A SCHEME FOR INITIAL EXPLORATORY DATA ANALYSIS OF MULTIVARIATE IMAGE DATA

Klaus Baggesen Hilger, Allan Aasbjerg Nielsen, Rasmus Larsen

IMM, Department of Mathematical Modelling
Technical University of Denmark, Building 321
DK-2800 Kgs. Lyngby, Denmark
e-mail: kbh@imm.dtu.dk, http://www.imm.dtu.dk/∼kbh

ABSTRACT

A new scheme is proposed for handling initial exploratory analyses of multivariate image data. The method is invariant to linear transformations of the original data and is useful for data fusion of multisource measurements. The scheme includes dimensionality reduction followed by unsupervised clustering of the data. A transformation is proposed which maximizes autocorrelation by projection onto subspaces with signal-to-noise ratio dependent variance. We apply the traditional fuzzy c-means algorithm and introduce two additional memberships enhancing the textural awareness of the algorithm. Cluster validation is performed by examining the partition density of the segmentations as a function of the number of classes applied. Results are presented for a synthetic 2-band noise degenerated image and for an 8-band SeaWiFS scene.

1. INTRODUCTION

In the GEOSONAR project, [5], we handle large amounts of multivariate image type data obtained from different sources. Looking for correlated patterns in the large amount of data, a natural need occured to develop tools for handling the initial exploratory analyses. The multisource data are obtained primarily from remote sensing, but they also include in-situ measurements and output from atmospherical models. An important part of the project is data fusion that reduces dimensionality and eliminates redundancy. This problem is addressed in Section 2. We apply what is known as the minimum/maximum autocorrelation factor transform (MAF), [13, 4, 9], but scale the output relative to the amount of signal contained in each component. The resulting components are termed SMAFs. The traditional MAFs are invariant to linear transformations. This is very useful when combining data from different sources. In the SMAFs this property is preserved. The principal components transform (PC) is also a candidate for dimensionality reduction, [1], but it is not invariant to linear transformations. Furthermore, it appears to be inferior in comparison with the MAF transform when attempting to separate signal from noise in data of spatial nature.

Following the SMAF transformation the exploratory analysis is dedicated to locating natural clusters in the data, see Section 3. Many clustering methods have been proposed that attempt to find natural groupings in multidimensional data. Measures used to quantify the algorithms are often related to speed, reliability and consistency, [8]. New clustering algorithms are constantly being developed. In this paper we present an extended version of the traditional spectral fuzzy c-means (FCM) algorithm, [2]. Two additional textural memberships are introduced and incorporated into the algorithm. The new memberships are general in the sense that they can be applied with different spectral fuzzy functionals other than the one proposed here.

A syntectic 2-band test image is analysed. Results can be found in Sections 2 and 3. A case study is presented using an 8-band SeaWiFS, [6], scene in Section 4. SMAFs are calculated and the data is analysed using the extended FCM algorithm. Section 5 contains some concluding remarks.

2. DIMENSIONALITY REDUCTION

In the following the PC, the MAF and the SMAF transformations are described. All the methods are applied to the test image presented in Figure 1 (top row) stretched linearly between its mean ±3 standard deviations (std). The test image consists of two bands (100^2 pixels in each) containing three classes: 72% a background class ($c = 1$), 24% a cross ($c = 2$), and 4% the center of the cross ($c = 3$). Consider a multivariate data set of P variables with grey levels $r_i(\boldsymbol{x}), i = 1, \cdots, P$, where $\boldsymbol{x}$ is the coordinate vector denoting the grid point of the sample. The 2-band ($P = 2$) test image is generated using

$$\begin{aligned} r_1(\boldsymbol{x}) &= s(\boldsymbol{x}) + n_1(\boldsymbol{x})/2 + n(\boldsymbol{x}) \\ r_2(\boldsymbol{x}) &= s(\boldsymbol{x}) + n_2(\boldsymbol{x})/2 + 2n(\boldsymbol{x}) \end{aligned} \quad (1)$$

where the signal $s(\boldsymbol{x}) = c(\boldsymbol{x}) - 1$, with the grid point dependent class index $c(\boldsymbol{x}) \in \{1, 2, 3\}$. Choosing n_1, n_2 and n as

This work was supported by the Danish Research Councils under the Earth Observation Program, [7]. Use of the SeaWiFS data is in accord with the SeaWiFS Research Data Use Terms and Conditions Agreement.

i.i.d. zero mean Gaussian with unit variance, the covariance matrix for the total noise is given by $[1.25\ \ 2;\ 2\ \ 4.25]$ (using semicolon separated rows).

2.1. Principal Components Analysis

Let us consider a random signal variable $\boldsymbol{r}(\boldsymbol{x}) = [r_1(\boldsymbol{x}); \cdots ; r_P(\boldsymbol{x})]$ and assume first and second order stationarity such that

$$\mathrm{E}\{\boldsymbol{r}(\boldsymbol{x})\} = \mathbf{0} \tag{2}$$

$$\mathrm{D}\{\boldsymbol{r}(\boldsymbol{x})\} = \boldsymbol{\Sigma}. \tag{3}$$

Determining the direction of maximum variation means finding the direction $\boldsymbol{a}$, with $\boldsymbol{a}^T\boldsymbol{a} = 1$, such that the linear combination $y(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{r}(\boldsymbol{x})$ possesses maximum variance.

