JOURNAL OF CHEMOMETRICS
J. Chemometrics **12,** 301–321 (1998)

ANALYSIS OF MULTIBLOCK AND HIERARCHICAL PCA AND PLS MODELS

JOHAN A. WESTERHUIS, THEODORA KOURTI* AND JOHN F. MACGREGOR
McMaster Advanced Control Consortium, Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada

SUMMARY

Multiblock and hierarchical PCA and PLS methods have been proposed in the recent literature in order to improve the interpretability of multivariate models. They have been used in cases where the number of variables is large and additional information is available for blocking the variables into conceptually meaningful blocks. In this paper we compare these methods from a theoretical or algorithmic viewpoint using a common notation and illustrate their differences with several case studies. Undesirable properties of some of these methods, such as convergence problems or loss of data information due to deflation procedures, are pointed out and corrected where possible. It is shown that the objective function of the hierarchical PCA and hierarchical PLS methods is not clear and the corresponding algorithms may converge to different solutions depending on the initial guess of the super score. It is also shown that the results of consensus PCA (CPCA) and multiblock PLS (MBPLS) can be calculated from the standard PCA and PLS methods when the same variable scalings are applied for these methods. The standard PCA and PLS methods require less computation and give better estimation of the scores in the case of missing data. It is therefore recommended that in cases where the variables can be separated into meaningful blocks, the standard PCA and PLS methods be used to build the models and then the weights and loadings of the individual blocks and super block and the percentage variation explained in each block be calculated from the results. © 1998 John Wiley & Sons, Ltd.

KEY WORDS: hierarchical models; latent variables; multiblock models; PCA; PLS

1. INTRODUCTION

Since the invention of the NIPALS (non-linear iterative partial least squares) method, principal component analysis (PCA) and partial least squares or projections to latent structures (PLS) have become very popular as methods to model and analyze large multivariable collinear data sets. Figure 1 shows the well-known arrow scheme of the PCA method, which is presented here for comparing it with the multiblock extension of PCA. The NIPALS algorithm corresponding to this figure is given in Appendix I.1. A tutorial on PCA and some chemical examples can be found in Reference 1.

Figure 2 shows the arrow scheme for the PLS method, which is also presented here for comparing it with multiblock extensions of the method. The algorithm is given in Appendix I.2. A data matrix with descriptors **X** and a data matrix of responses **Y** of several objects are represented by their scores **t** and **u**. The corresponding weights **w** and **q** are obtained by multiplying the scores through the specific matrix. Weight **w** is normalized to length one. New scores are obtained from the weights. This is

* Correspondence to: T. Kourti, Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada.

CCC 0886–9383/98/050301–21 $17·50

Received 16 February 1998
Accepted 13 May 1998

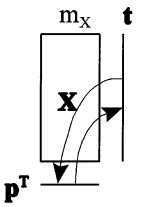

Figure 1. PCA method. A first estimate of the score **t** is multiplied through the data matrix **X** to get an approximation of the loading $\mathbf{p}^T$. The loading is normalized to length one and then multiplied back through **X** to get a new score **t.** This is repeated until convergence of **t**

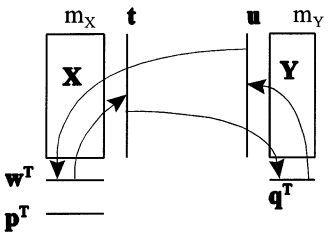

Figure 2. PLS method. **X** is represented by its score **t** and **Y** is represented by **u.** A first estimate of **u** is multiplied through **X** to get an approximation of the weight $\mathbf{w}^T$. The weight is normalized to length one and multiplied back through **X** to produce **t.** From **t** and **Y** the weight $\mathbf{q}^T$ is obtained which gives a new vector **u.** This is repeated until convergence of **t**

repeated until convergence of **t.** Loadings **p** are calculated for the deflation. Geladi described the history of PLS and its enormous impact on the chemometrics world.[2] Tutorials on PLS have appeared[3,4] and several authors have tried to describe the mathematical and statistical aspects of the method.[5–8]

The sizes of the score and weight vectors in PCA and PLS are undefined with respect to a multiplicative constant, c, as $\mathbf{tw} = (\mathbf{t}c)\cdot(\mathbf{w}/c)$. Hence it is necessary to anchor the solution in some way. In the standard PCA and PLS algorithms usually the weights (or loadings for PCA) are set to length one.[1] Then the squared length of the scores equals the variation explained by the principal component.

In recent years several PCA and PLS methods have been introduced which break the descriptors and/or responses into several blocks to improve the interpretation of the models. These multiblock methods have been used in cases where the number of variables is large and additional information is available for blocking the variables into conceptually meaningful blocks. Applications include modeling and monitoring of large chemical processes.[9–12] Several algorithms have appeared to deal with multiple-block data using different notation, under names such as hierarchical PCA (HPCA), consensus PCA (CPCA), hierarchical PLS (HPLS) and multiblock PLS (MBPLS). In this paper we review these algorithms and compare them from a theoretical point of view using a common notation. Undesirable properties such as loss of information and convergence problems of some of these methods are pointed out and corrected where possible. The differences in the results obtained when applying these methods to the same data are theoretically explained and illustrated with several case studies. It is also shown that the results of the consensus PCA and multiblock PLS methods can be calculated from the standard single-block PCA and two-block PLS methods when the same variable scalings are applied for these methods. The standard methods require less computation and give better estimation of the scores in the case of missing data. It is therefore recommended that the standard

PCA and PLS methods be used to build the models and then the weights and loadings of the individual blocks and super block and the percentage variation explained in each block be calculated from the results.

2. NOTATION

For the comparison of multiblock and hierarchical PCA and PLS methods a consistent notation scheme is developed in this paper for all the algorithms presented. Data matrices are written as boldface uppercase characters, vectors as boldface lowercase characters and scalars as lowercase characters.

$\|\ \|$	norm of a vector
$\mathbf{X}$	descriptor data
$\mathbf{X}_b$	descriptor block b
$\mathbf{E}_b$	residual of $\mathbf{X}_b$ after deflation
$\mathbf{Y}$	response data
$\mathbf{t}_b$	block score of $\mathbf{X}_b$
$\mathbf{u}$	score of $\mathbf{Y}$
$\mathbf{w}_b$	weight of variables in block $\mathbf{X}_b$
$\mathbf{q}$	weight of variables in $\mathbf{Y}$
$\mathbf{p}_b$	loadings of variables in block $\mathbf{X}_b$
$\mathbf{T}$	super block containing all $\mathbf{t}_b$s
$\mathbf{t}_\mathrm{T}$	super score
$\mathbf{w}_\mathrm{T}$	super weight
n	number of objects in all blocks
m_X	number of X variables
m_Y	number of Y variables
$m_{\mathrm{X}b}$	number of variables in block $\mathbf{X}_b$
b	block number ($b = 1,\ldots, B$)

Block variable weights $\mathbf{w}_b$ represent the contribution of variables to the block score. Block variable loadings $\mathbf{p}_b$ are used for the calculation of block scores in PCA and also for the deflation of blocks in PCA and PLS methods. Super weights $\mathbf{w}_\mathrm{T}$ give the contribution of the block scores $\mathbf{t}_b$ to the super score $\mathbf{t}_\mathrm{T}$.

PRESS	prediction error sum of squares of cross-validated predictions
RSS	residual sum of squares of calibration

3. MULTIBLOCK ALGORITHMS

3.1. Multiblock PCA methods

In 1987 at the Frankfurt PLS conference, Wold *et al.*[13] introduced the concept of using multiple blocks in PCA methods. Their PCA for multiple blocks was called consensus PCA (CPCA). It was introduced as a method for comparing several blocks of descriptor variables measured on the same objects. Figure 3 shows the arrow scheme for the CPCA method and the algorithm is given in Appendix I.3. The data are divided into B blocks $\mathbf{X}_1, \ldots, \mathbf{X}_B$. A consensus direction among all the blocks is sought. A starting consensus or super score is selected as a column of one of the blocks. This vector is regressed on all blocks $\mathbf{X}_b$ to give block variable loadings $\mathbf{p}_b$. From the block variable loadings, block scores $\mathbf{t}_b$ for all blocks are calculated. All block scores are combined into a super block $\mathbf{T}$. The super score $\mathbf{t}_\mathrm{T}$ is then regressed on the super block to give the super weight $\mathbf{w}_\mathrm{T}$ of each

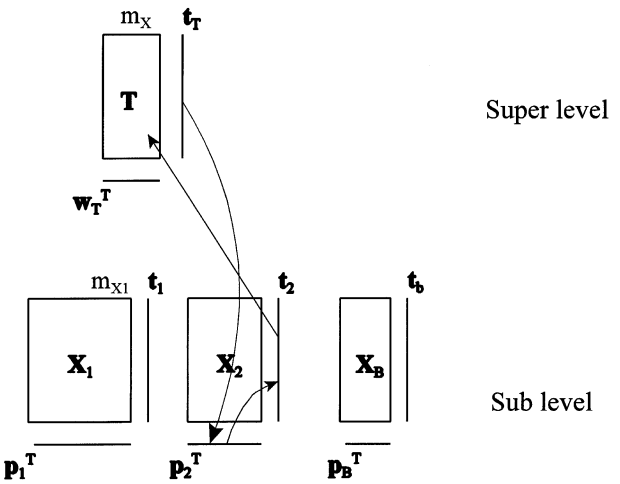

Figure 3. CPCA and HPCA methods. In CPCA a start super score $\mathbf{t}_T$ is regressed on all blocks $\mathbf{X}_b$ to give the block variable loadings $\mathbf{p}_b^T$. The block variable loadings are normalized to length one and then multiplied through the blocks to give the block scores $\mathbf{t}_b$. The block scores are combined in the super block $\mathbf{T}$. Then a PCA round on $\mathbf{T}$ is performed to give the super weight $\mathbf{w}_T^T$ (normalized to length one) and a new super score. This is repeated until convergence of $\mathbf{t}_T$. In HPCA $\mathbf{t}_b$ and $\mathbf{t}_T$ are normalized instead of $\mathbf{w}_T$ and $\mathbf{p}_b$

block score to the super score. The super weight is normalized to length one and a new $\mathbf{t}_T$ is calculated. A new iteration starts until the super score converges to a predefined precision. The super score is derived using all variables, whereas the block scores are derived using only the variables within the corresponding block. The super weight $\mathbf{w}_T$ gives the relative importance of the different blocks $\mathbf{X}_b$ for each dimension. After convergence, all the blocks are deflated using the super score, and a second super score, orthogonal to the first, can be determined by repeating the above iteration on the residual matrix. Wold *et al.*[13] suggested that CPCA could be applied for the analysis of sensory data, for example where a number of referees are judging the sensory quality of a number of wines. Each referee gives his/her judgement on various quality characteristics of the wines (body, bitterness, color, etc.). The results on all the tests for each judge are placed in separated blocks $\mathbf{X}_b$ and the summary scores of each judge are the block scores $\mathbf{t}_b$. The consensus of all judges is represented in the super score $\mathbf{t}_T$, while the super weight $\mathbf{w}_T$ shows the relative importance of each judge in the consensus score.

