# UCLA
## UCLA Previously Published Works

**Title**

Improved approximation and visualization of the correlation matrix.

**Permalink**

**Journal**

The American Statistician, 77(4)

**ISSN**

0003-1305

**Authors**

Graffelman, Jan
De Leeuw, Jan

**Publication Date**

2023

**DOI**

10.1080/00031305.2023.2186952

Peer reviewed

# Improved approximation and visualization of the correlation matrix

**Jan Graffelman**[a,b,*], **Jan de Leeuw**[c]

[a]Department of Statistics and Operations Research, Universitat Politècnica de Catalunya

[b]Department of Biostatistics, University of Washington

[c]Department of Statistics, University of California Los Angeles

## Abstract

The graphical representation of the correlation matrix by means of different multivariate statistical methods is reviewed, a comparison of the different procedures is presented with the use of an example data set, and an improved representation with better fit is proposed. Principal component analysis is widely used for making pictures of correlation structure, though as shown a weighted alternating least squares approach that avoids the fitting of the diagonal of the correlation matrix outperforms both principal component analysis and principal factor analysis in approximating a correlation matrix. Weighted alternating least squares is a very strong competitor for principal component analysis, in particular if the correlation matrix is the focus of the study, because it improves the representation of the correlation matrix, often at the expense of only a minor percentage of explained variance for the original data matrix, if the latter is mapped onto the correlation biplot by regression. In this article, we propose to combine weighted alternating least squares with an additive adjustment of the correlation matrix, and this is seen to lead to further improved approximation of the correlation matrix.

## Keywords

---

*Corresponding author: Jan Graffelman, jan.graffelman@upc.edu.

## 1 Introduction

The correlation matrix is of fundamental interest in many scientific studies that involve multiple variables, and the visualization of the correlation coefficient and the correlation matrix has been the focus of several studies. Hills (1969) proposed multidimensional scaling (MDS), using distances to approximate correlations. Rodgers and Nicewander (1988) review multiple formulas for the correlation coeffient, showing visualizations that use slopes, angles and ellipses. Murdoch (1996) proposed to visualize correlations using a table of elliptical glyphs. Friendly (2002) proposed *corrgrams* that use color and shading of tabular displays to represent the entries of a correlation matrix. Trosset (2005) developed the *correlogram*, which capitalizes on approximation of correlations by cosines. Obviously, a correlation matrix can be visualized in multiple ways, using different geometric principles. In the statistical environment R ( R Core Team 2022), visualizations by means of vector diagrams or biplots can be obtained using the R packages `FactoMineR` ( Lê et al. 2008) and `factoextra` ( Kassambara and Mundt 2020; Kassambara 2017); corrgrams can be made with the R packages `corrgram` ( Wright 2021) and `corrplot` ( Wei and Simko 2021). The visualization of the correlation matrix by means of a principal component analysis (PCA) is facilitated by many statistical programs, among them the R packages FactoMineR and factoextra. Figure 1 shows two popular pictures of the correlation matrix of the myocardial infarction or Heart attack data ( Saporta 1990), a colored tabular display or corrgram (Fig. 1A) and correlation circle (i.e., a correlation biplot, Fig. 1B) obtained by PCA.

Both plots reveal two groups of positively correlated variables (( *CI, SI*) and ( *Pulse, DBP, PA, VP, logPR*)) with negative correlations between the groups. In this article we focus on the visualization of correlations by means of vector and scatter diagrams, refraining from coloured tabular representations as in Figure 1A. Despite the popularity of the correlation circle, from a statistical point of view, PCA gives a suboptimal approximation of the correlation matrix. PCA provides a least-squares low-rank approximation to the *data* matrix (centred or standardized), and the visualization of the correlation matrix can be seen as a by-product of the analysis, but not its main goal. Principal factor analysis (PFA) and correlograms ( Trosset 2005) are multivariate methods more specifically designed for approximation and visualization of the correlation matrix. The main point of this article is to emphasize and illustrate the improvements offered by these and other methods and to stimulate their use over just using standard PCA for representing correlations. Also, as argued in the Discussion section, the correlation matrix requires goodness-of-fit measures that are different from ones shown in Figure 1B.

## 2 Materials and methods

In this section we briefly summarize methods for the visualization of the correlation matrix using a well-known multivariate data set, the Heart attack data ( Saporta 1990, pp. 452-454). The original data consists of 101 observations of patients who suffered a heart attack, and for which seven variables are registered: the *Pulse*, the cardiac index ( *CI*), the systolic index ( *SI*), the diastolic blood pressure ( *DBP*), the pulmonary artery pressure ( *PA*), the ventricular pressure ( *VP*) and the pulmonary resistance ( *logPR*). Pulmonary resistance was log-transformed in order to linearize its relationship with the other

variables. We successively address the representation of correlations by principal component analysis (PCA), the correlogram (CRG), multidimensional scaling (MDS), principal factor analysis (PFA) and weighted alternating least squares (WALS). An additive adjustment to the correlation matrix is proposed to improve its visualization by PCA and WALS.

## 2.1   Principal component analysis

Principal component analysis is probably the most widely used method to display correlation structure by means of a vector diagram as given in Figure 1B. A correlation-based PCA can be performed by the singular value decomposition of the standardized data matrix ($\mathbf{X}_s$, scaled by $1 / \sqrt{n}$)

$$\frac{1}{\sqrt{n}}\mathbf{X}_s = \mathbf{U}\mathbf{D}_s\mathbf{V}', \tag{1}$$

where the left singular vectors are identical to the standardized principal components, and the right singular vectors are eigenvectors of the sample correlation matrix $\mathbf{R}$ since

$$\mathbf{R} = \frac{1}{n}\mathbf{X}_s{}'\mathbf{X}_s = \mathbf{V}\mathbf{D}_s^2\mathbf{V}' = \mathbf{V}\mathbf{D}_\lambda\mathbf{V}', \tag{2}$$

where the eigenvalues of the correlation matrix are the squares of the singular values. The vectors (arrows) in the correlation circle are given by $\mathbf{G} = \mathbf{V}\mathbf{D}_s$, and represent the entries of the eigenvectors scaled by the singular values. A well-known property of this vector diagram is that cosines of angles $\theta_{ij}$ between vectors approximate correlations, as from $\mathbf{G}\mathbf{G}' = \mathbf{R}$ it follows that

$$cos(\theta_{ij}) = \frac{\mathbf{g}_i{}'\mathbf{g}_j}{\|\mathbf{g}_i\|\|\mathbf{g}_j\|} \approx r_{ij}, \tag{3}$$

where $\mathbf{g}_i$ is the $i$th row of $\mathbf{G}$. This equation holds true exactly in the full space when using all eigenvectors, but only approximately so if a subset of the first few (typically two) is used. Alternatively, one can use a PCA biplot ( Gabriel 1971) to approximate the correlation matrix. We define a biplot as a joint display of the rows and the columns of a matrix that is optimal in a (weighted) least squares sense. In biplots it is common practice to use scalar products to approximate the entries of a data matrix of interest; the entries of the matrix are approximated by the length of the projection of one vector onto another, multiplied by the length of the vector projected upon. The PCA biplot of the *data* matrix (centred or standardized), with observations represented by dots and variables by vectors, is most well-known, though PCA also allows the approximation of the correlations by using scalar products between vectors. In the case of a correlation matrix, we have, using Eq. (2),

$$r_{ij} \approx \mathbf{g}_i{}'\mathbf{g}_j = cos(\theta_{ij})\|\mathbf{g}_i\|\|\mathbf{g}_j\| = \|\mathbf{p}_i\|\|\mathbf{g}_j\|, \tag{4}$$

where $\mathbf{p}_i$ is the projection of $\mathbf{g}_i$ onto $\mathbf{g}_j$. Biplots have been developed for all classical multivariate methods, and several textbooks describe biplot theory and provide many examples ( Gower and Hand 1996; Yan and Kang 2003; Greenacre 2010; Gower et al.

