# Contents

**Abstract** 2

**1 Introduction** 3

**2 The Multivariate Alteration Detection (MAD) Transformation** 6

**3 Minimum/Maximum Autocorrelation Factors** 7

**4 Case Studies** 8

4.1 SPOT HRV Data, Kenya . . . . . . . . . . . . . . . . . . . . . . . . 8

4.1.1 Data and Univariate Change Detection . . . . . . . . . . . . . . 9

4.1.2 Multivariate Change Detection . . . . . . . . . . . . . . . . . . 11

4.1.3 Different Number of Bands . . . . . . . . . . . . . . . . . . . 18

4.1.4 Geometric Illustration of Canonical Variates . . . . . . . . . . . 19

4.1.5 MAFs of MADs . . . . . . . . . . . . . . . . . . . . . . . . . 21

4.1.6 No Change Situation . . . . . . . . . . . . . . . . . . . . . . . 22

4.2 Landsat TM Data, Sweden . . . . . . . . . . . . . . . . . . . . . . . . 24

**5 Conclusions** 29

**Acknowledgements** 31

**References** 31

**A Canonical Correlation Analysis** 34

A.1 An Interpretation of Canonical Variates . . . . . . . . . . . . . . . . . 36

A.2 A Minimizing Property of Canonical Variates . . . . . . . . . . . . . . 38

A.3 Redundancy Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . 40

**B Min/Max Autocorrelation Factor Analysis** 42

B.1 Linear Transformations of MAFs . . . . . . . . . . . . . . . . . . . . . 44

# Abstract

This paper introduces a new orthogonal transformation, the multivariate alteration detection (MAD) transformation, based on an established multivariate statistical technique canonical correlation analysis. The theory for canonical correlation analysis is sketched and a result necessary for the definition of the MAD transformation is proven. As opposed to traditional univariate change detection schemes our scheme transforms two sets of multivariate observations (e.g. two multispectral satellite images covering the same geographical area acquired at different points in time) into a difference between two linear combinations of the original variables explaining maximal change (i.e. the difference explaining maximal variance) in all variables simultaneously. The MAD transformation is invariant to linear scaling. The MAD transformation can be used iteratively. First, it can be used to detect outliers (such as drop-outs) or noise and in a second iteration, it can be used to perform the actual change detection after appropriate action on outliers or noise. Also, if an analyst has additional information such as geographical position of certain changes of interest that show up in certain bands only, our method can be applied to any spatial and/or spectral subset of the full data set to direct the analysis in any desired manner. In order to obtain a spatially more coherent representation of the detected change as obtained from the MAD analysis, post-processing by means of a minimum/maximum autocorrelation factor (MAF) transformation of the MAD variates can be performed. Whereas the traditionally used principal component (PC) transformation optimizes the data variance in each new component the MAF transformation optimizes the autocorrelation represented by each component. This post-processing introduces a new spatial element into our change detection scheme which is highly relevant for image data. Two case studies using multispectral SPOT HRV data from 5 February 1987 and 12 February 1989 covering coffee and pineapple plantations in central Kenya, and Landsat TM data from 6 June 1986 and 27 June 1988 covering a forested region in northern Sweden show the usefulness of these new concepts. Because of their ability to detect change in many channels simultaneously, the MAD transformation and the MAF post-processing are expected to be even more useful when applied to image data with more bands.

# 1 Introduction

When analyzing changes in panchromatic images taken at different points in time it is customary to analyze the difference between two images, possibly after some normalization. The idea is of course that areas with no or little changes come out with zero or low absolute values and areas with large changes come out with large absolute values in the difference image.

If we have two multivariate images written as vectors at a given pixel (without loss of generality we assume $\mathrm{E}\{\boldsymbol{X}\} = \mathrm{E}\{\boldsymbol{Y}\} = \boldsymbol{0}$)

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \quad \text{resp.} \quad \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} \tag{1}$$

where $k$ is the number of spectral bands, then a simple change detection transformation is

$$\boldsymbol{X} - \boldsymbol{Y} = \begin{bmatrix} X_1 - Y_1 \\ \vdots \\ X_k - Y_k \end{bmatrix}. \tag{2}$$

If our image data have more than three channels it is difficult to visualize changes in all channels simultaneously. To overcome this problem and to concentrate information on change, linear transformations of the image data that optimize some design criterion can be considered. A linear transformation that will maximize a measure of change in the simple multispectral difference image is one that maximizes deviations from no change, e.g. the variance

$$\mathrm{Var}\{v_1(X_1 - Y_1) + \cdots + v_k(X_k - Y_k)\} = \mathrm{Var}\{\boldsymbol{v}^T(\boldsymbol{X} - \boldsymbol{Y})\}. \tag{3}$$

A multiplication of vector $\boldsymbol{v}$ with a constant $c$ will multiply the variance with $c^2$. Therefore we must make a choice concerning $\boldsymbol{v}$. A natural choice is to request that $\boldsymbol{v}$ is a unit vector, $\boldsymbol{v}^T\boldsymbol{v} = 1$. This amounts to finding principal components of the simple difference images. Principal components analysis was developed by Hotelling (1933) based on a technique described by Pearson in 1901. A disadvantage of this technique is that principal components are sensitive to the scale at which the individual variables are measured. Therefore they depend on for instance gain settings of a measuring device. Other change detection schemes based on simple difference images include factor analysis, minimum/maximum autocorrelation factor analysis (Switzer and Green, 1984;

Section 3) and decorrelation methods. Of course these methods can be applied to simple ratio images also.

A more versatile measure of change that allows different coefficients for $\boldsymbol{X}$ and $\boldsymbol{Y}$ and different number of spectral bands in the two sets, $p$ and $q$ respectively, are linear combinations

$$\boldsymbol{a}^T \boldsymbol{X} = a_1 X_1 + \cdots + a_p X_p \tag{4}$$

$$\boldsymbol{b}^T \boldsymbol{Y} = b_1 Y_1 + \cdots + b_q Y_q \tag{5}$$

and the difference between them

$$\boldsymbol{a}^T \boldsymbol{X} - \boldsymbol{b}^T \boldsymbol{Y}. \tag{6}$$

This measure also accounts for situations where the spectral bands are not the same but cover different spectral regions, for instance if one set of data comes from Landsat Thematic Mapper and the other set comes from SPOT High Resolution Visible. In this case one must be more cautious when interpreting the multivariate difference as multivariate change. In principle, any choice of $\boldsymbol{a}$ and $\boldsymbol{b}$ will give a measure of change. One could use principal components analysis on $\boldsymbol{X}$ to find an optimal $\boldsymbol{a}$ and on $\boldsymbol{Y}$ to find an optimal $\boldsymbol{b}$ (independent of $\boldsymbol{a}$). An improvement of this technique is to use principal components analysis on $\boldsymbol{X}$ and $\boldsymbol{Y}$ considered as *one* variable, cf. Fung and LeDrew (1987). For (principal components analysis related) change detection work pre-1987, cf. Fung and LeDrew (1987). Shettigara and McGilchrist (1989) use a hybrid canonical correlation/principal components technique to enhance uncorrelated parts of Landsat TM equivalents of ATM data in a gold exploration study. An interesting study on NDVI change detection based on NOAA AVHRR decade (10 day) GAC data from Sudan covering a period of nearly 7 years was presented as a video by Stern (1989). Some later references on change detection work are Nicoloyanni (1990) on the simultaneous change in brilliance and vegetation indices in Landsat MSS data, Cihlar et al. (1992) on change detection with synthetic aperture radar (SAR), Thomson (1992) on change detection with Landsat TM data, Viovy et al. (1992) on noise reduction in NOAA GAC NDVI time-series, Gong et al. (1992) on registration-noise reduction in Landsat TM difference images, Price et al. (1992) on (among other things) change detection with Landsat MSS data, and Eidenshink (1992) on the availability of biweekly NOAA-11 AVHRR NDVI 1-km data over the conterminous U.S. Gong (1993) applies principal components analysis to simple difference images.

The Fung and LeDrew (1987) approach does not guarantee an optimal separation of $\boldsymbol{X}$ and $\boldsymbol{Y}$. It defines $\boldsymbol{a}$ and $\boldsymbol{b}$ simultaneously but the method does not have a clear design criterion. Also, bands are treated similarly whether or not they come from different points in time. The Gong (1993) approach depends on the scale at which the individual variables are measured (for instance it depends on gain settings of a measuring device). Also, it forces the two sets of variables to have the same coefficients (with opposite sign), and it does not allow for the case where the two sets of images have a different number of channels. A potentially better approach is to define an optimal set of $\boldsymbol{a}$ and $\boldsymbol{b}$ simultaneously in the fashion described below. Again, let us maximize the variance, this time Var$\{\boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}\}$. A multiplication of $\boldsymbol{a}$ and $\boldsymbol{b}$ with a constant $c$ will multiply the variance with $c^2$. Therefore we must make choices concerning $\boldsymbol{a}$ and $\boldsymbol{b}$, and natural choices in this case are requesting unit variance of $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{Y}$.

The criterion then is

$$\text{maximize Var}\{\boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}\} \tag{7}$$

subject to the constraints

$$\text{Var}\{\boldsymbol{a}^T\boldsymbol{X}\} = \text{Var}\{\boldsymbol{b}^T\boldsymbol{Y}\} = 1. \tag{8}$$

Under these constraints we have

$$\begin{aligned} \text{Var}\{\boldsymbol{a}^T\boldsymbol{X} - \boldsymbol{b}^T\boldsymbol{Y}\} &= \text{Var}\{\boldsymbol{a}^T\boldsymbol{X}\} + \text{Var}\{\boldsymbol{b}^T\boldsymbol{Y}\} - 2\text{Cov}\{\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}\} && (9)\\ &= 2(1 - \text{Corr}\{\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}\}). && (10) \end{aligned}$$

As we are talking difference (or change) detection here, we shall request that $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{Y}$ are positively correlated, i.e. Corr$\{\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}\} \geq 0$. Therefore, determining the difference between linear combinations with maximum variance corresponds to determining linear combinations with minimum positive correlation.

In the sequel we shall assume that $\boldsymbol{a}$ and $\boldsymbol{b}$ are chosen so that the correlation between $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{b}^T\boldsymbol{Y}$ is positive. Positive correlation may simply be obtained by a change of sign if necessary. Determining linear combinations with extreme correlations brings the theory of canonical correlation analysis to mind. Canonical correlation analysis investigates the relationship between two groups of variables. It finds two sets of linear combinations of the original variables, one for each group. The first two linear combinations are the ones with the largest correlation. This correlation is called the first canonical correlation and the two linear combinations are called the first canonical variates. The second two linear combinations are the ones with the largest correlation subject to the condition that they are orthogonal to the first canonical variates. This correlation is called the second canonical correlation and the two linear combinations are called the second canonical variates. Higher order canonical correlations and canonical variates are defined similarly. The technique was first described by Hotelling (1936) and a treatment is given in most standard textbooks on multivariate statistics (e.g. Cooley and Lohnes, 1971 and Anderson, 1984).

The main idea presented in this paper is that because corresponding pairs of canonical variates are linear combinations of the original variables ordered by correlation or similarity between pairs, it seems natural to base a change detection scheme on differences between pairs of variates that show minimum similarity, i.e. the higher order canonical variates.

