ASSESSMENT OF UNSUPERVISED CLASSIFICATION TECHNIQUES FOR INTERTIDAL SEDIMENTS

Elsy Ibrahim[1], Stefanie Adam[1], Daphne van der Wal[2], Aaike De Wever[3], Koen Sabbe[3], Rodney Forster[4] and Jaak Monbaliu[1]

1. Katholieke Universiteit Leuven, Hydraulics Laboratory, Civil Engineering Department, Heverlee, Belgium; elsy.ibrahim(at)bwk.kuleuven.be
2. Netherlands Institute for Ecology, Yerzeke, the Netherlands; d.vanderwal(at)nioo.knaw.nl
3. Universiteit Gent, Laboratory for Protistology and Aquatic Ecology, Gent, Belgium; Koen.Sabbe(at)UGent.be
4. The Centre for Environment, Fisheries, and Aquaculture Science, Suffolk, The United Kingdom; r.m.forster(at)cefas.co.uk

ABSTRACT

Intertidal sediment stability is related to physical, biological, and chemical properties of sediments. To characterize these properties, extensive field work is required. Since field sampling on intertidal flats can be inefficient, unsupervised analysis of remotely sensed data offers an alternative. In this study, three unsupervised classification techniques were explored for the extraction of sediment characteristics from airborne hyperspectral data: *k*-means, the Gustafson-Kessel algorithm, and the mixture of Gaussians model. Simulated datasets based on real data were built and utilised to examine the suitability of the techniques for sediment characterization. The issues of intra-class variability, spectral dimensionality, and the choice of the number of clusters were investigated. The study showed that unsupervised classification methodologies can be used for sediment characterization, and that their performance depends on intra-class variability and feature selection. The mixture of Gaussians model was revealed to be the most suitable of the three techniques. Finally, a hyperspectral image of an intertidal study area was successfully classified in an unsupervised manner using the mixture of Gaussians technique.

INTRODUCTION

An intertidal zone in a marine aquatic environment is the area of the foreshore and seabed that is exposed to air at low tide and submerged at high tide generating important geochemical processes (1). These processes define sediment stability in relation to the physical, biological, and chemical properties of sediments. The erodibility of non-cohesive sediments differs from that of cohesive ones (2), and the presence of microphytobenthic algae and macrofaunal species can either stabilize or destabilize sediments (3). To investigate these processes, good field knowledge is required. Since accurate data collection on intertidal flats is often costly, inefficient, or even unattainable, remote sensing can be a fine and resourceful alternative.

It has been shown that hyperspectral imagery can be used to successfully classify intertidal sediments by respectively supervised and unsupervised classification techniques (4,5). Supervised classification requires training data that can be attained either from the field, from the image itself, or from spectral libraries. When dealing with intertidal sediments, there are no "absolute" representative or standard spectra in spectral libraries that can be used for training as these training data are normally site dependent. Furthermore, some spectra can be extracted from an image and used for training when the study area is known and experienced. Yet, when dealing with new areas, the choice of training spectra without field knowledge would include too many assumptions. Therefore, basic field data or field knowledge is a requirement when dealing with the supervised classification of intertidal flats.

On the other hand, unsupervised techniques extract clusters from an image on the basis of its reflectance values, without taking field measurements into account. Field knowledge only plays a role in the

interpretation of clusters. Therefore, it is of high interest to investigate these unsupervised techniques and understand their suitability for sediment characterization.

Unsupervised classification might not be a replacement to supervised classification. Yet, when a suitable unsupervised classification technique is utilized, it offers a better understanding of the results of supervised classification, as it gives an overview regarding the distribution of different spectral patterns. Unsupervised classification can also assist in the planning of efficient field campaigns on a study area, as it indicates the locations of areas with distinct spectral properties. Furthermore, it helps in the choice of training pixels from the image itself for supervised classification.

The aim of this paper is to investigate the suitability of three unsupervised classification techniques for sediment characterization. These techniques are: *k*-means, Gustafson-Kessel algorithm (GK), and the mixture of Gaussians model (MG). In practice, there is globally no "absolute best" unsupervised classification methodology, where the reliability of a result is based on the aim of the classification and its use (6). Furthermore, measuring the accuracy of this type of classification can be very complex. Normally, the clustering accuracy is either represented qualitatively or measured indirectly by specific indices that take the overall separability of the clusters into account. In some studies, the accuracy of each technique is assessed by comparing its results to a supervised classification of the same image (7,4). In this paper, the accuracy of the techniques was explored by means of simulated data sets based on a hyperspectral airborne image and the reflectance behaviour of sediments. The choice of using simulated or synthetic datasets was due to the following reasons. First, the supervised classification taken as the basis of the accuracy assessment could entail several uncertainties while the use of synthetic data allowed a total knowledge of the data sets. Second, the use of synthetic datasets resulted in flexibility and control over the data to conduct different experiments. Finally, the classification accuracy was directly determined as the properties of each data point in the synthetic datasets were known.

The investigation of the techniques by means of the synthetic data sets included the effect of intra-class variability, the effect of cluster volume, the effect of feature selection, and the detection of the appropriate number of clusters in a dataset.

STUDY AREA AND DATA SET

The study area was the IJzermonding in Belgium, located at the outlet of the IJzer river in the North Sea. It is a nature reserve that consists of dunes, marshes, and intertidal flats with a mean tidal range of 4 metres. On the 17th of June 2005, an image was acquired of the study area by means of the Airborne Hyperspectral Sensor (AHS) at low tide and in cloud-free conditions (3.4 m × 3.4 m pixel size). The quality of the signal of several bands was relatively low due to problems with the radiometric calibration of the sensor. The useful bands are shown in Table 1. For each of the modules: visible (VIS), near-infrared (NIR), and short wave-infrared (SWIR), the number of bands, the spectral range, and the full width at half maximum (FWHM) are listed. Although noisy, the SWIR band was retained for the analysis due to its sensitivity to moisture content. Furthermore, a field campaign was carried out on the flat where samples were analyzed for specific properties. Figure 1 shows an image of the study area and the sampling locations.

Table 1: AHS relevant bands

Module	Spectral range (nm)	Band	FWHM (nm)
VIS	445 – 689	1 to 8	27 - 30
NIR	718 – 1004	9 to 18	27 - 30
SWIR	1419 – 1824	19	159

Figure 1: The IJzermonding and sampling locations.

METHODOLOGY

Field data

The unsupervised classification techniques were assessed using simulated data sets. Although these data were synthetic, their properties were based on the spectral behaviour of three major sediment types found on the IJzermonding. Field sampling at 28 sites was performed on the inter-tidal flat at low tide four days before the over-flight. The coordinates of these sites were determined by means of a differential geographical positioning system (DGPS). To account for the variability within one pixel and the hyperspectral image geometric correction uncertainty (± 0.5 pixel size or 2 m), two or three replicates were sampled at each site, leading to 80 measurements. Furthermore, spectral measurements by means of the Analytical Spectral Device (ASD) spectrometer were acquired for each replicate. Sediment properties including moisture content (MC), mud content (MUC), organic matter content, and chlorophyll *a* (chl *a*) content were collected for sample analyses. MC was determined as weight lost by drying at 85°C till constant weight was attained. Grain size distribution, that leads to the quantification of the mud content, was measured in a Coulter counter analyzer that determined volume concentrations in several particle size classes. From grain size distribution, mud content was obtained as the percentage of particles smaller than 63 µm. Organic matter (OM) was determined by calculating the difference in weight between the sample before and after burning at 600°C for 2 hours. Chl *a* content was measured using high performance liquid chromatography (HPLC) (8) and quantified using a calibration with commercial standards.

In order to define sediment classes, thresholds were set on sediment properties based on field knowledge, field data analysis, and reflectance spectra of the AHS image. Texture groups were based on mud volume fractions: sandy when the MUC was less than 10% of the volume, mixed sediments when the MUC was between 10 % and 40 % and muddy when MUC was greater than 40%. Three classes of moisture content were defined as: dry when MC was less than 20%, wet when MC was between 20% and 40%, and saturated when MC was larger than 40%. For organic matter content, the classes were defined as: low for OM less than 4%, and high when OM was higher than 4%. For chl *a* content, the following three classes were defined: low when the content was less than 40 mg/m^2, medium when the content was between 40 and 80 mg/m^2, and high when the content was higher than 80 mg/m^2.

Sediment types

The field campaign included sampling on specific regions of the IJzermonding (Figure 1) dominated by three major sediment types:

- Type1 was saturated, muddy, with high chl *a*, and high OM content
- Type2 was wet, contained mixed sediments, medium chl *a*, and high OM content
- Type3 was dry, sandy, with low amounts of chl *a*, and low OM content.

