Multiset Canonical Correlations Analysis and Multispectral, Truly Multitemporal Remote Sensing Data

Allan Aasbjerg Nielsen

***Abstract*—This paper describes two- and multiset canonical correlations analysis (CCA) for data fusion, multisource, multiset, or multitemporal exploratory data analysis. These techniques transform multivariate multiset data into new orthogonal variables called canonical variates (CVs) which, when applied in remote sensing, exhibit ever-decreasing similarity (as expressed by correlation measures) over sets consisting of 1) spectral variables at fixed points in time (R-mode analysis), or 2) temporal variables with fixed wavelengths (T-mode analysis). The CVs are invariant to linear and affine transformations of the original variables within sets which means, for example, that the R-mode CVs are insensitive to changes over time in offset and gain in a measuring device. In a case study, CVs are calculated from Landsat TM data with six spectral bands over six consecutive years. Both R- and T-mode CVs clearly exhibit the desired characteristic: they show maximum similarity for the low-order canonical variates and minimum similarity for the high-order canonical variates. These characteristics are seen both visually and in objective measures. The results from the multiset CCA R- and T-mode analyses are very different. This difference is ascribed to the noise structure in the data. The CCA methods are related to partial least squares (PLS) methods. This paper very briefly describes multiset CCA-based multiset PLS. Also, the CCA methods can be applied as multivariate extensions to empirical orthogonal functions (EOF) techniques. (Multiset) CCA is well-suited for inclusion in geographical information systems (GIS).**

***Index Terms*—Geographical information systems (GIS), minimum and maximum similarity variates, multiset partial least squares (PLS), multisource data fusion, multivariate empirical orthogonal functions (EOF).**

I. Introduction

THIS paper deals with multiset canonical correlations analysis (MCCA) for data fusion, multisource, multiset, or multitemporal exploratory data analysis. MCCA deals with data that naturally splits up into more (than two) groups of variables, e.g., multispectral satellite data covering the same geographical region over several points in time.

In Section II, ordinary two-set canonical analysis is described. Two-set canonical correlations analysis investigates the relationship between two groups of variables. It finds corresponding sets of linear combinations of the two groups of original variables with maximum correlation. Partial least squares regression is mentioned in Section II also.

In Section III, this analysis is generalized to deal with more than two sets of variables. The idea is to optimize characteristics of the dispersion matrix of the transformed variables to obtain high correlations between all new variables simultaneously. These characteristics include

- maximization of the sum of the elements;
- maximization of the sum of the squared elements;
- maximization of the largest eigenvalue;
- minimization of the smallest eigenvalue;
- minimization of the determinant.

These measures are not confined and the optimizations take place subject to different chosen constraints and orthogonality criteria. The latter are briefly mentioned in Section IV, which also describes computer implementations of the techniques.

Results from such analyses are linear combinations termed canonical variates (CVs) that when used with remote sensing data transform the original data into new orthogonal variables that show decreasing similarity over sets consisting of

1) spectral variables at fixed points in time (R-mode analysis);
2) temporal variables with fixed wavelengths (T-mode analysis).

The higher order canonical variates exhibit minimum similarity and they are therefore measures of differences in all variables simultaneously.

Multiset partial least squares methods emerge from this type of description with a special choice of optimization criteria, constraints, and orthogonality criteria leading to an optimization of covariance rather than correlation measures.

If applied to several variables that change over time, this type of analysis constitutes a multivariate extension to the technique of empirical orthogonal functions (EOF) [1] often applied in geophysical data analysis.

In Section V, a Landsat TM case with data from 1984 to 1989 covering a small forested region in northern Sweden is used to illustrate the technique. The purpose of the case study is to demonstrate the method and to suggest a possible way of interpreting the resulting transformed variables. It is not the purpose to assess the similarity (or lack of similarity, i.e., change) over time on the ground or in the atmosphere.

Multiset or multisource data analysis techniques such as the application of the Mahalanobis distance in joint distributions of multiset data to point out potential mineralization areas [2]

Manuscript received July 23, 1999; revised June 22, 2001. This work was supported by the European Union under Contract BRE2-CT92-0201. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael R. Frater.

The author is with Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark (e-mail: aa@imm.dtu.dk; http://www.imm.dtu.dk/~aa).

Publisher Item Identifier S 1057-7149(02)01323-4.

or Markov random field methods to fuse image data with large spatial resolution differences [3] are not dealt with here.

The methods described here are well suited for integration in a geographical information system (GIS).

II. Two-Set Canonical Correlations

Canonical correlations analysis was first introduced in [4] to analyze linear relations between two sets of variables. The technique is described in most standard textbooks on multivariate statistics, e.g., [5] and [6]. Work on nonlinear canonical correlations analysis is dealt with in [7]–[11]. This type of analysis will not be pursued here.

Two-set canonical correlations analysis investigates the relationship between two groups of variables. It finds corresponding sets of linear combinations of the original two groups of variables. The first set of 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 set of 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.

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, (without loss of generality we assume that $\mathbf{E}\{\boldsymbol{X}\} = \mathbf{E}\{\boldsymbol{Y}\} = \mathbf{0}$, where $\mathbf{E}\{\cdot\}$ denotes expectation)

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

and we assume that the relevant dispersion matrices are nonsingular. Of course $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T$.

We are searching for linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$U = \sum_{i=1}^{p} a_i X_i = \boldsymbol{a}^T\boldsymbol{X}, \mathrm{V}\{U\} = \boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}$$
$$V = \sum_{i=1}^{q} b_i Y_i = \boldsymbol{b}^T\boldsymbol{Y}, \mathrm{V}\{V\} = \boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b}$$

(where $\mathrm{V}\{\cdot\}$ denotes variance) with maximum correlation

$$\rho = \mathrm{Corr}\{U,V\} = \frac{\mathrm{Cov}\{U,V\}}{\sqrt{\mathrm{V}\{U\}\ \mathrm{V}\{V\}}} = \frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}\,\boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b}}}.$$

Let $R = \mathrm{Cov}\{U,V\}$ denote the covariance between U and V. To maximize ρ we set $\partial\rho/\partial\boldsymbol{a} = \partial\rho/\partial\boldsymbol{b} = \mathbf{0}$ and get

$$\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}\boldsymbol{\Sigma}_{12}\boldsymbol{b} = R\boldsymbol{\Sigma}_{11}\boldsymbol{a}$$
$$\boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b}\boldsymbol{\Sigma}_{21}\boldsymbol{a} = R\boldsymbol{\Sigma}_{22}\boldsymbol{b}.$$

Without loss of generality, we choose $[\boldsymbol{a},\boldsymbol{b}]$ so that $\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a} = \boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b} = 1$ which leads to

$$\rho^2 = \frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}} = \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{b}}{\boldsymbol{b}^T\boldsymbol{\Sigma}_{22}\boldsymbol{b}}$$

i.e., we find the desired projections for $\boldsymbol{X}$ by considering the conjugate eigenvectors $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_p$ corresponding to the eigenvalues $\rho_1^2 \geq \cdots \geq \rho_p^2$ of $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ with respect to $\boldsymbol{\Sigma}_{11}$. Similarly, we may find the desired projections for $\boldsymbol{Y}$ by considering the conjugate eigenvectors $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_p$ of $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ with respect to $\boldsymbol{\Sigma}_{22}$ corresponding to the same eigenvalues ρ_i^2. If $p = q$ this will be all the eigenvalues and -vectors of $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. If $q > p$ the last eigenvalue will be 0 with multiplicity $q - p$. (As the solutions $\boldsymbol{a}$ and $\boldsymbol{b}$ are interrelated we only need to find one of them.)

A. Partial Least Squares (PLS)

Canonical corelations analysis (CCA) is closely related to the method of partial least squares, PLS, in which $R = \mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{X}, \boldsymbol{b}^T\boldsymbol{Y}\} = \boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{b}$ (often with $\boldsymbol{Y}$ as a scalar response variable) is maximized with another choice of constraints, namely $\boldsymbol{a}^T\boldsymbol{a} = \boldsymbol{b}^T\boldsymbol{b} = 1$ leading to

$$R^2 = \frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{a}} = \frac{\boldsymbol{b}^T\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{12}\boldsymbol{b}}{\boldsymbol{b}^T\boldsymbol{b}}$$

(see [12]). We see that in this case matrix inversion is not needed which is good if we have many variables and few observations. Only the first pair of canonical variates (or latent variables) corresponding to the largest eigenvalue are calculated and the response CV is regressed on the predictor CV

$$V = cU + e.$$

If more information is present in the residuals, their projections replace the original response variables (i.e., replace $\boldsymbol{Y}$ by $\boldsymbol{Y} - cU\boldsymbol{b}$), the predictor variables are projected into a subspace orthogonal to the solution found (i.e., replace $\boldsymbol{X}$ by $\boldsymbol{X} - U\boldsymbol{p}$) with $\boldsymbol{p} = \boldsymbol{\Sigma}_{11}\boldsymbol{a}/\boldsymbol{a}^T\boldsymbol{\Sigma}_{11}\boldsymbol{a}$), and we iterate; see also [13]–[16].

B. The MAD Transformation

The above CCA technique is used in [17]–[22] to find linear combinations that give maximal multivariate differences. The name chosen for the transformation, multivariate alteration detection (MAD), is due to the application to change detection in remote sensing (and the acronym). Although it is presented as a change detection technique in remote sensing, the technique applies to nonspatial multivariate differences also. The MAD transformation has been used in an attempt to differentiate between geogenic and anthropogenic influences on soils in a mining processing area, see [23]. References [17] and [19] also suggest the use of the maximum autocorrelation factor (MAF) transformation, [24], to postprocess the MAD variates. Reference [25] uses MAF to process the simple differences. Reference [26] uses a hybrid canonical correlation/principal components technique to enhance uncorrelated parts of Landsat TM equivalents of ATM data in a gold exploration study. Change detection techniques based on canonical variates are also described in [27] and [28].

