Multivariate Change Detection in Multispectral, Multitemporal Images

Knut Conradsen
Allan Aasbjerg Nielsen

IMSOR
Institute of Mathematical Statistics and Operations Research
Technical University of Denmark, Building 321
DK-2800 Lyngby, Denmark

Abstract

This paper introduces a new orthogonal transform the multivariate change detection (MCD) transform based on an established multivariate statistical technique canonical correlation analysis. The theory for canonical correlation analysis is sketched and modified to be more directly applicable in our context. As opposed to traditional univariate change detection schemes our scheme transforms two sets of multivariate observations (e.g., two multispectral satellite images 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. A case study using multispectral SPOT data from 1987 and 1989 covering coffee and pineapple plantations near Thika, Kiambu District, Kenya, shows the usefulness of this new concept.

1 Introduction

When analyzing changes in panchromatic imagery taken at different points in time it is customary to analyze the difference between two images.

If we have multivariate images with outcomes at a given pixel

\[ X = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \quad \text{resp.} \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} \]
then a naive change detection transform would be

\[ X - Y = \begin{bmatrix} X_1 - Y_1 \\ \vdots \\ X_k - Y_k \end{bmatrix} \]

If we have seasonal changes in say vegetation cover this would cause shifts of "energy" from one wavelength to others and the changes may therefore be "smeared" out. This would make it difficult to actually see the changes in all the channels. The problem is obviously that the channels are correlated. We shall therefore try to overcome this problem by looking at linear combinations

\[ a'X = a_1X_1 + \cdots + a_kX_k \]

\[ b'Y = b_1Y_1 + \cdots + b_kY_k \]

and then consider the difference between these, i.e.

\[ a'X - b'Y. \]

For any choice of \( a \) and \( b \) this will give a measure of change. One could use principal components analysis on \( X \) to find an optimal \( a \) and on \( Y \) to find an optimal \( b \) (independent of \( a \)). An improvement of this technique is to use principal components analysis on \( X \) and \( Y \) considered as one variable, cf. Fung and LeDrew (1987). This approach does not, however, guarantee a separation of \( X \) and \( Y \). It might as well give, say, differences between shortwave and longwave image bands. A better approach is to define an optimal set of \( a \) and \( b \) simultaneously. Emphasizing changes is the same as saying we want the difference to show a great variation or in other words that we want to maximize the variance \( V(a'X - b'Y) \). Now, trivially a multiplication of \( a \) and \( b \) with a constant \( c \) will multiply the variance with \( c^2 \). Therefore we must put some restrictions on \( a \) and \( b \), and natural restrictions are requesting unit variance of \( a'X \) and \( b'Y \).

The criterion would thus be

maximize \( V(a'X - b'Y) \)

subject to the constraints

\[ V(a'X) = 1 \quad \text{and} \quad V(b'Y) = 1. \]

Under these constraints we have

\[ V(a'X - b'Y) = V(a'X) + V(b'Y) - 2Cov(a'X, b'Y) \]

\[ = 2(1 - Corr(a'X, b'Y)). \]

Therefore, determining the combinations with maximum variance corresponds to having minimum correlation.

In the sequel we shall assume that \( a \) and \( b \) are chosen so that the correlation between \( a'X \) and \( b'Y \) 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 into 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 second 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 is described in most standard textbooks on multivariate statistics, cf. e.g. Anderson (1984).

The main idea presented in this paper is now to try to modify the theory used in defining canonical variates. This could be viewed upon as a time analogue to the introduction of minimum/maximum autocorrelation factors in the spatial domain. Minimum/Maximum autocorrelation factor (MAF) analysis was first described by Switzer and Green, 1984. We shall therefore summarize the theory of canonical correlations and then modify the theorems so that they are more directly applicable in our context.

2 Canonical Correlation Analysis

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

$$
\begin{bmatrix}
X \\
Y
\end{bmatrix} \in N(\mu, \Sigma) = N\left(\begin{bmatrix}
\mu_1 \\
\mu_2
\end{bmatrix}, \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix}\right)
$$

and we assume that $\Sigma_{11}$ and $\Sigma_{22}$ (and $\Sigma$) are non-singular.

