# Software Tools for Robust Analysis of High-Dimensional Data

**Valentin Todorov**
UNIDO

**Peter Filzmoser**
Vienna University of Technology

### Abstract

The present work discusses robust multivariate methods specifically designed for high dimensions. Their implementation in R is presented and their application is illustrated on examples. The first group are algorithms for outlier detection, already introduced elsewhere and implemented in other packages. The value added of the new package is that all methods follow the same design pattern and thus can use the same graphical and diagnostic tools. The next topic covered is sparse principal components including an object oriented interface to the standard method proposed by Zou, Hastie, and Tibshirani (2006) and the robust one proposed by Croux, Filzmoser, and Fritz (2013). Robust partial least squares (see Hubert and Vanden Branden 2003) as well as partial least squares for discriminant analysis conclude the scope of the new package.

*Keywords*: high dimensions, robustness, classification, PLS, PCA, outliers.

## 1. Introduction

High-dimensional data are typical in many contemporary applications in scientific areas like genetics, spectral analysis, data mining, image processing, etc. and introduce new challenges to the traditional analytical methods. First of all, the computational effort for the anyway computationally intensive robust algorithms increases with increasing number of observations $n$ and number of variables $p$ towards the limits of feasibility. Some of the robust multivariate methods available in R (see Todorov and Filzmoser 2009) are known to deteriorate rapidly when the dimensionality of data increases and others are not applicable at all when $p$ is larger than $n$.

The present work discusses robust multivariate methods specifically designed for high dimensions. Their implementation in R is presented and their application is illustrated on examples. A key feature of this extension of the framework is the object model which follows the one already introduced by **rrcov** and based on statistical design patterns. The first group of classes are algorithms for outlier detection, already introduced elsewhere and implemented in other packages. The value added of the new package is that all methods follow the same pattern and thus can use the same graphical and diagnostic tools. The next topic covered is sparse principal component analysis including an object oriented interface to the standard method proposed by Zou *et al.* (2006) and the robust one proposed by Croux *et al.* (2013). These

are presented and illustrated in Section 2. Robust partial least squares (Hubert and Vanden Branden 2003; Sernels, Croux, Filzmoser, and van Espen 2005) as well as partial least squares for discriminant analysis are presented in Section 3 and Section 4. Section 5 concludes.

# 2. Robust sparse principal component analysis

Principal component analysis (PCA) is a prominent technique for dimension reduction, and the principle is to find a smaller number $q$ of linear combinations of the originally observed $p$ variables while retaining most of the variability of the data. Dimension reduction by PCA is mainly used for: (i) visualization of multivariate data by scatter plots (in a lower dimensional space); (ii) transformation of highly correlated variables into a smaller set of uncorrelated variables which can be used by other methods (e.g. multiple or multivariate regression, linear or quadratic discriminant analysis); (iii) combination of several variables characterizing a given process into a single or a few *characteristic* variables or *indicators*. In some cases–in particular if the original variables have physical meaning–it is important to be able to interpret these new variables. The interpretation of the principal components needs to be based on the *loadings matrix*, which links the original variables with the principal components.

The standard approach to PCA identifies new directions which are linear combinations of the original variables in such a way, that the data projected on these directions have maximal variance. The different directions need to be orthogonal to each other, and the variance measure used for classical PCA is the empirical sample variance. Practically, the PCA directions can be found by computing the eigenvectors of the sample covariance or correlation matrix. The disadvantage of this approach is that outlying observations may even artificially increase the variance measure, thus leading to essentially uninformative directions. In other words, outliers may attract PCA directions, and the pattern of the data majority will not be covered by the few extracted classical components.

In contrast, the goal of robust PCA is to retain as much of the information of the data majority (and not of single outliers) as possible with fewer directions–the robust PCs. Different approaches to robust PCA are discussed in many review papers, see for example Todorov and Filzmoser (2009) and Filzmoser and Todorov (2013), and examples are given how these robust analyses can be carried out in R. Details about the methods and algorithms can be found in the corresponding references. However, PCA usually tends to provide PCs which are linear combinations of *all* the original variables, even if some of the loadings are small in absolute size. This is a disadvantage for high-dimensional data analysis, since PC directions will then in general be affected by all the variables, even if they are noise variables. It would be more useful to have a method which completely suppresses the influence of potential noise variables by assigning loadings of exactly zero to them. This is the goal of sparse PCA, and there are several proposals available nowadays. A straightforward informal method is to set to zeros those PC loadings which have absolute values below a given threshold (*simple thresholding*). Jolliffe, Trendafilov, and Uddin (2003) proposed SCoTLASS which applies a *lasso* penalty on the loadings in a PCA optimization problem, and recently Zou *et al.* (2006) reformulated PCA as a regression problem and used the *elastic net* to obtain a sparse version - SPCA.

The above mentioned proposals for sparse PCA are not robust with respect to outlying observations. They suffer from the same problem as classical (non-sparse) PCA, namely that the new directions will be attracted by outliers. To cope with the possible presence of outliers in the data, recently Croux *et al.* (2013) proposed a method which is sparse and robust at the same time. It utilizes the *projection pursuit* approach where the PCs are extracted from the data by searching the directions that maximize a robust measure of variance of data projected on it. An efficient computational algorithm was proposed by Croux, Filzmoser, and Oliveira (2007).

