Investigation of Alternative Iteration Schemes for the IR-MAD Algorithm

Morton J. Canty[a] and Allan A. Nielsen[b]

[a]Institute for Chemistry and Dynamics of the Geosphere, Jülich Research Center, D-52425 Jülich, Germany, E-mail: m.canty@fz-juelich.de;
[b]Danish National Space Center, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark, E-mail: aa@space.dtu.dk

ABSTRACT

The iteratively re-weighted multivariate alteration detection (IR-MAD) transformation is proving to be very successful for multispectral change detection and automatic radiometric normalization applications in remote sensing. Various alternatives exist in the way in which the weights (no-change probabilities) are calculated during the iteration procedure. These alternatives are compared quantitatively on the basis of multispectral imagery from different sensors under a range of ground cover conditions exhibiting wide variations in the amount of change present, as well as with a partially artificial data set simulating truly time-invariant observations. A best re-weighting procedure is recommended.

Keywords: change detection, canonical correlations, radiometric normalization, IR-MAD

1. INTRODUCTION

The Multivariate Alteration Detection (MAD) algorithm consists of the canonical correlation analysis (CCA) of two co-registered, multi-spectral images followed by pairwise subtraction of the canonical variates.[1] Specifically, let the N-dimensional random vectors for the bitemporal scene be multivariate normally distributed with zero mean:

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_X), \quad \boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_Y), \quad \boldsymbol{\Sigma} = \left\langle \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \left(\boldsymbol{X}^\top, \boldsymbol{Y}^\top \right) \right\rangle = \begin{pmatrix} \boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_Y \end{pmatrix}. \tag{1}$$

The MAD variates are

$$\boldsymbol{M} = \boldsymbol{A}\boldsymbol{X} - \boldsymbol{B}\boldsymbol{Y}, \tag{2}$$

where $\boldsymbol{A} = (\boldsymbol{a}_N \dots \boldsymbol{a}_1)$, $\boldsymbol{B} = (\boldsymbol{b}_N \dots \boldsymbol{b}_1)$ and where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are the eigenvectors of the coupled, generalized eigenvalue problem

$$\begin{aligned} \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{XY}^\top \boldsymbol{a} &= \rho^2 \boldsymbol{\Sigma}_{XX}\boldsymbol{a} \\ \boldsymbol{\Sigma}_{XY}^\top\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} \boldsymbol{b} &= \rho^2 \boldsymbol{\Sigma}_{YY}\boldsymbol{b} \end{aligned} \tag{3}$$

with corresponding eigenvalues $\rho_1^2 \geq \rho_2^2 \geq \dots \geq \rho_N^2$. The MAD variates are mutually uncorrelated with covariance matrix

$$\boldsymbol{\Sigma}_M = \operatorname{diag}\left(2(1-\rho_N) \dots 2(1-\rho_1)\right). \tag{4}$$

Whereas the components of the original images are ordered by wavelength, the MAD variates order the change information in the bitemporal data according to similarity, i.e., correlation.

The MAD variares have nice statistical properties. For example they are invariant under affine transformations of the original images $\boldsymbol{X}$ and/or $\boldsymbol{Y}$, see 1, making the MAD transformation very suitable for automatic radiometric normalization.[2] Moreover, since the no-change observations are approximately normally distributed and uncorrelated, the sum of the squares of the standardized MAD variates, described by the random variable

$$Z = \sum_{i=1}^{N} \left(\frac{M_i}{\sigma_{M_i}} \right)^2, \tag{5}$$

will be approximately chi-square distributed with N degrees of freedom (distribution function $P_{\chi^2;N}(z)$).

Image and Signal Processing for Remote Sensing XIII, edited by Lorenzo Bruzzone,
Proc. of SPIE Vol. 6748, 674808, (2007) · 0277-786X/07/$18 · doi: 10.1117/12.736409

In 3 advantage was taken of the latter property in order to propose an iterative re-weighting scheme to further enhance the change information: In the first iteration all observations are given equal weights. Thereafter the pixels are weighted according to their probability of no change as determined by the chi-square distribution:

$$\Pr(\text{no change}) = 1 - P_{\chi^2;N}(z), \tag{6}$$

where z is a realization of the random variable Z corresponding to a given pixel, and where Pr(no change) is the probability that a sample z drawn from the chi-square distribution could be that large or larger. A small z thus implies a large probability of no change. The transformation is then repeated, iteration continuing until some stopping criterion is fulfilled, such as no further change in the eigenvalues ρ_i^2. This procedure provides an increasingly better background of no change against which to measure change.[3]