We consider the conjugate eigenvectors $a_1, \ldots, a_p$ corresponding to the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$ of $\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ with respect to $\Sigma_{11}$, i.e.

$$
(\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \lambda_i \Sigma_{11})a_i = 0.
$$

If we put

$$
b_i = \frac{1}{\sqrt{\lambda_i}} \Sigma_{22}^{-1} \Sigma_{21} a_i
$$

we have

$$
\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \frac{1}{\sqrt{\lambda_i}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} a_i
$$

$$
= \sqrt{\lambda_i} a_i
$$

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

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

\[
\begin{align*}
a_i' \Sigma_{11} a_j &= b_i' \Sigma_{22} b_j = \delta_{ij} \\
a_i' \Sigma_{12} b_j &= \sqrt{\lambda_i} \delta_{ij}.
\end{align*}
\]

**Proof.** The result for \( a_i \) follows by definition. We then obtain

\[
\begin{align*}
b_i' \Sigma_{22} b_j &= \frac{1}{\sqrt{\lambda_i \lambda_j}} a_i' \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{22} \Sigma_{22}^{-1} \Sigma_{21} a_j \\
&= \sqrt{\frac{\lambda_i}{\lambda_j}} a_i' \Sigma_{11} a_j \\
&= \delta_{ij}.
\end{align*}
\]

Similarly we obtain

\[
\begin{align*}
a_i' \Sigma_{12} b_j &= \frac{1}{\sqrt{\lambda_j}} a_i' \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} a_j \\
&= \sqrt{\lambda_j} \delta_{ij}.
\end{align*}
\]

We are now able to introduce the canonical variates

\[
U_i = a_i' X, \ i = 1, \ldots, p
\]

\[
V_i = b_i' Y, \ i = 1, \ldots, p
\]

and with an obvious choice of notation

\[
U = A' X \quad \text{and} \quad V = B' Y,
\]

where

\[
A = (a_1, \ldots, a_p) \text{ is } p \times p
\]

\[
B = (b_1, \ldots, b_p) \text{ is } q \times p.
\]

We then have

**Theorem.** We consider the random variable \( Z = U - V \) and have that the dispersion matrix is

\[
D(Z) = D(U - V) = 2 \begin{bmatrix} 1 - \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \sqrt{\lambda_p} \end{bmatrix} = 2(I - A^{\frac{1}{2}}).
\]
Proof. Straightforward

\[
D(U - V) = D(U) + D(V) - 2\text{Cov}(U, V) \\
= A'\Sigma_{11}A + B'\Sigma_{22}B - 2A'\Sigma_{12}B \\
= I + I - 2A\frac{1}{2}.
\]

\[\square\]

3 A Minimizing Property of 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 \(Z\) yielding the largest variance i.e. the canonical variates with the smallest correlation.

We assume that \(c'X\) is independent of \(U_{j+1}, \ldots, U_p\). We have

\[
\text{Cov}(c'X, a'_kX) = c'\Sigma_{11}a_k = 0, \ k = j + 1, \ldots, p.
\]

Now, \(c\) may be written as \(\gamma_1a_1 + \cdots + \gamma_pa_p\) and this implies

\[
c'\Sigma_{11}a_k = \gamma_k = 0, \ k = j + 1, \ldots, p
\]

i.e. \(c'X\) may be written as

\[
c'X = (\gamma_1a_1' + \cdots + \gamma_ja_j')X = \gamma_1U_1 + \cdots + \gamma_jU_j.
\]

Similarly, if \(d'Y\) is independent of \(V_{j+1}, \ldots, V_p\) we may write

\[
d'Y = \nu_1V_1 + \cdots + \nu_jV_j.
\]

We now want to minimize the absolute value of the correlation between \(c'X\) and \(d'Y\) i.e.