III. Multiset Canonical Correlations

Multiset canonical correlations analysis (MCCA) is a technique for analyzing linear relations between more (than two) sets of variables. Earlier work in this field comprise [29]–[32]. Reference [33] gives an interesting example using satellite data and two types of geochemical data. Work on nonlinear MCCA is reported in [48].

We consider an $m = m_1 + m_2 + \cdots + m_n$ dimensional random variable $\boldsymbol{X}$ ideally following a Gaussian distribution split into n groups of dimensions m_1, m_2 to m_n $(m_1 \leq m_2 \leq \cdots \leq m_n)$, respectively, (without loss of generality we assume that $\mathbf{E}\{\boldsymbol{X}_i\} = \mathbf{0}$)

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{bmatrix} \in N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$
$$= N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \right)$$

and we assume that the relevant dispersion matrices are nonsingular. Of course $\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Sigma}_{ji}^T$.

An obvious extension from the two-set case is to search for linear combinations $\boldsymbol{U}^T = [U_1, U_2, \ldots, U_n]$ of $\boldsymbol{X}^T = [\boldsymbol{X}_1^T, \boldsymbol{X}_2^T, \ldots, \boldsymbol{X}_n^T]$

$$\begin{aligned} U_1 &= \boldsymbol{a}_1^T \boldsymbol{X}_1,\ V\{U_1\} = \boldsymbol{a}_1^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}_1 \\ U_2 &= \boldsymbol{a}_2^T \boldsymbol{X}_2,\ V\{U_2\} = \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{22} \boldsymbol{a}_2 \\ &\vdots \\ U_n &= \boldsymbol{a}_n^T \boldsymbol{X}_n,\ V\{U_n\} = \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{nn} \boldsymbol{a}_n \end{aligned}$$

with dispersion matrix

$$\boldsymbol{\Sigma}_U = \begin{bmatrix} \boldsymbol{a}_1^T \boldsymbol{\Sigma}_{11} \boldsymbol{a}_1 & \boldsymbol{a}_1^T \boldsymbol{\Sigma}_{12} \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1^T \boldsymbol{\Sigma}_{1n} \boldsymbol{a}_n \\ \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{21} \boldsymbol{a}_1 & \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{22} \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2^T \boldsymbol{\Sigma}_{2n} \boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{n1} \boldsymbol{a}_1 & \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{n2} \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n^T \boldsymbol{\Sigma}_{nn} \boldsymbol{a}_n \end{bmatrix}$$

or $\boldsymbol{\Sigma}_U = \{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j\} = \{\rho_{ij}\}$ for short. As in the two-set case there is one $U_i = U_{ik}$, $\boldsymbol{a}_i = \boldsymbol{a}_{ik}$ and $\boldsymbol{\Sigma}_U = \boldsymbol{\Sigma}_{Uk}$ for each $k = 1, \ldots, m_1$, $m_1 = \min(m_1, \ldots, m_n)$.

In the two-set case we obtain new variables with a high measure of similarity by maximizing the scalar $\rho = \mathrm{Corr}\{\boldsymbol{a}_1^T \boldsymbol{X}_1, \boldsymbol{a}_2^T \boldsymbol{X}_2\}$. Here, we must maximize all correlations/covariances between the new variables simultaneously. To do this, the following measures of $\boldsymbol{\Sigma}_U$ can be optimized:

1) maximize sum of elements ($V = \sum_{i=1}^n \sum_{j=1}^n \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j$);
2) maximize sum of squared elements ($W = \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j)^2$);
3) maximize largest eigenvalue (λ_1);
4) minimize smallest eigenvalue (λ_n);
5) minimize determinant (det $\boldsymbol{\Sigma}_U = \prod_{i=1}^n \lambda_i$).

Reference [32] lists all these possibilities and names them

1) SUMCOR;
2) SSQCOR;
3) MAXVAR;
4) MINVAR;
5) GENVAR.

These measures are not constrained, but several natural choices for constraints under which to carry out the optimizations come to mind:

1) the projection vectors are unit vectors within each set ($\boldsymbol{a}_i^T \boldsymbol{a}_i = 1$);
2) the sum of the projection vectors is a unit vector ($\sum_{i=1}^n \boldsymbol{a}_i^T \boldsymbol{a}_i = 1$);
3) the new variables have unit variance ($\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = 1$);
4) the sum of the variances of the new variables is unity ($\sum_{i=1}^n \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = \mathrm{tr}\ \boldsymbol{\Sigma}_U = 1$);

where constraint 3 (causing ρ_{ij} to be correlations rather than covariances) is the natural extension from the two-set case. If the n sets analyzed are different variables measured over time this type of analysis constitutes a multivariate extension to the technique of empirical orthogonal functions (EOF) [1].

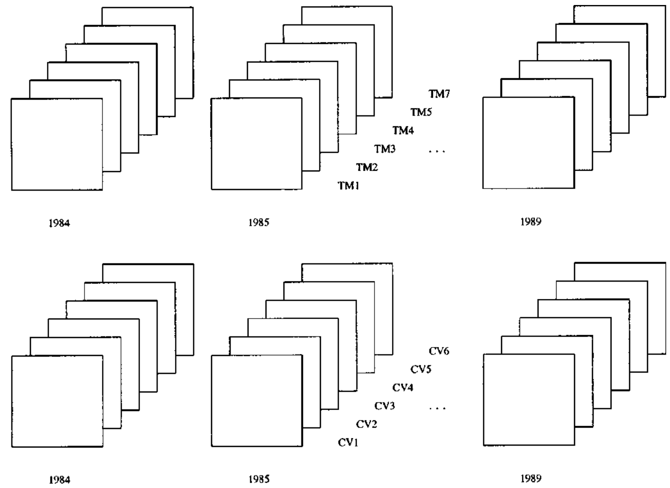

Fig. 1. Sketch of R-mode multiset canonical correlations analysis. Variables indicated in top row are transformed into CVs in bottom row.

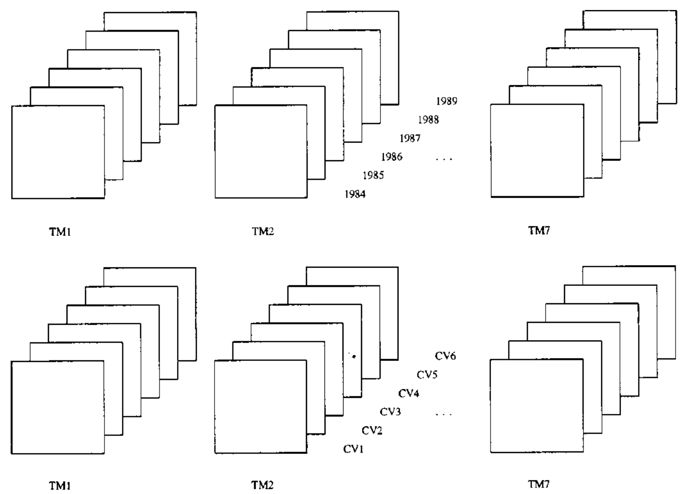

Fig. 2. Sketch of T-mode multiset canonical correlations analysis. Variables indicated in top row are transformed into CVs in bottom row.

TABLE I
CORRELATIONS BETWEEN R-MODE CANONICAL VARIATES 1 FOR ALL FIVE METHODS

SUMCOR					
1.0000	0.9114	0.8534	0.8768	0.8893	0.8852
0.9114	1.0000	0.9327	0.9248	0.8939	0.9037
0.8534	0.9327	1.0000	0.9202	0.8861	0.9048
0.8768	0.9248	0.9202	1.0000	0.8868	0.9127
0.8893	0.8939	0.8861	0.8868	1.0000	0.9544
0.8852	0.9037	0.9048	0.9127	0.9544	1.0000
SSQCOR					
1.0000	0.9114	0.8532	0.8765	0.8893	0.8851
0.9114	1.0000	0.9330	0.9250	0.8939	0.9038
0.8532	0.9330	1.0000	0.9204	0.8860	0.9049
0.8765	0.9250	0.9204	1.0000	0.8867	0.9126
0.8893	0.8939	0.8860	0.8867	1.0000	0.9545
0.8851	0.9038	0.9049	0.9126	0.9545	1.0000
MAXVAR					
1.0000	0.9114	0.8534	0.8767	0.8893	0.8852
0.9114	1.0000	0.9328	0.9249	0.8939	0.9037
0.8534	0.9328	1.0000	0.9202	0.8860	0.9048
0.8767	0.9249	0.9202	1.0000	0.8867	0.9127
0.8893	0.8939	0.8860	0.8867	1.0000	0.9544
0.8852	0.9037	0.9048	0.9127	0.9544	1.0000
MINVAR					
1.0000	0.8334	0.7259	0.7709	0.7651	0.7797
0.8334	1.0000	0.9195	0.9027	0.8472	0.8709
0.7259	0.9195	1.0000	0.8595	0.7692	0.8246
0.7709	0.9027	0.8595	1.0000	0.8667	0.9023
0.7651	0.8472	0.7692	0.8667	1.0000	0.9564
0.7797	0.8709	0.8246	0.9023	0.9564	1.0000
GENVAR					
1.0000	0.9067	0.8390	0.8645	0.8896	0.8792
0.9067	1.0000	0.9412	0.9276	0.8903	0.9022
0.8390	0.9412	1.0000	0.9241	0.8772	0.9031
0.8645	0.9276	0.9241	1.0000	0.8795	0.9076
0.8896	0.8903	0.8772	0.8795	1.0000	0.9577
0.8792	0.9022	0.9031	0.9076	0.9577	1.0000