An ambiguity arises in the method to be used to standardize the MAD variates on each iteration. In 3 the standard deviations σ_{M_i}, $i = 1 \ldots N$, are re-estimated from the recalculated MAD variates on each iteration. This is referred to as method A in the sequel. In 4 and 5 it is suggested to use the standard deviations determined by CCA, i.e., the diagonal elements in Equation (4), as these should more closely reflect the distribution of no-change pixels as the iteration progresses. This we shall refer to as method B. Alternatively one can apply, after each iteration, a clustering algorithm to discriminate change from no-change observations. In particular, for clustering based upon a Gaussian probability mixture model, the standard deviations for iterative re-weighting could be read from the diagonal elements of the covariance matrix for the no-change probability density. The parameters of such a mixture model may be conveniently estimated with the EM algorithm.[6] However, since the clustering algorithm estimates both no-change *and* change distributions, one can dispense with using the chi-square values altogether and choose the no-change probability weights as the probability of membership to the no-change cluster, see 7. This we call method C.

In the present paper these three alternatives A, B and C are compared quantitatively on the basis of multispectral imagery from different sensors under a range of ground cover conditions exhibiting wide variations in the amount of change present, as well as with a partially artificial data set simulating truly time-invariant pixels. In Section 2 the bitemporal images and associated preprocessing are described, the procedure chosen for comparing the iteration schemes is given in Section 3, results are presented in Section 4 and conclusions are drawn in Section 5.

2. IMAGERY

Three bitemporal satellite images were used for the study: two LANDSAT ETM+ scenes (729×754 pixels) over an arid area in Morocco (Figure 1), two 1000×1000-pixel ASTER scenes over an area southwest of Esfahan in Iran (the data had been processed to "level 1b - registered radiance at the sensor" and are shown in Figure 2) and two LANDSAT ETM+ images (1000×1000 pixels) over the town of Jülich, Germany (Figure 3). Extensive ground reference data were available for the Jülich acquisitions.

The image sets were registered to one another by applying an automatic contour matching algorithm[8] and using first-order polynomial, nearest-neighbor resampling. The RMS errors were less than 0.5 pixel. In addition, the six short wave infrared (SWIR) bands of the ASTER images were sharpened to the 15m resolution of the three visual near infrared (VNIR) bands with a wavelet fusion technique.[9] The ASTER images thus consisted of nine spectral bands each.

A partially artificial data set was generated from the two ASTER images shown in Figure 2 by copying a $316 \times 316 \times 9$ spatial/spectral subset from the upper left hand corner of the July 2001 image to the upper left hand corner of the May 2005 image, after adding a small amount of Gaussian noise (with standard deviation about 1% of the mean pixel intensity). The Gaussian noise prevents the solution of the coupled eigenvalue problem (3) from becoming degenerate once the no-change pixels have been selected. The subset chosen corresponds to a no-change fraction of 10% of the entire image.

All image processing was performed within the ENVI remote sensing image analysis environment (ITT Visual Information Solutions). Extensions to ENVI for image registration, wavelet fusion and the IR-MAD transformation were written in the IDL language, see 4 for software availability.

