C04351 Statistical Image Analysis, Spring 1999

Orthogonal Transformations

Allan Aasbjerg Nielsen

Department of Mathematical Modelling
Technical University of Denmark, Building 321
DK-2800 Lyngby, Denmark
www.imm.dtu.dk/∼aa
aa@imm.dtu.dk

1 Multichannel Imagery

When one collects multivariate data in some field of application a redundancy effect often arises because of covariation between variables. An interesting issue in reduction of dimensionality of the data is the desire to obtain simplicity for better understanding, visualizing and interpreting the data on the one hand, and the desire to retain sufficient detail for adequate representation on the other hand. E.g. a remote sensing device typically measures the emitted intensity at a number of discrete wavelengths or wavelength intervals for each element in a regular grid. This "repetition" of the measurement at different wavelengths induces a high degree of redundancy in the dataset. This can be used for noise reduction and data compression.

A traditional method used in this context is the celebrated principal components transformation. This is a pixel-wise operation that does not take the spatial nature of image data into account. Also, principal components will not always produce components that show decreasing image quality with increasing component number. It is perfectly imaginable that certain types of noise have higher variance than certain types of signal components.

First I shall briefly consider the theory of principal components, see also Conradsen (1984), Anderson (1984) who both describe several other orthogonal transformations. Following this, two procedures for transformation of multivariate data given on a spatial grid (images) with the purpose of isolating signal from noise and data compression. These are the minimum/maximum autocorrelation factor transformation, which was first described by Switzer & Green (1984) and the maximum noise fractions transformation which was described by Green, Berman, Switzer, & Craig (1988). The methods are also described in Conradsen, Nielsen, & Thyrsted (1985), Ersbøll (1989), Conradsen & Ersbøll (1991), Conradsen, Nielsen, & Nielsen (1991), Conradsen & al. (1991), Larsen (1991), Nielsen (1994). Also, an application of MAF/MNFs to remove periodic noise in hyper-channel airborne scanner data is described. A new concept of MAF/MNFs of irregularly spaced data is suggested. This also gives rise to a new form of kriging, namely maximum autocorrelation factorial or maximum signal-to-noise factorial kriging. Finally, canonical discriminant analysis is described.

2 Principal Components

Based on a technique described by Pearson in 1901 Hotelling (1933) developed principal components (PC) analysis. The principal components of a stochastic multivariate variable are based on a linear transformation which produces uncorrelated variables of decreasing variance.

The application of this transformation requires knowledge of or an estimate of the dispersion matrix. The PCs maximize the variance represented by each component. PC one is the linear combination of the original bands that explains maximum variance in the original data. A higher order PC is the linear combination of the original bands that explains maximum variance subject to the constraint that it is uncorrelated with lower order PCs. PC analysis performs an observation-wise operation that does not take the spatial nature of an image into account.

Let us consider a multivariate data set of m bands with grey levels $Z_i(\boldsymbol{x}), i = 1, \ldots, m$, where $\boldsymbol{x}$ denotes the coordinates of the sample, and the dispersion is $\mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{\Sigma}$, where $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$. Without loss of generality we assume that $\mathrm{E}\{\boldsymbol{Z}\} = \boldsymbol{0}$.

We are looking for linear combinations of $\boldsymbol{Z}$

$$X(\boldsymbol{x}) = \boldsymbol{p}^T \boldsymbol{Z}(\boldsymbol{x}) \tag{1}$$

with maximum variance

$$\mathrm{Var}\{X(\boldsymbol{x})\} = \boldsymbol{p}^T \boldsymbol{\Sigma} \boldsymbol{p}. \tag{2}$$

If $\boldsymbol{p}$ is a solution so is $c\boldsymbol{p}$ where c is any scalar. We choose $\boldsymbol{p}$ so that $\boldsymbol{p}^T\boldsymbol{p} = 1$.

Let $\lambda_1 \geq \lambda_2, \geq \cdots \geq \lambda_m$ be the eigenvalues and $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_m$ corresponding orthogonal eigenvectors of $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}\boldsymbol{p}_i = \lambda_i \boldsymbol{p}_i$. Then

$$X_i(\boldsymbol{x}) = \boldsymbol{p}_i^T \boldsymbol{Z}(\boldsymbol{x}), i = 1, \ldots, m \tag{3}$$

with variance λ_i is the i'th principal component.

We see that with $\boldsymbol{P} = [\boldsymbol{p}_1 \ldots \boldsymbol{p}_m]$

$$\boldsymbol{X} = \boldsymbol{P}^T \boldsymbol{Z} \text{ with } \boldsymbol{P}\boldsymbol{P}^T = \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}. \tag{4}$$

For the dispersion of $\boldsymbol{X}$ we get

$$\mathrm{D}\{\boldsymbol{X}\} = \boldsymbol{P}^T \boldsymbol{\Sigma} \boldsymbol{P} = \boldsymbol{\Lambda} \text{ with } \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix}. \tag{5}$$

This means that

$$\boldsymbol{\Sigma} = \boldsymbol{P}^{-T} \boldsymbol{\Lambda} \boldsymbol{P}^{-1} = \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^T = (\boldsymbol{P}\boldsymbol{\Lambda}^{\frac{1}{2}})(\boldsymbol{P}\boldsymbol{\Lambda}^{\frac{1}{2}})^T. \tag{6}$$

If we seek the i variables describing as much as possible of the total variance of the original variables the solution is the first i principal components. The fraction of the total variance described by these is given by

$$\frac{\lambda_1 + \cdots + \lambda_i}{\lambda_1 + \cdots + \lambda_i + \cdots + \lambda_m}. \tag{7}$$

A drawback of principal components analysis is that its results depend on the unit of measurement of the original variables. This problem can be circumvented by considering the standardized variables instead, i.e. by performing the PC transformation on the correlation matrix instead of on the dispersion matrix.