TABLE II
CORRELATIONS BETWEEN T-MODE CANONICAL VARIATES 1 FOR ALL FIVE METHODS

SUMCOR					
1.0000	0.9420	0.9548	0.6414	0.8919	0.9275
0.9420	1.0000	0.9531	0.7571	0.9059	0.9021
0.9548	0.9531	1.0000	0.6989	0.9038	0.9219
0.6414	0.7571	0.6989	1.0000	0.7366	0.6392
0.8919	0.9059	0.9038	0.7366	1.0000	0.9673
0.9275	0.9021	0.9219	0.6392	0.9673	1.0000
SSQCOR					
1.0000	0.9442	0.9574	0.6385	0.8935	0.9301
0.9442	1.0000	0.9547	0.7547	0.9064	0.9038
0.9574	0.9547	1.0000	0.6942	0.9049	0.9241
0.6385	0.7547	0.6942	1.0000	0.7326	0.6337
0.8935	0.9064	0.9049	0.7326	1.0000	0.9678
0.9301	0.9038	0.9241	0.6337	0.9678	1.0000
MAXVAR					
1.0000	0.9437	0.9566	0.6396	0.8931	0.9293
0.9437	1.0000	0.9543	0.7549	0.9064	0.9034
0.9566	0.9543	1.0000	0.6958	0.9047	0.9235
0.6396	0.7549	0.6958	1.0000	0.7333	0.6358
0.8931	0.9064	0.9047	0.7333	1.0000	0.9677
0.9293	0.9034	0.9235	0.6358	0.9677	1.0000
MINVAR					
1.0000	0.9451	0.8939	0.4978	0.8767	0.9316
0.9451	1.0000	0.8535	0.6725	0.8988	0.9055
0.8939	0.8535	1.0000	0.3950	0.7683	0.8438
0.4978	0.6725	0.3950	1.0000	0.6609	0.5046
0.8767	0.8988	0.7683	0.6609	1.0000	0.9666
0.9316	0.9055	0.8438	0.5046	0.9666	1.0000
GENVAR					
1.0000	0.9488	0.9666	0.5350	0.8903	0.9369
0.9488	1.0000	0.9566	0.6985	0.9036	0.9058
0.9666	0.9566	1.0000	0.5687	0.8953	0.9302
0.5350	0.6985	0.5687	1.0000	0.6778	0.5261
0.8903	0.9036	0.8953	0.6778	1.0000	0.9676
0.9369	0.9058	0.9302	0.5261	0.9676	1.0000

In the two-set case all of these methods with constraints 3 and 4 reduce to the standard Hotelling case described in Section II (except for a scaling factor for constraint 4).

Reference [31] examines the SUMCOR method and [32] examines all the above methods using constraint 3. Reference [17] examines all the above methods using all constraints. As an illustration, we consider the SUMCOR method with constraints 3 and 4 shown in the following.

A. Maximize Sum of Covariances

To maximize the sum of covariances under constraints we use a Lagrange multiplier technique.

1) Constraint 3: $\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 = 0$*:* Introduce

$$
\begin{aligned}
F =& V - \sum_{i=1}^{n} \lambda_i \left(\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 \right) \\
=& \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j - \sum_{i=1}^{n} \lambda_i \left(\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 \right)
\end{aligned}
$$

and maximize F without constraints. By setting $\partial F / \partial \boldsymbol{a}_i = \boldsymbol{0}$ we get

$$
\sum_{j=1}^{n} \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j = \lambda_i \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i, \; i = 1, \ldots, n
$$

or

$$
\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \boldsymbol{\Sigma}_{11} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_2 \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \lambda_n \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}. \tag{1}
$$

Setting $\partial F / \partial \lambda_i = 0$ merely reproduces the constraints. Because the λ_is are not equal this system of equations is more general than a generalized eigensystem. Invariance of the solution to linear transformations within sets is easily shown (after the transformation the λ_is will be the same, the $\boldsymbol{a}_i$s will not).

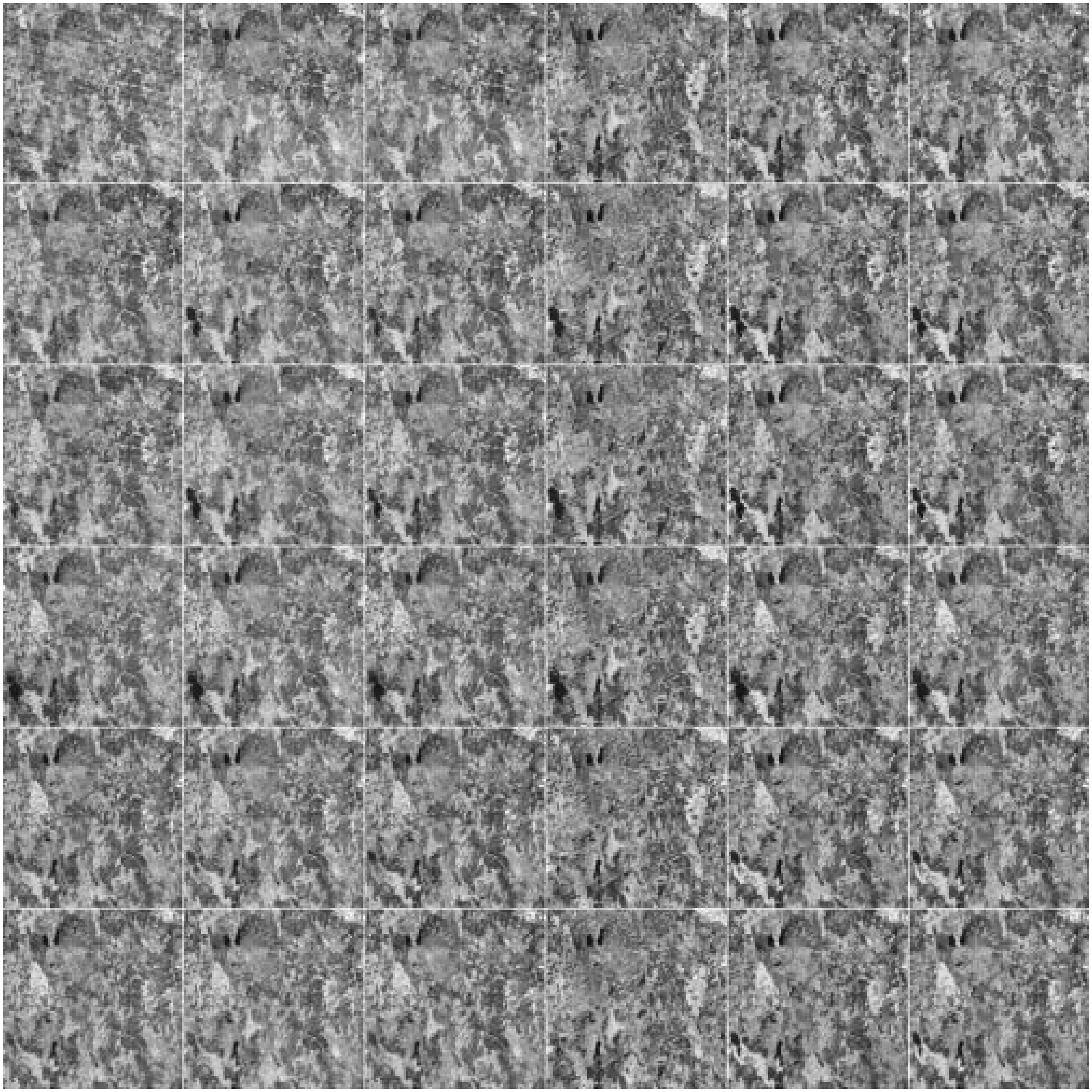

Fig. 3. Landsat TM data. Rows are years 1984 to 1989 and columns are TM bands 1 to 5 and 7.

2) Constraint 4: $\sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 = 0$*:* Introduce

$$\begin{aligned} F =& V - \lambda \left(\sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 \right) \\ =& \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j - \lambda \left(\sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i - 1 \right) \end{aligned}$$

and maximize F without constraints. By setting $\partial F / \partial \boldsymbol{a}_i = \boldsymbol{0}$ we get

$$\sum_{j=1}^{n} \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j = \lambda \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i, \; i = 1, \ldots, n$$

or

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1n} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n1} & \boldsymbol{\Sigma}_{n2} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Sigma}_{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}.$$

This is a (real, symmetric) generalized eigensystem, i.e., we find the desired projections for $\boldsymbol{X}_i$ by computing the conjugate eigenvectors $\boldsymbol{a}_{i1}, \boldsymbol{a}_{i2}, \ldots, \boldsymbol{a}_{im_1}$ corresponding to the first $m_1 =$