2011). A goodness-of-fit measure, based on least-squares, of the correlation matrix is given by

$$\frac{\text{tr}(\mathbf{\tilde{R}}\,,\mathbf{R})}{\text{tr}(\mathbf{R}\,,\mathbf{R})} = \frac{\lambda_1^2 + \lambda_2^2}{\sum_{i=1}^{p} \lambda_i^2}, \tag{5}$$

where $\mathbf{\tilde{R}}$ is the rank two approximation obtained from Eq. (2) by using two eigenvectors only. We note that this measure is based on the *squares* of the eigenvalues, as detailed in the seminal paper by Gabriel (1971), whereas the eigenvalues themselves are used to calculate the goodness-of-fit of the centered data matrix (See Section 3).

## 2.2 The correlogram

The correlogram, proposed by Trosset ( 2005), explicitly optimizes the approximation of correlations in a two or three dimensional subspace by cosines, by minimizing the loss function

$$\sigma(\mathbf{\theta}) = \|\mathbf{R} - \mathbf{C}(\mathbf{\theta})\|^2 \text{ with } \mathbf{C}(\mathbf{\theta})_{jk} = \cos(\theta_j - \theta_k), \tag{6}$$

where $\mathbf{\theta} = (0, \theta_2, \ldots, \theta_p)$ is a vector of angles with respect to the $x$ axis, one for each variable, the first variable being represented by the $x$ axis itself. Equation (6) can be minimized numerically, using R's function `nlminb` of the standard R package `stats` ( R Core Team 2022). In a correlogram, vector length is not used in the interpretation, and all variables are therefore represented by vectors that emanate from the origin, and that have unit length, falling all on a unit circle (see Figure 2B). A linearized version of the correlogram was proposed by Graffelman (2013).

## 2.3 Multidimensional scaling

Hills (1969) proposed to represent correlations by distances using MDS, and suggested to transform correlations to distances by using the transformation $d_{ij} = 2(1 - r_{ij})$, after which they are used as input for classical metric multidimensional scaling ( Mardia et al. 1979, Chapter 14), also known as principal coordinate analysis (PCO; Gower (1966)). As a historical note, in order to reproduces Hills' result, one actually needs to use the transformation $d_{ij} = \sqrt{2(1 - r_{ij})}$, implying Hills' article referred to the *squared* distances. Importantly, with this transformation the relationship between correlation and distance is ultimately non-linear. Using this distance, tightly positively correlated variables will be close ($d_{ij} \approx 0$), and tightly negatively correlated variables will be remote ($d_{ij} \approx 2$), whereas uncorrelated variables will appear at intermediate distance ($d_{ij} \approx \sqrt{2}$). Obviously, the diagonal of ones of the correlation matrix will always be perfectly fitted with this approach. In MDS, goodness-of-fit is usually assessed by looking at the eigenvalues. However, in this case the squared eigenvalues will be indicative of the goodness-of-fit of the *double-centered* correlation matrix, not of the original correlation matrix. In order to assess goodness-of-fit in terms of the root mean squared error (RMSE), as we will do for other methods, we will use the distances fitted by MDS (in two dimensions), and backtransform these to obtain fitted correlations in order to calculate the RMSE. A classical metric MDS of

correlations transformed to distances by Hills' transformation is equivalent to the spectral decomposition of the double-centered correlation matrix, which can be obtained from the ordinary correlation matrix by centering columns and rows with a centering matrix $\mathbf{H} = \mathbf{I} - (1 \ / \ p)\mathbf{1}\mathbf{1}'$. For the double-centered correlation matrix $\mathbf{R}_{dc}$, we have that

$$\mathbf{R}_{dc} = \mathbf{HRH} = (1 \ / \ n)\mathbf{HX}_{s}{}'\mathbf{X}_{s}\mathbf{H} = \mathbf{Z}'\mathbf{Z}, \tag{7}$$

with $\mathbf{Z} = (1 \ / \ \sqrt{n})\mathbf{X}_{s}\mathbf{H}$. It follows that $\mathbf{R}_{dc}$ is positive semidefinite, with rank no larger than $p - 1$, and consequently, a configuration of points whose interpoint distances exactly represent the correlation matrix in at most $p - 1$ dimensions can always be found ( Mardia et al. 1979, Section 14.2).

## 2.4 Principal factor analysis

The classical orthogonal factor model for a $p$-variate random vector $\mathbf{x}$, is given by $\mathbf{x} = \mathbf{Lf} + \mathbf{e}$, where $\mathbf{L}$ is the matrix of $p \times m$ factor loadings, $\mathbf{f}$ the vector with $m$ latent factors, and $\mathbf{e}$ a vector of errors. This model can be estimated in various ways. Currently, factor models are mostly fitted using maximum likelihood estimation, which also enables the comparison of different factor models by likelihood ratio tests.

PFA is an older iterative algoritm for estimating the orthogal factor model ( Johnson and Wichern 2002; Harman 1976). It is based on the iterated spectral decomposition of the reduced correlation matrix, which is obtained by subtracting the specificities from the diagonal of the correlation matrix. A classical factor loading plot is in fact a biplot of the correlation matrix, since the factor model implicitly decomposes the correlation matrix as

$$\mathbf{R} = \mathbf{LL}' + \mathbf{\Psi}, \tag{8}$$

where $\mathbf{\Psi}$ is the diagonal matrix of specificities (variances not accounted for by the common factors). A low-rank approximation to the correlation matrix, say of rank two, is obtained by $\mathbf{R} = \mathbf{LL}'$ after estimating the two-factor model. This approximation is known to be better than the approximation offered by PCA, for it avoids the fitting of diagonal of the covariance or correlation matrix ( Satorra and Neudecker 1998).