3 Min/Max Autocorrelation Factors

As opposed to the principal components transformation the minimum/maximum autocorrelation factors (MAF) transformation allows for the spatial nature of the image data. The application of this transformation requires knowledge of or an estimate of the dispersion matrix and the dispersion matrix of the difference between the original the spatially shifted image. The MAF transform minimizes the autocorrelation rather than maximizing the data variance (PC). In reverse order the MAFs maximize the autocorrelation represented by each component. MAF one is the linear combination of the original bands that contains minimum autocorrelation between neighboring pixels. A higher order MAF is the linear combination of the original bands that contains minimum 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. The MAF transformation is equivalent to a transformation of the data to a coordinate system in which the covariance matrix of the spatially shifted image data is the identity matrix followed by a principal components transformation.

An important property of the MAF procedure is its invariance to linear transforms, 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.

The minimum/maximum autocorrelation factors procedure was suggested by Switzer & Green (1984). PCs, MAFs and other orthogonal transforms are described in Ersbøll (1989), Conradsen & Ersbøll (1991), Nielsen (1994). Again we consider the random variable $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$ and we assume that

$$\begin{aligned} \mathrm{E}\{\boldsymbol{Z}(\boldsymbol{x})\} &= \boldsymbol{0} && (8)\\ \mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} &= \boldsymbol{\Sigma}. && (9) \end{aligned}$$

We denote a spatial shift by $\boldsymbol{\Delta}^T = [\Delta_1, \Delta_2]$. The spatial covariance function is defined by

$$\mathrm{Cov}\{\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{\Gamma}(\boldsymbol{\Delta}). \tag{10}$$

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

$$\begin{aligned} \boldsymbol{\Gamma}(\boldsymbol{0}) &= \boldsymbol{\Sigma} && (11)\\ \boldsymbol{\Gamma}(\boldsymbol{\Delta})^T &= \boldsymbol{\Gamma}(-\boldsymbol{\Delta}). && (12) \end{aligned}$$

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

$$\begin{aligned} \mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} &= \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})\boldsymbol{a} && (13)\\ &= (\boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})\boldsymbol{a})^T\\ &= \boldsymbol{a}^T\boldsymbol{\Gamma}(\boldsymbol{\Delta})^T\boldsymbol{a}\\ &= \boldsymbol{a}^T\boldsymbol{\Gamma}(-\boldsymbol{\Delta})\boldsymbol{a}\\ &= \frac{1}{2}\boldsymbol{a}^T\left(\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta})\right)\boldsymbol{a}. \end{aligned}$$

Introducing

$$\begin{aligned} \boldsymbol{\Sigma}_\Delta &= \mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} && (14)\\ &= \mathrm{E}\{[\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})][\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})]^T\}, \end{aligned}$$

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

$$\boldsymbol{\Gamma}(\boldsymbol{\Delta}) + \boldsymbol{\Gamma}(-\boldsymbol{\Delta}) = 2\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\Delta \tag{15}$$

and thus

$$\mathrm{Cov}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = \boldsymbol{a}^T(\boldsymbol{\Sigma} - \frac{1}{2}\boldsymbol{\Sigma}_\Delta)\boldsymbol{a} \tag{16}$$

wherefore

$$\mathrm{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}}. \tag{17}$$

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

$$R(\boldsymbol{a}) = \frac{\boldsymbol{a}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}}. \tag{18}$$

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

$$Y_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{Z}(\boldsymbol{x}) \tag{19}$$

is the i'th minimum/maximum autocorrelation factor or shortly the i'th MAF. The minimum/maximum autocorrelation factors satisfy

i) $\text{Corr}\{Y_i(\boldsymbol{x}), Y_j(\boldsymbol{x})\} = 0,\ i \neq j$,

ii) $\text{Corr}\{Y_i(\boldsymbol{x}), Y_i(\boldsymbol{x}+\boldsymbol{\Delta})\} = 1 - \frac{1}{2}\kappa_i$,

iii) $\text{Corr}\{Y_1(\boldsymbol{x}), Y_1(\boldsymbol{x}+\boldsymbol{\Delta})\} = \inf_a \ \text{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\}$,
$\text{Corr}\{Y_m(\boldsymbol{x}), Y_m(\boldsymbol{x}+\boldsymbol{\Delta})\} = \sup_a \ \text{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\}$,
$\text{Corr}\{Y_i(\boldsymbol{x}), Y_i(\boldsymbol{x}+\boldsymbol{\Delta})\} = \inf_{a\in\mathcal{M}_i} \ \text{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\}$,
$\mathcal{M}_i = \{\boldsymbol{a} \mid \text{Corr}\{\boldsymbol{a}^T\boldsymbol{Z}(\boldsymbol{x}), Y_j(\boldsymbol{x})\} = 0,\ j = 1, \ldots, i-1\}$.

The reverse numbering of MAFs so that the signal MAF is referred to as MAF1 is often used.