The main mathematical idea is to modify the theory used in defining canonical variates. This could be viewed as a time analog to the introduction of minimum/maximum autocorrelation factors (Switzer and Green, 1984; Section 3) in the spatial domain. In Appendix A we summarize the theory of canonical correlations and we modify the theorems so that they are more directly applicable in our context.

This type of multivariate change detection technique was first sketched in Conradsen and Nielsen (1991a, 1991b). Multivariate change detection techniques are also described in Hanaizumi and Fujimura (1992), Hanaizumi, Chino and Fujimura (1994) who work with multiple regression and canonical correlation methods applied to specific change detection.

# 2 The Multivariate Alteration Detection (MAD) Transformation

Having established the result in Appendix A we are able to define the multivariate alteration detection (MAD) transformation as

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{a}_p^T \boldsymbol{X} - \boldsymbol{b}_p^T \boldsymbol{Y} \\ \vdots \\ \boldsymbol{a}_1^T \boldsymbol{X} - \boldsymbol{b}_1^T \boldsymbol{Y} \end{bmatrix}, \tag{11}$$

where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are defined as in Appendix A, i.e. $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are the defining coefficients from a standard canonical correlation analysis. Without loss of generality we assume $\mathrm{E}\{\boldsymbol{X}\} = \mathrm{E}\{\boldsymbol{Y}\} = \boldsymbol{0}$. The MAD transformation has the very important property that if we consider linear combinations of two sets of $p$ resp. $q$ $(p \leq q)$ variables that are positively correlated then the $p$'th difference shows maximum variance among such variables. The $(p - j)$'th difference shows maximum variance subject to the constraint that this difference is uncorrelated with the previous $j$ ones. In this way we may sequentially extract uncorrelated difference images where each new image shows maximum difference (change) under the constraint of being uncorrelated with the previous ones.

If $p < q$ then the projection of $\boldsymbol{Y}$ on the eigenvectors corresponding to the eigenvalues 0 will be independent of $\boldsymbol{X}$. That part may of course be considered the extreme case of multivariate change detection.

As opposed to the principal components transformation the MAD transformation is invariant to linear scaling, which means that it is not sensitive to e.g. gain settings of a measuring device.

Another advantage of the MAD and canonical correlations procedures over methods based on simple difference images is that they are more easily extended to truly multi-temporal situations, i.e. situations where we have data from more than two points in time (Nielsen, 1994).

As the MAD transformation is a general technique for obtaining multivariate differences it can be applied to other types of data where such a difference is of interest. The method has been applied to differentiate between geogenic and anthropogenic influences in an environmental study based on soils geochemistry in eastern Germany (Pälchen, Rank, Kluge, Nielsen and Ersbøll, 1995).

# 3 Minimum/Maximum Autocorrelation Factors

In order to obtain a spatially more coherent representation of the detected change as obtained from the MAD analysis, a minimum/maximum autocorrelation factor (MAF) transformation of the MAD variates can be performed. The MAF procedure was suggested by Switzer and Green (1984). The application of MAFs in bi-temporal as well as in truly multi-temporal change detection is suggested in Nielsen (1994).

As opposed to the principal components (PC) transformation the minimum/maximum autocorrelation factor (MAF) transformation allows for the spatial nature of image data. The MAF transformation maximizes the autocorrelation rather than maximizing the data variance (PC). MAF one is the linear combination of the original bands that contains maximum autocorrelation between neighboring pixels. A higher order MAF is the linear combination of the original bands that contains maximum autocorrelation subject to the constraint that it is orthogonal to lower order MAFs. The MAF procedure thus constitutes a (conceptually) more satisfactory way of orthogonalizing image data than PC analysis. An important property of the MAF procedure—a property shared by the MAD procedure—is its invariance to linear transformations, a property not shared by ordinary PC analysis. This means that it doesn't matter whether the data have been scaled e.g. to unit variance before the analysis is performed.

PCs, MAFs and other orthogonal transformations are described in Ersbøll (1989), Conradsen and Ersbøll (1991) and Nielsen (1994). Also, MADs as well as MAFs of MADs have been applied to differentiate between geogenic and anthropogenic influences in an environmental study based on soils geochemistry in eastern Germany (Pälchen, Rank, Kluge, Nielsen and Ersbøll, 1995) mentioned above. MAF analysis is described in more detail in Appendix B.

# 4 Case Studies

This section gives two case studies to illustrate the techniques presented. One case uses SPOT High Resolution Visible data covering plantations in Kenya, another uses Landsat Thematic Mapper data covering a forested region in northern Sweden. Both case studies use raw data and are intended as illustrative examples on how calculations are performed and how interpretations of the MAD and MAF of MAD variates can be carried out. The case studies are not meant to carefully assess actual change on the ground.

## 4.1 SPOT HRV Data, Kenya

Two 512×512 SPOT High Resolution Visible (HRV) multispectral (XS) sub-scenes from 5 February 1987 and 12 February 1989 are used to test the procedure. The selected study area contains economically important coffee and pineapple fields near Thika, Kiambu District, Kenya. The analysis takes place on raw data (no atmospheric correction).

This case study is intended as an illustrative example showing how calculations are performed and how an interpretation of the canonical, MAD and MAF of MAD variates can be carried out. The case study is not meant as a careful assessment of the actual change that occurred in the study area chosen.

As an example of how to do the calculations consider the following source code if the software package Statistical Analysis System (SAS) is available:

```
data spot;
input xs1_87 xs2_87 xs3_87 xs1_89 xs2_89 xs3_89;
cards;
36 26 90 28 19 84
...
(more data lines)
...
;
```

```
proc cancorr data=spot all out=ccout
     vprefix=s87cv vname='SPOT 1987'
     wprefix=s89cv wname='SPOT 1989';
var  xs1_87 xs2_87 xs3_87;
with xs1_89 xs2_89 xs3_89;
title 'SPOT HRV XS Data 1987 vs. 1989';

data mad;
set ccout;
mad1=s89cv3-s87cv3;
mad2=s89cv2-s87cv2;
mad3=s89cv1-s87cv1;
```

#### 4.1.1 Data and Univariate Change Detection

In Figure 1 top-left (5 Feb 1987) and top-right (12 Feb 1989) we show false color composites of the multispectral SPOT HRV scenes, © CNES 1987 and 1989. The area is dominated by large pineapple fields to the northeast and small coffee fields to the northwest. To the south is the town of Thika. This is sketched in Figure 2 which also shows the positions of fields with pineapple in different phenological stages. Pineapple is a triennial crop and therefore we observe changes from one year to another.

In Figure 1 bottom-left we show the simple change detection image (differences between bands 3, bands 2 and bands 1 in red, green and blue). The major differences are due to the changes primarily in the pineapple fields. Because the changes are connected to change in vegetation, it seems natural to study the change using the normalized difference vegetation index (to avoid division by zero one may use NIR+R+1 as denominator if necessary)

$$\text{NDVI} = \frac{\text{NIR–R}}{\text{NIR+R}} \tag{12}$$

where NIR is the near-infrared channel (XS3) and R is the red channel (XS2). The philosophy behind the NDVI is that healthy green matter reflects the near-infrared light strongly and absorbs the red light. Therefore the NDVI will be large in vegetated areas and small in non-vegetated areas. In Figure 1 bottom-right we show the 1989 NDVI as red and 1987 NDVI as cyan (causing no change to be represented by a grey scale). This image enhances the differences between fields in a much clearer way than the simple change detection image. This enhancement is not necessarily due to changes from 1987 to 1989 but may also be explained by differences between, say, crops with no seasonal change at all.

Figure 1: 5 Feb 1987 SPOT HRV (top-left), 12 Feb 1989 SPOT HRV (top-right), Simple Change SPOT HRV (bottom-left), Change SPOT HRV NDVI (bottom-right)

Figure 2: Sketch of areas of interest

#### 4.1.2 Multivariate Change Detection

In Figure 3 top-left we show MAD 1, 2 and 3 as red, green and blue. In this image high and low values of the MADs corresponding to saturated colors show areas of a high degree of change. In Figure 3 top-right we show absolute values of MAD 1, 2 and 3 as red, green and blue with values below two standard deviations masked black and with values above three standard deviations saturated. These limits for stretching and masking are parameters that can be adjusted by the analyst. In this image only high values of the MADs corresponding to saturated colors other than black show areas of a high degree of change. Note that as with any technique based on eigenanalysis of covariance structures the sign of the transformed variables is arbitrary. An inspection of the MAD image (Figure 3 top-left) and a comparison with the simple change detection image (Figure 1 bottom-left) shows that there is a much better distinction between different types of changes. In the simple change detection image red and cyan are dominating but in the MAD image we see that a much better discrimination has been achieved. The image showing absolute values of MADs outlines the areas where large changes occurred and the color code in Figure 3 top says something about the nature of the change (change e.g. from vegetated to bare soil or vice versa, and dominating wavelength of change).

Below we give an interpretation of the numerical results from the computations of the MADs and a brief discussion. We discuss (1) correlations between original variables, (2) canonical correlations which are measures of similarity between the linear combinations found, (3) correlations between canonical variates and original variables in order to facilitate interpretation of the canonical variates, (4) correlations between

Figure 3: SPOT HRV MAD 1, 2 and 3 as RGB (top-left), SPOT HRV absolute values of MAD1, 2 and 3 as RGB, values below two standard deviations masked (top-right), SPOT HRV MAF 1, 2 and 3 of MADs as RGB (bottom-left), SPOT HRV absolute values of MAF 1, 2 and 3 of MADs as RGB, values below two standard deviations masked (bottom-right)

| | 1987 | | 1989 | |
|---|---|---|---|---|
| | Mean | Std Dev | Mean | Std Dev |
| XS1 | 45.00 | 5.40 | 32.27 | 4.79 |
| XS2 | 36.86 | 7.12 | 22.88 | 4.87 |
| XS3 | 74.15 | 12.55 | 62.33 | 10.66 |

Table 1: Simple statistics for 1987 and 1989 SPOT HRV XS data

| | 1987 | | | 1989 | | |
|---|---|---|---|---|---|---|
| | XS1 | XS2 | XS3 | XS1 | XS2 | XS3 |
| XS1 | 1.0000 | 0.9057 | –0.3336 | 0.5116 | 0.3955 | –0.0082 |
| 1987 XS2 | 0.9057 | 1.0000 | –0.4196 | 0.4352 | 0.4140 | –0.0381 |
| XS3 | –0.3336 | –0.4196 | 1.0000 | –0.3477 | –0.2644 | 0.2492 |
| XS1 | 0.5116 | 0.4352 | –0.3477 | 1.0000 | 0.8866 | –0.2609 |
| 1989 XS2 | 0.3955 | 0.4140 | –0.2644 | 0.8866 | 1.0000 | –0.4191 |
| XS3 | –0.0082 | –0.0381 | 0.2492 | –0.2609 | –0.4191 | 1.0000 |

Table 2: Correlations among original variables

MAD variates and original variables in order to facilitate interpretation of the MAD variates, (5) degrees of redundancy between the two sets of canonical variates, i.e. how much variance in either original data set is explained by the canonical variates, and (6) squared multiple correlations between one set of data and the canonical variates of the opposite set of data. Measures (5) and (6) assess other degrees of overlap or redundancy between the two sets of data than the canonical correlations themselves.

**Basic Statistics**

In any interpretation of statistical analysis of multivariate data it is of course important to look at the basic statistics such as means, standard deviations and correlations. The means and standard deviations are shown in Table 1.

