Taylor & Francis
Taylor & Francis Group

Visualization and unsupervised classification of changes in multispectral satellite imagery

MORTON J. CANTY*† and ALLAN A. NIELSEN‡

†Forschungszentrum Jülich, D-52425 Jülich, Germany
‡Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

The statistical techniques of multivariate alteration detection, minimum/ maximum autocorrelation factors transformation, expectation maximization and probabilistic label relaxation are combined in a unified scheme to visualize and to classify changes in multispectral satellite data. The methods are demonstrated with an example involving bitemporal LANDSAT TM imagery.

1. Introduction

To quote Singh's review article on change detection (Singh 1989): 'The basic premise in using remote sensing data for change detection is that changes in land cover must result in changes in radiance values … [which] must be large with respect to radiance changes from other factors'. When comparing multispectral images of a given scene taken at different times on a pixel-by-pixel basis, it is first necessary to correct the pixel intensities as much as possible for uninteresting differences such as those due to illumination and atmospheric conditions. The images being compared must also be co-registered to a high accuracy to avoid spurious signals resulting from registration errors. See for instance Richards and Jia (1999) and references therein.

Having performed the necessary preprocessing steps it is common to examine mathematical functions of the spectral bands involved (differences, ratios or linear combinations) which in some way bring the change information contained within them to the fore. We can define *unsupervised classification* of changes in this context as the estimation of probability distributions for change and no-change pixels in spectral band functions, in particular in the absence of a priori ground reference information. On the basis of the probabilities, appropriate labels may be assigned to the image pixels as being no-change or as belonging to one or more change categories.

In this paper we describe an analysis sequence, based on standard algorithms and essentially free of adjustable parameters, for the unsupervised classification of changes in bitemporal, multispectral satellite imagery. The procedure is founded on the *multivariate alteration detection* or MAD transformation for change enhancement proposed originally by Nielsen *et al.* (1998) which, because of its favourable statistical properties, provides a suitable starting point. The MAD transformation generates a set of mutually orthogonal difference images (MAD variates) of decreasing variance having the same spectral dimension as the original multispectral images to which it is applied. It is explained in §2, together with a further processing

---

step, the maximum autocorrelation factor (MAF) transformation, see Green *et al.* (1988), which serves to reduce dimensionality and to enhance spatial coherence as well as signal to noise ratio. In §3 the MAD/MAF variates are treated individually in terms of a mixture model for change and no-change. The parameters of the model are calculated with the expectation maximization (EM) algorithm to obtain better estimates of the variances of the no-change pixels and to set optimal decision thresholds. Section 4 extends this idea to cluster the change and no-change pixels in the full MAD/MAF feature space. Section 5 describes the use of probabilistic label relaxation (Richards and Jia 1999) to incorporate spatial information into the final change classification. In §6 examples of the techniques applied to the classification of changes in a LANDSAT TM bitemporal image are presented. Some conclusions are given in §7.

2. The MAD transformation

Briefly, the MAD transformation is as follows (see Nielsen *et al.* (1998) for a more detailed discussion). We first represent two N-dimensional multispectral images of a scene acquired at times t_1 and t_2 by random vectors $\mathbf{F}$, resp. $\mathbf{G}$, assumed to be multivariate normally distributed with zero mean. After forming the scalar difference

$$D = \mathbf{a}^\top \mathbf{F} - \mathbf{b}^\top \mathbf{G}, \tag{1}$$

the vectors $\mathbf{a}$ and $\mathbf{b}$ are chosen, analogously to the principal components transformation, so as to maximize the variance of D, with the additional constraints imposed that

$$\mathrm{var}(\mathbf{a}^\top \mathbf{F}) = var(\mathbf{b}^\top \mathbf{G}) = 1. \tag{2}$$

This means that the resulting difference image D will show maximum spread in its pixel intensities. If we assume that the spread is primarily due to actual changes that have taken place in the scene over the interval $t_2 - t_1$, then this procedure will enhance those changes as much as possible.

The determination of the $\mathbf{a}$ and $\mathbf{b}$ that satisfy (1) and (2) is equivalent to a canonical correlation analysis (CCA) which transforms each set of variables $\boldsymbol{F}$ and $\boldsymbol{G}$ such that their mutual correlation is displayed unambiguously, see Anderson (2003). CCA determines N eigenvalues ρ_i^2 and N pairs of eigenvectors $\mathbf{a}_i, \mathbf{b}_i$, $i=1,\ldots,N$ and, accordingly, N difference components of the form (1). These are referred to as the *MAD variates*. In addition, an iterative re-weighting scheme (Nielsen 2005) is used which establishes an increasingly better background of no-change against which to detect change. This is done by putting higher weights on observations of no-change in the calculation of the statistics for the CCA.

The quantities $\mathbf{a}_i^\mathrm{T}\mathbf{F}$ and $\mathbf{b}_i^\mathrm{T}\mathbf{G}$ are called *canonical variates*. The square root of the eigenvalues ρ_i are the correlations between them. The MAD variates themselves are orthogonal (uncorrelated):

$$\mathrm{cov}(D_i, D_j) = 2\delta_{ij}(1 - \rho_i) \tag{3}$$

and invariant under affine transformations. This invariance can be exploited, e.g. to determine time-invariant features for relative radiometric normalization (Canty *et al.* 2004). If the eigenvectors $\mathbf{a}_i, \mathbf{b}_i$, $i=1,\ldots,N$, are sorted according to increasing eigenvalue ρ_i^2, the MAD variates will be sorted according to decreasing variance.