In 1996, Wold *et al.*[11] introduced a slightly different multiblock PCA method called hierarchical PCA (HPCA, Appendix I.4). The arrow scheme of Figure 3 is also valid for this algorithm. The only difference is in the normalization. In CPCA the super weight $\mathbf{w}_T$ is normalized to have length one and therefore the variance explained by each principal component is equal to the squared length of the corresponding super score. In HPCA the super score is normalized to length one.

Both the CPCA and HPCA algorithms when they were initially presented had convergence problems and had to be corrected for the work presented in this paper. In the initial CPCA algorithm only the super weight $\mathbf{w}_T$ was normalized to length one. This does not seem to be enough for convergence; in this study the block variable loadings $\mathbf{p}_b$ were also normalized to length one. The HPCA algorithm was introduced with only the super score $\mathbf{t}_T$ normalized. The algorithm converges to different solutions depending on the starting vector. In this study we also normalize the block scores $\mathbf{t}_b$ to length one. Rännar *et al.*[14] have already used this adjustment in their adaptive version of HPCA for the monitoring of batch processes.

Additional research revealed that even with the normalization of the block scores in some cases HPCA still converges to several solutions depending on the initial guess of the super score $\mathbf{t}_T$. If the starting vector is highly correlated to a dominant direction in one of the blocks, the algorithm cannot escape from this direction and will select it as the direction of the super score $\mathbf{t}_T$. If the blocks are rather uncorrelated (which is often the reason for blocking), then this results in one of the blocks being favored to be described by the final super score. However, because of the extra normalization of the block scores in the algorithm, often an average between the solutions is found. In Appendix III we give an example to illustrate that HPCA converges to multiple solutions that depend on the initial guess of the super score. To prevent the initial guess from favoring any of the blocks, the eigenvector of the $\mathbf{X}^T\mathbf{X}$ matrix that corresponds to the largest eigenvalue was selected as initial guess in the iterations in these examples. This forces the algorithm to 'a specific solution'. However, the objective function of the HPCA method is not clear as it is for CPCA. The latter has the same objective function as standard PCA, i.e. maximization of the variance in $\mathbf{X}$, which is achieved by choosing the eigenvector of $\mathbf{X}^T\mathbf{X}$ that corresponds to the largest eigenvalue (Appendix II), and it is independent of the initial guess.

Two special cases must be noted for the HPCA algorithm. In any principal component where all block scores are orthogonal, the super score of HPCA of that principal component will be the exact mean of all block scores. In the case of only two blocks the HPCA super score will be the exact mean of both block scores. This averaging is not present in CPCA.

In the multiblock methods, where the variables are blocked, the normalization of the block variable loadings (or weights for PLS) seems more appealing than that of the scores. This allows one to compare the different block scores with their respective lengths on the same objects. If the block scores are normalized to length one, the lengths of the latent variables are in the loadings (and, in PLS, the weights) of the different variables for each block. This makes a comparison between the blocks rather difficult.

Finally, a point to note is that in the CPCA method the block scores are divided by the square root of the number of variables in the specific block to achieve a block scaling. This means that each block starts with the same variance irrespective of its size. However, if additional knowledge is available, important blocks can be scaled up and less important blocks can be scaled down by defining additional scaling factors.[11]

3.2. Multiblock PLS methods

Extensions of PLS to use multiple blocks have also been found to be useful. Using additional information the variables can be divided into blocks to improve the interpretability of the model. The first application of PLS with multiple $\mathbf{X}$ blocks, called PLS path modeling, was a non-predictive one.[15] Water quality parameters of five sites on Trout Greek in Colorado were connected by a path model. Predictive multiblock PLS models, to predict quality of adhesive tapes and quality and geographic origin of wines, were introduced by Frank and co-workers.[16,17] Furthermore, Frank and Kowalski suggested two different algorithms for the multiblock method.[18] In the averaging MBPLS algorithm a weighted average of the block scores of the separate $\mathbf{X}$ blocks, according to their correlation with the response, was used as the super score to describe the response score $\mathbf{u}$. The second algorithm was a stepwise algorithm. For each latent variable only the block with the highest correlation with $\mathbf{u}$ was selected to provide the super score $\mathbf{t}_T$.

A refinement of the PLS method with multiple blocks was introduced by Wold *et al.* at the Frankfurt conference.[13] The method was called hierarchical two-block predictive PLS or PLS-2H. This method could be used for investigation of complicated samples with many different physical, chemical and other variables to characterize the samples. This PLS method with multiple $\mathbf{X}$ blocks

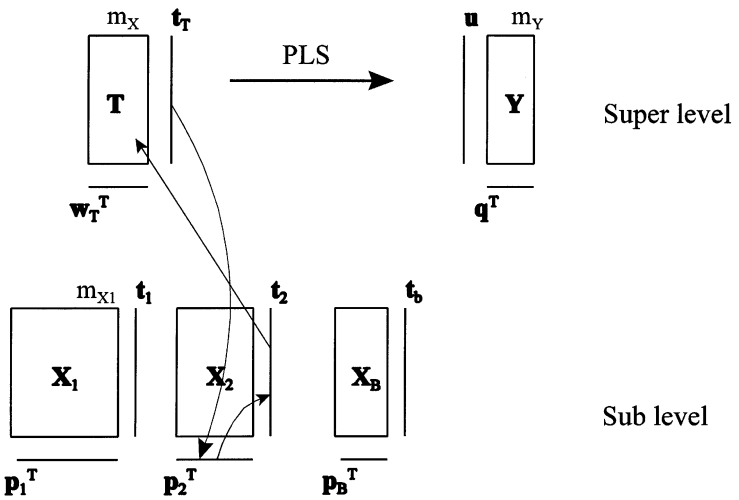

Figure 4. HPLS method. HPLS is an extension of the HPCA method. First an HPCA cycle is performed without normalization of the block scores $\mathbf{t}_b$. Then, instead of the PCA cycle on **T**, a PLS cycle is done between **T** and **Y** from which a super weight $\mathbf{w}_T^T$ and an updated super score $\mathbf{t}_T$ are obtained. These cycles are repeated until convergence of $\mathbf{t}_T$

was an extension of the CPCA method. First a CPCA cycle is performed on the multiple **X** blocks. Then a PLS cycle is done with the super block **T** and **Y**. In this PLS cycle both the super weight $\mathbf{w}_T$ and the **Y** weight **q** are normalized to length one. The CPCA and PLS cycles are repeated until convergence of the super score $\mathbf{t}_T$.

Several other variations of this hierarchical PLS (HPLS) method have been used since that paper. Slama[9] used the HPLS method for the modeling of a catalytic cracker and the subsequent fractionation section of a petroleum refinery. That specific HPLS algorithm used a normalization of the super score $\mathbf{t}_T$ instead of the normalization of the super weight $\mathbf{w}_T$ indicated by Wold *et al.*[13] Figure 4 shows the arrow scheme of this hierarchical PLS model with only one **Y** block. In 1996, Wold *et al.*[11] published a slightly different HPLS algorithm which was used for modeling process data from a catalytic cracker. Here also the super score $\mathbf{t}_T$ was normalized, but an additional step to assure orthogonality of the super scores was added. However, this latter step seems unnecessary, since the super scores are already orthogonal. The algorithm used by Slama[9] is presented in Appendix I.5 of this paper. All these algorithms[9,11,13] gave identical results and so they are all referred to in the following as HPLS.

In 1988, Wangen and Kowalski[19] introduced another PLS algorithm for multiple blocks that was based on an algorithm originally presented by Wold *et al.*[20] This multiblock PLS (MBPLS) algorithm could handle many types of pathway relationships between the blocks. Blocks can be left end blocks, which only predict subsequent blocks, right end blocks, which are only predicted by preceding blocks, or interior blocks, which are predicted by other blocks to their left but also predict blocks to the right of themselves. The arrow scheme of the most basic MBPLS method is shown in Figure 5 and the algorithm is presented in Appendix I.6. The main difference between this method and the HPLS method is that in MBPLS each block $\mathbf{X}_b$ is used in a PLS cycle with **Y** to calculate the block scores $\mathbf{t}_b$, whereas in HPLS $\mathbf{t}_b$ is calculated as in CPCA. Then the block scores are again combined in the super block **T** and a PLS cycle between **T** and **Y** is performed. These cycles are repeated until convergence of the super score $\mathbf{t}_T$. In this MBPLS algorithm both the block variable weights $\mathbf{w}_b$ and the super

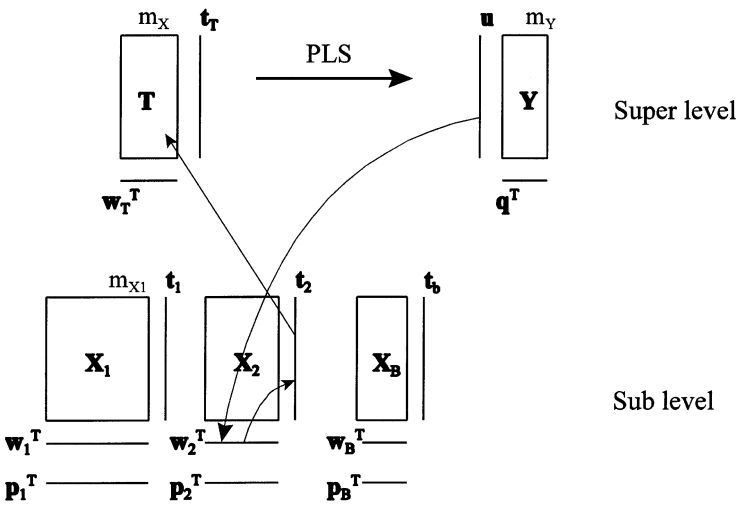

Figure 5. MBPLS method. A start score **u** is regressed on all blocks $\mathbf{X}_b$ to give the block variable weights $\mathbf{w}_b^{\mathrm{T}}$. The block variable weights are normalized to length one and multiplied through the blocks to give the block scores $\mathbf{t}_b$. The block scores are combined into the super block **T**. A PLS cycle between **T** and **Y** is performed to give the super weight $\mathbf{w}_{\mathrm{T}}^{\mathrm{T}}$, which is also normalized to length one, and the super score $\mathbf{t}_{\mathrm{T}}$. This is repeated until convergence of $\mathbf{t}_{\mathrm{T}}$

weight $\mathbf{w}_{\mathrm{T}}$ are normalized. Block variable loadings $\mathbf{p}_b$ are calculated for the deflation step. The MBPLS method has been used in several applications for the monitoring of chemical processes.[10,21] The advantage of the blocking approach is that in addition to a monitoring space for the whole process, one also obtains monitoring spaces for each process block. When a fault occurs in the process, this approach makes it much easier to detect, isolate and identify causes for the fault. Kourti *et al.*[21] used the MBPLS method to combine data from preprocessing conditions and other process information with the trajectories of the variables in the batch process to monitor the batch operation and predict the quality of the product in a polymerization process.