The values from 1989 are considerably lower than the values from 1987. Whether this is due to calibration problems in the sensors or to actual changes in albedo is not known.

The correlations among the original variables are shown in Table 2.

Despite the differences in means and standard deviations it is noted that the correlation structure is remarkably similar in the two years considered. The crosscorrelations between years are less similar and decreasing with increasing wavelength, in this case indicating that changes in vegetation are the most important ones.

| | Canonical Correlation ($\sqrt{\lambda}$) | Approx Standard Error | Squared Canonical Correlation ($\lambda$) |
|---|---|---|---|
| 1 | 0.6505 | 0.0011 | 0.4232 |
| 2 | 0.4024 | 0.0016 | 0.1619 |
| 3 | 0.2403 | 0.0018 | 0.0577 |

Table 3: Canonical correlations

| | 1987 | | | 1989 | | |
|---|---|---|---|---|---|---|
| | CAN1 | CAN2 | CAN3 | CAN1 | CAN2 | CAN3 |
| XS1 | 0.3487 | –0.1272 | 0.2370 | 0.4269 | –0.1702 | 0.0887 |
| XS2 | –0.2154 | 0.2374 | –0.1323 | –0.3103 | 0.3669 | –0.0909 |
| XS3 | –0.0473 | 0.0325 | 0.0672 | –0.0245 | 0.0603 | 0.0850 |

Table 4: Raw canonical coefficients

**Canonical Correlation Analysis**

The magnitude of the canonical correlation coefficients shown in Table 3 can be used in assessing the degree of change in the bi-temporal imagery.

We see from the canonical correlations that only 6% of the variation in canonical variate 3 from one year may be explained by the variation in the other canonical variate 3. This indicates a considerable degree of change. For canonical variates 2 the number is 16%, still a rather small number. Finally, canonical variates 1 show a common predictability of 42%.

The raw canonical coefficients are shown in Table 4.

Thus the canonical variates for the 1987 XS data are

$$\begin{bmatrix} \text{CAN1} \\ \text{CAN2} \\ \text{CAN3} \end{bmatrix} = \begin{bmatrix} 0.3487 & -0.2154 & -0.0473 \\ -0.1272 & 0.2374 & 0.0325 \\ 0.2370 & -0.1323 & 0.0672 \end{bmatrix} \begin{bmatrix} \text{XS1} - 45.00 \\ \text{XS2} - 36.86 \\ \text{XS3} - 74.15 \end{bmatrix} \tag{13}$$

and the canonical variates for the 1989 XS data are

$$\begin{bmatrix} \text{CAN1} \\ \text{CAN2} \\ \text{CAN3} \end{bmatrix} = \begin{bmatrix} 0.4269 & -0.3103 & -0.0245 \\ -0.1702 & 0.3669 & 0.0603 \\ 0.0887 & -0.0909 & 0.0850 \end{bmatrix} \begin{bmatrix} \text{XS1} - 32.27 \\ \text{XS2} - 22.88 \\ \text{XS3} - 62.33 \end{bmatrix}. \tag{14}$$

| | 1987 | | | 1989 | | |
|---|---|---|---|---|---|---|
| | CAN1 | CAN2 | CAN3 | CAN1 | CAN2 | CAN3 |
| XS1 | 0.6915 | 0.7078 | 0.1442 | 0.4499 | 0.2848 | 0.0347 |
| 1987 XS2 | 0.4206 | 0.8967 | –0.1377 | 0.2736 | 0.3609 | –0.0331 |
| XS3 | –0.5784 | –0.0719 | 0.8126 | –0.3763 | –0.0289 | 0.1952 |
| XS1 | 0.5021 | 0.2423 | –0.0491 | 0.7718 | 0.6021 | –0.2045 |
| 1989 XS2 | 0.2667 | 0.3201 | –0.1072 | 0.4099 | 0.7955 | –0.4462 |
| XS3 | –0.1050 | 0.0429 | 0.2357 | –0.1613 | 0.1067 | 0.9811 |

Table 5: Correlations between original variables and canonical variables

The coefficients for computing the canonical variates are hard to interpret directly. The correlations between the original variables and the canonical variates are better for interpretation, cf. below.

**Canonical Structure**

The correlations between the original variables and the canonical variates may be used in the interpretation of the canonical variates. These correlations are shown in Table 5.

In both years we see that canonical variate 2 is strongly correlated with the visible channels, i.e. MAD2 measures changes in the visible part of the spectrum. In both years canonical variate 3 is positively correlated with the near-infrared channel and negatively correlated with or at least almost not correlated with the visible channels. This conforms with a vegetation index. Therefore, in this case MAD1 measures vegetation changes. A similar pattern but with less emphasis on the near-infrared channel is seen for canonical variates 1 and MAD3.

**MAD Structure**

In order to interpret the MADs we give the correlations between the original variables and the MADs. These values will not be supplied by a canned canonical correlations computer program. The values are computed by means of the expressions given in Appendix A. It is easier—and more CPU time consuming—to use an ordinary correlation program on the estimated MAD image. The correlations between original variables and MADs are shown in Table 6.

The most dominant correlations are MAD1 with 1987 XS3 (–0.50) and with 1989 XS3 (0.60). Pixels showing extreme values of MAD1 will predominantly have high values of 1987 XS3 and low values of 1989 XS3 or vice versa. Thus MAD1 basically describes changes in XS3, the photo-infrared channel, which again is strongly related to vegetation. Areas showing extreme values in MAD1 will then most likely have very different vegetation cover in 1987 and 1989. Changes orthogonal to (i.e. uncorrelated with) these changes are described by MAD2 and MAD3. Similar considerations on magnitudes of correlations show that MAD2 and MAD3 describe changes in the

| | MAD1 | MAD2 | MAD3 |
|---|---|---|---|
| XS1 | –0.0889 | –0.3868 | –0.2890 |
| 1987 XS2 | 0.0849 | –0.4901 | –0.1757 |
| XS3 | –0.5008 | 0.0393 | 0.2418 |
| XS1 | –0.1260 | 0.3292 | 0.3227 |
| 1989 XS2 | –0.2750 | 0.4349 | 0.1714 |
| XS3 | 0.6047 | 0.0583 | –0.0674 |

Table 6: Correlations between original variables and MADs

| Channel<br>Area | Red<br>MAD1 | Green<br>MAD2 | Blue<br>MAD3 | <br>MAD |
|---|---|---|---|---|
| 1 | High | High | High | Light Gray |
| 2 | Low | High | High | Cyan |
| 3 | High | Low | Low | Red |
| Town | Low | High | Low | Green |

Table 7: Levels of MADs in three pineapple areas and in the town

shorter wavelengths, MAD2 with the emphasis on XS2 and MAD3 with the emphasis on XS1.

Correlations between original variables and MADs are quite low. Alternatively they could be calculated only where the absolute values of MADs are high.

At this point it should be emphasized again that the interpretations presented are scene dependent. In other scenes the interpretations of the MADs will very likely be different. Where a technique as the NDVI change detection "looks for" changes in vegetation cover the present method detects general alterations in the scene no matter what caused the alteration. Once established the MADs may be interpreted by means of the correlations between the original and the transformed variables as presented above.

To illustrate the concept further we shall examine the pineapple fields north of Thika somewhat closer. In Figure 2 some pineapple fields are outlined along with the center of Thika. In Table 7 we have indicated the relative level of the three MAD variables mapped as red, green and blue in Figure 3 top-left.

First we consider area 2, bare soil in 1989 and healthy pineapple in 1987. This is an area that shows extreme deviation between the two scenes. The area is strongly outlined in all change schemes used, a.o. simple difference change detection, NDVI change detection, decorrelated simple difference change detection (not shown), principle components, rotated factors and MAFs of simple difference images (not shown),

| 1987 Canonical Variables | | | 1989 Canonical Variables | | |
|---|---|---|---|---|---|
| | Proportion | Cumulative Proportion | $\lambda$ | Proportion | Cumulative Proportion |
| CV1 | 0.3299 | 0.3299 | 0.4232 | 0.1396 | 0.1396 |
| CV2 | 0.4368 | 0.7667 | 0.1619 | 0.0707 | 0.2103 |
| CV3 | 0.2333 | 1.0000 | 0.0577 | 0.0135 | 0.2238 |

Table 8: Variance of 1987 XS explained by the individual canonical variates for 1987 and 1989

MAD and MAF of MAD. Area 3 shows the opposite pattern, pineapple in 1989 and bare soil in 1987. The values related with these patterns are consistent with the general interpretation of the MADs given before. Since pineapple is a triennial crop and the time difference between the two scenes is two years area 1 must be pineapple in different phenological stages. The alterations are strongly related to vegetation change and are therefore clearly visible in the NDVI change image. In the NDVI change image we see a totally black area in the vegetation free center of Thika. This is very consistent with the notion of a vegetation index. In the same area the MAD change image reveals a considerable alteration. No information is available to us on the probable causes for these changes and we shall not speculate on their nature. Whatever the causes, the differences described illustrate the fact that the MADs may be used in general detection of alterations irrespective of the nature of the alterations.

As a concluding remark we therefore suggest the usage of the MAD transformation in the analysis of multispectral, bi-temporal imagery. The MADs give an optimal (in the sense of maximal variance with linear transformation invariance) detection of alterations from one scene to the other. Also, it provides a statistical analysis and an interpretation of the nature of the alterations.

**Canonical Redundancy Analysis**

A more detailed assessment of the degree of change may be obtained from a deeper study of the correlations between the variates involved. The standardized variance of 1987 XS explained by the individual canonical variates for 1987 and 1989 are shown in Table 8.

The standardized variance of 1989 XS explained by the individual canonical variates for 1989 and 1987 are shown in Table 9.

The squared multiple correlations ($R^2$) between 1987 XS and the first $M$ canonical variates of 1989 XS, and squared multiple correlations ($R^2$) between 1989 XS and the first $M$ canonical variates of 1987 XS are shown in Table 10.

The canonical redundancy analysis confirms that we have considerable changes between the two years. The degrees of explanation of one set of original variables by the

| 1989 Canonical Variables | | | 1987 Canonical Variables | | |
|---|---|---|---|---|---|
| | Proportion | Cumulative Proportion | $\lambda$ | Proportion | Cumulative Proportion |
| CV1 | 0.2632 | 0.2632 | 0.4232 | 0.1114 | 0.1114 |
| CV2 | 0.3356 | 0.5988 | 0.1619 | 0.0543 | 0.1658 |
| CV3 | 0.4012 | 1.0000 | 0.0577 | 0.0232 | 0.1889 |

Table 9: Variance of 1989 XS explained by the individual canonical variates for 1989 and 1987

| | $R^2$(1987 XS, 1989 CAN) | | | $R^2$(1989 XS, 1987 CAN) | | |
|---|---|---|---|---|---|---|
| $M$ | 1 | 2 | 3 | 1 | 2 | 3 |
| XS1 | 0.2024 | 0.2835 | 0.2847 | 0.2521 | 0.3108 | 0.3132 |
| XS2 | 0.0749 | 0.2051 | 0.2062 | 0.0711 | 0.1736 | 0.1851 |
| XS3 | 0.1416 | 0.1424 | 0.1805 | 0.0110 | 0.0129 | 0.0684 |

Table 10: Squared multiple correlations ($R^2$) between 1987 (1989) XS and the first $M$ canonical variates of 1989 (1987) XS

opposite canonical variates range from 1% to 14%, very low numbers that indicate this change. Similarly, we see from the squared multiple correlations between the original 1987 variables and the first $M$ 1989 canonical variates and the squared multiple correlations between the original 1989 variables and the first $M$ 1987 canonical variates that the degree of explanation is poorest in the near-infrared band, again indicating that vegetation changes are dominating.