## 2.5 Weighted alternating least squares

In general, a low-rank approximation for a rectangular matrix $\mathbf{X}$ can be found by weighted alternating least squares, by minimizing the loss function

$$\sigma(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{n} \sum_{j=1}^{p} w_{ij}(x_{ij} - \mathbf{a}_{i}{}'\mathbf{b}_{j})^{2}, \tag{9}$$

where $\mathbf{a}_{i}$ is the $i$th row of $\mathbf{A}$, $\mathbf{b}_{j}$ the $j$th row of $\mathbf{B}$, $\mathbf{W}$ a matrix of weights and where we seek the factorization $\mathbf{X} = \mathbf{AB}'$. The unweighted case ($w_{ij} = 1$) is solved by the singular value decomposition ( Eckart and Young 1936). Keller (1962) also addressed the unweighted case, and explicitly considered the symmetric case. Bailey & Gower (1990) considered the symmetric case with differential weighting of the diagonal. A general-purpose algorithm

for the weighted case based on iterated weighted regressions (''criss cross regressions'') was proposed by Gabriel and Zamir (1979); Gabriel (1978) also presented an application in the context of the approximation of a correlation matrix, where we have that $\mathbf{X} = \mathbf{R}$. Pietersz and Groenen (2004) present a majorization algorithm for minimizing Eq. (9) for the correlation matrix. The fit of the diagonal of the correlation matrix can be avoided by using a weight matrix $\mathbf{W} = \mathbf{J} - \mathbf{I}$, where $\mathbf{J}$ is a $p \times p$ matrix of ones, and $\mathbf{I}$ a $p \times p$ identity matrix. This weighting gives weight 1 to all off-diagonal correlations and effectively turns off its diagonal by assigning it zero weight. An efficient algorithm and R code for using WALS with a symmetric matrix has been developed by De Leeuw (2006). The WALS approach can outperform PFA, for not being subject to the restrictions of the factor model. Communalities in factor analysis cannot exceed 1, which implies that the rows of the matrix of loadings $\mathbf{L}$ are vectors that are constrained to be inside, or in the limit, on the unit circle. In practice, PFA and WALS give the same RMSE if all variable vectors in PFA fall inside the unit circle, but WALS achieves a lower RMSE whenever a variable vector reaches the unit circle in PFA. This typically happens when a variable, in terms of the factor model, has a communality of one, or equivalently, zero specifity, a condition known as a *Heywood case* in maximum likelihood factor analysis ( Johnson and Wichern 2002; Heywood 1931). In WALS, the length of the variable vectors is unconstrained, and vectors can obtain a length larger than one if that produces a better fit to the correlation matrix. Indeed, if a factor analysis produces a Heywood case, then this indicates that a representation of the correlation matrix by WALS that outperforms PFA is possible.

## 2.6 Weighted alternating least squares with an additive adjustment

We propose a modification of the WALS procedure in order to further improve the approximation of the correlation matrix. By default, all vector diagrams (i.e. biplots) of the correlation matrix have vectors that emanate from the origin, the latter representing zero correlation for all variables; the fitted plane is constrained to pass through the origin. This does generally not provide the best fit to the correlation matrix. We propose an *additive adjustment* to improve the fit of the correlation matrix. By using an additive adjustment $\delta$, the origin of the plot no longer represents zero correlation but a certain level of correlation. Consequently, the scalar products between vectors represent the deviation from this level. The optimal adjustment ($\delta$) and the corresponding factorization of the adjusted correlation matrix can be found simultaneously by minimizing the loss function

$$\sigma(\mathbf{A}, \mathbf{B}, \delta) = \sum_{i=1}^{n} \sum_{j=1}^{p} w_{ij} (x_{ij} - \delta - \mathbf{a}_{i}{}_{,}\mathbf{b}_{j})^2, \tag{10}$$

where the notation is again kept general (for rectangular $\mathbf{X}$; in this article $\mathbf{X} = \mathbf{R}$ and $n = p$). The adjustment amounts to subtracting an optimal constant $\delta$ from all entries of the correlation matrix, and factoring the so obtained adjusted correlation matrix $\mathbf{R}_a = \mathbf{R} - \delta\mathbf{J} = \mathbf{AB}'$. The minimization can be carried out using the R program `wAddPCA` program developed by de Leeuw (https://jansweb.netlify.app/), and included in the `Correlplot` package for the purpose of this article. For a correlation matrix, the minimization does, in general, yield $\mathbf{A} \neq \mathbf{B}$, though unique biplot vectors for each

variable are easily obtained by a posterior spectral decomposition: $\mathbf{AB}' = \mathbf{VDV}' = \mathbf{GG}'$ with $\mathbf{G} = \mathbf{VD}^{1/2}$. The weighted least-squares approximation to $\mathbf{R}$ is then given by $\delta\mathbf{J} + \mathbf{GG}'$, and the WALS biplot is made by plotting the first two columns of $\mathbf{G}$; the origin of that plot represents correlation $\delta$ (See Figure 3D).

We note that the additive adjustment is different from the usual column (or row) centering operation, employed by many multivariate methods like PCA, consisting of the subtraction of column means (or row means) from each column (or respectively, each row). It is also different from the double centering operation, used in MDS, that subtracts row and column means, but adds the overall mean. The adjusted correlation matrix preserves the property of symmetry. The additive adjustment can also be used in the unweighted approximation of the correlation matrix by the iterated spectral decomposition of $\mathbf{R}_a$, calculating $\delta$ on each iteration as $\delta = tr((\mathbf{R} - \mathbf{GG}')\mathbf{J}) / p^2$, $\mathbf{G}$ containing the scaled eigenvectors of $\mathbf{R}_a$ (analogously to Eq. 2) in which case it can improve the fit to the correlation matrix obtained by PCA (See Figure 3B for an example), though this will not solve PCA's problem of fitting the diagonal. It is thus most appealing to use the additive adjustment in the weighted approach.

## 3    Results

We present some examples of biplots of correlation matrices obtained by PCA and by using WALS with the δ adjustment. First, we successively apply all methods reviewed in the previous section to the Heart attack data, whose correlation matrix is given in Table 1, and compare them in terms of goodness-of-fit. Second, we provide some additional illustrative examples, using the Aircraft data ( Gower and Hand 1996), the Swiss banknote data ( Weisberg 2005) and an artificial equicorrelation matrix.

### 3.1    Heart attack data

We use the root mean squared error (RMSE) of the off-diagonal elements of the correlation matrix given by $\mathrm{rmse} = \sqrt{\dfrac{1}{\frac{1}{2}p(p-1)}\sum_{i<j}\left(r_{ij}-\hat{r}_{ij}\right)^2}$ as a measure of fit for those methods that do not aim to approximate the diagonal (PFA and WALS), whereas we will include the diagonal for those methods that do try to fit the diagonal (PCA, CRG and MDS). The panel plot in Figure 2 shows the results for PCA, CRG, MDS, PFA, WALS and WALS with scalar adjustment $\delta$.

Figure 2A shows a PCA biplot of the correlation structure, where correlations are approximated by the scalar products between vectors. The RMSE when using scalar products is 0.1808, whereas the RMSE obtained in PCA by cosines is 0.2945. Figure 2B shows the correlogram for the Heart attack data, which capitalizes on the representation by cosines. This representation decreases the RMSE to 0.2671 in comparison with cosines in PCA. Figure 2C shows an MDS plot of the correlation structure. To facilitate interpretation, intervariable distances larger than $\sqrt{2}$ are marked with dotted lines to stress that they represent negative correlations. Variables that are not connected thus have a positive correlation. The plot indicates that *SI* and *CI* are positively correlated, and that these variables have negative correlations with all other variables. The plot also shows that *PA,*

*DBP* and *PR* form a positively correlated group. Figure 2D shows the factor loading plot obtained by PFA; this achieves a considerably lower RMSE of 0.0755 for not fitting the diagonal. Note that variable *CI* reaches the unit circle.