The MBPLS method of Wangen and Kowalski used the block scores $\mathbf{t}_b$ for the deflation of the data blocks $\mathbf{X}_b$.[19] This forces the block scores to be orthogonal, but the super scores $\mathbf{t}_{\mathrm{T}}$ become correlated. Westerhuis and Coenegracht[12] showed that by using the block score deflation method, some of the information in **X** may be lost in the deflation step. The entire variation in the direction of the block score $\mathbf{t}_b$ is subtracted from each block $\mathbf{X}_b$, even though only $w_{\mathrm{T}}(b)\cdot\mathbf{t}_b$ is used for prediction of the response **Y**. Here $w_{\mathrm{T}}(b)$ is the super weight of the corresponding block score $\mathbf{t}_b$, and $w_{\mathrm{T}}(b) \leq 1$. This undesirable effect becomes worse as the number of blocks increases. In that case each $w_{\mathrm{T}}(b)$ decreases, because $\Sigma(w_{\mathrm{T}}(b))^2 = 1$. Westerhuis and Coenegracht[12] suggested using the super score $\mathbf{t}_{\mathrm{T}}$ for the deflation step. Then one deflates only the information from the blocks $\mathbf{X}_b$ that was used for the prediction of the response **Y**. Now the super scores become orthogonal and the block scores are slightly correlated. A similar deflation that produces orthogonal super scores was used by Frank and Kowalski[18] in their averaging algorithm. The loss of information when deflating with the block scores can lead to poor performance. Therefore it is recommended that the super score deflation approach always be used in MBPLS.

To summarize, there are two differences between the MBPLS method and the HPLS method. The first one is the normalization. In MBPLS the block variable weights $\mathbf{w}_b$ and super weight $\mathbf{w}_{\mathrm{T}}$ are normalized to length one, whereas in the HPLS model only the super score $\mathbf{t}_{\mathrm{T}}$ is normalized. The

second difference is that in MBPLS **Y** is regressed on all descriptor blocks $\mathbf{X}_b$, whereas in HPLS **Y** is only regressed on the super block **T**. This causes the block scores to be different for the two methods.

In both HPLS and MBPLS the block scores are divided by the square root of the number of variables in the block to achieve block scaling, although additional scaling factors can be introduced to scale up or down the importance of various blocks.[11]

4. RELATIONSHIPS BETWEEN MULTIBLOCK AND SINGLE-BLOCK ALGORITHMS

For CPCA and MBPLS it is possible to calculate the scores and variable weights and loadings of the individual blocks and super block and the percentage explained variation in each block by using the standard PCA and PLS methods. When the variables are scaled with the same scaling factors, the super scores in the multiblock methods are exactly the same as the scores in the standard methods (see Appendix II). Therefore the block scores and block variable weights and loadings can be calculated from the results of the standard PCA and PLS analyses.

4.1. CPCA and PCA

Let **X** be a data matrix that will be divided into B blocks ($\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_B$) with block b having $m_{\mathrm{X}b}$ variables. To maintain the same scaling in PCA as is used in CPCA, we apply the following scaling to **X** for the PCA analysis:

$$\mathbf{X} = [\mathbf{X}_1/m_{\mathrm{X1}}^{1/2}, \ldots, \mathbf{X}_B/m_{\mathrm{X}B}^{1/2}] \tag{1}$$

The PCA method will be applied on **X** as defined by (1) and the CPCA method on $\mathbf{X}_1$, …, $\mathbf{X}_B$. Let $\mathbf{t}_{\mathrm{T}}$ be the super score of the CPCA method, $\mathbf{t}_b$ the block scores, $\mathbf{w}_{\mathrm{T}}$ the super weight and $\mathbf{p}_b$ the block variable loading of block $\mathbf{X}_b$. Let **t** be the score of the standard PCA method.

The super score of CPCA equals the score of PCA, i.e. $\mathbf{t}_{\mathrm{T}} \equiv \mathbf{t}$ (see Appendix II.1). The block variable loadings $\mathbf{p}_b$, block scores and super weight can be obtained from the **t** score of standard PCA:

$$\mathbf{p}_b = \mathbf{X}_b \cdot \mathbf{t}/\mathbf{t}^{\mathrm{T}} \cdot \mathbf{t} \tag{2}$$

$$\text{normalize } \mathbf{p}_b \text{ to } \| \mathbf{p}_b \| = 1 \tag{3}$$

$$\mathbf{t}_b = \mathbf{X}_b \cdot \mathbf{p}_b/m_{\mathrm{X}b}^{1/2} \tag{4}$$

$$\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_B] \tag{5}$$

$$\mathbf{w}_{\mathrm{T}} = \mathbf{T}^{\mathrm{T}} \cdot \mathbf{t}/\mathbf{t}^{\mathrm{T}} \cdot \mathbf{t} \tag{6}$$

The residuals for the variables in block $\mathbf{X}_b$ **are**

$$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_{\mathrm{T}} \cdot \mathbf{p}_{\mathrm{b}}^{\mathrm{T}} \tag{7}$$

The percentage explained for each block $\mathbf{X}_b$ equals

$$[1 - \ \mathrm{trace}(\mathbf{E}_b^{\mathrm{T}} \cdot \mathbf{E}_b)/\mathrm{trace}(\mathbf{X}_b^{\mathrm{T}} \cdot \mathbf{X}_b)] \times 100\% \tag{8}$$

If the block scaling is not applied in the CPCA method, i.e. $\mathbf{t}_b = \mathbf{X}_b \cdot \mathbf{p}_b$ in equation (4), then **X** should be scaled as [$\mathbf{X}_1$, …, $\mathbf{X}_B$] for the PCA method to obtain the same score as with CPCA.

4.2. MBPLS and PLS

In the same way, MBPLS results using the super score deflation method[12] can be obtained from the

standard PLS method. Let $\mathbf{X}$ be a data matrix of descriptors that is going to be divided into B blocks with m_{Xb} variables for block b. To maintain the same scaling in PLS as is used in MBPLS, we apply the following scaling for the PLS analysis:

$$\mathbf{X} = [\mathbf{X}_1/m_{X1}^{1/2}, ..., \mathbf{X}_B/m_{XB}^{1/2}]$$

The PLS method will be applied on $\mathbf{X}$ as defined above and the MBPLS method on $\mathbf{X}_1$, …, $\mathbf{X}_B$. Let $\mathbf{t}_T$ be the super score of the MBPLS method, $\mathbf{t}_b$ the block scores, $\mathbf{w}_T$ the super weight and $\mathbf{w}_b$ the block variable weight for block $\mathbf{X}_b$. Let $\mathbf{t}$ be the score of $\mathbf{X}$ and $\mathbf{u}$ the score of $\mathbf{Y}$ of the standard PLS method, where $\mathbf{Y}$ is a data matrix of response variables.

The PLS scores $\mathbf{t}$ and $\mathbf{u}$ equal the MBPLS super scores $\mathbf{t}_T$ and $\mathbf{u}$ (see Appendix II.2). From these results the block scores $\mathbf{t}_b$, block variable weights $\mathbf{w}_b$ and super weight $\mathbf{w}_T$ of the MBPLS method for each block can be calculated:

$$\mathbf{w}_b = \mathbf{X}_b^T \cdot \mathbf{u}/\mathbf{u}^T \cdot \mathbf{u} \tag{9}$$

$$\mathbf{t}_b = \mathbf{X}_b \cdot \mathbf{w}_b^T/m_{Xb}^{1/2} \tag{10}$$

$$\mathbf{T} = [\mathbf{t}_1, ..., \mathbf{t}_B] \tag{11}$$

$$\mathbf{w}_T = \mathbf{T}^T \cdot \mathbf{u}/\mathbf{u}^T \cdot \mathbf{u} \tag{12}$$

Just as with PCA, the explained variation for each block $\mathbf{X}_b$ can be calculated. Again, if no block scaling is used in the MBPLS method, i.e. $\mathbf{t}_b = \mathbf{X}_b \cdot \mathbf{w}_b$ in equation (10), then $\mathbf{X}$ should be scaled as [$\mathbf{X}_1$, …, $\mathbf{X}_B$] for the PLS method in order to obtain the same scores as with MBPLS.

4.3. Discussion

The above equivalences between CPCA and PCA and between MBPLS and PLS are valid only when there are no missing values. In the case of missing data in the PLS and PCA methods,[22] information on the correlations among the variables is used to estimate the scores. The blocking methods can only use the correlations among the variables that are in the specific block to give estimates for the scores, whereas the standard PCA and PLS methods make use of the correlation of all the variables for the score estimation. In cases where variables in different blocks correlate, missing values will decrease the performance of the blocking methods as compared with the standard PCA and PLS methods.

The above analysis has shown the equivalence between CPCA and ordinary PCA and between MBPLS (with super score deflation) and ordinary PLS. However, no such equivalence exists with HPCA and HPLS.

The HPLS method of Wold *et al.*[11] and MBPLS with super score deflation[12] can also be used for the modeling of several $\mathbf{Y}$ blocks. The case of several $\mathbf{Y}$ blocks might be interesting when the number of response variables is large and additional knowledge for blocking is present. For example, when a chemical process produces several products, characteristics of the products can be placed into different $\mathbf{Y}$ blocks to improve the interpretability. However, just as in the case of one $\mathbf{Y}$ block, in the MBPLS method the same results can be obtained by first calculating the standard PLS model using the appropriate variable scaling. Instead of blocking the response into several blocks, it is better to first run an ordinary PLS and then calculate the response block scores and block variable weights. This has the same advantages for calculation speed and presence of missing values as mentioned before.

5. SIMULATION STUDIES AND DISCUSSION

In this section, consensus principal component analysis (CPCA) is compared with hierarchical PCA

(HPCA) and the multiblock projections to latent structures (MBPLS) method is compared with hierarchical PLS (HPLS) as used by Slama.[9] Both multiblock PLS models are considered with only one response block **Y**.

The differences in the results obtained when these algorithms are applied to the same set of data are illustrated with examples from simple simulations and real data sets.