The sum of squares of standardized variates

$$\chi_N^2 = \left(\frac{D_1}{\sigma_{D_1}}\right)^2 + \cdots + \left(\frac{D_N}{\sigma_{D_N}}\right)^2, \tag{4}$$

is approximately chi-square distributed with N degrees of freedom where σ_{oi} is the standard deviation of D_i.

The MAD transformation can be augmented by subsequent application of the minimum/maximum autocorrelation factors (MAF) transformation, in order to improve the spatial coherence of the difference components, see Nielsen *et al.* (1998). When image noise is estimated as the difference between intensities of neighbouring pixels, the MAF transformation is equivalent to a minimum noise fraction (MNF) transformation (Green *et al.* 1988), which generates image components with maximum signal to noise ratio (SNR). In fact, the eigenvalues λ_i of the MAF transformation are related to the SNR approximately by (Green *et al.* 1988)

$$\text{SNR} = \frac{2}{\lambda_i} - 1, \quad i = 1, \ldots, N. \tag{5}$$

The MAD/MAF variates thus generated are also orthogonal and invariant under affine transformations.

3. Decision thresholds

Decision thresholds for change pixels can be set in terms of standard deviations about the mean for each MAD/MAF component separately. This may be done subjectively, for example by saying that all pixels in an image component D whose intensities are in the interval $[-2\sigma_D, 2\sigma_D]$ are no-change pixels and constitutes what is usually understood as *change vector analysis*, see Singh (1989).

We can do better than this, however, using a Bayesian technique proposed by Bruzzone and Prieto (2000) in a similar context, which tries to minimize the misclassification error. Let us consider the following simple mixture model for the probability density $p(d)$ of a random variable D representing one of the MAD/MAF variates:

$$p(d) = p(d|NC)p(NC) + p(d|C-)p(C-) + p(d|C+)p(C+), \tag{6}$$

where $C+$, $C-$ and NC denote positive change, negative change and no change, respectively. Thus $p(d|NC)$ is the probability density of the no-change pixels and $p(NC)$ is the corresponding mixing coefficient, etc. The set of n observations $S = \{d_i | i = 1, \ldots, n\}$ may be grouped into four disjoint sets:

$$S_{NC}, \quad S_{C-}, \quad S_{C+}, \quad S_U = S \backslash S_{NC} \cup S_{C-} \cup S_{C+},$$

with S_U indicating the set of ambiguous pixels. (The symbols $\cup$ and $\backslash$ denote set union and set difference, respectively.) These sets can be determined in practice by setting generous, scene-independent thresholds for change and no-change pixel intensities, e.g. $-\sigma_D/2 < d_i < \sigma_D/2$ for *NC,* $d_i < -3\sigma_D$ and $d_i > 3\sigma_D$ for $C-$ and $C+$, respectively.

Let $p(NC|d_i)$ be the *a posteriori* probability for a no-change pixel conditional on measurement d_i. We have the following rules for determining $p(NC|d_i)$:

1. $d_i \in S_{NC}$: $p(NC|d_i)=1$
2. $d_i \in S_{C\pm}$: $p(NC|d_i)=0$
3. $d_i \in S_U$: $p(NC|d_i)=p(d_i|NC)p(NC)/p(d_i)$

Rule (3) follows from Bayes' Theorem, and we assume

$$p(d_i|NC) = \frac{1}{\sqrt{2\pi}\cdot\sigma_{NC}}\cdot exp\left(-\frac{(d_i-\mu_{NC})^2}{2\sigma_{NC}^2}\right).$$

Starting values for the moments and mixture coefficient σ_{NC}, and $p(nc)$ can be estimated from the set S_{NC}. Applying the EM algorithm, we then update the parameters according to the prescription

$$\mu'_{NC} = \sum_{i \in S} p(NC|d_i)d_i \Big/ \sum_{i \in S} p(NC|d_i)$$

$$(\sigma'_{NC})^2 = \sum_{i \in S} p(NC|d_i)(d_i - \mu'_{NC})^2 \Big/ \sum_{i \in S} p(NC|d_i) \tag{7}$$

$$p'(NC) = \frac{1}{|S|}\cdot\sum_{i \in S} p(NC|d_i),$$

where the prime denotes the new values. Equations (7) are derived e.g. in Bishop (1995). They can now be iterated together with the corresponding equations for $C+$ and $C-$ to estimate the change and no-change densities. One then determines the upper change threshold d_U which minimizes the probability of misclassification as a solution of

$$p(d|NC)p(NC) = p(d|C+)p(C+).$$

Taking logarithms,

$$\frac{1}{2\sigma_{C+}^2}\left(d-\mu_{C+}\right)^2 - \frac{1}{2\sigma_{NC}^2}(d-\mu_{NC})^2 = \log\left[\frac{\sigma_{NC}}{\sigma_{C+}}\cdot\frac{p(C+)}{P(NC)}\right] =: A$$

with solutions

$$d_U = \frac{\mu_{C+}\sigma_{NC}^2 - \mu_{NC}\sigma_{C+}^2 \pm \sigma_{NC}\sigma_{C+}\sqrt{(\mu_{NC}-\mu_{C+})^2 + 2A(\sigma_{NC}^2 - \sigma_{C+}^2)}}{\sigma_{NC}^2 - \sigma_{C+}^2}.$$