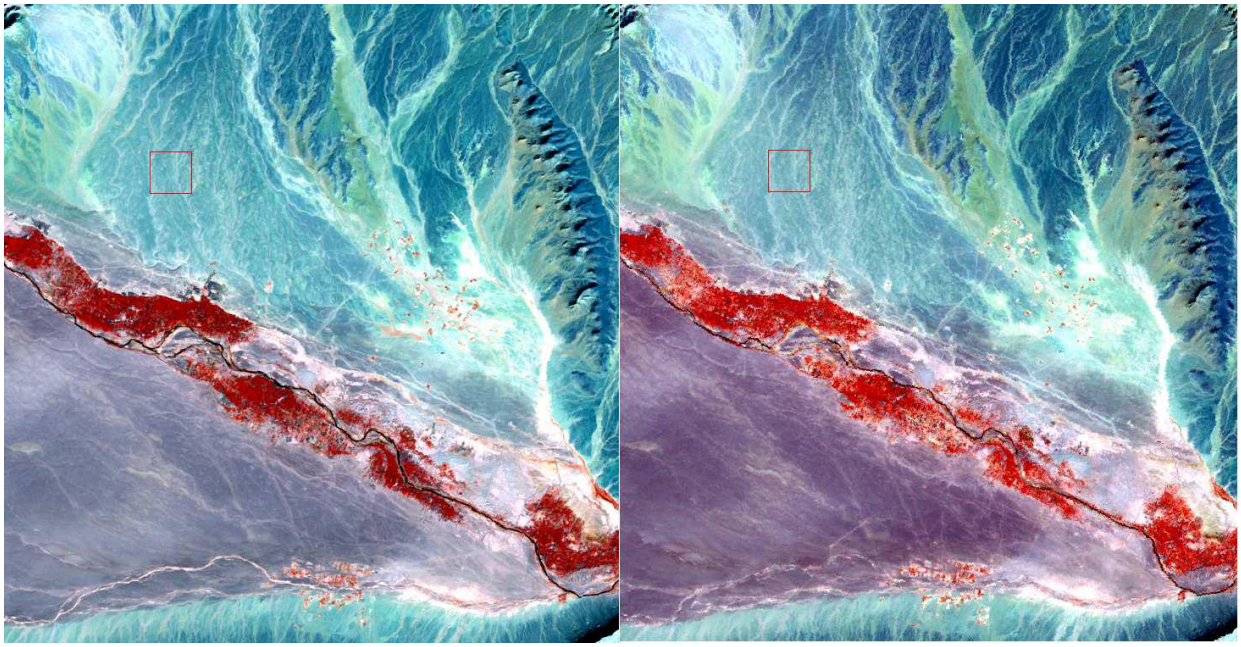

Figure 1. Landsat ETM+ images (729×754, RGB composite of bands 4,5,7) over east central Morocco, left: December 19, 1999, right: October 18, 2000.

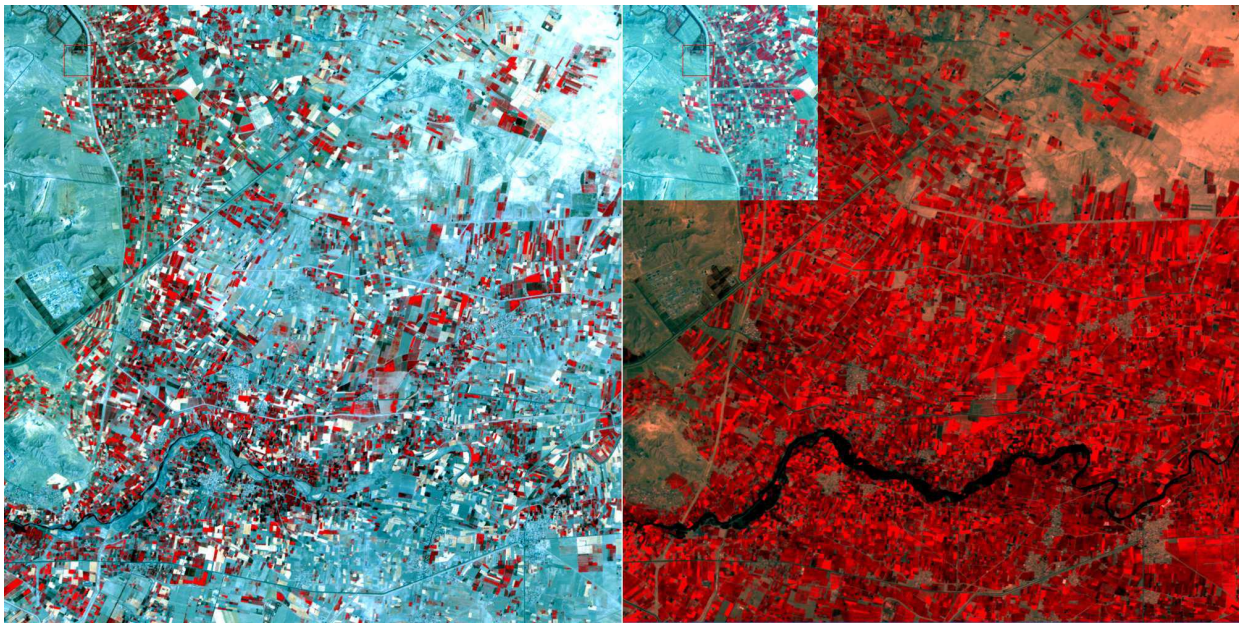

Figure 2. ASTER images (1000×1000, RGB composite of bands 3,2,1) near Esfahan, Iran, left: July 30, 2001, right: May 22, 2005. For some experiments, a $316 \times 316 \times 9$ spatial/spectral subset was copied from the left hand to the right hand image.

