not be carried out.

CONCLUSIONS

This paper describes a system for classification of minerals based on SEM EDS images. In spite of the noisy visual appearance of the input data, 29 different mineral classes are successfully classified, covering the most common minerals in both siliciclastic and carbonate rocks. The success of the system is a result of improvements in the data acquisition and the systematic use of multivariate statistical methods such as the semiautomatic training and validation set generation, the Jeffreys–Matusita distance measure, canonical discriminant analysis, and different supervised classification methods.

Although we have linked the methods closely with an application in SEM EDS imagery, we believe they can be used for other types of data as well, such as air- or satellite-borne remote sensing data.

Some mineral classes are found to be overlapping. It is not possible to discriminate between ferrous and nonferrous calcite. Likewise, the garnets show too high a degree of overlap to separate them. Both siderite and biotite are divided into two subclasses, and chlorite into three.

The analysis prior to classification covers construction of training and validation areas and Jeffreys–Matusita analysis. These steps in the analysis give a clear indication of the performance of the classification model. Much of the success in separating the classes is attributed to the seed algorithm used for defining training and validation areas.

Selection of classification method should be done based on requirements of accuracy and processing time. Analysis shows that the data are close to orthogonal. This suggests use of a minimum distance classifier. It is also shown that due to the Poisson nature of the data, linear classification can be performed after taking the square root of the data.

Distances to other classes are calculated as the posterior probability of a class center belonging to the other classes. By comparing the posterior probability of the assumed class with that of the nearest class, we obtain a measure of the uniqueness of the class. If a class is deleted from the model, observations from this class are likely to be classified as the nearest class.

The most important quality control measure of the classification model is the validation set based confusion matrix. In most cases, the validation areas are similar to the training areas with regard to the degree of misclassification. Validation of the biotites has proven difficult, as the assumed validation area for biotite is mostly classified as chlorite. This is explained as an effect of alteration of biotite. With combination of the two calcite classes and the three garnet classes, and also disregarding the validation of biotite, the misclassification rates of the validation set is 0.65–1.13%. These results are considered remarkably good.

The amount of rejects in the classified images gives an indication of the performance of the classification model. This is used both in the model-building

phase and later when the classification is run as a standard analysis. Tests from 28 wells show a constant rate of rejects.

This model is found suitable for classification of 29 classes covering 24 different minerals (and porosity) that are common in both siliciclastic and carbonate rocks.

A separate sensitivity study shows that the seed growing is very robust, only in the case of spatially dispersed classes tuning of the parameters for the initial Euclidean growing is necessary. The study also shows that updating parameters during the Mahalanobis distance growing should not be carried out.

The classification model described here is now used for routine analyses of mineral composition in rock samples at Norsk Hydro Research Centre.

ACKNOWLEDGMENTS

We would like to thank Dr. Mogens Ramm, Norsk Hydro Exploration, for suggesting SEM EDS image analysis as a petrographical analysis method at Norsk Hydro and later for input on mineralogy; Professor Knut Conradsen and Dr. Bjarne Ersbøll, both of the Department of Mathematical Modelling, for many good discussions on multivariate analysis; Johannes Rykkje from Norsk Hydro Research Centre for expertise on SEM; and Johan Doré Hansen, Department of Mathematical Modelling, for coding part of the seed algorithm. We would also like to thank the reviewer for constructive comments on the manuscript. Finally, we would like to thank Norsk Hydro for permission to publish these results.

REFERENCES

Anderson, T. W., 1984, An introduction to multivariate statistical analysis, 2nd ed.: John Wiley & Sons, New York, 675 p.

Clelland, W. D., and Fens, T. W., 1991, Automated rock characterization with SEM/image-analysis technique: SPE Formation Evaluation, p. 437–443.

ERDAS Inc., 1990, ERDAS Version 7.4.

Ersbøll, B. K., 1989, Transformations and classifications of remotely sensed data: Theory and geological cases: doctoral dissertation 45, Department of Mathematical Modelling, Technical University of Denmark, 297 p. ISSN 0107-525x.

Fisher, R. A., 1936, The utilisation of multiple measurements in taxonomic problems: Annals of Eugenics, v. 7, p. 179–188.

Green, A. A., Berman, M., Switzer, P., and Craig, M. D., 1988, A transformation for ordering multispectral data in terms of image quality with implications for noise removal: IEEE Transactions on Geoscience and Remote Sensing, v. 26, no. 1, p. 65–74.