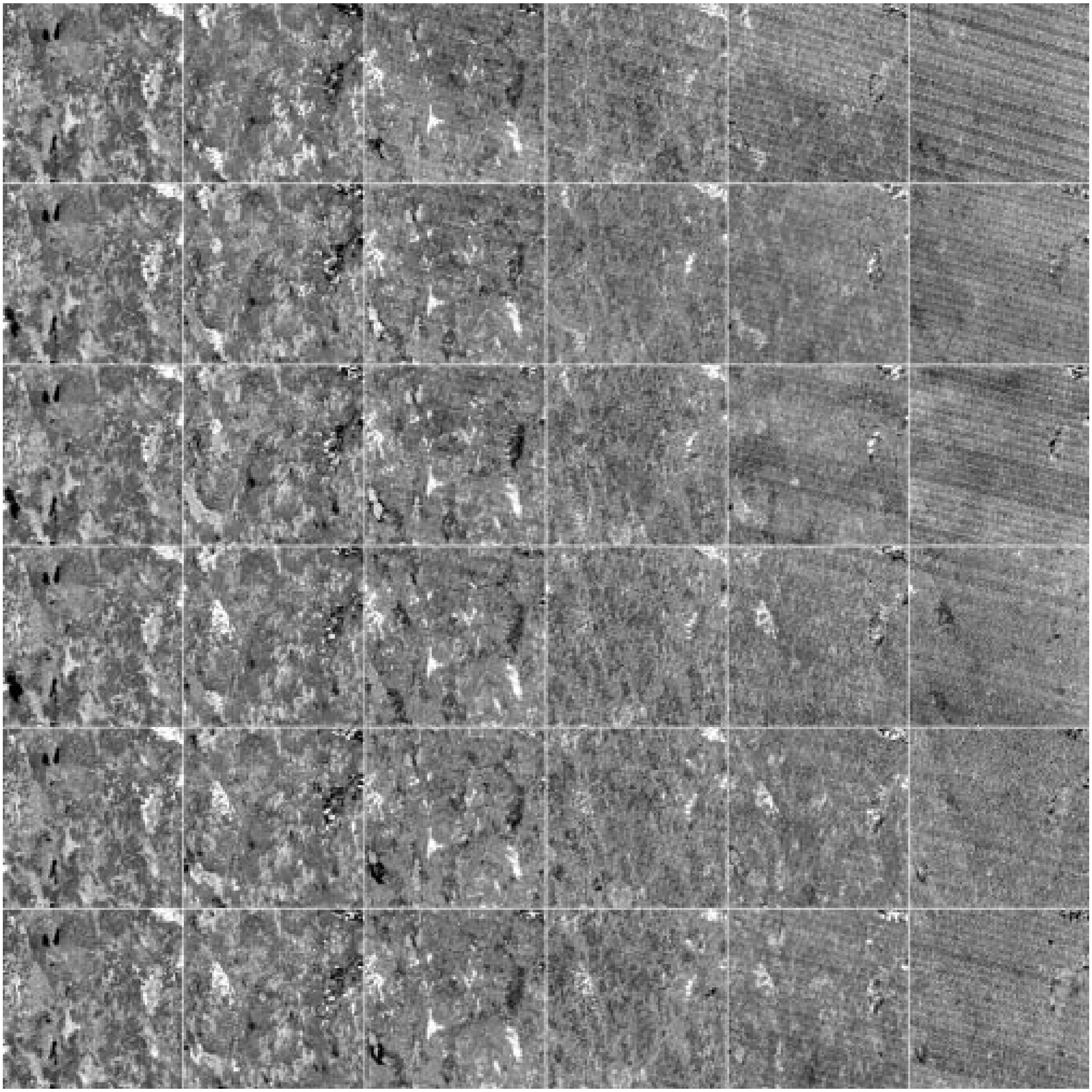

Fig. 4. R-mode CVs. Rows are years 1984 to 1989 and columns are CV1–CV6.

$\min(m_1, m_2, \ldots, m_n)$ eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{m_1}$ of the above eigensystem.

B. Multiset Canonical Variates

We are now able to define the *multiset canonical variates*

$$\begin{aligned} U_{1k} &= \boldsymbol{a}_{1k}^T \boldsymbol{X}_1, \ k = 1, 2, \ldots, m_1 \\ U_{2k} &= \boldsymbol{a}_{2k}^T \boldsymbol{X}_2, \ k = 1, 2, \ldots, m_1 \\ &\vdots \\ U_{nk} &= \boldsymbol{a}_{nk}^T \boldsymbol{X}_n, \ k = 1, 2, \ldots, m_1 \end{aligned}$$

where the $\boldsymbol{a}_{ik}$s come from either of the above solutions. With an obvious choice of notation we get

$$\begin{aligned} \boldsymbol{U}_1 &= \boldsymbol{A}_1^T \boldsymbol{X}_1 \\ \boldsymbol{U}_2 &= \boldsymbol{A}_2^T \boldsymbol{X}_2 \\ &\vdots \\ \boldsymbol{U}_n &= \boldsymbol{A}_n^T \boldsymbol{X}_n \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{A}_1 &= [\boldsymbol{a}_{11}, \boldsymbol{a}_{12}, \ldots, \boldsymbol{a}_{1m_1}] \text{ is } m_1 \times m_1 \\ \boldsymbol{A}_2 &= [\boldsymbol{a}_{21}, \boldsymbol{a}_{22}, \ldots, \boldsymbol{a}_{2m_1}] \text{ is } m_2 \times m_1 \\ &\vdots \\ \boldsymbol{A}_n &= [\boldsymbol{a}_{n1}, \boldsymbol{a}_{n2}, \ldots, \boldsymbol{a}_{nm_1}] \text{ is } m_n \times m_1. \end{aligned}$$

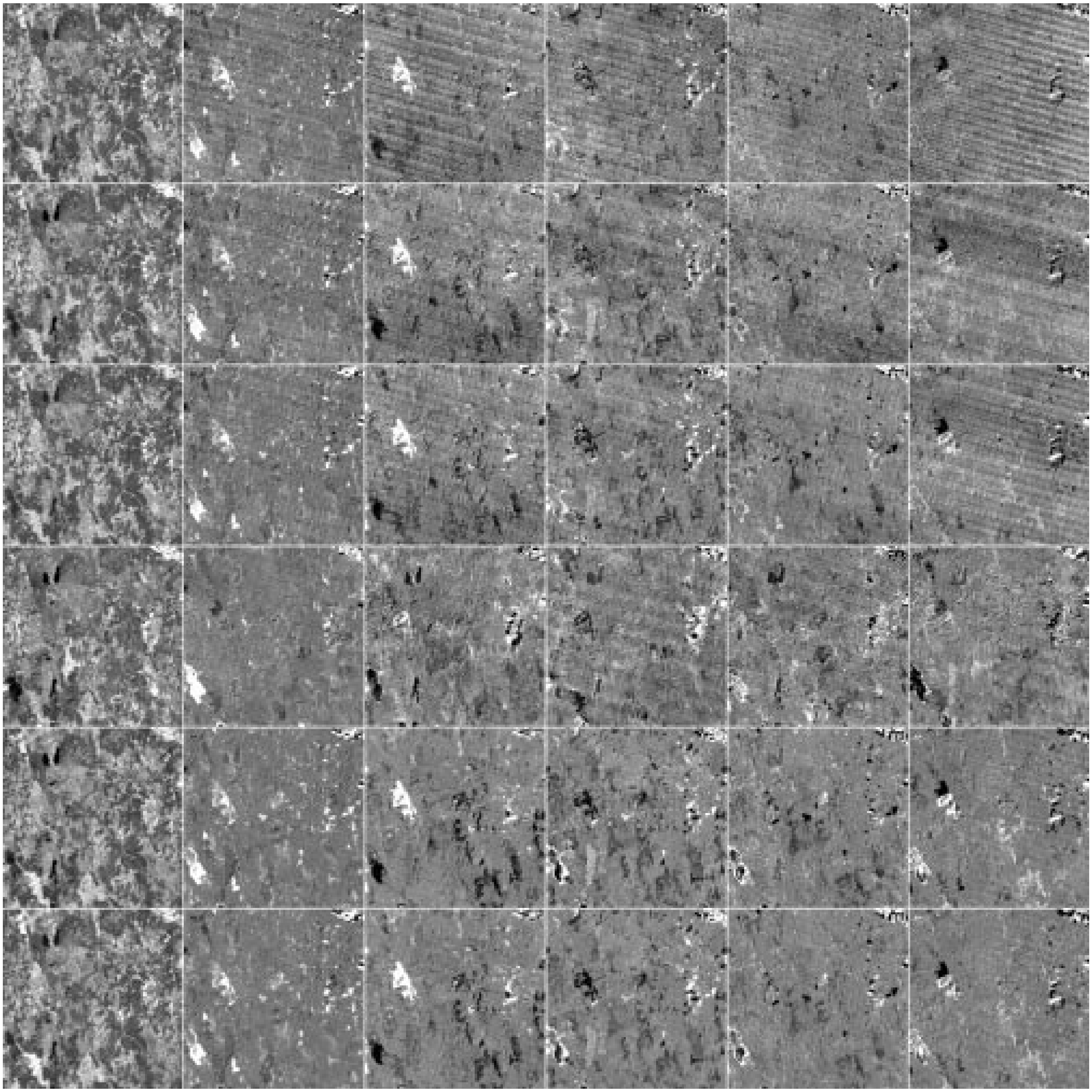

Fig. 5. T-mode CVs. Rows are TM bands 1 to 5 and 7 and columns are CV1–CV6.

