

# DECOMPOSITION USING MAXIMUM AUTOCORRELATION FACTORS

Rasmus Larsen

Informatics and Mathematical Modelling, Technical University of Denmark,  
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark  
Phone: +45 45253415; Fax: +45 45881397; E-Mail: rl@imm.dtu.dk

Original research article

## Summary

This article presents methods for the analysis and decomposition of multivariate datasets where a given ordering/structure of the observations or the variables exist. Examples of such data sets are remote sensing imagery where observations (pixels) each consisting of a reflectance spectrum are organised in a two-dimensional grid. Another example is biological shape analysis. Here each observation (e.g. human bone, cerebral ventricle) is represented by a number of landmarks the coordinates of which are the variables. Here we do not have an ordering of the observations (individuals). However, normally we have an ordering of landmarks (variables) along the contour of the objects. For the case with observation ordering the maximum autocorrelation factor (MAF) transform was proposed for multivariate imagery in [18]. This corresponds to a R-mode analyse of the data matrix. We propose to extend this concept to situations with variable ordering. This corresponds to a Q-mode analysis of the data matrix. We denote this methods Q-MAF decomposition. It turns out that in many situations the new variables resulting from the MAF and the Q-MAF analyses can be interpreted as a decomposition of (spatial) frequency. However, contrary to Fourier decomposition these new variables are located in frequency as well as location (space, time, wavelength etc).

**Keywords:** maximum autocorrelation factors, shape, ordered variables, ordered observations.

# 1 Introduction

The maximum autocorrelation factor (MAF) analysis was originally proposed as an alternative transformation of multivariate spatial imagery to the celebrated PCA transform by Paul Switzer [18]. In the MAF analysis we seek a transformation that maximizes the autocorrelation between neighbouring observations (i.e. pixels). The basic assumption of the MAF analysis is that the interesting signal exhibits high autocorrelation, whereas the noise exhibits low autocorrelation. By building the additional information of the structure of the observations into the model application examples show a more satisfying ordering and compression of the data (cf. [3, 7, 16, 15]). This is particularly the case when some noise components have higher variance than some signal components. In this case the principal components will fail to give an intuitive order of image quality. The MAF analysis requires knowledge of or estimation of the variance-covariance matrix of the data as well as the variance-covariance matrix of the difference between the original data and a spatially shifted version of the data. It may be formulated as a canonical correlation analysis problem [9]. A similar approach to filter design is shown in [11]. The power of the MAF transform is illustrated on a multivariate remotely sensed dataset, where a 62 channel sun-reflectance spectrum is measured in each pixel of an image. In an Appendix A we show that the Molgedey-Schuster algorithm for independent components analysis (ICA) is equivalent to MAF analysis.

We show how the MAF transform can be extended to situations where we exploit not the ordering of the observations, but the ordering of the variables. A large application area for this is biological shape analysis. In this context we are concerned with shapes that are represented by sets of corresponding points. The landmarks, the object representation, often consists of (using the terminology of [5]) combinations of anatomical landmarks, that correspond between organisms in a biologically meaningful way, mathematical landmarks that are allocated on an object according to some mathematical or geometrical property, and pseudo landmarks that are constructed points, e.g. dispersed along the outline of an object between anatomical or mathematical landmarks. For use for simulation, prediction, or segmentation [8, 2, 4] a fair number of landmarks are necessary in order to achieve sufficiently good or realistic models.

Given their representation the objects are aligned with respect to translation, rotation, and scale (e.g. by means of a Procrustes analysis [6]). Finally, the residual variation is decomposed into latent variables and a low dimensional representation is obtained by retaining only the most important of these. The decomposition of the variation has been based on a number of transformations, most importantly PCA [4, 2]. The use of Fourier modes and wavelets are also reported [17, 14].

By representing the shape coordinates of each observation as the rows of a data matrix, the PCA usually applied is an R-mode analysis of this matrix. In the statistical sense - in this analysis - the variables are the point coordinates and the observations are the training shapes. The estimated eigenvectors are used to deform the mean shape. If instead we make a Q-mode analysis then the variables are the (unordered) training shapes and the observations are the point coordinates. Solving in Q-mode provides us with an ordering of our “new observations”. In this case the deformation of the mean shape is introduced

not by the eigenvectors but by the transformed variables. This allow for application of the MAF transform. This analysis we denote the Q-MAF transformation. We illustrate this on a dataset consisting of 50 landmarks on metacarpal-2 (a bone in the hand) extracted from x-rays of the hands of 24 women.

## 2 Maximum autocorrelation factors (MAF)

Let the spatial covariance function of a multivariate stochastic variable,  $\mathbf{Z}_k$ , where  $k$  denotes spatial position and  $\Delta$  a spatial shift, be  $\Gamma(\Delta) = \text{Cov}\{\mathbf{Z}_k, \mathbf{Z}_{k+\Delta}\}$ . Evidently  $\Gamma^T(\Delta) = \Gamma(-\Delta)$ . Then by letting the variance-covariance matrix of  $\mathbf{Z}_k$  be  $\Sigma$  and defining the variance-covariance matrix  $\Sigma_\Delta = D\{\mathbf{Z}_k - \mathbf{Z}_{k+\Delta}\}$ ,  $D\{\cdot\}$  is the variance-covariance matrix of its argument, we find

$$\Sigma_\Delta = 2\Sigma - \Gamma(\Delta) - \Gamma(-\Delta) \quad (1)$$

We are now able to compute the covariance between a linear combination of the original variables and the shifted variables

$$\begin{aligned} \text{Cov}\{\mathbf{w}_i^T \mathbf{Z}_k, \mathbf{w}_i^T \mathbf{Z}_{k+\Delta}\} &= \mathbf{w}_i^T \Gamma(\Delta) \mathbf{w}_i = \mathbf{w}_i^T \Gamma^T(\Delta) \mathbf{w}_i \\ &= \frac{1}{2} \mathbf{w}_i^T (\Gamma(\Delta) + \Gamma(-\Delta)) \mathbf{w}_i \\ &= \mathbf{w}_i^T (\Sigma - \frac{1}{2} \Sigma_\Delta) \mathbf{w}_i, \end{aligned} \quad (2)$$