**Example** Sparse classical and robust PCA is illustrated here by the (low-dimensional) `cars` data set (Consumer Reports 1990, pp. 235–288); (Chambers and Hastie 1992, pp. 46–47), which is available in the package **rrcovHD** as the data frame `cars`. For $n = 111$ cars, $p = 11$ characteristics were measured, including the length, width, and height of the car. After looking at pairwise scatterplots of the variables, and computing pairwise Spearman rank correlations $\rho(X_i, X_j)$ we see that there are high correlations among the variables, for example, $\rho(X_1, X_2) = .83$ and $\rho(X_3, X_9) = .87$. Thus, PCA will be useful for reducing the dimensionality of the data set (see also Hubert, Rousseeuw, and Vanden Branden (2005)). The first four classical PCs explain more than 96% of the total variance and the first four robust PCs explain more than 95%, therefore we decide to retain four components in both cases. Next we need to choose the degree of sparseness which is controlled by a regularization parameter ($\lambda$). With sparse PCA we take a trade-off between sparseness of the loadings matrix and maximization of the explained variability. The appropriate tuning parameter can be chosen by computing sparse PCA for many different values of $\lambda$ and plotting the percentage of explained variance against $\lambda$. We choose $\lambda = 0.78$ for classical PCA and $\lambda = 2.27$ for robust PCA, thus attaining 83 and 82 percent of explained variance, respectively, which is only an acceptable reduction compared to the non-sparse PCA. Retaining $k = 4$ principal components as above and using the selected parameters $\lambda$, we can construct the so called *diagnostic plots* which are especially useful for identifying outlying observations. The *diagnostic plot* shows the orthogonal distances versus the score distance, and indicates with a horizontal and vertical line the cut-off values that allow to distinguish regular observations (those with small score and small orthogonal distance) from the different types of outliers: *bad leverage points* with large score and large orthogonal distance, *good leverage points* with large score and small orthogonal distance and *orthogonal outliers* with small score and large orthogonal distance (for detailed description see Hubert *et al.* (2005)). In Figure 1 the classical and robust diagnostic plot as well as their sparse counterparts are presented. The diagnostic plot for classical PCA reveals only several orthogonal outliers and identifies two observations as bad leverage points. Three more observations are identified as bad leverage points by sparse classical PCA which is already an improvement, but only the robust methods identify a large cluster of outliers. These outliers are masked by the non-robust score and orthogonal distances and cannot be identified by the classical methods. It is important to note that the sparsity feature added to the robust PCA did not influence its ability to detect properly the outliers.

# 3. Robust linear regression in high dimensions

The toolbox of linear regression methods and their robust counterparts becomes limited when the number of explanatory variables $p$ exceeds the number of observations $n$. The matrix of explanatory variables $\boldsymbol{X}$ is then said to be "flat". In that case, partial least squares (PLS) regression is known to work very well, in particular if the explanatory variables are highly correlated. In this section we will focus on PLS regression and robust versions thereof, since these are widely used tools in various areas.

PLS regression Wold (1975) can be used for the case of a univariate response (PLS1) as well as for a matrix of response variables (PLS2), here denoted by the $n \times q$ matrix $\boldsymbol{Y}$. In the latter case, the regression problem is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \tag{1}$$

with the regression coefficient matrix $\boldsymbol{B}$ and the errors $\boldsymbol{E}$. The basic idea is to decompose $\boldsymbol{X}$ and $\boldsymbol{Y}$ as follows,

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^\top + \boldsymbol{E_X} \tag{2}$$