minimize

\[
\frac{c'\Sigma_{12}d}{c'\Sigma_{11}d\Sigma_{22}d} = \frac{\gamma'\Lambda^{\frac{1}{2}}\nu}{\sqrt{\gamma'\gamma}\nu'\nu}.
\]

or equivalently minimize

\[
(\gamma_1, \ldots, \gamma_j) \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_j} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_j \end{bmatrix} = \gamma'\Lambda^{\frac{1}{2}}\nu
\]

subject to the constraints

\[
\gamma'\gamma = 1 \quad \text{and} \quad \nu'\nu = 1.
\]
We introduce the Lagrange expression 
\[ F = \gamma^* \Lambda_{1/2} \nu^* - \frac{\alpha}{2} (\gamma^* \gamma_\nu - 1) - \frac{\beta}{2} (\nu^* \nu_\gamma - 1), \]
and have at optimum 
\[ \frac{\partial F}{\partial \gamma^*} = \Lambda_{1/2} \nu^* - \alpha \gamma^* = 0 \Leftrightarrow \alpha \gamma^* = \Lambda_{1/2} \nu^*. \]
\[ \frac{\partial F}{\partial \nu^*} = \Lambda_{1/2} \gamma^* - \beta \nu^* = 0 \Leftrightarrow \beta \nu^* = \Lambda_{1/2} \gamma^*. \]
We insert this in Equation 1 and obtain the expression 
\[ \frac{\beta}{\gamma^* \Lambda_{1/2} \gamma^*} \gamma^* \Lambda_{1/2} \gamma^* = \text{sign}(\beta) \sqrt{\gamma^* \Lambda_{1/2} \gamma^*}. \]
The square of this expression is 
\[ \frac{\gamma^2_1 \gamma^2_2 + \cdots + \gamma^2_j}{\gamma^2_1 + \cdots + \gamma^2_j} \]
and this has a minimum for \( \gamma_1 = \cdots = \gamma_{j-1} = 0 \) and \( \gamma_j = 1 \), which corresponds to choosing \( c^* X \) 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 \( U_j \)'s and the projection on the eigenvectors of \( \Sigma_1, \Sigma_1^{-1} \Sigma_{12} \) with respect to \( \Sigma_{22} \) corresponding to the eigenvalue 0 are exactly equal to 0.

### 4 The Multivariate Change Detection (MCD) Transform

Having established this result we are now ready to define the multivariate change detection (MCD) transform as 
\[
\begin{bmatrix}
X \\
Y
\end{bmatrix} \rightarrow U - V = A'X - B'Y = \begin{bmatrix}
a'_1X - b'_1Y \\
\vdots \\
a'_pX - b'_pY
\end{bmatrix},
\]
where \( A \) and \( B \) etc. are defined as above i.e. \( A \) and \( B \) are the defining coefficients from a standard canonical correlation analysis. The MCD transform has the very important property that if we consider linear combinations of the two sets of variables that are positively correlated then the \( p \)'th difference will show
maximum variance among such variables. The \((p - j)\)th difference will show maximum variance subject to the constraint that it should be uncorrelated with the previous \(j\) ones etc. In this way we may sequentially extract uncorrelated difference images where each new image shows maximum change under the constraint of being uncorrelated with the previous ones.

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

5 Case Study – SPOT XS Data, Kenya

A 512×512 sub-scene from multispectral 1987 and 1989 SPOT XS data is used as a pilot area. This area contains economically important coffee and pineapple fields near Thika, Kiambu District, Kenya. In Plate 1 we show false colour composites of multispectral SPOT scenes from 1987 and 1989. The area is dominated by large pineapple fields to the northeast and coffee fields to the northwest. To the south is Thika town. We also show the naïve change detection image. The major differences are due to the changes primarily in the pineapple fields. Pineapple is a triennial crop and therefore we observe changes from one year to another. Since the changes are connected to change in vegetation it seems natural to study the change in the normalized difference vegetation index

\[
\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R + 1}
\]