After categorizing the field samples into the three types, their geographical positions were determined on the AHS image, and their corresponding spectra were extracted. The number of these representative spectra was between 20 and 50 (depending on the sediment type), thus too low to give a statistical overview of the sediment properties. Without using field data, it is a difficult task to choose all representative image spectra of each type due to the gradual changes in sediment properties on intertidal flats from one type to another. Therefore, spectral angles were calculated between the representative spectra and the remaining spectra of the image (9). The pixels closest in reflectance to the representative pixels were then also extracted leading to a total of 500 pixels per type. The mean ($\boldsymbol{\mu}_i$) and standard deviation ($\boldsymbol{\sigma}_i$) of sediment type (*i*) were calculated per spectral band (*j*) (μ_{ij} and σ_{ij}) (Figure 2).
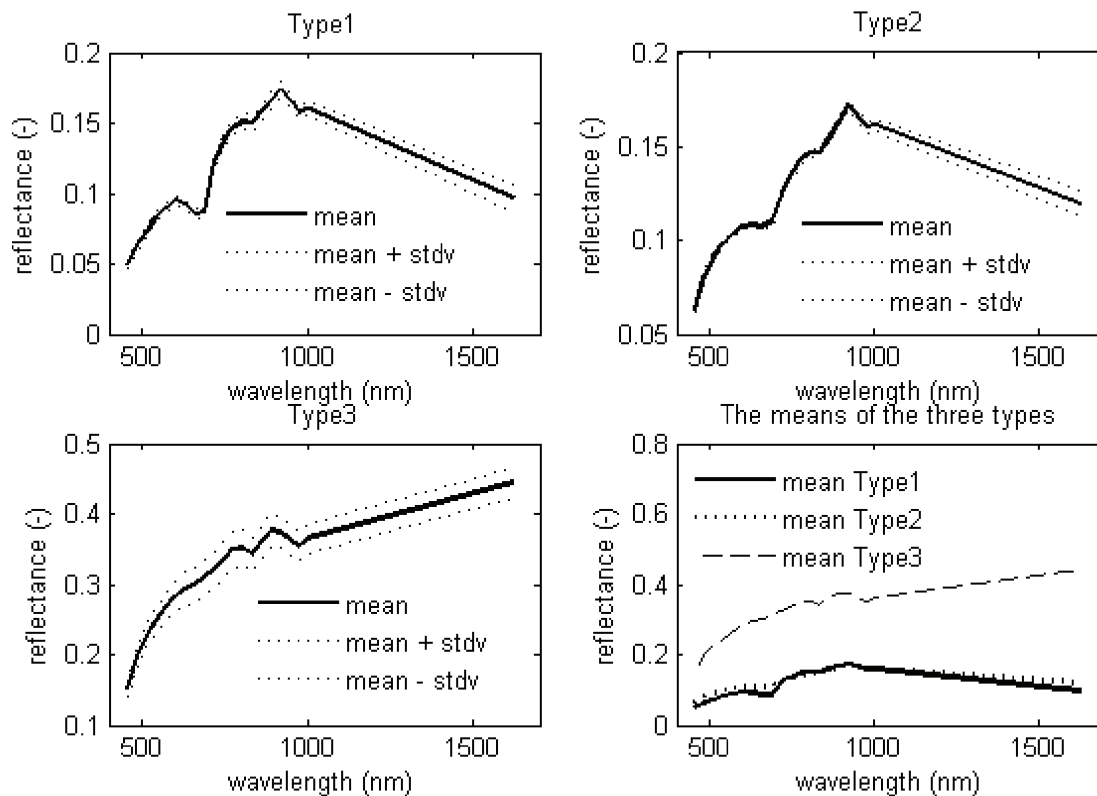

Figure 2: Mean spectra of each sediment type taken from AHS 2005.

Building synthetic data

In general, a class is any group of pixels sharing a common attribute with a known reference label. This general use of the term was utilized in this study, where a "class" is a synthetic group of pixels referring to a specific sediment type. Starting from statistical information obtained for each sediment type in the previous section, a synthetic class was generated for each of those types. A synthetic dataset was then produced as a combination of these classes. Spatial information was not taken into account to build the synthetic data. This means that a pixel was only described by its reflectance value per spectral band without including the effect of neighbouring pixels or location. Therefore, a synthetic dataset was an $N \times n$ matrix $\mathbf{X}$, where N was the number of pixels in the dataset and n the number of bands or features. $\mathbf{X}$ was built as a combination of a number of classes (*I*), C_i represented a specific class.

First, the effect of moisture content was considered. Reflectance decreases as moisture content increases, where reflectance at longer wavelengths shows more change than at shorter wavelengths (10,11). Yet, at a soil dependent value of MC referred to as "the cut-off thickness", reflectance increases as MC increases (12).

Second, the effect of mud content was investigated. The spectral response of increasing mud content in dry sediments consists of an increase in overall reflectance (13) and an increase in absorption at specific clay absorption features (14). However, these effects of clay absorption features were not observed on intertidal flats since muddy intertidal sediments have high moisture content, hampering the scattering of clay particles (15). On the other hand, a high correlation could be seen in the field data between mud content and moisture content (15). Therefore, the effect of mud content would be incorporated indirectly with the effect of moisture content, as the two properties are highly correlated and lead to general changes in the overall brightness of the spectra.

Third, the effect of chl a content was considered. Samples taken from the IJzermonding showed that MC and chl a content were positively correlated. The chl a absorption could be characterized by the ratio between minimal reflectance in the absorption feature and reflectance outside the absorption feature, where the ratio is negatively correlated to chl a content (16,17). When dealing the bands of the AHS image (Figure 3), band 8 of 689 nm represented the minimal reflectance in the absorption feature and band 9 of 718 nm represented the reflectance outside the absorption feature. A positive correlation was found between chl a content of field samples and the ratio between the AHS bands 8 and 9 of 689 nm and 718 nm, respectively. Therefore, the effect of chl a was introduced into the simulated data by varying the chl a dip through the band ratio discussed above. On the other hand, the increase in chl a content was accompanied by an increase in MC content. Due to the effect of the cut-off thickness in MC, the chl a ratio could either decrease or increase randomly over all the spectra.
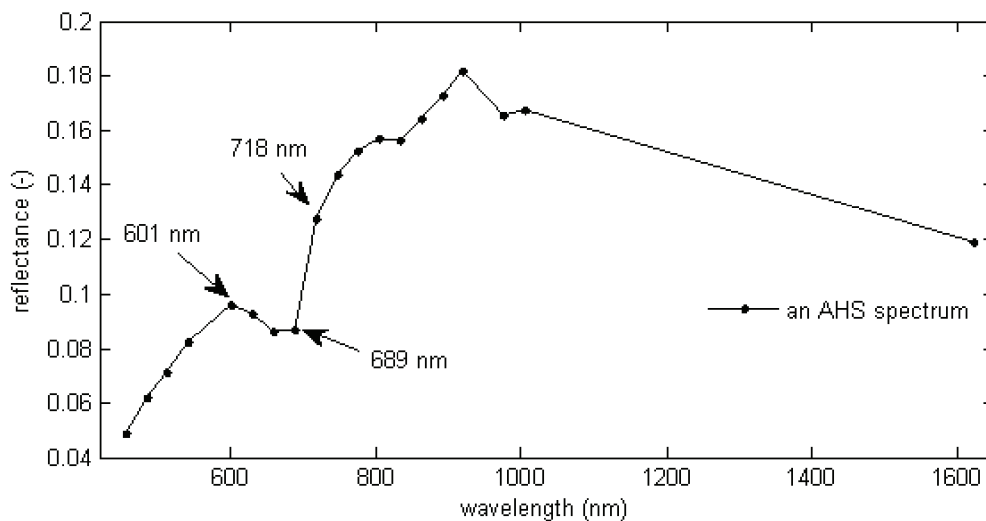

Figure 3: An example of a chl a absorption dip in AHS spectra.

Fourth, the effect of organic matter content was considered. The reflectance in the 400-2500 nm region decreases as organic matter content increases (18). The effect of organic matter on spectra is partially comparable to that of moisture content (15). Two studies indicated that the slopes or first derivatives at certain locations of a spectrum are important for organic matter content (19,20). This was tested with the field data obtained from the IJzermonding, but no significant correlations were found. A probable cause for low correlations is that the types of soil described in the mentioned literature are different from the soil in the IJzermonding. Therefore, since the major effect of organic matter is a general decrease or increase of the spectra, this is then already incorporated by the effect of moisture content, thus not added separately to the data. Finally, the effect of instrument noise was investigated by checking the variation in a homogeneous water area.

The above properties were then implemented in building the data. Each of the 19 bands of the AHS image was checked for normality by means of probability plots, where several bands could be consid-

ered following normal distributions. A probability plot is a technique that allows a graphical assessment of whether or not data follow a normal - or any other - distribution (21). The bands that did not seem to follow a normal distribution were tested for the following: Weibull, Rayleigh, and exponential distributions. Yet, these distributions were not representative as well. For calculation purposes, the assumption of normality was considered for all the bands.

By means of correlation analysis, it was seen that some bands were highly correlated, especially in similar parts of the spectrum; yet, some did not show any correlation to any other band in the image. Therefore, for some types, different parts of the spectrum needed to be considered separately while building the synthetic data. For type1 and type2, there was a high correlation between: (a) the first bands in VIS (bands 1 to 5), (b) the bands of the chl a dip (bands 6 to 8), and the bands of the NIR (bands 9 to 18). Finally, there was no high correlation between the SWIR band and any other band. The independency of the SWIR band is due to its high and overwhelming sensitivity to moisture content in a pixel. Therefore, a spectrum of type1 or type2 would be divided into four individually treated parts. As for type3, most of the bands showed degrees of correlation with each other. Therefore, all the bands were treated simultaneously. Therefore, a spectrum c_{ik}, of the class represented by the $N_i \times n$ matrix $\mathbf{c}_i$, is generated as follows:

$$\mathbf{c}_{ik} = \mathbf{\mu}_i + r \cdot \mathbf{\sigma}_i \cdot s \tag{1}$$

where $\mathbf{\mu}_i$ is a matrix containing the mean values per band of $\mathbf{c}_i$, $1 \leq i \leq I$, I is the number of classes, $1 \leq k \leq N_i$, $\mathbf{\sigma}_i$ is a matrix containing the type standard deviation per band, s is a positive number and r is a random value following a normal distribution (with a zero mean and a standard deviation of 1, it allows the variation of the spreading of spectra around the means spectrum).