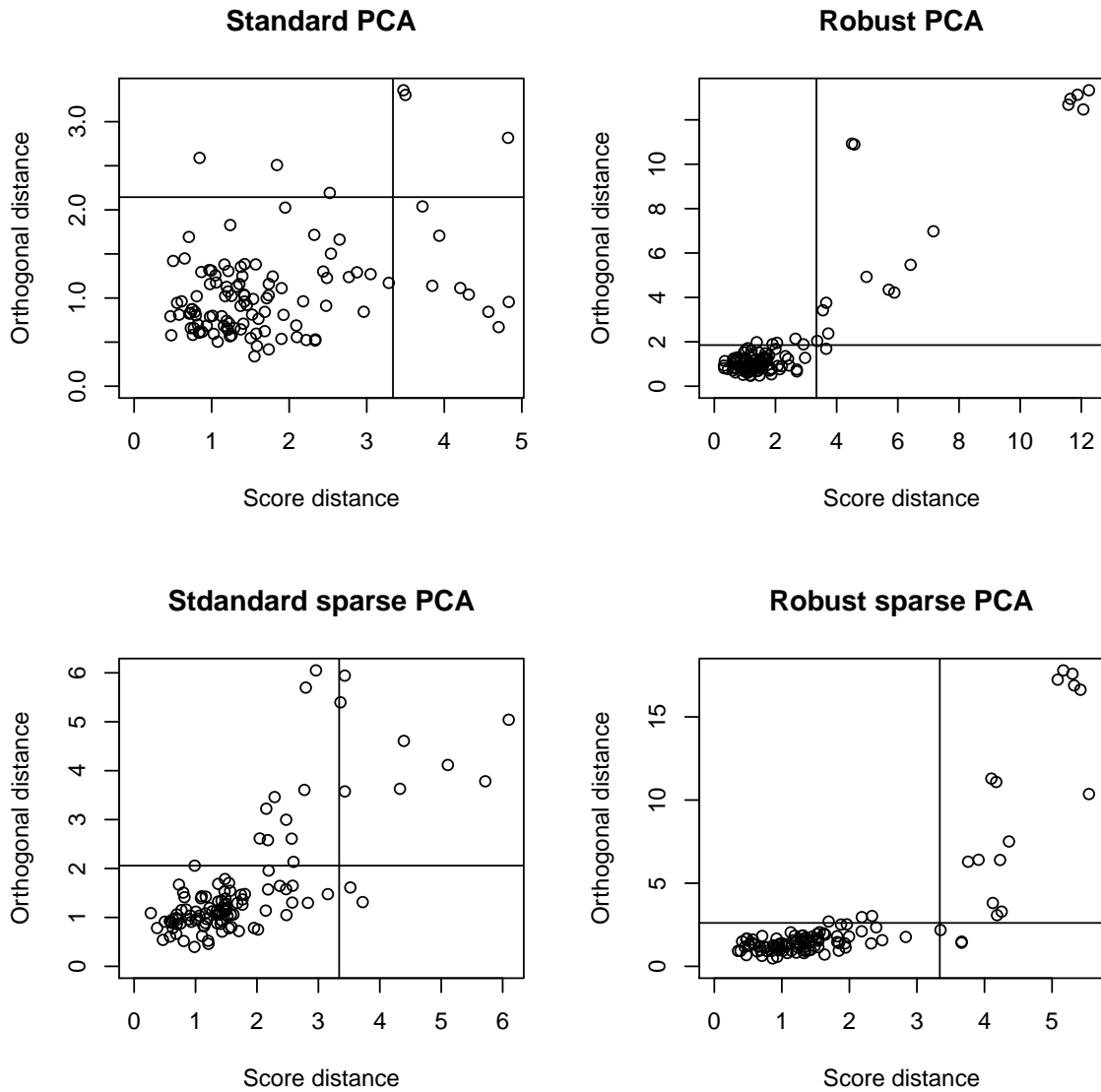$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}^\top + \boldsymbol{E_Y}, \tag{3}$$

Figure 1: Distance-distance plots for standard and sparse PCA and their robust versions for the `cars` data.

with the scores matrices $\boldsymbol{T}$ and $\boldsymbol{U}$, and the loadings matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, each having $K$ columns, and the error matrices $\boldsymbol{E_X}$ and $\boldsymbol{E_Y}$. The number of components $K$ for the factorization is limited with $K \leq \min\{n, p, q\}$. The *inner relationship* connecting the scores is given by

$$U = TD + H, \tag{4}$$

with the diagonal matrix $\boldsymbol{D}$ and a residual matrix $\boldsymbol{H}$.

The key idea in PLS regression is to find a direction $\boldsymbol{w}$ in the $x$-space and a direction $\boldsymbol{c}$ in the $y$-space such that

$$\text{cov}(\boldsymbol{Xw}, \boldsymbol{Yc}) \longrightarrow \max \quad \text{with} \quad \|\boldsymbol{t}\| = \|\boldsymbol{Xw}\| = 1 \quad \text{and} \quad \|\boldsymbol{u}\| = \|\boldsymbol{Yc}\| = 1, \tag{5}$$

where "cov" is an estimator for the covariance. The resulting $\boldsymbol{t}$ and $\boldsymbol{u}$ then form a column in the matrix $\boldsymbol{T}$ and $\boldsymbol{U}$, respectively.

The above procedure is carried out in a sequential manner. This means that the score vectors are computed one after the other, until $K$ vectors are extracted, hereby imposing appropriate constraints (e.g. uncorrelatedness). There are different proposals to solve problem (5), like

the NIPALS algorithm, the kernel algorithm, or the SIMPLS algorithm. For details we refer to Varmuza and Filzmoser (2009).

Hubert and Vanden Branden (2003) suggested a robust version of the SIMPLS algorithm. Since this algorithm is based on estimates of the covariance matrix of the $x$-variables, and of the joint covariance matrix between the $x$- and the $y$-variables, a first step is to robustify these estimates by employing robust PCA. In a second step, a multivariate robust regression method is used.

In case of PLS1, Sernels *et al.* (2005) proposed a robust version that is called partial robust M (PRM) regression. The main idea is to perform robust regression using an M-estimator of the response $\boldsymbol{y}$ on latent variables which are summarizing the explanatory variables. These latent variables, representing only partial information of the $x$-variables, are found in the same spirit as shown in criterion (5),

$$\operatorname{cov}(\boldsymbol{y}, \boldsymbol{X}\boldsymbol{a}) \longrightarrow \max, \tag{6}$$

with appropriate constraints on the loadings vector $\boldsymbol{a}$, and a robust estimator for "cov" using a certain weighting scheme for outlying observations. The loadings and scores are extracted sequentially, again with appropriate side constraints (see also Filzmoser and Todorov 2011).