The PC transformation thus chooses P linear transformations

$$y_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{r}(\boldsymbol{x}),\ \ i = 1, \cdots, P \tag{4}$$

such that the variance for $y_i(\boldsymbol{x})$ is maximum among all linear transforms orthogonal to $y_j(\boldsymbol{x}), j = 1, \cdots, i-1$. The variance is given by

$$\lambda_i = \boldsymbol{a}_i^T\boldsymbol{\Sigma}\boldsymbol{a}_i. \tag{5}$$

We see that the basis for the PCs is identified as the conjugate eigenvectors of the dispersion matrix. Let $\lambda_1 \geq \cdots \geq \lambda_P \geq 0$ be the eigenvalues with the corresponding conjugate eigenvectors $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_P$. Then $y_i(\boldsymbol{x})$ is the i'th PC (PCi) with variance λ_i.

In Figure 1 (middle row) the PCs of the test image are shown. Approximately 95% of the variance in the data is explained by PC1. The images are stretched linearly between their mean ± 3 std. Notice that the PC transformation has problems separating the signal from the noise. In the test image the variance of the noise is much larger than the variance of the signal, therefore the eigenvalue analysis is dominated by the dispersion matrix of the noise. The PC transformation is a rotation of the original basis. With the test image being two-dimensional, the scatter plot in Figure 2 (first row/first column) can also serve to illustrate the spectral space of the calculated PCs.

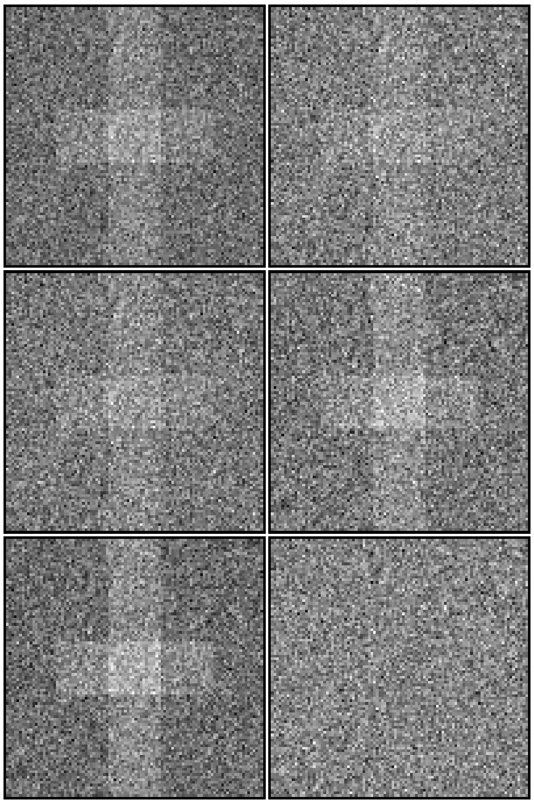

Fig. 1. Top row contains the test image. The middle row contains the calculated PCs and the bottom row contains the MAFs. All images are scaled to mean ± 3 std. Notice that the MAF transform performs a much better job at separating signal from noise when compared to the PC transform.

2.2. Maximum Autocorrelation Factors Analysis

Let us again consider the random signal variable $\boldsymbol{r}(\boldsymbol{x})$ and assume first and second order stationarity by imposing Eq. 2 and 3. The MAF transformation chooses P linear transformations

$$z_i(\boldsymbol{x}) = \boldsymbol{b}_i^T\boldsymbol{r}(\boldsymbol{x}),\ \ i = 1, \cdots, P \tag{6}$$

such that the spatial autocorrelation for $z_i(\boldsymbol{x})$ is maximum among all linear transforms orthogonal to $z_j(\boldsymbol{x}), j = 1, \cdots, i-1$.

Let $\boldsymbol{\Delta} = [\Delta_1;\ \Delta_2]$ represent a spatial shift. The spatial covariance function is then defined by

$$\mathrm{Cov}\{\boldsymbol{r}(\boldsymbol{x}), \boldsymbol{r}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{\Gamma}(\boldsymbol{\Delta}). \tag{7}$$

$\boldsymbol{\Gamma}$ has the following properties $\boldsymbol{\Gamma}(\mathbf{0}) = \boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}(\boldsymbol{\Delta})^T = \boldsymbol{\Gamma}(-\boldsymbol{\Delta})$.

Looking at the correlations between projections of the variables and the shifted variables, we find

$$\begin{aligned}&\mathrm{Cov}\{\boldsymbol{b}_i^T\boldsymbol{r}(\boldsymbol{x}), \boldsymbol{b}_i^T\boldsymbol{r}(\boldsymbol{x}+\boldsymbol{\Delta})\}\\ &\quad = \frac{1}{2}\,\boldsymbol{b}_i^T(\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta}))\boldsymbol{b}_i.\end{aligned} \tag{8}$$

Introducing $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}$, gives

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\Delta}} &= \mathrm{D}\{\boldsymbol{r}(\boldsymbol{x}) - \boldsymbol{r}(\boldsymbol{x}+\boldsymbol{\Delta})\}\\ &= 2\,\boldsymbol{\Sigma} - (\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta}))\end{aligned} \tag{9}$$

which when considered as a function of $\boldsymbol{\Delta}$ is a multivariate variogram. Combining Eq. 8 and 9 we obtain

$$\mathrm{Cov}\{\boldsymbol{b}_i^T\boldsymbol{r}(\boldsymbol{x}), \boldsymbol{b}_i^T\boldsymbol{r}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{b}_i^T(\boldsymbol{\Sigma} - \frac{1}{2}\,\boldsymbol{\Sigma}_{\boldsymbol{\Delta}})\boldsymbol{b}_i. \tag{10}$$