For type1 and type2, instead of one random value, r, four different random values were used to create the reflectance spectrum of a pixel k, dividing the spectra into α, β, χ, and δ sub-matrices (Figure 4). Yet, throughout the process, the following conditions were to be fulfilled for each spectrum. First, to preserve the chl a absorption dip, the reflectance value at band 6 did not exceed the reflectance at band 5, and the reflectance value at band 8 did not exceed the reflectance at band 9. Second, due to the high moisture levels of two sediment types, there was higher absorption at the SWIR than at the NIR. Therefore, the reflectance value at band 19 was restricted from exceeding the reflectance value of band 18 of each spectrum. Figure 5, shows an example of the simulated spectra.
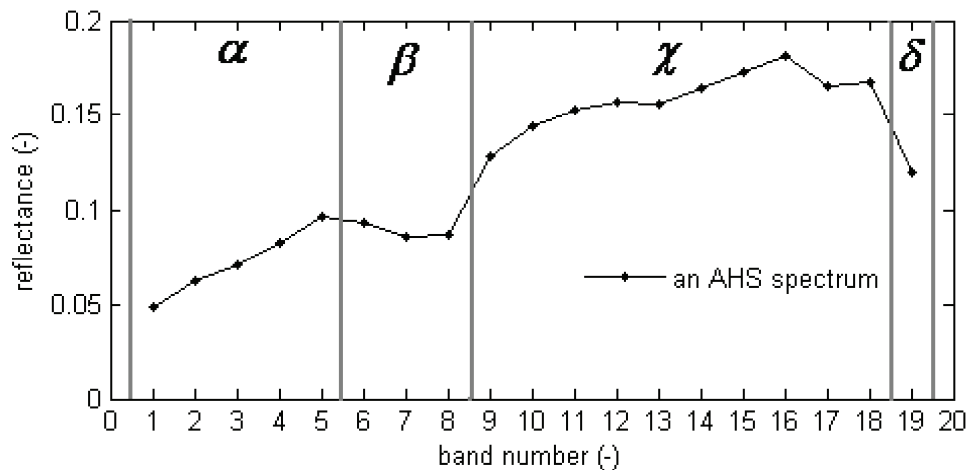

Figure 4: An example of the division of a spectrum into four sub-matrices when building the synthetic data of type1.
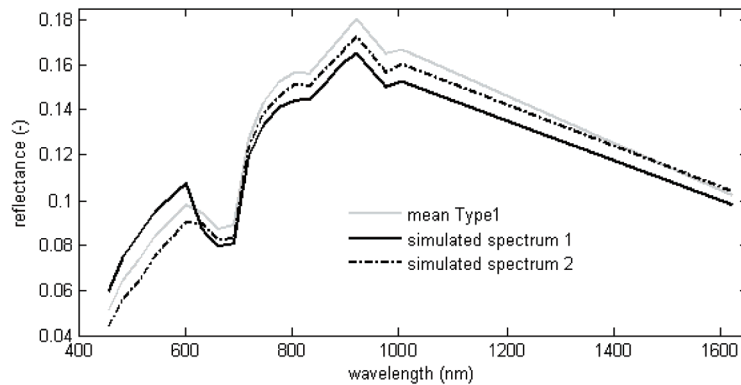
Figure 5: An example of two spectra built by the synthetic data for type1 with s=2.

Unsupervised classification techniques

The first step in unsupervised classification is finding structure in a collection of unlabeled data, or clustering. Clustering partitions an $N \times n$ data set **X** into a defined number of clusters I'. In this analysis, **X** is the studied image or dataset; N is the number of pixels, n is the spectral dimensionality (number of bands or channels), I' is a defined number of clusters to be produced from the data, and $\mathbf{C}'_i$ are the resulting clusters where $i \in [1, I']$. The three clustering techniques used for this study entail various properties, capabilities, and limitations, mainly based on the concept of dissimilarity or distance measures (6). The notion of distance measures between data points is an essential aspect of clustering as it determines the shape of clusters. These measures or norms are of various forms. One of the basic and popular forms of distance norms is:

$$\mathbf{d}_{M(\mathbf{C}'_{ik})} = \left((\mathbf{C}'_{ik} - v_i)^T \mathbf{A} (\mathbf{C}'_{ik} - v_i) \right) \tag{2}$$

v_i represents the cluster centres, $i \in [1, I']$, $k \in [1, N'_i]$, N'_i is the number of pixels in cluster i, and **A** is a positive symmetric definite matrix referred to as the norm-inducing matrix. Two popular distance norms can be used in the clustering techniques and are introduced by the matrix **A**. The Euclidean distance results when **A** is an identity matrix. This Euclidean distance measure is popular as it represents real physical distances between the variables. When **A** is the inverse of the covariance matrix of **X**, the Mahalanobis distance is resulted. This distance measure is commonly used for clustering as it takes correlations between different bands into account (22).

When fixed distance norms (non-adaptive) are used for finding clusters, i.e. **A** is the same for all the clusters, limitations arise as the techniques then impose defined geometrical structures. For example, the Euclidean norm-based methods would only be able to find *(hyper)*spherical clusters while the Mahalanobis norm-based methods would only find *(hyper)*ellipsoidal clusters. Even if those shapes of clusters are not present in the dataset at all, a fixed distance norm will only give those outcomes. Yet, when an adaptive norm is used, i.e. a different and suitable **A** for each cluster, more freedom in cluster geometry is possible (22).

The three clustering techniques chosen, *k*-means, GK and MG, were selected as they are of different specifications and types of clustering. *K*-means is a popular and simple technique with a non-adaptive distance norm, GK is a popular technique with an adaptive distance norm, and MG is a more sophisticated technique with an adaptive distance norm.

The first clustering technique considered was the hard clustering *k*-means. Exclusive or hard clustering techniques are the simplest of all the clustering methods. This makes them popular though their results may not always be reliable. *K*-means is one of those techniques, and it allocates each pixel to a cluster minimizing the within-cluster sum of squares (Euclidean distance). With the user defining the number of clusters, the algorithm randomly guesses the location of cluster centres (seed point). Then, data sets are grouped according to pixel location with respect to those centres. When the first clusters are formed, each centre finds the centroid of the points it owns and refers to it. These centroids would

then become the cluster centres and the procedure is repeated until the within-cluster sum of squares is minimized (6). From experience with *k*-means, it was seen that its performance was quite dependent on its initialization i.e. the selection of its seed points. Therefore, the algorithm was executed several times with random initialization and then the relatively "best" results were chosen. For the classification of the synthetic data, a partition index (SC) was utilized to select the run with the best partitioning, where the lowest value of SC showed the best clustering. SC describes the compactness and separability of clusters (23). A random selection of seed points is used to initialize the clustering, and the calculation stops when there is no more change in the cluster means. This hard clustering was based on the "Fuzzy Clustering and Data Analysis Toolbox" (23).

The second clustering technique considered was the Gustafson-Kessel algorithm, an example of overlapping clustering. Overlapping, soft, or fuzzy partition is a generalization of hard partitioning. The notion of fuzzy clustering actually refers to the fact that each pixel can belong to more than one cluster yet in a different degree of belonging, i.e. a value from 0 to 1 (24). For this study, each pixel was referred to the cluster with the highest degree of belonging. The Gustafson-Kessel algorithm (GK) detects clusters of different geometrical shapes in one data set by employing an adaptive Mahalanobis distance norm between a pixel and the cluster means (25,26). Random selection of seed points is used for initializing the clustering and the calculation stops when there was no change in the cluster means. Similarly to *k*-means, the algorithm was executed several times and the partition index, SC, was used to select the run with the best partitioning. This overlapping clustering is based on the "Fuzzy Clustering and Data Analysis Toolbox" (23).

The third clustering technique is the mixture of Gaussians (probabilistic clustering). The mixture of Gaussians (MG) is model-based where it is constituted of models suitable for describing each cluster. It attempts to optimize the fit between the data and the models. Clusters are considered as various Gaussian distributions centred according to their covariance structure (27). This type of data description by Gaussian distributions for clustering purposes was performed by means of the Mixture Modelling software, MIXMOD (28). To find the mixture components, the expectation-maximization algorithm (EM) was utilized to find the mixture components, where EM is a popular tool used to derive maximum likelihood estimates in mixture models (29). Depending on the parameterization of **A**, MIXMOD includes 28 Gaussian models that allow fixing cluster properties in various combinations (30). In order to choose the most descriptive of those models for the data, the Bayesian Information Criterion (BIC) was used. It is a commonly used criterion to compare models with different parameterization and/or components, since it effectively describes the data without the use of too many parameters (31). Therefore, a BIC value was calculated for each model under consideration, and the model with the smallest BIC value was chosen as the most suitable. Seed points were selected randomly to initialize the clustering. The calculation was stopped using a threshold of 10^{-4} for the relative change of the likelihood criterion.

Labelling clusters

Unsupervised classification techniques were applied on the synthetic data sets resulting in clusters without any reference labels. The process of labelling the resulting clusters can be straightforward if the clustering techniques resulted in high-quality classification, that is if all the pixels of each class were grouped in exclusive clusters, each having the label of its corresponding class. In many cases, the clustering might not be of high accuracy. Therefore the cluster that contained the spectra with the closest distance to a class mean was labelled according to that class.

Accuracy assessment

Since "a classification is not complete until its accuracy is assessed" (9), two associated measures of accuracy based on confusion matrices were used in this study: the percentage of correct classification of pixels followed by the kappa (κ) statistic. Kappa statistic was used to measure the difference between the clustering accuracy and the chance agreement between the classes and the clusters. The comparison between the synthetic sediment classes as the reference data on one hand and the "corresponding" resulting clusters on the other hand indicated the accuracy of the classification. From a confusion matrix, with rows containing the resulting clusters, and columns containing the classes of the synthetic data, the kappa (κ) statistic was calculated. When working with κ, the number of classes

should be equal to the number of clusters, resulting in a value between 0 and 1 for each classification, where 0 indicates that the clustering is no better than grouping the data by chance. However, the closer κ was to 1, the better the classification results are:

$$\kappa = \frac{N \sum_{i=1}^{I} y_{ii'} - \sum_{i=1}^{I} (y_i, y_{i'})}{N^2 - \sum_{i=1}^{I} (y_i, y_{i'})} \tag{3}$$

where I is the number of classes or clusters, $1 \le i \le I$, N is the total number of pixels in the synthetic dataset classified, i' refers to the cluster corresponding to class i, $y_{ii'}$ is the number of observations in row i and column i' in the confusion matrix, y_i is the number of observations in row i in the confusion matrix, and $y_{i'}$ is the number of observations in column i' in the confusion matrix.