C. Multiset Partial Least Squares

We can base a true multiset or multiblock partial least squares (PLS) method on MCCA with modified optimization criteria 1 and 2 mentioned above if we use constraint 1, $\boldsymbol{a}_i^T \boldsymbol{a}_i = 1$, with $\boldsymbol{\Sigma}_{ii}$ replaced by the null matrix since in this case we do not want to include the diagonal terms of $\boldsymbol{\Sigma}_U$. To see this, consider for example the maximization of the sum of all nondiagonal elements in $\boldsymbol{\Sigma}_U$

$$V_1 = \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j.$$

To maximize V_1 under constraint 1, again use Lagrange multipliers and maximize

$$F_1 = V_1 - \sum_{i=1}^{n} \lambda_i \left(\boldsymbol{a}_i^T \boldsymbol{a}_i - 1 \right)$$

without constraints. By setting $\partial F_1 / \partial \boldsymbol{a}_i = \mathbf{0}$ we get

$$\sum_{j=1, j \neq i}^{n} \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j = \lambda_i \boldsymbol{a}_i, \; i = 1, \ldots, n$$

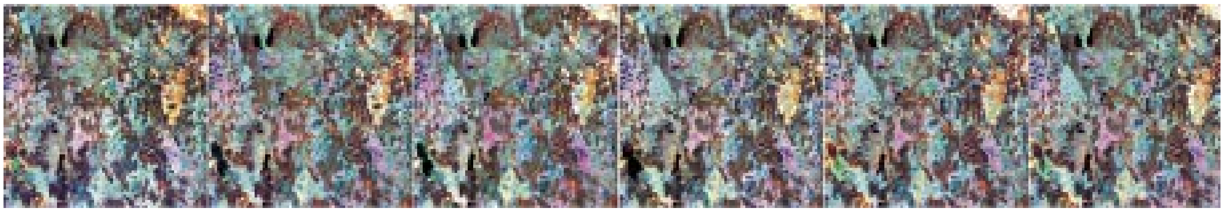

Fig. 6. Landsat TM bands 4, 5, and 3 as red, green, and blue (1984–1989).

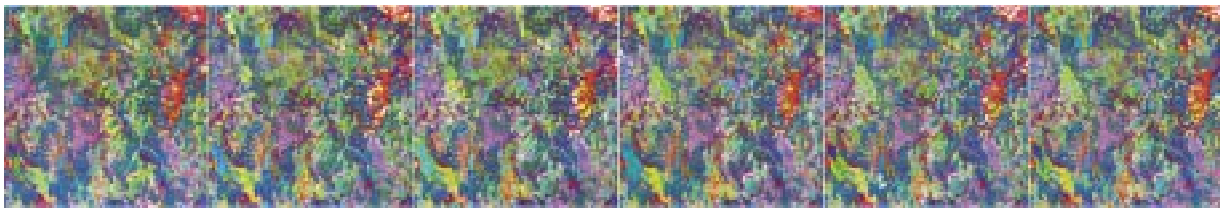

Fig. 7. R-mode CVs 1, 2, and 3 as red, green, and blue (1984–1989).

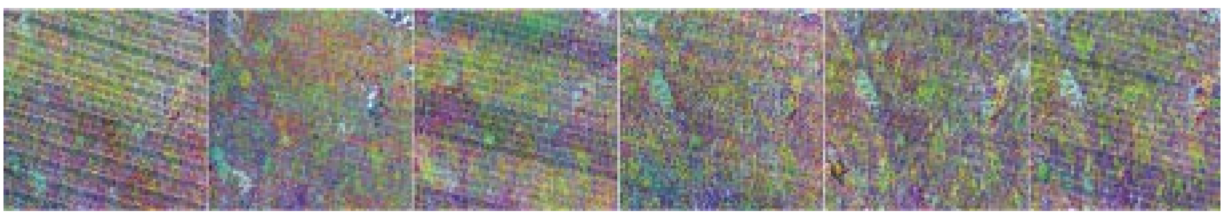

Fig. 8. R-mode CVs 6, 5, and 4 as red, green, and blue (1984–1989).

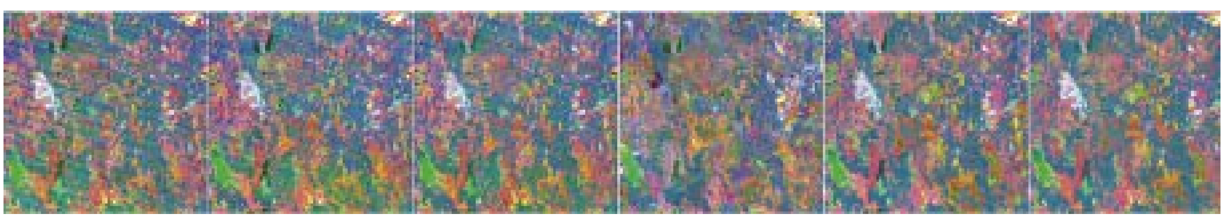

Fig. 9. T-mode CVs 1, 2, and 3 as red, green, and blue (TM bands 1–5 and 7).

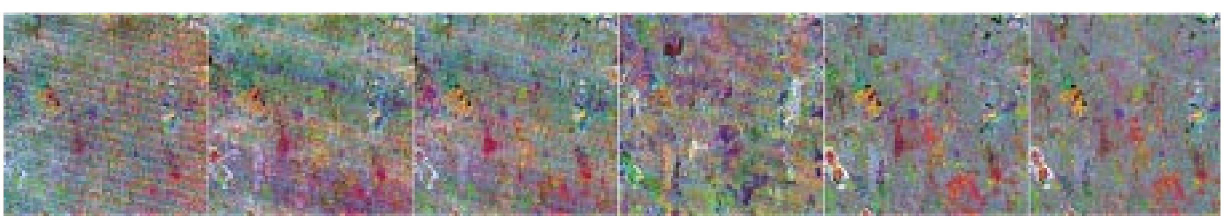

Fig. 10. T-mode CVs 6, 5, and 4 as red, green, and blue (TM bands 1–5 and 7).

or

$$\begin{bmatrix} \mathbf{0} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1n} \\ \mathbf{\Sigma}_{21} & \mathbf{0} & \cdots & \mathbf{\Sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{n1} & \mathbf{\Sigma}_{n2} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \boldsymbol{a}_1 \\ \lambda_2 \boldsymbol{a}_2 \\ \vdots \\ \lambda_n \boldsymbol{a}_n \end{bmatrix}.$$

Now, calculate all first sets of latent variables, i.e., the first canonical variates U_i, $i = 1, \ldots, n$. Without loss of generality we place the response variables in the first set and perform a multiple regression analysis

$$U_1 = c_2 U_2 + \cdots + c_n U_n + e.$$

If more information is present in the residuals their projections replace the original response variables (i.e., replace $\boldsymbol{X}_1$ by $\boldsymbol{X}_1 - \boldsymbol{a}_1 \sum_{i=2}^{n} c_i U_i$), the predictor variables are projected into a subspace orthogonal to the solution found (i.e., for $i = 2, \ldots, n$ replace $\boldsymbol{X}_i$ by $\boldsymbol{X}_i - U_i \boldsymbol{p}_i$ with $\boldsymbol{p}_i = \mathbf{\Sigma}_{ii} \boldsymbol{a}_i / \boldsymbol{a}_i^T \mathbf{\Sigma}_{ii} \boldsymbol{a}_i$) and we iterate. For ordinary two-set PLS, see [13]–[16].

IV. Computer Implementations

Two-set canonical correlations analysis is implemented in a computer program, `maf`, which is a general orthogonalization program that also finds principal components, (rotated) principal factors, maximum autocorrelation factors (MAF), [24], scaled MAFs [34], minimum noise fractions (MNF), [35], [36], multivariate alteration detection (MAD) variates, [17]–[19] and canonical discriminant functions, etc. Also the multiset canonical correlations analysis methods of maximizing the sum of covariances under constraints 2 ($\sum \boldsymbol{a}_i^T \boldsymbol{a}_i = 1$) and 4 ($\sum \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = 1$) are implemented in `maf`.

All dispersion (variance–covariance) matrices are found by the method of provisional means, [37]. The eigenvalue problems associated with the analysis are solved by means of LAPACK routines, [38]. Good general descriptions of the methods used are given in, e.g., [39]–[41].

The remaining optimization problems concerning MCCA (in fact, all of them, including the eigenvalue problems) are solved by means of the general algebraic modeling system (GAMS), [42], NLP solver CONOPT, [43]. A computer program `musecc` that calls GAMS to perform the analysis is implemented. As a parallel to the solution of the eigensystem involved in the two-set case the orthogonality criteria chosen in `musecc` are similar to the normalization criteria, for example for constraint 3 $\boldsymbol{a}_{ik}^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_{jl} = \rho_{ij}\delta_{kl}$, where $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ if not is Kronecker's delta.

V. Case: Landsat TM Data in Forestry

The utility of multiset canonical correlations analysis to multivariate and truly multi-temporal data is demonstrated in a case study using Landsat-5 Thematic Mapper (TM) data covering a small forested area approximately 20 kilometers north of Umeå in northern Sweden. The data consist of six by six spectral bands with 512×512 20-m pixels from the summers 1984–1989 rectified to the Swedish national grid. The acquisition dates are 1 August 1984, 26 June 1985, 6 June 1986, 12 August 1987, 27 June 1988, and 21 June 1989. These data are also analyzed in [44]–[46].

As an illustration, all results reported here (except the correlations in Tables I and II) relate to the Sumcor method with constraint and orthogonality criterion 3, i.e., the CVs have unit variance. In R-mode analysis, we consider Landsat TM bands 1 to 5 and 7 for 1984 as one set of variables and similarly for 1985, etc. In T-mode analysis we consider TM bands 1 for all years 1984–1989 as one set of variables, TM bands 2 for all years 1984–1989 as another set of variables, etc. [1]. For a sketch of R- and T-mode analysis setup; see Figs. 1 and 2. In both figures the six sets of variables indicated on the top are transformed into six sets of new variables on the bottom. For example, in T-mode analysis the variables 1984 TM1, 1985 TM1, …, 1989 TM1 are transformed into TM1 CVs and similarly for TM2, etc.