5.1. Consensus PCA and hierarchical PCA

The difference between the two methods is caused by the different way of normalization in the corresponding algorithm. In both algorithms the super score will be the direction most dominant in the consensus block **T**. However, because in HPCA the block scores are normalized to length one, the algorithm searches for the most dominant direction in these normalized scores. In CPCA the block scores enter **T** as they are calculated for each block and therefore the super score will just be the direction most dominant in the block scores. Differences between the methods can be expected when a strong direction exists in only a single block (Example 1). When the directions are spread among the blocks, the methods are expected to give similar results (Example 2).

Example 1

In the first simulation we consider four blocks of five variables where one block contains a strong direction that is not available in the other blocks:

$$\mathbf{X}_1 = [\mathbf{d}_1\ \ \mathbf{d}_1\ \ \mathbf{d}_1\ \ \mathbf{d}_1\ \ \mathbf{d}_1], \qquad \mathbf{X}_2 = [\mathbf{d}_2\ \ \text{randn}(4)]$$
$$\mathbf{X}_3 = [\mathbf{d}_2\ \ \text{randn}(4)], \qquad \mathbf{X}_4 = [\mathbf{d}_2\ \ \text{randn}(4)]$$

Here randn(4) stands for four columns of normally distributed random data which are different in each of the blocks. All blocks have 50 observations. Block $\mathbf{X}_1$ consists of only one direction $\mathbf{d}_1$. Blocks $\mathbf{X}_2$ to $\mathbf{X}_4$ all have one common direction $\mathbf{d}_2$ with four random variables. All directions are selected to be orthogonal. Twenty per cent of random noise was added to each variable in each of the blocks. All variables in all the blocks are mean-centered and scaled to unit variance. Table 1 shows the cumulative percentages of explained variation of the first two PCs for the four blocks for the CPCA and HPCA methods

The first super score in the CPCA method, just as we would expect from a standard PCA, is the direction most dominant in all the data. This is direction $\mathbf{d}_1$ and therefore block $\mathbf{X}_1$ is described completely by this super score. The second super score represents direction $\mathbf{d}_2$ which accounts for 20% of the variation in blocks $\mathbf{X}_2$ to $\mathbf{X}_4$. In the HPCA method the first super score is direction $\mathbf{d}_2$ which is present in three blocks and their block scores and therefore more available in the super block than direction $\mathbf{d}_1$. Direction $\mathbf{d}_1$ does not appear in the second super score either, because the block scores for the second component for $\mathbf{X}_2$ to $\mathbf{X}_4$, which are now random numbers, are set by the algorithm to be equally important in the super block. Considering these results, the hierarchical PCA

Table 1. Cumulative percentage explained variation of four blocks with CPCA and HPCA for the first two PCs

		Block $\mathbf{X}_1$	Block $\mathbf{X}_2$	Block $\mathbf{X}_3$	Block $\mathbf{X}_4$
CPCA	PC1	96·7	0·1	0·2	0·9
	PC2	96·8	19·8	19·8	20·3
HPCA	PC1	0·1	19·7	19·6	19·5
	PC2	6·3	26·2	25·7	30·4

Table 2. Cumulative percentage explained variation of four blocks with simulated data for the first two PCs for CPCA and HPCA

		Block $\mathbf{X}_1$	Block $\mathbf{X}_2$	Block $\mathbf{X}_3$	Block $\mathbf{X}_4$
CPCA	PC1	58·1	38·6	19·5	0·1
	PC2	77·3	57·6	58·1	0·4
HPCA	PC1	56·4	38·4	20·4	0·2
	PC2	77·0	57·6	57·9	0·6

seems to emphasize the consensus more than the consensus PCA, but the strong direction in only one of the blocks is never found. In CPCA the strong belief of a single judge is preferred to the weak beliefs of many.

Example 2

The second simulation shows a case where the two directions are spread over the blocks. In this case the results of the two methods are nearly the same. Now the two directions $\mathbf{d}_1$ and $\mathbf{d}_2$ are spread among three of the four blocks and are not present in the last block $\mathbf{X}_4$:

$$\mathbf{X}_1 = [\mathbf{d}_1\ \mathbf{d}_1\ \mathbf{d}_1\ \mathbf{d}_2\ \text{randn}(1)], \qquad \mathbf{X}_2 = [\mathbf{d}_1\ \mathbf{d}_1\ \mathbf{d}_2\ \text{randn}(2)]$$

$$\mathbf{X}_3 = [\mathbf{d}_1\ \mathbf{d}_2\ \mathbf{d}_2\ \text{randn}(2)], \qquad \mathbf{X}_4 = [\text{randn}(5)]$$

Table 2 shows the cumulative percentage explained variation of the second simulation for all blocks for the CPCA and HPCA methods.

The percentage explained variations are nearly the same for both methods and equal to that expected given the frequency of $\mathbf{d}_1$ and $\mathbf{d}_2$ in the four data sets. In both methods the first super score is similar to direction $\mathbf{d}_1$ and the second super score to direction $\mathbf{d}_2$. Direction $\mathbf{d}_1$ accounts for 60%, 40%, 20% and 0% and direction $\mathbf{d}_2$ for 20%, 20%, 40% and 0% of the variation in the four blocks respectively.

The super weights $\mathbf{w}_T$ of both methods show how the blocks behave compared to each other. Figure 6 shows the relative importance of the four blocks for CPCA and HPCA according to their super weights. HPCA just signals that the direction of the super score is present in the block. Therefore blocks $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ get the same super weight of 1·0 for PC1 and PC2, because the direction of the super score is present in the blocks, irrespective of how much of this direction is present. In CPCA the super weight depends on how much of the super score is present in the blocks.

Table 3 shows the correlations between the consensus directions $\mathbf{d}_1$ and $\mathbf{d}_2$ and the super score $\mathbf{t}_T$ and block scores $\mathbf{t}_1$ to $\mathbf{t}_4$ of the first and second PCs. The correlations between the consensus directions and the scores are somewhat higher for CPCA than for HPCA, which would indicate that the consensus directions are estimated more accurately by CPCA; this is because in CPCA the block scores maintain their length in the super block.

Consensus PCA and hierarchical PCA are both methods that deal with multiple blocks of variables based on the same objects to find latent directions in the data. The above simulations show that both methods give similar results when the latent directions are spread among the blocks. However, when a very strong direction exists in only one of the blocks, CPCA finds this direction, whereas the HPCA algorithm does not find it. The super weight in CPCA shows how the blocks are related to each other. The HPCA super weight only shows whether the direction of the super score is present in the block.

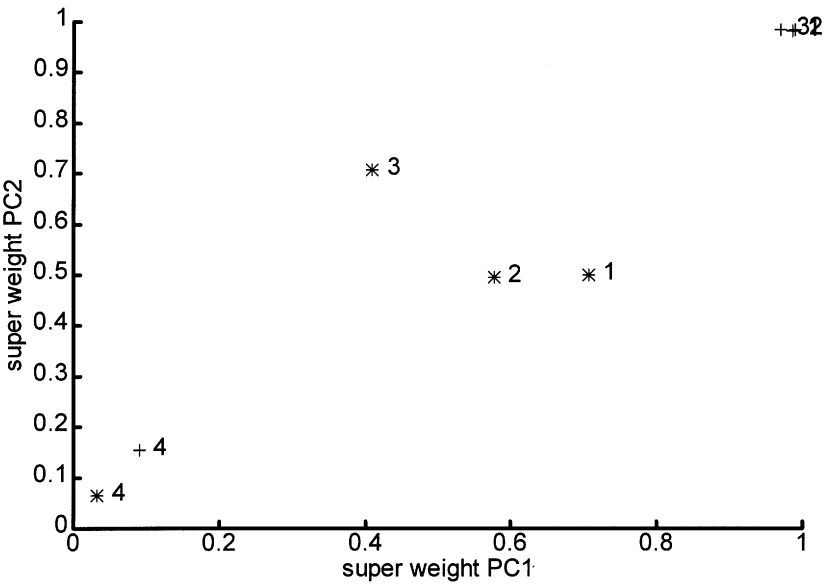

Figure 6. Super weight $\mathbf{w}_T$ for four blocks with CPCA (∗) and HPCA (+). In HPCA $\mathbf{w}_T$ is nearly one for all blocks in which the consensus direction is present. CPCA discriminates between the blocks by the amount the consensus score is present in the blocks

Table 3. Correlations between directions $\mathbf{d}_1$ and $\mathbf{d}_2$ and super score $\mathbf{t}_T$ and block scores $\mathbf{t}_1$ to $\mathbf{t}_4$ for PC1 and PC2

		$\mathbf{t}_T$	$\mathbf{t}_1$	$\mathbf{t}_2$	$\mathbf{t}_3$	$\mathbf{t}_4$
CPCA	$\mathbf{d}_1$ (PC1)	0·990	0·990	0·981	0·961	0·067
	$\mathbf{d}_2$ (PC2)	0·987	0·973	0·971	0·980	0·080
HPCA	$\mathbf{d}_1$ (PC1)	0·966	0·986	0·971	0·890	0·054
	$\mathbf{d}_2$ (PC2)	0·958	0·948	0·946	0·952	0·078

5.2. Multiblock PLS and hierarchical PLS

In this subsection we compare HPLS with MBPLS. Both implementations of MBPLS, block score deflation[19] and super score deflation,[12] will be considered. The super score is calculated in exactly the same way by both MBPLS methods, but the results differ after the first latent variable because of the different deflation methods.

The block scores in MBPLS are chosen to have affinity with the response variable **Y**, whereas the block scores in HPLS are chosen to have affinity with the consensus score $\mathbf{t}_T$. Furthermore, the normalization in HPLS is done on the super score, whereas for MBPLS a normalization of the block variable weights and super weight is used. In Examples 3 and 4 it is shown that HPLS selects only one or a few descriptor blocks to predict the response. Example 3 shows that the block score deflation MBPLS uses more variation of the **X** blocks but predicts less of **Y**. In Example 4, simulated data from a hypothetical process are used to show that the objective of HPLS is not predictability of **Y** but rather explanation of **X**.