### 4.1.3 Different Number of Bands

As a simulated example of an application where we don't have the same number of spectral bands at the two points in time we analyze what happens if we omit SPOT HRV XS1 from the 1989 data but not from the 1987 data. This situation cannot be dealt with by any technique based on simple difference images without dropping the information from XS1 in 1987 also. The result from such analysis can be handled in the same fashion as above. Here we shall confine ourselves to a visual inspection of the absolute value of MAD1 with MAD1 from the analyses above with all three spectral bands retained, Figure 4. Bright greytones represent areas of maximum change in all bands simultaneously independent of what caused the change and independent of direction of change. A comparison shows that basically the same areas emerge, specifically the fields showing most change are the same.

Figure 4: SPOT HRV MAD1, absolute values, three bands in 1987 and in 1989 (left), three bands in 1987 and two bands in 1989 (right)

This is of course a somewhat constructed example but in historical change detection studies comparisons between data from for instance Landsat MSS (four bands) and Landsat TM (six or seven bands) or SPOT HRV (three bands) may be relevant.

### 4.1.4 Geometric Illustration of Canonical Variates

To hopefully give a better feel for what canonical variates are and to illustrate geometrically the solution to the real, symmetric, generalized (RSG) eigenproblem involved in finding them (cf. Appendix A), we have generated two sets of data both consisting of two variables. The data consist of every 50'th row and every 50'th column of the image data analyzed above. The first set of variables are bands 1 and 2 from the 1987 data and the second set of variables are bands 2 and 3 from the 1989 data. The 1987 (1989) data are estimated from the 1989 (1987) data by regression.

The two top plots in Figure 5 show scatterplots and ellipses corresponding to $\chi^2_{0.95}(2) = 5.991$ contours for the 1987 and 1989 data. These contour ellipses are (top-left; see Appendix A for a description of the mathematics illustrated here) $a^T \hat{\Sigma}_{11}^{-1} a = 5.991$ (for the data) and $a^T (\hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21})^{-1} a = 5.991$ (for the regressions), and (top-right) $b^T \hat{\Sigma}_{22}^{-1} b = 5.991$ (for the data) and $b^T (\hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12})^{-1} b = 5.991$ (for the regressions). The open circles symbolize observations and the crosses symbolize regressions made from the opposite set of variables.

The two bottom plots show the solution to the eigenproblem. The ellipses shown are contours for $a^T \mathsf{D} a = 1$, $a^T \mathsf{N} a = 1$, $a^T \mathsf{N} a = \lambda_1$, $a^T \mathsf{N} a = \lambda_2$, where $\mathsf{N}$ means the matrix in the numerator of the Rayleigh coefficient identifying the canonical correlation problem and $\mathsf{D}$ means the matrix in the denominator. In this case $\lambda_1 = 0.2730$

Figure 5: Canonical variates geometrically

and $\lambda_2 = 0.05147$ corresponding to canonical correlations $0.5199$ and $0.2269$. In the bottom-right plot the contour lines are identified.

In the two bottom plots the eigenvectors to the canonical correlation problem are vectors with end points in the center of the ellipses and the points where the ellipses have a common tangent (indicated with short lines). The square root of the eigenvalues (the canonical correlations) $\sqrt{\lambda_i}$ are the ratios of the the lengths of the major (or minor) axes in the ellipses corresponding to $a^T\mathsf{N}a = 1$ and $a^T\mathsf{N}a = \lambda_i$. The major axes of $a^T\mathsf{N}a = 1$ are indicated with long lines. The major and minor axes of $\mathsf{D}$ corresponds to the axes of projection for principal components. We see that the directions of projection are different for the canonical variates.

#### 4.1.5 MAFs of MADs

In order to obtain a spatially more coherent representation of the detected change as obtained from the MAD analysis illustrated in Figure 3 top, post-processing by means of a MAF transformation of the MAD variates was performed.

In Figure 3 bottom-left we show MAF 1, 2 and 3 of MADs as red, green and blue. In this image high and low values of the MAFs corresponding to saturated colors show areas of high spatial coherence in the change pattern. In Figure 3 bottom-right we show absolute values of MAF 1, 2 and 3 of MADs as red, green and blue with values below two standard deviations masked black and with values above three standard deviations saturated. In this image only high values of the MAFs corresponding to saturated colors other than black show areas of high spatial coherence in the change pattern. This image outlines areas where large changes occurred and the color code in Figure 3 bottom says something about the nature of the change (change e.g. from vegetated to bare soil or vice versa; the dominating wavelength of change can be read from the correlations between original spectral bands and the MAFs).

The largest visual difference between MADs (Figure 3 top) and MAFs of MADs (Figure 3 bottom) is the emergence of area 3 (Figure 2) and a general impression of better spatial "togetherness", especially in the masked absolute values (Figure 3 right). The fact that area 3 emerges in the MAFs is assuring as the change in this field seems significant from inspection of Figure 1 top.

In order to interpret the MAFs of MADs we give the correlations between the MADs and the MAFs (Table 11), and the correlations between the original variables and the MAFs (Table 12).

Table 11 shows that MAF1 concentrates information from MADs 2 and 3 vs. MAD1, MAF2 concentrates information from MADs 1 and 2, and MAF3 is basically MAD3.

Table 12 shows that MAF1 for 1987 complies with a vegetation index and for 1989 with a negative vegetation index. MAF1 therefore basically contains information on

| | MAF1 | MAF2 | MAF3 |
|---|---|---|---|
| Autocorr. | 0.9243 | 0.8738 | 0.5780 |
| MAD1 | –0.7053 | 0.6655 | 0.2444 |
| MAD2 | 0.5812 | 0.7401 | –0.3383 |
| MAD3 | 0.4060 | 0.0965 | 0.9088 |

Table 11: Correlations between MADs and MAFs of MADs

| | MAF1 | MAF2 | MAF3 |
|---|---|---|---|
| Autocorr. | 0.9243 | 0.8738 | 0.5780 |
| XS1 | –0.2795 | –0.3734 | –0.1535 |
| 1987 XS2 | –0.4161 | –0.3233 | 0.0268 |
| XS3 | 0.4742 | –0.2809 | 0.0840 |
| XS1 | 0.4111 | 0.1909 | 0.1510 |
| 1989 XS2 | 0.5162 | 0.1554 | –0.0586 |
| XS3 | –0.4200 | 0.4391 | 0.0668 |

Table 12: Correlations between original variables and MAFs of MADs

vegetation changes. This is confirmed by inspection of the MAF1 image (red in Figure 3 bottom). MAF2 has medium negative correlations with all bands in 1987 and low positive correlations with visual bands in 1989 and a relatively high positive correlation with 1989 XS3. Inspection of the MAF2 image (green in Figure 3 bottom) shows that change in lakes and in the town is picked up by MAF2. MAF3 is quite noisy, it has low correlations with the original bands (highest with opposite signs for the two XS1s perhaps indicating differences in atmosphere and water); inspection of the MAF3 image (blue in Figure 3 bottom) reveals that it contains some information on change in lakes and in the town.

Correlations between original variables and MAFs of MADs are quite low. Alternatively they could be calculated only where the absolute values of MAFs of MADs are high.

#### 4.1.6 No Change Situation

As a simple simulation of a situation with no change in all bands we pad the two $512 \times 512$ scenes described above into the central part of $600 \times 600$ backgrounds with values 20, 30 and 40 in bands 1, 2 and 3 in both years. Change between the two $600 \times 600$ scenes is estimated by means of

- simple differences, dif,

| Band/Component | 1 | 2 | 3 |
|---|---|---|---|
| dif | 0 | 0 | 0 |
| Ddif(cov) | 0.86 | 0.87 | 0.59 |
| Ddif(corr) | 0.95 | 0.81 | 0.56 |
| PCdif(cov) | 0.86 | 1.05 | 0.08 |
| PCdif(corr) | 1.32 | 0.35 | –0.04 |
| Fdif | 1.23 | 0.58 | 0.07 |
| MAFdif | 1.32 | 0.31 | 0.12 |
| MAD | 0.02 | 0.10 | –0.10 |
| MAF/MAD | 0.09 | –0.11 | –0.00 |

Table 13: Change detected in no change region for SPOT HRV data, value should be 0

- decorrelation of simple differences (based on covariance matrix), Ddif(cov),
- decorrelation of simple differences (based on correlation matrix), Ddif(corr),
- principal components of simple differences (based on covariance matrix), PCdif(cov),
- principal components of simple differences (based on correlation matrix), PCdif(corr),
- varimax rotated factors of simple differences, Fdif,
- MAFs of simple differences, MAFdif,
- MADs, MAD, and
- MAFs of MADs, MAF/MAD.

Change detected in the region with no change (the 44 pixels broad edge around the actual image data) as indicated by standardized values of the results from the above change detection methods is given in Table 13.

The value in this region in the simple difference images is zero. Therefore the change detected here is due to the subtraction of the mean value of the entire image before calculating the relevant linear combinations of the original bands. This can easily be identified in this very simple situation, but had the no-change pixels been scattered as several fields inside the image this would not have been possible and the statistics for calculating the MADs would have been exactly the same. Therefore the situation simulated though very simple is indeed a realistic one.

It is obvious that MAD and MAF of MAD are the only multivariate techniques that perform well in this situation. All other methods give change values much higher than zero.

## 4.2 Landsat TM Data, Sweden

As another application example of the MAD and the combined MAD/MAF technique two 512×512 Landsat Thematic Mapper (TM) sub-scenes from 6 June 1986 and 27 June 1988 are used. The thermal TM band 6 is not included in the analysis. Resampling to 20 meter pixels rectified to the Swedish national grid was performed by the Swedish Space Corporation who also provided the data. The area covered is a small forested area in northern Sweden near Umeaa. The study area and previous work done is described in Olsson (1993, 1994, 1995). Though described in a less elaborate fashion our aim here is the same as in the previous case study, namely to show the application of the new techniques and to give ideas on how to interpret results, this time with Landsat TM data. Again, we mean to illustrate the techniques presented, the case study is not meant as a careful assessment of actual change on the ground.

The top two images in Figure 6 are TM bands 4, 5 and 3 as red, green, and blue in 1986 (left) and in 1988 (right).

Figure 7 shows all six MADs row-wise. Figure 8 shows all six MAFs of MADs row-wise.

Correlations between the MADs and the original TM bands given in Table 14 are generally quite low. However, the pattern revealed shows that MAD1 is associated with TM1, i.e. probably differences in atmospheric conditions. MAD4 is positively correlated with 1986 TM4 and negatively correlated with 1986 TM1, 2 and 3. The opposite correlation structure is true for MAD4 and the 1988 data. Therefore MAD4 is a sort of vegetation index change detector. With reverse signs for the correlations this is true for MAD5 also.

Correlations between the MADs and their MAFs are shown in Table 15. We see that low order MAFs (signal) are associated with high order MADs, i.e. maximum similarity canonical variates and vice versa.