Dimensionality reduction

When dealing with supervised classification, supervised and unsupervised, the issue of dimensionality reduction is essential. Even though hyperspectral data is useful for sediment characterization, its many contiguous bands carry too much redundancy in information. Dimensionality reduction reduces high dimensional data sets into a lower dimensional space while preserving the useful structure of the original data. It assumes that data points entail a certain structure in their locations. However, this does not necessarily indicate that hyperspectral data are not required for sediment characterization and can be replaced by multispectral data. An investigation regarding spectral resolutions and the effect of using fewer and broader bands could tackle this issue, but this was not in the scope of this paper.

Several studies showed that the dimensionality of data can be a critical factor for image classification. Each band in an image is a dimension, and when a successful reduction of the number of dimensions is performed, computational efficiency and classification accuracy can increase (e.g., 5,32,33). To reduce the number of bands or dimensions in imagery, several methods exist. The most popular techniques are of two groups: feature extraction and feature selection.

Feature extraction involves the transformation of the original dimensions into a new set of features. There are several popular and powerful feature extraction techniques (34,35). The drawback of this type of feature reduction for this study is that it transforms the data to "non-real" systems. Although in some cases the original features that contributed most to this transformation can be identified, the specific features that gave the most contribution to the transformation are unknown. Therefore, the bands that are the most informative about the data are not directly known.

Feature selection refers to the selection of a subset of relevant dimensions for a stable and robust representation of a dataset (e.g. 36,37). Feature selection not only improves computational efficiency but can also improve classification accuracy due to the removal of non-useful information, leading to a stable representation (38,39). Most of feature selection methodologies are in the context of supervised classification due to the issues of the unknown number of clusters in unsupervised classification and the difficulty in assessing the benefit or loss of information when specific features are selected. The AHS image of the IJzermonding was classified in a supervised manner with feature selection based on the methodology used for the classification of an intertidal flat in the Westerschelde estuary in the Netherlands (5). This supervised feature selection triggered an investigation of the effect of simple selection techniques on the unsupervised classification accuracy.

In this paper, experiments were carried out to investigate the effect of simple feature selections that involve bands from major parts of the spectrum. In most multispectral data used for sediment classification, the few bands available are a combination of all or some bands of the following parts of the spectrum: blue (~400 nm – ~500 nm), green (~500 nm – ~600 nm), red (~600 nm – ~700 nm), NIR (~700nm – ~1 400 nm) and SWIR (~1 400 nm -~3 000 nm). Therefore, it was essential to test, using the synthetic data based on the AHS image, the effect of choosing these specific bands on the classification accuracy.

The number of clusters

One of the basic steps of cluster analysis is to specify the number of clusters that the data is to be grouped into. Yet, this is not a simple task. There exist several estimation methods that investigate the number of clusters in a data set (6). These are quantitative measures that enable the evaluation of the clustering results. This is referred to as the "validation" of clusters. There are different approaches to clustering validity. The approach considered in this paper is based on "relative criteria". This leads to an evaluation of a certain clustering structure by comparing it to other clustering schemes of the same data but with different number of clusters (40). The aspects taken into account for this comparison are compactness and separation. Compactness refers to the relationship between data within each cluster. For a better clustering, the data should be as close to each other as possible. Separation, describes the inter-class variability between the clusters, where for better clustering, the separation of clusters should be high. Therefore, a good clustering is represented by compacted and separate clusters.

For k-means and GK, three validity measures were used simultaneously to give a suitable overview of the best number of clusters. The first two measures were the partition index (SC) utilized earlier and the separation index (S). They represent the ratio of the sum of compactness and separation of the clusters (23,41). The third measure was the Xie and Beni's Index (XB), a robust and stable measure that measures the compactness and separation of the clusters (42) where the compactness of a cluster is defined as the distance between each sample to the cluster centre (weighted by its membership value in case of fuzzy clustering). The separation of partitions is defined as the minimum distance between class centres (41). As the three measures are minimized, the better is the choice of the number of clusters (23). In case of no correspondence between the three measures, the results of SC and S are considered more useful.

For MG, the normalized entropy criterion (NEC) was used to determine the well separated clusters. It measures the ability of a mixture model to provide well-separated clusters. It is derived from a relation showing the differences between the maximum likelihood approach and the classification maximum likelihood approach to the mixture problem (43,44). A minimization of the NEC would refer to a suitable number of clusters.

EXPERIMENTS AND RESULTS

Intra-class variability

Experiments

Variability in a dataset can be represented by overall, inter-class, and intra-class variability. Intra-class variability affects the separability of clusters in a dataset and is therefore essential in classification procedures (9). It is expected that an increase in intra-class variability could lead to a decrease in classification accuracy. Therefore, in order to determine the applicability of the techniques to intertidal sediment characterization, intra-class variability was altered by the standard deviation used to build the classes of synthetic data or by modifying the number of pixels per class.

The first experiment carried out aimed to investigate the effect of changing the variability in the data on the clustering techniques. This was applied by varying the standard deviations used to build each class. A standard deviation would be $s \times \boldsymbol{\sigma}_i$, where $1 \leq i \leq I$, $0 \leq s \leq 5$, and $\boldsymbol{\sigma}_i$ is the standard deviation used to build $\mathbf{C}_i$. Synthetic datasets with different class standard deviations and different class combinations were built accordingly. The analysis was limited to $s \leq 5$ as higher standard deviations resulted in spectra that were much beyond the spectral responses in the AHS image. When a dataset was a combination of type1, type2, and type3, it, therefore, contained three classes of the three sediment types. For these experiments, the classes contained the same number of pixels, $N_i = 500$, $1 \leq i \leq I$ with a varying standard deviation. Each of the datasets was then clustered into three clusters. Similarly, data with two-class combinations were built with $N_i = 500$ and $I = I' = 2$ with different standard deviations.

The second experiment aimed at investigating the effect of relative cluster volumes in the data. Intra-class variability, controlled by the number of pixels per class, gives the volume property of clusters. The synthetic datasets were built in the following manner. First, two sediment types were used to build two classes, $\mathbf{C}_1$ and $\mathbf{C}_2$, with an equal number of pixels, $N_1 = N_2 = 500$, and $s = 1$. While the number of pixels of the first class (N_1) was fixed to 500 pixels, the number of pixels in the other (N_2) varied in steps of 100 pixels. The data were then clustered.

Results and discussion

The three investigated unsupervised classification techniques were mostly successful in distinguishing between the different sediment types. However, the performance of each technique differed with changing intra-class variability as well as the sediment types being classified. When investigating the effect of intra-class variability by changing the standard deviation, the three techniques resulted in the accuracy values shown in Figure 6.
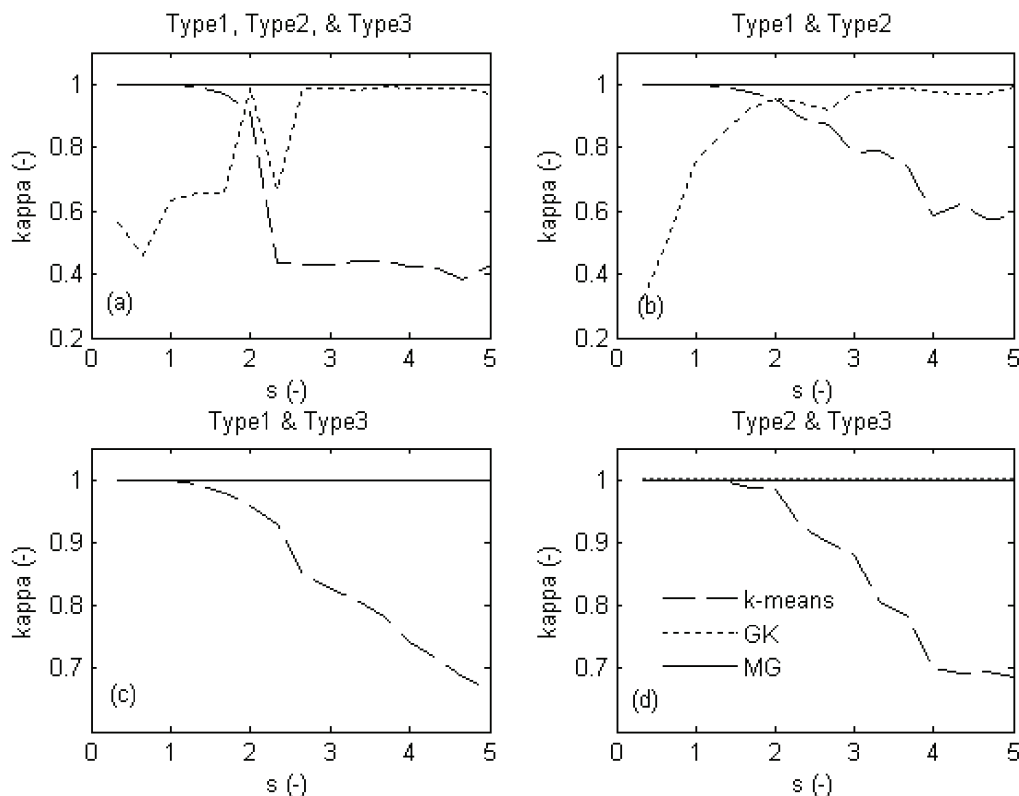

Figure 6: Kappa accuracy with respect to standard deviation of (a) the three types (b) type1 and type2 (c) type2 and type3 (d) type1 and type3.