Thus the autocorrelation in shift  $\Delta$  of a linear combination of the mean-centered original variables,  $Z_k$ , is

$$\text{Corr}\{\mathbf{w}_i^T \mathbf{Z}_k, \mathbf{w}_i^T \mathbf{Z}_{k+\Delta}\} = 1 - \frac{1}{2} \frac{\mathbf{w}_i^T \Sigma_\Delta \mathbf{w}_i}{\mathbf{w}_i^T \Sigma \mathbf{w}_i}. \quad (3)$$

In order to minimize that correlation we must maximize the Rayleigh coefficient

$$R(\mathbf{w}) = \frac{\mathbf{w}^T \Sigma_\Delta \mathbf{w}}{\mathbf{w}^T \Sigma \mathbf{w}}. \quad (4)$$

The MAF transform is given by the set of conjugate eigenvectors of  $\Sigma_\Delta$  with respect to  $\Sigma$ ,  $W = [\mathbf{w}_1, \dots, \mathbf{w}_m]$ , corresponding to the eigenvalues  $\kappa_1 \leq \dots \leq \kappa_m$  [18]. The resulting new variables are ordered so that the first MAF is the linear combination that exhibits maximum autocorrelation. The  $i$ th MAF is the linear combination that exhibits the highest autocorrelation subject to it being uncorrelated to the previous MAFs. The autocorrelation of the  $i$ th component is  $1 - \frac{1}{2} \kappa_i$ .

Usually we assume first and second order stationarity of our data. So mean-centering consists of estimating and subtracting a global mean over all observations (pixels). In other situations other types of mean-centering may be applicable, e.g. no centering (for calibrated data), centering across variables (empirical orthogonal functions), double mean-centering.

One problem now arise, namely, how should we choose  $\Delta$ . Switzer suggested that for images we estimate  $\Sigma_{\Delta}$  for a horizontal shift in lag 1 and for a vertical shift in lag 1, followed by a pooling of these two covariances. Blind source separation by independent components analysis using the Molgedey-Schuster (MS-ICA) algorithm [13] is equivalent to MAF, as is shown in Appendix A. The purpose of this algorithm is to separate independent signals from linear mixings. MS-ICA does this by exploiting differences in autocorrelation structure between the independent signals. Kolenda et al. [12] use an iterative procedure for identifying the optimal lags based on the sum of pairwise absolute differences between the autocorrelations of the estimated independent components. In this study we use Switzer's original suggestion. This is based on the assumption that the noise is separated from the interesting latent variables in terms of autocorrelation already in lag 1.

An additional problem will arise when the number of training examples is less than the dimensionality of the problem. Then the variance-covariance matrix in the nominator of Eq. 4 is not positive definite. In this case the optimization must be carried out in the subspace spanned by the eigenvectors corresponding to non-zero eigenvalues of this matrix. This may be achieved by use of generalized singular value decomposition algorithms [1].

## 2.1 MAF on multivariate imagery

We will illustrate the MAF transform on a multivariate remotely sensed dataset, where a 62 channel sun-reflectance spectrum is measured in each pixel of an image. The original data are shown in Fig. 1. The images are recorded using the GER II Imaging Spectrometer from the Geophysical Environmental Research Corp., New York. Using a rotating mirror a line is scanned perpendicularly to the flight direction of the aircraft. The second dimension is generated by the forward motion of the aircraft. Correction for roll, pitch, and yaw of the aircraft as measured by a gyroscope hardmounted on the scanner has been made.

The GER II sensor consists of 3 spectrometers. The spectral coverage and resolution of these are shown in Tab. 1. The second spectrometer records data in a wavelength where the water in the atmosphere absorbs most light, rendering these channels very noisy.

In Figs. 2 and 3 the principal components (PC) and maximum autocorrelation factors resulting from transformation of the original 62-dimensional observations are shown. It is evident that the PCs (shown rowwise with PC1 in the upper-left corner) do a poor job of separating the interesting signal from the noise. The MAF transform does this in a very satisfactory way.

The reason for this lies in the design criteria for the methods. Principal components analysis seeks linear combination that maximize variance. Whereas the MAF transform seeks maximum autocorrelation between neighbouring observations. Because the phenomena that we want to study in the images have spatial extent the latter method is better at extracting these. PC fails because some (interesting) components with high autocorrelation have smaller variance than some of the noise components.

Furthermore, we can see that the intrinsic dimensionality of this dataset is much less than the original 62 channels, say on the order of 10. For interpretation purposes we may compute the correlations between the new variables and the original channels. These are shown in Fig. 4 for the first 4 MAFs. We see that the first component is a mean of all spectral channel from spectrometer 1 (visual) and spectrometer 3 (near infrared). The second MAF is a contrast between the low number and the high number channels of spectrometer 1. MAF3 is a mean of all near infrared channels, etc.

### 3 Shape analysis

Biological shape analysis based on landmark data has developed rapidly in the past decade. Before entering into the matter of things, we shall define precisely what we mean by shape: shape is all the geometrical information that is left when translation, rotation, and isotropical scaling is filtered out, i.e. shape is the geometrical information up to a Euclidean similarity transformation. In the following we shall shortly present how a dataset is constructed and how it is decomposed into latent variables.

We will illustrate the proposed methods on a dataset that consists of annotations of the contour of 24 metacarpals, i.e. a bone in the human hand. An example is shown in Figure 5(a). The annotations are based on 2-D wrist radiographs of human hands. The annotations are prone to errors in the proximal and distal ends due to the bones being overlaid in the projection of the radiograph and thus difficult to discern.

#### 3.1 Alignment of shapes

Let there be given  $p$  training examples for a given shape class, and let each example be represented by a complex vector,  $\mathbf{u}_k$  of  $n$  landmark points in complex coordinates  $u_{kj} + iv_{kj}$ ,  $k = 1, \dots, p$  and  $j = 1, \dots, n$ . Without loss of generality we will assume that each training example is centered on  $(0, 0)$ , i.e.  $\sum_{j=1}^n u_{kj} + iv_{kj} = 0$ .