Table 4. Results of modeling of LDPE data with HPLS and MBPLS methods. The cumulative percentages of explained variation for both **X** blocks and **Y** are given, as is the CSV/SD factor for all three multiblock PLS methods

	HPLS				MBPLS (block)				MBPLS (super)			
	$\%\mathbf{X}_1$	$\%\mathbf{X}_2$	%**Y**	CSV/SD	$\%\mathbf{X}_1$	$\%\mathbf{X}_2$	%**Y**	CSV/SD	$\%\mathbf{X}_1$	$\%\mathbf{X}_2$	%**Y**	CSV/SD
LV1	8	**41**	46	0·74	39	41	61	0·64	24	30	61	0·64
LV2	**44**	44	65	0·82	57	56	83	0·72	37	48	87	0·63
LV3	64	53	76	0·94	69	70	86	0·95	50	61	91	0·89
LV4	78	66	**82**	0·93	87	89	90	0·91	60	75	**94**	0·86

Example 3. LDPE data

The first comparison between PLS with multiple blocks is done with data from a multizone tubular reactor for the production of low-density polyethylene (LDPE), which was described by MacGregor *et al*.[10] The process variables are divided into two blocks of eight variables. The response consists of five product variables that characterize the LDPE. A review on LDPE processes can be found in Reference 23 and details of the simulation can be found in Reference 24. Table 4 shows the results of the modeling for HPLS, MBPLS with block score deflation and MBPLS with super score deflation. The cumulative percentages of explained variation for both **X** blocks and **Y** are given for the first four latent variables (LV1 to LV4). Furthermore, cross-validation results are indicated by the CSV/SD factor, which is indicative of the predictive abilities of each principal component. It is the ratio $\mathrm{PRESS}_r/\mathrm{RSS}_{r-1}$ and values larger than unity suggest that the rth component does not improve the prediction power of the model.[25] Here PRESS_r is the sum of the squared prediction errors after the rth latent variable and RSS_{r-1} is the residual sum of squares after the $(r-1)$th latent variable.

Table 4 clearly shows that HPLS tends to select only one of the blocks to use information from, in each latent variable. In the first latent variable LV1, 41% of block $\mathbf{X}_2$ is used and only 8% of block $\mathbf{X}_1$ while in LV2 almost all information comes from block $\mathbf{X}_1$. The MBPLS methods use both **X** blocks in all LVs. This may improve the interpretability of the HPLS model, because only one **X** block is important for each LV, but the amount of explained **Y** is only 82% after four components, whereas MBPLS (super score deflation method) explains 94% of the variation in **Y**. The block score deflation MBPLS method describes more of the descriptor blocks, but the response is predicted less than with the super score deflation method. This illustrates the previously described problem with deflation using block scores. Only part of the block scores $(w_{\mathrm{T}}(b)\cdot\mathbf{t}_b)$ is used for the prediction of **Y** in each component and yet the entire direction $\mathbf{t}_b$ is removed from $\mathbf{X}_b$ at the deflation stage. As a result, variation in each $\mathbf{X}_b$ is removed by deflation prior to its use in predicting **Y**; hence the relatively high percentage $\mathbf{X}_b$ explained and the low percentage **Y** explained.

Table 5 shows the correlations between the super score $\mathbf{t}_{\mathrm{T}}$ and **u** and the block scores $\mathbf{t}_1$ and $\mathbf{t}_2$ for the MBPLS (super score deflation method) and HPLS methods for LV1 and LV2. In HPLS the

Table 5. Correlations between $\mathbf{t}_{\mathrm{T}}$, $\mathbf{t}_1$ and $\mathbf{t}_2$, and $\mathbf{t}_{\mathrm{T}}$ and **u** for MBPLS (super) and HPLS methods

		$\mathbf{t}_1$ MBPLS	$\mathbf{t}_2$ MBPLS	$\mathbf{t}_{\mathrm{T}}$ MBPLS	$\mathbf{t}_1$ HPLS	$\mathbf{t}_2$ HPLS	$\mathbf{t}_{\mathrm{T}}$ HPLS
$\mathbf{t}_{\mathrm{T}}$	LV1	0·77	0·85	1	0·50	**1·00**	1
u	LV1	0·73	0·82	**0·96**	0·59	0·80	**0·82**
$\mathbf{t}_{\mathrm{T}}$	LV2	0·97	0·98	1	**1·00**	0·40	1
u	LV2	0·92	0·88	**0·92**	0·90	0·58	**0·90**

Table 6. Results of modeling of simulated data with HPLS and MBPLS. The cumulative percentage explained sum of squares for the **X** blocks and **Y** and the CSV/SD factor are given for LV1 to LV4

		CSV/SD	%$\mathbf{X}_1$	%$\mathbf{X}_2$	%$\mathbf{X}_3$	%**Y**
HPLS	LV1	1·01	0	**78**	**73**	2
	LV2	0·90	**82**	79	74	26
	LV3	1·01	82	**99**	**99**	**29**
	LV4	0·24	**99**	99	99	96
MBPLS (super)	LV1	0·78	59	14	15	54
	LV2	0·69	68	53	62	79
	LV3	0·50	97	79	78	**95**
	LV4	0·94	99	99	99	96

correlation of one of the two block scores to the super score is 1·00, whereas the correlation with the other block is rather low. The correlations of the super scores $\mathbf{t}_T$ with $\mathbf{u}$ (0·82 and 0·90) are less for HPLS than for the MBPLS method (0·96 and 0·92).

Example 4. Simulated hypothetical process

The last example is a very simple simulation of a hypothetical process consisting of two units. The data available are temperatures from the first unit ($\mathbf{X}_1$), temperatures from the second unit ($\mathbf{X}_2$) and pressures from the second unit ($\mathbf{X}_3$). $\mathbf{X}_1$ can be described by two latent directions $\mathbf{d}_1$ and $\mathbf{d}_2$. $\mathbf{X}_2$ and $\mathbf{X}_3$ can also be described by two directions $\mathbf{d}_3$ and $\mathbf{d}_4$ for both blocks. $\mathbf{X}_1$ consists of three columns and $\mathbf{X}_2$ and $\mathbf{X}_3$ consist of four columns as follows:

$$\mathbf{X}_1 = [0{\cdot}50\mathbf{d}_1 + 0{\cdot}50\mathbf{d}_2,\ 0{\cdot}75\mathbf{d}_1 + 0{\cdot}25\mathbf{d}_2,\ 0{\cdot}25\mathbf{d}_1 + 0{\cdot}75\mathbf{d}_2]$$

$$\mathbf{X}_2 = [0{\cdot}50\mathbf{d}_3 + 0{\cdot}50\mathbf{d}_4,\ 0{\cdot}25\mathbf{d}_3 + 0{\cdot}75\mathbf{d}_4,\ 0{\cdot}75\mathbf{d}_3 + 0{\cdot}25\mathbf{d}_4,\ 0{\cdot}99\mathbf{d}_3 + 0{\cdot}01\mathbf{d}_4]$$

$$\mathbf{X}_3 = [0{\cdot}50\mathbf{d}_3 + 0{\cdot}50\mathbf{d}_4,\ 0{\cdot}75\mathbf{d}_3 + 0{\cdot}25\mathbf{d}_4,\ 0{\cdot}25\mathbf{d}_3 + 0{\cdot}75\mathbf{d}_4,\ 0{\cdot}01\mathbf{d}_3 + 0{\cdot}99\mathbf{d}_4]$$

The quality of the product mainly depends on the one direction of the temperatures of the first process unit:

$$\mathbf{Y} = \mathbf{d}_1 + 0{\cdot}2\mathbf{d}_3$$

Ten per cent of normally distributed noise was added to all variables in all three blocks and to the response. The latent directions $\mathbf{d}_1$, $\mathbf{d}_2$, $\mathbf{d}_3$ and $\mathbf{d}_4$ were nearly mutually orthogonal. All variables in all blocks were mean-centered and scaled to unit variance. Table 6 shows the results of the modeling with HPLS and MBPLS (super score deflation) for the first four latent vectors (LV1 to LV4).

The results in Table 6 show that in the first latent variable HPLS describes the strong information from blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ which has only a weak relation with **Y**, instead of the smaller amount of information from block $\mathbf{X}_1$ which has a strong relation with **Y**. The CSV/SD parameter is even higher than 1·00 for HPLS for LV1 and LV3, suggesting that these LVs do not improve the model for **Y**. After four LVs all the information in the **X** blocks is described, but only in the last LV does HPLS make a large jump to explain 96% of **Y**. This means that a weak latent variable is used to explain most of the variation of the response variable.

Wold *et al.*[11] suggested that in order to estimate the number of model components to collect in HPLS, 'the normal procedure is to first run an "unblocked" model and then use the same number of components in the hierarchical model as in the corresponding "unblocked" model.' However, our findings contradict this statement. It is not always sufficient for HPLS to use the same number of LVs

Table 7. Results of modeling of simulated data with HPLS and MBPLS with only two blocks. The cumulative percentage explained sum of squares for the **X** blocks and **Y** and the CSV/SD factor are given for LV1 to LV4

		CSV/SD	$\%\mathbf{X}_1$	$\%\mathbf{X}_2\mathbf{X}_3$	$\%\mathbf{Y}$
HPLS	LV1	1·01	0	**76**	2
	LV2	0·91	**82**	76	26
	LV3	0·92	82	**99**	29
	LV4	0·24	**99**	99	96
MBPLS (super)	LV1	0·73	70	4	56
	LV2	0·45	94	6	95
	LV3	1·03	98	77	96
	LV4	1·04	99	98	96

as for an unblocked PLS model. MBPLS, which equals the standard PLS, only needs three LVs to explain 95% of the variation of **Y**, whereas HPLS explained only 29% of the variance after three LVs.

In summary, HPLS acts more like an HPCR method. This might be expected from the way the block scores are calculated. It means that HPLS prefers to describe the variation in **X** rather than the covariance between **X** and **Y**. In agreement with this, HPLS sometimes may need more LVs to reach the same amount of explained variation in **Y** than MBPLS. HPLS also acts like a block selector. In the first and third LVs only blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ are used to predict **Y**, whereas in the second and fourth LVs only block $\mathbf{X}_1$ is used. In MBPLS the blocks are combined in all LVs to give the best predicting power.

One might argue with the blocking chosen in this example. Because blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ carry somewhat similar information (because of the relation between temperature and pressure in the second unit), they should not be divided. Table 7 shows the results of the same data when blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ are combined into one new block of eight variables. Combining blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ hardly makes any difference to the percentage explained variation for the **X** blocks and also for the response variable. The amount of explained variation for blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ is just combined into one new block. However, because of the block scaling, they have even less influence than they had as two separate blocks.