In Figure 2E we show the WALS biplot, which also avoids fitting the diagonal, but is also freed from the constraints of the factor model. Variable *CI* is now seen to slightly move out of the unit circle. The RMSE of WALS is the lowest of all methods in comparison with all previous methods; its RMSE is slightly below the RMSE obtained by PFA (0.075519 versus respectively 0.075523). Maximum likelihood factor analysis of the data produces a Heywood case, with precisely *CI* achieving a communality of 100%. When an additive adjustment is used, we obtain $\delta = -0.2706$. The corresponding biplot is shown in Figure 2F. With the adjustment, *CI* appears to stretch further, and the obtuse angles between the pair *CI* and *SI* and the remaining variables become smaller. The scalar adjustment further reduces the RMSE of the approximation to 0.06622, providing the best approximation to the correlation matrix.

The effect of the proposed adjustment is illustrated in more detail in Figure 3, by using biplot axis calibration ( Graffelman and van Eeuwijk 2005), for both PCA and WALS. Calibration of a biplot axis refers to the process of drawing tick marks with numeric labels along (mostly oblique) biplot vectors. Calibration can be used to illustrate the biplot interpretation rules, and to highlight a particular variable of interest. The earliest example of biplot calibration stems from Gabriel and Odoroff (1990); formulae for carrying out the calibration have been developed by several authors ( Gower and Hand 1996; Graffelman and van Eeuwijk 2005; Gower et al. 2011). In the R environment, calibration can be carried out with the package `calibrate`. In all panels of Figure 3, variable *SI* has been calibrated in order to show the change in interpretation, and the calibrated scale for *SI* is shifted towards the margin of the plot ( Graffelman 2011) to improve the visualization. Note that in the analyses with the $\delta$ adjustment (panels B and D), the origin of the scale for *SI* is no longer zero, but shifted by $\delta$, as is emphasized by the projection of the origin onto the calibrated scale. The origin of the plot, where the biplot vectors emanate from, is represented by the values $\delta = 0.14$ and $\delta = -0.27$ for panels B and D respectively. The sample correlations of *SI* with all other variables and the approximations by the different types of analysis are shown in Table 2; generally WALS with the adjusted correlation matrix most closely approximates the sample correlations, and has the lowest RMSE. The RMSE of all variables are shown in Table 3; this shows that WALS considerably lowers the RMSE in comparison with PCA and that the representation of variable *Pulse* benefits from using the adjustment.

Finally, we map observations onto the WALS correlation biplot by regression, as shown in Figure 4B, and compare the results with those obtained by PCA in Figure 4A. In PCA, the goodness-of-fit of the standardized data matrix is calculated from the eigenvalues and is 0.736; the goodness-of-fit of the correlation matrix (including the diagonal), calculated from the squared eigenvalues is 0.913. The correlation matrix thus has better fit than the standardized data matrix (see Discussion). In PCA, both principal components contribute to the goodness-of-fit, and these contributions neatly add up. Figure 4A shows the contributions of both axes to both the representation of the standardized data matrix (0.560 + 0.176 = 0.736) and to the correlation matrix 0.832 + 0.082 = 0.913. For WALS, the

goodness-of-fit of the standardized data matrix is 0.729, only slightly below the maximum achieved by PCA. The scores of the patients for the principal components in Figure 4A have been scaled by multiplying by $1 / \sqrt{\chi_2^2(0.95)} = 0.409$; this way the unit circle drawn for the variables coincides exactly with the 95% contour of a multivariate normal density for the principal components. PCA actually provides a *double biplot*, since the projections of dots onto vectors approximates the (standardized) data matrix, and the projections of vectors onto vectors approximates the correlation matrix.

The WALS solution has different properties. First of all, there is no nesting of axes, i.e. the first dimension obtained of a two-dimensional approximation is different from the single dimension obtained in a one-dimensional solution. Moreover, extracted axes are generally not uncorrelated. The goodness-of-fit of the correlation matrix is best judged by the RMSE of the off-diagonal correlations. In PCA the RMSE is 0.181 if the diagonal is included. By using WALS, the RMSE is more than halved, giving 0.0755 or 0.0662 if the adjustment is used. In this case, one could say the WALS solution halves the RMSE of the correlations at the expensive of sacrificing less than 1% of the goodness-of-fit of the standardized data matrix in comparison with PCA. WALS reduces the RMSE of all variables with respect to PCA (see Table 3), *VP, PA, DBP* and *Pulse* in particular. In the WALS biplot, the vectors for *Pulse* and *VP* appear shorter, reducing the exaggeration observed in PCA of the correlations of these two variables with DBP and PA, and *CI* and *SI* respectively. The Supplementary Materials provide approximations of the correlation matrix obtained by all methods discussed.

### 3.2   Aircraft data

The Aircraft data consists of four variables, the specific power ( *SPR*), the flight range factor ( *RGF*), the payload ( *PLF*) and sustained load factor ( *SLF*), registered for 21 fighter aircrafts. This data has been described and analysed by Gower and Hand (1996). We here focus on the correlations between these four variables. Figure 5A shows the biplot of the correlation matrix obtained by a PCA of the standardized data. The RMSE of the representation of the correlation matrix (ones on the diagonal included) is 0.1362. If we apply WALS with the additive adjustment, we obtain an estimate for $\delta$ that is almost zero (3.3e-05), indicating that for this data, there is no benefit in using the adjustment. The corresponding biplot is shown in Figure 5B, and has a RMSE that is very small (0.0003), meaning the correlation structure is virtually perfectly represented in this two dimensional display. All variable vectors fall within the unit circle. For this case, the WALS representation is equivalent to the factor loading plot obtained in PFA. The biplot vectors obtained by WALS are shorter, which translates into smaller scalar products and therefore lower estimates of the correlation coefficients. Because PCA tries to fit the diagonal of the correlation matrix, it tends to stretch the biplot vectors towards unit length, and thereby exaggerates the correlations between the variables.

### 3.3   Swiss banknote data

The Swiss banknote data consists of six measurements of different size aspects of a banknote: the top margin ( *Top*) and the bottom margin ( *Bottom*) surrounding the image, the diagonal of the image on the banknote ( *Diagonal*), the left and right height of the

banknote ( *Left* and *Right* respectively) and the horizontal length ( *Length*) of the banknote. The original data consists of counterfeits and non-counterfeits (a hundred of each), of which we use the counterfeits only. Figure 6A shows the PCA correlation circle, and Figure 6B the correlation circle obtained by WALS with zero weights for the diagonal and using the $\delta$ adjustment. By using WALS we obtain a considerable lower RMSE (0.0467), and the origin of the WALS biplot represents a correlation of $\delta = 0.07$. To facilitate the interpretation of the WALS biplot, zero correlation is marked on the biplot vector of each variable with a black dot and the positive part of the correlation scale is shown in blue. Variable *Bottom* appears stretched outside the unit circle with respect to PCA; the data set would have a produced a Heywood case in ML factor analysis. PCA and WALS biplots show two quite interpretable dimensions in the data: an apparent *banknote* size dimension represented by the outer size measurements *Left*, *Right* and *Length*, and an *vertical centring of the image* dimension represented by the opposition between *Top* and *Bottom*. Detailed RMSE statistics for each variable in Table 4 show all variables improve their representation when the WALS biplot is used. The largest improvement is due to avoiding the fit of the diagonal of the correlation matrix. The PCA biplot exaggerates the negative correlation between *Top* and *Diagonal*, and also exaggerates the positive correlations of the group ( *Left*, *Right* and *Length*). In general, WALS reduces the estimates of the correlations obtained by PCA towards zero. When a rank three approximation to the correlation matrix is considered, the RMSE obtained by PCA with three principal components decreases to 0.1447, whereas WALS obtains an almost perfect representation of **R** with a RMSE below 0.0002.