Table 16 shows correlations between MAFs of MADs and the original TM bands. MAF1 is positively correlated with 1986 TM4 (and slightly negatively correlated with 1986 TM1 and 3). The opposite correlation structure is true for MAF1 and the 1988 data. Therefore MAF1 is a sort of vegetation index change detector and it concentrates the information from MADs 4 and 5. MAF2 is a change detector of the weighted mean of all bands except TM4, i.e. a change detector of the non-vegetation related level. MAF6 measures change in TM1 which is likely to represent changes in atmospheric conditions. We see that the MAF analysis of the MADs has isolated the changes related to TM1 (presumably atmospheric conditions) and changes related to TM4 (presumably vegetation) in each end of the autocorrelation "spectrum".

The bottom two images in Figure 6 are MAFs 1, 2 and 3 of MADs as red, green and blue (left) and absolute values of MAFs 1, 2 and 3 of MADs as red, green and blue with values below 2 standard deviations masked black and values above three standard values saturated (right). These images beautifully depict locations and strength of

Figure 6: 6 June 1986 Landsat TM bands 4, 5 and 3 as RGB (top-left), 27 June 1988 Landsat TM bands 4, 5 and 3 as RGB (top-right), Landsat TM MAF 1, 2 and 3 of MADs as RGB (bottom-left), Landsat TM absolute values of MAF 1, 2 and 3 of MADs as RGB, values below two standard deviations masked (bottom-right)

Figure 7: Landsat TM MAD 1–6 row-wise

Figure 8: Landsat TM MAF 1–6 of MADs row-wise

| | MAD1 | MAD2 | MAD3 | MAD4 | MAD5 | MAD6 |
|---|---|---|---|---|---|---|
| Can.corr. | 0.0464 | 0.0978 | 0.3199 | 0.5957 | 0.7255 | 0.8966 |
| 1986 TM1 | 0.2583 | –0.0010 | –0.0789 | –0.0759 | 0.1078 | 0.1941 |
| 1986 TM2 | –0.0389 | 0.1994 | –0.1150 | –0.0512 | 0.0313 | 0.2096 |
| 1986 TM3 | –0.0280 | 0.0202 | 0.0634 | –0.0896 | 0.1008 | 0.2123 |
| 1986 TM4 | 0.0061 | 0.0781 | –0.0103 | 0.1105 | –0.2109 | 0.1764 |
| 1986 TM5 | –0.0167 | 0.0355 | –0.0431 | 0.1380 | 0.1089 | 0.2047 |
| 1986 TM7 | –0.0592 | –0.0676 | –0.1041 | 0.0638 | 0.1493 | 0.1993 |
| 1988 TM1 | –0.1453 | 0.0239 | 0.0778 | 0.1197 | –0.1621 | –0.1867 |
| 1988 TM2 | 0.0986 | –0.0480 | 0.0399 | 0.1069 | –0.0978 | –0.2089 |
| 1988 TM3 | 0.0068 | 0.0922 | –0.0676 | 0.1583 | –0.1567 | –0.1854 |
| 1988 TM4 | –0.0089 | –0.0218 | –0.0211 | –0.1216 | 0.1930 | –0.1838 |
| 1988 TM5 | 0.0122 | 0.0567 | –0.0765 | –0.0662 | –0.1588 | –0.1995 |
| 1988 TM7 | 0.0390 | 0.1630 | 0.0069 | 0.0072 | –0.2069 | –0.1798 |

Table 14: Correlations between MADs and original TM bands

| | MAF1 | MAF2 | MAF3 | MAF4 | MAF5 | MAF6 |
|---|---|---|---|---|---|---|
| Autocorr. | 0.9130 | 0.8769 | 0.6822 | 0.5402 | 0.4770 | 0.4696 |
| MAD1 | 0.0403 | –0.0784 | –0.1402 | –0.1820 | 0.2033 | 0.9477 |
| MAD2 | 0.2501 | –0.1095 | –0.2162 | 0.7958 | 0.4954 | –0.0051 |
| MAD3 | 0.0331 | 0.0074 | –0.0456 | 0.4942 | –0.8262 | 0.2646 |
| MAD4 | 0.4769 | 0.1545 | 0.8523 | 0.0701 | 0.0540 | 0.1205 |
| MAD5 | –0.7485 | 0.4915 | 0.2780 | 0.2759 | 0.1661 | 0.1309 |
| MAD6 | 0.3834 | 0.8464 | –0.3576 | –0.0913 | –0.0162 | –0.0133 |

Table 15: Correlations between MAFs of MADs and MADs

| | MAF1 | MAF2 | MAF3 | MAF4 | MAF5 | MAF6 |
|---|---|---|---|---|---|---|
| Autocorr. | 0.9130 | 0.8769 | 0.6822 | 0.5402 | 0.4770 | 0.4696 |
| 1986 TM1 | –0.0349 | 0.1848 | –0.1365 | –0.0801 | 0.1278 | 0.2263 |
| 1986 TM2 | 0.0771 | 0.1653 | –0.1423 | 0.0949 | 0.1849 | –0.0731 |
| 1986 TM3 | –0.0307 | 0.2159 | –0.1276 | 0.0547 | –0.0396 | –0.0103 |
| 1986 TM4 | 0.2976 | 0.0536 | –0.0448 | –0.0105 | 0.0166 | –0.0140 |
| 1986 TM5 | 0.0696 | 0.2452 | 0.0713 | 0.0310 | 0.0720 | 0.0007 |
| 1986 TM7 | –0.0277 | 0.2632 | 0.0522 | –0.0670 | 0.0655 | –0.0587 |
| 1988 TM1 | 0.1095 | –0.2099 | 0.1354 | 0.0646 | –0.0994 | –0.1215 |
| 1988 TM2 | 0.0374 | –0.2105 | 0.1334 | –0.0368 | –0.0438 | 0.1071 |
| 1988 TM3 | 0.1428 | –0.2206 | 0.1398 | 0.0236 | 0.0884 | –0.0109 |
| 1988 TM4 | –0.2794 | –0.0765 | 0.0227 | 0.0353 | 0.0333 | –0.0009 |
| 1988 TM5 | 0.0229 | –0.2649 | –0.0397 | –0.0251 | 0.0671 | –0.0351 |
| 1988 TM7 | 0.1319 | –0.2737 | –0.0281 | 0.0858 | 0.0519 | 0.0141 |

Table 16: Correlations between MAFs of MADs and original TM bands

change. Different types of change are indicated by different colors and they can be interpreted from shift in mean values, dispersion matrices, and from correlations with the original data as indicated above. Noise related change is isolated in the higher order MAFs of MADs.

The simple simulation of a situation with no change in all bands performed for the SPOT HRV data above is repeated with the Landsat TM data. Here we pad the two 512×512 scenes into the central part of 600×600 backgrounds with values 0 in all bands in both years. Apart from the exclusion of the decorrelation methods change between the two 600×600 scenes is estimated by means of the same methods as in the SPOT HRV case above.

Change detected in the region with no change (the 44 pixels broad edge around the actual image data) as indicated by standardized values of the results from the different change detection methods is given in Table 17.

As in the previous case with SPOT HRV data it is obvious that MAD and MAF of MAD are the only multivariate techniques that perform well in this situation. All other methods give values much higher than zero.

# 5 Conclusions

Based on the established multivariate statistical technique canonical correlation analysis, we have developed a new tool for the analysis of multispectral, bi-temporal

| Band/Component | 1 | 2 | 3 | 4 | 5 | 6/7 |
|---|---|---|---|---|---|---|
| dif | 0 | 0 | 0 | 0 | 0 | 0 |
| PCdif(cov) | –1.28 | 0.06 | 0.09 | 0.91 | –0.03 | 0.26 |
| PCdif(corr) | –0.06 | –1.53 | 0.19 | 0.40 | 0.03 | –0.01 |
| Fdif | 0.37 | 1.01 | –0.27 | 1.10 | 0.29 | 0.05 |
| MAFdif | –1.52 | 0.05 | 0.46 | 0.03 | 0.01 | 0.01 |
| MAD | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | –0.03 |
| MAF/MAD | –0.00 | 0.01 | 0.02 | 0.00 | –0.02 | 0.02 |

Table 17: Change detected in no change region for Landsat TM data, value should be 0

imagery. The multivariate alteration detection (MAD) transformation gives an optimal (in the sense of maximal variance) detection of alterations (differences, changes) from one scene to the other in all spectral channels simultaneously. As opposed to the principal components transformation the MAD transformation is invariant to linear scaling, which means that it is not sensitive to e.g. gain settings of a measuring device. It also provides a statistical analysis and it offers an interpretation of the nature of the alterations. The MAD transformation can be used iteratively (or in an exploratory fashion, if you like). First, it can be used to detect outliers (such as drop-outs) or noise and in a second iteration, it can be used to perform the actual change detection after appropriate action on outliers or noise. Also, if an analyst has additional information such as geographical position of certain changes of interest that show up in certain bands only, our method can be applied to any spatial and/or spectral subset of the full data set to direct the analysis in any desired manner. In classification-type methods the spatial subset is referred to as training area(s) and the spectral subset is a simple form of feature selection.

Irrespective of the application in question and the individual analyst's favorite change detection scheme, the absolute value of the MADs will always outline the areas with the largest overall changes in all channels simultaneously. The correlations between the original image channels and the MADs form a basis for interpretation of the MADs. Based on the absolute values of the MADs and the interpretation, a more physically oriented change study can be performed if so desired. The same is true for MADs post-processed with the MAF transformation.

The data in the case studies are raw, not pre-processed (e.g. atmospherically corrected). Analysis of corrected data would be interesting also. However, our purpose here is to introduce a new change detection technique.

We conclude that the MADs and post-processing with MAFs of MADs with their linear scaling invariance—which allows us to detect also subtle change and not just dominant over-all change—combined with the stretching and masking chosen constitute very useful supplements to univariate and existing multivariate change detection schemes. This is supported by the results reported in section 4.1.6 "No Change Situation" for the SPOT HRV case and similar results for the Landsat TM case.

The MAD and the MAF of MADs techniques are believed to be useful with multi- and hyper-channel data in monitoring and surveillance in the fields of in environmental studies, oceanography, agriculture, forestry, geobotany etc.

# Acknowledgements

We wish to thank our colleagues Dr. Bjarne Kjær Ersbøll and Dr. Rasmus Larsen, both IMM, for many good discussions on multivariate statistics and analysis of spatial data.

Most calculations were done by means of software written by Dr. Rasmus Larsen on our initiative and under our supervision.

Also, we would like to thank Dr. James J. Simpson, Scripps Institution of Oceanography, for the immediate interest he took in these new change detection schemes and for reading and commenting very thoroughly on our manuscript.

The data for the Landsat TM forestry case study was kindly provided by the Swedish Space Corporation who also carried out the geometric rectification.

The work reported was funded in part by Danida, the Danish International Development Agency, under grant No. 104.Dan.8/410, the Commission of European Communities under contracts No. MA2M-CT90-0010 and BRE2-CT92-0201, and the Danish National Technical Research Council under grant No. 5.26.09.07. This funding is highly appreciated.

# References

T.W. Anderson (1984): *An Introduction to Multivariate Statistical Analysis*. 2nd Edition. John Wiley & Sons.

J. Cihlar, T.J. Pultz and A.L. Gray (1992): Change detection with synthetic radar. *International Journal of Remote Sensing*, Vol. 13, No. 3, pp. 401–414.