Dividing this by the variance of the data projected on the subspace corresponding to $\boldsymbol{b}_i$, the expression for the correlation is given by

$$\text{Corr}\{\boldsymbol{b}_i^T\boldsymbol{r}(\boldsymbol{x}), \boldsymbol{b}_i^T\boldsymbol{r}(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\frac{\boldsymbol{b}_i^T\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}\boldsymbol{b}_i}{\boldsymbol{b}_i^T\boldsymbol{\Sigma}\boldsymbol{b}_i}. \quad (11)$$

If we want to maximize the correlation, we must minimize the Rayleigh coefficient

$$\chi = \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}\boldsymbol{b}}{\boldsymbol{b}^T\boldsymbol{\Sigma}\boldsymbol{b}}. \quad (12)$$

Let $\chi_1 \leq \cdots \leq \chi_P$ be the eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}$ with respect to $\boldsymbol{\Sigma}$ with the corresponding conjugate eigenvectors $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_P$. Then $z_i(\boldsymbol{x})$ is the i'th MAF (MAFi).

It is possible to calculate MAFs using two-set canonical correlation analysis (CCA) comparing the original data to a spatially shifted set. In CCA the resulting components are conditioned to have unit variance, [1], and the MAFs have traditionally been imposed the same restriction.

In order to obtain an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}$, Switzer and Green, [13], recommend the formation of two sets of difference images. The two sets are $\boldsymbol{r}(\boldsymbol{x}+\boldsymbol{\Delta_h})$ and $\boldsymbol{r}(\boldsymbol{x}+\boldsymbol{\Delta_v})$ where $\boldsymbol{\Delta_h}$ is a unit horizontal shift and $\boldsymbol{\Delta_v}$ is a unit vertical shift. Calculate the dispersion matrices for both sets of images and pool them as an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\Delta}}$.

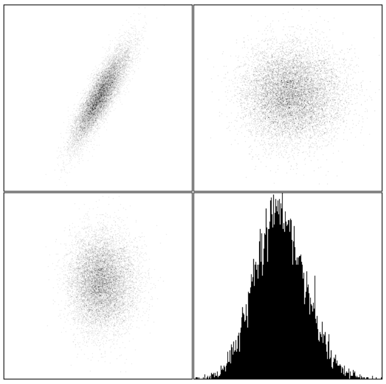

Fig. 2. Scatterplots 1-3 and the histogram of SMAF1, rowwise ordering. The scatterplots are calculated from 1) the original data, 2) the traditional MAFs, and 3) MAFs preserving Euclidean distance. Notice that classes are not necessarily well separated in the SMAF space.

The MAF transform is applied to the test image, producing the images of Figure 1 (bottom row) scaled to mean ± 3 std. Comparing the MAFs to the PCs, we notice that they do a much better job at separating the signal from noise. The reason why MAF performs so much better than the PC is related to the fact that the noise is constructed to be highly correlated in the two bands of the test image. The signal contained in MAF1 has an autocorrelation of 0.26 and in MAF2 of -0.01. In Figure 2 (first row/second column) is included a scatter plot of the traditional MAFs shown in Figure 1.

Analysis of high dimensional data using the MAF transform can be found in [9]. Conceptually it produces a nice ordering of the new components and can often be perceived as a decomposition of spatial frequency. Unlike the PC the MAF transformation is invariant to rescaling and other linear rank preserving transformations of the original data. Notice that MAF reduces to ordinary PC only when the covariance matrix of the multidimensional noise-signal is proportional to the identity matrix.

2.3. Signal-to-Noise Ratio Scaling

When we assume that an additive noise structure applies $\boldsymbol{r}(\boldsymbol{x}) = \boldsymbol{s}(\boldsymbol{x}) + \boldsymbol{n}(\boldsymbol{x})$ with $\text{Corr}\{\boldsymbol{s}(\boldsymbol{x}), \boldsymbol{n}(\boldsymbol{x})\} = 0$, the dispersion structure can be separated into $\text{D}\{\boldsymbol{r}(\boldsymbol{x})\} = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_n$. We define a signal-to-noise ratio (SNR) for the i'th component by

$$\text{SNR}_i = \text{V}\{s_i(\boldsymbol{x})\}/\text{V}\{n_i(\boldsymbol{x})\}, \quad (13)$$

and evaluate it for the MAFs producing

$$\begin{aligned} \text{SNR}_i &= \frac{\text{V}\{\boldsymbol{b}_i^T\boldsymbol{s}(\boldsymbol{x})\}}{\text{V}\{\boldsymbol{b}_i^T\boldsymbol{n}(\boldsymbol{x})\}} \\ &= \frac{\boldsymbol{b}_i^T\boldsymbol{\Sigma_s}\boldsymbol{b}_i}{\boldsymbol{b}_i^T\boldsymbol{\Sigma_n}\boldsymbol{b}_i} = \frac{\boldsymbol{b}_i^T(\boldsymbol{\Sigma}-\boldsymbol{\Sigma_n})\boldsymbol{b}_i}{\boldsymbol{b}_i^T\boldsymbol{\Sigma_n}\boldsymbol{b}_i}. \end{aligned} \quad (14)$$

Assuming a proportional covariance model, [13], an estimate of the dispersion structure for the noise can be found as

$$\boldsymbol{\Sigma_n} = \boldsymbol{\Sigma}_{\boldsymbol{\Delta}}/2 \quad (15)$$

and we obtain the SNR as a function of the eigenvalues of the MAF analysis

$$\text{SNR}_i = \frac{2}{\chi_i} - 1. \quad (16)$$

In [10] several models are presented for estimating noise in images. Using Eq. 15, the estimate of the covariance structure for the noise of the test image is found as $[1.24 \;\; 1.98; 1.98 \;\; 4.23]$ which can be compared to the one presented in the beginning of this section.