**Example**  To illustrate robust PLS regression we use a real data example, known from other studies on robust methods. The data set originates from 180 glass vessels Janssens, Deraedt, Freddy, and Veeckman (1998) and was analyzed also in Sernels *et al.* (2005); Hubert, Rousseeuw, and van Aelst (2008); Filzmoser, Maronna, and Werner (2008). In total, 1920 characteristics are available for each vessel, coming from an analysis by an electron-probe X-ray micro-analysis. The data set includes four different materials comprising the vessels, and we focus on the material forming the larger group of 145 observations. It is known from other studies on this data set that these 145 observations should form two groups, because during the measurement process the detector efficiency has been changed. In the original analysis, univariate PLS calibration was performed for all of the main constituents of the glass but here we will consider only the prediction of the sodium oxide concentration and will carry out classical (SIMPLS) and robust (RSIMPLS) PLS with $K = 8$ components. Since the response variable is univariate, regression diagnostic plots for both classical and robust PLS can be created, as shown in Figure 2. The vertical axis represents the standardized residuals $r_i/\hat{\sigma}$
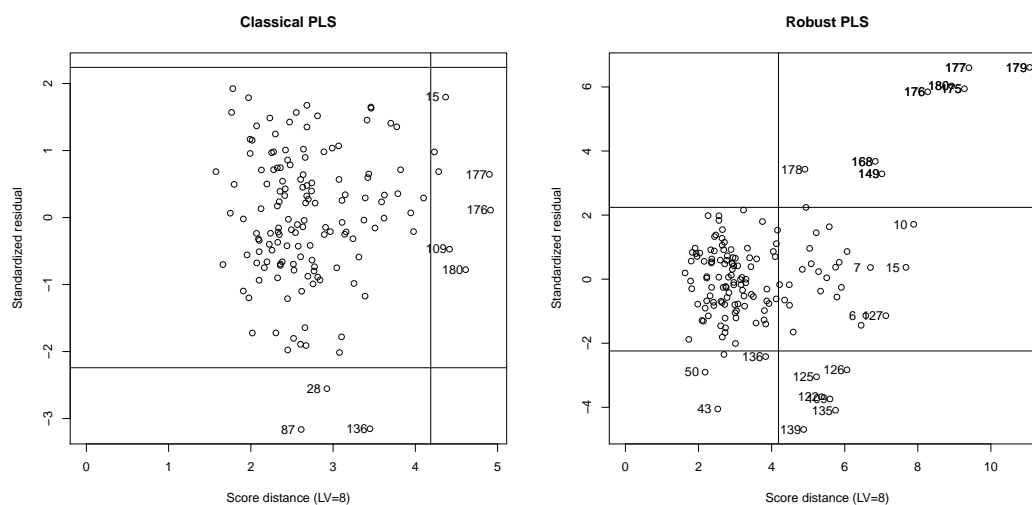


Figure 2: Regression diagnostic plots for the glass data set with (left) SIMPLS and (right) RSIMPLS.

with $r_i = y_i - \hat{\boldsymbol{\beta}}\boldsymbol{x}_i$ while on the horizontal axis the Mahalanobis distances of the data points

in the score space (therefore called *score distances*) are displayed. Outliers in the *t*-space are identified as data points with score distances exceeding the cutoff value of $\sqrt{\chi^2_{K,0.975}}$. Data points which have an absolute standardized residual exceeding $\sqrt{\chi^2_{1,0.975}}$ are flagged as regression outliers. The SIMPLS regression diagnostic plot identifies only three observations as regression outliers and several more as outlying according to the score distances while the robust plot identifies most of the outliers know from other studies. A detailed definition of
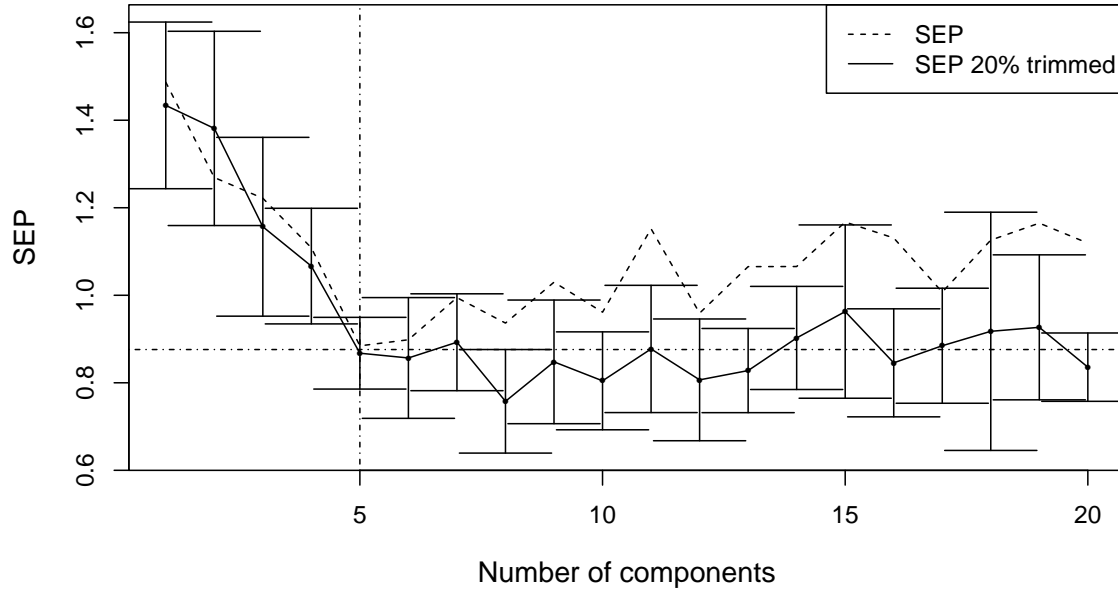