HPLS shows the same convergence problems as HPCA. However, in this case no extra normalization of the block scores was applied. Therefore for each of the latent vectors the final solution of the super score $\mathbf{t}_T$ favors one of the blocks. If in Example 4 a column of block $\mathbf{X}_1$ was used as the initial guess for the super score, then the final super score would favor the first block and hardly anything of blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ would be explained. If any column of block $\mathbf{X}_2$ or $\mathbf{X}_3$ was selected as the initial guess of the super score, then the same results would be obtained as the ones presented in Table 6. This explains why HPLS acts as a block selector. To force the algorithm to 'a specific solution' in these examples, we chose the initial guess to be the eigenvector of $\mathbf{X}^T\mathbf{X}$ that corresponds to the largest eigenvalue. However, the objective function of HPLS is not clear as it is for MBPLS. The latter has the same objective functions as standard PLS, i.e. maximization of the covariance between $\mathbf{t}_T$ and $\mathbf{u}$, which is achieved by selecting the eigenvector of $\mathbf{X}\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T$ that corresponds to the largest eigenvalue as $\mathbf{t}_T$ (Appendix II). As mentioned above, HPLS also acts like a hierarchical principal component regression (HPCR) method. The scores of HPLS hardly change when the response variable **Y** is replaced by any random variable. Thus the response has almost no influence on the selection of the scores. This explains in Example 4 why latent vectors are selected that hardly describe the response variable.

6. CONCLUSIONS

In this paper several multiblock and hierarchical PCA and PLS methods that have recently appeared in the literature have been compared. The same notation was used to clearly show the similarities and differences between the methods. Some convergence problems were indicated. On some occasions hierarchical PCA and hierarchical PLS converged to different solutions depending on the initial guess of the super score. The final super score mostly describes the block that is closest to the initial guess. In the case of HPCA, where the block scores are also normalized, often an average between the solutions is found. Still, the objective function of the hierarchical PCA and PLS methods is not clear, as opposed to CPCA and MBPLS where the objective functions are the same as those of standard PCA and PLS respectively.

Hierarchical PCA and consensus PCA have been compared. For HPCA the initial guess of the super score was set to be equal to the eigenvector of $\mathbf{X}^T\mathbf{X}$ that corresponds to the largest eigenvalue. When the consensus directions are spread over the blocks, the methods produce similar results. HPCA disregards a very strong direction from a single block to favor a weaker direction that is present in multiple blocks; in fact, it never selects this strong direction. When all blocks are orthogonal and no consensus can be found, CPCA looks for the block score with the highest variance, whereas HPCA gives the mean of all the block scores, regardless of their length. In the case of only two blocks HPCA gives the exact mean of the two block scores as the consensus, whereas the consensus score of CPCA is a weighted average of both block scores, dependent on their length.

Hierarchical PLS and multiblock PLS were also compared. Two options of deflation were compared for the MBPLS method. The block score deflation method performs worse than the super score deflation method because of the removal of information from the descriptor blocks that is never used for prediction. It is therefore recommended that the super score deflation method always be used in MBPLS. HPLS acts like a block selection method. For each latent variable one or a few blocks are selected to be used for the regression and the other blocks are almost ignored. This may facilitate the interpretation of the model, but the method performs less well for prediction of the response. HPLS looks more like an HPCR method. The super score hardly changes when the response is replaced by any random variable. As a result, HPLS may need more latent variables to predict the same percentage of variance of $\mathbf{Y}$ as MBPLS.

The results of both CPCA and MBPLS can also be calculated by using the standard PCA and PLS methods respectively when the same variable scaling is used. This requires less computation and gives better estimation for the scores in the case where there are missing data, because correlations among all variables can be used for estimation of the scores instead of only the correlations among the variables in the specific block. Therefore we recommend that one should use the standard PCA and PLS methods with appropriate scaling of the variables to obtain the model and then calculate the block scores, block variable weights, super weight and percentage variation explained by each block.

APPENDIX I

I.1. Principal Component Analysis (PCA)

Transform, center and scale
For each dimension
  choose start $\mathbf{t}$
  loop until convergence of $\mathbf{t}$
    $\mathbf{p} = \mathbf{X}^T\cdot\mathbf{t}/\mathbf{t}^T\cdot\mathbf{t}$ % $\mathbf{X}$ loading
    normalize $\mathbf{p}$ to $\|\mathbf{p}\| = 1$
    $\mathbf{t} = \mathbf{X}\cdot\mathbf{p}/\mathbf{p}^T\cdot\mathbf{p}$ % $\mathbf{X}$ score


```
  end
  DEFLATION
  X = X − t·p^T
end
```

I.2. Projections to Latent Structures (PLS)

```
Transform, center and scale
For each dimension
  choose start u
  loop until convergence of t
    w = X^T·u/u^T·u                 % X weight
    normalize w to ||w|| = 1
    t = X·w/w^T·w                   % X score
    q = Y^T·t/t^T·t                 % Y weight
    u = Y·q                         % Y score
  end
  DEFLATION
  p = X^T·t/t^T·t                   % X loading
  X = X − t·p^T
  Y = Y − t·q^T
end
```

I.3. Consensus PCA (CPCA)

```
Transform, center and scale
For each dimension
  choose start t_T
  loop until convergence of t_T
    p_b = X_b^T·t_T/t_T^T·t_T       % X_b block variable loadings
    normalize p_b to ||p_b|| = 1
    t_b = X_b·p_b/m_Xb^1/2          % X_b block scores (block scaling)
    T = [t_1…t_B]                   % Combine all block scores in T
    w_T = T^T·t_T/t_T^T·t_T         % Super weight
    normalize w_T to ||w_T|| = 1
    t_T = T·w_T                     % Super score
  end
  DEFLATION
  p_b = X_b^T·t_T/t_T^T·t_T
  X_b = X_b − t_T·p_b^T
end
```

Some scaling factor additional to the block scaling can be introduced. Then in the loop $\mathbf{t}_b$ becomes $\mathbf{t}_b = s_b \cdot \mathbf{X}_b \cdot \mathbf{p}_b / m_{\mathrm{X}b}^{1/2}$, where s_b is the additional scaling factor for block $\mathbf{X}_b$.

I.4. Hierarchical PCA (HPCA)

```
Transform, center and scale
For each dimension
  choose t_T to be the eigenvector of X^TX corresponding to the largest eigenvalue
```


loop until convergence of $\mathbf{t}_T$
$\mathbf{p}_b = \mathbf{X}_b^T \cdot \mathbf{t}_T / \mathbf{t}_T^T \cdot \mathbf{t}_T$ % $\mathbf{X}_b$ block variable loadings
$\mathbf{t}_b = \mathbf{X}_b \cdot \mathbf{p}_b$ % $\mathbf{X}_b$ block scores
normalize $\mathbf{t}_b$ to $\|\mathbf{t}_b\| = 1$
$\mathbf{T} = [\mathbf{t}_1 \ldots \mathbf{t}_B]$ % Combine all block scores in $\mathbf{T}$
$\mathbf{w}_T = \mathbf{T}^T \cdot \mathbf{t}_T / \mathbf{t}_T^T \cdot \mathbf{t}_T$ % Super weight
$\mathbf{t}_T = \mathbf{T} \cdot \mathbf{w}_T$ % Super score
normalize $\mathbf{t}_T$ to $\|\mathbf{t}_T\| = 1$
end
DEFLATION
$\mathbf{X}_b = \mathbf{X}_b - \mathbf{t}_T \cdot \mathbf{p}_b^T$
end

I.5. Hierarchical PLS (HPLS) (for one **Y** block)

Transform, center and scale
For each dimension
choose $\mathbf{t}_T$ to be the eigenvector of $\mathbf{X}^T\mathbf{X}$ corresponding to the largest eigenvalue
loop until convergence of $\mathbf{t}_T$
$\mathbf{p}_b = \mathbf{X}_b^T \cdot \mathbf{t}_T / \mathbf{t}_T^T \cdot \mathbf{t}_T$ % $\mathbf{X}_b$ block variable loadings
$\mathbf{t}_b = \mathbf{X}_b \cdot \mathbf{p}_b / m_{Xb}^{1/2}$ % $\mathbf{X}_b$ block scores (block scaling)
$\mathbf{T} = [\mathbf{t}_1 \ldots \mathbf{t}_B]$ % Combine all $\mathbf{X}_b$ block scores in $\mathbf{T}$
$\mathbf{q} = \mathbf{Y}^T \cdot \mathbf{t}_T / \mathbf{t}_T^T \cdot \mathbf{t}_T$ % $\mathbf{Y}$ weight
$\mathbf{u} = \mathbf{Y} \cdot \mathbf{q} / \mathbf{q}^T \cdot \mathbf{q}$ % $\mathbf{Y}$ score
$\mathbf{w}_T = \mathbf{T}^T \cdot \mathbf{u} / \mathbf{u}^T \cdot \mathbf{u}$ % $\mathbf{X}$ super weight
$\mathbf{t}_T = \mathbf{T} \cdot \mathbf{w}_T / \mathbf{w}_T^T \cdot \mathbf{w}_T$ % $\mathbf{X}$ super score
normalize $\mathbf{t}_T$ to $\|\mathbf{t}_T\| = 1$
end
DEFLATION
$\mathbf{X}_b = \mathbf{X}_b - \mathbf{t}_T \cdot \mathbf{p}_b^T$ % Deflation of $\mathbf{X}_b$ with $\mathbf{X}$ super score
$\mathbf{Y} = \mathbf{Y} - \mathbf{t}_T \cdot \mathbf{q}^T$ % Deflation of $\mathbf{Y}$ with $\mathbf{X}$ super score
end

Some scaling factor additional to the block scaling can be introduced. Then in the loop $\mathbf{t}_b$ becomes $\mathbf{t}_b = s_b \cdot \mathbf{X}_b \cdot \mathbf{p}_b / m_{Xb}^{1/2}$, where s_b is the additional scaling factor for block $\mathbf{X}_b$.