We can now scale the variance of the MAFs relative to the SNR in each component. The justification for doing this is to stretch the subspaces spanned by the basis of the MAFs favoring those rich in autocorrelated signal and low in noise. The stretching/compression is done to increase the distance between natural clusters in the data reducing Bayes Risk, [12] , and improving the robustness when handling high dimensional data. A natural choice is to scale the new components such that the covariance structure of the noise becomes the indenty matrix. Thus we impose the condition $\text{V}\{\boldsymbol{b}_i^T\boldsymbol{n}(\boldsymbol{x})\} = 1$ and obtain new components with variance $\text{SNR}_i + 1 = 2/\chi_i$. In order to discard components that have little signal and to reduce the dimensionality of the data, we suggest to eliminate SMAFs that have negative SNRs that can occur due to model breakdown. For the MAFs presented in Figure 1, the SNRs are calculated to be 0.35 and -0.01.

In the second row of Figure 2 two different transformations of the original MAFs are applied. In the second row/first column a stretch is imposed to produce components that preserve the Euclidean distance in the data. It is not invariant to linear transformation of the original data and does not qualify as a SMAF. Second row/second column contains a histogram of SMAF1 generated by using the transformation described above. Since SMAF2 has negative SNR, it is eliminated. Notice that classes are not necessarily well separated looking only at spectral properties of the data in the SMAF space.

3. UNSUPERVISED IMAGE SEGMENTATION

In this section the notion of fuzzy partitioning is introduced followed by the different types of memberships that we apply, i.e., a spectral and two textural memberships. The SMAF transform of the test image is analysed using the extended algorithm.

3.1. Fuzzy Partitioning

Given N P-dimensional observations that we wish to classify into C classes, the fuzzy c-means algorithm applies. Let $\boldsymbol{U}$ be an $N \times C$ matrix with the elements $u_{ic} = u_c(\boldsymbol{x})$ that describe the membership for observation i at grid point $\boldsymbol{x}$ to class c, and let $\boldsymbol{R}$ be a $P \times C$ matrix that contains the cluster centers, i.e., the centroids of the classes. The spectral FCM algorithm minimizes the within class sum of squared errors functional $J(\boldsymbol{U}, \boldsymbol{R})$ under the conditions

$$u_{ic} \in [0,1] \quad \forall\, i = 1,\ldots,N; c = 1,\ldots,C \tag{17}$$

$$\sum_{c=1}^{C} u_{ic} = 1 \qquad \forall\, i = 1,\ldots,N \tag{18}$$

$$\sum_{i=1}^{N} u_{ic} > 0 \qquad \forall\, c = 1,\ldots,C \tag{19}$$

Any $\boldsymbol{U}$ that satisfies the conditions is referred to as a fuzzy partitioning of N individuals into C classes. The functional proposed by Bezdek, [2], is given by

$$J(\boldsymbol{U}, \boldsymbol{R}) = \sum_{i=1}^{N}\sum_{c=1}^{C} u_{ic}^{m} d_{ic}^{2}. \tag{20}$$

Here $m \geq 1$ is a fixed parameter that determines the degree of fuzziness of the final solution, that is the degree of overlap between groups. With $m = 1$ the solution is a hard partitioning. The degree of fuzziness increases with m. The higher the value of m, the lower the membership degrees of observations which are far from all cluster centers. As m approaches infinity, the solution approaches its highest degree of fuzziness, with $u_{ic} = 1/C$ for every pair of i and c. The squared distance between the i'th observation and the c'th cluster center is denoted d_{ic}^2.

3.2. Spectral FCM

The spectral fuzzy c-means algorithm

1. assigns values to P-dimensional feature vectors for C cluster centers, $\tilde{\boldsymbol{r}}_c,\ c = 1,\ldots,C$;

2. assigns to each observation $i = 1,\ldots,N$ membership weights for clusters $c = 1,\ldots,C$

$$u_{spec,ic} = \frac{1/d_{ic}^{2/(m-1)}}{\sum_{j=1}^{C} 1/d_{ij}^{2/(m-1)}} \tag{21}$$

where the squared spectral distance from the running observation $\boldsymbol{r}_i$ to each cluster center $d_{ic}^2 = (\boldsymbol{r}_i - \tilde{\boldsymbol{r}}_c)^T F_c^{-1}(\boldsymbol{r}_i - \tilde{\boldsymbol{r}}_c)$, and $m > 1$. With $F_c = \boldsymbol{I}$ the Euclidean distance measure is applied. Using the Mahalanobis distance, F_c can be calculated from

$$F_c = \frac{\sum_{i=1}^{N} u_{ic}^{m}(\boldsymbol{r}_i - \tilde{\boldsymbol{r}}_c)(\boldsymbol{r}_i - \tilde{\boldsymbol{r}}_c)^T}{\sum_{j=1}^{N} u_{ic}^{m}}; \tag{22}$$

3. calculates new cluster centers from

$$\tilde{\boldsymbol{r}}_c = \frac{\sum_{i=1}^{N} u_{ic}^{m}\, \boldsymbol{r}_i}{\sum_{i=1}^{N} u_{ic}^{m}}. \tag{23}$$

Steps 2 and 3 are iterated until the largest change in cluster membership becomes small or zero.