Following Goodall [6] the alignment problem in 2D consists of estimating pose parameters for each shape. Let the pose parameters be scale:  $\beta_k \in \mathbb{R}_+$ , rotation:  $\psi_k \in [0; 2\pi[$ , and translation:  $a_k + ib_k \in \mathbb{C}$ . For each example the alignment of an example,  $\mathbf{u}_k$ , to the (yet unknown) mean shape,  $\boldsymbol{\mu}$ , consists of choosing  $(\hat{\beta}_k, \hat{\psi}_k, \hat{a}_k, \hat{b}_k)$  to minimize the least squares objective

$$\left\| \frac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|} - \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|} \beta_k e^{i\psi_k} - a - ib \right\| \quad (5)$$

The minimizing parameters are  $\hat{\beta}_k = |\boldsymbol{\mu}^* \mathbf{u}_k| / (\mathbf{u}_k^* \mathbf{u}_k)$ ,  $\hat{\psi}_k = -\arg(\boldsymbol{\mu}^* \mathbf{u}_k)$ , and  $\hat{a}_k = \hat{b}_k = 0$ , yielding the full Procrustes distance

$$d_F(\mathbf{u}_k, \boldsymbol{\mu}) = \inf_{\beta_k, \psi_k, a_k, b_k} \left\| \frac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|} - \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|} \beta_k e^{i\psi_k} - a - ib \right\| = \left\{ 1 - \frac{\boldsymbol{\mu}^* \mathbf{u}_k \mathbf{u}_k^* \boldsymbol{\mu}}{\mathbf{u}_k^* \mathbf{u}_k \boldsymbol{\mu}^* \boldsymbol{\mu}} \right\}^{1/2}. \quad (6)$$

\* denote complex conjugation and transposition. Note, that due to the standardization,  $d_F$  is symmetric in  $\boldsymbol{\mu}$  and  $\mathbf{u}_k$  rendering it a suitable measure of shape distance.

The full Procrustes estimate of the mean,  $\hat{\boldsymbol{\mu}}$ , is obtained by minimizing the sum of square full Procrustes distances from each example shape to an unknown unit size mean configuration.

$$\sum_{k=1}^p d_F^2(\mathbf{u}_k, \boldsymbol{\mu}) = \sum_{k=1}^p \left\{ 1 - \frac{\boldsymbol{\mu}^* \mathbf{u}_k \mathbf{u}_k^* \boldsymbol{\mu}}{\mathbf{u}_k^* \mathbf{u}_k \boldsymbol{\mu}^* \boldsymbol{\mu}} \right\} = n - \frac{\boldsymbol{\mu}^* S \boldsymbol{\mu}}{\boldsymbol{\mu}^* \boldsymbol{\mu}}, \text{ where } S = \sum_{k=1}^p \frac{\mathbf{u}_k \mathbf{u}_k^*}{\mathbf{u}_k^* \mathbf{u}_k}, \quad (7)$$

i.e. the complex sum of squares and products matrix. Hence,  $\hat{\boldsymbol{\mu}}$  is given by the eigenvector corresponding to the largest eigenvalue of  $S$ . All rotations of  $\hat{\boldsymbol{\mu}}$  in the complex plane are also solutions. If the example shapes are annotated in the same coordinate system a reasonable choice of rotation would be  $-\arg(\sum_{k=1}^p \hat{\beta}_k e^{i\psi_k})$ .

The Procrustes fitted shape example live on a Riemannian manifold called shape space. This shape space is approximated by a tangent plane at the mean shape. In this way we can use the Euclidean distances of the tangent space to approximate the Procrustes distances in shape space. This will allow to use standard multivariate analysis on the tangent space coordinates. The tangent space is usually a very good approximation to shape space.

The tangent space coordinates are the projections of the full Procrustes coordinates into the tangent plane to the shape space at the full Procrustes mean. Let the tangent space coordinates with the origin placed at the pole of the tangent space for each training example be

$$\mathbf{x}_k = (x_{k1}, \dots, x_{kn}, y_{k1}, \dots, y_{kn})^T. \quad (8)$$

Note, that the Procrustes analysis filters out translations of the observations, thus mean-centering the variables. When we subsequently analyse the deviations from the estimated mean shape we additionally mean-center across observations.

The aligned metacarpal dataset as well as the estimated mean shape is shown in Fig. 5(b). The mean shape is shown as an interpolated curve. The points of the aligned shapes are shown as scatter on this mean shape. The variation in this scatter in what we try to decompose into (a few) latent variables.

### 3.2 Shape decomposition using Principal Components

Cootes et al. [4] describe how a low dimensional model of shape variability can be obtained by decomposition using principal components. They call the resulting model an point distribution model (PDM)

Let the tangent space coordinates of the training examples be organised in a  $p \times 2n$  data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_p^T \end{bmatrix}$$

This matrix may then be decomposed using Eckart-Young's theorem [10]

$$\mathbf{X} = \mathbf{V} \boldsymbol{\Lambda} \mathbf{U}^T.$$

Where  $\mathbf{U}$  ( $2n \times r$ ) and  $\mathbf{V}$  ( $p \times r$ ) are orthogonal matrices, and  $\mathbf{\Lambda}$  ( $r \times r$ ) is a diagonal matrix with positive diagonal elements. The diagonal elements of  $\mathbf{\Lambda}$  are called the singular values of  $\mathbf{X}$ . This decomposition is also called the singular value decomposition (SVD).

By direct calculation using Eckart-Young's theorem we have the following two eigenvalue decompositions

$$\begin{aligned}\mathbf{X}\mathbf{X}^T &= \mathbf{V}\mathbf{\Lambda}^2\mathbf{V}^T \\ \mathbf{X}^T\mathbf{X} &= \mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^T\end{aligned}$$

The diagonal elements of  $\mathbf{\Lambda}^2$  are the squared diagonal elements of  $\mathbf{\Lambda}$  and these are the positive eigenvalues of  $\mathbf{X}\mathbf{X}^T$  and  $\mathbf{X}^T\mathbf{X}$ . The analysis of  $\mathbf{X}^T\mathbf{X}$  is called a R-mode analysis, and the analysis of  $\mathbf{X}\mathbf{X}^T$  is called a Q-mode analysis. The relation between the eigenvectors corresponding to the positive eigenvalues for the two problems are given by

$$\begin{aligned}\mathbf{V} &= \mathbf{X}\mathbf{U}\mathbf{\Lambda}^{-1} \\ \mathbf{U} &= \mathbf{X}^T\mathbf{V}\mathbf{\Lambda}^{-1}\end{aligned}\tag{9}$$

The estimated variance-covariance matrix of the tangent space coordinates of the training examples in Eq. (8) is

$$\hat{\mathbf{\Sigma}} = \frac{1}{p-1}\mathbf{X}^T\mathbf{X}\tag{10}$$

the eigenvectors (i.e. the principal components) of which are given by the columns of  $\mathbf{U}$ .