Figure 3: Results of 10-fold cross-validation for robust PLS for the glass data set. A model with 5 components is optimal.

this plot as well as its version for multivariate response variable, can be found in Hubert and Vanden Branden (2003).

For choosing an optimal number of PLS components, 10-fold cross-validation (CV) is used for a maximum of e.g. 20 components and the result is presented graphically in Figure 3. As a performance measure the *standard error of prediction* (SEP) value is used, and its 20% trimmed version.

$$\text{SEP} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{K}(y_{ij}-\hat{y}_{ij}-\text{bias})^2} \quad \text{with} \quad \text{bias} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{q}(y_{ij}-\hat{y}_{ij}). \tag{7}$$

Here, $\{\hat{y}_{ij}\} = \hat{Y} = X\hat{B}$ are the predicted values of the response variable, using the estimated regression parameters $\hat{B}$ (see Varmuza and Filzmoser 2009). Note that the performance measure in (7) is not robust against outliers, because each observation gets the same contribution in the formulas. The influence of outliers to the performance measure can be reduced by trimming for example the 20% of the largest contributions. The dashed line presents the mean of SEP values from CV and the solid part presents the mean and standard deviation of 20% trimmed SEP values from CV. The vertical and horizontal lines correspond to the optimal number of components (after standard-error-rule) and the corresponding 20% trimmed SEP mean, respectively. The optimal number of components is selected as the lowest number whose prediction error mean is below the minimal prediction error mean plus one standard

error, see Varmuza and Filzmoser (2009). Here, 5 components are selected, leading to a prediction error of 0.95.

A more detailed model selection can be done with repeated double cross-validation (rdCV) (see Filzmoser, Liebmann, and Varmuza (2009); Liebmann, Filzmoser, and Varmuza (2010) for details). However, the procedure is rather time consuming. Within an "inner loop", $k$-fold CV is used to determine an optimal number of components, which then is applied to a "test set" resulting from an "outer loop". The procedure is repeated a number of times. The frequencies of the optimal numbers of components are shown in Figure 4. There is a clear peak at 5 components, meaning that a model with 5 components has been optimal in most of the experiments within rdCV. Note that here we obtain the same result as for single CV. In
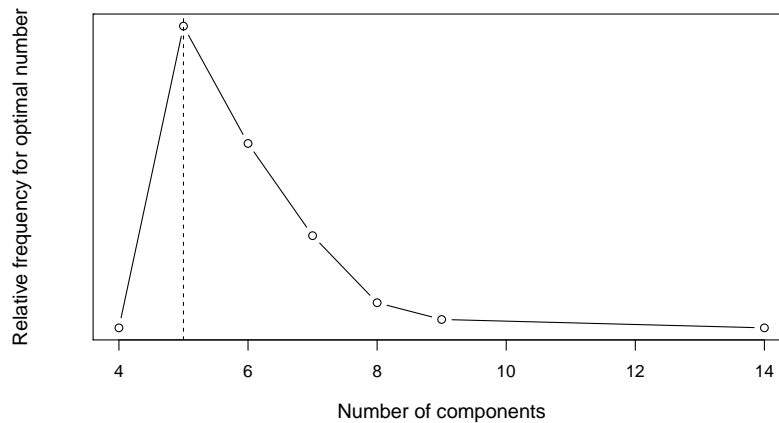


Figure 4: Results of rdCV of RSIMPLS. The optimal number of components is indicated by the vertical dashed line.

a next plot, Figure 5, the prediction performance measure, the 20% trimmed SEP, is shown. The gray lines correspond to the results of the 20 repetitions of the double CV scheme, while the black line represents the single CV result. Obviously, single CV is much more optimistic than rdCV. The estimated prediction error for 5 components is 0.85. Using the optimal number of 5 components, predictions and residuals can be computed. However, for rdCV there are predictions and residuals available for each replication (we used 20 replications). The diagnostic plot shown in 6 presents the predicted versus measured response values. The left panel is the prediction from a single CV, while in the right panel the resulting predictions from rdCV are shown. The latter plot gives a clearer picture of the prediction uncertainty. A similar plot can be generated for predicted values versus residuals (not shown here).