K. Conradsen and B.K. Ersbøll (1991): Data Dependent Orthogonal Transformations of Multichannel Image Data. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark. Technical Report.

K. Conradsen and A.A. Nielsen (1991a): Remote Sensing in Forecasting Agricultural Statistics in Kenya. Danida, the Danish International Development Agency. Technical Report. Contract No. 104.Dan.8/410. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

K. Conradsen and A.A. Nielsen (1991b): Multivariate Change Detection in Multispectral, Multitemporal Images. In *Abstracts and Notes from Seminar on Near Real-Time Remote Sensing for Land and Ocean Applications*. Eurimage and ESA/Earthnet, Rome, Italy.

W.W. Cooley and P.R. Lohnes (1971): *Multivariate Data Analysis*. John Wiley & Sons.

J.C. Eidenshink (1992): The 1990 Conterminous U.S. AVHRR Data Set. *Photogrammetric Engineering and Remote Sensing*, Vol. 58, No. 6, pp. 809–813.

B.K. Ersbøll (1989): *Transformations and Classifications of Remotely Sensed Data*. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark. Ph.D. Thesis.

T. Fung and E. LeDrew (1987): Application of Principal Components Analysis to Change Detection. *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 12, pp. 1649–1658.

P. Gong, E.F. LeDrew and J.R. Miller (1992): Registration-noise reduction in difference images for change detection. *International Journal of Remote Sensing*, Vol. 13, No. 4, pp. 773–779.

P. Gong (1993): Change Detection Using Principal Component Analysis and Fuzzy Set Theory. *Canadian Journal of Remote Sensing*, Vol. 19, No. 1, pp. 22–29.

H. Hanaizumi, S. Chino and S. Fujimura (1994): A Method for Change Analysis with Weight of Significance Using Multi-temporal, Multi-spectral Images. In Desachy, J. (Ed.) *Proceedings SPIE 2315 European Symposium on Satellite Remote Sensing, Image and Signal Processing for Remote Sensing*, Rome, Italy, pp. 282–288.

H. Hanaizumi and S. Fujimura (1992): Change Detection from Remotely Sensed Multi-temporal Images Using Multiple Regression. In *Proceedings from the 1992 International Geoscience and Remote Sensing Symposium, IGARSS*, Houston, Texas, USA, pp. 564–566.

H. Hotelling (1933): Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, Vol. 24, pp. 417–441.

H. Hotelling (1936): Relations between Two Sets of Variates. *Biometrika*, Vol. XXVIII, pp. 321–377.

E. Nicoloyanni (1990): Un indice de changement diachronique appliqué à deux scènes Landsat MSS sur Athènes (Grèce). *International Journal of Remote Sensing*, Vol. 11, No. 9, pp. 1617–1623.

A.A. Nielsen (1994): *Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data*. Department of Mathematical Modelling, Technical University of Denmark. Lyngby. Ph.D. Thesis.

H. Olsson (1993): Regression Functions for Multitemporal Relative Calibration of Thematic Mapper Data over Boreal Forest. *Remote Sensing of Environment*, Vol. 46, No. 1, pp. 89–102.

H. Olsson (1994): *Monitoring of Local Reflection Changes in Boreal Forests Using Satellite Data*. Remote Sensing Laboratory, Swedish University of Agricultural Sciences. Umeaa. Ph.D. Thesis.

H. Olsson (1995): Reflectance Calibration of Thematic Mapper Data for Forest Change Detection, *International Journal of Remote Sensing*, Vol. 66, No. 1, pp. 81–96.

W. Pälchen, G. Rank, A. Kluge, A.A. Nielsen and B.K. Ersbøll (1995): A New Approach to Differentiation between Geogenic and Anthropogenic Influences on Soils in a Mining Processing Area. In K. Haarstad (Ed.): *Proceedings of the Nordic Symposium on Variability in Polluted Soil and Groundwater*. Oslo, Norway.

K.P. Price, D.A. Pyke and L. Mendes (1992): Shrub Dieback in a Semiarid Ecosystem: The Integration of Remote Sensing and Geographic Information Systems for Detecting Vegetation Change. *Photogrammetric Engineering and Remote Sensing*, Vol. 58, No. 4, pp. 455–463.

V.K. Shettigara and C.A. McGilchrist (1989): A Principal Component and Canonical Correlation Hybrid Technique for Change Detection in Two-Image Sets. In R.F. Barrett (Ed.) *ASSPA 89, signal processing, theories, implementations and applications*, Adelaide, Australia, pp. 47–52.

M. Stern (personal communication, 1989): Video on NDVI change detection in Sudan based on 7 years of NOAA AVHRR decade GAC data. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

P. Switzer and A.A. Green (1984): Min/Max Autocorrelation Factors for Multivariate Spatial Imagery. Technical Report. Department of Statistics, Stanford University.

A.G. Thomson (1992): A multi-temporal comparison of two similar Landsat Thematic Mapper images of upland North Wales, U.K. *International Journal of Remote Sensing*, Vol. 13, No. 5, pp. 947–955.

N. Viovy, O. Arino and A.S. Belward (1992): The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series. *International Journal of Remote Sensing*, Vol. 13, No. 8, pp. 1585–1590.

# A Canonical Correlation Analysis

We consider a $p+q$ dimensional random variable $(p \leq q)$ ideally following a Gaussian distribution split into two groups of dimensions $p$ and $q$ respectively

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \in N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = N(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}) \tag{15}$$

and we assume that $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ (and $\boldsymbol{\Sigma}$) are non-singular. $\boldsymbol{X}$ and $\boldsymbol{Y}$ are multivariate images written as vectors at a given location

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \quad \text{resp.} \quad \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_q \end{bmatrix}. \tag{16}$$

We consider the conjugate eigenvectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ ($\boldsymbol{a}_i$ is $p \times 1$) corresponding to the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$ of $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ with respect to $\boldsymbol{\Sigma}_{11}$, i.e.

$$(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \lambda_i \boldsymbol{\Sigma}_{11})\boldsymbol{a}_i = \mathbf{0}. \tag{17}$$

If we put

$$\boldsymbol{b}_i = \frac{1}{\sqrt{\lambda_i}} \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{a}_i \tag{18}$$

($\boldsymbol{b}_i$ is $q \times 1$) we have

$$\begin{aligned} \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{b}_i &= \frac{1}{\sqrt{\lambda_i}} \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{a}_i & (19) \\ &= \sqrt{\lambda_i}\boldsymbol{\Sigma}_{21}\boldsymbol{a}_i & (20) \\ &= \lambda_i \boldsymbol{\Sigma}_{22}\boldsymbol{b}_i & (21) \end{aligned}$$

i.e. $\boldsymbol{b}_i$ is an eigenvector of $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ with respect to $\boldsymbol{\Sigma}_{22}$ corresponding to eigenvalue $\lambda_i$. If $p = q$ this will be all the eigenvalues and -vectors of $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. If $q > p$ then then last eigenvalue will be 0 with multiplicity $q - p$.

**Theorem**. Letting $\delta_{ij}$ be the Kronecker delta ($\delta_{ij} = 1$ for $i = j$, $\delta_{ij} = 0$ otherwise) we have

$$\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}_j = \boldsymbol{b}_i^T \boldsymbol{\Sigma}_{22} \boldsymbol{b}_j = \delta_{ij} \tag{22}$$

$$\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{12} \boldsymbol{b}_j = \sqrt{\lambda_i}\delta_{ij}. \tag{23}$$

**Proof**. The result for $\boldsymbol{a}_i$ follows by definition. We then obtain

$$\begin{aligned}
\boldsymbol{b}_i^T \boldsymbol{\Sigma}_{22} \boldsymbol{b}_j &= \frac{1}{\sqrt{\lambda_i \lambda_j}} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{22} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{a}_j &\quad (24)\\
&= \sqrt{\frac{\lambda_i}{\lambda_j}} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}_j &\quad (25)\\
&= \delta_{ij}. &\quad (26)
\end{aligned}$$

Similarly we obtain

$$\begin{aligned}
\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{12} \boldsymbol{b}_j &= \frac{1}{\sqrt{\lambda_j}} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{a}_j &\quad (27)\\
&= \sqrt{\lambda_j}\delta_{ij}. &\quad (28)
\end{aligned}$$

We are now able to introduce the canonical variates

$$U_i = \boldsymbol{a}_i^T \boldsymbol{X},\ i = 1, \ldots, p \tag{29}$$

$$V_i = \boldsymbol{b}_i^T \boldsymbol{Y},\ i = 1, \ldots, p \tag{30}$$

and with an obvious choice of notation

$$\boldsymbol{U} = \boldsymbol{A}^T \boldsymbol{X} \quad \text{and} \quad \boldsymbol{V} = \boldsymbol{B}^T \boldsymbol{Y}, \tag{31}$$

where

$$\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p] \text{ is } p \times p \tag{32}$$

$$\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p] \text{ is } q \times p. \tag{33}$$

We then have

**Theorem**. We consider the random variable $\boldsymbol{Z} = \boldsymbol{U} - \boldsymbol{V}$ and have that the dispersion matrix is

$$\mathrm{D}\{\boldsymbol{Z}\} = \mathrm{D}\{\boldsymbol{U} - \boldsymbol{V}\} = 2 \begin{bmatrix} 1 - \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \sqrt{\lambda_p} \end{bmatrix} = 2(\boldsymbol{I} - \boldsymbol{\Lambda}^{\frac{1}{2}}). \tag{34}$$

**Proof**. Straightforward

$$\begin{aligned} \mathrm{D}\{\boldsymbol{U} - \boldsymbol{V}\} &= \mathrm{D}\{\boldsymbol{U}\} + \mathrm{D}\{\boldsymbol{V}\} - 2\mathrm{Cov}\{\boldsymbol{U}, \boldsymbol{V}\} & (35) \\ &= \boldsymbol{A}^T \boldsymbol{\Sigma}_{11} \boldsymbol{A} + \boldsymbol{B}^T \boldsymbol{\Sigma}_{22} \boldsymbol{B} - 2\boldsymbol{A}^T \boldsymbol{\Sigma}_{12} \boldsymbol{B} & (36) \\ &= \boldsymbol{I} + \boldsymbol{I} - 2\Lambda^{\frac{1}{2}}. & (37) \end{aligned}$$

The covariance between the original and the MAD transformed variables are

$$\mathrm{Cov}\{\boldsymbol{X}, \boldsymbol{A}^T\boldsymbol{X} - \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{11}\boldsymbol{A} - \boldsymbol{\Sigma}_{12}\boldsymbol{B} \tag{38}$$

(39)

$$\mathrm{Cov}\{\boldsymbol{Y}, \boldsymbol{A}^T\boldsymbol{X} - \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{21}\boldsymbol{A} - \boldsymbol{\Sigma}_{22}\boldsymbol{B}. \tag{40}$$

## A.1 An Interpretation of Canonical Variates

Consider a regression $\hat{\boldsymbol{X}}$ of $\boldsymbol{X}$ based on $\boldsymbol{Y}$ respectively $\hat{\boldsymbol{Y}}$ of $\boldsymbol{Y}$ based on $\boldsymbol{X}$

$$\hat{\boldsymbol{X}} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{Y} \tag{41}$$

$$\hat{\boldsymbol{Y}} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X} \tag{42}$$

For the dispersions we get

$$\mathrm{D}\{\hat{\boldsymbol{X}}\} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{43}$$

$$\mathrm{D}\{\hat{\boldsymbol{Y}}\} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \tag{44}$$

Linear combinations $\boldsymbol{a}^T\hat{\boldsymbol{X}}$ respectively $\boldsymbol{b}^T\hat{\boldsymbol{Y}}$ that maximize $\frac{\mathrm{Var}\{\boldsymbol{a}^T\hat{\boldsymbol{X}}\}}{\mathrm{Var}\{\boldsymbol{a}^T\boldsymbol{X}\}}$ respectively $\frac{\mathrm{Var}\{\boldsymbol{b}^T\hat{\boldsymbol{Y}}\}}{\mathrm{Var}\{\boldsymbol{b}^T\boldsymbol{Y}\}}$ fulfill

$$\frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{a}} = \lambda \tag{45}$$

$$\frac{\boldsymbol{b}^T\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{b}}{\boldsymbol{b}^T\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{b}} = \lambda \tag{46}$$

As these so-called Rayleigh coefficients are identical to the defining equations above, the canonical variates can be interpreted as being new variables that maximize the ratio of the variances between linear combinations of predicted values of one set of variables from the other set of variables and the same linear combinations of the actual values of the one set of variables.