Fig. 3 shows the original TM data. Column one is TM1, column two is TM2, etc. Row one is 1984, row two is 1985, etc. Fig. 4 shows the R-mode CVs. Column one is CV1, column two is CV2, etc. Row one is 1984, row two is 1985, etc. Fig. 5 shows the T-mode CVs. Column one is CV1, column two is CV2, etc. Row one is TM1, row two is TM2, etc.

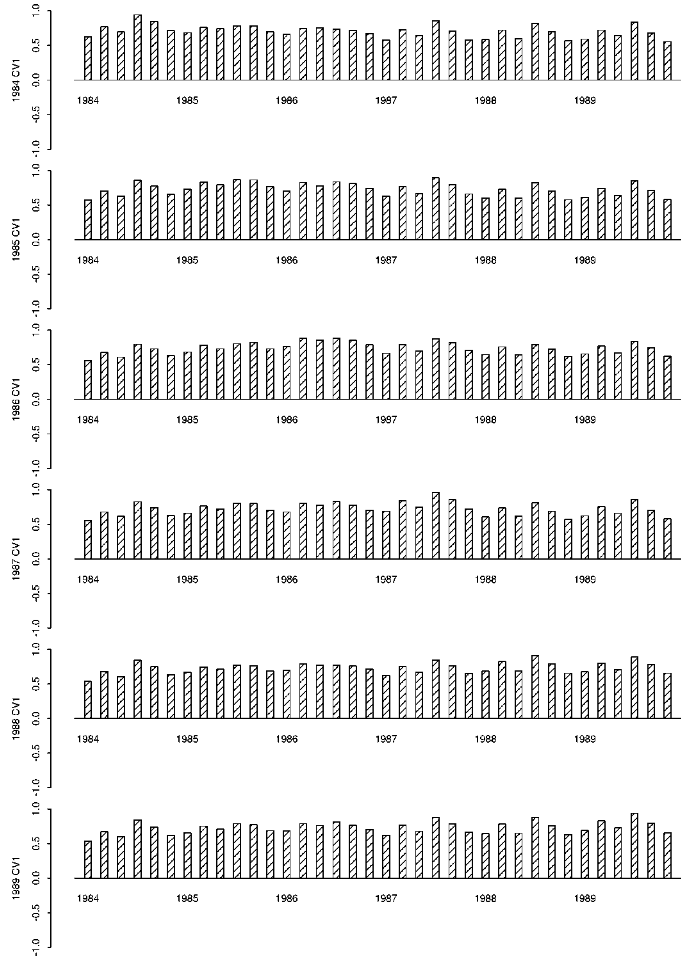

Fig. 11. Correlations between R-mode CVs 1 and original data.

Fig. 6 shows Landsat TM bands 4, 5, and 3 as red, green, and blue, respectively. Fig. 7 shows R-mode canonical variates 1, 2, and 3 as red, green, and blue, respectively. We see that we have indeed obtained a high degree of similarity over years. Fig. 8 shows R-mode canonical variates 6, 5, and 4 as red, green, and blue, respectively. This is the RGB combination that shows minimum similarity over years. We see that noise (striping and dropouts) is depicted well as is to be expected: if data from one year is noisy and data from another year is not (or if the noise patterns are different) then certainly the largest difference could be that noise (or that difference). This observation inspires an iterative use of the procedure: first identify noise, restore data or exclude areas with noise from further analysis and carry out the analysis once more. This iterative use is not illustrated here.

Fig. 9 shows T-mode canonical variates 1, 2, and 3 as red, green, and blue, respectively. Again, we see that we have obtained a high degree of similarity, this time over TM bands. Fig. 10 shows T-mode canonical variates 6, 5, and 4 as red, green, and blue, respectively. We see that striping is strongly present in TM bands 1 and 2.

The transformation matrices containing the weights applied to the original variables to obtain the CVs are not shown as these weights are difficult to interpret because of inter-correlation between the original variables. Instead we show correlations between the original variables and the CVs. (It is often seen that

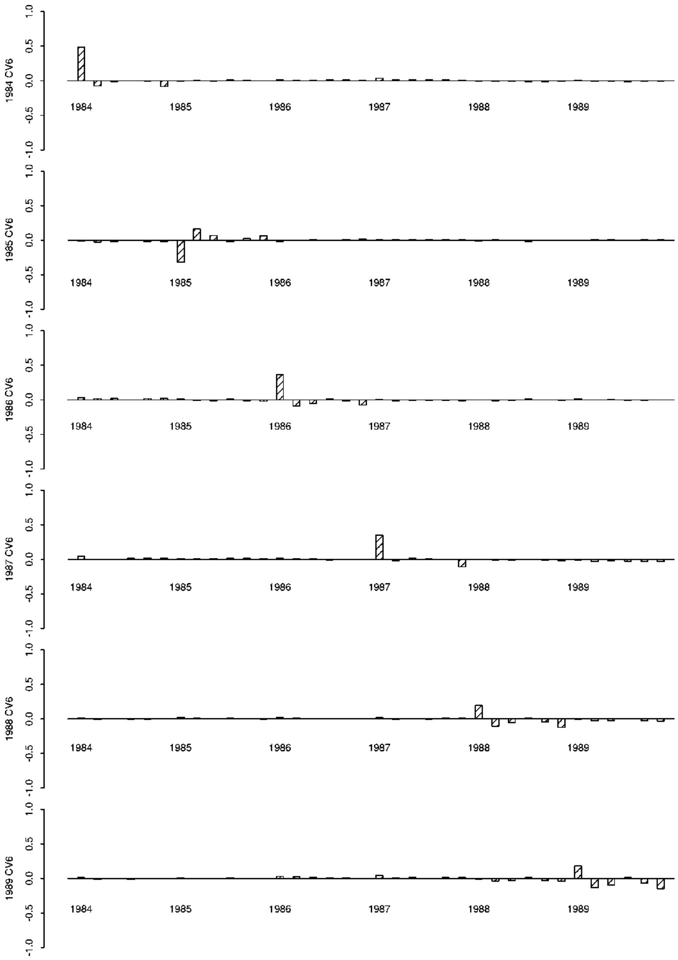

Fig. 12. Correlations between R-mode CVs 6 and original data.

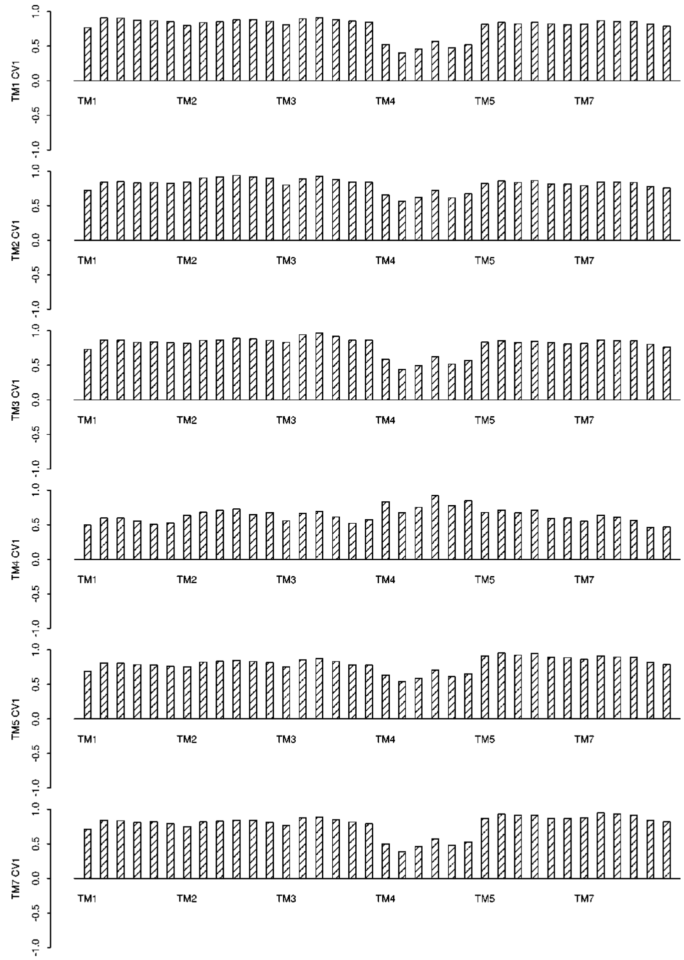

Fig. 13. Correlations between T-mode CVs 1 and original data.

an original variable that has a, say, negative weight in the calculation of some transformed variable has a positive correlation with that transformed variable.) Correlations between R-mode CVs 6 and the original data given in Fig. 12 show that dissimilarities (differences between years) are associated with TM bands 1 especially from 1984 to 1987. This is probably because of differences in atmospheric conditions. Therefore analysis of atmospherically corrected data would be interesting. Correlations between T-mode CVs 6 and the original variables given in Fig. 14, for TM bands 1, 2, 3, 5, and 7 reveal a pattern of positive correlation with 1984, negative correlation with 1985 and again positive correlation with 1986 (but not as high as with 1984) combined with (nearly) no correlation with 1987–1989. T-mode CV6 for TM4 is positively correlated with TM4 in 1984–1986, uncorrelated with TM4 in 1987 and negatively correlated with TM4 in 1988 and 1989. This could indicate that vegetation related changes occurred from 1986 to 1988. This finding is confirmed by an observation in [45]: "Several stands with Scots Pine (*Pinus Sylvestris*) had been damaged by the snow-break in the winter 1987/1988." Correlations between T-mode CVs 1 and TM4 given in Fig. 13 are (except for TM4 CV1) lower than correlations between T-mode CVs 1 and the other bands. Again, this indicates changes that are related with TM4, possibly vegetation changes. For completeness Fig. 11 gives correlations between R-mode CVs 1 and the original data.