## 4. Robust classification in high dimensions

The prediction of group membership and/or describing group separation on the basis of a data set with known group labels (training data set) is a common task in many applications and linear discriminant analysis (LDA) has often been shown to perform best in such classification problems. However, very often the data are characterized by far more variables than objects and/or the variables are highly correlated which renders LDA (and the other similar standard methods) unusable due to their numerical limitations. Let us assume that $Y$ is univariate and categorical, i.e. $\forall i, 1 \leq i \leq n : y_i \in \{1, \ldots, G\}$ where $G$ is the number of groups. For high dimensional data sets, classical linear discriminant analysis cannot be performed due to the singularity of the estimated covariance matrix $\hat{\Sigma}$, as it requires the inverse of $\hat{\Sigma}$. To overcome the high dimensionality problem in classification context one can reduce the dimensionality by either selecting a subset of "interesting" variables (variable selection)
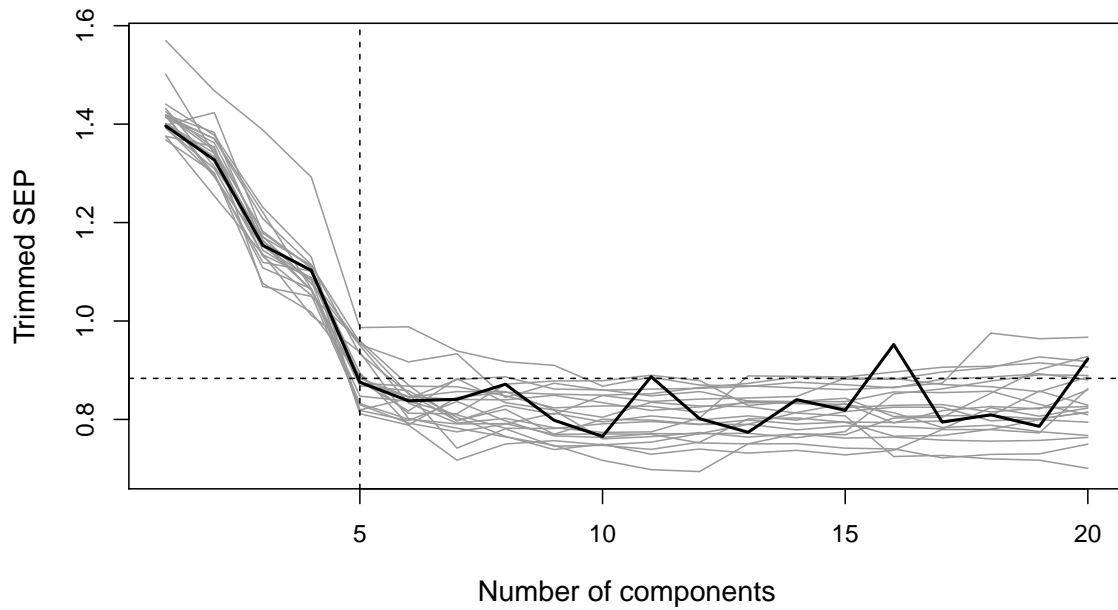
Figure 5: Results of rdCV for RSIMPLS. The gray lines result from repeated double CV, the black line from single CV.

or construct $K$ new components, $K \ll p$ which represent the original data with minimal loss of information (feature extraction, dimension reduction). Many methods for dimension reduction were considered in the literature but the two most popular are principal component analysis (PCA) and partial least squares (PLS). It is intuitively clear that a supervised method (which uses the group information while constructing the new components) like PLS should be preferred to unsupervised methods like PCA.

**SIMCA:**   Instead of applying the dimension reduction method (e.g. PCA) to the full set of observations, one could fit a model to each of the groups (possibly with different number of components) and use these models to classify new observations. This method, called *Soft Independent Modeling of Class Analogies* (SIMCA), was introduced by Wold (1976) and nowadays is widely used as a discriminant technique in chemometrics, where typically $p$ is large relative to $n$. Since in SIMCA PCA is performed on each group, it provides additional information on the different groups, including the relevance of the different variables for groups separation, i.e. their discrimination power. In the original SIMCA method new observations are classified based on their deviation from the different PCA models. These deviations are the Euclidean distances of the observations to the PCA subspace, and thus they are called orthogonal distances. Vanden Branden and Hubert (2005) propose a slightly modified classification rule which better exploits the benefit of applying PCA to each group. This rule includes additionally the score distances, i.e. the Mahalanobis distances measured in the PCA (score) subspace. Furthermore, as a guard against outliers in the data, they propose to replace the classical PCA by a robust alternative. Both the classical and the robust version of the SIMCA method are available in the R package **rrcovHD**.

**Robust PLS-DA:**   PLS was not originally designed to be used in the context of statistical discrimination but nevertheless was routinely applied with evident success by practitioners for this purpose. Taking into account the grouping variable(s) when decomposing the data one