The PDM model then consists of retaining the  $t \leq r$  first principal components. Deviations from the Procrustes mean (in tangent space) can then be modelled by

$$\mathbf{x} = \mathbf{U}'\mathbf{b}$$

where  $\mathbf{U}'$  is a matrix consisting of the first  $t$  columns of  $\mathbf{U}$ , and  $\mathbf{b}$  defines a set of  $t$  parameters of the deformable model.

However, from Eq. (9) we see that by solving the problem in Q-mode, i.e. solve for  $\mathbf{V}$  we could generate the same PDM by

$$\mathbf{x} = \mathbf{X}^T\mathbf{V}'\mathbf{b}\tag{11}$$

where  $\mathbf{V}'$  is a matrix consisting of the first  $t$  columns of  $\mathbf{V}$ .

Solving the problem in Q-mode corresponds to an eigenvalue decomposition of the covariance matrix of a stochastic variable,  $\mathbf{Z}$ , examples of which are given by the coordinates of each point across the shape training examples, i.e.

$$\begin{aligned}\mathbf{z}_j &= (x_{1j}, \dots, x_{pj}), \quad \text{for } j = 1, \dots, n \\ \mathbf{z}_j &= (y_{1j}, \dots, y_{pj}), \quad \text{for } j = n + 1, \dots, 2n.\end{aligned}\tag{12}$$

This matrix is given by

$$\hat{\mathbf{\Pi}} = \frac{1}{2n-2}\mathbf{X}\mathbf{X}^T\tag{13}$$

The eigenvectors (i.e. the principal components) of this matrix corresponding to the positive eigenvalues are given by the columns of  $\mathbf{V}$ .



### 3.3 Shape Q-MAF

In this section we will describe how to use the maximum autocorrelation factors (MAF) [18] transform instead of principal components for formulation of an PDM. The approach assumes an ordering of the landmarks, and is based on a the Q-mode analysis. In general, it should be applicable since an ordering of the landmark points is almost always present.

In the MAF analysis in Section 3.2 we substitute the matrix  $\hat{\Pi}$  from Eq. (13) for the variance-covariance matrix  $\Sigma$ . The estimate of the difference variance-covariance matrix is given by

$$\hat{\Sigma}_{\Delta} = \frac{1}{2n-2} \mathbf{E} \mathbf{E}^T \quad (14)$$

where

$$\mathbf{E}^T = \begin{bmatrix} (\mathbf{z}_1 - \mathbf{z}_2)^T \\ \vdots \\ (\mathbf{z}_{n-1} - \mathbf{z}_n)^T \\ (\mathbf{z}_n - \mathbf{z}_1)^T \\ (\mathbf{z}_{n+1} - \mathbf{z}_{n+2})^T \\ \vdots \\ (\mathbf{z}_{2n-1} - \mathbf{z}_{2n})^T \\ (\mathbf{z}_{2n} - \mathbf{z}_{n+1})^T \end{bmatrix} \quad (15)$$

and  $\mathbf{z}_j$  is defined in Eq (12).

The MAF PDM is built by retaining the  $t \leq r$  first maximum autocorrelation factors. From Eq. (11) deviations from the Procrustes mean (in tangent space) is then modelled by

$$\mathbf{x} = \mathbf{X}^T \mathbf{W}' \mathbf{b} \quad (16)$$

where  $\mathbf{W}'$  is a matrix consisting of the first  $t$  columns of  $\mathbf{W}$ , and  $\mathbf{b}$  defines the set of  $t$  parameters of the model.

The Q-MAF transformation is performed on the metacarpal dataset and compared to ordinary PCA. The resulting eigen modes are shown in Fig. 6. In the figure all eigen modes for the two transformations are shown. The eigen modes are visualised as deviations from the (full Procrustes) mean shape. The mean shape and the deviations from the mean shape is shown for each eigen mode as  $\pm 5$  standard deviations across the training, respectively. Note that if we assume that the variation from the mean value across the training set can be modelled by a Gaussian distribution, then we would expect (almost) all deviations to be within  $\pm 3$  standard deviation.

From the plots in Fig. 6 we see that the Q-MAF transformation results in different eigen modes from the PCA transformation. The Q-MAF modes constitute a decomposition of (localized) spatial frequency along the contour with frequency increasing with mode number. Furthermore, the first two modes are easily interpreted as aspect ratio (the metacarpals have been scaled to the same size) of the cortical bone, mode three as bending, and mode four as thickness of the proximal (top) end. In the high order number modes variations composed of neighbouring points deforming in opposite directions are concentrated.

The PCA eigen modes are less easily interpreted and it seems that many low number modes are devoted to descriptions of variations of the proximal end. These are variations that may well stem from annotation arbitrariness.

## 4 Discussion

We have demonstrated the power of the maximum autocorrelation factor analysis over principal components for datasets where the observations are organised in some structure. Furthermore, we have devised how a new transformation – the Q-MAF transformation – can utilize that the variables of multivariate problem are organised in some structure. It has been demonstrated how this new transformation when applied to biological shape analyses can ease interpretation. It is suspected that the Q-MAF procedure is better at isolating effects of the (stochastic) annotation process in high number latent variables than the featured principal components. Finally, we have shown how apply the Q-MAF transformation to datasets consisting of spectra from samples. In addition we shown that the Molgedey-Schusters algorithm for performing independent components analysis is equivalent to the maximum autocorrelation factor transformation.