The following could be distinguished among the techniques:

(a) *k*-means showed the expected behaviour of decrease in accuracy as intra-class variability increased

(b) GK showed a poor performance when the intra-class variability was too low. When the standard deviation was higher than two times $\boldsymbol{\sigma}_i$, GK was able to classify the three types with high accuracy. A possible reason for this behaviour is that GK utilizes the Mahalanobis distance, and this requires the inversion of the covariance matrix. Besides the fact that most of the consecutive bands (columns of the data matrix) are highly correlated, the data was also highly correlated when low intra-class variability was involved. A result of this alignment is a covariance matrix whose inverse was either difficult or impossible to calculate. This might lead to low accuracy in classification by GK when low intra-class variability was involved

(c) MG resulted in a stable high classification accuracy of the spectra throughout the range of standard deviations considered. Therefore, despite the fact that the three techniques were suitable for sediment classification, MG was revealed to be the most accurate of all and the

least sensitive to the underlying standard deviation used to build the spectral classes (Figure 6.a).

For more detail, the classification accuracies with respect to different pairwise class combinations were compared, and the following could be deduced:

(a) MG has the best performance in all cases in the range of $0 < s \le 5$ (Figure 6 b,c,d)

(b) GK resulted in high classification accuracy when handling type1 and type3, and type2 and type3 ($\kappa > 0.9$)

(c) and k-means was highly sensitive to class standard deviations (Figure 6.b,c,d).

Therefore, an unsupervised classification method that would best distinguish the three sediment types with high accuracy for all through the range of standard deviations was revealed to be the MG technique.

Figure 7: Kappa accuracy of the different techniques as one class contained 500 fixed pixels and the other varied in volume.

The second experiment aimed at investigating the dependency of the techniques on relative class/cluster volume. The results can be summarized as follows (Figure 7):

a) there was no major reduction in the classification quality of k-means as the relative class/cluster volumes varied

b) the performance of GK was more affected by the relative cluster volume, especially between type2 and type3; and

c) MG still resulted in high classification accuracy throughout the experiment.

Figure 6 shows the results in pairs as one cluster was set to 500 pixels while the other cluster varied in volume. GK showed the lowest classification accuracy of the three techniques. When discriminating between type1 and type2, GK classification accuracy was already low (kappa=0.6) when dealing with the same cluster volume. With increasing difference in cluster volume, the classification accuracy did not decrease further. Yet, for discriminating between type1 and type3 or type2 and type3, the classification accuracy of GK decreased dramatically. This can be due to the fact that in some cases, GK requires prior information of the shape and volume of each cluster (45). Else, it can detect only the classes with the same volume. Therefore, this might be a reason for this lower classification accuracy.

Feature selection

Based on training data

A successful classification of intertidal sediments by means of the sequential floating forward selection algorithm (SFFS) to perform the feature selection and the linear discriminant analysis (LDA) for classification was successfully carried out on an intertidal flat on the Westerschlede (5). This procedure was applied on the AHS image and the bands selected for the classification of each individual sediment property are listed in Table 2.

Table 2: list of bands selected and used for the supervised classification

	Selected bands used in supervised classification (central wavelength)
Moisture content	455 nm (band 1), 513 nm (band 3), 918 nm (band 16), 1622 nm (band 19)
Chlorophyll *a* content	689 nm (band 8), 718 nm (band 9), 774 nm (band 11)
Mud content	689 nm (band8), 833 nm (band 13), 1004 nm (band 18)

Synthetic datasets were built where each set contained three classes of the three sediment types, the same number of pixels $N_i = 500$, the same number of bands (*n* = 19 bands), and a specific standard deviation. From these datasets, new datasets were formed by selecting the bands specified in Table 2 for each individual property. Then, each of the datasets was classified into three classes.

Based on four band selection

Adjacent bands in the AHS image are, as in hyperspectral imagery in general, highly correlated. Therefore, this experiment aimed at testing the effect of selecting a few bands from major parts of the spectrum (blue, green, red, NIR, and SWIR) on the clustering accuracy.

The synthetic datasets for this experiment were built with $I = I' = 3$, $N_i = 500$, *n* = 19, and 0<*s*≤ 5. From these datasets, new sets were formed by selecting all the possible combinations of bands such that one band was selected randomly from each of the blue, green, and NIR parts of the spectrum. Therefore, each of these datasets contained four bands. Finally, the whole procedure is repeated with the SWIR band added to each of the four band combinations.

Based on five band selection

This band selection is based on the previous "four band selection" , yet in the band selection, the chl a absorption dip, with its lowest point around band 8 and bounded by band 5 and band 9 (Figure 3), was also taken into account.

The synthetic datasets for this experiment were built with $I = I' = 3$, $N_i = 500$, *n* = 19, and 0<*s*≤5. From these datasets, new sets were formed by selecting all the possible combinations of bands such that: one band is selected randomly from each of the blue, green, and NIR parts of the spectrum. To represent the chl *a* dip, two specific fixed bands were selected: band 8 and band 5. There was no need to fix band 9, the other bound of the chl *a* dip, since a random NIR band was chosen instead. Therefore, each of these images contained five bands. Finally, the whole procedure is repeated with the SWIR band added to each of the five bands.

Results and discussion

The classification accuracy of the clustering techniques that resulted from using the bands selected by the SFFS (Table 2) is shown in Figure 8. This figure shows a comparison between the accuracy resulting from using these selected bands and using all the 19 bands of the AHS image. K-means did not result in a better classification when using the bands selected by the SFFS than using all original bands. On the other hand, the accuracy of GK increased considerably, and MG remained consistent with high accuracy but slightly lower kappa values when the MC bands were used.
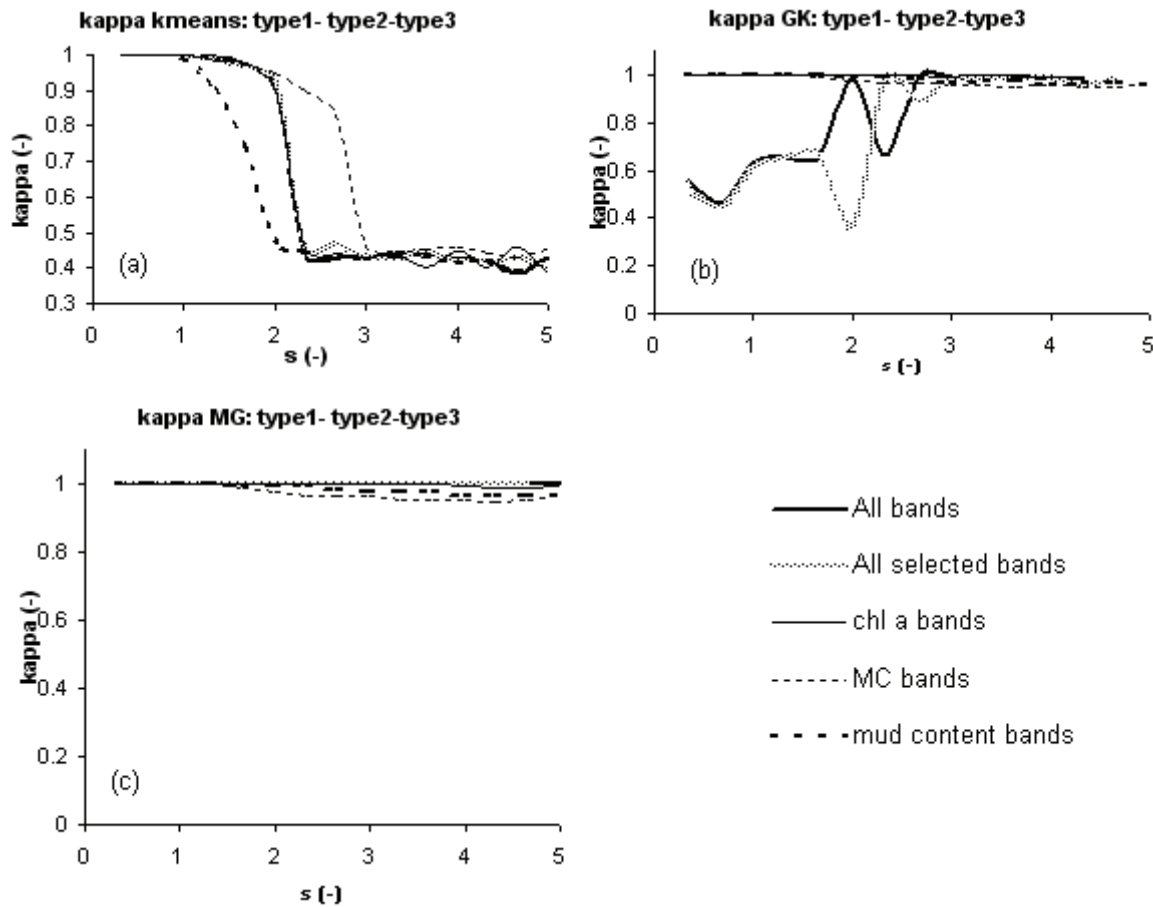
Figure 8: Kappa accuracy for (a) k-means, (b) GK and (c) MG using the bands selected by the SFFS.

Table 3 shows an overview of the resulting classification accuracy of each of the techniques with the two different feature selection methodologies, as the standard deviation, $s \times \sigma_i$, increases. Since there were various combinations of bands taken into account for each case of feature selection, the table shows the minimum (min), maximum (max), and mean values obtained for each feature selection. These results show that, generally, simple feature selections from the VIS and NIR (VNIR) or VNIR + SWIR can be successfully used with the three clustering techniques to classify the sediments. This is due to the fact that these selections resulted in classification accuracies comparable to using all the bands of the AHS image. In some cases, the classification accuracy was even improved. Table 4 shows the results of using all the bands at the specific values of standard deviations represented by *s*. The performance of k-means can be summarized as follows: (a) the use of these fewer band combinations resulted in similar results to using all the bands. (b) In all band combinations, a decrease in classification accuracy was noted as intra-class variability increased. The performance of GK was generally better when fewer bands were considered. When the SWIR band was used, the classification accuracy was highest. Finally, MG could still characterize the three types with very high accuracy. The least accuracy of MG was mainly noted when the four band selection was used at high *s* values.