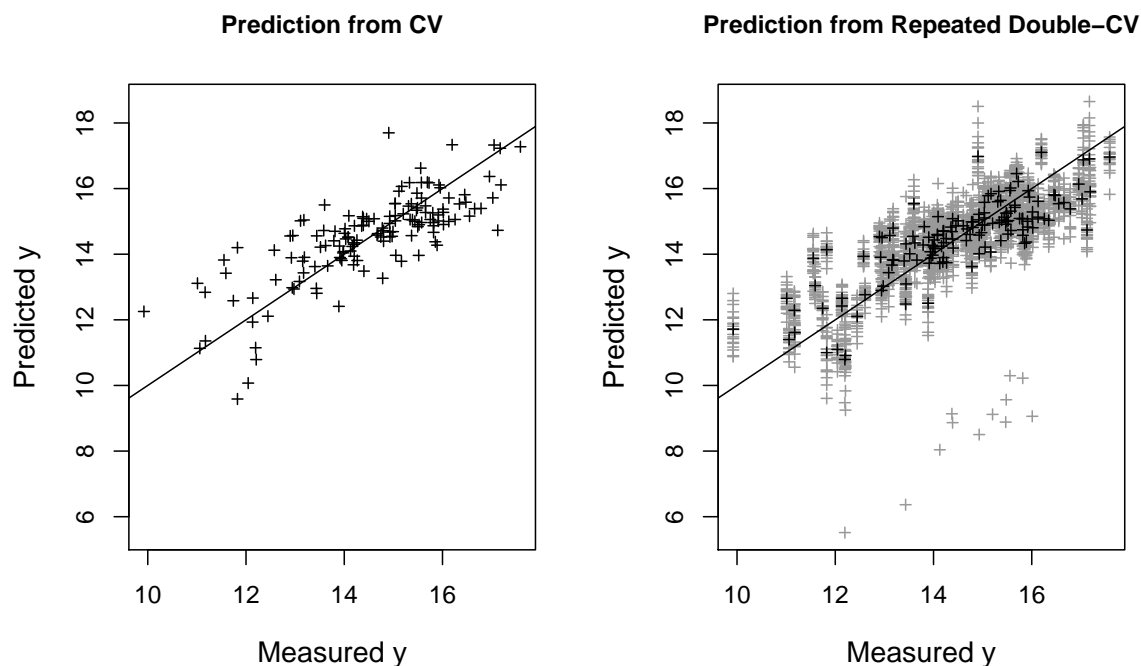**Prediction from CV**    **Prediction from Repeated Double–CV**

Figure 6: Predicted versus measured response values for RSIMPLS. The left panel shows the results from single CV, the right panel visualizes the results from repeated double CV.

would intuitively expect an improved performance for group separation. Since the response variable in case of a classification problem is a categorical variable, none of the robust PLS methods proposed above can be used. Therefore, in order to obtain a robust PLS-DA we proposed to apply any of the outlier detection methods described in Filzmoser and Todorov (2013), which are implemented in package **rrcovHD**, and then use classical PLS on the already cleaned data set. Hubert and Van Driessen (2004) used a data set containing the spectra of three different cultivars of the same fruit. The three cultivars (groups) are named D, M and HA, and their sample sizes are 490, 106 and 500 observations, respectively. The spectra are measured at 256 wavelengths. The fruit data is thus a high-dimensional data set which was used to illustrate a new approach for robust linear discriminant analysis, and it was studied again by Vanden Branden and Hubert (2005). From these studies it is known that the first two cultivars D and M are relatively homogenous and do not contain atypical observations, but the third group HA contains a subgroup of 180 observations which were obtained with a different illumination system. In Figure 7 are shown the prediction histograms for class D for the `fruit` data using classical and robust PLS-DA.

## 5. Summary and conclusions

An object oriented framework for robust multivariate analysis developed in the S4 class system of the programming environment R already exists implemented in the package **rrcov** and is described in Todorov and Filzmoser (2009). The main goal of this framework is to support the usage, experimentation, development and testing of robust multivariate methods as well as simplifying comparisons with related methods. In this article we investigated several robust multivariate methods specifically designed for high dimensions. The focus was on PCA and its sparse version, PLS, PLS for discrimination, and SIMCA. All considered methods and data sets are available in the R package **rrcovHD**. A key feature of this extension of the framework is that the object model follows the one already introduced by **rrcov** which is

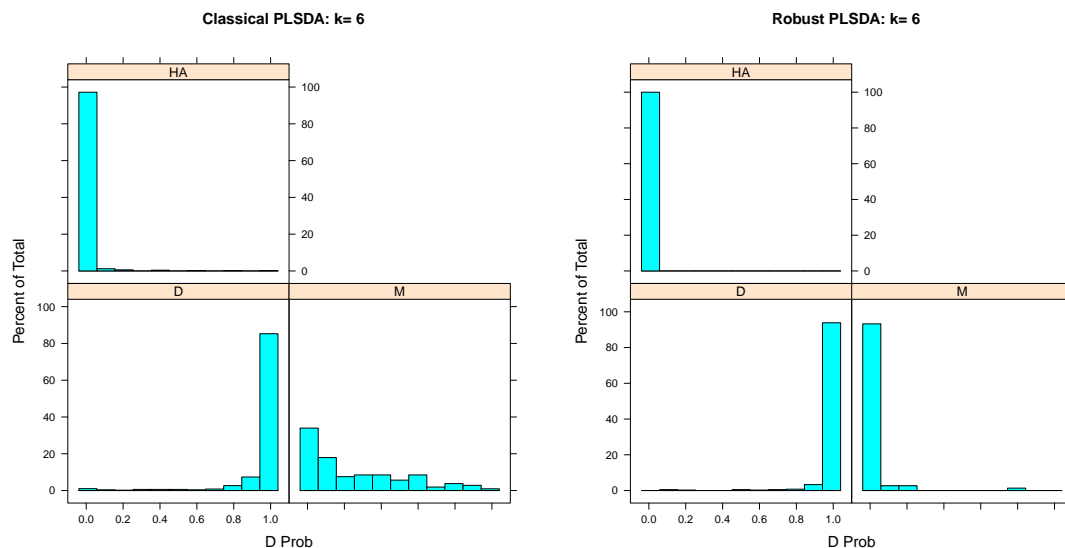**Classical PLSDA: k= 6**     **Robust PLSDA: k= 6**