(The appropriate sign must be chosen.) A similar expression obtains for the lower threshold d_L. The values of any three MAD/MAF variates lying above d_U or below d_L can be conveniently displayed in RGB coordinates, thus giving a colour-coded representation of significant changes that have taken place between the two scenes. Since it is the no-change pixels which we expect to obey the chi-square distribution, see equation (4), we can replace the denominators σ_{D_i} with σ_{NC_i}, i.e.,

$$\chi_N^2 = \left(\frac{D_1}{\sigma_{NC_1}}\right)^2 + \cdots + \left(\frac{D_N}{\sigma_{NC_N}}\right)^2, \tag{8}$$

where $\sigma^2_{NC_i}$ is the variance of the no-change component for the ith MAD/MAF variate as determined in the mixture model. The probability that a given observation could exhibit a chi-square (8) or smaller can be interpreted as the change probability and be represented as a grey-scale image.

According to Bruzzone and Prieto (2002) good results can also be obtained with non-parametric, kernel-based initial estimates of the probability densities of the change and no-change pixels, again with the help of the EM algorithm. Rather than pursue this idea, however, we extend in the following the above parametric approach.

4. Clustering of changes

The procedure described in the preceding section has the disadvantage that only changes signalled by at most three MAD/MAF variates can be displayed at one time. Furthermore, only two change categories (negative and positive) are modelled in each variate. Clustering of the change pixels can of course be applied in the full dimensional feature space, where the number of clusters chosen determines the number of change categories. The approximate chi-square distribution of the sum of squares of the standardized variates (4) allows the labelling of pixels with high no-change probability. These can be excluded from the clustering process in the same manner as in the preceding section, i.e. by 'freezing' their membership probabilities to 1 for the no-change class.

We apply the algorithm of Gath and Geva (1989), which they refer to as *fuzzy maximum likelihood estimation* (FMLE), and which is essentially equivalent to the EM algorithm; see, e.g. Hilger (2001). We begin by defining the change pixels as those having a chi-square exceeding a threshold

$$t = \chi^2_{N,P},\tag{9}$$

where P is the probability of observing a value lower than t (typically we choose $P=0.99$). Denote the MAD/MAF observations by the vectors

$$\mathbf{d}_i, \quad i=1,\dots,n$$

and let u_{ki} be the membership probability of the ith pixel to the kth cluster. If there are K change clusters in all, with $k=0$ denoting the no-change cluster, we have

$$\sum_{k=0}^{K} u_{ki} = 1, \quad i=1,\dots,n,\tag{10}$$

and the probability of cluster k is given by

$$P(k) = \frac{1}{n}\sum_{i=1}^{n} u_{ki}, \quad k=0,\dots,K,\tag{11}$$

The cluster sample means $\mathbf{m}_k$ and sample covariance matrices $\mathbf{F}_k$ can then be written in terms of the membership probabilities as

$$\mathbf{m}_k = \frac{1}{nP(k)}\sum_{i=1}^{n} u_{ki}\,\mathbf{d}_i$$

$$\mathbf{F}_k = \frac{1}{nP(k)}\sum_{i=1}^{n} u_{ki}(\mathbf{d}_i - \mathbf{m}_k)(\mathbf{d}_i - \mathbf{m}_k)^\top, \quad k=0,\dots,K.\tag{12}$$

These in turn themselves determine the membership probabilities according to

$$u_{ki} = P(k \mid \mathbf{d}_i),$$

the posterior probability for cluster k conditional upon observation d_i. That is, invoking Bayes' Theorem and assuming each cluster to have a multivariate normal density distribution, we have

$$u_{ki} = P(k \mid \mathbf{d}_i) = CP(k)P(\mathbf{d}_i \mid k)$$

$$= CP(k)\frac{1}{\sqrt{\mid \mathbf{F}_k \mid}} exp\left(-\frac{1}{2}(\mathbf{d}_i - \mathbf{m}_k)^{\top}\mathbf{F}_k^{-1}(\mathbf{d}_i - \mathbf{m}_k)\right), \tag{13}$$

where C is a normalization constant independent of k which can be determined from (10). Starting from some initial choice of the memberships u_{ki}, the FMLE algorithm consists of the iteration of equations (10)–(13) until convergence. In this calculation the no-change pixels, as determined from the chi-square threshold (9), have their class memberships frozen in order to reduce the number of iterations necessary for convergence. Finally, all pixels are labelled according to the maximum membership probability:

$$\mathrm{label}(i) = arg \max_{k \in \{0,\dots,K\}} u_{ki}. \tag{14}$$

Because of the exponential distance dependence of the fuzzy cluster membership in (13), the algorithm is very sensitive to initialization conditions, and can even become unstable. To avoid this problem we follow a suggestion of Gath and Geva (1989) and first obtain initial values for the u_{ki} by preceding the calculation with a fuzzy C-means clustering for which the class memberships follow an inverse square law (Dunn 1973).

In order to discriminate among different choices for the number of change clusters we again follow Gath and Geva (1989) and first define a 'fuzzy hypervolume' given by

$$F_{HV} = \sum_{k=0}^{K} \sqrt{\mid \mathbf{F}_k \mid}.$$

The *partition density* P_D is then defined according to

$$P_D = S/F_{HV}, \tag{15}$$

where S is the number of change pixels within unit Mahalanobis distance of a cluster prototype:

$$S = \sum_{i \in R}^{n} \sum_{k=0}^{K} u_{ki}, \quad R = \left\{i \mid (\mathbf{d}_i - \mathbf{m}_k)^{\top}\mathbf{F}_k^{-1}(\mathbf{d}_i - \mathbf{m}_k) < 1\right\}.$$

An optimal choice for K is one which maximizes the partition density.