Table 3: An overview of the classification of the three techniques with feature selection

kappa	features	standard deviation used to build the data (represented by *s*)														
		1			2			3			4			5		
		min	max	mean	min	max	mean	min	max	mean	min	max	mean	min	max	mean
k- means	4 bands	0.99	1	0.99	0.86	0.92	0.9	0.46	0.47	0.46	0.42	0.43	0.43	0.41	0.42	0.41
	4 bands + SWIR	0.99	1	0.99	0.94	0.97	0.96	0.45	0.47	0.46	0.65	0.66	0.66	0.42	0.43	0.42
	5 bands	0.99	0.99	0.99	0.88	0.92	0.9	0.46	0.46	0.46	0.44	0.44	0.44	0.42	0.43	0.43
	5 bands + SWIR	1	1	1	0.95	0.96	0.95	0.43	0.43	0.43	0.41	0.43	0.42	0.41	0.41	0.41
GK	4 bands	1	1	1	0.95	1	0.99	0.94	1	0.99	0.94	1	0.98	0.92	1	0.98
	4 bands + SWIR	1	1	1	0.98	1	1	0.96	1	0.99	1	1	1	0.94	1	0.98
	5 bands	0.45	1	0.95	1	1	1	0.73	1	1	1	1	1	1	1	1
	5 bands + SWIR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MG	4 bands	1	1	1	1	1	1	0.61	1	0.99	0.93	1	0.98	0.57	1	0.97
	4 bands + SWIR	1	1	1	0.98	1	1	0.95	1	0.99	1	1	1	0.6	1	0.98
	5 bands	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5 bands + SWIR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 4: Classification results using all the AHS bands.

Kappa	standard deviation (represented by s)				
	1	2	3	4	5
k-means	1	0.95	0.45	0.44	0.41
GK	0.47	0.97	0.99	1	0.99
MG	1	1	1	1	1

Choosing the number of clusters

The synthetic data was used to test the different validity measures to obtain estimates regarding a suitable number of clusters. When the synthetic data was used, the number of clusters found in the data was known. In the following experiment, tests were carried out to check the ability of the validity measures to retrieve the number of clusters from the data. Synthetic datasets of three classes, corresponding to the three types, were used with a standard deviation equals to $2 \times \sigma_i$.

For the case of *k*-means, the three validity measures did not result in a good estimate of the number of clusters. SC was monotonically decreasing, and XB and S reached a local minimum at 5 clusters (Figure 9). For GK, the validity measures SC and S resulted in a reasonable estimate of the number of clusters, as they resulted in the minimum values at two and three clusters (Figure 10). In the case of MG, the NEC was able to retrieve that two and three clusters were suitable for partitioning the data (Figure 11), resulting in a good estimate of the number of clusters. The results of NEC were further investigated. As it resulted in more than one value for the suitable number of clusters, the smallest value was chosen, therefore, two clusters. When the datasets were partitioned into two clusters, it was seen that the spectra of type1 and type2 (the two types that are the most similar) were merged together in the first cluster, and the spectra of type3 constituted the second cluster.
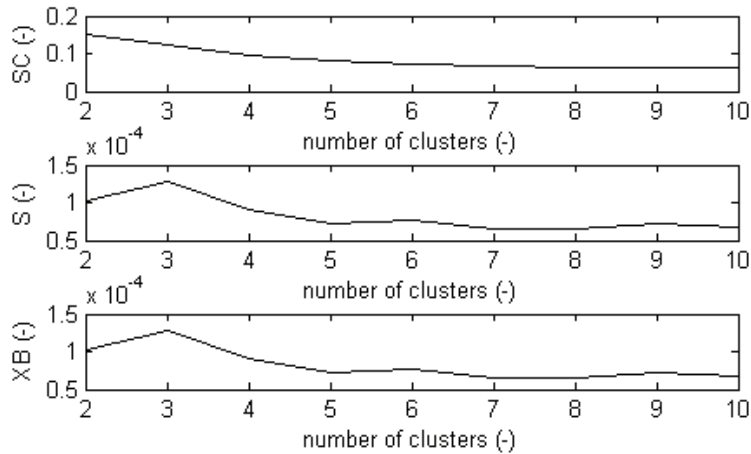
Figure 9: Validity measures of different cluster numbers for k-means with deviation = $2 \times \sigma_i$
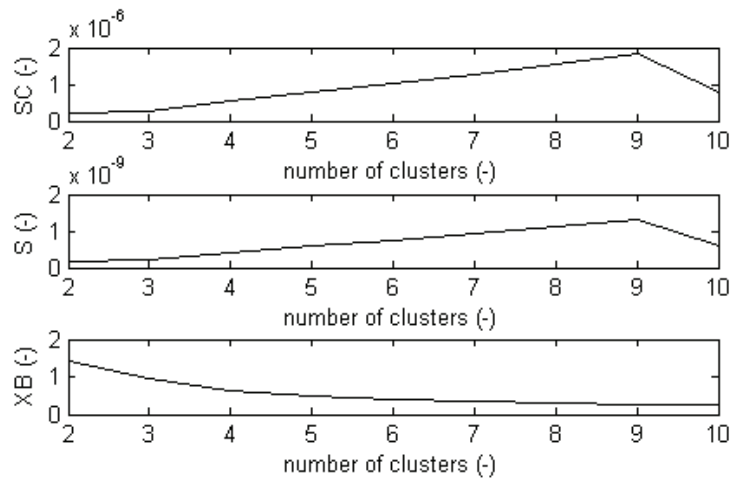

Figure 10: Validity measures of different cluster numbers for GK with deviation = $2 \times \sigma_i$
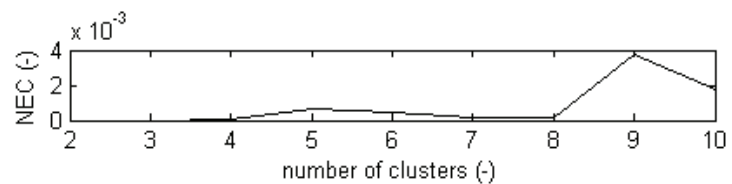

Figure 11: Validity measure of different cluster numbers for MG with deviation = $2 \times \sigma_i$

Applying the techniques on the real image

On real spectra of the three types

This section is used to validate the results obtained by the synthetic data. Instead of creating an image of synthetic spectra, a synthetic image was built by means of the "real" spectra, extracted from the AHS image, as representative spectra for each type. When the images of "real" spectra were built, all the bands or selected bands (as explained in the feature selection section) were used. Table 5 shows the classification accuracy resulting from this experiment. These results are compared to the results obtained by the synthetic data at $s = 1$. The results are seen to be comparable to the experiments using the synthetic data. Only GK did not show similar results. This can be due to the assumption of normality considered for all the bands of the AHS image. As GK previously showed sensitivity to low intra-class variability, such a modification can trigger a large variation in the results of GK at this considered low class variability ($s = 1$).

Table 5: mean classification accuracy with respect to selected features

features used	kappa (κ)		
	k-means	GK	MG
All bands	1	>0.8	1
4 random bands (VNIR)	0.97	0.67	0.98
4 random bands (VNIR) + SWIR	0.99	0.68	0.94
5 bands including chl *a* dip	1	0.69	1
5 bands including chl *a* dip + SWIR	1	0.78	0.99

On the whole AHS image

The assessment of the clustering techniques showed that the mixture of Gaussians outperforms GK and *k*-means. Therefore, it was used to classify the AHS image of the IJzermonding. This illustrates how the method can be used on real imagery of intertidal flats. During the assessment of the clustering techniques, the number of clusters was known as the number of types per classes in a synthetic dataset. When clustering real data, the number of components that is best described by the data is unknown. Yet, few steps are needed in order to reach the final number of clusters.

In the first step, the number of clusters should vary between 1 and the smallest integer larger than $N^{0.3}$ where N is the number of pixels in an image (46). For that purpose, the normalized entropy criterion, tested above, was used to choose the most suitable number of clusters for the AHS image (28). Therefore, NEC uncovered the most suitable number of clusters *(c)*, where $1 < c < N^{0.3}$.

The second step was mainly the treatment of each of the resulting clusters individually as in step1. The whole procedure of steps 1 and 2 was repeated until, based on field knowledge, a decision is made to stop the clustering. This resulted in a hierarchical clustering based on the NEC criterion.

Third, after obtaining all the clusters, separability measures and field knowledge were used to group the clusters that were least separable, leading to a lower but meaningful number of clusters. The separability of the clusters was measured by means of the Jeffries-Matusita (JM) separability measure. It calculates a mean distance for each cluster to all other clusters by averaging the distance between a specific cluster and the rest of the clusters (47,48). It was computed in a pair-wise manner - values ranged from 0 to 2.0 - and indicated how well the clusters were statistically separate. Values greater than 1.9 indicated very good separability for a pair of clusters. When the final clusters are defined, field knowledge based on field data and the mean spectrum of each partition are used to identify and label each cluster.

The AHS image was classified by MG into fourteen final clusters. The Jeffries-Matusita (JM) separability measure was calculated for each pair of clusters, and it ranged between 1.8 and 2, referring to high general separability. Using the JM values, a merging of clusters with the lowest JM values was carried out. The final JM values ranged from 1.95 to 2.0. Figure 12 and Figure 13 show the final clusters of the AHS image and spectra respectively.

The same procedure was carried out on the AHS image with six bands. The band selection was based on the previously explained simple five band feature selection from the VNIR part of the spectrum plus the SWIR band. The image was classified into sixteen final clusters. The Jeffries-Matusita (JM) separability measure was calculated for each pair of clusters, and it ranged between 1.4 and 2.0. Using the JM values, a merging of clusters with the lowest JM values was carried out. The final JM values ranged from 1.87 to 2.0. Figure 14 and Figure 15 show the final clusters of the AHS image and spectra respectively.