Figure 7: Prediction histograms for class D for the `fruit` data using classical and robust PLS-DA.

based on statistical design patterns. This makes it easy for the user to apply the methods, since they are following the same structure. A further advantage is that summaries, results, as well as diagnostic plots follow the same structure.

Finally, the strict design pattern used in the package **rrcovHD** is an advantage for extending the package with other methods developed for high-dimensional data–and for sure their robust versions will follow.

# Acknowledgements

# References

Chambers JM, Hastie TJ (1992). *Statistical Models in S*. Wadsworth and Brooks, Cole, Pacific Grove, CA.

Consumer Reports (1990). "Annual Auto Issue: The 1990 cars." http://backissues.com/issue/Consumer-Reports-April-1990. April.

Croux C, Filzmoser P, Fritz H (2013). "Robust Sparse Principal Component Analysis." *Technometrics*, **55**(2), 202–214.

Croux C, Filzmoser P, Oliveira M (2007). "Algorithms for Projection-pursuit Robust Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems*, **87**(218), 218–225.

Filzmoser P, Liebmann B, Varmuza K (2009). "Repeated Double Cross Validation." *Journal of Chemometrics*, **23**(4), 160–171.

Filzmoser P, Maronna R, Werner M (2008). "Outlier Identification in High Dimensions." *Computational Statistics and Data Analysis*, pp. 1694–1711.

Filzmoser P, Todorov V (2011). "Review of Robust Multivariate Statistical Methods in High Dimension." *Analytica Chimica Acta*, **705**, 2–14.

Filzmoser P, Todorov V (2013). "Robust Tools for the Imperfect World." *Information Sciences*, **245**, 4–20.

Hubert M, Rousseeuw P, Vanden Branden K (2005). "ROBPCA: A New Approach to Robust Principal Component Analysis." *Technometrics*, **47**, 64–79.

Hubert M, Rousseeuw PJ, van Aelst S (2008). "High-Breakdown Robust Multivariate Methods." *Statistical Science*, **23**, 92–119.

Hubert M, Van Driessen K (2004). "Fast and Robust Discriminant Analysis." *Computational Statistics & Data Analysis*, **45**, 301–320.

Hubert M, Vanden Branden K (2003). "Robust Methods for Partial Least Squares Regression." *Journal of Chemometrics*, **17**(10), 537–549.

Janssens K, Deraedt I, Freddy A, Veeckman J (1998). "Composition of 15-17$^{th}$ Century Archæological Glass Vessels Excavated in Antwerp, Belgium." *Mikrochimica Acta*, **15 (Suppl.)**, 253–267.

Jolliffe IT, Trendafilov NT, Uddin M (2003). "A Modified Principal Component Technique based on the LASSO." *J. Comput. Graph. Statist.*, **12**(3), 531–547.

Liebmann B, Filzmoser P, Varmuza K (2010). "Robust and Classical PLS Regression Compared." *Journal of Chemometrics*, **24**, 111–120.

Sernels S, Croux C, Filzmoser P, van Espen P (2005). "Partial Robust M-reression." *Chemometrics and Intellegent Laboratory Systems*, **79**, 55–64.

Todorov V, Filzmoser P (2009). "An Object Oriented Framework for Robust Multivariate Analysis." *Journal of Statistical Software*, **32**(3), 1–47. URL http://www.jstatsoft.org/v32/i03/.

Vanden Branden K, Hubert M (2005). "Robust Classification in High Dimensions Based on the SIMCA Method." *Chemometrics and Intellegent Laboratory Systems*, **79**, 10–21.

Varmuza K, Filzmoser P (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics.* Taylor and Francis - CRC Press, Boca Raton, FL.

Wold H (1975). "Soft Modeling by Latent Variables: the Non-Linear Iterative Partial Least Squares Approach." In J Giani (ed.), *Perspectives in probability and statistics, papers in honor of M.S. Bartlett*, pp. 117–142. Academic Press, London.

Wold S (1976). "Pattern Recognition by means of Disjoint Principal Component Models." *Pattern Recognition*, **8**, 127–139.

Zou H, Hastie T, Tibshirani R (2006). "Sparse Principal Component Analysis." *Journal of Computational and Graphical Statistics*, **15**(2), 265–286.

**Affiliation:**

Valentin Todorov
United Nations Industrial Organization (UNIDO), Austria
E-mail: valentin.todorov@unido.org
URL: http://www.unido.org

Peter Filzmoser
Department of Statistics and Probability Theory
Vienna University of Technology, Austria
E-mail: p.filzmoser@tuwien.ac.at
URL: http://www.statistik.tuwien.ac.at/public/filz