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. 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 the result of which was presented as a video was made by Stern (1989). In Plate 1 we finally 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 naïve change detection image. However, this enhancement of differences may also be due to differences between fields in the same season. In Plate 2 we show the canonical variates for the 1987 and the 1989 data. Enhancing many interesting features it readily follows that these images are useful by themselves. We also show the MCD₃ variable, i.e. the difference between linear combinations from the two years showing maximal difference. In Plate 2 we finally show all three MCDs. An inspection of these images and a comparison with the naïve change detection image show that there is a much better distinction between different types of changes. In the naïve change detection image cyan is dominating but in the MCD image we see that a much better discrimination has been achieved. We therefore conclude that the MCD transformation is a useful extension of more naïve multivariate change detection schemes.
6 Annex

In this annex we show the numerical results from the canonical correlation analysis of the two scenes. We first report the basic statistics.

**Means and Standard Deviations**

<table>
<thead>
<tr>
<th></th>
<th>1987 Mean</th>
<th>1989 Mean</th>
<th>Std Dev</th>
<th>1987 Std Dev</th>
<th>1989 Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>XS1</td>
<td>45.00</td>
<td>32.27</td>
<td>5.40</td>
<td>4.79</td>
<td></td>
</tr>
<tr>
<td>XS2</td>
<td>36.86</td>
<td>22.88</td>
<td>7.12</td>
<td>4.87</td>
<td></td>
</tr>
<tr>
<td>XS3</td>
<td>74.15</td>
<td>62.33</td>
<td>12.55</td>
<td>10.66</td>
<td></td>
</tr>
</tbody>
</table>

**Correlations Among the Original Variables**

<table>
<thead>
<tr>
<th></th>
<th>1987</th>
<th>1989</th>
<th>1987</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>XS1</td>
<td>1.0000</td>
<td>0.5116</td>
<td>0.9057</td>
<td>0.4352</td>
</tr>
<tr>
<td>XS2</td>
<td>0.9057</td>
<td>0.4352</td>
<td>1.0000</td>
<td>0.4140</td>
</tr>
<tr>
<td>XS3</td>
<td>-0.3336</td>
<td>-0.0082</td>
<td>-0.4196</td>
<td>-0.0381</td>
</tr>
</tbody>
</table>

It is noted that the correlation structure is basically the same in the two years considered.

**Canonical Correlation Analysis**

<table>
<thead>
<tr>
<th></th>
<th>Approx CAN Correlation</th>
<th>Squared Canonical Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6505</td>
<td>0.4232</td>
</tr>
<tr>
<td>2</td>
<td>0.4024</td>
<td>0.1619</td>
</tr>
<tr>
<td>3</td>
<td>0.2403</td>
<td>0.0577</td>
</tr>
</tbody>
</table>

Test of $H_0$: The canonical correlations in the current row and all that follow are zero.

<table>
<thead>
<tr>
<th></th>
<th>Likelihood Ratio</th>
<th>Approx F</th>
<th>DF</th>
<th>Numerator DF</th>
<th>Denominator DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4555</td>
<td>27039</td>
<td>9</td>
<td>637975</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7897</td>
<td>16425</td>
<td>4</td>
<td>524278</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.9423</td>
<td>16058</td>
<td>1</td>
<td>262140</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>
Standardized canonical coefficients

<table>
<thead>
<tr>
<th></th>
<th>1987</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAN1</td>
<td>CAN2</td>
</tr>
<tr>
<td>XS1</td>
<td>-1.8816</td>
<td>-0.6862</td>
</tr>
<tr>
<td>XS2</td>
<td>1.5328</td>
<td>1.6894</td>
</tr>
<tr>
<td>XS3</td>
<td>0.5938</td>
<td>0.4081</td>
</tr>
</tbody>
</table>

Canonical Structure
Correlations between original variables and canonical variables