From field knowledge, the clustering techniques were able to reveal various essential parts of the intertidal flat, especially, the saturated muddy area. The results using all 19 bands or 6 bands resulted in similar clusters. The separability of the clusters used indicated slightly higher separability of the clusters when all the bands were used.
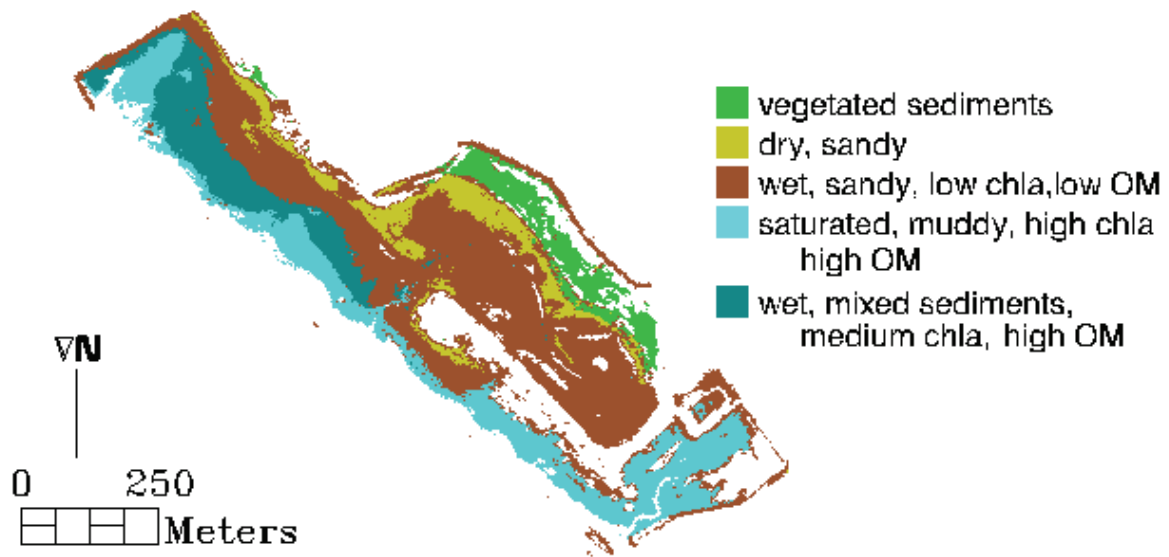
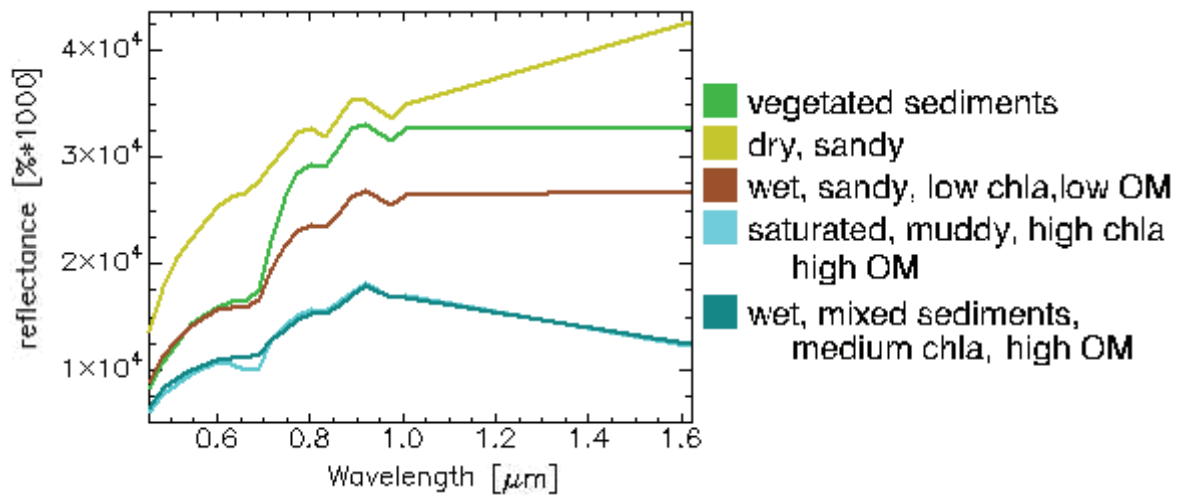*Figure 12: Classification of AHS 2005 by mixture of Gaussians with 19 bands.*

Figure 13: Mean spectra of each cluster resulting by the mixture of Gaussians clustering with 19 bands.
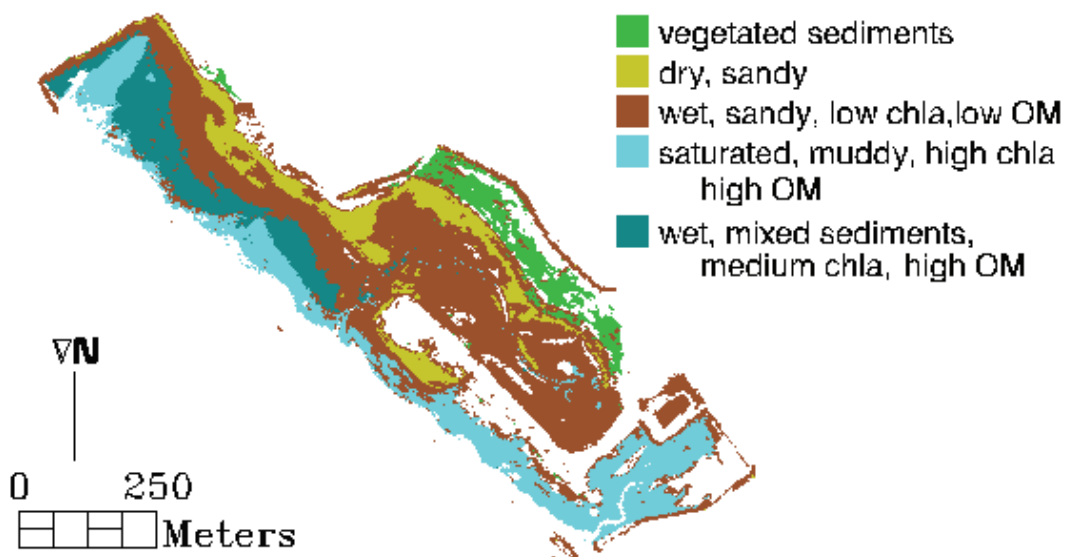

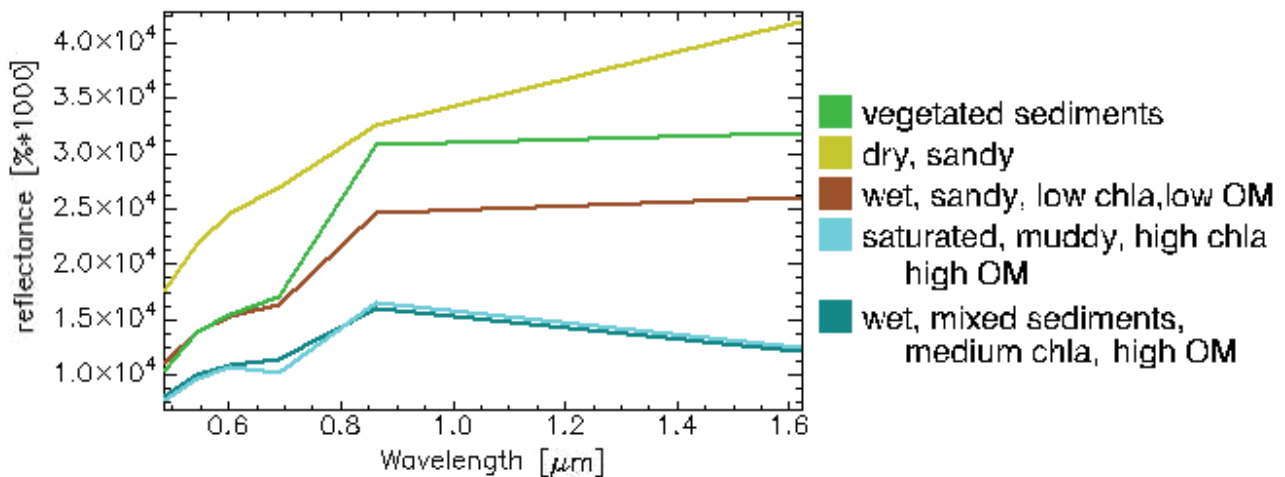*Figure 14: Classification of AHS 2005 by mixture of Gaussians and 6 bands.*

Figure 15: Mean spectra of each cluster resulting by the mixture of Gaussians clustering with six bands.

CONCLUSIONS

This investigation of unsupervised classification techniques in this paper was carried by means of synthetic data based on sediment properties and real spectra. An overview of the reliability of three clustering techniques for classifying sediments was given. The study included the effects of intra-class variability, feature selection, and the choice of the number of clusters. The three types chosen in this study were the most essential types of the intertidal flat under study.

The classification techniques considered in this paper were capable and suitable for discriminating three chosen sediment types successfully using simulated datasets. The techniques were dependent on variability in imagery. Comparing the overall performance of MG, GK, and k-means, the powerful MG approach showed a good performance in all experiments. It was seen that k-means performed well with and without feature selection. On the other hand, GK performed relatively better when feature selection was involved than when all the bands were used. Finally, MG was shown to be robust and with very high classification accuracy with or without feature selection and throughout a wide range of intra-class variability.

The classification techniques were also applied to read data sets. First, they were applied to groups of real spectra chosen from the image, representing the three sediment types. In this case k-means and MG outperformed GK, with and without feature selection. Finally, the most suitable technique, MG, was performed on the real AHS image. First, the clustering was done using all the nineteen bands of the image. Then, it was carried out using 6 chosen bands. Field data was used to identify the properties of corresponding clusters. The results using all bands or six lead to similar results that corresponded to field knowledge. This illustrated the ability of MG for unsupervised classification of sediments on the IJzermonding. Furthermore, it showed that when a simple feature selection was involved, MG still attained a suitable clustering of the data.