I.6. Multiblock PLS (MBPLS) (for one **Y** block only)

Transform, center and scale data
For each dimensions
take $\mathbf{u}$ = some column of $\mathbf{Y}$
loop until convergence of $\mathbf{t}_T$
$\mathbf{w}_b = \mathbf{X}_b^T \cdot \mathbf{u} / \mathbf{u}^T \cdot \mathbf{u}$ % $\mathbf{X}_b$ block variable weights
normalize $\mathbf{w}_b$ to $\|\mathbf{w}_b\| = 1$
$\mathbf{t}_b = \mathbf{X}_b \cdot \mathbf{w}_b / m_{Xb}^{1/2}$ % $\mathbf{X}_b$ block score
$\mathbf{T} = [\mathbf{t}_1 \ldots \mathbf{t}_B]$ % Combine all blocks scores in $\mathbf{T}$
$\mathbf{w}_T = \mathbf{T}^T \cdot \mathbf{u} / \mathbf{u}^T \cdot \mathbf{u}$ % $\mathbf{X}$ super weight
normalize $\mathbf{w}_T$ to $\|\mathbf{w}_T\| = 1$
$\mathbf{t}_T = \mathbf{T} \cdot \mathbf{w}_T / \mathbf{w}_T^T \cdot \mathbf{w}_T$ % $\mathbf{X}$ super score

$\mathbf{q} = \mathbf{Y}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}}$ % **Y** weight
$\mathbf{u} = \mathbf{Y} \cdot \mathbf{q} / \mathbf{q}^{\mathrm{T}} \cdot \mathbf{q}$ % **Y** score
end
DEFLATION
$\mathbf{p}_{b\mathrm{T}} = \mathbf{X}_b^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}}$ % Deflation with **X** super score
$\mathbf{X}_b = \mathbf{X}_b - \mathbf{t}_{\mathrm{T}} \cdot \mathbf{p}_{b\mathrm{T}}^{\mathrm{T}}$ % Westerhuis and Coenegracht[12]
$\mathbf{Y} = \mathbf{Y} - \mathbf{t}_{\mathrm{T}} \cdot \mathbf{q}^{\mathrm{T}}$
or
$\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}} \cdot \mathbf{t}_b / \mathbf{t}_b^{\mathrm{T}} \cdot \mathbf{t}_b$ % Deflation with $\mathbf{X}_b$ block score
$\mathbf{X}_b = \mathbf{X}_b - \mathbf{t}_b \cdot \mathbf{p}_b^{\mathrm{T}}$ % Wangen and Kowalski[19]
$\mathbf{Y} = \mathbf{Y} - \mathbf{t}_{\mathrm{T}} \cdot \mathbf{q}^{\mathrm{T}}$
end

Some scaling factor additional to the block scaling can be introduced. Then in the loop $\mathbf{t}_b$ becomes $\mathbf{t}_b = s_b \cdot \mathbf{X}_b \cdot \mathbf{w}_b / m_{\mathrm{X}b}^{1/2}$, where s_b is the additional scaling factor for block $\mathbf{X}_b$.

APPENDIX II

II.1. Proof that **t** score of PCA equals $\mathbf{t}_{\mathrm{T}}$ score of CPCA (in case of no missing data)

PCA on **X** ($\mathbf{X} = [\mathbf{X}_1 / m_{\mathrm{X1}}^{1/2} \;\; \mathbf{X}_2 / m_{\mathrm{X2}}^{1/2}] = [\mathbf{X}_1^* \;\; \mathbf{X}_2^*]$).

$$\mathbf{t} = \mathbf{X} \cdot \mathbf{p} \tag{13}$$

$$\mathbf{t} = \mathbf{X}_1^* \cdot \mathbf{p}_1^{\mathrm{o}} + \mathbf{X}_2^* \cdot \mathbf{p}_2^{\mathrm{o}} \tag{14}$$

where $\mathbf{p}_1^{\mathrm{o}}$ is the part of **p** belonging to the variables in $\mathbf{X}_1^*$

$$\mathbf{t} = (\mathbf{X}_1^* \cdot \mathbf{X}_1^{*\mathrm{T}} \cdot \mathbf{t} + \mathbf{X}_2^* \cdot \mathbf{X}_2^{*\mathrm{T}} \cdot \mathbf{t}) / \mathbf{t}^{\mathrm{T}} \cdot \mathbf{t} \tag{15}$$

$$\mathbf{t} = (\mathbf{X}_1^* \cdot \mathbf{X}_1^{*\mathrm{T}} + \mathbf{X}_2^* \cdot \mathbf{X}_2^{*\mathrm{T}}) \mathbf{t} / \mathbf{t}^{\mathrm{T}} \cdot \mathbf{t} \tag{16}$$

At convergence $(\mathbf{X}_1^* \cdot \mathbf{X}_1^{*\mathrm{T}} + \mathbf{X}_2^* \cdot \mathbf{X}_2^{*\mathrm{T}})\mathbf{t} = (\mathbf{t}^{\mathrm{T}} \cdot \mathbf{t})\mathbf{t}$. This shows that $\mathbf{t}^{\mathrm{T}} \cdot \mathbf{t}$ is the largest eigenvalue of $\mathbf{X}_1^* \cdot \mathbf{X}_1^{*\mathrm{T}} + \mathbf{X}_2^* \cdot \mathbf{X}_2^{*\mathrm{T}}$ and **t** is its corresponding eigenvector.
CPCA on $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\mathbf{t}_{\mathrm{T}} = \mathbf{T} \cdot \mathbf{w}_{\mathrm{T}} \tag{17}$$

$$\mathbf{t}_{\mathrm{T}} = \mathbf{t}_1 \cdot w_{\mathrm{T}}(1) + \mathbf{t}_2 \cdot w_{\mathrm{T}}(2) \tag{18}$$

$$\mathbf{t}_{\mathrm{T}} = (\mathbf{t}_1 \cdot \mathbf{t}_1^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} + \mathbf{t}_2 \cdot \mathbf{t}_2^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}}) / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{19}$$

$$\mathbf{t}_{\mathrm{T}} = [(\mathbf{X}_1 / m_{\mathrm{X1}}^{1/2}) \cdot \mathbf{p}_1 \cdot \mathbf{p}_1^{\mathrm{T}} \cdot (\mathbf{X}_1^{\mathrm{T}} / m_{\mathrm{X1}}^{1/2}) \cdot \mathbf{t}_{\mathrm{T}} + (\mathbf{X}_2 / m_{\mathrm{X2}}^{1/2}) \cdot \mathbf{p}_2 \cdot \mathbf{p}_2^{\mathrm{T}} \cdot (\mathbf{X}_2^{\mathrm{T}} / m_{\mathrm{X2}}^{1/2}) \cdot \mathbf{t}_{\mathrm{T}}] / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{20}$$

$$\mathbf{t}_{\mathrm{T}} = (\mathbf{X}_1^* \cdot \mathbf{p}_1 \cdot \mathbf{p}_1^{\mathrm{T}} \cdot \mathbf{X}_1^{*\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} + \mathbf{X}_2^* \cdot \mathbf{p}_2 \cdot \mathbf{p}_2^{\mathrm{T}} \cdot \mathbf{X}_2^{*\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}}) / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{21}$$

where $\mathbf{X}_1^* = \mathbf{X}_1 / m_{\mathrm{X1}}^{1/2}$

$$\mathbf{t}_{\mathrm{T}} = [(\mathbf{X}_1^* \cdot \mathbf{p}_1 \cdot \mathbf{p}_1^{\mathrm{T}} \cdot \mathbf{X}_1^{*\mathrm{T}} + \mathbf{X}_2^* \cdot \mathbf{p}_2 \cdot \mathbf{p}_2^{\mathrm{T}} \cdot \mathbf{X}_2^{*\mathrm{T}}) \mathbf{t}_{\mathrm{T}}] / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{22}$$

$\mathbf{X}_1^*$ is in fact composed of its first component $\mathbf{X}_1^* \cdot \mathbf{p}_1 \cdot \mathbf{p}_1^{\mathrm{T}}$ and the residual after deflation, $\mathbf{E}_1^*$. Thus

$$\mathbf{t}_{\mathrm{T}} = [(\mathbf{X}_1^* \cdot \mathbf{X}_1^{*\mathrm{T}} + \mathbf{X}_2^* \cdot \mathbf{X}_2^{*\mathrm{T}}) \mathbf{t}_{\mathrm{T}}] / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{23}$$

$$\mathbf{t}_{\mathrm{T}} = \{[(\mathbf{X}_1^* \cdot \mathbf{p}_1 \cdot \mathbf{p}_1^{\mathrm{T}} + \mathbf{E}_1^*) \mathbf{X}_1^{*\mathrm{T}} + (\mathbf{X}_2^* \cdot \mathbf{p}_2 \cdot \mathbf{p}_2^{\mathrm{T}} + \mathbf{E}_2^*) \mathbf{X}_2^{*\mathrm{T}}] \mathbf{t}_{\mathrm{T}}\} / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{24}$$

$$\mathbf{t}_{\mathrm{T}} = \{[\mathbf{X}_1^* (\mathbf{X}_1^* \cdot \mathbf{p}_1 \cdot \mathbf{p}_1^{\mathrm{T}} + \mathbf{E}_1^*)^{\mathrm{T}} + \mathbf{X}_2^* (\mathbf{X}_2^* \cdot \mathbf{p}_2 \cdot \mathbf{p}_2^{\mathrm{T}} + \mathbf{E}_2^*)^{\mathrm{T}}] \mathbf{t}_{\mathrm{T}}\} / \mathbf{t}_{\mathrm{T}}^{\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{25}$$

$$\mathbf{t}_{\mathrm{T}} = [\mathbf{X}_1^* (\mathbf{p}_1 \cdot \mathbf{p}_1^{\mathrm{T}} \mathbf{X}_1^{*\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}} + \mathbf{E}_1^{*\mathrm{T}} \cdot \mathbf{t}_{\mathrm{T}}) + \mathbf{X}_2^* (\mathbf{p}_2 \cdot \mathbf{p}_2^{\mathrm{T}} \cdot \mathbf{X}_2^{*\mathbf{T}} \cdot \mathbf{t}_{\mathrm{T}} + \mathbf{E}_2^{*\mathbf{T}} \cdot \mathbf{t}_{\mathrm{T}})] / \mathbf{t}_{\mathrm{T}}^{\mathbf{T}} \cdot \mathbf{t}_{\mathrm{T}} \tag{26}$$

This equals equation (22), because $\mathbf{E}_1^{*\mathrm{T}}\cdot\mathbf{t}_\mathrm{T}$ and $\mathbf{E}_2^{*\mathrm{T}}\cdot\mathbf{t}_\mathrm{T}$ are both zero.

$$\mathbf{E}_1^{*\mathrm{T}}\cdot\mathbf{t}_\mathrm{T} = \mathbf{X}_1^{*\mathrm{T}}\cdot\mathbf{t}_\mathrm{T} - \mathbf{p}_1\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T} \tag{27}$$

$$\mathbf{E}_1^{*\mathbf{T}}\cdot\mathbf{t}_\mathrm{T} = \mathbf{p}_1\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T} - \mathbf{p}_1\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T} = 0 \tag{28}$$

From equations (16) and (23) we see that $\mathbf{t}$ and $\mathbf{t}_\mathrm{T}$ are the eigenvectors of $\mathbf{X}_1^*\cdot\mathbf{X}_1^{*\mathrm{T}} + \mathbf{X}_2^*\cdot\mathbf{X}_2^{*\mathrm{T}}$ corresponding to the same largest eigenvalue and therefore are equal.