We also see that canonical correlation analysis can be considered as a type of regression analysis with several independent as well as dependent variables.

## A.2 A Minimizing Property of Canonical Variates

In this section we prove a property of the canonical variates. Normally the stepwise definition of canonical variates starts at the set with maximal correlation as mentioned in the introduction. From our point of view it will be more natural to start with the component of $\boldsymbol{Z}$ yielding the largest variance i.e. the canonical variates with the smallest correlation.

We assume that $\boldsymbol{c}^T\boldsymbol{X}$ is independent of $U_{j+1},\ldots,U_p$. We have

$$\mathrm{Cov}\{\boldsymbol{c}^T\boldsymbol{X},\boldsymbol{a}_k^T\boldsymbol{X}\} = \boldsymbol{c}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}_k = 0,\ k = j+1,\ldots,p. \tag{47}$$

Now, $\boldsymbol{c}$ may be written as $\gamma_1\boldsymbol{a}_1 + \cdots + \gamma_p\boldsymbol{a}_p$ and this implies

$$\boldsymbol{c}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}_k = \gamma_k = 0,\ k = j+1,\ldots,p \tag{48}$$

i.e. $\boldsymbol{c}^T\boldsymbol{X}$ may be written as

$$\boldsymbol{c}^T\boldsymbol{X} = (\gamma_1\boldsymbol{a}_1^T + \cdots + \gamma_j\boldsymbol{a}_j^T)\boldsymbol{X} = \gamma_1U_1 + \cdots + \gamma_jU_j. \tag{49}$$

Similarly, if $\boldsymbol{d}^T\boldsymbol{Y}$ is independent of $V_{j+1},\ldots,V_p$ we may write

$$\boldsymbol{d}^T\boldsymbol{Y} = \nu_1V_1 + \cdots + \nu_jV_j. \tag{50}$$

We now want to minimize the absolute value of the correlation between $\boldsymbol{c}^T\boldsymbol{X}$ and $\boldsymbol{d}^T\boldsymbol{Y}$ i.e. minimize

$$\frac{\boldsymbol{c}^T\boldsymbol{\Sigma}_{12}\boldsymbol{d}}{\boldsymbol{c}^T\boldsymbol{\Sigma}_{11}\boldsymbol{d}\boldsymbol{\Sigma}_{22}\boldsymbol{d}} = \frac{\boldsymbol{\gamma}_*^T\boldsymbol{\Lambda}_*^{\frac{1}{2}}\boldsymbol{\nu}_*}{\sqrt{\boldsymbol{\gamma}_*^T\boldsymbol{\gamma}_*\boldsymbol{\nu}_*^T\boldsymbol{\nu}_*}} \tag{51}$$

or equivalently minimize

$$(\gamma_1,\ldots,\gamma_j)\begin{bmatrix}\sqrt{\lambda_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{\lambda_j}\end{bmatrix}\begin{bmatrix}\nu_1\\ \vdots\\ \nu_j\end{bmatrix} = \boldsymbol{\gamma}_*^T\boldsymbol{\Lambda}_*^{\frac{1}{2}}\boldsymbol{\nu}_* \tag{52}$$

subject to the constraints

$$\boldsymbol{\gamma}_*^T \boldsymbol{\gamma}_* = 1 \quad \text{and} \quad \boldsymbol{\nu}_*^T \boldsymbol{\nu}_* = 1. \tag{53}$$

We introduce the Lagrange expression

$$F = \boldsymbol{\gamma}_*^T \boldsymbol{\Lambda}_*^{\frac{1}{2}} \boldsymbol{\nu}_* - \frac{\alpha}{2}(\boldsymbol{\gamma}_*^T \boldsymbol{\gamma}_* - 1) - \frac{\beta}{2}(\boldsymbol{\nu}_*^T \boldsymbol{\nu}_* - 1), \tag{54}$$

and have at optimum

$$\frac{\partial F}{\partial \gamma_*} = \boldsymbol{\Lambda}_*^{\frac{1}{2}} \boldsymbol{\nu}_* - \alpha \boldsymbol{\gamma}_* = 0 \Leftrightarrow \alpha \boldsymbol{\gamma}_* = \boldsymbol{\Lambda}_*^{\frac{1}{2}} \boldsymbol{\nu}_* \tag{55}$$

$$\frac{\partial F}{\partial \nu_*} = \boldsymbol{\Lambda}_*^{\frac{1}{2}} \boldsymbol{\gamma}_* - \beta \boldsymbol{\nu}_* = 0 \Leftrightarrow \beta \boldsymbol{\nu}_* = \boldsymbol{\Lambda}_*^{\frac{1}{2}} \boldsymbol{\gamma}_*. \tag{56}$$

We insert this in Equation 51 and obtain the expression

$$\frac{|\beta|}{\beta} \frac{\boldsymbol{\gamma}_*^T \boldsymbol{\Lambda}_* \boldsymbol{\gamma}_*}{\sqrt{\boldsymbol{\gamma}_*^T \boldsymbol{\gamma}_* \boldsymbol{\gamma}_*^T \boldsymbol{\Lambda}_* \boldsymbol{\gamma}_*}} = \text{sign}(\beta) \sqrt{\frac{\boldsymbol{\gamma}_*^T \boldsymbol{\Lambda}_* \boldsymbol{\gamma}_*}{\boldsymbol{\gamma}_*^T \boldsymbol{\gamma}_*}}. \tag{57}$$

The square of this expression is

$$\frac{\lambda_1 \gamma_1^2 + \cdots + \lambda_j \gamma_j^2}{\gamma_1^2 + \cdots + \gamma_j^2} \tag{58}$$

and this has a minimum for $\gamma_1 = \cdots = \gamma_{j-1} = 0$ and $\gamma_j = 1$, which corresponds to choosing $\boldsymbol{c}^T \boldsymbol{X}$ as $U_j$. We have now proven the following

**Theorem**. The canonical variates have the property that the $j$'th canonical variate shows minimal correlation amongst linear combinations independent of the previous $p - j$ least correlated canonical variates. In the case $q > p$ correlations between any of the $U$'s and the projection on the eigenvectors of $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ with respect to $\boldsymbol{\Sigma}_{22}$ corresponding to the eigenvalue 0 are exactly equal to 0.

## A.3 Redundancy Analysis

Here we shall study some further ways of analyzing the degree of overlap between the two sets of data. In order to clarify matters let us recall a fundamental property of the correlation coefficient $\rho$, namely that the squared correlation between two variables equals the fraction of the variation in one variable $Y$ that may be explained by an affine expression in the other variable $X$. If we call this predicted value $\hat{Y}$ we have

$$\text{Var}\{\hat{Y}\} = \rho_{YX}^2 \text{ Var}\{Y\}. \tag{59}$$

If the correlation equals 1 we have the same variance of the predicted value as of the original, i.e. we can make a perfect prediction of $Y$ based on $X$. If on the other hand the correlation equals 0 the predictor $X$ contains no information on $Y$.

If we have $k$ predictors $X_1, \ldots, X_k$ the expression is still valid if we replace the correlation with the multiple correlation coefficient

$$\text{Var}\{\hat{Y}\} = \rho_{Y|X_1,\ldots,X_k}^2 \text{ Var}\{Y\}. \tag{60}$$

The squared multiple correlation coefficient is thus the degree of variation in $Y$ that can be explained by $X_1, \ldots, X_k$.

After these more general remarks we shall revert to investigate some properties of the correlation structure of canonical variates. From the definitions we get

$$\text{Cov}\{\boldsymbol{X}, \boldsymbol{U}\} = \text{Cov}\{\boldsymbol{X}, \boldsymbol{A}^T\boldsymbol{X}\} = \boldsymbol{\Sigma}_{11}\boldsymbol{A} \tag{61}$$

$$\text{Cov}\{\boldsymbol{X}, \boldsymbol{V}\} = \text{Cov}\{\boldsymbol{X}, \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{12}\boldsymbol{B} \tag{62}$$

$$\text{Cov}\{\boldsymbol{Y}, \boldsymbol{V}\} = \text{Cov}\{\boldsymbol{Y}, \boldsymbol{B}^T\boldsymbol{Y}\} = \boldsymbol{\Sigma}_{22}\boldsymbol{B} \tag{63}$$

$$\text{Cov}\{\boldsymbol{Y}, \boldsymbol{U}\} = \text{Cov}\{\boldsymbol{Y}, \boldsymbol{A}^T\boldsymbol{X}\} = \boldsymbol{\Sigma}_{21}\boldsymbol{A} \tag{64}$$

We shall for simplicity and without lack of generality assume that the $X$'s an the $Y$'s are standardized, i.e. they have variance 1. Then the above matrices are correlation matrices.

The expression

$$f_j = \frac{1}{p} \sum_{i=1}^{p} [\text{Corr}\{X_i, U_j\}]^2 = \frac{1}{p}\, \boldsymbol{a}_j^T \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{11} \boldsymbol{a}_j \tag{65}$$

equals the fraction of the (standardized) variance of the original $X$ variables that are explained by canonical variate $U_j$. It follows that

$$\sum_{j=1}^{p} f_j = \sum_{j=1}^{p} \frac{1}{p}\, \boldsymbol{a}_j^T \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{11} \boldsymbol{a}_j = \frac{1}{p} \,\text{tr}\, (\boldsymbol{A}^T \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{11} \boldsymbol{A}), \tag{66}$$

where tr denotes the trace of a matrix. From $\boldsymbol{A}^T \boldsymbol{\Sigma}_{11} \boldsymbol{A} = \boldsymbol{I}$ we obtain $\boldsymbol{A}^T = \boldsymbol{A}^{-1} \boldsymbol{\Sigma}_{11}$ and therefore

$$\sum_{j=1}^{p} f_j = \frac{1}{p} \,\text{tr}\, (\boldsymbol{A}^{-1} \boldsymbol{\Sigma}_{11} \boldsymbol{A}) = \frac{1}{p} \,\text{tr}\, (\boldsymbol{\Sigma}_{11} \boldsymbol{A} \boldsymbol{A}^{-1}) = 1. \tag{67}$$

This corresponds to the fact that all of the variation in the original variables $X_1, \ldots, X_p$ may be explained by the whole set of canonical variates $U_1, \ldots, U_p$.