The simplification of the simulated data by assuming normality for the data was an issue to keep in mind with the interpretation of results. The experiments were carried out on both simulated spectra and real spectra chosen from the image. A comparison between these experiments showed similar results for MG and k-means, yet not for GK. GK was able to cluster the sediments in a higher accuracy when dealing with real data with low intra-class variability. This can be due to the greater variation in the real data with low intra-class variability compared to simulated data. As explained earlier, an obstacle for GK was the high correlation between the data at low intra-class variability as explained earlier.

For this study, three essential and representative sediment types were chosen for the IJzermonding. When dealing other study sites, more or other sediment types can be added to the simulated data if required. The user mainly needs to know the effect of a type's properties on sediment spectra. Furthermore, the user needs to keep in mind to check pairwise clustering as well as the clustering of all the simulated types simultaneously (as shown in Figure 6). This would explain where the clustering

difficulty or simplicity occurs. Therefore, the assessment methodology can be rebuilt for other sites, and the application of the clustering mixture of Gaussians approach was shown to be successful and powerful when utilized for both simulated and real data.

ACKNOWLEDGEMENTS

REFERENCES

1   Silva J, R Santos, M Calleja & C Duarte, 2005. Submerged versus air-exposed intertidal macrophyte productivity: From physiological to community-level assessments. Journal of Experimental Marine Biology and Ecology, 31: 87-95

2   Mitchener H & H Torfs, 1996. Erosion of mud/sand mixtures. Coastal Engineering, 29: 1-25

3   Austen I, T J Andersen & K Edelvang, 1999. The influence of benthic diatoms and invertebrates on the erodability of an intertidal mudflat, the Danish Wadden Sea. Estuarine, Coastal, and Shelf Science, 49: 99-111

4   Adam S, I Vitse, C Johannsen & J Monbaliu, 2006. Sediment type unsupervised classification of the Molenplaat, Westerschelde estuary, The Netherlands. EARSeL eProceedings, 5: 146-160

5   Deronde B, P Kempeneers & R Forster, 2006. Imaging spectroscopy as a tool to study sediment characteristics on a tidal sandbank in the Westerschelde. Estuarine, Coastal and Shelf Science, 69: 580-590

6   Everitt B S, 2001. Cluster Analysis (Edward Arnold, London) 237 pp.

7   Duda T & M Canty, 2002. Unsupervised classification of satellite imagery: choosing a good algorithm. International Journal of Remote Sensing, 23: 2193–2212

8   Wright S, S Jeffrey, R Mantoura, C Llewellyn, T Bjornland, D Repeat & N Welshmeyer, 1991. Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. Marine Ecology Progress Series, 77: 183-196

9   Lillesand,T M & R W Kiefer, 2000. Remote Sensing and Image Interpretation (John Wiley & Sons) 724 pp.

10 Weidong L, F Baret, G Xingfa, T Qingxi, Z Lanfen & Z Bing, 2002. Relating soil surface moisture to reflectance. Remote Sensing of Environment, 81: 238-246

11 Muller E & H Décamps, 2000. Modelling soil moisture-reflectance. Remote Sensing of Environment, 76: 173-180

12 Neema D, A Shah & A Patel, 1987. A statistical model for light reflection and penetration through sand, International Journal of Remote Sensing, 8: 1209-1217

13 Baumgardner M F, 1985. Reflectance properties of soils, Advances in Agronomy, 38: 1-44

14 Hunt G, 1977. Spectral signatures of particulate minerals in the visible and near-infrared. Journal of Geophysical Research, 95: 12653-12680

15 Adam S, 2009. Bio-physical characterization of indicators of sediment stability of mudflats using remote sensing. PhD thesis, Katholieke Universiteit Leuven, 198 pp.

16 Carrère V, 2003. Mapping microphytobenthos in the intertidal zone of Northern France using high spectral resolution field and airborne data, In: 3rd EARSeL Workshop on Imaging Spectroscopy, edited by M Habermeyer, A Müller, & S. Holzwarth (EARSeL, Herrsching) 395-409

17 Murphy R, T Tolhurst, M Chapman & A Underwood, 2005. Estimation of surface chlorophyll a on an emersed mudflat using field spectrometry: accuracy of ratios and derivative-based approaches. International Journal of Remote Sensing, 26: 1835-1859

18 Kooistra L, J Wanders, G F Epema, R S E W Leuven, R Wehrens, & S M C Buydens, 2003. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. Analytica Chimica Acta, 484: 189 - 200

19 Ben-Dor E, Y Inbar, & Y Chen, 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infra-red region (400-2500 nm) during a controlled decomposition process. Remote Sensing of Environment, 61: 1-15

20 Kooistra L, R Wehrens, R S E W Leuven, & L M C Buydens, 2001. Possibilities of visible-near infrared spectroscopy for the assessment of soil contamination in river floodplains. Analytica Chimica Acta, 446: 97-105

21 Chambers J, W Cleveland, B Kleiner & P Tukey, 1983. Graphical Models for Data Analysis (Wadsworth-Duxbury Press) 395 pp.

22 Kim D, K Lee & D Lee, 2005. Detecting clusters of different geometrical shapes in microarray gene expression data. Bioinformatics Advance Access, Bioinformatics, 21: 1927-1934

23 Balasko B, J Abonyi & B Feil, 2004. Fuzzy Clustering and Data Analysis Toolbox: For Use with Matlab. University of Veszprem, Hungary. http://www.fmt.vein.hu/softcomp/fclusttoolbox/ (last date accessed: 15 December 2008)

24 Balzano W & M R Del Sorbo, 2007. Genomic comparison using data mining techniques based on a possibilistic fuzzy sets model. Biosystems, 88: 343-349

25 Gustafson D & W Kessel, 1979. Fuzzy clustering with fuzzy covariance matrix. IEEE Conference on Decision and Control (San Diego) 761-766

26 Babuska R, P J van der Veen & U Kaymak, 2002. Improved co-variance estimation for Gustafson-Kessel clustering. IEEE International Conference on Fuzzy Systems (Honolulu, USA) 2: 1081-1085

27 Beaven S G , D Stein & L E Hoff, 2000. Comparison of Gaussian mixture and linear mixture models for classification of hyperspectral data. IEEE Geoscience and Remote Sensing Society (Honolulu, USA) 1597-1599

28 Biernacki C, G Celeux, G Govaert & F Langrognet, 2006. Model-based cluster and discriminant analysis with the MIXMOD software. Computational Statistics & Data Analysis, 51: 587-600

29 Dempster A, N Laird & D Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 39: 1-38

30 Banfield J & A Raftery, 1993. Model-based Gaussian and non-Gaussian clustering. Biometrics, 49: 803-821

31 Fraley C & A Raftery 1998. How many clusters? Which clustering method? Answers via model- based cluster analysis. The Computer Journal, 41: 578-588

32 Jain A & D Zongker, 1997. Feature selection: evaluation, application, and small sample per-formance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19: 153-158

33 Forman G, 2004. A pitfall and solution in multi-class feature selection for text classification, In: 21st International Conference on Machine Learning, 38

34 Kumar S, J Ghosh & M Crawford, 2001. Best-bases feature extraction algorithms for classifi-cation of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 39: 1368-1379

35 Jolliffe, 1986. Principal Component Analysis (Springer, New York) 271 pp.

36 Liu H, E Dougherty, J Dy, K Torkkola, E Tuv, H Peng, C Ding, F Long, M Berens, L Parsons, Z Zhao, L Yu & G Forman, 2005. IEEE Intelligent Systems, Evolving feature selection, 20: 64-76

37 Serpico S & L Bruzzone, 2001. A new search algorithm for feature selection in hyperspectral remote sensing images. IEEE Transactions and Geoscience and Remote Sensing, 39: 1360-1367

38 Trunk G, 1979. A problem of dimensionality: A simple example, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1: 306-307

39 Pudil P, J Novovicova & J Kittler, 1994. Floating search methods in feature selection. Pattern Recognition Letters, 15: 1119 -1125

40 Berry M J A & G Linoff, 2004. Data Mining Techniques For Marketing, Sales and Customer Support (John Wiley & Sons, Inc., USA.) pp 643

41 Xu Y & R G Brereton, 2005. A comparative study of cluster validation indices applied to geno-typing data. Chemometrics and Intelligent Laboratory Systems, 78: 30-40

42 Hammah R E & J Curran, 2000. Validity measures for the fuzzy cluster analysis of orienta-tions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22: 1467-1472

43 Celeux G & G Soromenho, 1996. An entropy criterion for assessing the number of clusters in a mixture model. Journal of Classification, 13: 95-212

44 Biernacki C, G Celeux & G Govaert, 1999. An improvement of the NEC criterion for assessing the number of components arising from a mixture. Pattern Recognition Letters, 20: 267-272

45 Liu H, J Yih, D Wu & S Liu, 2008. Fuzzy c-mean algorithm based on "complete" Mahalanobis distances. In: The Seventh International Conference on Machine Learning and Cybernetics, 6: 3569-3574

46 Bozdogan H, 1993. Choosing the number of clusters in the mixture-model using a new infor-mational complexity criterion of the Inverse-Fisher Information Matrix. In: Opitz O, B Klausen & R Klar (Editors) Information and Classification (Springer, Heidelberg) 40-54

47 Klein D, A Moll & G Menz, 2005. Land Cover/Land Use Classification in a Semiarid Environ-ment in East Africa Using Multi-Temporal Alternating Polarisation Envisat ASAR Data. In: 2004 Envisat & ERS Symposium (ESA SP-572)

48 Richards J A, 1999. Remote Sensing Digital Image Analysis (Springer-Verlag, Berlin) 240 pp.