### 3.4   The equicorrelation matrix

We analyze a $10 \times 10$ equicorrelation matrix with a correlation of 0.5 between all pairs of variables. The PCA and WALS biplots are shown in Figure 7. The RMSE of the approximation obtained with two principal components is 0.1414, whereas for WALS a the approximation of the correlation matrix turns out to be perfect (RMSE=0). Though we use two dimensional plots in Figure 7, we show a rank one approximation for WALS, which already achieves a perfect approximation to **R**. Quite obviously, the perfect rank one approximation is obtained by 10 coincident biplot vectors with norm $\sqrt{r_{ij}} = 0.7071$. Note that in order to achieve zero RMSE in PCA, all 10 dimensions would be needed. Again, PCA is hampered by having to fit the diagonal of ones of the correlation matrix.

## 4   Discussion

Principal component analysis is widely used for making graphical representations, correlation circles, better termed biplots, of a correlation matrix. However, as shown in this article, the approximation to the correlation matrix offered by using the first two dimensions extracted by PCA is suboptimal, and a weighted alternating least squares algorithm that avoids the fitting of the diagonal outperforms PCA and other approaches. It is commonplace to analyse and visualize a quantitative data matrix by PCA, but it is questionable if that is really the best way to proceed. Indeed, it is shown with an example that an approximation of the correlation matrix by WALS and adding the observations to the biplot of the correlation matrix posteriorly by regression may be more interesting than PCA itself, because it capitalizes on the display of the correlation structure, possibly at the expense of only

sacrificing a minor portion of explained variance of the data matrix. The WALS algorithm is powerful and deserves more attention; we suggest it as the default method for depicting correlation structure. The various approaches detailed in the article are all available in the R environment combined with the `Correlplot` R package. The examples shown in the article shows the largest reduction in RMSE is obtained by turning off the diagonal of the correlation matrix; an additional but smaller gain results from using the proposed adjustment. The adjustment, when different from zero, changes the interpretation rules of the biplot: orthogonal vectors no longer correspond to correlation zero, but to correlation δ. Consequently, biplot vectors with sharp and obtuse angles no longer necessarily correspond to positive and negative correlations. To facilitate interpretation, marking the zero on the biplot vector is therefore recommended (See Figure 6B). More practical data analysis, and eventually some simulation studies, will provide more insight on the benefit of the adjustment.

In a correlation-based PCA, the first eigenvalue is always larger or at worst equal to 1, as the first component will have a variance larger than that of one standardized variable. Trailing eigenvalues are smaller than 1, because the eigenvalues sum to p, the number of variables. Consequently, when eigenvalues are squared, the relative contribution of the first dimension increases. This implies that PCA generally does a better job at representing the correlation matrix than it does at representing the standardized data matrix, as is also the case for the Heart attack data studied in this paper, at least if the diagonal of ones is included. The ''variables plot'' of the `FactoMineR` and `factoextra` R packages report, on its axes, the explained variability of the standardized *data* matrix, *not* of the correlation matrix (See Figure 1B), for which the squared eigenvalues (Fig. 4A, second entry on each axis) are needed. Ultimately, to report the goodness-of-fit (or error) of the approximation to the correlation matrix, the off-diagonal RMSE is the preferred measure for avoiding the innecessary approximation of the ones, and for being directly interpretable as the average amount of error in the correlation scale.

Principal components are usually centered, therefore have zero mean, they are orthogonal and uncorrelated. The axes extracted by the WALS algorithm generally do not have these properties. A centering of the WALS solution is not always convenient, because scalar products are not invariant under the centering operation, and centering the solution can therefore worsen the approximation. The scores obtained by WALS can, if desired, be orthogonalized by using the singular value decomposition; this has been used for most WALS biplots shown in this article.

The elegance and power of the WALS algorithm resides in its generality, as it encompasses most, if not all, types of analysis considered in this article. PCA can be performed by applying the WALS algorithm (Eq. (9)) to the column-centred data matrix. Hills (1969) MDS of the transformed correlations can be carried out by WALS of a double-centred correlation matrix.

With regard to the preprocessing of the correlation matrix prior to analysis, obvious alternatives to using the proposed *δ* adjustment are column (or row) centering or a double centering operation prior to biplot construction. Column (or row) centering alone is not

recommended, because this yields a non-symmetric transformed correlation matrix. For biplot construction, the singular value decomposition of this matrix would be needed, leading to different biplot markers for rows and for columns. Consequently, each variable would be represented twice, by both a column and a row marker that differ numerically, leading to a more dense plot that is less intuitive to interpret. A double centering of the correlation matrix retains symmetry, but due to the double centering operation the origin no longer has a unique interpretation and represents a different value for each scalar product. If double centering is applied, then the representation of the correlations by distances, as proposed by Hills (1969), is more convenient than the use of scalar products. The proposed additive adjustment retains symmetry and preserves the use of the scalar product for interpretation.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

## References

Bailey R and Gower J (1990), 'Approximating a symmetric matrix', Psychometrika 55(4), 665–675.

De Leeuw J (2006), 'A decomposition method for weighted least squares low-rank approximation of symmetric matrices', Department of Statistics, UCLA. Retrieved from https://escholarship.org/uc/item/1wh197mh.

Eckart C and Young G (1936), 'The approximation of one matrix by another of lower rank', Psychometrika 1(3), 211–218.

Friendly M (2002), 'Corrgrams: exploratory displays for correlation matrices', The American Statistician 56(4), 316–324.

Gabriel KR (1971), 'The biplot graphic display of matrices with application to principal component analysis', Biometrika 58(3), 453–467.

Gabriel KR (1978), The complex correlational biplot, *in* Shye S, ed., 'Theory construction and data analysis in the behavioral sciences', Jossey-Bass, pp. 350–370.

Gabriel KR and Odoroff CL (1990), 'Biplots in biomedical research', Statistics in Medicine 9(5), 469–485. [PubMed: 2349401]

Gabriel K and Zamir S (1979), 'Lower rank approximation of matrices by least squares with any choice of weights', Technometrics 21(4), 489–498.

Gower JC (1966), 'Some distance properties of latent root and vector methods used in multivariate analysis', Biometrika 53(3), 325–338.

Gower JC, Gardner Lubbe E and Le Roux N (2011), Understanding biplots, John Wiley, Chichester.

Gower JC and Hand DJ (1996), Biplots, Chapman & Hall, London.

Graffelman J (2011), A universal procedure for biplot calibration, *in* Ingrassia S, Rocci R and Vichi M, eds, 'New Perspectives in Statistical Modeling and Data Analysis, Studies in Classification, Data Analysis and Knowledge organization', Springer-Verlag Berlin Heidelberg, pp. 195–202.

Graffelman J (2013), 'Linear-angle correlation plots: new graphs for revealing correlation structure', Journal of Computational and Graphical Statistics 22(1), 92–106.