Table I shows correlations between R-mode canonical variates 1 ($\boldsymbol{\Sigma}_U$) for all five methods investigated. The same correlations for T-mode analysis are shown in Table II. Again, we see a special behavior for TM4 indicating vegetation changes.

Table III shows the values of objective function, i.e., the quantity which is maximized, namely the sum of the elements in $\boldsymbol{\Sigma}_U$, $V = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ij} \boldsymbol{a}_j$ under the constraint that $\boldsymbol{a}_i^T \boldsymbol{\Sigma}_{ii} \boldsymbol{a}_i = 1$. We see that although R-mode analysis obtains the highest objective function value (i.e., level for all correlations simultaneously), T-mode analysis maintains a high level for this measure for higher order CVs than does R-mode. Also, for R-mode the highest difference between objective function values occurs between CVs 3 and 4, whereas for T-mode it occurs between CVs 5 and 6. Also, there is a big reduction in the objective function values between T-mode CVs 1 and 2. This difference is clearly visible in the imagery also (Fig. 5).

In the comparisons performed in Tables I and II, SUMCOR, SSQCOR, and MAXVAR seem to perform similarly. MINVAR and GENVAR seem to perform differently and not in the same fashion. [47] observes a similar different behavior for MINVAR. This is understandable when contemplating the design criteria behind the individual methods. SUMCOR and SSQCOR both focus on all correlations between CVs, i.e., all elements in $\boldsymbol{\Sigma}_U$. MAXVAR maximizes the largest eigenvalue, again a focus on all elements in $\boldsymbol{\Sigma}_U$. MINVAR relies heavily on the smallest eigenvalue, whereas

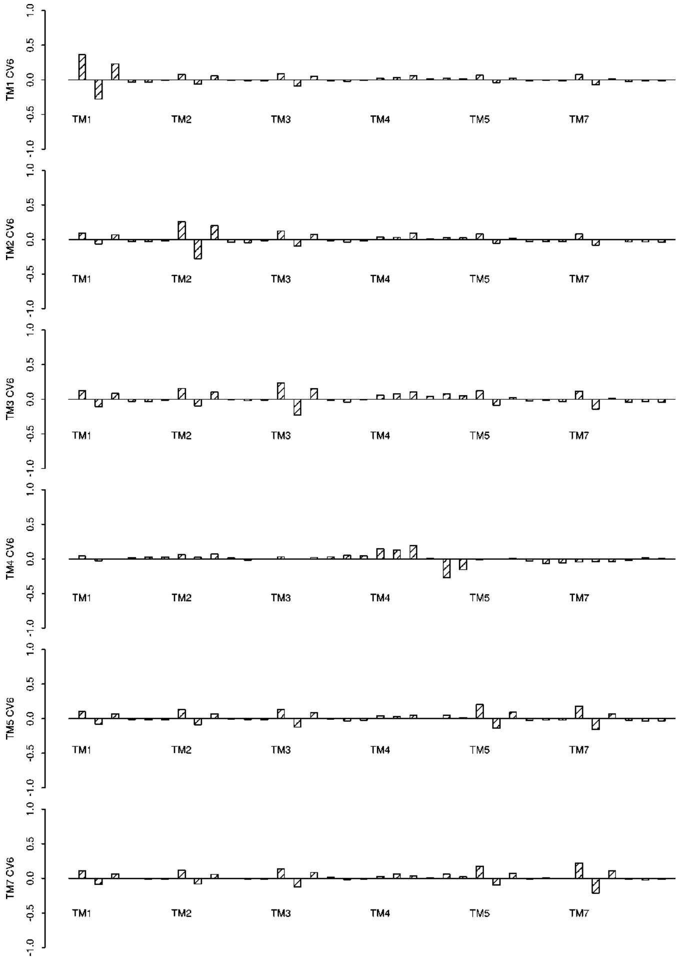

Fig. 14. Correlations between T-mode CVs 6 and original data.

TABLE III
VALUE OF OBJECTIVE FUNCTION: SUM OF ELEMENTS IN $\boldsymbol{\Sigma}_U$

Mode	CV1	CV2	CV3	CV4	CV5	CV6
R	33.0725	28.1649	24.9316	14.4707	9.3141	6.0432
T	31.4869	23.5503	21.0998	19.2327	18.5041	9.2767

GENVAR minimizes the determinant of $\boldsymbol{\Sigma}_U$ and therefore relies on several small eigenvalues. Due to lack of ground truth data, it has not been possible to determine empirically which of the five methods (if any) perform best in this context.

Tables IV and V show comparisons of the actual values of the optimization criteria for the five methods discussed for R- and T-mode canonical variates 1. The optimization criteria are not contradicted, e.g., for MINVAR $\lambda_{\min}$ is smaller than for the other methods. Also in this comparison, SUMCOR, SSQCOR, and MAXVAR seem to perform similarly and MINVAR and GENVAR seem to perform differently and not in the same fashion.

Fig. 15 shows λ_i for both R- and T-mode CV1–CV6 [see (1)]. Some differences within individual CVs are seen especially for T-mode.

The difference between the results of the R- and T-mode analyses is believed to be due to the expected correlation in noise over bands in the same year and the expected lack of correlation in noise over years in the same bands.

TABLE IV
VALUE OF OBJECTIVE FUNCTION FOR ALL FIVE METHODS, R-MODE CV1

Method	$\sum\sum\boldsymbol{\Sigma}_{Uij}$	$\sum\sum(\boldsymbol{\Sigma}_{Uij})^2$	λ_{max}	λ_{min}	det $\boldsymbol{\Sigma}_U$
SUMCOR	33.0725	30.4481	5.5125	0.0415	$2.3488\ 10^{-5}$
SSQCOR	33.0724	30.4482	5.5125	0.0414	$2.3347\ 10^{-5}$
MAXVAR	33.0725	30.4482	5.5125	0.0415	$2.3448\ 10^{-5}$
MINVAR	31.1882	27.2723	5.2038	0.0336	$1.0191\ 10^{-4}$
GENVAR	32.9787	30.2877	5.4971	0.0371	$2.0069\ 10^{-5}$

TABLE V
VALUE OF OBJECTIVE FUNCTION FOR ALL FIVE METHODS, T-MODE CV1

Method	$\sum\sum\boldsymbol{\Sigma}_{Uij}$	$\sum\sum(\boldsymbol{\Sigma}_{Uij})^2$	λ_{max}	λ_{min}	det $\boldsymbol{\Sigma}_U$
SUMCOR	31.4869	28.0484	5.2730	0.0177	$1.2683\ 10^{-5}$
SSQCOR	31.4812	28.0566	5.2732	0.0167	$1.0872\ 10^{-5}$
MAXVAR	31.4842	28.0562	5.2734	0.0171	$1.1480\ 10^{-5}$
MINVAR	29.2292	24.9382	4.9373	0.0073	$1.6008\ 10^{-5}$
GENVAR	30.6156	26.9877	5.1558	0.0078	$3.4272\ 10^{-6}$

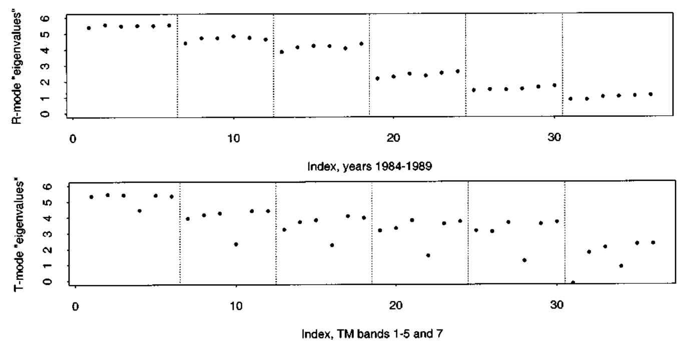

Fig. 15. λ_i for both R-mode (top) and T-mode (bottom) CV1 (indexes 1–6), CV2 (indexes 7–12), . . ., CV6 (indexes 31–26) [see (1)].

If we assign variables to $\boldsymbol{X}_i$ (Section III) in an appropriate fashion the methods described can be used for simultaneous optimization of more objectives such as interset correlation and spatial correlation.

VI. CONCLUSIONS

Two- and multiset canonical correlations analysis for data fusion, multisource, multiset, or multitemporal exploratory data analysis is described and applied to six spectral bands from Landsat TM summer data from 1984 to 1989. The resulting canonical variates are invariant to linear and affine transformations of the original data within sets. This means, for example, that the R-mode CVs are insensitive to changes over time in offset and gain in a measuring device. The CVs show the desired characteristic, namely that they exhibit ever decreasing similarity (as measured by correlation) with increasing order of the CVs. There is a big (visual) difference between the results of the R- and T-mode analyses. This difference is ascribed to the noise structure of the data.

Choosing other optimization criteria and constraints than the ones usually chosen for canonical correlation analysis, the methods described also form a basis for true multiset PLS.

If the data analyzed are variables measured over time this type of analysis constitutes a multivariate extension to the technique of EOF.

Although applied to Landsat TM data here, the methods are suitable to the analysis of any data that are naturally divided into several multivariate groups.

The methods described are well suited for integration in a GIS.

ACKNOWLEDGMENT

The author would like to thank Prof. K. Conradsen, IMM, for his cooperation in the development of the MAD concept and for establishing a professional contact to Dr. A. Drud whom the author thanks for the immediate interest he took in the noneigenvalue type optimization problems. Dr. Drud wrote the GAMS NLP solver CONOPT applied and he also wrote the generic GAMS code that is the heart of the computer program `musecc`. The author would like to thank Dr. R. Larsen, IMM, for good discussions on many aspects of multivariate statistics and for programming `maf` in close cooperation with him. The access to six years of geometrically corrected Landsat TM data from northern Sweden given by the Swedish Space Corporation is acknowledged.