5. Post-classification processing

With Richards and Jia (1999) we note that a possible misclassification of a pixel in (14) could in principle be corrected by examining the membership probabilities of

pixels in its neighbourhood. They describe a method referred to as *probabilistic label relaxation*, which we have adapted here to improve potentially our classification result and take spatial information into account.

Let $Q_m(k)$ be a *neighbourhood function* for some pixel m which corrects its cluster membership probability u_{km} according to the prescription

$$u'_{km} = u_{km} \frac{Q_m(k)}{\sum_{j=1}^{K} u_{jm} Q_m(j)}, \quad k=1,\ldots,K,$$

or, written as a vector equation,

$$\mathbf{u}'_m = \mathbf{u}_m \frac{\mathbf{Q}_m}{\mathbf{u}_m^\top \mathbf{Q}_m}, \tag{16}$$

where $\mathbf{u}_m = (u_{0m},\ldots,u_{Km})^\mathrm{T}$ and $\mathbf{Q}_m = (Q_m(0),\ldots,Q_m(K))^\mathrm{T}$. The denominator ensures that the corrected result is also a probability, i.e. that

$$\sum_{k=1}^{K} u'_{km} = 1.$$

The neighbourhood function should reflect the spatial coherence of the image. In order to define that function, a *compatibility measure* for a neighbourhood N of pixel m is first postulated:

$$P_{mn}(k|\ell), \quad n \in N,$$

namely, the conditional probability that pixel m belongs to class k, given that a neighbouring pixel $n \in N$ belongs to class l. Evidence that m should be classified to k would then be (Richards and Jia 1999)

$$P_{mn}(k|\ell)u_{\ell n}, \quad n \in N. \tag{17}$$

This is the conditional probability that pixel m is in class k if neighbouring pixel n is in class l, multiplied by the probability that n actually is in class l. The neighbourhood function $Q_m(k)$ is obtained by summing over all terms like (17):

$$\mathbf{Q}_m(k) = \sum_{\ell=1}^{K} P_{mn}(k|\ell)u_{\ell n}$$

or in vector notation,

$$\mathbf{Q}_m = \mathbf{P}_{mn} \mathbf{u}_n,$$

where $\mathbf{P}_{mn}$ is a $K \times K$ matrix of compatibility measures. Equation (16) can thus be written finally as

$$\mathbf{u}'_m = \mathbf{u}_m \otimes \frac{\mathbf{P}_{mn} \mathbf{u}_n}{\mathbf{u}_m^\top \mathbf{P}_{mn} \mathbf{u}_n}, \tag{18}$$

where $\otimes$ indicates component-by-component multiplication. The matrix $\mathbf{P}_{mn}$ is easily estimated directly from the originally classified image.

The probabilistic label relaxation procedure can be iterated arbitrarily. However too many iterations may lead to a widening of the effective neighbourhood of a pixel

to such an extent that irrelevant spatial information may falsify the final classification. Experience shows that the best results are obtained after 3–4 iterations, see Richards and Jia (1999).

6.   Application to bitemporal satellite imagery

The classification procedures described in the foregoing sections will be illustrated with a bitemporal multispectral image. The image chosen, figures 1 and 2, consists of two Landsat 5 TM scenes acquired on March 29 and then again on May 16, 1998 over a semi-arid agricultural area in Hindustan, India.

The scenes were registered to one another by applying an automatic contour matching algorithm due to Li *et al.* (1995) and using first-order polynomial, nearest-neighbour resampling. The RMS errors were less than or equal to 0.5 pixel. Because of the robustness of the MAD transformation, no radiometric corrections were applied to the two scenes. The most evident change that took place between the acquisitions is the shallow flooding at the western end of the reservoir.

After performing the MAD/MAF transformations as described in §2, the MAD/MAF variates were ordered according to decreasing signal to noise ratio, see equation (5). The values obtained are shown in table 1. The last three components with low signal to noise were eliminated from subsequent analysis.

Figure 3 shows a scatter plot of the two MAD/MAF variates derived from the change components with highest signal to noise ratios, i.e. components 1 and 2.