Graffelman J (2022), Fifty years of biplots: some remaining enigmas and challenges, *in* '17th Conference of the International Federation of Classification Societies (IFCS 2022). Classification and Data Science in the Digital Age. Book of Abstracts', CLAD - Associação Portuguese de Classificação e Análise de Dados (ed.), p. 40. **URL:** https://ifcs2022.fep.up.pt**URL:**

Graffelman J and van Eeuwijk FA (2005), 'Calibration of multivariate scatter plots for exploratory analysis of relations within and between sets of variables in genomic research', Biometrical Journal 47(6), 863–879. [PubMed: 16450858]

Greenacre MJ (2010), Biplots in Practice, BBVA Foundation. Rubes Editorial.

Harman HH (1976), Modern Factor Analysis, third edn, University of Chicago Press.

Heywood H (1931), 'On finite sequence of real numbers', Proceedings of the Royal Society of London. Series A 134, 486–501.

Hills M (1969), 'On looking at large correlation matrices', Biometrika 56(2), 249–253.

Johnson RA and Wichern DW (2002), Applied Multivariate Statistical Analysis, fifth edn, New Jersey: Prentice Hall.

Kassambara A (2017), Practical Guide To Principal Component Methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra, Vol. 2, STHDA.

Kassambara A and Mundt F (2020), factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. **URL:** https://CRAN.R-project.org/package=factoextra**URL:**

Keller J (1962), 'Factorization of matrices by least-squares', Biometrika 49(1/2), 239–242.

Lê S, Josse J and Husson F (2008), 'FactoMineR: A package for multivariate analysis', Journal of Statistical Software 25(1), 1–18.

Mardia KV, Kent JT and Bibby JM (1979), Multivariate Analysis, Academic Press London.

Murdoch D and Chow E (1996), 'A graphical display of large correlation matrices', The American Statistician 50(2), 178–180.

Pietersz R and Groenen PJF (2004), 'Rank reduction of correlation matrices by majorization', Quantitative Finance 4, 649–662.

R Core Team (2022), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. **URL:** https://www.R-project.org/**URL:**

Rodgers JL and Nicewander WA (1988), 'Thirteen ways to look at the correlation coefficient', The American Statistician 42(1), 59–66.

Saporta G (1990), Probabilités analyse des données et statistique, éditions technip, Paris.

Satorra A and Neudecker H (1998), 'Least-squares approximation of off-diagonal elements of a variance matrix in the context of factor analysis', Econometric Theory 14(1), 156–157.

Trosset MW (2005), 'Visualizing correlation', Journal of Computational and Graphical Statistics 14(1), 1–19.

Wei T and Simko V (2021), R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.92). **URL:** https://github.com/taiyun/corrplot**URL:**

Weisberg S (2005), Applied Linear Regression, third edn, John Wiley & Sons, New Jersey.

Wright K (2021), corrgram: Plot a Correlogram. R package version 1.14. **URL:** https://CRAN.R-project.org/package=corrgram**URL:**

Yan W and Kang M (2003), GGE biplot analysis : a graphical tool for breeders, geneticists, and agronomists, CRC Press, Boca Raton, FL.
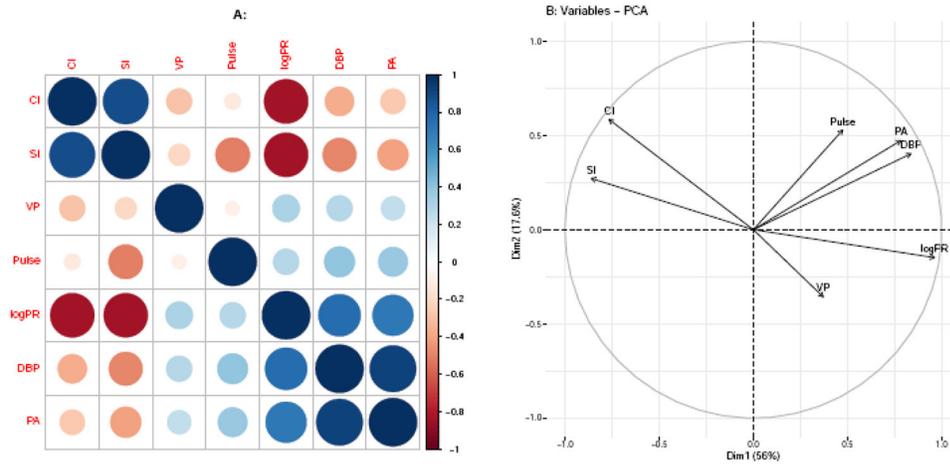
**Fig. 1.**