3.3. Including Spatial Information

Two additional memberships, the spatial and the parental, are introduced to include spatial context information. Before calculating the new cluster centers, we merge the spectral, the spatial, and the parental memberships into joint memberships which are applied as weights for the spectral observations.

We define spatial membership by

$$u_{spat,ic} = \frac{1}{Z}\exp(-\beta E_{\mathcal{N}}) \tag{24}$$

where $E_{\mathcal{N}} = |\mathcal{N}|^{-1}\sum_{j\in\mathcal{N}}(1-u_{jc})$ corresponds to a Markov random field (MRF) energy function, $\beta \geq 0$ is a weight parameter, and Z is a normalizing constant. The sum over $\mathcal{N}$ indicates a sum over the neighbourhood of an observation and $|\mathcal{N}|$ is the number of neighbours in $\mathcal{N}$. With $\beta = 0$ no spatial context information is included. The spatial membership to a class is large, if the observations in the neighbourhood have large memberships to the same class and small, if the neighbours tend to belong to other classes. In each iteration we estimate the spatial membership degrees from the spectral memberships.

Using a combination of blurring and subsampling, a multiresolution scale-space pyramid is constructed from the input data. We apply a nearly Gaussian blurring kernel (a 5-tap filter with values 0.05, 0.25, 0.4, 0.25, and 0.05) and perform subsampling such that a parent pixel at level j has four children at level $j-1$. We use reflection in order to handle border effects. In a j-level pyramid, level 0 corresponds to the level of highest resolution and level $j-1$ to the lowest. Applying the FCM algorithm, the scale-pyramid is processed top down. The resulting cluster centers from one level are passed down to the next as initial cluster centers. Passing the memberships found at a higher level down through the pyramid introduces additional spatial awareness into the segmentation algorithm. We introduce the additional membership as an external field that corresponds to an a priori knowledge of the memberships of the given observation to the different classes. The membership of a parent of an observation i to a class c is denoted $u_{parent,ic}$. The parental memberships applied at level j are generated by linear interpolation among the nearest neighbours of the converged fuzzy field at level $j+1$.

3.4. Merging Memberships

The joint spectral-spatial-parental membership can be calculated as

$$u_{ic} = \frac{u_{spec,ic}\ u_{spat,ic}\ u_{parent,ic}}{\sum_{j=1}^{C} u_{spec,ij}\ u_{spat,ij}\ u_{parent,ij}}. \qquad (25)$$

Numerous functionals have been proposed for calculating spectral membership, almost all of these are somehow related to Bezdek's original formulation.

3.5. Test Results

The SMAF transform of the 2-band test image is analysed. Before applying the FCM algorithm, the parameter m has to be set by the user. With default $m = 2$ the algorithm is optimized computationally. The value of β and the number of levels used in computing the parental membership can be chosen depending on the image size and the nature of the input data. Default is $\beta = 1$, using a 2-level scale-pyramid. The SMAFs provides valuable information about the noise level in the data and could be used to set up dynamical parameter control. Even when using SMAF as a preprocessor, clusters can still overlap in the feature space. The test image is meant to illustrate this problem and demonstrate the usefulness of the additional textural fuzzy memberships. Moreover, spectral FCM has already been reported successful in detecting compact, hyperspherical, well-separated clusters in the feature space. If the Mahalanobis distance measure is used, the algorithm is even able to detect hyperellipsoidally

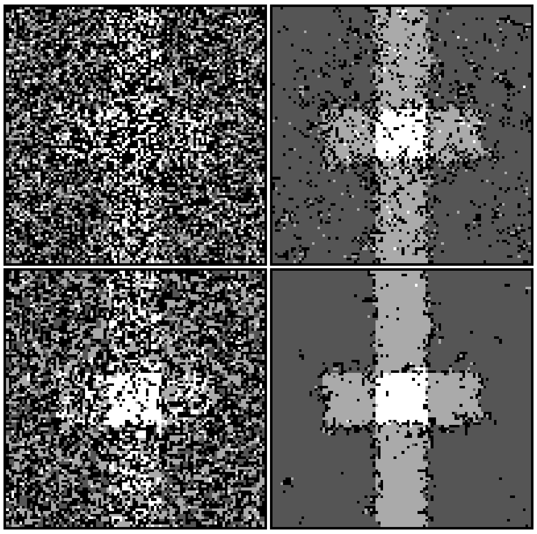

Fig. 3. Image segmentation 1-4 (row-wise) using the input of the SMAF transformation. The results are based on the memberships: 1) spectral, 2) spectral-parental, 3) spectral-spatial, 4) spectral-spatial-parental. Black=reject class, greyscale $\propto$ class index. Including textual information improves the segmentation results.

shaped clusters, [8]. In order to benefit from the SMAF transformation, we apply the Euclidean distance resulting in a fast algorithm. With a low SNR of 0.3 in SMAF1 a relatively high $\beta = 4$ and a 3-level scale-pyramid are applied. Results are shown in Figure 3 in which hard classifications are made by selecting the class with the highest membership above a given reject threshold, here 0.9. If the assignment fails, the pixel is assigned to a reject class. Inspecting Figure 3 we see that the resulting segmentations are significantly improved by including textural memberships. Combining the different memberships favours segmentation into homogeneous regions, and mainly pixels at the class boundaries are assigned to the reject class.

One-hundred test images, each generated as described in Section 2, are analysed together with the corresponding SMAFs by means of the extended FCM algorithm. The above described parameter settings are applied using spectral-spatial-parental memberships. The results are shown in Table 1. Using the SMAF transformation as a preprocessor improves the classifications. The segmentations improve from an average of 86% to 93% correctly classified pixels, and the percentage of rejected pixels is reduced from 11% to 5%. Evidence can be found of improved reliability of the algorithm's performance by inspecting the std in the table.

3.6. Cluster Validation

We assume good partitioning consists of i) clear separation between the resulting clusters, ii) minimal volume of the clusters, and iii) maximal number of data points concentrated in the vicinity of the cluster center.

Table 1. Summary of 100 segmentation results using the original data and using SMAFs as input. The percentage (mean and std) of correctly classified pixels in the different class regions and of rejected pixels are included. Spectral-spatial-parental memberships are applied and hard results are obtained using a reject threshold of 0.9. Using the SMAF transform as preprocessor improves the performance of the clustering algorithm.

Region of	Original data		SMAFs	
interest:	mean %	std %	mean %	std %
Class 1	91.7	5.4	96.3	0.5
Class 2	71.8	7.9	84.7	1.5
Class 3	78.9	6.6	85.1	3.5
The whole image	86.4	5.5	93.0	0.4
Rejected pixels	11.4	2.4	5.4	0.5

Introducing the fuzzy hypervolume

$$F_{HV} = \sum_{i=c}^{C} [\det(F_c)]^{1/2} \tag{26}$$

where F_c is defined in Eq. 22, and summing the memberships of the observations within one standard deviation of each cluster center

$$\begin{aligned} S &= \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ic} \\ \forall \quad & \boldsymbol{r}_i \in \{\boldsymbol{r}_i : (\boldsymbol{r}_i - \tilde{\boldsymbol{r}}_c)^T F_c^{-1} (\boldsymbol{r}_i - \tilde{\boldsymbol{r}}_c) < 1\} \end{aligned} \tag{27}$$

the partition density is defined by $P_D = S/F_{HV}$, [3]. Local maxima in the partition density are expected where good partitionings af the data are obtained. Other measures of cluster validation can be found in [11].

4. A SEAWIFS CASE STUDY

SeaWiFS is an 8 channel optical scanner on the SeaStar spacecraft which orbits sun-synchronously at 705 km altitude. On a daily basis SeaWiFS provides 10 bit data in the 402-422, 433-453, 480-500, 500-520, 545-565, 660-680, 745-785 and 845-885 nm regions. The pixel size is 1.1 km $\times$ 1.1 km. See also [6].

A SeaWiFS scene acquired on 13 May 1998 is analysed using the proposed exploratory analysis. In Figure 4 we show the original data scaled to mean ± 3 std, and in Figure 5 the SMAFs are scaled in the same manner. In order to interpret the different components of the SMAFs, Table 2 shows the autocorrelations, the SNRs, and the correlations between the SMAFs and the original data. Since the components in the original data may be correlated, one should look at correlations rather than the weights from the linear transformation. High SNR values are present in the SMAFs. Therefore the default settings were used applying the extended FCM algorithm. Hard results into 5 and 6 classes, with no reject class, are shown in the top row of Figure 6. For the segmentation into 6 classes, the figure also contains the corresponding membership degrees ordered row-wise. Performing segmentation only based on spectral membership produces meaningful results (not shown). However, adding spatial and parental context leads to more distinct membership degrees, i.e., membership degrees closer to 0 and 1 indicating a better classification.

In Figure 7 the partition density is plotted vs. the number of classes applied. For 5 and 6 classes a local maxima indicate good partitioning. In the figure the class descriptors, i.e., the cluster centroids, are included for the segmentation into 6 classes. The memberships presented in Figure 6 are those used in calculating the class descriptors. Inspecting the spectral signatures and the segmentation results, a user can identify the dominating signal-source for the classes 1-6 as 1) ocean water, 2) water/clouds, 3-4) land vegetation, 5) vegetation/clouds, and 6) thick clouds and ice. Examining other segmentations into different number of classes, the algorithm still appears to produce meaningful results (not shown). Inspecting the different segmentations with emphasis on the ones with local maxima in the partition density and comparing them to the original data and the SMAFs (and possibly PCs), a user can obtain valuable information about the natural clusters and correlation structures in the data.

5. CONCLUSION

A data driven scheme for exploratory analysis is proposed. The initial SMAF orthogonalization corresponds to the maximum autocorrelation factors transform but includes a scaling of the components proportional to the signal-to-noise ratio estimated in each subspace. The motivation for the stretching/compression is to increase the distance between natural clusters in the data. By applying the SMAFs, the following unsupervised cluster analysis weights the variates rich in autocorrelated signal higher. Improved segmentation results for simulated image data are obtained using the SMAF transformation as preprocessor. An extended version of the fuzzy c-means algorithm enhanced for segmentation of image type data is proposed. In addition to the traditional spectral membership, a spatial and a parental membership are introduced enhancing the textural awareness of the algorithm. The extensions are easily incorporated into other spectral fuzzy clustering algorithms. Improved segmentation results for the simulated image data are obtained by applying a joint membership. The resulting segmentations are invariant to linear transformation of the original data. A case study is presented in which an 8-band SeaWiFS scene is analysed. The analysis includes an automated evaluation of segmentation as a function of the number of classes. This is helpful for detecting "good" partitionings and is useful

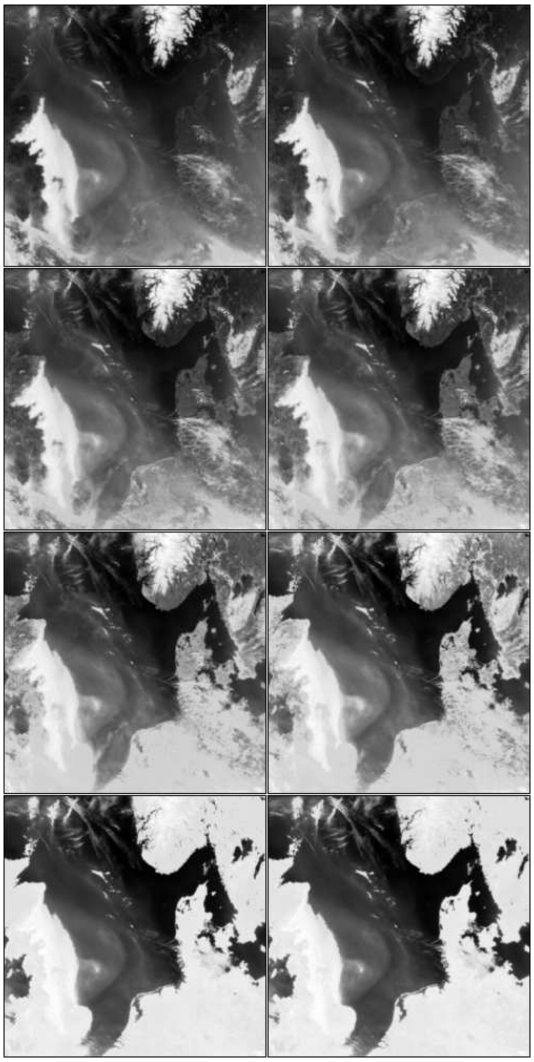

Fig. 4. The SeaWiFS bands 1-8 ordered row-wise. Each band is scaled to mean ± 3 std.

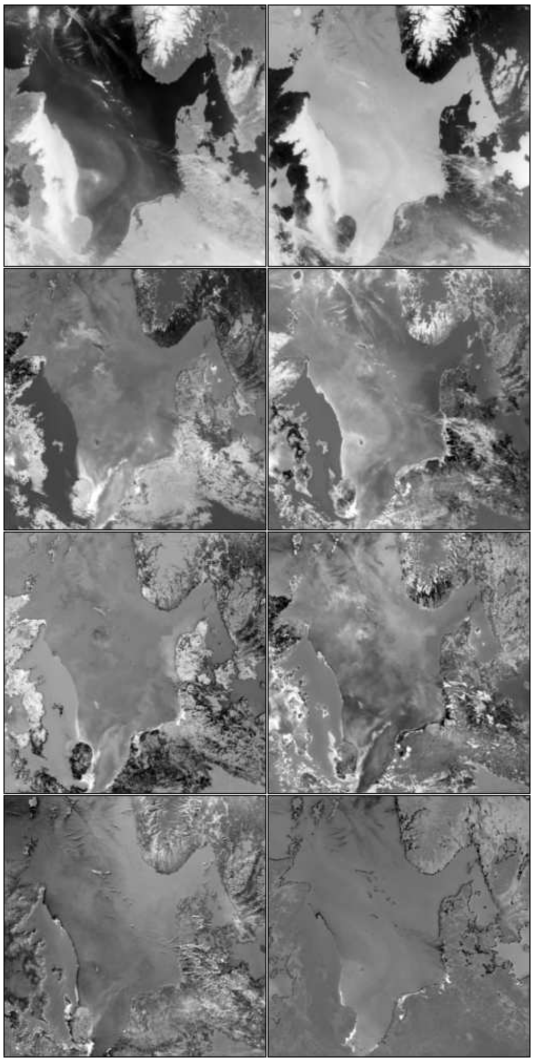

Fig. 5. The SMAFs 1-8 of the SeaWiFS data ordered row-wise. Each SMAF is scaled to mean ± 3 std.

when inspecting the data prior to supervised data analyses. The suggested exploratory approach has been developed in the GEOSONAR project for handling multivariate image type data obtained from both multisource measurements and model outputs.

	SMAF1	SMAF2	SMAF3	SMAF4	SMAF5	SMAF6	SMAF7	SMAF8
Autocorr.	0.995	0.989	0.959	0.941	0.876	0.849	0.702	0.505
SNR	207.9	93.5	23.1	15.9	7.1	5.6	2.4	1.0
Band1	0.838	0.539	-0.068	-0.014	0.033	0.010	0.010	0.001
Band2	0.873	0.478	-0.054	0.030	0.035	0.034	0.043	0.001
Band3	0.910	0.381	0.032	0.151	0.021	0.049	0.019	0.002
Band4	0.943	0.277	0.103	0.146	-0.004	0.019	0.030	0.003
Band5	0.980	-0.005	0.171	0.076	0.072	0.013	0.003	0.007
Band6	0.962	-0.160	0.043	0.173	0.099	-0.087	0.022	0.005
Band7	0.908	-0.403	-0.093	0.009	-0.038	0.006	-0.010	0.058
Band8	0.904	-0.413	-0.098	0.005	-0.047	0.012	-0.012	0.001

Table 2. Autocorrelations, SNRs, and correlations between the SMAFs and the original variables. The first SMAFs contains the highest SNRs and correlations.