3.1 Linear Transformations of MAFs

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

$$\boldsymbol{U}(\boldsymbol{x}) = \boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) \tag{20}$$

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

$$R_1(\boldsymbol{b}) = \frac{\boldsymbol{b}^T\boldsymbol{T}\boldsymbol{\Sigma}_\Delta\boldsymbol{T}^T\boldsymbol{b}}{\boldsymbol{b}^T\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\boldsymbol{b}}. \tag{21}$$

The equation for solving the eigenproblem is

$$\begin{aligned} \boldsymbol{T}\boldsymbol{\Sigma}_\Delta\boldsymbol{T}^T\boldsymbol{b}_i &= \lambda_i\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T\boldsymbol{b}_i \Leftrightarrow \\ \boldsymbol{\Sigma}_\Delta(\boldsymbol{T}^T\boldsymbol{b}_i) &= \lambda_i\boldsymbol{\Sigma}(\boldsymbol{T}^T\boldsymbol{b}_i) \end{aligned} \tag{22}$$

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

$$\begin{aligned} \boldsymbol{b}_i^T\boldsymbol{U}(\boldsymbol{x}) &= \boldsymbol{b}_i^T\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) \\ &= (\boldsymbol{T}^T\boldsymbol{b}_i)^T\boldsymbol{Z}(\boldsymbol{x}) \\ &= \boldsymbol{a}_i^T\boldsymbol{Z}(\boldsymbol{x}) \\ &= Y_i(\boldsymbol{x}). \end{aligned} \tag{23}$$

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—inspired by Equation 6—set

$$\boldsymbol{T}^T = \boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}} \tag{24}$$

we have for the dispersion of the transformed variables

$$\text{D}\{\boldsymbol{U}(\boldsymbol{x})\} = \text{D}\{\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}^T = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{\Sigma}\boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \boldsymbol{I}. \tag{25}$$

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

$$\begin{aligned} \boldsymbol{T}\boldsymbol{\Sigma}_\Delta\boldsymbol{T}^T &= \text{D}\{\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}) - \boldsymbol{T}\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{\Delta})\} \\ &= \text{D}\{\boldsymbol{U}(\boldsymbol{x}) - \boldsymbol{U}(\boldsymbol{x}+\boldsymbol{\Delta})\} \end{aligned} \tag{26}$$

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 & 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 $\hat{\boldsymbol{\Sigma}}_\Delta$.

4 Maximum Noise Fractions

Principal components do not always produce components of decreasing image quality. When working with spatial data the maximization of variance across bands is not an optimal approach if the issue is this ordering. In this section we will maximize a measure of image quality, namely a signal-to-noise ratio. This should ensure achievement of the desired ordering in terms of image quality. In the previous section another measure of image quality namely spatial autocorrelation was dealt with.

If we estimate the noise at a pixel site as the difference of the pixel value at that site and the value of a neighboring pixel, we obtain the same eigenvectors as in the MAF analysis.

The maximum noise fractions (MNF) transformation can be defined in several ways. It can be shown that the same set of eigenvectors is obtained by procedures that maximize the signal-to-noise ratio and the noise fraction. The procedure was first introduced by Green et al. (1988) where the authors in continuation of the MAF work by Switzer & Green (1984) choose the latter. Hence the name maximum noise fractions.

The MNF transformation maximizes the noise content rather than maximizing the data variance (PC) or minimizing the autocorrelation (MAF). The application of this transformation requires knowledge of or an estimate of the signal and noise dispersion matrices. In reverse order the MNFs maximize the signal-to-noise ratio represented by each component MNF one is the linear combination of the original bands that contains minimum signal-to-noise ratio. A higher order MNF is the linear combination of the original bands that contains minimum signal-to-noise ratio subject to the constraint that it is orthogonal to lower order MNFs. The MNF transformation is equivalent to a transformation of the data to a coordinate system in which the noise covariance matrix is the identity matrix followed by a principal components transformation. The MNFs therefore also bear the name noise adjusted principal components (NAPC), cf. Lee, Woodyatt, & Berman (1990). The MNFs share the MAFs' property of invariance to linear transforms.

First we will deduce the maximum noise fraction transformation. We will then briefly describe methods for estimating the dispersion of the signal and the noise.

Let us as before consider a multivariate data set of m bands with grey levels $Z_i(\boldsymbol{x}), i = 1, \ldots, m$, where $\boldsymbol{x}$ denotes the coordinates of the sample. We will assume an additive noise structure

$$\boldsymbol{Z}(\boldsymbol{x}) = \boldsymbol{S}(\boldsymbol{x}) + \boldsymbol{N}(\boldsymbol{x}), \tag{27}$$

where $\boldsymbol{Z}^T = [Z_1(\boldsymbol{x}), \ldots, Z_m(\boldsymbol{x})]$, and $\boldsymbol{S}(\boldsymbol{x})$ and $\boldsymbol{N}(\boldsymbol{x})$ are the uncorrelated signal and noise components. As before we assume that $\mathrm{E}\{\boldsymbol{Z}\} = \boldsymbol{0}$. Therefore

$$\mathrm{D}\{\boldsymbol{Z}(\boldsymbol{x})\} = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_S + \boldsymbol{\Sigma}_N, \tag{28}$$

where $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_N$ are the dispersion matrices for $\boldsymbol{S}(\boldsymbol{x})$ and $\boldsymbol{N}(\boldsymbol{x})$ respectively. Note that the techniques described in this section can in principle be applied to multiplicative noise also by first taking logarithms of the observations.