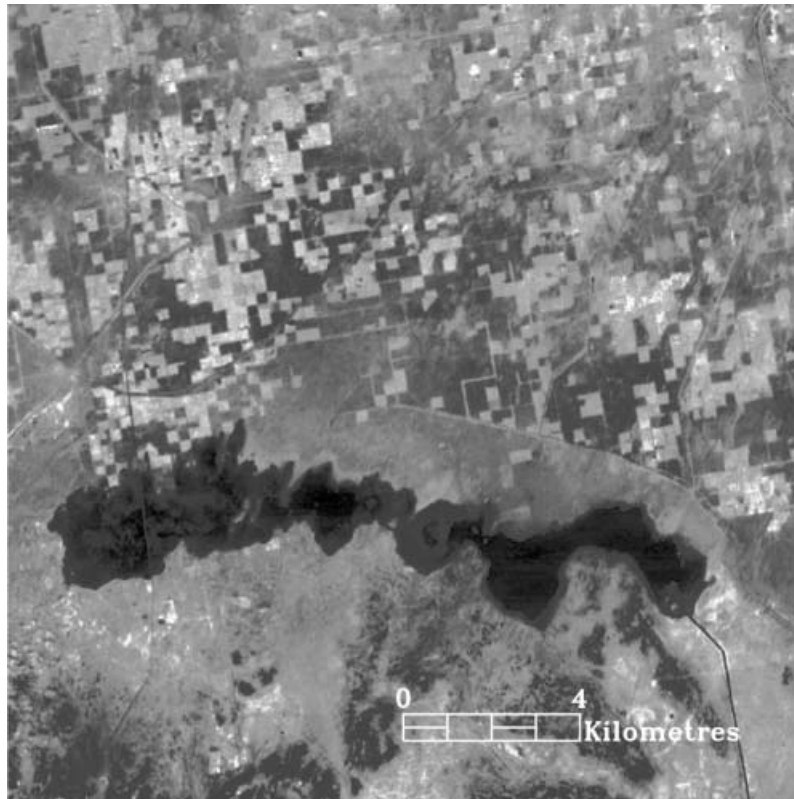Figure 1.   LANDSAT 5 TM image (band 4, Gaussian stretch) acquired on 29 March 1998.

Figure 2.   The same scene as in figure 1 but acquired on 16 May 1998.

Change pixels scattered away from the intense no-change cluster centred at the origin are clearly evident.

We next illustrate the automatic determination of decision thresholds as discussed in §3. Figure 4 shows a fit of the mixture model (5) to a MAD/MAF variate. The thresholds are chosen at the left and right intersections of the three normal distribution functions.

In figure 5 we see the result of thresholding the difference components with the automatically determined decision thresholds. A MAD/MAF component lying within the lower and upper no-change thresholds is coloured middle grey (RGB intensity 127). This serves to make negative changes more visible. Moreover the pixel intensities are stretched across ± 16 (!) standard deviations in the respective component so that changes less than or exceeding this amount are saturated at 0 or 255.

Table 1. Signal to noise ratios.

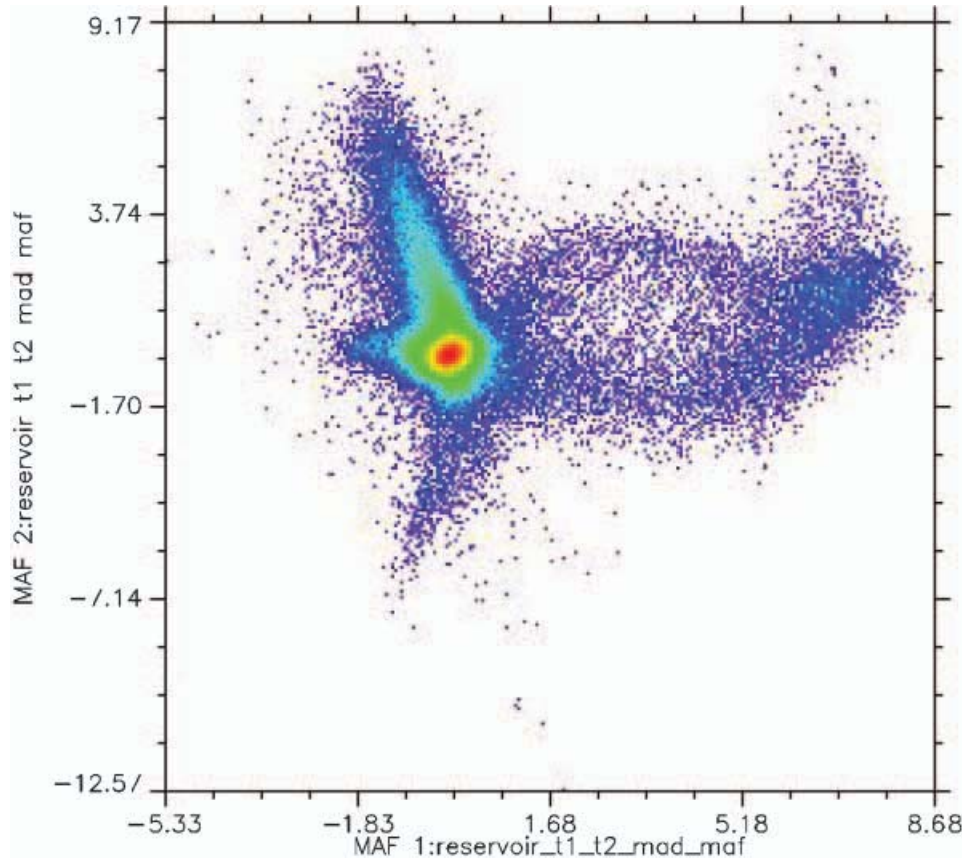| MAD/MAF	SNR
1	45.4
2	13.4
3	6.8
4	2.4
5	2.0
6	1.7

Figure 3.   Scatter plot of MAD/MAF variates 2 vs. 1.