Displays of the correlation matrix of the Heart attack data obtained in R by using the `corrplot`, `FactoMineR` and `factoextra` packages. A: Colored tabular display or corrgram. B: Correlation circle or correlation biplot. See Section 2 for the abbreviations of the names of the variables.
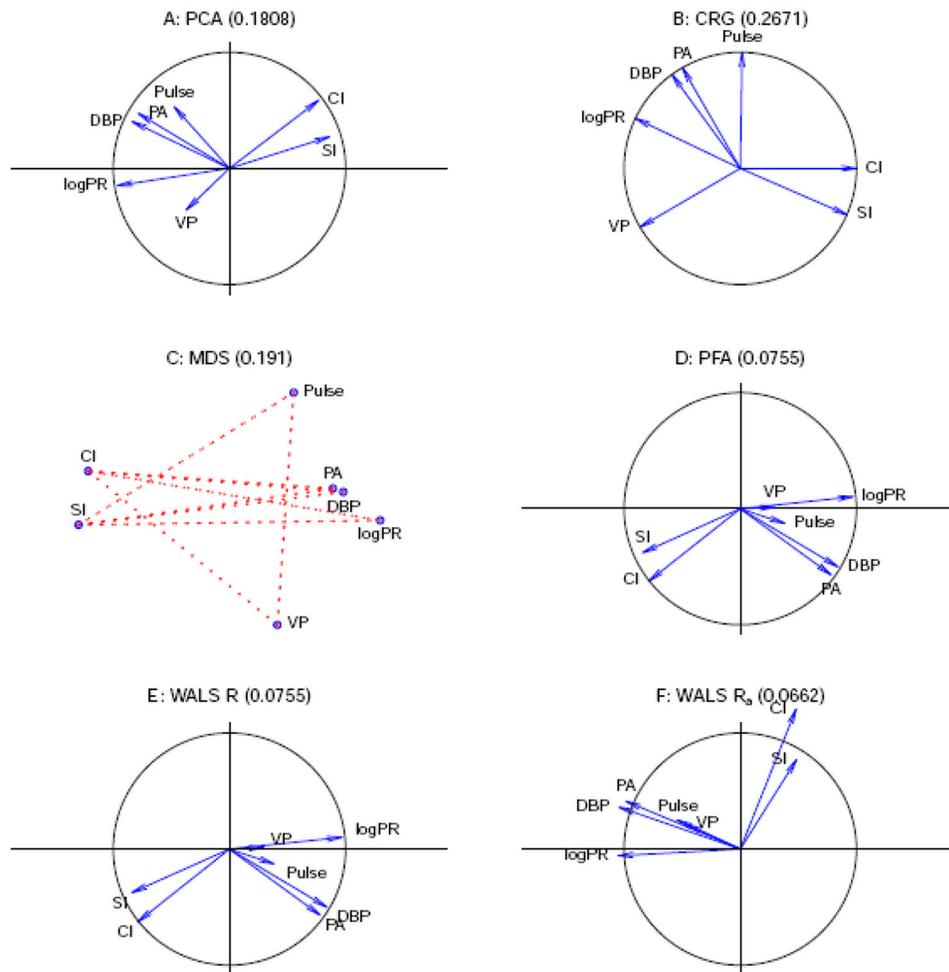
**Fig. 2.**
Visualizations of correlation structure, using the Heart attack data. A: PCA biplot of the correlation matrix; B: Trosset's correlogram; C: MDS map, with negative correlations indicated by dotted lines. D: Biplot obtained by PFA. E: Biplot obtained by WALS. F: Biplot obtained by WALS with scalar adjustment $\delta$. The RMSE of the approximation is given between parentheses in the title of each panel.
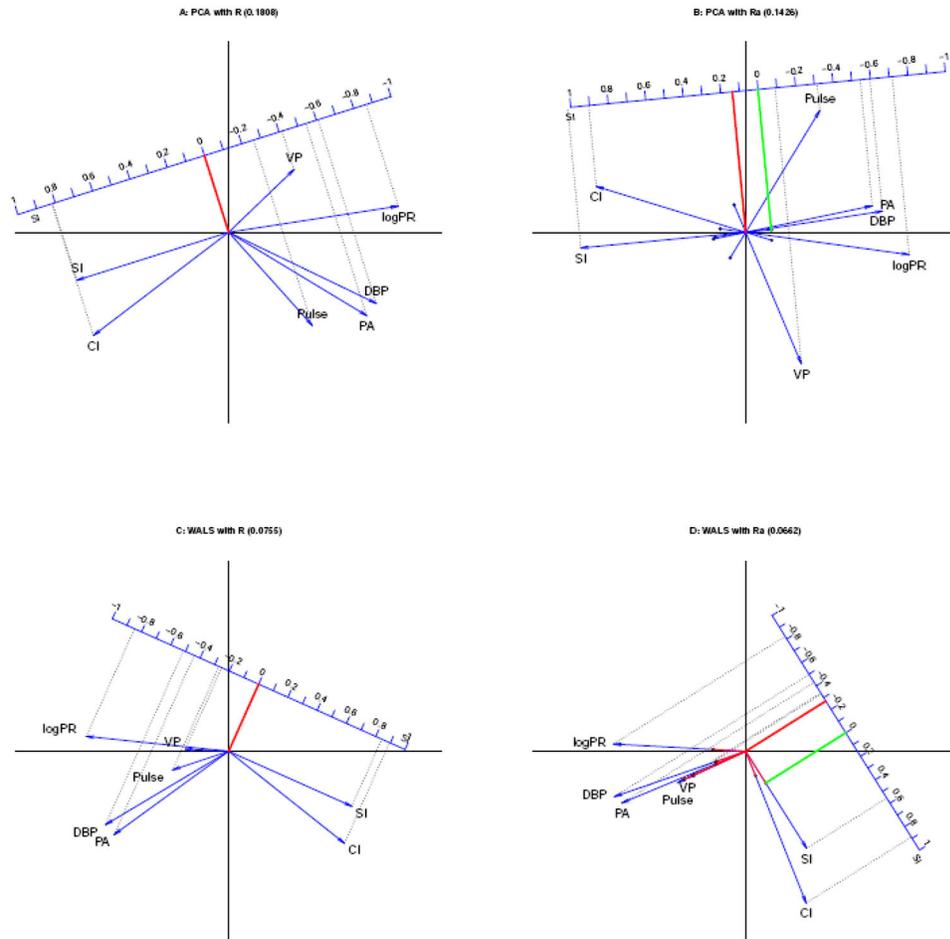
**Fig. 3.**
Adjustment of the correlation matrix. All biplots have a calibrated correlation scale for *SI*. A: PCA biplot; B: PCA biplot using the $\delta$ adjustment; C: WALS biplot; D: WALS biplot using the $\delta$ adjustment. The interpretation of the biplot origin is shown by its projection (in red) onto the calibrated scale. Black dots on biplot vectors correspond to zero correlation for the corresponding variable. The interpretation of the black dot for *SI* is shown by its projection (in green) onto the calibrated scale. For positive $\delta$, biplot vectors are extended beyond the biplot origin towards their zero point. For negative $\delta$, tails of biplot vectors are colored in red for the negative part of the correlation scale.
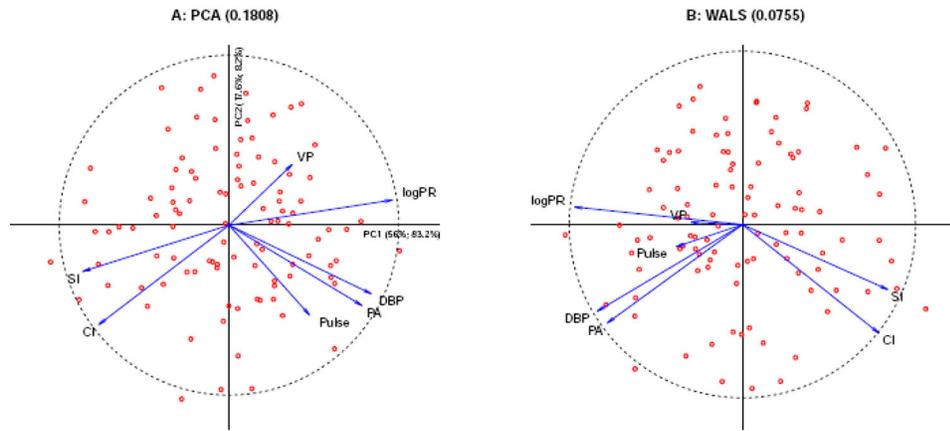
**Fig. 4.**
Comparison of the double biplots of PCA and WALS. A: PCA biplot, reporting goodness-of-fit of both data and correlation matrix on each axis using eigenvalues. B: WALS biplot (without $\delta$ adjustment).
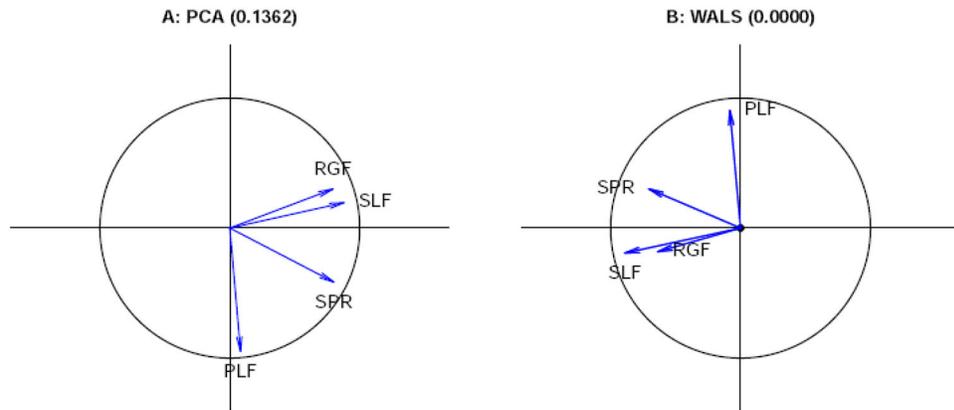
**Fig. 5.**
Biplots of the correlation matrix of the Aircraft data. A: PCA biplot B: WALS biplot. The RMSE of the approximation is given in the title of each panel.
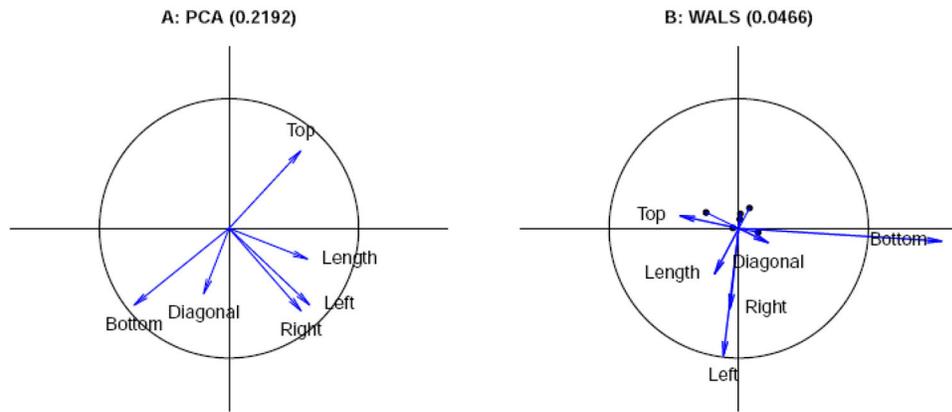
**Fig. 6.**
Biplots of the Swiss banknote data. A: PCA biplot B: WALS biplot. The RMSE of the approximation is given in the title of each panel.
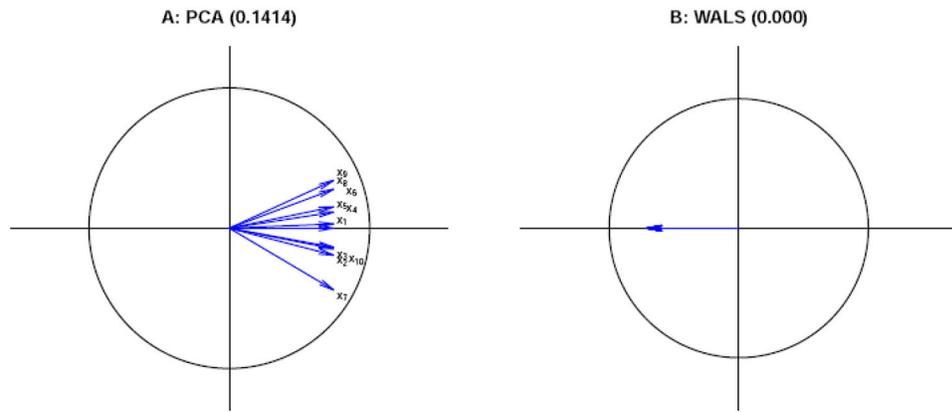
**Fig. 7.**
Biplots of a $10 \times 10$ equicorrelation matrix. A: PCA biplot B: WALS biplot. Variable labels for the WALS biplot are not shown, as the 10 vectors coincide.