We define the signal-to-noise ratio of the i'th band as

$$\frac{\mathrm{Var}\{S_i(\boldsymbol{x})\}}{\mathrm{Var}\{N_i(\boldsymbol{x})\}}, \tag{29}$$

i.e. the ratio of the signal variance and the noise variance. We define the noise fraction of the i'th band as

$$\frac{\mathrm{Var}\{N_i(\boldsymbol{x})\}}{\mathrm{Var}\{Z_i(\boldsymbol{x})\}}, \tag{30}$$

i.e. the ratio of the noise variance and the total variance. We define the maximum noise fraction transformation as the linear transformations

$$Y_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{Z}(\boldsymbol{x}), i = 1, \ldots, m \tag{31}$$

such that the signal-to-noise ratio for $Y_i(\boldsymbol{x})$ is maximum among all linear transforms orthogonal to $Y_j(\boldsymbol{x}), j = 1, \ldots, i-1$. Furthermore we shall assume that the vectors $\boldsymbol{a}_i$ are normed so that

$$\boldsymbol{a}_i^T \boldsymbol{\Sigma} \boldsymbol{a}_i = 1, i = 1, \ldots, m. \tag{32}$$

Maximization of the noise fraction leads to the opposite numbering, namely a numbering that gives increasing image quality with increasing component number.

The SNR for $Y_i(\boldsymbol{x})$ is

$$
\begin{aligned}
\frac{\text{Var}\{\boldsymbol{a}_i^T \boldsymbol{S}(\boldsymbol{x})\}}{\text{Var}\{\boldsymbol{a}_i^T \boldsymbol{N}(\boldsymbol{x})\}} &= \frac{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_S \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i} \qquad (33) \\
&= \frac{\boldsymbol{a}_i^T (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_N) \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i} \\
&= \frac{\boldsymbol{a}_i^T \boldsymbol{\Sigma} \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i} - 1.
\end{aligned}
$$

If we work on the noise fraction instead, we get

$$
\frac{\text{Var}\{\boldsymbol{a}_i^T \boldsymbol{N}(\boldsymbol{x})\}}{\text{Var}\{\boldsymbol{a}_i^T \boldsymbol{Z}(\boldsymbol{x})\}} = \frac{\boldsymbol{a}_i^T \boldsymbol{\Sigma}_N \boldsymbol{a}_i}{\boldsymbol{a}_i^T \boldsymbol{\Sigma} \boldsymbol{a}_i}. \qquad (34)
$$

In both cases we will find the vectors $\boldsymbol{a}_i$ as eigenvectors to the real, symmetric, generalized eigenproblem

$$
\det(\boldsymbol{\Sigma}_N - \lambda \boldsymbol{\Sigma}) = 0. \qquad (35)
$$

Thus the SNR for $Y_i(\boldsymbol{x})$ is given by

$$
\frac{1}{\lambda_i} - 1, \qquad (36)
$$

where λ_i is the eigenvalue of $\boldsymbol{\Sigma}_N$ with respect to $\boldsymbol{\Sigma}$.

An important characteristic of the MNF transformation which is not shared by the PC transformation is the invariability to linear scaling (the signal-to-noise *ratio* is maximized).

As for the MAFs the reverse numbering of MNFs so that the signal MNF is referred to as MNF1 is often used (in this case the transformation is often termed the minimum noise fraction transformation).

4.1 Estimation of the Noise Covariance Matrix

The central problem in the calculation of the MNF transformation is the estimation of the noise with the purpose of generating a dispersion matrix that approximates $\boldsymbol{\Sigma}_N$. In this process we will make use of the spatial characteristics of the image. Five methods are suggested, see also Olsen (1993)

- Simple differencing. The noise is estimated as for MAFs as the difference between the current and a neighboring pixel. In this case we refer to $\boldsymbol{\Sigma}_N$ as $\boldsymbol{\Sigma}_\Delta$.
- Causal SAR. The noise is estimated as the residual in a simultaneous autoregressive (SAR) model involving the neighboring pixel to the W, NW, N and NE of the current pixel.
- Differencing with the local mean. More pixels could be entered in to the estimation by differencing between the current pixel and the local mean.
- Differencing with local median. Mean filters blur edges and other details. This could be avoided by using the local median instead of the local mean.
- Quadratic surface. The noise is estimated as the residual from a fitted quadratic surface in a neighborhood.

4.2 Periodic Noise

As satellite images and images obtained from airborne scanners often are corrupted by striping we will consider methods for eliminating this form of noise. As periodic noise such as various forms of striping often has a high degree of spatial correlation, it will often be considered signal by the MAF and MNF transformations. To some extent the striping will be isolated in some of the factors. Periodic noise can be removed by Fourier methods. It should be noted that periodic noise can be very disturbing as the regular pattern catches the viewer's eyes.

A simple bandwise Fourier filtering may corrupt significant parts of the relevant signal. Therefore we shall minimize the amount of filtering by eliminating this noise by filtering out the relevant structures in the MNF Fourier domain. In order not to create an inverse pattern by setting the Fourier values to zero we keep the phase and fill the magnitude values by an iterative algorithm that takes means of the neighboring values. If we want to remove other types of noise also (e.g. salt-and-pepper noise) we can filter or skip the MNFs that contain the noise pattern in question before transforming back from MNF space to the original image space, Nielsen (1994).

5 MAF/MNFs of Irregularly Spaced Data

Consider the above formulation of the MAF/MNF problem. For irregularly spaced data an alternative to the estimate of the noise dispersion $\mathbf{\Sigma}_N$ (or rather $\mathbf{\Sigma}_\Delta$) is simply the dispersion matrix of a new variable consisting of the difference between a data value and its nearest neighbor for all variables. This defines a new data analytical concept namely minimum/maximum autocorrelation factors (MAFs) for irregularly spaced multivariate data, Nielsen, Conradsen, Pedersen, & Steenfelt (1997).

Analogous to the extension of MAFs into MNFs for gridded data, a more elaborate model for $\mathbf{\Sigma}_N$ based on each observation's neighborhood can be defined. With gridded data the neighborhood is easily defined. With non-gridded or irregularly spaced data, the Voronoi tessellation of the plane and its dual concept, the Delaunay triangulation are useful. To each point in the plane we associate a Voronoi polygon which is the part of the plane that is nearer to that point than to any other point. From the Voronoi tessellation we can construct the Delaunay triangulation by joining points with common Voronoi polygon edges.

Two ways of estimating $\mathbf{\Sigma}_N$ come directly to mind: the use of the dispersion matrix of a new variable consisting of the difference between a data value and the mean or the median of all its, say, first order Delaunay neighbors for all variables. This defines a new data analytical concept namely maximum noise fractions (MNFs) for irregularly spaced multivariate data.

Of course, both the MAFs and MNFs defined in this fashion can be extended to allow for other neighborhoods, e.g. confined by distance and/or direction constraints.

Grunsky & Agterberg (1988, 1991) circumvent the problem of irregularity of the sampling pattern by fitting parametric surfaces to observed correlations. Because of the lack of positive definiteness of the joint correlation structures this approach seems less satisfactory than the method proposed here.

The MAF/MNFs for irregularly spaced data possess the characteristics that they are orthogonal and they are ordered by decreasing autocorrelation/signal-to-noise ratio. Typically, low order factors will contain a lot of signal, high order factors will contain a lot of noise. The MAF/MNFs thus relate beautifully to interpolation by kriging. Because the low order MAF/MNFs contain signal they are expected to have low nugget effects and long ranges of influence. This tendency is expected to develop towards higher nugget effects and shorter ranges of influence as the higher order MAF/MNFs contain increasingly more noise. As the factors are orthogonal there is no cross-correlation for small lags or displacements. I therefore suggest the use of separate (as opposed to co-) kriging for interpolation purposes and introduce a new data analytic concept: maximum autocorrelation factorial or maximum signal-to-noise factorial kriging based on the above new MAF/MNF analysis of non-gridded, multivariate data. In order to obtain kriged versions of the original data the inverse MAF/MNF transformation can be applied.

6 Example, Noise in Hyperspectral GERIS Data

In this section the above orthogonal transformations are applied to a hyperspectral scene, namely a dataset recorded over central Spain using the Geophysical Environmental Research Corporation (GER) airborne scanner with the purpose of isolating signal from noise.

The GER imaging spectrometer (GERIS) actually consists of three spectrometers, that view the ground through the same aperture via an optoelectronic scanning device. The three spectrometers record a total of 63 bands through the visible, near infrared and short wave-infrared wavelength range between 0.47 and 2.45 μm. The spectral resolution in the visible region between 0.47 and 0.84 μm is 12.3 nm. In the near infrared region from 1.40 to 1.90 μm it is much broader, around 120 nm. In the short wave-infrared region between 2.00 and 2.45 μm the sampling frequency is 16.2 nm.

The scanner uses a rotating mirror perpendicular to the flight direction to scan a line of 512 pixels with a scan angle of 45° to either side of the flight track. A flight altitude of 3,000 meters and an aperture setting of 2.5 mrad leads to a nominal pixel size of 7.5 meters. The recorded data are stored in 16 bits with a dynamic range of 12 bits. After recording, the dataset is corrected for aircraft roll by the use of roll data recorded by a gyroscope hard mounted to the scanner optics.

Apart from noise introduced by the atmosphere, the instrumentation, and from quantization and sampling, the GERIS data are corrupted by a heavy two line and four line banding. This is due to slight differences of the surfaces of the rotating mirrors in the scanning device. These differences in the optical properties probably stem from dirt and oil on the surfaces.

6.1 PC versus MAF

In Figures 1 and 2 the 62 principal components and the 62 minimum/maximum autocorrelation factors are shown. The images are ordered row wise with component/factor 1 in the top-left corner (paper in landscape mode). Each subimage consists of 160×240 7.5×7.5 m^2 pixels. Because of the extreme noise content channel 28 is omitted from the analyses. It is evident that the principal components transformation is not capable of producing a natural ordering of image quality. The minimum/maximum autocorrelation factors do a much better job in terms of ordering as well as separating signal from noise. One might describe the MAF transformation as a decomposition of spatial frequency.

7 Canonical Discriminant Analysis

Consider k groups (or classes or populations) with $n_1, \ldots, n_k$ multivariate (p-dimensional) observations $\{\boldsymbol{X}_{ij}\}$, where i is the group index and the j is the (multivariate) observation number. The group means are denoted $\bar{\boldsymbol{X}}_1, \ldots, \bar{\boldsymbol{X}}_k$ and the overall mean is denoted $\bar{\boldsymbol{X}}$, i.e.

$$\bar{\boldsymbol{X}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{X}_{ij},\ i = 1, \ldots, k$$

and

$$\bar{\boldsymbol{X}} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \boldsymbol{X}_{ij} \text{ with } N = \sum_{i=1}^{k} n_i.$$

As in a one-way analysis of variance the "total" sum of squares matrix is

$$\boldsymbol{T} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}})(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}})^T.$$