If we multiply $f_j$ with the $j$'th squared canonical correlation $\lambda_j$ we obtain the so called redundancy factor $R_j$. It may be instructive to note that

$$\begin{aligned}
R_j &= \lambda_j f_j && (68)\\
&= \lambda_j \frac{1}{p}\, \boldsymbol{a}_j^T \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{11} \boldsymbol{a}_j && (69)\\
&= \frac{1}{\lambda_j} \frac{1}{p}\, \boldsymbol{a}_j^T \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{a}_j && (70)\\
&= \frac{1}{p}\, \boldsymbol{b}_j^T \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{12} \boldsymbol{b}_j && (71)\\
&= \frac{1}{p} \sum_{i=1}^{p} [\text{Corr}\{X_i, V_j\}]^2 && (72)
\end{aligned}$$

Introducing the mnemotechnical expressions $CVX_i$ for $U_i$ and $CVY_i$ for $V_i$ and we obtain

$$
\begin{aligned}
R_j &= \frac{1}{p}\,\mathrm{cancor}_j^2 \sum_{i=1}^{p} [\mathrm{Corr}\{X_i, CVX_j\}]^2 &(73)\\
&= \frac{1}{p} \sum_{i=1}^{p} [\mathrm{Corr}\{X_i, CVY_j\}]^2. &(74)
\end{aligned}
$$

The squared canonical correlation is the shared variability between the two sets of canonical variates. Consequently we have the following interpretation of the redundancy factor of the first set of variables given the availability of the second set of variables: $R_j$ expresses the amount of variation in the original variables that is explained by the $j$'th canonical variate adjusted with the shared variation between the $j$'th canonical variates. This equals the amount of variation of variation explained by the opposite $j$'th canonical variate.

Related to these considerations are properties of multiple squared correlations between the original variables and the set of opposite canonical variates. If $\boldsymbol{\sigma}_i^T$ denotes the $i$'th row in $\boldsymbol{\Sigma}_{12}$ the squared multiple correlation between $X_i$ and $CVY_1, \ldots, CVY_k$ is

$$
\begin{aligned}
\rho_{X_i|CVY_1,\ldots,CVY_k} &= \boldsymbol{\sigma}_i^T [\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k] \begin{bmatrix} \boldsymbol{b}_1^T \\ \vdots \\ \boldsymbol{b}_k^T \end{bmatrix} \boldsymbol{\sigma}_i &(75)\\
&= \sum_{j=1}^{k} [\boldsymbol{\sigma}_i \boldsymbol{b}_j]^2 &(76)\\
&= \sum_{j=1}^{k} [\mathrm{Corr}\{X_i, CVY_j\}]^2, &(77)
\end{aligned}
$$

a quantity that is also useful in assessing overlap between the two sets of variables.

# B Min/Max Autocorrelation Factor Analysis

We consider the random variable $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$ and we assume that

$$
\begin{aligned}
\mathrm{E}\{\boldsymbol{Z}(\boldsymbol{x})\} &= \boldsymbol{0} &(78)\\
\mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} &= \boldsymbol{\Sigma}. &(79)
\end{aligned}
$$

We denote a spatial shift by $\boldsymbol{\Delta}^T = [\Delta_x, \Delta_y]$. The spatial covariance function is defined by

$$\mathrm{Cov}\{\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{\Gamma}(\boldsymbol{\Delta}). \tag{80}$$

$\boldsymbol{\Gamma}$ has the following properties

$$\begin{aligned} \boldsymbol{\Gamma}(\boldsymbol{0}) &= \boldsymbol{\Sigma} && (81)\\ \boldsymbol{\Gamma}(\boldsymbol{\Delta})^T &= \boldsymbol{\Gamma}(-\boldsymbol{\Delta}). && (82) \end{aligned}$$

We are interested in the correlations between projections of the variables and the shifted variables. Therefore we find

$$\begin{aligned} \mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} &= \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})\boldsymbol{a} && (83)\\ &= (\boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})\boldsymbol{a})^T\\ &= \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})^T\boldsymbol{a}\\ &= \boldsymbol{a}^T\boldsymbol{\Gamma}(-\boldsymbol{\Delta})\boldsymbol{a}\\ &= \frac{1}{2}\boldsymbol{a}^T\left(\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta})\right)\boldsymbol{a}. \end{aligned}$$

Introducing

$$\begin{aligned} \boldsymbol{\Sigma}_\Delta &= \mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} && (84)\\ &= \mathrm{E}\{[\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})][\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})]^T\}, \end{aligned}$$

(which considered as a function of $\boldsymbol{\Delta}$ is a multivariate variogram) we have

$$\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta}) = 2\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\Delta \tag{85}$$

and thus

$$\mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{a}^T(\boldsymbol{\Sigma} - \frac{1}{2}\boldsymbol{\Sigma}_\Delta)\boldsymbol{a} \tag{86}$$

wherefore

$$\mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}}. \tag{87}$$

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

$$R(\boldsymbol{a}) = \frac{\boldsymbol{a}^T \boldsymbol{\Sigma}_\Delta \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}}. \tag{88}$$

Let $\kappa_1 \geq \cdots \geq \kappa_m$ be the eigenvalues and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ corresponding conjugate eigenvectors of $\boldsymbol{\Sigma}_\Delta$ with respect to $\boldsymbol{\Sigma}$. Then

$$\boldsymbol{Y}_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{Z}_i(\boldsymbol{x}) \tag{89}$$

is the $i$'th minimum/maximum autocorrelation factor or shortly the $i$'th MAF.

The minimum/maximum autocorrelation factors satisfy

i) $\mathrm{Corr}\{\boldsymbol{Y}_i(\boldsymbol{x}), \boldsymbol{Y}_j(\boldsymbol{x})\} = 0,\ i \neq j$,

ii) $\mathrm{Corr}\{\boldsymbol{Y}_i(\boldsymbol{x}), \boldsymbol{Y}_i(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\kappa_i$,

iii) $\mathrm{Corr}\{\boldsymbol{Y}_1(\boldsymbol{x}), \boldsymbol{Y}_1(\boldsymbol{x}+\boldsymbol{\Delta})\} = \inf_a\ \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\}$,
$\mathrm{Corr}\{\boldsymbol{Y}_m(\boldsymbol{x}), \boldsymbol{Y}_m(\boldsymbol{x}+\boldsymbol{\Delta})\} = \sup_a\ \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\}$,
$\mathrm{Corr}\{\boldsymbol{Y}_i(\boldsymbol{x}), \boldsymbol{Y}_i(\boldsymbol{x}+\boldsymbol{\Delta})\} = \inf_{a\in\mathcal{M}_i}\ \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\}$,
$\mathcal{M}_i = \{\boldsymbol{a} \mid \mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{Y}_j(\boldsymbol{x})\} = 0,\ j = 1, \ldots, i-1\}$.

The reverse numbering of MAFs so that the signal MAF is referred to as the first MAF or MAF1 is often used, also in this paper.

## B.1 Linear Transformations of MAFs

We now consider the problem of transforming the original variables. If we set

$$\boldsymbol{U}(\boldsymbol{x}) = \boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) \tag{90}$$

where $\boldsymbol{T}$ is a transformation matrix, we have that the MAF solution for $\boldsymbol{U}$ is obtained by investigating

$$R_1(\boldsymbol{b}) = \frac{\boldsymbol{b}^T\boldsymbol{T}\boldsymbol{\Sigma}_\Delta\boldsymbol{T}^T\boldsymbol{b}}{\boldsymbol{b}^T\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\boldsymbol{b}}. \tag{91}$$

The equation for solving the eigenproblem is

$$\begin{aligned}
\boldsymbol{T}\boldsymbol{\Sigma}_\Delta\boldsymbol{T}^T\boldsymbol{b}_i &= \lambda_i\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\boldsymbol{b}_i \Leftrightarrow \\
\boldsymbol{\Sigma}_\Delta(\boldsymbol{T}^T\boldsymbol{b}_i) &= \lambda_i\boldsymbol{\Sigma}(\boldsymbol{T}^T\boldsymbol{b}_i)
\end{aligned} \tag{92}$$

i.e. the eigenvalues are unchanged and $\boldsymbol{T}^T\boldsymbol{b}_i = \boldsymbol{a}_i$ is an eigenvector for $\boldsymbol{\Sigma}_\Delta$ with respect to $\boldsymbol{\Sigma}$. We find that the MAFs in the transformed problem are

$$\begin{aligned}
\boldsymbol{b}_i^T\boldsymbol{U}(\boldsymbol{x}) &= \boldsymbol{b}_i^T\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) \\
&= (\boldsymbol{T}^T\boldsymbol{b}_i)^T\boldsymbol{Z}(\boldsymbol{x}) \\
&= \boldsymbol{a}_i^T\boldsymbol{Z}(\boldsymbol{x}) \\
&= \boldsymbol{Y}_i(\boldsymbol{x}).
\end{aligned} \tag{93}$$

Therefore the MAF solution is invariant to linear transformations, which can be useful in computations. Let $\lambda_1 \geq \cdots \geq \lambda_m$ be the ordinary eigenvalues and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ corresponding orthonormal eigenvectors of $\boldsymbol{\Sigma}$. If we set ($\boldsymbol{P} = [\boldsymbol{p}_1 \cdots \boldsymbol{p}_m]$)

$$\boldsymbol{T}^T = \boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}} \tag{94}$$

we have for the dispersion of the transformed variables

$$\mathrm{D}\{\boldsymbol{U}(\boldsymbol{x})\} = \mathrm{D}\{\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{\Sigma}\boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \boldsymbol{I}. \tag{95}$$

With this transformation the original generalized eigenproblem is reduced to an ordinary eigenproblem for

$$\begin{aligned}
\boldsymbol{T}\boldsymbol{\Sigma}_\Delta\boldsymbol{T}^T &= \mathrm{D}\{\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} \\
&= \mathrm{D}\{\boldsymbol{U}(\boldsymbol{x}) - \boldsymbol{U}(\boldsymbol{x}+\boldsymbol{\Delta})\}
\end{aligned} \tag{96}$$

and the MAF solution can be obtained by solving two ordinary eigenproblems as follows

- calculate principal components from the usual dispersion matrix $\boldsymbol{\Sigma}$,
- calculate dispersion matrix for shifted principal components $\boldsymbol{P}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{P}$,
- calculate principal components for transformed data corresponding to $\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}$.

The original generalized eigenproblem can be solved by means of Cholesky factorization of $\boldsymbol{\Sigma}$ also.

As far as the practical computation of $\hat{\boldsymbol{\Sigma}}_\Delta$ is concerned Switzer and Green (1984) recommend the formation of two sets of difference images. The two sets are $\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x} + \boldsymbol{\Delta}_h)$ and $\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x} + \boldsymbol{\Delta}_v)$ where $\boldsymbol{\Delta}_h$ is a unit horizontal shift and $\boldsymbol{\Delta}_v$ is a unit vertical shift. Calculate $\hat{\boldsymbol{\Sigma}}_{\Delta_h}$ and $\hat{\boldsymbol{\Sigma}}_{\Delta_v}$ and pool them to obtain $\hat{\boldsymbol{\Sigma}}_\Delta$.