<table>
<thead>
<tr>
<th></th>
<th>1987</th>
<th>1989</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAN1</td>
<td>CAN2</td>
</tr>
<tr>
<td>XS1</td>
<td>-0.6915</td>
<td>0.7078</td>
</tr>
<tr>
<td></td>
<td>-0.4206</td>
<td>0.8967</td>
</tr>
<tr>
<td></td>
<td>0.5784</td>
<td>-0.0719</td>
</tr>
<tr>
<td>XS2</td>
<td>-0.5021</td>
<td>0.2423</td>
</tr>
<tr>
<td></td>
<td>-0.2667</td>
<td>0.3201</td>
</tr>
<tr>
<td></td>
<td>0.1050</td>
<td>0.0429</td>
</tr>
</tbody>
</table>

Canonical Redundancy Analysis
Standardized variance of 1987 XS

<table>
<thead>
<tr>
<th>Explained by</th>
<th>Their Own Canonical Variables</th>
<th>The Opposite Canonical Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cumulative Proportion</td>
<td>Canonical R^2</td>
</tr>
<tr>
<td>1</td>
<td>0.3299</td>
<td>0.4232</td>
</tr>
<tr>
<td>2</td>
<td>0.4368</td>
<td>0.1619</td>
</tr>
<tr>
<td>3</td>
<td>0.2333</td>
<td>0.0577</td>
</tr>
</tbody>
</table>

Standardized variance of 1989 XS

<table>
<thead>
<tr>
<th>Explained by</th>
<th>Their Own Canonical Variables</th>
<th>The Opposite Canonical Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cumulative Proportion</td>
<td>Canonical R^2</td>
</tr>
<tr>
<td>1</td>
<td>0.2632</td>
<td>0.4232</td>
</tr>
<tr>
<td>2</td>
<td>0.3356</td>
<td>0.1619</td>
</tr>
<tr>
<td>3</td>
<td>0.4012</td>
<td>0.0577</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>XS1</td>
<td>0.2024 0.2835 0.2847</td>
<td>0.2521 0.3108 0.3132</td>
</tr>
<tr>
<td>XS2</td>
<td>0.0749 0.2051 0.2062</td>
<td>0.0711 0.1736 0.1851</td>
</tr>
<tr>
<td>XS3</td>
<td>0.1416 0.1424 0.1805</td>
<td>0.0110 0.0129 0.0684</td>
</tr>
</tbody>
</table>

Thus the canonical variates for the 1987 XS data are

$$\begin{bmatrix}
\text{CAN1} \\
\text{CAN2} \\
\text{CAN3}
\end{bmatrix} = \begin{bmatrix}
-1.8816 & 1.5238 & 0.5938 \\
-0.6862 & 1.6894 & 0.4081 \\
1.2787 & -0.9417 & 0.8441
\end{bmatrix} \begin{bmatrix}
\text{XS1} - 45.00) / 5.40 \\
\text{XS2} - 36.86) / 7.12 \\
\text{XS3} - 74.15) / 12.6
\end{bmatrix}$$

and the canonical variates for the 1989 XS data are

$$\begin{bmatrix}
\text{CAN1} \\
\text{CAN2} \\
\text{CAN3}
\end{bmatrix} = \begin{bmatrix}
-2.0441 & 1.5120 & 0.2616 \\
-0.8151 & 1.7877 & 0.6431 \\
0.4247 & -0.4430 & 0.9063
\end{bmatrix} \begin{bmatrix}
\text{XS1} - 32.27) / 4.79 \\
\text{XS2} - 22.88) / 4.87 \\
\text{XS3} - 62.33) / 10.7
\end{bmatrix}$$

Acknowledgements

The authors wish to thank Bjarne Kjær Ersbøll, IMSOR, for many good discussions on multivariate statistics and analysis of spatial data.

The work reported was funded by Danida, the Danish International Development Agency, under grant No. 104 Dan.8/410.

References


M. Stern (1989): *pers. comm.*, Space Media Networks/IMSOR.

Plate 1  (left-to-right top-to-bottom): (a) 1987 SPOT XS, (b) 1989 SPOT XS, (c) Naive Change SPOT XS, (d) Change SPOT NDVI.
Plate 2  (right-to-left top-to-bottom): (a) 1987 SPOT Can. Var., (b) 1989 SPOT Can. Var., (c) SPOT MCD$_3$, (d) SPOT MCD.