We define the "among groups" matrix as

$$\boldsymbol{A} = \sum_{i=1}^{k} n_i (\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})^T$$

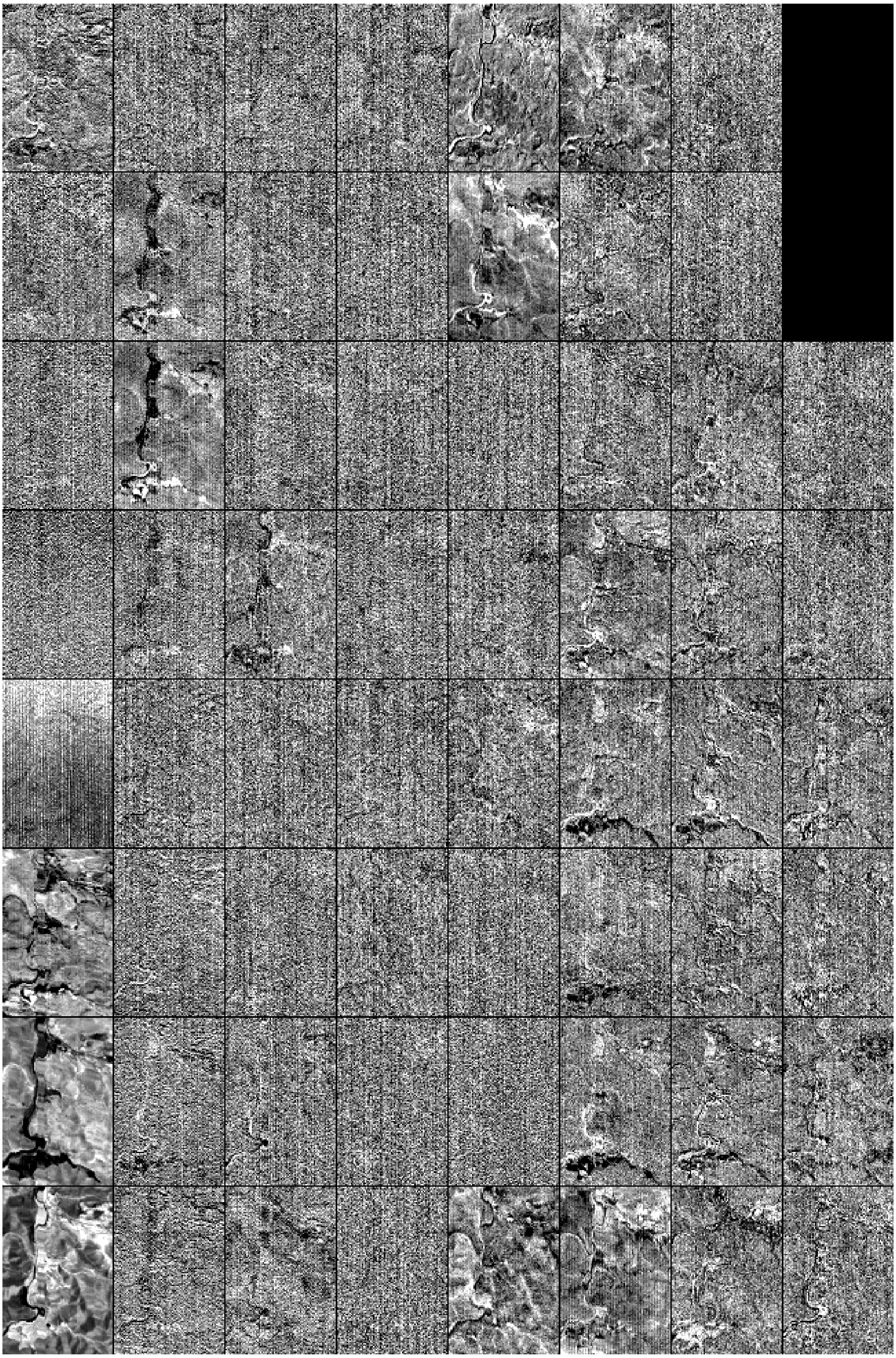

Figure 1: Principal components of 62 GERIS bands

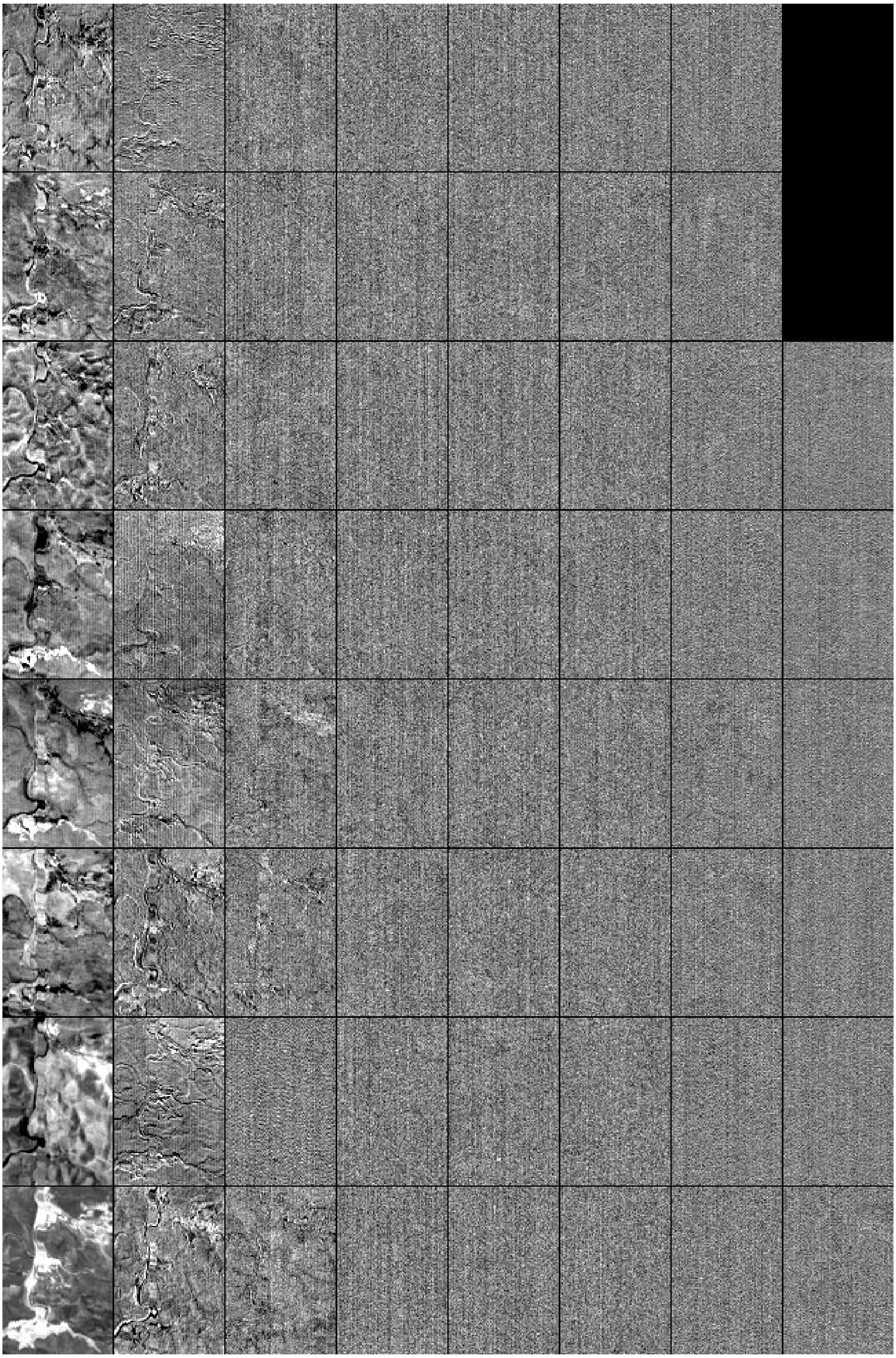

Figure 2: Minimum/maximum autocorrelation factors of 62 GERIS bands

and the "within groups" matrix as

$$\boldsymbol{W} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)^T.$$

With these definitions we have

$$\boldsymbol{T} = \boldsymbol{A} + \boldsymbol{W}$$

or in words: the total variation can be written as a sum of the variation of the group means around the overall mean and the variation around the group means.

We are looking for new variates $Y = \boldsymbol{d}^T(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}})$ that maximise the ratio between variation among groups and variation within groups; the latter can be considered as the natural level of variance of the variables $\boldsymbol{X}_{ij}$. The idea of maximising this ratio is due to Fisher (1936). This ratio equals the Rayleigh coefficient $\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d}/\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}$, i.e., the transformation is defined by the conjugate eigenvectors $\boldsymbol{d}_\ell$ of $\boldsymbol{A}$ with respect to $\boldsymbol{W}$

$$\boldsymbol{A}\boldsymbol{d}_\ell = \lambda_\ell \boldsymbol{W}\boldsymbol{d}_\ell.$$

Again inspired by one-way analysis of variance we define the canonical correlation coefficients R_ℓ by their squares, $R_\ell^2 = \boldsymbol{d}_\ell^T\boldsymbol{A}\boldsymbol{d}_\ell/\boldsymbol{d}_\ell^T\boldsymbol{T}\boldsymbol{d}_\ell$. This gives the relation $R_\ell^2 = \lambda_\ell/(\lambda_\ell+1)$. The new variates $Y_\ell = \boldsymbol{d}_\ell^T(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}})$ are called canonical discriminant functions (CDF). The first CDF defined by $\boldsymbol{d}_1$ is the affine transformation of the original variables that gives the best discrimination between the k groups. A higher order CDF is the affine transformation of the original variables that gives the best discrimination between the k groups subject to the constraint that it is orthogonal (with respect to $\boldsymbol{A}$ and $\boldsymbol{W}$) to the lower order CDFs. Note, that the number of CDFs is given by rank considerations for $\boldsymbol{A}$ and $\boldsymbol{W}$. If $\boldsymbol{A}$ and $\boldsymbol{W}$ have full rank this number equals $\min(k-1, p)$.

Scatterplots of the first few CDFs give a good visual impression of the separability of the observations.

8 Canonical Correlations Analysis

This type of analysis will not be described here; see Conradsen (1984), Nielsen & Conradsen (1997), Nielsen, Conradsen, & Simpson (1998).

9 The Multivariate Analysis Program - `maf`

The HIPS program `maf` performs several kinds of orthogonal transformations, including principal component, factor, canonical correlations (multivariate alteration detection), minimum/maximum autocorrelation factor, maximum noise fractions, and canonical discriminant analysis.

It is recommended that you study the manual page for this program (`man maf`). You can print the manual page by means of

```
groff -man -Tps < /usr/local/man/man1/maf.1 | lp -dps2 -onb
```

(or another postscript printer you may wish to use).