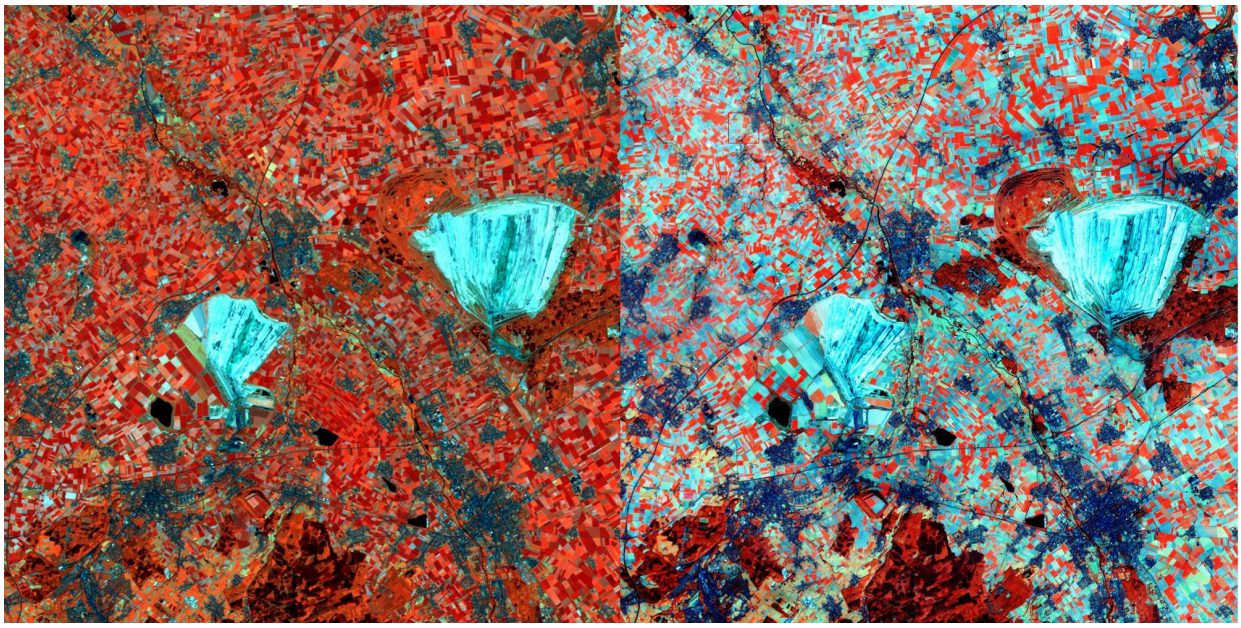

Figure 3. Landsat ETM+ images (1000×1000, RGB composite of bands 4,5,7) over Jülich, Germany, left: June 26, 2001, right: August 29, 2001.

3. METHOD OF COMPARISON

Ideally the accuracy of any change detection procedure should be assessed on the basis of reliable ground reference data. Alternatively, in the absence of such information, one can examine the observations identified by the method as unchanged. These should be highly correlated in the two scenes being compared. This will in fact form the basis of our comparison: band-wise regressions of the pixels having a high no-change probability as determined from Equation (6) are calculated and correlation coefficients (Pearson's r) and RMS errors are compared for the three methods A, B and C. Orthogonal regression, rather than ordinary linear regression, is the preferred method for calculating the regression coefficients[2] and is the one chosen here. In the case of the images over Jülich, considerable ground reference information was available and aided the evaluation. Of course the no-change observations for the simulated data are known *a priori*.

4. RESULTS

In the following, comparison of the three iteration methods is made first on the basis of real LANDSAT ETM+ and ASTER data, then with the partially artificial data generated from the ASTER imagery.

4.1 Real data

Tables 1, 2 and 3 give, band-wise, the correlation coefficients r and the RMS errors for orthogonal regression fits to no-change pixels identified by the IR-MAD transformation for the three bitemporal images and for each of the three methods A, B and C. Figures 4, 5 and 6 show corresponding scatterplots and regression lines for spectral band 4 for each bitemporal image. The no-change pixels were selected using Equation (6) and setting a lower threshold of 0.95 on Pr(no change). The number of no-change pixels actually selected differs greatly among the methods A, B and C. This will be discussed in Section 5 below.

In all three cases, method B is seen to perform better than A or C. The r values for method B are always close to unity and the RMS errors are, with the exception of the RMSE entries in the first row in Table 1, smallest. Only in the case of the Morocco scenes (Table 1 and Figure 1) do methods A and C perform satisfactorily. This is an arid terrain with a large proportion of no-change. The Iranian and German scenes, on the other hand, involve very considerable ground reflectance changes due to seasonal vegetation differences and agricultural activity. Here, methods A and C perform poorly.

Table 1. Comparison of correlation r and RMS error for the Morocco images.