**Table 1**

Correlation matrix of the Heart attack data.

|        | CI     | SI     | VP     | Pulse  | logPR  | DBP    | PA     |
|--------|--------|--------|--------|--------|--------|--------|--------|
| CI     | 1.000  | 0.887  | −0.282 | −0.112 | −0.839 | −0.361 | −0.269 |
| SI     | 0.887  | 1.000  | −0.201 | −0.503 | −0.833 | −0.483 | −0.405 |
| VP     | −0.282 | −0.201 | 1.000  | −0.085 | 0.318  | 0.285  | 0.244  |
| Pulse  | −0.112 | −0.503 | −0.085 | 1.000  | 0.287  | 0.399  | 0.370  |
| logPR  | −0.839 | −0.833 | 0.318  | 0.287  | 1.000  | 0.761  | 0.716  |
| DBP    | −0.361 | −0.483 | 0.285  | 0.399  | 0.761  | 1.000  | 0.928  |
| PA     | −0.269 | −0.405 | 0.244  | 0.370  | 0.716  | 0.928  | 1.000  |

**Table 2**

Sample correlations ($r_{(SI,i)}$) of *SI* with all other variables, and estimates of the sample correlations according to four biplots. The bottom line gives the RMSE of *SI* for each method. For PCA, the RMSE calculation includes the correlation of *SI* with itself, for WALS, it does not.

| | $r_{(SI,i)}$ | PCA | | WALS | |
|---|---|---|---|---|---|
| | | $\delta = 0$ | $\delta = 0.14$ | $\delta = 0$ | $\delta = -0.27$ |
| | | **R** | **R**$_a$ | **R** | **R**$_a$ |
| Pulse | −0.503 | −0.264 | −0.316 | −0.271 | −0.340 |
| CI | 0.887 | 0.818 | 0.905 | 0.894 | 0.889 |
| SI | 1.000 | 0.814 | 1.017 | 0.842 | 0.557 |
| DBP | −0.483 | −0.609 | −0.597 | −0.514 | −0.497 |
| PA | −0.405 | −0.544 | −0.546 | −0.440 | −0.430 |
| VP | −0.201 | −0.416 | −0.092 | −0.252 | −0.330 |
| logPR | −0.833 | −0.867 | −0.717 | −0.848 | −0.823 |
| RMSE (SI) | | 0.160 | 0.116 | 0.099 | 0.086 |

**Table 3**

RMSE of all variables for four methods. The bottom line gives the overall RMSE for each method. For PCA, the RMSE calculation includes the correlations of the variables with themselves whereas for WALS these are excluded.

|  | PCA | | WALS | |
|---|---|---|---|---|
|  | $\delta = 0$ | $\delta = 0.14$ | $\delta = 0$ | $\delta = -0.27$ |
|  | **R** | **R**$_a$ | **R** | **R**$_a$ |
| Pulse | 0.2469 | 0.1618 | 0.1345 | 0.0948 |
| CI | 0.0945 | 0.1078 | 0.0482 | 0.0530 |
| SI | 0.1598 | 0.1158 | 0.0988 | 0.0857 |
| DBP | 0.1212 | 0.1540 | 0.0242 | 0.0239 |
| PA | 0.1390 | 0.1828 | 0.0196 | 0.0218 |
| VP | 0.3103 | 0.1336 | 0.0877 | 0.0883 |
| logPR | 0.0564 | 0.1275 | 0.0329 | 0.0521 |
| All | 0.1808 | 0.1426 | 0.0755 | 0.0662 |

**Table 4**

RMSE of all variables for four methods. The bottom line gives the overall RMSE for each method. For PCA, the RMSE calculation includes the correlations of the variables with themselves.

|          | PCA | | WALS | |
|----------|----------------|-------------------|----------------|-------------------|
|          | $\delta = 0$ | $\delta = 0.16$ | $\delta = 0$ | $\delta = 0.07$ |
|          | **R** | **R**$_a$ | **R** | **R**$_a$ |
| Length   | 0.2725 | 0.2734 | 0.0380 | 0.0170 |
| Left     | 0.1624 | 0.1172 | 0.0616 | 0.0440 |
| Right    | 0.1661 | 0.1855 | 0.0608 | 0.0641 |
| Bottom   | 0.0665 | 0.0986 | 0.0071 | 0.0109 |
| Top      | 0.2084 | 0.1080 | 0.0355 | 0.0186 |
| Diagonal | 0.3352 | 0.2890 | 0.0826 | 0.0790 |
| All      | 0.2192 | 0.1949 | 0.0533 | 0.0466 |