Satellite scanners are calibrated to produce non-saturated images globally, i.e. over land, water and ice. Therefore we often need to stretch local satellite images in order to get a proper result. This can be done by using

```
saturate -s -l -3 3 -b < image.hips | xshow
```

or by using

```
scale -e image.hips | histobe -a 2 -b 2 | xshow
```

10 Problems

The image sequence sotiel.hips consists of an 256×256 extracted section of 12 bands of a 63 channel image sequence recorded over central Spain. The first 6 bands are from the visual part of the spectrum and the last 6 are from the infrared part of the spectrum.

The image sequences thikaxs87.hips and thikaxs89.hips are co-registered SPOT HRV data recorded over Kenya in the years 1987 and 1989, respectively.

The estimation of covariance matrices requires access to pixels from all bands at the same time. By band interleaving the input sequence by line a more RAM economic way of calculating the transformations is obtained. This can be done by use of `bil`.

To avoid unnecessary computation you should note the possibility `maf` gives you to store computed covariance matrices. Also you should be aware of the size of the produced output.

When displaying and printing images you might want to show several frames at the same time. You can obtain this by using the HIPS program `comicstrip`.

1. Compare the principal components of the Sotiel sequence calculated using the raw and the standardized data. What is the difference? Why is this? Apart from the visual differences you might want to compare the correlations of the components with the original variables. Why is this a better measure than the actual weight?

2. Calculate the minimum/maximum autocorrelation factors and compare these to the principal components. What are the differences? Why is this?

3. Calculate MAFs of the PCs and comment.

4. Calculate the canonical variables of the two GER datasets from Sotiel. What are the differences between the two sets? Comment on the components.

5. Calculate the canonical variables of the two SPOT datasets from Kenya. What are the differences between the two sets? Comment on the components.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (second edition). John Wiley, New York. 675 pp.

Conradsen, K. & al. (1991). *Development and Testing of New Techniques for Mineral-Exploration Based on Remote Sensing, Image Processing Methods and Multivariate Analysis. Final Report.* The Commission of the European Communities, Contract No. MA1M-0015-DK(B). 196 pp.

Conradsen, K. & Ersbøll, B. K. (1991). Data dependent orthogonal transformations of multichannel image data. Tech. rep., IMSOR, Technical University of Denmark. 35 pp.

Conradsen, K., Nielsen, B. K., & Thyrsted, T. (1985). A comparison of min/max autocorrelation factor analysis and ordinary factor analysis. In *Proceedings from Symposium in Applied Statistics*, pp. 47–56. Lyngby, Denmark.

Conradsen, K., Nielsen, B. K., & Nielsen, A. A. (1991). Noise removal in multichannel image data by a parametric maximum noise fractions estimator. In ERIM (Ed.), *Proceedings of the 24th International Symposium on Remote Sensing of Environment*, pp. 403–416. Rio de Janeiro, Brazil.

Conradsen, K. (1984). *Introduktion til Statistik*. IMSOR, DTH. Vol. 2B, In Danish.

Ersbøll, B. K. (1989). *Transformations and classifications of remotely sensed data*. Ph.D. thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby. 297 pp.

Fisher, R. A. (1936). The utilization of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179–188.

Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, **26**(1), 65–74.

Grunsky, E. C. & Agterberg, F. P. (1988). Spatial and multivariate analysis of geochemical data from metavolcanic rocks in the Ben Nevis area, Ontario. *Mathematical Geology*, **20**(7), 825–861.

Grunsky, E. C. & Agterberg, F. P. (1991). SPFAC: a Fortran program for spatial factor analysis of multivariate data. *Computers & Geosciences*, **17**(1), 133–160.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, **24**, 417–441.

Larsen, R. (1991). MAF and other transformations applied in remote sensing. Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby. In Danish. 130 pp.

Lee, J. B., Woodyatt, A. S., & Berman, M. (1990). Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, **28**(3), 295–304.

Nielsen, A. A. & Conradsen, K. (1997). Multivariate alteration detection (MAD) in multispectral, bi-temporal image data: a new approach to change detection studies. Tech. rep. 1997-11, Department of Mathematical Modelling, Technical University of Denmark. Internet http://www.imm.dtu.dk/documents/users/aa/tech-rep-1997-11/.

Nielsen, A. A., Conradsen, K., Pedersen, J. L., & Steenfelt, A. (1997). Spatial factor analysis of stream sediment geochemistry data from South Greenland. In Pawlowsky-Glahn, V. (Ed.), *Proceedings of ther Third Annual Conference of the International Association for Mathematical Geology* Barcelona, Spain. pp. 955–960.

Nielsen, A. A., Conradsen, K., & Simpson, J. J. (1998). Multivariate alteration detection (MAD) and MAF post-processing in multispectral, bi-temporal image data: new approaches to change detection studies. *Remote Sensing of Environment*, **64**, 1–19.

Nielsen, A. A. (1994). *Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data*. Ph.D. thesis, Institute of Mathematical Modelling, Technical University of Denmark, Lyngby. 213 pp. Internet http://www.imm.dtu.dk/documents/users/aa/phd/.

Olsen, S. I. (1993). Estimation of noise in images: an evaluation. *Graphical Models and Image Processing*, **55**(4), 319–323.

Switzer, P. & Green, A. A. (1984). Min/max autocorrelation factors for multivariate spatial imagery. Tech. rep. 6, Stanford University. 10 pp.