Method	A		B		C	
Pixels	45285		1527		20736	
Band	r	RMSE	r	RMSE	r	RMSE
1	0.9478	0.8456	0.9872	0.8839	0.9547	0.7456
2	0.9651	1.1137	0.9771	0.8216	0.9720	0.9479
3	0.9708	1.7643	0.9830	1.2360	0.9787	1.4424
4	0.9756	1.3048	0.9864	0.8922	0.9820	1.0757
5	0.9723	1.7741	0.9889	1.0765	0.9817	1.3972
7	0.9660	1.5651	0.9852	0.9609	0.9765	1.2460

Table 2. Comparison of correlation r and RMS error for the Iran images.

Method	A		B		C	
Pixels	70951		293		46207	
Band	r	RMSE	r	RMSE	r	RMSE
1	0.7852	8.9709	0.9661	0.6171	0.6778	9.0124
2	0.7870	11.3843	0.9871	1.0935	0.6937	11.0410
3	0.8725	4.3727	0.9861	1.0779	0.9240	3.4391
4	0.8077	6.1143	0.9769	0.8465	0.7566	5.3718
5	0.8452	5.6042	0.9517	0.8139	0.7528	4.1831
6	0.8326	6.4759	0.9595	0.9168	0.7421	4.8424
7	0.8129	0.9705	0.7252	0.9168	0.6585	5.0212
8	0.7377	8.0546	0.9827	0.9228	0.5413	7.9596
9	0.8106	6.2123	0.9271	0.8815	0.6975	4.2707

Table 3. Comparison of correlation r and RMS error for the Germany images.

Method	A		B		C	
Pixels	144667		1182		41559	
Band	r	RMSE	r	RMSE	r	RMSE
1	0.5768	5.5472	0.9701	1.4583	0.6285	4.7324
2	0.4874	7.9934	0.9802	1.3355	0.5323	6.8752
3	0.3454	13.9987	0.9866	1.6138	0.3915	11.4347
4	0.8660	7.6981	0.9727	3.1040	0.8903	6.8301
5	0.3831	16.0136	0.9888	1.6384	0.2891	13.6277
7	0.3291	15.9422	0.9944	1.1007	0.2784	13.1595

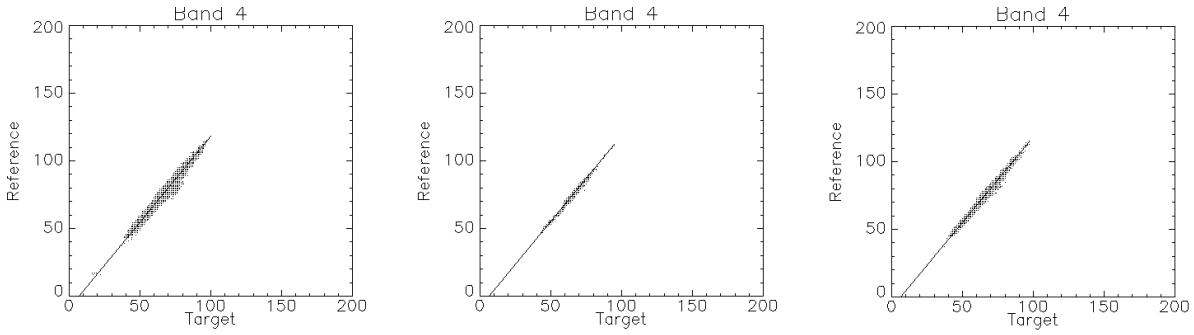

Figure 4. Orthogonal regressions for bands 4 of the Morocco images, left: method A, center: method B, right: method C.

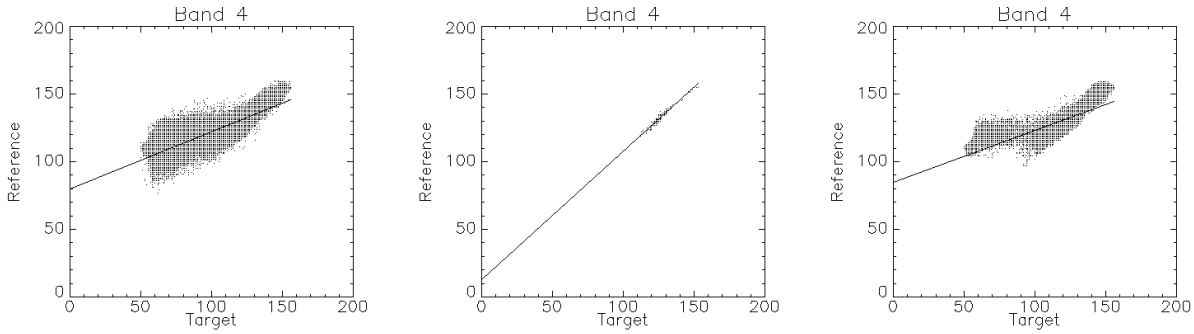

Figure 5. Orthogonal regressions for bands 4 of the Iran images, left: method A, center: method B, right: method C.

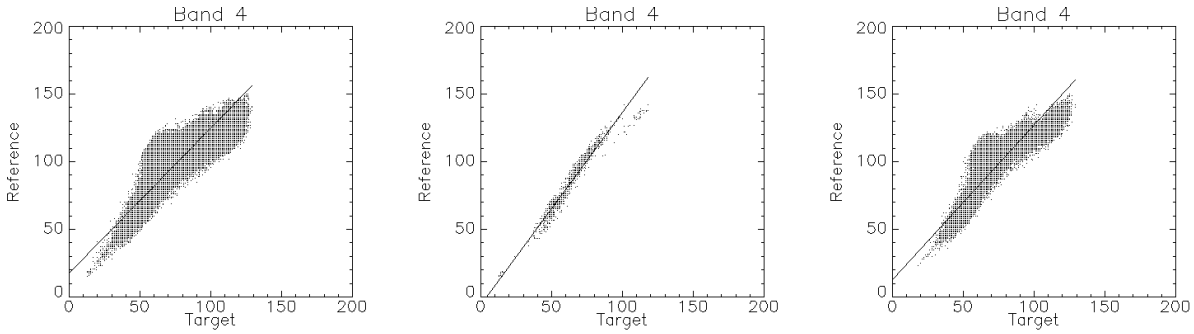

Figure 6. Orthogonal regressions for bands 4 of the Germany images, left: method A, center: method B, right: method C.

4.2 Partially artificial data

Table 4 similarly gives the correlation coefficients r and the RMS errors for orthogonal regression fits to no-change pixels for the partially artificial data set generated from the Iran data for the three methods A, B and C. Scatterplots and regression lines are shown in Figure 7.

Table 4. Comparison of correlation r and RMS error for the partially artificial data.

Method	A		B		C	
Pixels	65650		395		29171	
Band	r	RMSE	r	RMSE	r	RMSE
1	-0.0462	26.5316	0.9999	0.2904	0.7415	15.1568
2	0.2822	25.2415	0.9999	0.2920	0.8769	11.5564
3	0.8792	4.3860	0.9995	0.2963	0.9970	0.7047
4	0.7013	7.9720	0.9995	0.3048	0.9958	0.8976
5	0.6385	8.6213	0.9994	0.3850	0.9869	1.8517
6	0.6370	9.5818	0.9996	0.3892	0.9900	1.8543
7	0.6306	8.9946	0.9995	0.3693	0.9923	1.4028
8	0.5484	11.0874	0.9996	0.3183	0.9874	1.8313
9	0.5007	10.1485	0.9996	0.3790	0.9684	3.1666

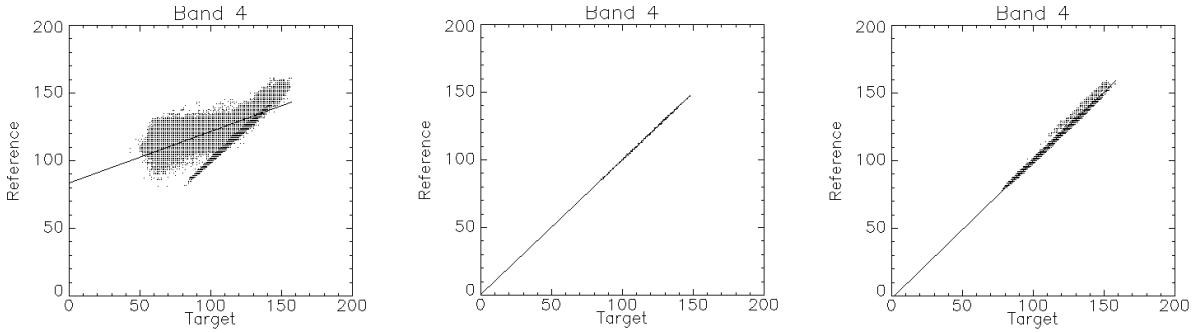

Figure 7. Orthogonal regressions for bands 4 of the partially artificial ASTER image data, left: method A, center: method B, right: method C.

Method B selects the artificial no-change pixels perfectly, providing a very convincing demonstration of its superiority over the other two alternatives.