## 5 Acknowledgements

We thank Hans Henrik Thodberg, Ph.D., Pronosco A/S for annotations of the metacarpal dataset.

## A Equivalence of ICA and MAF

It turns out that Molgedey-Schusters algorithm for performing ICA [13] is the same as the MAF analysis [18].

Assuming the linear mixing model of independent components analysis  $\mathbf{X} = \mathbf{A}\mathbf{S}$ , where  $\mathbf{X}$  is the  $(P \times N)$  data matrix with each row consituting a signal,  $\mathbf{S}$  is a matrix of the same form as  $\mathbf{X}$  containing independent signals in the rows, and  $\mathbf{A}$  is a linear mixing matrix. Furthermore, let  $\mathbf{X}_\Delta$  and  $\mathbf{S}_\Delta$  be  $\mathbf{X}$  and  $\mathbf{S}$  cyclicly shifted  $\Delta$  steps rowwise. Then the solution is found by forming

$$\mathbf{Q} = \frac{1}{2} [\mathbf{X}_\Delta \mathbf{X}^T + \mathbf{X} \mathbf{X}_\Delta^T] (\mathbf{X} \mathbf{X}^T)^{-1} = \mathbf{A} \left[ \frac{1}{2} (\mathbf{S}_\Delta \mathbf{S}^T + \mathbf{S} \mathbf{S}_\Delta^T) (\mathbf{S} \mathbf{S}^T)^{-1} \right] \mathbf{A}^{-1} \quad (17)$$

Due to the independence of the source signals the latter bracketed parenthesis is diagonal. Therefore the mixing matrix can be determined by an eigenvalue decomposition of the matrix  $\mathbf{Q}$ , and the source signals up to a scale factor are estimated by  $\mathbf{S} = \mathbf{A}^{-1} \mathbf{X}$ . An

estimator for the crosscovariance function for a shift  $\Delta$  is  $\frac{1}{N} \mathbf{X} \mathbf{X}_{\Delta}^T$ , and an estimator for the covariance matrix,  $\Sigma$ , is  $\frac{1}{N} \mathbf{X} \mathbf{X}^T$ . Therefore using Eq. (1)

$$\mathbf{Q} = \frac{1}{2} [2\Sigma - \Sigma_{\Delta}] \Sigma^{-1} = \left[ \mathbf{I} - \frac{1}{2} \Sigma_{\Delta} \Sigma^{-1} \right]$$

The unity matrix  $\mathbf{I}$  has no effect on the eigenvectors, so  $\mathbf{A}$  simply consists of the conjugate eigenvectors of  $\Sigma_{\Delta}$  with respect to  $\Sigma$ , i.e. the MAF problem given in Eq. (3).

It is easily shown that the MAF transform is invariant to affine transformations. Therefore we may execute a prewhitening beforehand, thus obtaining  $\Sigma = \mathbf{I}$ . Then  $\mathbf{Q}$  becomes symmetric yielding  $\mathbf{A}^{-1} = \mathbf{A}^T$ , and the MAF factors become  $\mathbf{A}^T \mathbf{X}$  prewhitened, i.e. the independent components.

## References

- [1] E. Anderson, Z. Bai, C. Bischoff, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorenson. *LAPACK Users Guide*. SIAM, Philadelphia, 2 edition, 1995.
- [2] P. R. Andresen, F. L. Bookstein, K. Conradsen, B. K. Ersbøll, J. L. Marsh, and S. Kreiborg. Surface-bounded growth modeling applied to human mandibles. *IEEE Transactions on Medical Imaging*, 19(11), Nov. 2000. 1053–1063.
- [3] K. Conradsen, B. K. Nielsen, and T. Thyrted. A comparison of min/max autocorrelation factor analysis and ordinary factor analysis. In *Proceedings from Symposium in Applied Statistics*, pages 47–56, Lyngby, Jan. 1985.
- [4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision, Graphics and Image Processing*, 61(1):38–59, Jan. 1995.
- [5] I. L. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, Chichester, 1998. xx + 347 pp.
- [6] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, Series B*, 53(2):285–339, 1991.
- [7] A. A. Green, M. Berman, P. Switzer, and M. D. Craig. 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, Jan. 1988.
- [8] D. C. Hogg, N. Johnson, R. Morris, D. Buesching, and A. Galata. Visual modes of interaction. In *2nd International Workshop on Cooperative Distributed Vision, Kyoto, Japan*, 1998.
- [9] H. Hotelling. Relations between two sets of variables. *Biometrika*, 28:321–377, 1936.

- [10] R. M. Johnson. On a theorem stated by Eckart and Young. *Psychometrika*, 28:259–263, 1963.
- [11] H. Knutsson, M. Andersson, M. Borga, and J. Wiklund. Automated generations of representations in vision. In A. Sanfeliu, J. J. Villanueva, M. Vanrell, R. Alquézar, T. Huang, and J. Serra, editors, *Proceedings of the 15th International Conference on Pattern Recognition*, volume 3, pages 63–70, Los Alamitos, California, USA, 2000. IEEE Computer Society.
- [12] T. Kolenda, L. K. Hansen, and J. Larsen. Signal detection using ica: Application to chat room topic spotting. In Lee, Jung, Makeig, and Sejnowski, editors, *Proceedings of the 3rd International Conference on Independent Components Analysis and Blind Signal Separation (ICA'2001), San Diego, USA, December 9-13, 2001*.
- [13] L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72(23):3634–3637, 1994.
- [14] A. Neuman and C. Lorenz. Statistical shape model based segmentation of medical images. *Computerized Medical Imaging and Graphics*, 22:133–143, 1998.
- [15] 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 Sensing of Environment*, 19:1–19, 1998.
- [16] A. A. Nielsen and R. Larsen. Restoration of GERIS data using the maximum noise fractions transform. In *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition*, volume 2, pages 557–568, Strasbourg, France, 1994.
- [17] L. H. Staib and J. S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1061–1075, 1992.
- [18] P. Switzer. Min/max autocorrelation factors for multivariate spatial imagery. In L. Billard, editor, *Computer Science and Statistics*, pages 13–16. Elsevier Science Publishers B.V. (North Holland), 1985.