II.2. Proof that **u** score of MBPLS equals **u** score of PLS

PLS on $\mathbf{X}$ and $\mathbf{Y}$ ($\mathbf{X}=[\mathbf{X}_1/m_{\mathrm{X1}}^{1/2}\ \ \mathbf{X}_2/m_{\mathrm{X2}}^{1/2}] = [\mathbf{X}_1^*\ \ \mathbf{X}_2^*]$):

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{q}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q} \tag{29}$$

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{t}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}^\mathrm{T}\cdot\mathbf{t} \tag{30}$$

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}\cdot\mathbf{w}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}^\mathrm{T}\cdot\mathbf{t} \tag{31}$$

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_1^*\cdot\mathbf{w}_1^o + \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_2^*\cdot\mathbf{w}_2^o/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}^\mathrm{T}\cdot\mathbf{t} \tag{32}$$

where $\mathbf{w}_1^o$ is the part of $\mathbf{w}$ belonging to the variables in $\mathbf{X}_1^*$

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_1^*\cdot\mathbf{X}_1^{*\mathrm{T}}\cdot\mathbf{u} + \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_2^*\cdot\mathbf{X}_2^{*\mathrm{T}}\cdot\mathbf{u}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}^\mathrm{T}\cdot\mathbf{t}\cdot\mathbf{u}^\mathrm{T}\cdot\mathbf{u} \tag{33}$$

$$\mathbf{u} = (\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_1^*\cdot\mathbf{X}_1^{*\mathrm{T}} + \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_2^*\cdot\mathbf{X}_2^{*\mathrm{T}})\mathbf{u}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}^\mathrm{T}\cdot\mathbf{t}\cdot\mathbf{u}^\mathrm{T}\cdot\mathbf{u} \tag{34}$$

MBPLS with $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{Y}$:

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{q}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q} \tag{35}$$

$$\mathbf{u} = \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T} \tag{36}$$

$$\mathbf{u} = (\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{t}_1\cdot w_\mathrm{T}(1) + \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{t}_2\cdot w_\mathrm{T}(2))/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T} \tag{37}$$

$$\mathbf{u} = (\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{t}_1\cdot\mathbf{t}_1^\mathrm{T}\cdot\mathbf{u} + \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{t}_2\cdot\mathbf{t}_2^\mathrm{T}\cdot\mathbf{u})/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T}\cdot\mathbf{u}^\mathrm{T}\cdot\mathbf{u} \tag{38}$$

$$\begin{aligned}\mathbf{u} = [&\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot(\mathbf{X}_1/m_{\mathrm{X1}}^{1/2})\cdot\mathbf{w}_1\mathbf{w}_1^\mathrm{T}\cdot(\mathbf{X}_1^\mathrm{T}/m_{\mathrm{X1}}^{1/2})\cdot\mathbf{u}+\\ &\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot(\mathbf{X}_2/m_{\mathrm{X2}}^{1/2})\cdot\mathbf{w}_2\cdot\mathbf{w}_2^\mathrm{T}\cdot(\mathbf{X}_2^\mathrm{T}/m_{\mathrm{X2}}^{1/2})\cdot\mathbf{u}]/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T}\cdot\mathbf{u}^\mathrm{T}\cdot\mathbf{u}\end{aligned} \tag{39}$$

$$\mathbf{u} = (\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_1^*\cdot\mathbf{w}_1\cdot\mathbf{w}_1^\mathrm{T}\cdot\mathbf{X}_1^{*\mathrm{T}}\cdot\mathbf{u} + \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_2^*\cdot\mathbf{w}_2\cdot\mathbf{w}_2^\mathrm{T}\cdot\mathbf{X}_2^{*\mathrm{T}}\cdot\mathbf{u})/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T}\cdot\mathbf{u}^\mathrm{T}\cdot\mathbf{u} \tag{40}$$

where $\mathbf{X}_1^* = \mathbf{X}_1/m_{\mathrm{X1}}^{1/2}$

$$\mathbf{u} = (\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_1^*\cdot\mathbf{w}_1\cdot\mathbf{w}_1^\mathrm{T}\cdot\mathbf{X}_1^{*\mathrm{T}} + \mathbf{Y}\cdot\mathbf{Y}^\mathrm{T}\cdot\mathbf{X}_2^*\cdot\mathbf{w}_2\cdot\mathbf{w}_2^\mathrm{T}\cdot\mathbf{X}_2^{*\mathrm{T}})\mathbf{u}/\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T}\cdot\mathbf{u}^\mathrm{T}\cdot\mathbf{u} \tag{41}$$

In the same way as in the CPCA case it can be shown that equations (34) and (41) are equal. This shows that the **u** score of PLS equals the **u** score of MBPLS. The same can be shown for the **t** score of PLS and the $\mathbf{t}_\mathrm{T}$ super score of MBPLS, where both derivations come to the following equation:

$$\mathbf{t}_\mathrm{T} = (\mathbf{X}_1^*\cdot\mathbf{X}_1^{*\mathrm{T}}\cdot\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T} + \mathbf{X}_2^*\cdot\mathbf{X}_2^{*\mathrm{T}}\cdot\mathbf{Y}\cdot\mathbf{Y}^\mathrm{T})\mathbf{t}_\mathrm{T}/\mathbf{u}^\mathrm{T}\cdot\mathbf{u}\cdot\mathbf{q}^\mathrm{T}\cdot\mathbf{q}\cdot\mathbf{t}_\mathrm{T}^\mathrm{T}\cdot\mathbf{t}_\mathrm{T} \tag{42}$$

APPENDIX III

In this appendix an example is used to illustrate that HPCA converges to different solutions depending on the initial guess of the super score $\mathbf{t}_\mathrm{T}$. The data used are from the simulated hypothetical process that was shown in Example 4. There are three blocks, where $\mathbf{X}_1$ consists of temperatures of the first process unit and $\mathbf{X}_2$ and $\mathbf{X}_3$ contain respectively temperatures and pressures of the second process unit. Blocks $\mathbf{X}_2$ and $\mathbf{X}_3$ are combined into one block ($\mathbf{X}_2\mathbf{X}_3$) as was done for the results of

Table 8. Cumulative percentage explained variation of block $\mathbf{X}_1$ and combined block $\mathbf{X}_2\mathbf{X}_3$ with HPCA when a different column is selected as initial guess of super score

	$\mathbf{X}_1$, col. 1 or 3		$\mathbf{X}_1$, col. 2		$\mathbf{X}_2\mathbf{X}_3$, col. 1, 3, 4, 5 or 6		$\mathbf{X}_2\mathbf{X}_3$, col. 2, 7 or 8	
	$\%\mathbf{X}_1$	$\%\mathbf{X}_2\mathbf{X}_3$	$\%\mathbf{X}_1$	$\%\mathbf{X}_2\mathbf{X}_3$	$\%\mathbf{X}_1$	$\%\mathbf{X}_2\mathbf{X}_3$	$\%\mathbf{X}_1$	$\%\mathbf{X}_2\mathbf{X}_3$
PC1	42	39	61	18	42	39	1	75
PC2	82	77	82	76	82	77	48	88
PC3	91	89	91	89	82	97	82	99
PC4	99	99	99	99	99	99	94	99

Table 7.

Table 8 shows the cumulative percentage variation explained per block at each PC when a different column of a block is selected as the initial guess of the super score $\mathbf{t}_T$. The percentages of explained variation in the first component (PC1) change from 1% to 61% for block $\mathbf{X}_1$ and from 18% to 75% for combined block $\mathbf{X}_2\mathbf{X}_3$ depending on the initial guess.

When the initial guess is the eigenvector of $\mathbf{X}^T\mathbf{X}$ that corresponds to the largest eigenvalue, as was done throughout this study, the HPCA solution equals the solution when column 1 or column 3 of $\mathbf{X}_1$ is chosen to be the initial guess.

REFERENCES

1. S. Wold, K. Esbensen and P. Geladi, *Chemometrics Intell. Lab. Syst.* **2**, 37–52 (1987).
2. P. Geladi, *J. Chemometrics*, **2**, 231–246 (1988).
3. P. Geladi and B. R. Kowalski, *Anal. Chim. Acta*, **185**, 1–17 (1986).
4. A. Höskuldsson, *J. Chemometrics*, **2**, 211–228 (1988).
5. R. Manne, *Chemometrics Intell. Lab. Syst.* **2**, 187–197 (1988).
6. A. Lorber, L. E. Wangen and B. R. Kowalski, *J. Chemometrics*, **1**, 19–31 (1987).
7. I. E. Frank and J. H. Friedman, *Technometrics*, **35**, 109–135 (1994).
8. A. J. Burnham, R. Viveros and J. F. MacGregor, *J. Chemometrics*, **10**, 31–45 (1996).
9. C. F. Slama, *M. Eng. Thesis*, McMaster University, Hamilton, Ontario (1991).
10. J. F. MacGregor, C. Jeackle, C. Kiparissides and M. Koutoudi, *AIChE J.* **40**, 826–838 (1994).
11. S. Wold, N. Kettaneh and K. Tjessem, *J. Chemometrics*, **10**, 463–482 (1996).
12. J. A. Westerhuis and P. M. J. Coenegracht, *J. Chemometrics*, **11**, 379–392 (1997).
13. S. Wold, S. Hellberg, T. Lundstedt, M. Sjostrom and H. Wold, *Proc. Symp. on PLS Model Building: Theory and Application*, Frankfurt am Main, 1987; also *Tech. rep*., Department of Organic Chemistry, Umeå University (1987).
14. S. Rännar, J. F. MacGregor and S. Wold, *Chemometrics Intell. Lab. Syst.* in press (1998).
15. R. W. Gerlach, B. R. Kowalski and H. O. A. Wold, *Anal. Chim. Acta*, **112**, 417–421 (1979).
16. I. E. Frank, J. Feikema, N. Constantine and B. R. Kowalski, *J. Chem. Info. Comput. Sci*. **24**, 20–24 (1984).
17. I. E. Frank and B. R. Kowalski, *Anal. Chim. Acta*, **162**, 241–251 (1984).
18. I. E. Frank and B. R. Kowalski, *Anal. Chim. Acta*, **167**, 51–63 (1985).
19. L. E. Wangen and B. R. Kowalski, *J. Chemometrics*, **3**, 3–20 (1988).
20. S. Wold, H. Martens and H. Wold, in *MULDAST Proc*., ed. by S. Wold, *Tech. Rep*. Research Group for Chemometrics, Umeå University (1984).
21. T. Kourti, P. Nomikos and J. F. MacGregor, *J. Process Control*, **5**, 277–284 (1995).
22. P. R. C. Nelson, P. A. Taylor and J. F. MacGregor, *Chemometrics Intell. Lab. Syst*. **35**, 45–65 (1996).
23. C. Kiparissides, G. Veros and J. F. MacGregor, *J. Macromol. Sci., Rev. Macromol. Chem. Phys*. **C33**, 437–527 (1993).
24. G. Veros, M. Papadakis and C. Kiparissides, *Polym. React. Engng*. **3**, 427–460 (1993).
25. S. Wold, *Technometrics*, **20**, 397–405 (1978).