REFERENCES

[1] R. W. Preisendorfer, "Principal component analysis in meteorology and oceanography," in *Developments in Atmospheric Science*, C. D. Mobley, Ed. Amsterdam, The Netherlands: Elsevier, 1988, vol. 17.

[2] K. Conradsen, B. K. Ersbøll, and A. A. Nielsen, "Integration of multi-source data in mineral exploration," in *Proc. 8th Thematic Conf. Geologic Remote Sensing*, Denver, CO, Apr. 1991, pp. 1053–1066.

[3] B. K. Ersbøll, K. Conradsen, R. Larsen, A. A. Nielsen, and T. H. Nielsen, "Fusion of SPOT HRV XS and orthophoto data using a Markov random field model," in *Proc. Fusion of Earth Data: Merging of Point Measurements, Raster Maps, and Remotely Sensed Images*, T. Ranchin and L. Wald, Eds., Sophia Antipolis, France, Jan. 1998.

[4] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321–377, 1936.

[5] W. W. Cooley and P. R. Lohnes, *Multivariate Data Analysis*. New York: Wiley, 1971.

[6] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.

[7] E. van der Burg and J. de Leeuw, "Non-linear canonical correlation," *Brit. J. Math. Statist. Psychol.*, vol. 36, pp. 54–80, 1983.

[8] L. Breiman and J. H. Friedman, "Estimating optimal transformations for multiple regression and correlation," *J. Amer. Statist. Assoc.*, vol. 80, no. 391, pp. 580–619, 1985.

[9] A. Buja, "Remarks on functional canonical variates, alternating least squares methods and ACE," *Ann. Statist.*, vol. 18, pp. 1032–1069, 1990.

[10] S. G. Shi and W. Taam, "Non-linear canonical correlation analysis with a simulated annealing solution," *J. Appl. Statist.*, vol. 19, no. 2, pp. 155–165, 1992.

[11] K. Windfeld, "Application of computer intensive data analysis: Methods to the analysis of digital images and spatial data," Ph.D. dissertation, Inst. Math. Statist. Oper. Res., Tech. Univ. Denmark, Lyngby, Denmark, 1992.

[12] M. Borga, "Learning multidimensional signal processing," Ph.D. dissertation, Dept. Elect. Eng., Linköping Univ., Linköping, Sweden, 1998.

[13] S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses," *SIAM J. Sci. Statist. Comput.*, vol. 5, no. 3, pp. 735–743, 1984.

[14] P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Anal. Chimica Acta*, vol. 185, pp. 1–17, 1986.

[15] A. Höskuldsson, "PLS regression methods," *J. Chemom.*, vol. 2, pp. 211–228, 1986.

[16] I. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.

[17] A. A. Nielsen, "Analysis of regularly and irregularly sampled spatial, multivariate and multi-temporal data," Ph.D. dissertation, Dept. Inform. Math. Modeling, Tech. Univ. Denmark, Lyngby, Denmark, http://www.imn.dtu.dk/~aa/phd/, 1994.

[18] A. A. Nielsen and K. Conradsen, "Multivariate alteration detection (MAD) in multispectral, bi-temporal image data: a new approach to change detection studies," Dept. Math. Modeling, Tech. Univ. Denmark, Lyngby, Denmark, http://www.imn.dtu.dk/~aa/tech-rep-1997-11/, Tech. Rep. 1997-11, 1997.

[19] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 19, pp. 1–19, 1998.

[20] A. A. Nielsen, "Multi-channel remote sensing data and orthogonal transformations for change detectection," in *Proc. Concerted Action MAVIRIC (MAchine Vision in Remotely Sensed Image Comprehension)*, I. Kanellopoulos, G. G. Wilkinson, and T. Moons, Eds., 1999, pp. 37–48.

[21] I. Niemeyer, M. Canty, and D. Klaus, "Possibilities and limits of remote sensing for the verification of international agreements: Algorithms to detect changes in nuclear plants," in *Proc. Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Seattle, WA, July 6–10, 1998, pp. 819–821.

[22] ——, "Unsupervised change detection techniques using multispectral satellite images," in *Proc. Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Hamburg, Germany, 1999, pp. 327–329.

[23] W. Pälchen, G. Rank, A. Kluge, A. A. Nielsen, and B. K. Ersbøll, "A new approach to differentiation between geogenic and anthropogenic influences on soils in a mining processing area," in *Proc. Nordic Symp. Variability in Polluted Soil and Groundwater*, K. Haarstad, Ed., June 1995.

[24] P. Switzer and A. A. Green, "Min/Max autocorrelation factors for multivariate spatial imagery," Stanford Univ., Stanford, CA, Tech. Rep. 6, 1984.

[25] P. Switzer and S. E. Ingebritsen, "Ordering of time-difference data from multispectral imagery," *Remote Sens. Environ.*, vol. 20, pp. 85–94, 1986.

[26] K. V. Shettigara and C. A. McGilchrist, "A principal component and canonical correlation hybrid technique for change detection in two-image sets," in *Proc. ASSPA 89, Signal Processing, Theories, Implementations and Applications*, R. F. Barrett, Ed., 1989, pp. 47–52.

[27] H. Hanaizumi and S. Fujimura, "Change detection from remotely sensed multi-temporal images using multiple regression," in *Proc. Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 1992, pp. 564–566.

[28] H. Hanaizumi, S. Chino, and S. Fujimura, "A method for change analysis with weight of significance using multi-temporal, multi-spectral images," in *Proc. SPIE 2315 Eur. Symp. Satellite Remote Sensing, Image and Signal Processing for Remote Sensing*, J. Desachy, Ed., 1994, pp. 282–288.

[29] B. Vinograde, "Canonical positive definite matrices under internal linear transformations," in *Proc. Amer. Math. Soc.*, 1950, pp. 159–161.

[30] R. G. D. Steel, "Minimum generalized variance for a set of linear functions," *Ann. Math. Statist.*, vol. 22, pp. 456–460, 1951.

[31] P. Horst, "Relations among m sets of measures," *Psychometrika*, vol. 26, pp. 129–149, 1961.

[32] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, pp. 433–451, 1971.

[33] J. J. Royer and J. L. Mallet, "Imagerie Spatielle et Cartographie Géochemique. Étude des Correlation Entre Géochemie en Roches et Télédétetion: Application au Massif de la Marche Orientale (Massif Central),", Tech. Rep. 80/CNES/279, CNES et CNRS, 1982.

[34] K. B. Hilger, A. A. Nielsen, and R. Larsen, "A scheme for initial exploratory analysis of multivariate image data," in *Proc. Scand. Conf. Image Analysis (SCIA)*, Bergen, Norway, June 11–14, 2001, pp. 717–724.

[35] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "Transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sensing*, vol. 26, pp. 65–74, Jan. 1988.

[36] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 295–304, May 1990.

[37] W. J. Dixon, Ed., *BMDP Statistical Software*. Los Angeles, CA: Univ. California Press, 1985.

[38] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorenson, "LAPACK users' guide," in *Proc. Soc. Ind. Appl. Math.*, Philadelphia, PA, 1999.

[39] G. Strang, *Linear Algebra and Its Applications*, 2nd ed. New York: Academic, 1980.

[40] P. S. Hansen, *Linear Algebra—Datamatorienteret*. Lyngby, Denmark: Matematetisk Institut og Numerisk Institut, Danmarks Tekniske Højskole, 1987.

[41] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. R. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[42] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman, *GAMS: A User's Guide*. Washington, DC: GAMS Development Corp., 1998.
[43] A. Drud, "CONOPT—A GRG code for large sparse dynamic nonlinear optimization problems," *Math. Progr.*, vol. 31, pp. 153–191, 1985.
[44] H. Olsson, "Regression functions for multitemporal relative calibration of thematic mapper data over boreal forest," *Remote Sens. Environ.*, vol. 46, no. 1, pp. 89–102, 1993.
[45] ——, "Monitoring of local reflection changes in boreal forests using satellite data," Ph.D. dissertation, Remote Sensing Lab., Swedish Univ. Agricultural Sciences, Umeå, Sweden, 1994.
[46] ——, "Reflectance calibration of thematic mapper data for forest change detection," *Int. J. Remote Sensing*, vol. 16, no. 1, pp. 81–96, 1995.
[47] R. Gnanadesikan, *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley, 1977.
[48] K. B. Hilger, "Exploratory analysis of multivariate data: Unsupervised image segmentation and data driven linear and nonlinear decomposition," Ph.D. dissertation, Dept. Inform. Math. Modeling, Tech. Univ. Denmark, Lyngby, Denmark, http://www.imm.dtu.dk/~kbh/phd/phdkbh.pdf/, 2001.

Allan Aasbjerg Nielsen received the M.Sc. degree in 1978 from the Department of Electrophysics, Technical University of Denmark, Lyngby, and the Ph.D. degree in 1994 from Informatics and Mathematical Modelling (IMM), Technical University of Denmark.

He is an Associate Professor with IMM. He was with the Danish Defense Research Establishment in 1977–1978. He worked on energy conservation in housing with the Thermal Insulation Laboratory, Technical University of Denmark, from 1978 to 1985. Since 1985, he has been with the Section for Image Analysis, IMM. He has worked on several national and international projects on development, implementation, and application of statistical methods and remote sensing in mineral exploration, mapping, geology, environment, oceanography, geodesy, and agriculture funded by industry, the European Union, Danida (the Danish International Development Agency), and the Danish National Research Councils.