5. DISCUSSION

On the basis of the comparisons made in the preceding section, we conclude that method B for determining no-change probability weights for iteration of the MAD transformation is superior to the two alternatives A and C. This is especially evident from Subsection 4.2, where only method B was able to discriminate the artificially generated no-change observations perfectly.

Figure 8 gives a more subjective comparison. The IR-MAD change images for the LANDSAT ETM+ scene over Jülich, Germany are compared for methods A, B and C, left to right. Middle gray indicates no change, while colored pixels signify change. Regions of prominent change are cultivated fields (principally due to maturing corn and sugar beet crops as well as grain harvesting), land reclamation in the wake of open pit mining, and mining activities themselves. These are most clearly detected with method B.

Method B, while giving by far the best results, underestimates the variances of the weighted (i.e., no-change) MAD variates after convergence of the iteration procedure. This effect may be demonstrated with a simple

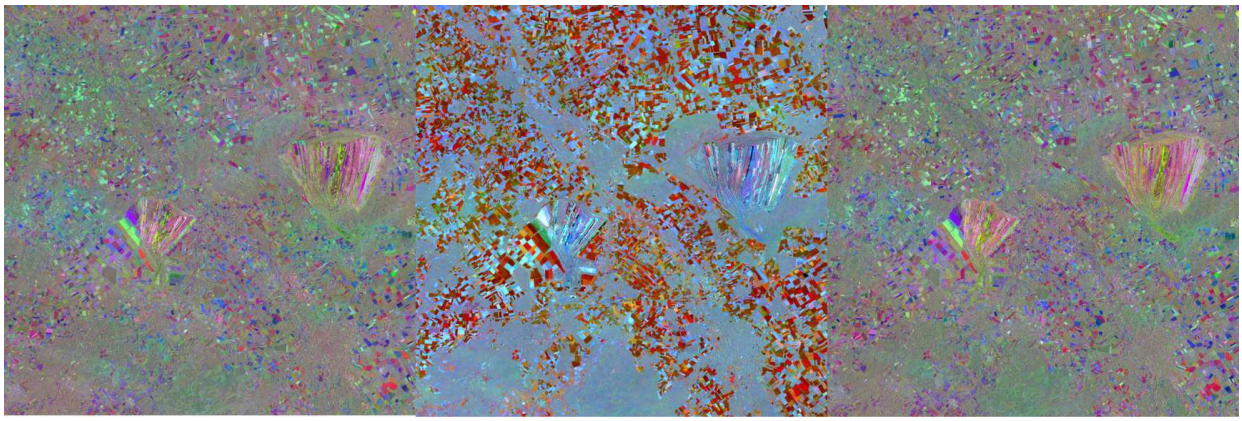

Figure 8. RGB composites of IR-MAD variates 4,5 and 6 for the Jülich, Germany scene, left: method A, center: method B, right: method C. The intensities are stretched to ±8 standard deviations of the no-change pixels.

simulation. Let us suppose that there are no changes whatsoever, and furthermore that

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{W}, \quad \boldsymbol{W} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}), \quad \langle \boldsymbol{W}\boldsymbol{X}^\top \rangle = \boldsymbol{0}. \tag{7}$$

The random vector $\boldsymbol{W}$ represents measurement noise with $\sigma^2 < 1$ and the within-image bands are uncorrelated. Then

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & (1+\sigma^2)\boldsymbol{I} \end{pmatrix} \tag{8}$$

and the MAD transformation reduces to

$$\boldsymbol{A} = \boldsymbol{I}, \quad \boldsymbol{B} = \frac{1}{\sqrt{1+\sigma^2}}\boldsymbol{I}, \quad \rho_j^2 = \frac{1}{1+\sigma^2} = \rho^2, \; j = 1 \ldots N. \tag{9}$$

The MAD variates are given by

$$M_j = X_{N-j+1} - \frac{Y_{N-j+1}}{\sqrt{1+\sigma^2}} = X_{N-j+1} - \rho Y_{N-j+1} \tag{10}$$

and their variances by

$$\mathrm{var}(M_j) = 2(1-\rho), \tag{11}$$

$j = 1 \ldots N$. The quantity

$$Z = \sum_{j=1}^{N} \frac{M_j^2}{2(1-\rho)} \tag{12}$$

is chi-square distributed with N degrees of freedom.

Since all observations are no change, we should stop here. The result of iteration to convergence of the IR-MAD procedure, method B, for the above situation is shown in Figure 9 for $\sigma = 0.5$, 100,000 observations and $N = 6$. The ratio of the iterated standard deviation of the first IR-MAD variate* to the correct value after 50 iterations is

$$\frac{\sigma_{M_1}(\text{iteration} = 50)}{\sqrt{2(1-\rho)}} = 0.657. \tag{13}$$

The chi-square statistic (12) is thus "stretched" by a factor of $0.657^{-2} \approx 2.3$ after convergence of IR-MAD. This accounts for the much smaller number of no-change observations having Pr(no change) > 0.95 for method B relative to the numbers obtained for the other two methods. On the other hand, method A will "compress" the

*The standard deviations are the same for all of the variates in this simple case.

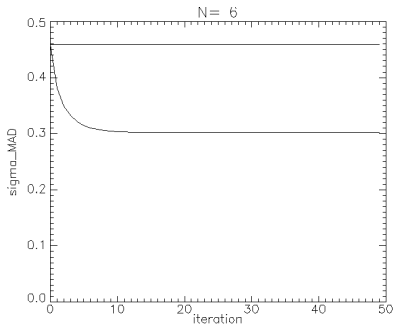

Figure 9. Iterated values of σ_{M_1} for $N = 6$. The horizontal line is the correct value.

chi-square values for the no-change observations, since the standard deviations are determined at each iteration from all of the MAD variates, both change and no-change. The variances of the change observations may be very large compared to those for no-change. This effect is evident in Tables 1 to 4, where the number of pixels exceeding the 95% threshold is always largest for method A.

Table 5. Standard deviations of the no-change IR-MAD variates for the Morocco images.

	band 1	band 2	band 3	band 4	band 5	band 7
σ_{M_C}	1.400	1.267	0.669	0.544	0.330	0.260
$\sigma_{M_B} = \sqrt{2(1-\rho)}$	1.208	0.945	0.499	0.358	0.202	0.164
ratio	0.86	0.74	0.74	0.65	0.61	0.63

For comparison with real data, Table 5 shows the standard deviations of the no-change IR-MAD variates for methods C (first row) and B (second row) for the Morocco images, as well as their ratios (third row). Method C might be expected to give the best estimate of the standard deviations of the no-change observations since it calculates the parameters of the change and no-change clusters separately. The average value of the ratio over the 6 spectral bands is 0.71. Figure 10 shows a recalculation of the regression line for the Morocco scene using method B in which the MAD variates are standardized by dividing by $\sqrt{4(1-\rho_i)}$ rather than $\sqrt{2(1-\rho_i)}$, the ratio being 0.71. The correlation is still good (> 0.94 for all 6 bands) but now 27868 rather than 1527 no-change pixels (Table 1) are selected.

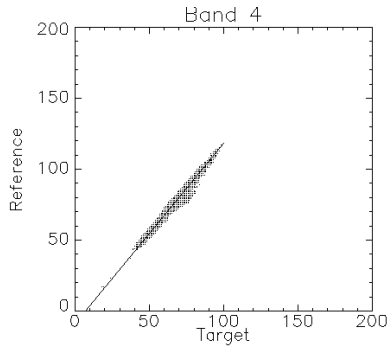

Figure 10. Orthogonal regressions for bands 4 of the Morocco images, method B, with corrected chi-square values (see text).

REFERENCES

1. A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sensing of Environment* **64**, pp. 1–19, 1998. Internet http://www.imm.dtu.dk/pubdb/p.php?1220.
2. M. J. Canty, A. A. Nielsen, and M. Schmidt, "Automatic radiometric normalization of multitemporal satellite imagery," *Remote Sensing of Environment* **91**, pp. 441–451, June 2004. Internet http://www.imm.dtu.dk/pubdb/p.php?2815.
3. A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data.," *IEEE Transactions on Image Processing* **16**(2), pp. 463–478, 2007. Internet http://www.imm.dtu.dk/pubdb/p.php?4695.
4. M. J. Canty, *Image Analysis, Classification, and Change Detection in Remote Sensing, With Algorithms for ENVI/IDL*, Taylor and Francis, 2007.
5. M. J. Canty and A. A. Nielsen, "Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation," 2007. Accepted for *Remote Sensing of Environment* Internet http://www.imm.dtu.dk/pubdb/p.php?5362.
6. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
7. M. J. Canty and A. A. Nielsen, "Visualization and unsupervised classification of changes in multispectral satellite imagery," *International Journal of Remote Sensing* **27**(18), pp. 3961–3975, 2006. Internet http://www.imm.dtu.dk/pubdb/p.php?3389.
8. H. Li, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Transactions on Image Processing* **4**(3), pp. 320–334, 1995.
9. B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Transactions on Geoscience and Remote Sensing* **40**(10), pp. 2300–2312, 2002.