In addition to the changes associated with the lake level and water quality, substantial changes in the agricultural fields to the north of the lake are evident. These changes are predominantly in intense green, that is in MAD/MAF variate 2. Figure 6 shows the correlation of the three variates with the earlier LANDSAT
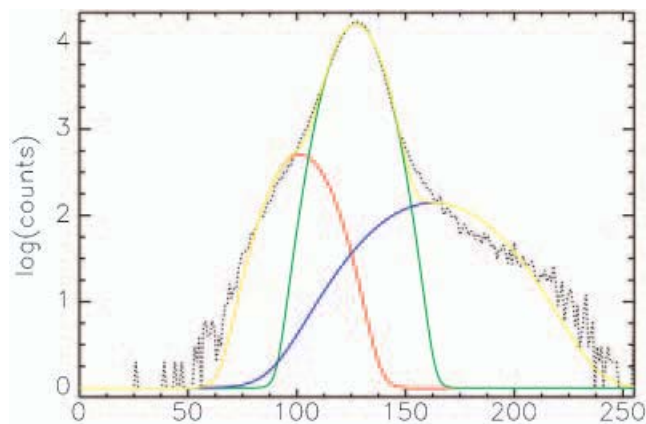

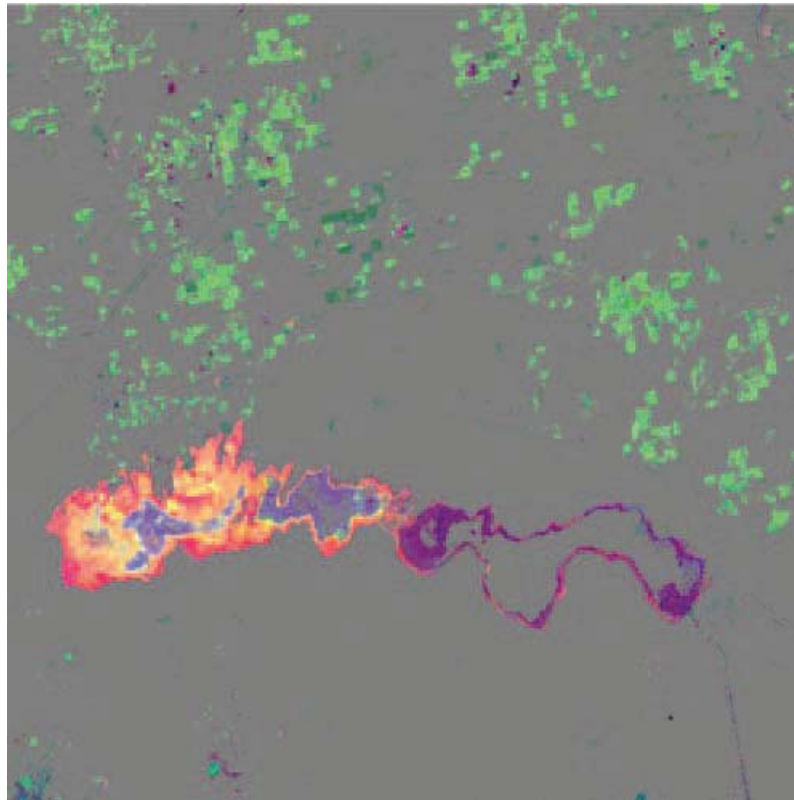Figure 4.   Mixture model fit to the histogram of a MAD/MAF variate.

Figure 5. Change image with thresholds on MAD/MAF components 1, 2 and 3 displayed in red, green and blue, respectively.
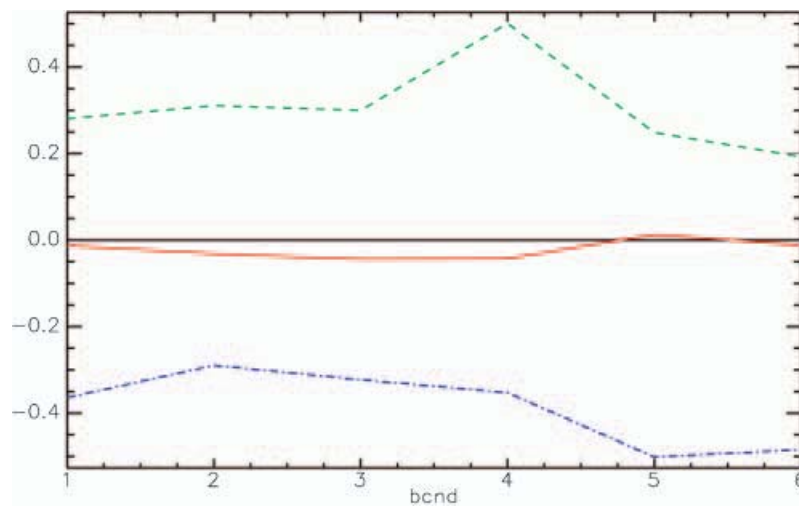

Figure 6. Correlation of three MAD/MAF variates with t_1 scene, 1 red, 2 green and 3 blue.

image acquired on March 29, 1998 (figure 1). Note that variate 2 is most strongly correlated with the Landsat TM vegetation band 4, as is to be expected.

Next we consider clustering the no-change pixels with the procedure of §4. The first question is the appropriate number of clusters to use. Figure 7 displays the partition density, equation (15), as a function of number of change clusters ranging from 1 to 8, determined after convergence of the FMLE algorithm. A fairly convincing local maximum is evident for five change clusters and this number was accordingly chosen.

Figure 8 shows the clustered change pixels after convergence for five change clusters, projected onto the 1–2 MAD/MAF plane. It is to be compared with the scatter plot in figure 3. The no-change cluster is shown in black.

Figure 9 illustrates the effect of post-classification processing with probabilistic label relaxation (§5). The spatial coherence of the change classes improves remarkably.

The final product, that is, the unsupervised change classification including post-classification, is shown in figure 10. By complementing this representation with ground reference data, the five classes may be assigned to real change categories, thereby giving a quantitative analysis of ground cover changes that have taken place.

7.   Conclusion

We have introduced and demonstrated a generally applicable and straightforward method for visualization and classification of changes in multispectral satellite imagery, based on well-established algorithms in image analysis and pattern recognition and building upon the robustness of the MAD and MAF transformations for extracting change information from bitemporal multispectral data and for enhancing signal to noise.
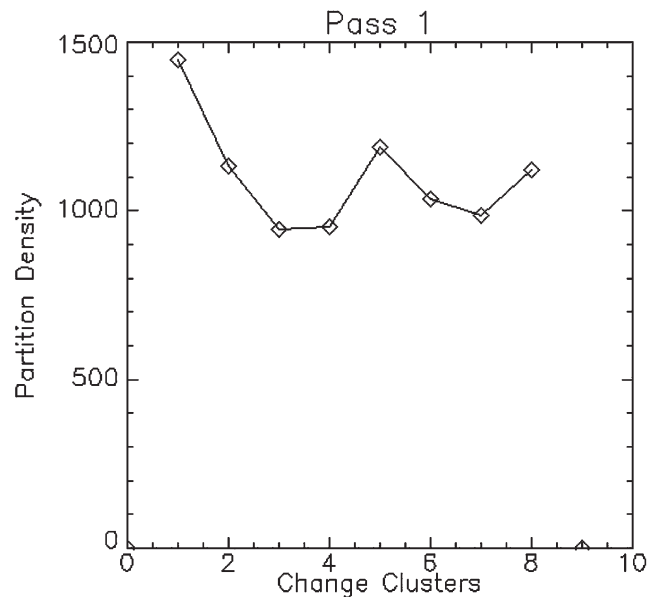

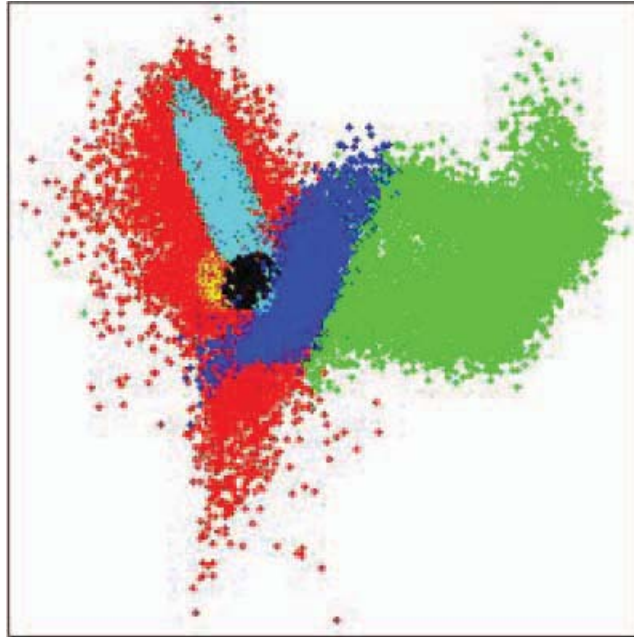Figure 7.   Plot of partition density vs. cluster number.

Figure 8.   Projection of five change clusters onto the 1–2 MAD/MAF plane.

For visualization we applied a probability mixture model and the EM algorithm to variates of the MAD/MAF transformations to allow the automatic determination of optimal (error-minimizing) decision thresholds for component-wise discrimination of change from no-change pixels. The result was a colour-coded representation of significant changes that had taken place. This procedure is closely related to change vector analysis (Singh 1989), a technique most commonly used with untransformed spectral bands or with the tasselled cap transformation. In such cases a direct physical interpretation of the change signals is sometimes possible. Invariably, however, the change–no-change decision thresholds are treated as adjustable parameters. The automatic threshold determination procedures of §3 could clearly be applied here as well, although we are not aware of anyone having
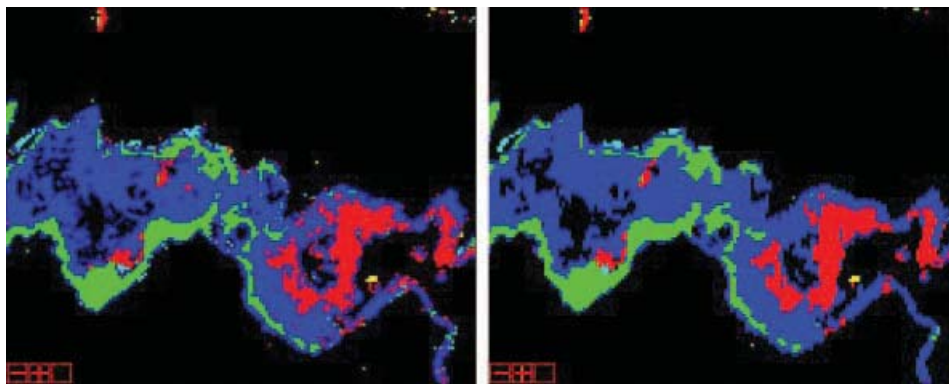

Figure 9.   Portion of clustered image before (left) and after four iterations of probabilistic label relaxation.
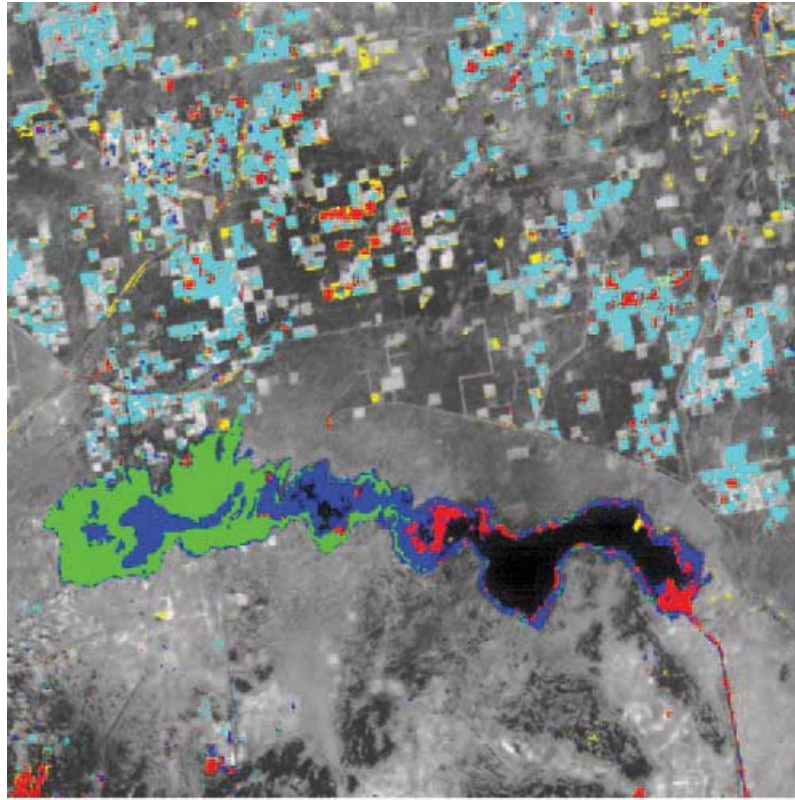
Figure 10. Final classification of changes with five change clusters.

done so. We have preferred to use the MAD/MAF transformed variables because of their robustness, their orthogonality, their approximate normal distribution and their ability to enhance signal to noise. Physical interpretation of change can still be accomplished by examining correlations with the spectral bands in the original scenes, as was illustrated in figure 6, see also Nielsen *et al.* (1998).

We then performed an unsupervised classification of change and no-change pixels with the FMLE method of Gath and Geva (1989), a variant of the EM algorithm, which allowed for hyper-ellipsoidal clusters and clusters of differing sizes, and included a criterion for choosing the best number of classes. Apart from the choice of the chi-square percentile for fixation of no-change pixels, no other adjustable parameters were involved. Advantage was then taken of the probabilistic interpretation of class memberships established by the FMLE algorithm to apply post-classification processing to improve the spatial coherence of the change classes obtained.

Quantitative comparisons of unsupervised classification methods are notoriously difficult, see for example Duda and Canty (2002). In any case we are unaware of other work which could be critically compared with the clustering approach described here.

References

ANDERSON, T.W., 2003, *An Introduction to Multivariate Statistical Analysis*, 3rd Edn. Wiley Series in Probability and Statistics (New York: John Wiley).

BISHOP, C.M., 1995, *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press).

BRUZZONE, L. and PRIETO, D.F., 2000, Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, pp. 1171–1182.

BRUZZONE, L. and PRIETO, D.F., 2002, An adaptive semi-parametric and context-based approach to unsupervised change detection in multitemporal remote sensing images. Technical Report No. DIT-02-0030, Department of Information and Communication Technology, University of Trento.

CANTY, M.J., NIELSEN, A.A. and SCHMIDT, M., 2004, Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sensing of Environment*, **91**, pp. 441–451. Internet: http://www.imm.dtu.dk/pubdb/p.php?123.

DUDA, T. and CANTY, M.J., 2002, Unsupervised classification of satellite imagery: choosing a good algorithm. *International Journal of Remote Sensing*, **23**, pp. 2193–2212.

DUNN, J.C., 1973, A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, PAM1-1, pp. 32–57.

GATH, I. and GEVA, A.B., 1989, Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intellegence*, **3**, pp. 773–781.

GREEN, A.A., BERMAN, M., SWITZER, P. and CRAIG, M.D., 1988, A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, **26**, pp. 65–74.

HILGER, K.B., 2001, Exploratory Analysis of Multivariate Data. PhD Thesis, Technical University of Denmark.

LI, H., MANJUNATH, B.S. and MITRA, S.K., 1995, A contour-based approach to multisensor image registration. *IEEE Transactions on Image Processing*, **4**, pp. 320–334.

NIELSEN, A.A., 2005, Iteratively re-weighted multivariate alteration detection in multi- and hyperspectral data. Submitted for publication.

NIELSEN, A.A., CONRADSEN, K. and SIMPSON, J.J., 1998, Multivariate alteration detection (MAD) and MAF processing in multispectral, bitemporal image data: new approaches to change detection studies. *Remote Sensing of Environment*, **64**, pp. 1–19. Internet: http://www.imm.dtu.dk/pubdb/p.php?123.

RICHARDS, J.A. and JIA, X., 1999, *Remote Sensing Digital Image Analysis* (Berlin: Springer-Verlag).

SINGH, A., 1989, Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, **10**, pp. 989–1002.