6. REFERENCES

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York, second edition, 1984. 675 pp.

[2] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981.

[3] I. Gath and A. B. Geva. Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI 11(7):773–781, July 1989.

[4] A. A. Green, M. Berman, P. Switzer, and M. D. Craig. Transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, 1988.

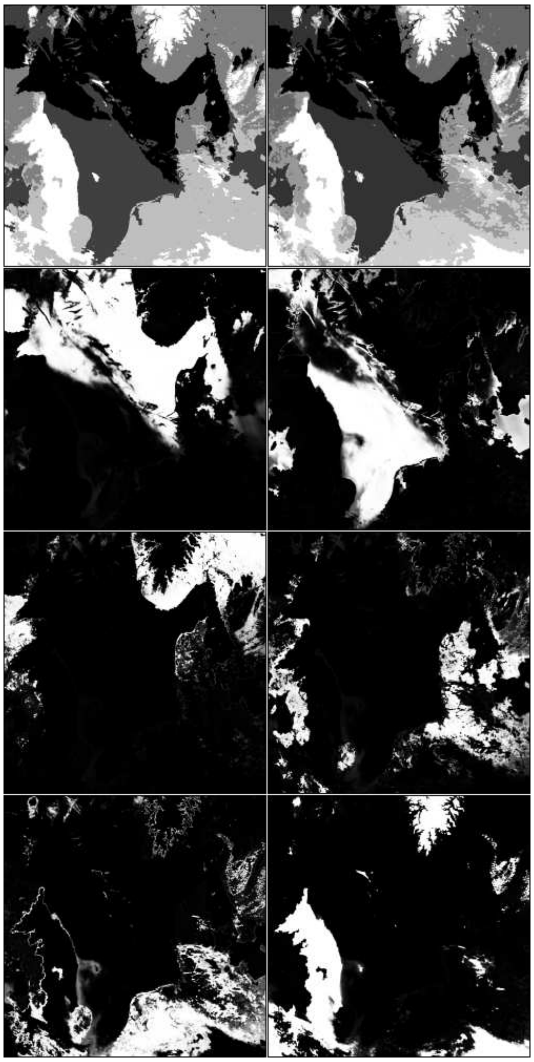

Fig. 6. Top row: Hard segmentations into 5 and 6 classes without a rejcet class, greylevel $\propto$ class number. Second row and down: Membership degrees for segmentation into 6 classes (row-wise ordering) of the SMAFs 1-8.

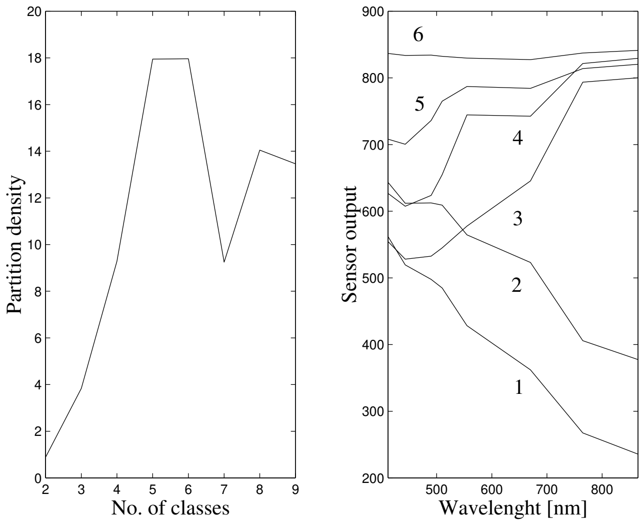

Fig. 7. Left: The partition density as a function of the number of classes. Local maxima indicate good partitioning. Right: Resulting class descriptors using 6 classes.

[5] http://manicoral.kms.min.dk/PK/geosonarproj.html. The GEOSONAR (GEOid and Sea level Of the North Atlantic Region) project description, serviced by Per Knudsen, KMS - Geodetic Division, Denmark.

[6] http://seawifs.gsfc.nasa.gov/SEAWIFS.html. Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, Maryland, USA.

[7] http://www.forskraad.dk/saerforsk/jordobs/joobs0.html. The Earth Observation Programmes funded by the joint Danish Research Councils.

[8] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI 22(1):4–37, January 2000.

[9] A. A. Nielsen. *Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data*. PhD thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, 1994. Internet http://www.imm.dtu.dk/∼aa/phd/.

[10] S. I. Olsen. Estimation of noise in images: An evaluation. *Graphical Models and Image Processing*, 55(4):319–323, 1993.

[11] M. Razaee, P. van der Zwet, B. Lelieveldt, R. van der Geest, and J. Reiber. A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI 9(7):1238–1248, July 2000.

[12] B.D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1995.

[13] P. Switzer and A. A. Green. Min/max autocorrelation factors for multivariate spatial imagery. Technical Report 6, Dept. of Statistics, Stanford University, 1984.