Haslett, J., 1985, Maximum likelihood discriminant analysis on the plane using a Markovian model of spatial context: Pattern Recognition, v. 18, no. 3, p. 287–296.

Hjort, N. L., 1985, Estimating parameters in neighbourhood based classifiers for remotely sensed data, using unclassified vectors: *in* Sæbø, H. V., Bråten, K., Hjort, N. L., Llewellyn, B., and Mohn, E.,

eds., Contextual classification of remotely sensed data: Statistical methods and development of a system: Norwegian Computing Center Technical Report No. 768.

Hjort, N. L., and Mohn, E., 1984, A comparison of some contextual methods in remote sensing classification: The 18th International Symposium on Remote Sensing of Environment, Paris, France, October 1984.

Hjort, N. L., Mohn, E., and Storvik, G., 1985, Contextual classification of remotely sensed data, based on an auto-correlated model, *in* Sæbø, H. V., Bråten, K., Hjort, N. L., Llewellyn, B., and Mohn, E., eds., Contextual classification of remotely sensed data: Statistical methods and development of a system: Norwegian Computing Center Technical Report No. 768.

Hughes, G. F., 1968, On the mean accuracy of statistical pattern recognition: IEEE Transactions on Information Theory, v. IT-14, no. 1, p. 55–63.

Jia, X., and Richards, J. A., 1996, Feature reduction using a supervised hierarchical classifier: The 8th Australasian Remote Sensing Conference, Canberra, Australia.

Larsen, R., Nielsen, A. A., and Flesche, H., 1999, Sensitivity study of a semi-automatic supervised classifier applied to minerals from x-ray mapping images: Proceedings of the 11th Scandinavian Conference on Image Processing (SCIA'99), June 7–11, 1999, Kangerlussuaq, Greenland, p. 785–792.

Matusita, K., 1966, A distance and related statistics in multivariate analysis, *in* Krishnaiah, P. R., ed., Multivariate analysis: Academic Press, New York, p. 187–200.

Minnis, M. M., 1984, An automatic point-counting method for mineralogical assessment: Am. Assoc. Petroleum Geologists Bull., v. 68, no. 6, p. 744–752.

Nielsen, A. A., 1994, Analysis of regularly and irregularly sampled spatial, multivariate, and multi-temporal data: doctoral dissertation 6, Department of Mathematical Modelling. Technical University of Denmark. ISSN 0909-3192. Internet address: http://www.imm.dtu.dk/documents/users/aa/phd/

Nielsen, A. A., Conradsen, K., and Simpson, J. J., 1998, Multivariate alteration detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: New approaches to change detection studies: Remote Sensing of Environment 64, p. 1–19.

Nielsen, A. A., Flesche, H., and Larsen, R., 1998, Semiautomatic supervised classification of minerals from X-ray mapping images: Proceedings of the 4th Annual Conference of the International Association for Mathematical Geology (IAMG'98), Ischia, Italy, October 1998, p. 473–478.

Owen, A., 1984, A neighbourhood-based classifier for LANDSAT data: The Can. Jour. Statistics, v. 12, p. 191–200.

Richards, J. A., 1993, Remote sensing digital images analysis: An introduction: Springer Verlag, Berlin, 340 p.

Safarian, S. R., and Landgrebe, D., 1991, A survey of decision tree classifier methodology: IEEE Transactions on Systems, Man, and Cybernetics, v. 21, no. 3, p. 660–674.

Swain, P. H., and Davis, S. M., 1978, Remote sensing: The quantitative approach: McGraw-Hill, New York, 369 p.

Switzer, P., 1965, A random set process in the plane with a Markovian property: Annals of Mathematical Statistics, v. 36, p. 1859–1863.

Switzer, P., and Green, A. A., 1984, Min/Max Autocorrelation factors for multivariate spatial imagery: Technical Report 6, Department of Statistics, Stanford University.

Tovey, N. K., and Krinsley, D. H., 1991, Mineralogical mapping of scanning electron micrographs: Sedimentary Geology, v. 75, p. 109–123.

Welch, J. R., and Salter, K. G., 1971, A context algorithm for pattern recognition and image interpretation: IEEE Transactions on Systems, Man, and Cybernetics, v. 1, p. 24–30.