Table 1: Spectral coverage and resolution of the GER II sensor

	Spectral coverage $\mu\text{m}$	Number of channels	Sampling interval nm
Spectrometer I	0.47-0.84	30	12.3
Spectrometer II	1.40-1.90	4	120
Spectrometer III	2.00-2.45	28	16.2

Fig. 1: Sun-reflectance in 62 spectral channels captured by the GER II Imaging Spectrometer. The spectral channels are shown rowwise

Fig. 2: Principal components shown rowwise with PC1 in the upper left corner.

Fig. 3: Maximum autocorrelation factors shown rowwise with MAF1 in the upper left corner.

Fig. 4: Correlations between the first 4 MAFs and the original variables. On the left in each subplot the autocorrelation of the MAF components is shown.

Fig. 5: (a) Metacarpal annotation using 50 landmarks. (b) The metacarpal dataset aligned and projected into tangent space.

Fig. 6: The eigen modes for the shape Q-MAF and the PCA transformation are visualised as deviations from the (full Procrustes) mean shape. The mean shape is drawn in blue. The deviation from the mean shape is shown for each eigen mode as  $\pm 5$  standard deviations across the training set as a red and a green curve, respectively.

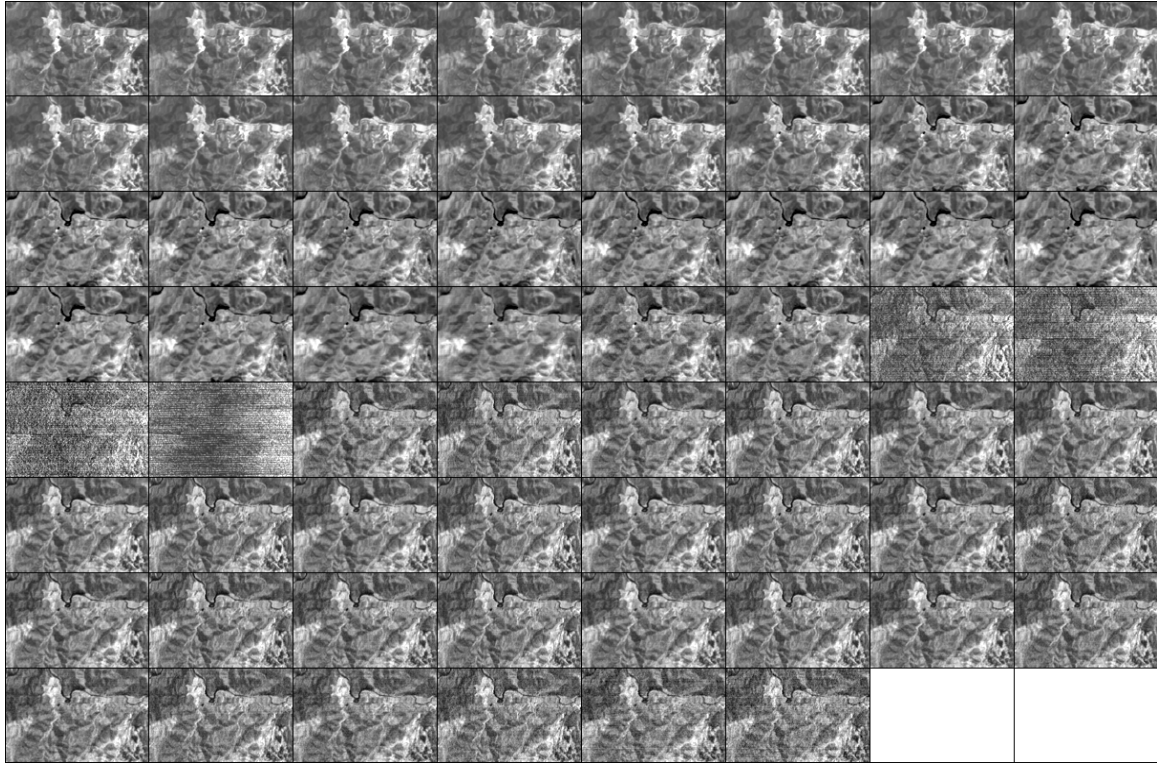


Fig. 1:

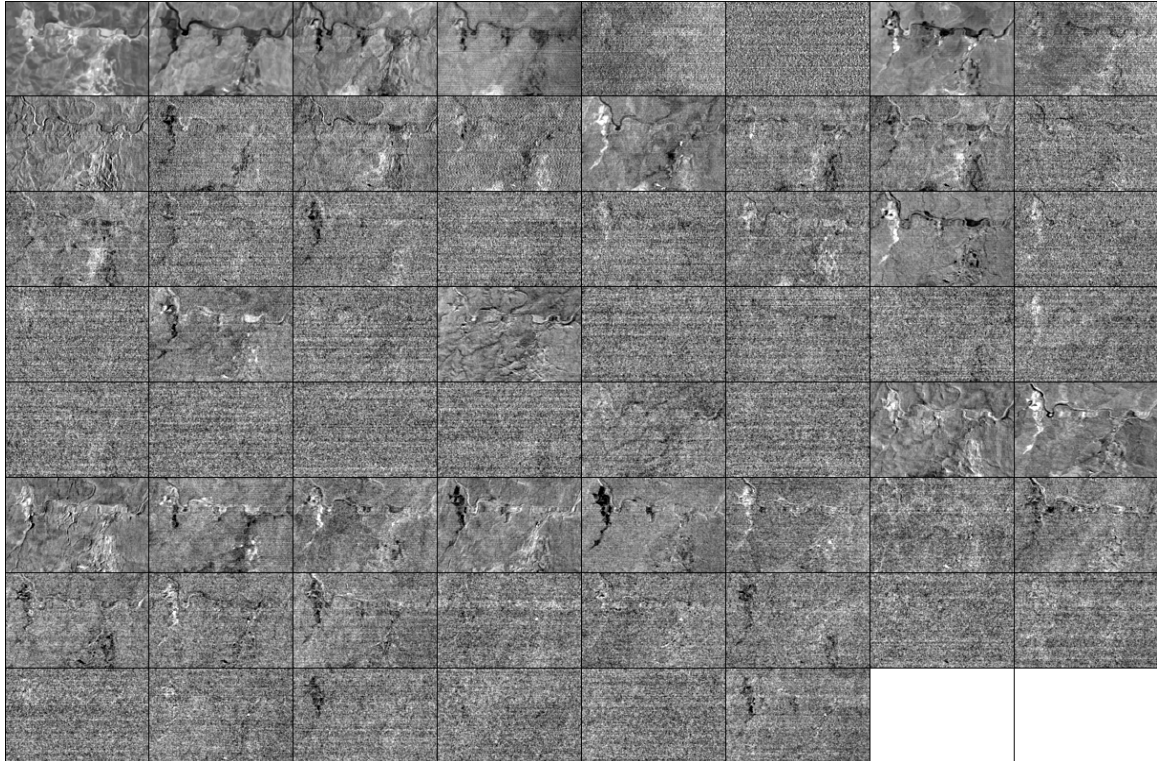


Fig. 2:



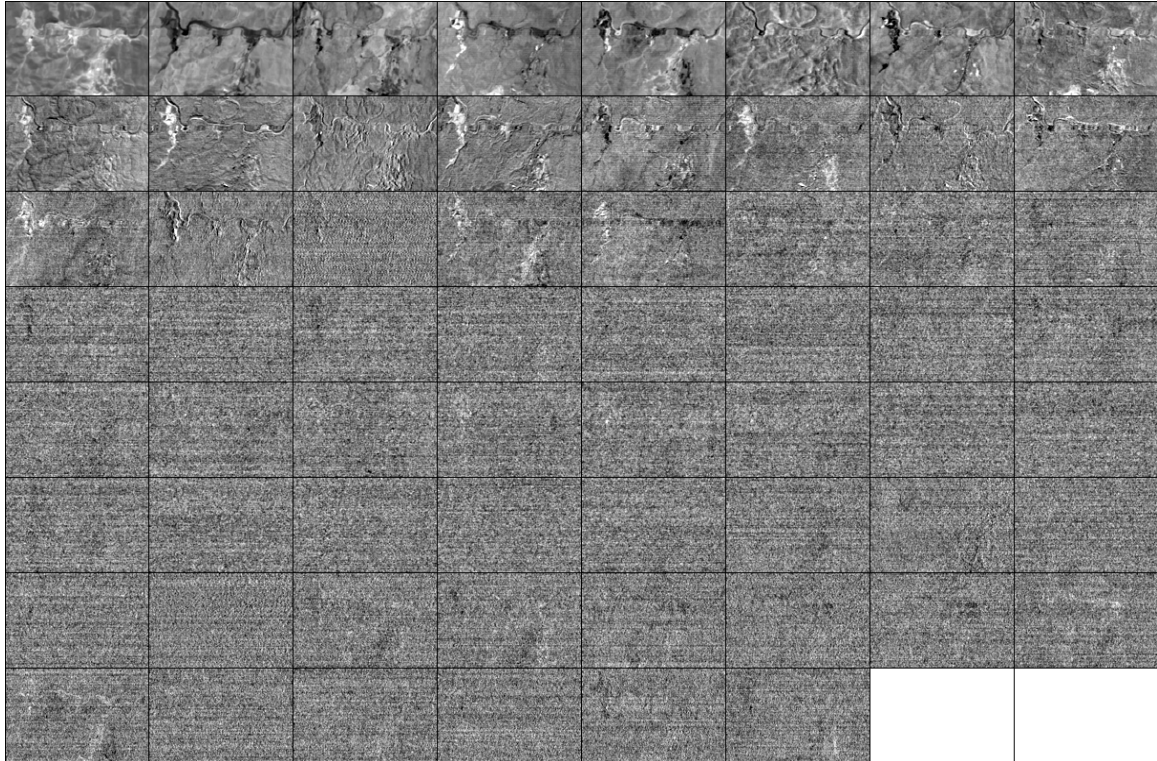


Fig. 3:

### Maximum Autocorrelation Factor Analysis Correlations Between MAFs and Bands

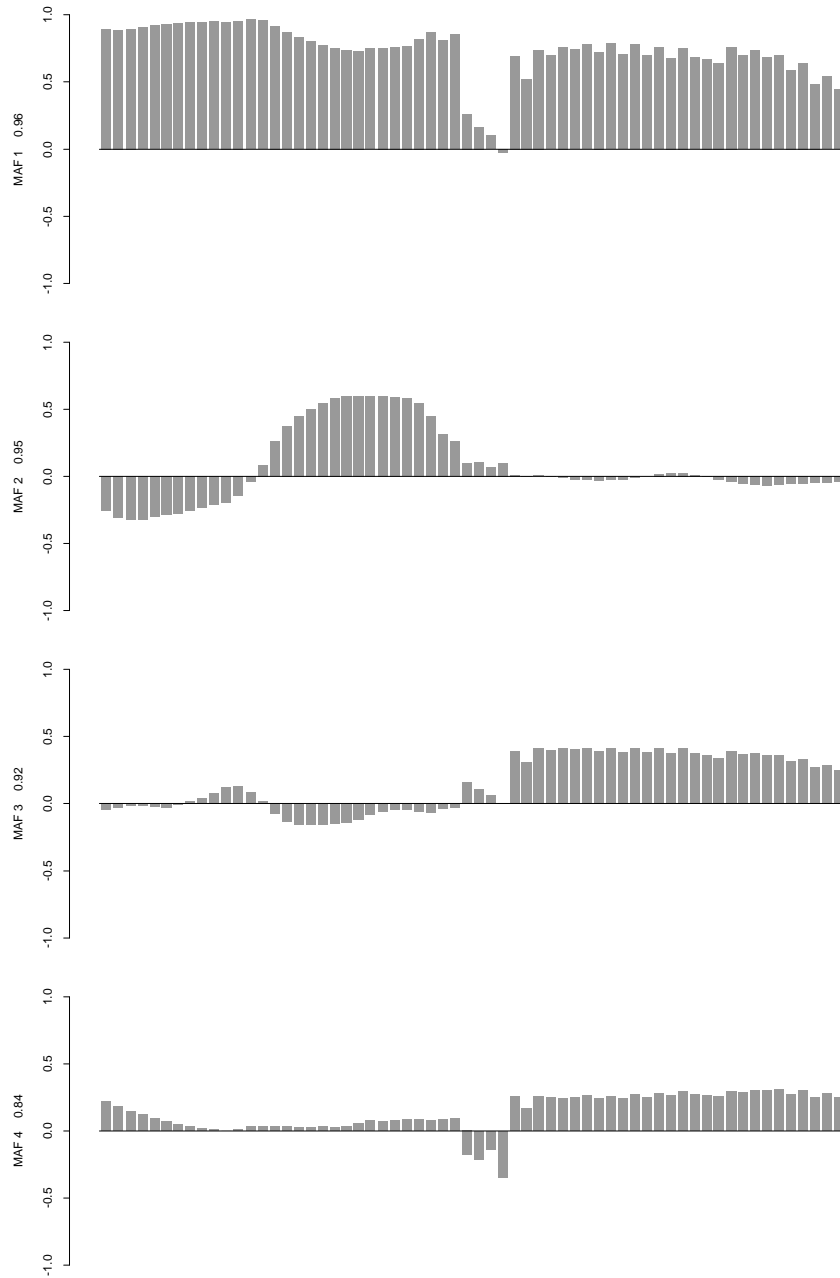
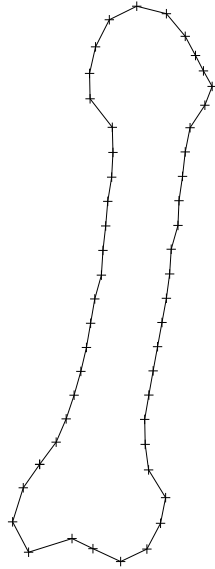
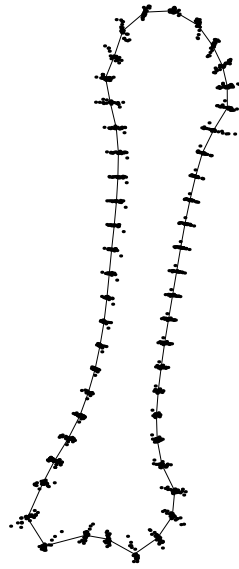


Fig. 4:



(a)



(b)

Fig. 5:

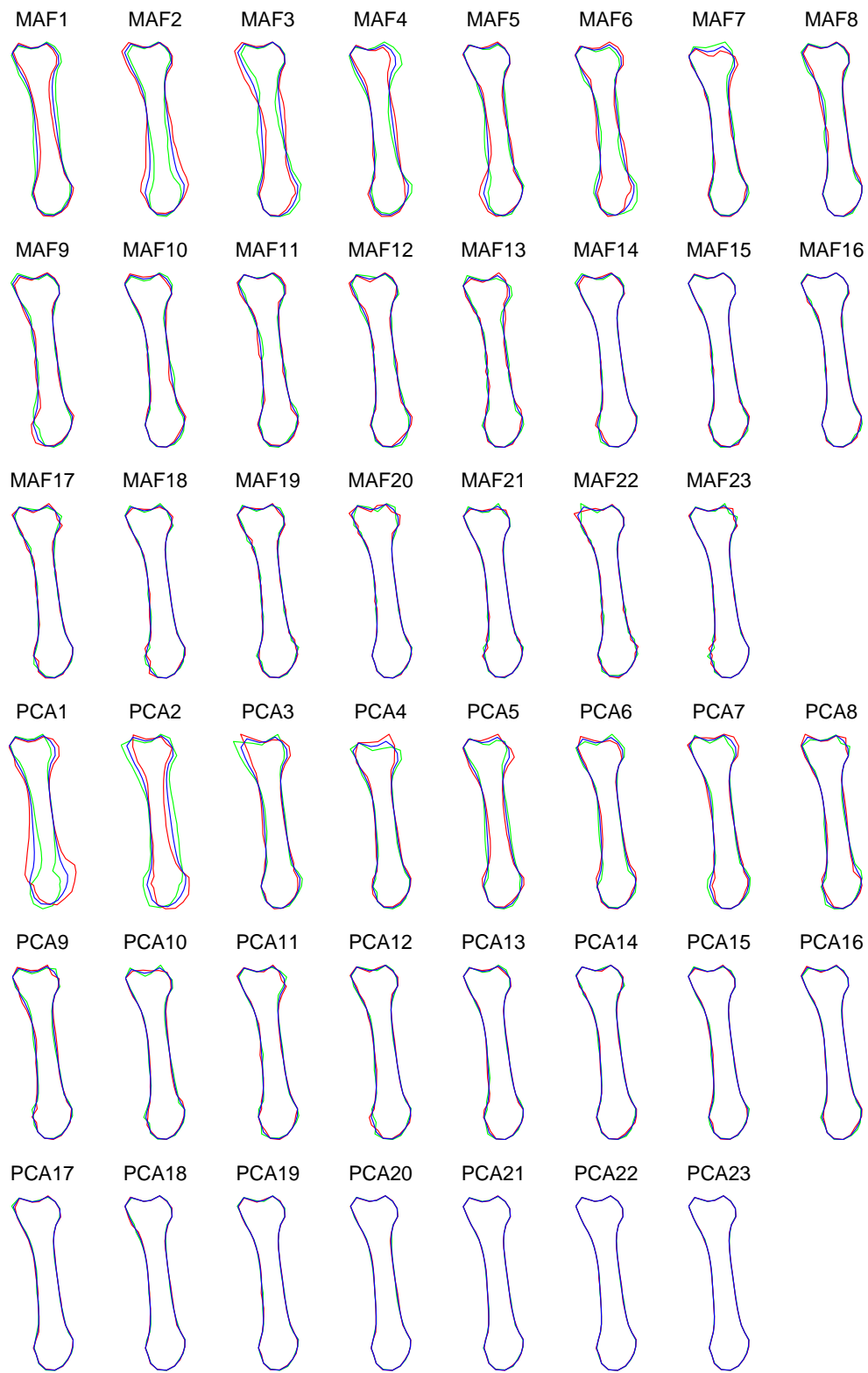


Fig. 6: