JOURNAL OF CHEMOMETRICS
J. Chemometrics 2003; **17**: 323–337
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.811

A framework for sequential multiblock component methods

Age K. Smilde[1,2]*, Johan A. Westerhuis[1] and Sijmen de Jong[3]

[1]Process Analysis and Chemometrics, Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, The Netherlands
[2]TNO Nutrition and Food Research, Utrechtseweg 48, NL-3700 AJ Zeist, The Netherlands
[3] Unilever Research and Development Vlaardingen, PO Box 114, NL-3130 AC Vlaardingen, The Netherlands

Received 4 October 2002; Revised 18 May 2003; Accepted 18 May 2003

Multiblock or multiset methods are starting to be used in chemistry and biology to study complex data sets. In chemometrics, sequential multiblock methods are popular; that is, methods that calculate one component at a time and use deflation for finding the next component. In this paper a framework is provided for sequential multiblock methods, including hierarchical PCA (HPCA; two versions), consensus PCA (CPCA; two versions) and generalized PCA (GPCA). Properties of the methods are derived and characteristics of the methods are discussed. All this is illustrated with a real five-block example from chromatography. The only methods with clear optimization criteria are GPCA and one version of CPCA. Of these, GPCA is shown to give inferior results compared with CPCA. Copyright © 2003 John Wiley & Sons, Ltd.

KEYWORDS: multiblock methods; hierarchical PCA; consensus PCA; generalized PCA; multiway methods; stationary phases; reversed phase liquid chromatography

1. INTRODUCTION

With the availability of sophisticated instruments and computers, collecting large amounts of data is becoming increasingly popular. In science and engineering, often problems occur that have to do with analyzing several blocks of collected data. This has generated the need for two new types of data analysis methods, namely multiway analysis methods and multiblock analysis methods. The first class of methods deals with data sets that can be arranged into a multiway array (i.e. the natural extension of a two-way array, a matrix) [1,2], whereas the second class of methods has been developed to analyze several data sets simultaneously [3]. Recently, a method has also been developed for multiway multiblock analysis [4].

Multiblock analysis methods already have a long history in psychometrics [3,5,6] and are still the subject of active research [7–11]. Also in computational statistics and chemometrics, multiblock methods have been developed [4,12–16] and investigated theoretically [17,18]. Applications in different areas have been reported [13,19,20]. It is expected that in the future the need for multiblock methods will become even more urgent, given the large amounts of data generated in e.g. genomics and high-throughput experimentation.

*Correspondence to: A. K. Smilde, Process Analysis and Chemometrics, Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, The Netherlands
E-mail: asmilde@science.uva.nl

The generic problem of multiblock analysis is to find underlying relationships between several sets of possibly related data sets. This paper discusses multiblock *component* models, where components or latent variables are used to summarize the relevant information between and within the blocks. Hence multiblock models designed to predict a certain response in a multiblock setting (multiblock *regression* problems or latent path models) are not discussed. There are two essentially different ways of calculating components in multiblock component models. One approach is to calculate the first component, then remove the influence of this component from all matrices involved (e.g. by deflation) and subsequently calculate the next component from the corrected matrices. This is called the sequential approach and is much used in chemometrics [16]. Another approach is to calculate all components simultaneously by solving a global optimization problem [4]. Since there is much confusion about the exact background of sequential algorithms, these will be the focus of this paper.

The goals of this paper are (i) to give a framework for sequential multiblock component methods, (ii) to discuss the properties of these multiblock methods and (iii) to show the workings of these multiblock methods and some of their properties using a worked-out case study. The paper starts with the theoretical background. Then the background of the case study is explained, followed by a presentation of the results of applying the different multiblock methods on the real example. The paper ends with some conclusions.

2. THEORY

2.1. General

There are different classes of multiblock problems, depending on the structure of the collected data sets. A *mode* of a data set is defined as one of the categories used to order the data, e.g. an object mode categorizes the different objects or samples of whichever data are collected. A classification of multiblock problems is possible by analyzing how many and which modes are in common between the data sets. Figure 1 summarizes the different possibilities.

The first class of multiblock problems is encountered when data sets with different objects and variables are considered (Figure 1(a)). Although there might be a relationship between the different blocks of data, this is not directly evident from their structure.

In the second class of multiblock problems the objects are in common (Figure 1(b)), but the variables measured on these objects are different. For each data set the objects form a configuration in a multivariate space. For finding (dis)similarities between these configurations, generalized Procrustes analysis can be used [21], which finds the maximal agreement (in the least squares sense) allowing translation, rotation and optional isotropic scaling of the individual blocks. Often low-dimensional solutions are wanted. This can be achieved either by applying PCA postprocessing to the average (consensus) configuration obtained in full space or by imposing low dimensionality on the consensus from the outset [22].

An alternative is to perform generalized Procrustes analysis on score matrices obtained from separate principal component analyses of each data set, but this has serious limitations [8]. Multiblock component methods developed for the second type of problems are consensus PCA [15], hierarchical PCA [16] and multiblock covariate component models [4].

The third class of multiblock problems is encountered when the variable mode is in common (Figure 1(c)). For this type of problem, simultaneous component analysis methods have been developed [8,10]. In chemistry, this class of problems is encountered e.g. when analyzing different chemical systems with the same spectroscopic methods. Multivariate curve resolution has been developed to analyze such multiblock data sets [23,24]. Monitoring of batch processes can also be done in this fashion [25].

In the fourth class of multiblock problems there are two modes in common (Figure 1(d)). As shown in Figure 1(d), stacking the data sets on top of each other gives a three-way array, for which there exist data analysis methods [1,2]. Hence multiblock methods can be seen as generalizations of multiway methods.

It is possible to use a class 3 method for a class 2 problem and *vice versa*. For instance, transposing a class 2 problem fits such a problem in the framework of simultaneous component analysis. Whether or not this is useful to consider depends on the situation. The focus of this paper will be on class 2 problems.

The different multiblock methods will be explained with a B-block example, where blocks $\mathbf{X}_1\ (I \times J_1), \ldots, \mathbf{X}_b\ (I \times J_b), \ldots, \mathbf{X}_B\ (I \times J_B)$ are the subject of research. The concatenation of $\mathbf{X}_1, \ldots, \mathbf{X}_b, \ldots, \mathbf{X}_B$ is denoted as the supermatrix $\mathbf{X}(I \times \sum_{b=1}^{B} J_b)$. The cross-product matrices $\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}} = \mathbf{S}_b \quad (I \times I)$ and $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}\ (I \times I)$ will be used in the sequel.

Extensions to multiway arrays will be considered using the three-way arrays $\underline{\mathbf{X}}_1\ (I \times J_1 \times K_1), \ldots, \underline{\mathbf{X}}_b\ (I \times J_b \times K_b), \ldots, \underline{\mathbf{X}}_B\ (I \times J_B \times K_B)$. These three-way arrays

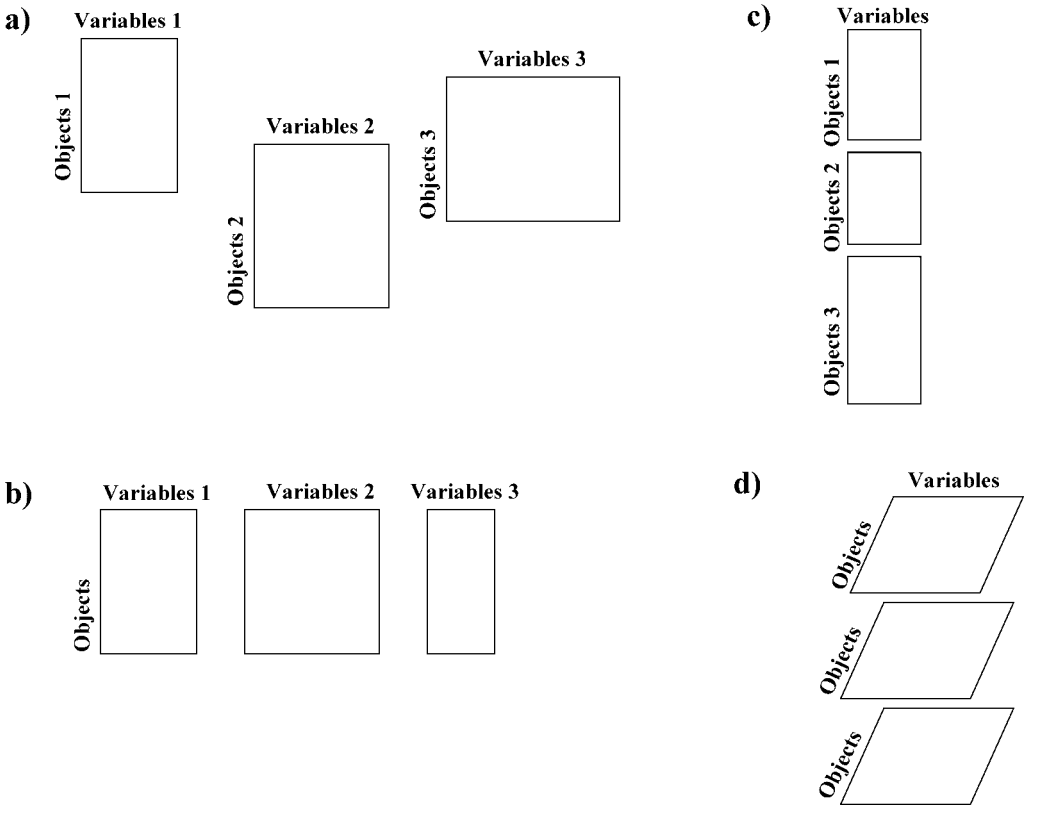

Figure 1. Different types of multiblock problems (see text).

can be matricized to $\mathbf{X}_1\,(I \times J_1K_1), \ldots, \mathbf{X}_b\,(I \times J_bK_b), \ldots, \mathbf{X}_B(I \times J_BK_B)$ and subsequently concatenated to $\mathbf{X} = [\mathbf{X}_1|\ldots|\mathbf{X}_B]\,(I \times (J_1K_1 + \cdots + J_BK_B))$. All matrices are assumed to be properly preprocessed; that is, usually centered in the object direction and possibly scaled. It is also possible to use block scaling; that is, giving each block *a priori* a prespecified sum of squares.

2.2. Properties of multiblock models

There are two important choices to make when considering multiblock methods: variance explained and fairness [3]. The first choice relates to whether only the relationship between blocks is interesting or whether the multiblock solution should also describe variation within each block of data. It is not difficult to find components that describe linear relationships between blocks, but these components are not necessarily good at explaining the variation within blocks. For example, the right singular vector $\mathbf{v}$ associated with the smallest singular value of $\mathbf{X}$ defines a component $\mathbf{Xv} \approx \mathbf{0}$ that describes an almost perfect linear relationship. The second choice relates to whether the solution found is fair, that is, whether each block contributes in approximately the same amount to the solution.

Another important property of a multiblock method is the criterion it optimizes in order to find the solution. For some of the multiblock methods this criterion is not clear, as will be shown below. Flexibility of the multiblock methods with respect to generalizability to multiway data and possibilities to incorporate constraints reflecting *a priori* knowledge are also discussed. These issues become increasingly important when dealing with real-life chemical and biological data.

In the following subsections the different multiblock methods are presented. Table I summarizes the algorithms and Table II summarizes some properties of the multiblock methods, among which the ones discussed above.

2.3. SUM-PCA

The most straightforward way of multiblock analysis is to perform a PCA on $\mathbf{X}$ (see Table I). Hence each block $\mathbf{X}_b$ is modeled with the same superscores $\mathbf{T}_{\text{sup}}$ and its blockloadings $\mathbf{P}_b$, with residuals $\mathbf{E}_b$. SUM-PCA solves the general PCA problem

$$\min_{\mathbf{T}_{\text{sup}},\,\mathbf{P}_{\text{sup}}} \left\|\mathbf{X} - \mathbf{T}_{\text{sup}}\mathbf{P}_{\text{sup}}^{\mathrm{T}}\right\|^2 = \min_{\mathbf{T}_{\text{sup}},\,\mathbf{P}_b} \sum_{b=1}^{B} \left\|\mathbf{X}_b - \mathbf{T}_{\text{sup}}\mathbf{P}_b^{\mathrm{T}}\right\|^2 \quad (1)$$

or, stated differently,

$$\max_{\mathbf{T}_{\text{sup}}^{\mathrm{T}}\mathbf{T}_{\text{sup}}=\mathbf{I}} \operatorname{tr}\left(\mathbf{T}_{\text{sup}}^{\mathrm{T}}\mathbf{S}\mathbf{T}_{\text{sup}}\right) = \max_{\mathbf{T}_{\text{sup}}^{\mathrm{T}}\mathbf{T}_{\text{sup}}=\mathbf{I}} \operatorname{tr}\left(\mathbf{T}_{\text{sup}}^{\mathrm{T}} \sum_{b=1}^{B} \mathbf{S}_b\mathbf{T}_{\text{sup}}\right) \quad (2)$$

and the right-hand side of (2) explains the name 'SUM-PCA'. The constraint $\mathbf{T}_{\text{sup}}^{\mathrm{T}}\mathbf{T}_{\text{sup}} = \mathbf{I}$ identifies the solution up to permutation and sign.

The solution of the SUM-PCA problem (2) is found by solving the eigenvalue problem

$$\left(\sum_{b=1}^{B} \mathbf{S}_b\right)\mathbf{T}_{\text{sup}} = \mathbf{T}_{\text{sup}}\mathbf{\Lambda} \quad (3)$$

Table I. Algorithms of multiblock methods

SUM-PCA	CPCA	CPCA-W	HPCA	HPCA-W	GPCA
$\mathbf{X} = [\mathbf{X}_1\ \ldots\ \mathbf{X}_B]$		(a) choose start $\mathbf{t}_{\text{sup}}$ with $\|\mathbf{t}_{\text{sup}}\| = 1$			(a) $\mathbf{X} = \mathbf{USV}^{\mathrm{T}}$; $\mathbf{t}_{\text{sup}} = \mathbf{u}_1$
$= \mathbf{T}_{\text{sup}}\mathbf{P}_{\text{sup}}^{\mathrm{T}} + \mathbf{E}$		(b) for all b iterate until stable $\mathbf{t}_{\text{sup}}$			(b) $\|\mathbf{t}_{\text{sup}}\| \rightarrow 1$, For all b:
$= \mathbf{T}_{\text{sup}}[\mathbf{P}_1^{\mathrm{T}}\ \ldots\ \mathbf{P}_B^{\mathrm{T}}] + \mathbf{E}$					(c) $\mathbf{t}_b = \mathbf{X}_b\mathbf{X}_b^{+}\mathbf{t}_{\text{sup}}$
	(c) $\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\text{sup}}$	(c) $\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\text{sup}}/\|\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\text{sup}}\|$	(c) $\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\text{sup}}$	(c) $\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\text{sup}}$	(d) $\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}}\mathbf{t}_b/(\mathbf{t}_b^{\mathrm{T}}\mathbf{t}_b)$
	(d) $\mathbf{t}_b = \mathbf{X}_b\mathbf{p}_b/(\mathbf{p}_b^{\mathrm{T}}\mathbf{p}_b)$	(d) $\mathbf{t}_b = \mathbf{X}_b\mathbf{p}_b$	(d) $\mathbf{t}_b = \mathbf{X}_b\mathbf{p}_b$	(d) $\mathbf{t}_b = \mathbf{X}_b/\|\mathbf{X}_b\mathbf{p}_b\|$	(e) $\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_b\mathbf{p}_b^{\mathrm{T}}$
		(e) $\mathbf{T} = [\mathbf{t}_1\ \ldots\ \mathbf{t}_B]$			(f) $\mathbf{E} = [\mathbf{E}_1 \ldots\ \mathbf{E}_B]$
		(f) $\mathbf{w} = \mathbf{T}^{\mathrm{T}}\mathbf{t}_{\text{sup}}$			(g) go to (a) with $\mathbf{E} = \mathbf{X}$
		(g) $\mathbf{t}_{\text{sup}} = \mathbf{Tw}$			
		(h) $\mathbf{w} = \mathbf{w}/\|\mathbf{t}_{\text{sup}}\|$			
		(i) $\mathbf{t}_{\text{sup}} = \mathbf{t}_{\text{sup}}/\|\mathbf{t}_{\text{sup}}\|$			
		(j) end			
		(k) deflation:			
		$\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\text{sup}}$			
		$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_{\text{sup}}\mathbf{p}_b$			
		(l) continue with $\mathbf{E}_b$			

Table II. Properties of multiblock methods

	Fairness	Variance	Criterion	Superscores	Blockscores
SUM-PCA	Block scaling	Overall: yes; individual: no control	$\max_{\mathbf{T}_{sup}^T\mathbf{T}_{sup}=\mathbf{I}} \text{tr}(\mathbf{T}_{sup}^T\mathbf{S}\mathbf{T}_{sup})$ or solve $\mathbf{S}\mathbf{T}_{\mathbf{sup}} = \Lambda\mathbf{T}_{sup}$	Orthogonal	Not available
CPCA	Built in	No guarantee	Solve sequentially $\left(\sum_{b=1}^{B} c_b^{-1}\mathbf{S}_b\right)\mathbf{t}_{sup} = B\mathbf{t}_{sup}$	Orthogonal	Not orthogonal
CPCA-W	Block scaling	Overall: yes; Individual: no control	Solve sequentially $\left(\sum_{b=1}^{B}\mathbf{S}_b\right)\mathbf{t}_{sup} = \left(\sum_{b=1}^{B} c_b\right)\mathbf{t}_{sup}$ or simultaneously $\max_{\mathbf{T}_{sup}^T\mathbf{T}_{sup}=\mathbf{I}} \text{tr}(\mathbf{T}_{sup}^T\mathbf{S}\mathbf{T}_{sup})$	Orthogonal	Not orthogonal
HPCA	No	No guarantee	Solve sequentially $\left(\sum_{b=1}^{B} c_b\mathbf{S}_b\right)\mathbf{t}_{sup} = \left(\sum_{b=1}^{B} c_b^2\right)\mathbf{t}_{sup}$	Orthogonal	Not orthogonal
HPCA-W	No	No guarantee	Solve sequentially $\left(\sum_{b=1}^{B} c_b d_b^{-1}\mathbf{S}_b\right)\mathbf{t}_{sup} = \left(\sum_{b=1}^{B} c_b^2 d_b^{-1}\right)\mathbf{t}_{sup}$	Orthogonal	Not orthogonal
GPCA	Block scaling	Built in	Solve sequentially $\max_{\mathbf{t}} \sum_{b=1}^{B} R^2(\mathbf{t}_{sup}, \mathbf{t}_b) \sum_{j_b=1}^{J_b} R^2(\mathbf{t}_b, \mathbf{X}_{bj_b})$	Orthogonal	Orthogonal

and, given $\mathbf{T}_{sup}$, the matrix of superloadings $\mathbf{P}_{sup}$ follows directly from regressing $\mathbf{X}$ on $\mathbf{T}_{sup}$, i.e. $\mathbf{P}_{sup} = \mathbf{X}^T\mathbf{T}_{sup}\left(\mathbf{T}_{sup}^T\mathbf{T}_{sup}\right)^{-1} = \mathbf{X}^T\mathbf{T}_{sup}$. The superloading matrix also has orthogonal columns. Note that this does not translate to the individual loading matrices $\mathbf{P}_b$: these do not have orthogonal columns.

The amount of variation in $\mathbf{X}$, more specifically the sum of squares, explained by $\mathbf{T}_{sup}$ can be separated into amounts of variation in the different blocks explained by the superscores $\mathbf{T}_{sup}$. This is shown for the first superscore $\mathbf{t}_{sup,1}$ but can be extended to all superscores. The amount of variation explained in $\mathbf{X}$ by $\mathbf{t}_{sup,1}$ is

$$\lambda_1 = \frac{\mathbf{t}_{sup,1}^T\mathbf{S}\mathbf{t}_{sup,1}}{\mathbf{t}_{sup,1}^T\mathbf{t}_{sup,1}} = \mathbf{t}_{sup,1}^T\mathbf{S}\mathbf{t}_{sup,1} \text{ because } \mathbf{t}_{sup,1}^T\mathbf{t}_{sup,1} = 1 \quad (4)$$

and the value λ_1 can be written as the sum of separate contributions c_b (the suffix indicating component 1 is dropped for convenience):

$$\lambda_1 = \mathbf{t}_{sup,1}^T\mathbf{S}\mathbf{t}_{sup,1} = \mathbf{t}_{sup,1}^T\left(\sum_{b=1}^{B}\mathbf{S}_b\right)\mathbf{t}_{sup,1} = \sum_{b=1}^{B} c_b \quad (5)$$

where $c_b = \mathbf{t}_{sup,1}^T\mathbf{S}_\mathbf{b}\mathbf{t}_{sup,1}$ is the amount of variation (sum of squares) explained in block $\mathbf{X}_b$ by the superscore $\mathbf{t}_{sup,1}$.

There is no guarantee that each block contributes to the final superscores; also, the variation explained by the superscores in each block is not under control. The only guarantee is that the superscores explain the maximum amount of variation in $\mathbf{X}$ and hence will also explain variation in at least one of the blocks. Of course, by using block scaling, the fairness and variance explained in each block can be controlled to a certain extent (by giving 'low variance blocks' a higher initial sum of squares). The number of components for each block is the same.

Constraints on scores and loadings can be built in by solving problem (1) subject to those constraints. In that case the problem cannot be solved anymore with an eigenvalue approach, but an alternating least squares approach should be used. An extension to multiway blocks is possible after matricizing the three-way arrays and building a SUM-PCA-like model for the concatenated $\mathbf{X}$ where each block obtains its own loading matrix structure. Consider e.g. three blocks $\mathbf{X}_1$ $(I \times J_1)$, $\underline{\mathbf{X}}_2$ $(I \times J_2 \times K_2)$ and $\mathbf{X}_3$ $(I \times J_3)$, of which $\underline{\mathbf{X}}_2$ is the multiway block. Then a SUM-PCA can be built as

$$\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3] = \mathbf{T}_{sup}[\mathbf{P}_1^T\ \mathbf{G}_2\,(\mathbf{C}_2 \otimes \mathbf{B}_2)^T\ \mathbf{P}_3^T] + \mathbf{E} \quad (6)$$

where $\mathbf{X}_2$ $(I \times J_2K_2)$ is the proper matricization of $\underline{\mathbf{X}}_2$ and a Tucker3 structure is imposed on $\underline{\mathbf{X}}_2$. Likewise, other three-way structures can be imposed on $\underline{\mathbf{X}}_2$ and this multiway generalization can cover any combination of two-way and multiway arrays. Again, alternating least squares schemes can be used to perform the estimations.

2.4. Consensus PCA

Consensus PCA (CPCA) was first proposed by Wold *et al.* [15]. The CPCA algorithm is given in Table I, where the normalization of $\mathbf{t}_{sup}$ is omitted (this normalization was not performed in Wold *et al.*'s original CPCA algorithm; see later). As can be seen from Table I, four of the algorithms are very similar: they share common steps (a), (b) and (e)–(l). Several remarks regarding CPCA are appropriate.

1. A rational start $\mathbf{t}_{sup}$ might be the first left singular vector of $\mathbf{X}$, i.e. the SUM-PCA superscore vector.
2. In the original algorithm [15] a weighting factor was introduced in step (d) to correct for the different sizes of each block. This was done to avoid the situation that blocks with a large size tend to dominate the solution. Such a weighting can also be done on the block itself by using an appropriate block scaling. For convenience, it is assumed that the blocks are properly scaled and the weighting factor is omitted.
3. The algorithm is continued until the *a priori* selected number of superscores is reached.
4. The deflation step (k) is not by definition rank-reducing, because $\mathbf{t}_{sup}$ is not necessarily in the range of $\mathbf{X}_b$. It is rank-reducing, however, for $\mathbf{X} = [\mathbf{X}_1|\ldots|\mathbf{X}_B]$.

Some properties of this CPCA algorithm can be derived and these are summarized in Table II. Upon convergence of CPCA it holds that $\mathbf{w}$ is a vector of equal weights:

$$\mathbf{w} = B^{-1/2}\mathbf{1} \tag{7}$$

where $\mathbf{1}$ is a $B \times 1$ vector of ones (see Appendix II). This result was already realized by Geladi *et al*. [26] and it means that, in each cycle, $\mathbf{t}_{\text{sup}}$ is simply proportional to the algebraic mean of the blockscores $\mathbf{t}_b$ ($b = 1, \ldots, B$). Moreover, it can be shown that, in each cycle, CPCA solves the problem (see Appendix II)

$$\left(\sum_{b=1}^{B} c_b^{-1}\mathbf{S}_b\right)\mathbf{t}_{\text{sup}} = B\mathbf{t}_{\text{sup}} \tag{8}$$

where the term

$$c_b = \frac{\mathbf{t}_{\text{sup}}^{\text{T}}\mathbf{S}_b\mathbf{t}_{\text{sup}}}{\mathbf{t}_{\text{sup}}^{\text{T}}\mathbf{t}_{\text{sup}}} \tag{9}$$

again represents the amount of variation in $\mathbf{X}_b$ explained by the superscore $\mathbf{t}_{\text{sup}}$. This means that blocks which have a tendency to contribute strongly to $\mathbf{t}_{\text{sup}}$ will be downweighted and *vice versa*. Hence CPCA has a built-in tendency for fairness. Equation (8) shows that CPCA has no closed solution (as e.g. SUM-PCA has), because c_b depends on the unknown solution $\mathbf{t}_{\text{sup}}$ and *vice versa*. There is no guarantee that the variance explained by the superscores in each block is high: this aspect is not considered in the optimization criterion.

Constraints are not easily added to the CPCA algorithm. Owing to the sequential nature of CPCA, a generalization of CPCA to multiway arrays has to be done by assuming a Tucker1 structure for the multiway arrays and using CPCA on the properly matricized arrays.

The properties of the eigenvalue problem (8) are not clear, because the matrix involved changes during the iterations owing to the presence of c_b. This also hampers the interpretation of the superscores $\mathbf{t}_{\text{sup}}$, which is markedly different from SUM-PCA. There a clear problem is solved of finding a set of orthogonal superscores explaining the maximum amount of variance in the combined data $\mathbf{X}$. CPCA solves a similar problem but now the blocks are weighted differently by a weight factor that is only known afterwards. A further complication of CPCA is that the set of weights will differ for each dimension. The question is whether fairness is a desirable property. For example, it gives undue importance to blocks that have little in common with other blocks. It seems more reasonable to introduce fairness by prescaling the blocks to e.g. equal total variance and then applying SUM-PCA.

In real-life applications it was noticed that CPCA had poor convergence properties [18]. It was observed e.g. that $\mathbf{p}_b$ becomes infinitely small and $\mathbf{t}_b$ very large (the norm of the product $\mathbf{t}_b\mathbf{p}_b^{\text{T}}$ does not change much). This effect is not cancelled by normalizing $\mathbf{w}$ in step (g), because $\mathbf{w}$ only contains the relative contributions of the $\mathbf{t}_b$s and not their lengths. If the superscores are normalized to unit length (which does not change the solution essentially), then the convergence problems disappear. Of course, $\mathbf{w}$ should not be normalized in that case. This modified CPCA algorithm is the one presented in Table I. For this modified CPCA the weights become (see Appendix II)

$$\mathbf{w} = B^{-1}\mathbf{1} \tag{10}$$

Apart from this normalization of $\mathbf{w}$ or $\mathbf{t}_{\text{sup}}$ there is no real difference with respect to the original CPCA method; for example, Equation (8) and Equations (22)–(25) in Appendix II still hold. From now on this modified version of CPCA will be used in the calculations and derivations.

A new CPCA method was suggested by normalizing $\mathbf{p}_b$ to length one in step (c) [18]. This normalization is markedly different from normalizing $\mathbf{t}_{\text{sup}}$, because the suggested normalization of the $\mathbf{p}_b$s gives different normalization constants for each block which do not cancel out at the superlevel. Hence the solution of the resulting CPCA-W (see Table I) method is essentially different from the CPCA solution. This difference is also illustrated by investigating the weights $\mathbf{w} = (w_1, \ldots, w_B)^{\text{T}}$ in CPCA-W. It can be shown (see Reference [17] and Appendix II) that, apart from the normalization of $\mathbf{w}$ in step (h), these weights are

$$w_b = \|\mathbf{p}_b\| = c_b^{1/2} \tag{11}$$

where $\mathbf{p}_b$ is the blockloading (i.e. the loadings from step (c)) prior to normalization. Hence the new $\mathbf{t}_{\text{sup}}$ becomes a linear combination of the blockscores with weights equal to the norms of the blockloadings prior to normalization. This is obviously very different from the weights used in CPCA, which are equal for all blocks. When the superscores are normalized to length one, then the explained variance of the supercomponents is carried by the superloadings $\mathbf{p}$.

It can be shown that the resulting method CPCA-W is exactly the same as SUM-PCA and hence its properties are also equal [17,18]. In terms of Equation (8), for each dimension, CPCA-W solves (see Appendix II)

$$\left(\sum_{b=1}^{B}\mathbf{S}_b\right)\mathbf{t}_{\text{sup}} = \lambda_{\max}\,\mathbf{t}_{\text{sup}} \tag{12}$$

which is exactly the same problem as (2) solved sequentially in a NIPALS-like fashion. The maximum eigenvalue equals $\sum_{b=1}^{B} c_b$, the total amount of variance explained by the first component, showing again that CPCA-W equals SUM-PCA. The similarity between SUM-PCA and CPCA-W allows for a much simpler algorithm for CPCA-W [17,18]:

(a) $\mathbf{X} = \mathbf{T}_{\text{sup}}\mathbf{P}_{\text{sup}}^{\text{T}} + \mathbf{E}$
$\mathbf{T}_{\text{sup}} = [\mathbf{t}_{\text{sup},1} \ \ldots \mathbf{t}_{\text{sup},R}]$
$\mathbf{P}_{\text{sup}} = [\mathbf{p}_{\text{sup},1} \ \cdots \ \mathbf{p}_{\text{sup},R}]$

(b) for all b

(c) $\mathbf{p}_b = \mathbf{X}_b^{\text{T}}\mathbf{t}_{\text{sup}}/\|\mathbf{X}_b^{\text{T}}\mathbf{t}_{\text{sup}}\|$

(e) $\mathbf{t}_b = \mathbf{X}_b\mathbf{p}_b$

(f) $\mathbf{T} = [\mathbf{t}_1 \ldots \mathbf{t}_B]$

(g) $\mathbf{w} = \mathbf{T}^{\text{T}}\mathbf{t}_{\text{sup}}/\|\mathbf{T}\mathbf{T}^{\text{T}}\mathbf{t}_{\text{sup}}\|$

(h) deflation: $\mathbf{p}_b = \mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} / \left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)$

$$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_{\mathrm{sup}}\mathbf{p}_b^{\mathrm{T}}$$

(i) continue with $\mathbf{E}_b$ (13)

where the $\mathbf{t}_{\mathrm{sup}}$ are again normalized to length one.* The weights **w** in CPCA-W have a different role than in CPCA. Whereas in CPCA the weights are used to define the superscores, this is not the case for CPCA-W. In CPCA-W, all superscores and blockscores can be calculated without using weights. Nevertheless, a weight vector **w** can still be defined as the weights transforming the blockscores into the superscore: $\mathbf{t}_{\mathrm{sup}} = \mathbf{Tw}$.

The same remarks with respect to imposing contraints and generalizations to accommodate for multiway arrays are relevant for CPCA-W as for SUM-PCA.

2.5. Hierarchical PCA

Another class of methods for multiblock PCA was termed hierarchical PCA (HPCA) [16]. Hierarchical PCA is able to deal with several blocks of data in a hierarchical order. In this paper, only the multiblock problem as stated in the introduction is considered. The HPCA algorithm for this problem is given in Table I. The remarks 1–4 under CPCA are equally valid for HPCA.

Upon convergence, it holds for each cycle that HPCA solves the eigenproblem (see Appendix II)

$$\left(\sum_{b=1}^{B} c_b\mathbf{S}_b\right)\mathbf{t}_{\mathrm{sup}} = \left(\sum_{b=1}^{B} c_b^2\right)\mathbf{t}_{\mathrm{sup}} \qquad (14)$$

This means that blocks having a tendency to be important receive even more weight. Hence HPCA does not support fairness and tries to discriminate between blocks: it acts as a block selector. Equation (14) shows that HPCA again has no closed solution, because c_b depends on the unknown $\mathbf{t}_{\mathrm{sup}}$. In practice it was noticed that the HPCA solution was very sensitive to the starting vector of $\mathbf{t}_{\mathrm{sup}}$, indicating the presence of local minima in solving (14) iteratively [18].

A new version of HPCA was suggested in which the blockscores in step (d) were normalized to length one [18,20]. This version will be called HPCA-W (see Table I). This again radically changes the algorithm and it can be shown (see Appendix II) that each cycle of HPCA-W solves the problem

$$\left(\sum_{b=1}^{B} c_b d_b^{-1}\mathbf{S}_b\right)\mathbf{t}_{\mathrm{sup}} = \left(\sum_{b=1}^{B} c_b^2 d_b^{-1}\right)\mathbf{t}_{\mathrm{sup}} \qquad (15)$$

where $d_b = \mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{S}_b^2\mathbf{t}_{\mathrm{sup}}$ (see Appendix II). The particular weighting of the cross-product matrices in (15) lacks a simple interpretation, let alone a convincing justification. Note, however, that this weighting makes HPCA-W insensitive to block scaling, since multiplying e.g. all values in block b by 10 cancels out in both $c_b d_b^{-1}\mathbf{S}_b$ and $c_b^2 d_b^{-1}$. Even after the extra normalization the sensitivity to starting values of $\mathbf{t}_{\mathrm{sup}}$ remained, as shown with a simple example in Reference [18]. These problems might be due to the complicated optimization criterion of HPCA-W.

*Note that both Qin *et al*. [17] and Westerhuis *et al*. [18] reported a wrong value for **w** (step g).

For both HPCA and HPCA-W it is unclear how to impose constraints. Owing to the sequential nature of HPCA(-W), a generalization of HPCA(-W) to multiway arrays has to be done by assuming a Tucker1 structure for the multiway arrays and using HPCA(-W) on the properly matricized arrays.

The methods CPCA, CPCA-W, HPCA and HPCA-W can be summarized by the general equation

$$\left(\sum_{b=1}^{B} \lambda_b\mathbf{S}_b\right)\mathbf{t}_{\mathrm{sup}} = \left(\sum_{b=1}^{B} c_b\lambda_b\right)\mathbf{t}_{\mathrm{sup}} \qquad (16)$$

where CPCA corresponds to $\lambda_b = c_b^{-1}$, CPCA-W to $\lambda_b = 1$, HPCA to $\lambda_b = c_b$ and HPCA-W to $\lambda_b = c_b d_b^{-1}$. Writing Equation (16) in a more general form gives

$$\left(\sum_{b=1}^{B} \lambda_b\mathbf{S}_b\right)\mathbf{t}_{\mathrm{sup}} = \left(\sum_{b=1}^{B} \mu_b\right)\mathbf{t}_{\mathrm{sup}} \qquad (17)$$

which suggests many alternative methods.

2.6. Generalized PCA

Recently, a new multiblock method termed generalized principal component analysis (GPCA) has appeared [12]. GPCA also works sequentially and the algorithm is shown in Table I. Several remarks regarding the GPCA algorithm are appropriate.

- The algorithm is continued until the *a priori* chosen number of dimensions is reached.
- The deflation step (e) is always rank-reducing, because $\mathbf{t}_b$ is in the range of $\mathbf{X}_b$ (owing to the projection in step (c)).
- In chemometrics applications there will often be blocks $\mathbf{X}_b$ with $I \leq J_b$, e.g. when $\mathbf{X}_b$ contains spectral data. For such blocks, $\mathbf{t}_{\mathrm{sup}}$ will always be in the range of $\mathbf{X}_b$ and therefore $\mathbf{t}_{\mathrm{sup}} = \mathbf{t}_{\mathrm{b}}$. To avoid this situation, it is possible to restrict $\mathbf{X}_b$ to a certain subspace $\widetilde{\mathbf{X}}_b$ and use $\mathbf{t}_b = \widetilde{\mathbf{X}}_b\widetilde{\mathbf{X}}_b^{+}\mathbf{t}_{\mathrm{sup}}$. Of course, this changes the problem to that of finding a suitable subspace.
- For $r > s$ it holds that $\mathbf{t}_{br}^{\mathrm{T}}\mathbf{t}_{\mathrm{sup},s} = \mathbf{0}$ for $b = 1, \ldots, B$, i.e. blockscores are orthogonal to previous superscores [12].

The proof of the orthogonality of the blockscores (see Table II) goes as follows. Owing to the deflation steps (d) and (e), $\mathbf{E}_{br}$ is orthogonal to $\mathbf{t}_{br}$ [27]. The next blockscore $\mathbf{t}_{b(r+1)}$ is found by projecting $\mathbf{t}_{\mathrm{sup},r}$ onto the range of $\mathbf{E}_b$ and is therefore orthogonal to $\mathbf{t}_{br}$.

Using autoscaled $\mathbf{X}_1, \ldots, \mathbf{X}_B$, it can be shown that GPCA solves the following problem in the first dimension:

$$\max_{\mathbf{t}} \sum_{b=1}^{B} R^2(\mathbf{t}_{\mathrm{sup}}, \mathbf{t}_b) \sum_{j_b=1}^{J_b} R^2(\mathbf{t}_b, \mathbf{x}_{bj_b}) \qquad (18)$$

and after deflation a similar optimization is performed for the second dimension. In PCA the first principal component $\mathbf{t}_1$ of an autoscaled matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_J]$ maximizes the sum of squared correlations between $\mathbf{t}_1$ and $\mathbf{x}_j$. The next principal component $\mathbf{t}_2$ maximizes the sum of squared correlations between $\mathbf{t}_2$ and $\mathbf{x}_j$, given that $\mathbf{t}_2$ and $\mathbf{t}_1$ are orthogonal (see Appendix II). In (18) the term $\sum_{j_b=1}^{J_b} R^2(\mathbf{t}_b, \mathbf{x}_{bj_b})$ represents the

amount of explained variation in $\mathbf{X}_b$ by $\mathbf{t}_b$; this is the variance explained by the blockscores within block $\mathbf{X}_b$. The term $R^2(\mathbf{t}_{\text{sup}}, \mathbf{t}_b)$ represents the variation in $\mathbf{t}_b$ explained by the common component $\mathbf{t}_{\text{sup}}$, which is a part of the variance explained by the superscores between blocks. Hence Equation (18) mixes explained variation between and within blocks summed over the total number of blocks and dimensions. In order to make the GPCA results comparable with the other results, the superscores are normalized to length one. This does not change the solution essentially.

In the original publication on GPCA, no superweights were given. Such weights, however, can be defined analogously to CPCA(-W) and HPCA(-W):

$$w_{br} = \mathbf{t}_{\text{sup},r}^{\text{T}}\mathbf{t}_{br}, \quad r = 1, \ldots, R \text{ and } b = 1, \ldots, B \qquad (19)$$

because the length of $\mathbf{t}_{\text{sup},r}$ is one by construction. For blocks $\mathbf{X}_b$ with $I \leq J_b$, $\mathbf{t}_{\text{sup},r}$ will always be in the range of $\mathbf{X}_b$ and therefore $\mathbf{t}_{\text{sup},r} = \mathbf{t}_{\text{br}}$. Consequently, all weights will be one for such blocks. If $J_b \geq I$ for all $b = 1, \ldots, B$, then all weights will be equal to one for all blocks. This resembles CPCA (where the weights are also equal), but in GPCA all $\mathbf{t}_b$s are equal to $\mathbf{t}_{\text{sup}}$ and this is not the case in CPCA (note the difference between the operators $\mathbf{X}_b\mathbf{X}_b^{\text{T}}$ and $\mathbf{X}_b\mathbf{X}_b^{+}$).

It is not clear how to impose constraints on the GPCA solution. Owing to the sequential nature of GPCA, a generalization of GPCA to multiway arrays has to be done by assuming a Tucker1 structure for the multiway arrays and using GPCA on the properly matricized arrays.

2. CASE STUDY

3.1. Description of problem and data

The case study of this paper deals with stationary phases for reversed phase liquid chromatographic columns. One of the well-known problems with these stationary phases is that they differ between manufacturers and even between batches from the same manufacturer. This hampers the use of liquid chromatography, one of the most widespread routine methods of chemical analysis.

To assess the characteristics of a new stationary phase material, tests have been proposed and compared [28–30]. These tests usually involve eluting a set of test solutes on the column packed with a certain stationary phase and measuring the characteristics of the chromatographic peaks. For each peak these characteristics might include (i) the retention factor k (the position of the peak), (ii) the plate height h (a function of the width of the peak) and (iii) the asymmetry factor *USP* (the skewness of the peak). The retention factors of the test solutes give information about the *hydrophobicity* and *selectivity* of the stationary phase material. The plate heights of the peaks give information about the *efficiency* of the packed column. Finally, the asymmetry factors give information about *secondary retention mechanisms* due to e.g. free silanol groups at the surface of the stationary phase material.

In a large comparison study, 22 stationary phase materials were tested using different tests proposed in the literature [31]. Five of these tests were selected in the current case study and these are summarized in Table III. The variables characterizing the stationary phase were ln$k$, h and *USP* (for details see Reference [28]). The logarithm of k was used in the analyses for two reasons: (i) it linearizes the data [32]; (ii) it transforms a constant relative error as present in the k values into a constant absolute error in the ln$k$ values [33]. This transformation is therefore sensible if linear models are used for the subsequent analysis. After removing several outlying stationary phase materials (based on prior knowledge), this results in five data sets $\mathbf{X}_{\text{A}}$ (19×12), $\mathbf{X}_{\text{B3}}$ (19×12), $\mathbf{X}_{\text{B7}}$ (19×12), $\mathbf{X}_{\text{B}}$ (19×18) and $\mathbf{X}_{\text{N}}$ (19×15) (for the abbreviations see Table III).

The goal of the data analysis is to answer the following questions.

1. Which tests are similar (redundancy) and dissimilar (uniqueness)?
2. What are the differences between the tests?
3. What are the differences between the stationary phase materials?
4. In which tests do these differences show up?

Roughly stated, the answers to questions 1–4 are given by the superweights, blockloadings, superscores and blockscores respectively. This will be illustrated in the following, although a very detailed analysis of the tests is beyond the scope of this paper.

3.2. Analysis of individual data matrices

Before discussing the results of the multiblock methods, it is insightful to look at the data matrices separately. Prior to analysis, all data sets were column autoscaled and subsequently block scaled to unit sum of squares. The block scaling does not affect the results of performing a PCA on the separate data sets, but was done to be consistent with the multiblock analyses later which are all run with this type of scaling.

Table III. Summary of chromatographic tests

Name of test (abbreviation)	Solutes (abbreviations)	Measured properties	Mobile phase
Acidic (A)	4 acidic compounds (2nb, 4hb, 3cb, 2ab)	Lnk, h and *USP* ($4 \times 3 = 12$)	Water/organic modifier
Basic pH 3 (B3)	4 basic compounds (2hmp-b3, pht-b3, 3aq-b3, 4ea-b3)	Lnk, h and *USP* (total $4 \times 3 = 12$)	Buffer pH 3/organic modifier
Basic pH 7 (B7)	4 basic compounds (2hmp-b7, pht-b7, 3aq-b7, 4ea-b7)	Lnk, h and *USP* (total $4 \times 3 = 12$)	Buffer pH 7/organic modifier
Basic (B)	6 basic compounds (nor, dox, pro, tez, am, tim)	Lnk, h and *USP* (total $6 \times 3 = 18$)	Water/organic modifier
Neutral (N)	5 neutral compounds (t, e, p, a, py)	Lnk, h and *USP* (total $5 \times 3 = 15$)	Water/organic modifier

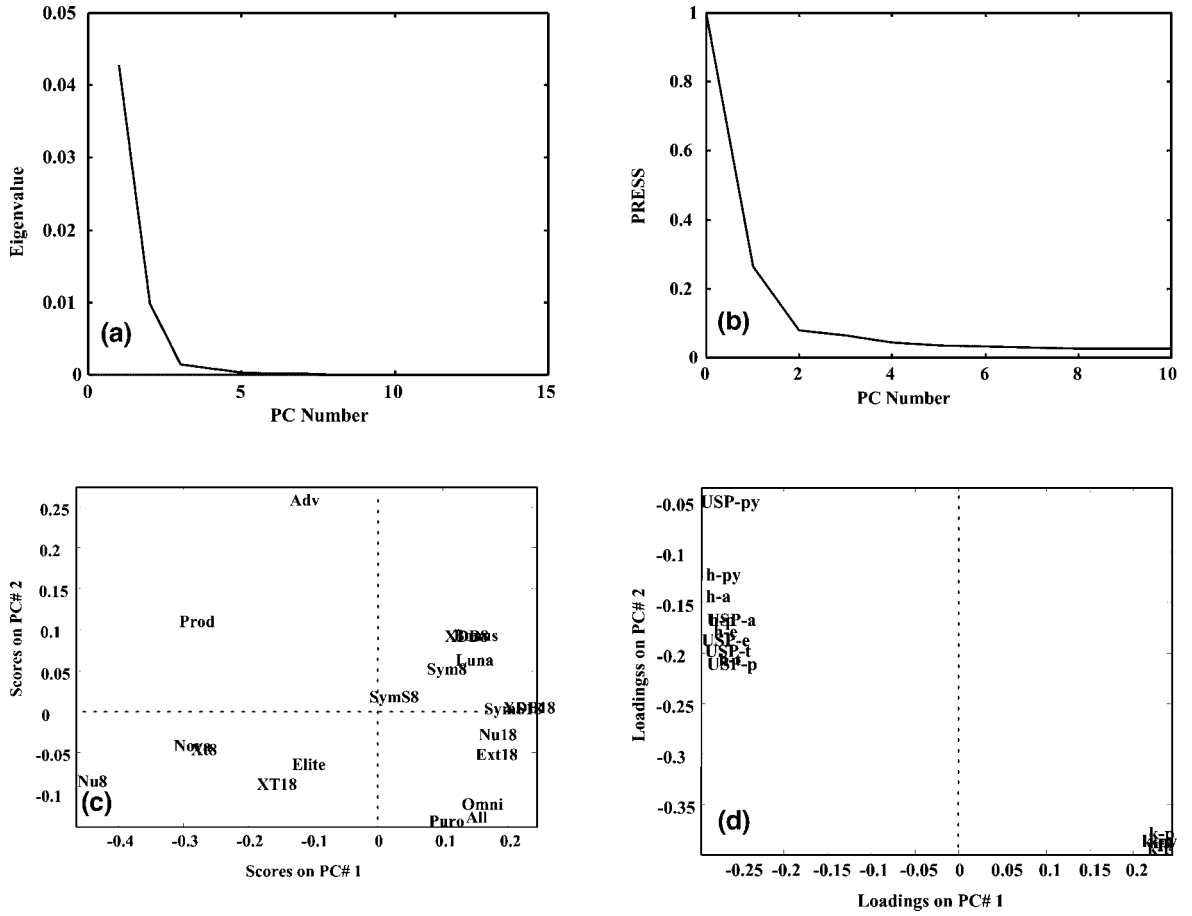

Figure 2. Results for neutral block: (a) eigenvalue plot; (b) cross-validation (PRESS = prediction error sum of squares); (c) score plot; (d) loading plot.

The results of $\mathbf{X}_N$ and $\mathbf{X}_{B7}$ are shown as representative of the set of five matrices. Figure 2 shows the results of analyzing the neutral data block ($\mathbf{X}_N$). In this figure the eigenvalues of $\mathbf{X}_N^T\mathbf{X}_N/(I\text{-}1)$ (a), the prediction error sums of squares (PRESS) using cross-validation (b), the score plot (c) and the loading plot (d) are shown. There are two dominant eigenvalues, and cross-validation shows that two PCs are significant. The loading plot shows a clear cluster of the ln$k$ values, and the first PC seems to distinguish between hydrophobicity/selectivity on the one hand and efficiency/secondary mechanism on the other hand.

Figure 3 shows the results for the basic pH 7 data block ($\mathbf{X}_{B7}$). Again, eigenvalues (a), PRESS values (b), scores (c) and loadings (d) are shown. The eigenvalue and PRESS plots suggest three components. The score and loading plots are different from those of $\mathbf{X}_N$, and this difference is not due to a simple rotation. Hence the basic pH 7 test gives a different view on the stationary phases than the neutral test.

3.3. Analysis of similarity between data sets

A first quick idea about the similarity between the data sets (tests) can be obtained by calculating their correlation. Matrix correlation is a generalization of variable correlation, and several matrix correlation methods exist [34]. Of these methods, the RV coefficient is the most appropriate for the current situation because it is orientation-independent (i.e. RV = 1 for $\mathbf{A} = \mathbf{B}$ up to scaling and orthogonal rotation) and easily calculated from the data [34,35]. The RV coefficient is calculated as

$$\mathrm{RV}(\mathbf{A},\mathbf{B}) = \frac{\mathrm{tr}(\mathbf{B}^T\mathbf{A}\mathbf{A}^T\mathbf{B})}{\{\mathrm{tr}[(\mathbf{A}^T\mathbf{A})^2]\mathrm{tr}[(\mathbf{B}^T\mathbf{B})^2]\}^{1/2}} \qquad (20)$$

where tr($\mathbf{A}$) stands for the trace of $\mathbf{A}$. There is clearly a resemblance between the correlation coefficient between two variables and Equation (20). The RV coefficient runs from 0 (no correlation) to 1 (perfect correlation).

The RV coefficients were calculated for the five data sets with $R = 4$ and the results are shown in Table IV. The basic tests have a reasonable intercorrelation, whereas the acidic and neutral tests are the most deviating. Table IV does not tell the whole story, because it only shows the *pairwise* (i.e. matrix-by-matrix) correlations. A type of multiple correlation coefficient can be obtained by regressing one block of data (e.g. $\mathbf{X}_A$) on the concatenation of all the others ([$\mathbf{X}_{B3}$ $\mathbf{X}_{B7}$ $\mathbf{X}_B$ $\mathbf{X}_N$]) and calculating how much variation in that block ($\mathbf{X}_A$) is explained by these other blocks. In order to make this regression feasible and non-trivial, each block is summarized by its first four principal components and these are used subsequently in the regression. From this regression a multiple correlation coefficient can be calculated. For the data sets $\mathbf{X}_A$, $\mathbf{X}_{B3}$, $\mathbf{X}_{B7}$, $\mathbf{X}_B$ and $\mathbf{X}_N$ these numbers are 0.9780, 0.9931, 0.9978, 0.9945 and 0.9935 respectively. This shows that the acidic test is probably the most atypical.

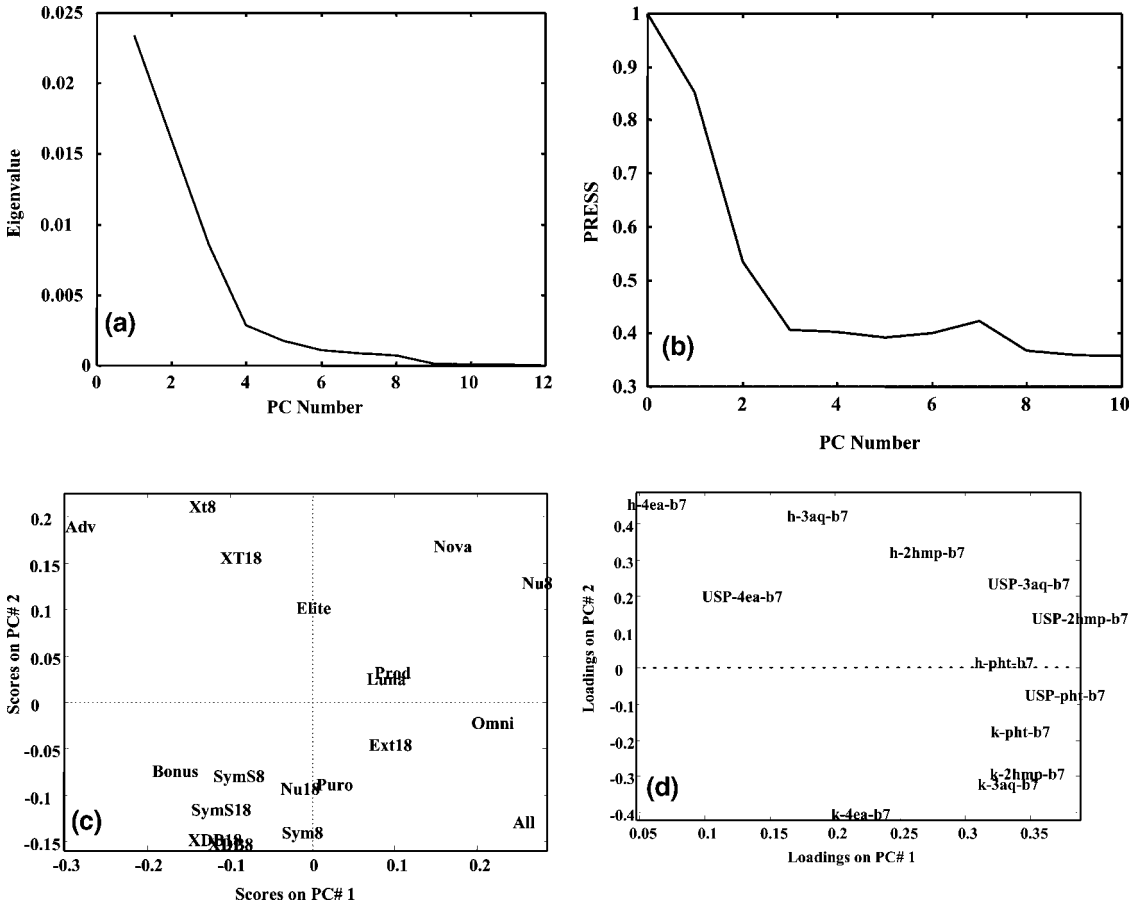

Figure 3. Results for basic pH 7 block: (a) eigenvalue plot; (b) cross-validation (PRESS = prediction error sum of squares); (c) score plot; (d) loading plot.

3.4. Consensus PCA

The results of applying CPCA to the data are reported in Table V. Each row in Table V shows the amount of explained variation by the (sequentially calculated) superscores (a) and blockscores (b) in each block of data and the total amount explained in all blocks. Concentrating on the amounts of variation explained by the superscores (Table V(a)), it seems that CPCA tries to explain information in all the blocks for each successive calculated superscore. This could be called 'consensus behavior'. Table V(b) shows the amount of explained variation in each block by the blockscores. These values are higher than the corresponding ones for the superscores, as expected, but they show a similar pattern (Table V(b)). The amount of explained variation in the acid block lags behind for all dimensions. The superweights are not shown in a plot because they are all equal to 0.2 for all dimensions and hence are not very informative.

The CPCA algorithm was also initiated with random numbers as starting vectors of the superscores. This showed that also modified CPCA was very sensitive to different initializations. The results obtained for the superscores and blockscores were different in each run; not only in their directions, but also in the spaces they spanned. Hence serious initialization problems are also present in CPCA.

Table IV. RV coefficients

	Acid	Basic3	Basic7	Basic	Neutral
Acid	1.00				
Basic3	0.29	1.00			
Basic7	0.38	0.78	1.00		
Basic	0.20	0.72	0.71	1.00	
Neutral	0.55	0.25	0.48	0.23	1.00

3.5. Consensus PCA-W (SUM-PCA)

The results of applying CPCA-W to the data are reported in Table VI. Table VI(a) shows a different pattern from Table V(a): larger 'jumps' in explained variation between the blocks as a function of the number of components are visible in CPCA-W. The first superscore explains mostly the similarity between the basic blocks, the second superscore explains mainly the neutral block, and the third superscore

Table V. CPCA results

# Comp.	Acid	Basic3	Basic7	Basic	Neutral	Total
(a) Superscores						
1	20.9	40.5	33.8	38.1	42.5	35.2
2	46.6	66.4	63.9	71.0	80.0	65.6
3	56.9	75.4	73.7	77.2	84.5	73.5
4	70.4	78.9	78.4	85.6	88.7	80.4
(b) Blockscores						
1	34.2	57.5	40.1	66.7	75.6	54.8
2	73.6	77.2	70.5	80.1	89.2	78.1
3	86.3	82.8	85.0	90.9	91.7	87.3
4	87.9	89.1	90.5	94.3	96.0	91.6

Table VI. CPCA-W results

# Comp.	Acid	Basic3	Basic7	Basic	Neutral	Total
(a) Superscores						
1	15.2	47.7	36.8	49.4	28.3	35.5
2	44.4	66.7	64.1	72.7	80.6	65.7
3	78.4	69.8	67.8	75.1	82.4	74.7
4	82.2	78.4	77.5	84	86.4	81.7
(b) Blockscores						
1	34.1	58.0	41.4	67.4	74.1	55.0
2	73.3	78.5	70.5	80.6	88.9	78.4
3	86.4	82.2	85.4	87.0	91.9	86.6
4	87.9	85.8	88.6	94.5	96.3	90.6

Table VII. HPCA results

# Comp.	Acid	Basic3	Basic7	Basic	Neutral	Total
(a) Superscores						
1	8.1	52.1	39.1	63.7	7.9	34.2
2	40.4	63.2	62.3	75.5	82.5	64.8
3	81.4	64.6	63.1	76.3	83.3	73.7
4	83.4	77.9	74.8	85.2	84.1	81.1
(b) Blockscores						
1	37.3	57.7	42.0	67.8	52.3	51.4
2	75.0	80.5	70.7	83.2	93.3	80.5
3	86.8	82.7	84.9	85.5	95.6	87.1
4	88.4	86.8	90.3	94.6	96.9	91.4

almost only explains the acid block. This behavior is similar to the results of the first simulation example in Reference [18], where the blocks had different underlying types of variation. To reduce the number of plots, only the ones illustrating differences between the methods are shown. Figure 4(a) shows the superweights of the second and third supercomponents of CPCA-W. This plot reflects the findings above.

After three components, each block is explained almost equally well. The CPCA-W results follow very closely the pattern of the multiple correlations between the blocks (see Section 3.3). Note that the total amount of explained variation (last column of Table VI(a)) is an upper bound for the analogous numbers for all other methods, because CPCA-W equals SUM-PCA and this method optimizes such explained variations.

3.6. Hierarchical PCA

The results of applying HPCA to the data are reported in Table VII. HPCA works as a 'block selector': the first superscore selects the basic blocks, the second mainly the acid and neutral blocks, whereas the third superscore selects almost exclusively the acid block. The pattern of explained variation by the superscores resembles that of CPCA-W but is more pronounced. This is also visible in Figure 4(b): it is very similar to Figure 4(a). Also the explained variation by the blockscores is very similar for HPCA and CPCA-W.

3.7. Hierarchical PCA-W

The five data sets were also subjected to a block-scaled HPCA-W analysis, and Table VIII shows the results. Clearly, the amount of explained variation in the acid block lags behind. HPCA-W discriminates between blocks and is able

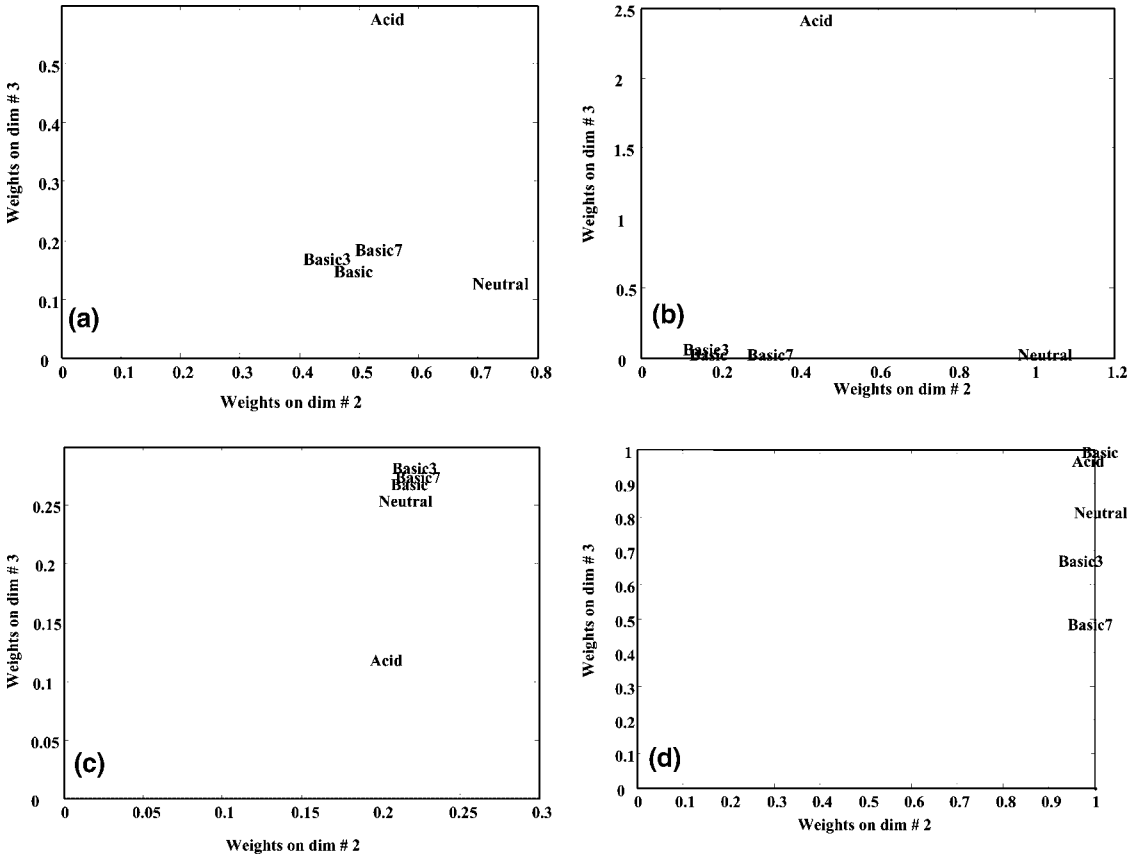

Figure 4. Plots of superweights: (a) CPCA-W; (b) HPCA; (c) HPCA-W; (d) GPCA.

Table VIII. HPCA-W results

# Comp.	Acid	Basic3	Basic7	Basic	Neutral	Total
(a) Superscores						
1	18.7	44.6	35.7	42.3	35.1	35.3
2	45.7	67.4	64.8	70.8	78.6	65.5
3	48.8	76.9	75.5	78.8	84.8	72.9
4	57	81.4	81.7	87.6	87.7	79.1
(b) Blockscores						
1	34.2	57.8	40.8	67.2	74.8	55.0
2	73.5	78.1	70.5	79.7	89.1	78.2
3	86.0	83.0	84.9	90.1	91.8	87.2
4	88.1	89.8	90.6	94.5	95.7	91.7

to find the deviating block. Also in the explained variation by the blockscores the acid block is lagging behind, although not so pronouncedly as in the results of the superscores.

The superweights of HPCA-W for the second and third dimensions are shown in Figure 4(c). This plot clearly shows that the acidic test is different and that this difference is described by superweight 3. This is also reflected in Table VIII(a), where, in going from two to three components, not much variance is explained by the superscores in the acid block, whereas all other blocks have a substantial increase in explained variation.

3.8. Generalized PCA

GPCA was applied to all block-scaled data sets, and the results are shown in Table IX. The amounts of explained variation by the superscores follow closely the ones of CPCA-W. The explained variations by the blockscores of GPCA (Table IX(b)) are considerably lower than the corresponding ones of HPCA-W and CPCA-W. The superweights for the second and third dimensions are shown in Figure 4(d). This shows the problem of blocks whose number of variables is large compared with the number of observations. This is e.g. the case with the basic block (19×18). The weights of this block are almost one. Figure 4(d) does not show a very realistic view on the structure of the current data set.

3.9. Comparing the five different methods

CPCA, CPCA-W, HPCA, HPCA-W and GPCA are calculated componentwise. Hence these methods can be compared componentwise. For CPCA-W and GPCA, each component has a specific meaning (see Section 2); this is less clear for CPCA, HPCA and HPCA-W. The componentwise comparison of the superscores of CPCA, CPCA-W, HPCA, HPCA-W and GPCA can be done simply by calculating the correlation coefficients between the corresponding successive superscores. The results are reported in Table X.

Table IX. GPCA results

# Comp.	Acid	Basic3	Basic7	Basic	Neutral	Total
(a) Superscores						
1	15.2	47.7	36.8	49.5	28.3	35.5
2	44.8	66.9	63.9	72.3	80.0	65.6
3	75.1	71.0	68.4	74.9	82.6	74.4
4	80.9	77.1	75.3	83.5	88.6	81.1
(b) Blockscores						
1	16.3	47.9	37.6	49.5	30.4	36.3
2	46.1	67.7	65.4	72.3	81.7	66.7
3	77.5	74.5	73.6	74.9	83.6	76.8
4	82.4	80.2	79.9	83.5	90.0	83.2

Table X. Correlation coefficients between supercomponents of CPCA, CPCA-W, HPCA, HPCA-W and GPCA[a]

	CPCA	CPCA-W	HPCA	HPCA-W	GPCA
CPCA	1				
	1				
	1				
	1				
CPCA-W	0.97	1			
	0.98	1			
	0.70	1			
	0.55	1			
HPCA	0.78	0.90	1		
	0.78	0.90	1		
	0.51	0.94	1		
	0.30	0.93	1		
HPCA-W	0.99	0.99	0.84	1	
	0.99	0.99	0.84	1	
	0.95	0.45	0.22	1	
	0.93	0.26	0.02	1	
GPCA	0.97	1	0.90	0.99	1
	0.97	1	0.89	0.99	1
	0.81	0.98	0.90	0.60	1
	0.77	0.92	0.75	0.55	1

[a]Correlation coefficients are reported dimensionwise (top to bottom: dimensions 1 to 4).

In all cases, dimensions 1 and 2 are very similar. CPCA-W is very similar to HPCA, which corroborates the findings in Tables VI and VII. CPCA is very similar to HPCA-W; this is to some extent reflected in Tables V and VIII. CPCA-W is also similar to GPCA, although the amount of explained variation by the blockscores is considerably lower for GPCA than for CPCA-W (see Tables VI(b) and IX(b)). The most deviating methods are HPCA and HPCA-W, showing the impact of the modification suggested by Westerhuis *et al*. [18].

The differences between the multiblock methods are also visible in the plots of the superscores (Figure 5). Only the plots of superscores 2 versus 3 are shown, because these are the most interesting for illustrating differences. The similarity between CPCA-W (Figure 5(b)) and HPCA (Figure 5(c)) is clear. Also CPCA (Figure 5(a)) and HPCA-W (Figure 5(d)) are again similar. HPCA (Figure 5(c)) and HPCA-W (Figure 5(d)) are very dissimilar (see e.g. the position of Bonus). Also the similarity between CPCA-W and GPCA is visible (Figures 5(b) and 5(e)). A detailed analysis of the superscores in terms of the properties of the stationary phases involved is beyond the scope of this paper.

Again a matrix correlation method can be used to compare the subspaces generated by the first four superscores of the different methods [34]. In this case the GCD coefficient is most suitable, since the interest is in the closeness of subspaces spanned by the superscores of the different methods [36]. The GCD coefficient of matrices **A** and **B** equals the average of the squared canonical correlation coefficients between these matrices. A GCD value of 1 means that two subspaces completely overlap, whereas a GCD value of 0 means that the two subspaces are orthogonal. The GCD results are presented in Table XI. This table provides a quick overview of the (dis)similarities between the methods and corroborates the findings in Table X and Figure 5.

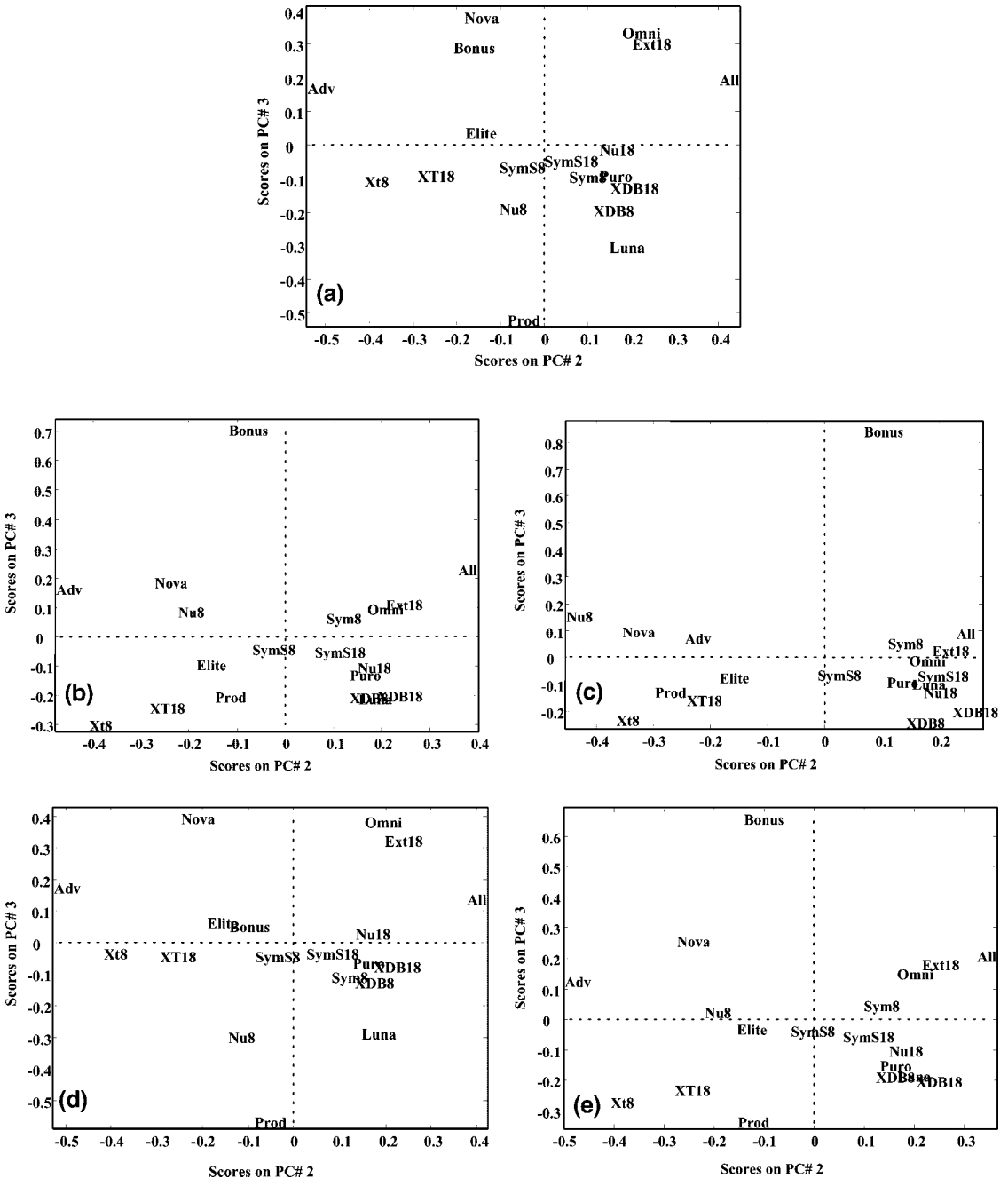

Figure 5. Plots of superscores: (a) CPCA; (b) CPCA-W; (c) HPCA; (d) HPCA-W; (e) GPCA.

3.10. Are the methods really multiblock?

An interesting question is whether the methods really give a multiblock solution or merely a 'collection of single-block principal component solutions'. One way to answer that question is to look at the differences between the blockscores generated by the different methods and the PCA scores of the individual blocks. The GCD measure was again used to investigate the differences between the subspaces spanned by the first four blockscores and the first four principal components of the individual blocks. Table XII shows the results.

All methods are really multiblock: there are clear differences between the blockscores and the individual principal

Table XI. GCD coefficients between superscore solutions (four components)

	CPCA	CPCA-W	HPCA	HPCA-W	GPCA
CPCA	1				
CPCA-W	0.92	1			
HPCA	0.85	0.97	1		
HPCA-W	0.96	0.85	0.79	1	
GPCA	0.94	0.96	0.91	0.85	1

Table XII. GCD coefficients of blockscores and PCA solutions (four components)

	Acid	Basic3	Basic7	Basic	Neutral
CPCA	0.73	0.76	0.93	0.75	0.66
CPCA-W	0.73	0.66	0.83	0.75	0.66
HPCA	0.74	0.69	0.92	0.78	0.74
HPCA-W	0.74	0.78	0.93	0.75	0.66
GPCA	0.69	0.58	0.72	0.60	0.56

components. The blockscores generated by GPCA differ the most from the individual PCA solutions. Again this corroborates the results of Table IX(b).

4. CONCLUSIONS AND FINAL REMARKS

Some multiblock methods have been described mathematically and illustrated with an example. CPCA-W and GPCA have clear optimization criteria, but this does not hold for CPCA, HPCA and HPCA-W. The methods CPCA and HPCA-W behave very similarly. This holds also for CPCA-W and HPCA on the one hand and for CPCA-W and GPCA on the other hand. The explained variation by the blockscores in GPCA is lower than in CPCA-W, and GPCA has undesirable properties when blocks are present with fewer objects than variables. This makes GPCA an unattractive alternative to CPCA-W.

One of the difficulties with multiblock methods is how to evaluate their performance. The data set in this study is too small to perform all kinds of stability tests. However, initial calculations indicate that CPCA, HPCA and HPCA-W are sensitive to starting values. A thorough analysis of the cause of this problem is not possible owing to the limited sample size of the case study, but an extensive simulation study is in progress.

For this problem, CPCA-W is probably the best choice. It has a clear optimization criterion, has built-in fairness and is not sensitive to starting values owing to its non-iterative nature.

Answers to the questions posed in Section 3.1 cannot easily be given. Concentrating on the CPCA-W results, the plot of the superweights (Figure 4(a))shows that the basic tests are interchangeable. Probably only one basic test, the acidic test and the neutral test are sufficient. The interpretation of the blockscores and blockloadings of CPCA-W is hampered by their non-orthogonality. This is the subject of a follow-up paper.

APPENDIX I. NOTATION

Scalars in italic lowercase; vectors in bold lowercase; matrices in bold uppercase; three-way arrays in bold underlined uppercase.

Symbol	*Size*	*Content*
$\mathbf{X}_b$	$I \times J_b$	bth block
$\mathbf{x}_{bj_b}$	$I \times 1$	j_bth variable in block $\mathbf{X}_b$
$\mathbf{X}$	$I \times \sum_{b=1}^{B} J_b$	concatenation of $\mathbf{X}_1, \ldots, \mathbf{X}_B$
$\mathbf{T}_b$	$I \times R_b$	blockscores of block b
$\mathbf{P}_b$	$J_b \times R_b$	blockloadings of block b
$\mathbf{W}_b$	$J_b \times R_b$	blockweights of block b
$\mathbf{T}_{\text{sup}}$	$I \times R$	superscores
$\mathbf{t}_{\text{sup},r}$	$I \times 1$	rth superscore
$\mathbf{t}_{br}$	$I \times 1$	rth blockscore of block b
$\mathbf{p}_{br}$	$J_b \times 1$	rth blockloading of block b
$\mathbf{T}_r$	$B \times 1$	collected blockscores of rth dimension (CPCA and HPCA)
$\mathbf{w}_r$	$B \times 1$	superweights of rth dimension (CPCA and HPCA)
$b = 1, \ldots, B$	scalar	running index for blocks
$i = 1, \ldots, I$	scalar	running index for objects
$j = 1, \ldots, J_b$	scalar	running index for variables in block b
$r = 1, \ldots, R$	scalar	running index for dimension (latent variable)
B	scalar	number of blocks
I	scalar	number of samples
J_b	scalar	number of variables in block b
R	scalar	number of superscores
c_b	scalar	amount of explained variation in block b

APPENDIX II. PROPERTIES

AII. 1. Proof of Equation (8)

At convergence of the CPCA algorithm (steps (e), (f) and (h) it holds that

$$\mathbf{t}_{\text{sup}} \propto \mathbf{T}\mathbf{T}^{\text{T}}\mathbf{t}_{\text{sup}} = \sum_{b=1}^{B} \mathbf{t}_b \mathbf{t}_b^{\text{T}} \mathbf{t}_{\text{sup}} \tag{21}$$

where the symbol $\propto$ indicates 'proportional to'. Anticipating other methods, Equation (21) also holds for CPCA-W, HPCA and HPCA-W. For CPCA it holds that

$$\begin{aligned} \mathbf{t}_b &= \mathbf{X}_b \mathbf{p}_b \left(\mathbf{p}_b^{\text{T}} \mathbf{p}_b\right)^{-1} \\ &= \mathbf{X}_b \mathbf{X}_b^{\text{T}} \mathbf{t}_{\text{sup}} \left(\mathbf{t}_{\text{sup}}^{\text{T}} \mathbf{t}_{\text{sup}}\right)^{-1} \left(\mathbf{t}_{\text{sup}}^{\text{T}} \mathbf{X}_b \mathbf{X}_b^{\text{T}} \mathbf{t}_{\text{sup}}\right)^{-1} \left(\mathbf{t}_{\text{sup}}^{\text{T}} \mathbf{t}_{\text{sup}}\right)^{2} \\ &= c_b^{-1} \mathbf{S}_b \mathbf{t}_{\text{sup}} \end{aligned} \tag{22}$$

using Equation (9). It also follows that

$$\mathbf{t}_b^{\text{T}} \mathbf{t}_{\text{sup}} = c_b^{-1} \mathbf{t}_{\text{sup}}^{\text{T}} \mathbf{S}_b \mathbf{t}_{\text{sup}} = \mathbf{t}_{\text{sup}}^{\text{T}} \mathbf{t}_{\text{sup}} \tag{23}$$

Hence

$$\mathbf{t}_{\text{sup}} \propto \sum_{b=1}^{B} \mathbf{t}_b \mathbf{t}_b^{\text{T}} \mathbf{t}_{\text{sup}} \propto \sum_{b=1}^{B} \mathbf{t}_b = \sum_{b=1}^{B} c_b^{-1} \mathbf{S}_b \mathbf{t}_{\text{sup}} \tag{24}$$

which shows that $\mathbf{t}_{\text{sup}}$ is an eigenvector of $\sum_{b=1}^{B} c_b^{-1} \mathbf{S}_b$, a weighted sum of the block cross-product matrices. The corresponding eigenvalue λ can be found by premultipying the eigenvalue problem by $\mathbf{t}_{\text{sup}}^{\text{T}}$:

$$\begin{aligned} &\sum_{b=1}^{B} c_b^{-1} \mathbf{S}_b \mathbf{t}_{\text{sup}} = \lambda \mathbf{t}_{\text{sup}} \\ &\Rightarrow \mathbf{t}_{\text{sup}}^{\text{T}} \sum_{b=1}^{B} c_b^{-1} \mathbf{S}_b \mathbf{t}_{\text{sup}} = \lambda \mathbf{t}_{\text{sup}}^{\text{T}} \mathbf{t}_{\text{sup}} \\ &\Rightarrow \lambda = \sum_{b=1}^{B} 1 = B \end{aligned} \tag{25}$$

From (24) it is clear that $\mathbf{t}_{\text{sup}}$ is proportional to a simple summation of the blockscores $\mathbf{t}_b$. Hence the weight vector $\mathbf{w}$ must be proportional to the unity vector $\mathbf{1}$. Normalization of $\mathbf{w}$ in original CPCA then requires $\mathbf{w} = B^{-1/2}\mathbf{1}$ (Equation (7)).

In modified CPCA (where the superscores are normalized to length one) it holds that $\mathbf{t}_{\text{sup}} = B^{-1} \sum_{b=1}^{B} c_b^{-1} \mathbf{S}_b \mathbf{t}_{\text{sup}} = B^{-1} \sum \mathbf{t}_b$, hence $w_b = B^{-1}$, giving $\mathbf{w} = B^{-1}\mathbf{1}$ (Equation (10)).

The weights for CPCA-W can be calculated along similar lines. Assuming that $\mathbf{t}_{\text{sup}}$ is normalized to length one, and

omitting the normalization of $\mathbf{w}$, the element w_b of $\mathbf{w} = (w_1, \ldots, w_B)^{\mathrm{T}}$ becomes

$$\begin{aligned} w_b = \mathbf{t}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} &= \mathbf{p}_b^{\mathrm{T}}\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} \\ &= \left(\frac{\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}}{\left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)^{1/2}}\right)\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} \\ &= \frac{\mathbf{t}_{\mathrm{sup}}^{T}\mathbf{X}_b\mathbf{X}_b^{T}\mathbf{t}_{\mathrm{sup}}}{(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}})^{1/2}} \\ &= \left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)^{1/2} = \|\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\| = \|\mathbf{p}_b\| = c_b^{1/2} \end{aligned} \tag{26}$$

AII. 2. Proof of equivalence of CPCA-W and SUM-PCA (Equation (12))

After convergence of the CPCA-W algorithm it holds that

$$\begin{aligned} \mathbf{t}_b &= \mathbf{X}_b\mathbf{p}_b \\ &= \mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)^{-1}\left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)\left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)^{-1/2} \\ &= \mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)^{-1/2} = c_b^{-1/2}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} \end{aligned} \tag{27}$$

using again that $\mathbf{t}_{\mathrm{sup}}$ has length one. Inserting this result into (21) yields the eigenvector problem

$$\begin{aligned} \mathbf{t}_{\mathrm{sup}} \propto \sum_{b=1}^{B}\mathbf{t}_b\mathbf{t}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} &= \sum_{b=1}^{B}c_b^{-1}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}}\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} \\ &= \sum_{b=1}^{B}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} \end{aligned} \tag{28}$$

which is equivalent to the SUM-PCA problem (Equation (3)).

AII. 3. Proof of Equation (14)

At convergence of the HPCA algorithm it holds that

$$\mathbf{t}_b = \mathbf{X}_b\mathbf{p}_b = \mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t} = \mathbf{S}_b\mathbf{t} \tag{29}$$

Inserting this result into (21) yields the eigenvector problem

$$\begin{aligned} \mathbf{t}_{\mathrm{sup}} \propto \sum_{b=1}^{B}\mathbf{t}_b\mathbf{t}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} &= \sum_{b=1}^{B}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}}\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} \\ &= \sum_{b=1}^{B}c_b\mathbf{S}_b\mathbf{t} \end{aligned} \tag{30}$$

showing $\mathbf{t}_{\mathrm{sup}}$ as an eigenvector of $\sum_{b=1}^{B}c_b\mathbf{S}_b$. The corresponding eigenvalue can be found by premultiplying Equation (30) by $\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}$:

$$\begin{aligned} \lambda\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} = \mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\left(\sum_{b=1}^{B}c_b\mathbf{S}_b\right)\mathbf{t}_{\mathrm{sup}} &\Leftrightarrow \\ \lambda = \sum_{b=1}^{B}c_b\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} = \sum_{b=1}^{B}c_b^2 \end{aligned} \tag{31}$$

AII. 4. Proof of Equation (15)

At convergence of the HPCA-W algorithm it holds that

$$\begin{aligned} \mathbf{t}_b &= \mathbf{X}_b\mathbf{p}_b/\|\mathbf{X}_b\mathbf{p}_b\| \\ &= \mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}/\left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{X}_b\mathbf{X}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}}\right)^{1/2} \\ &= \mathbf{S}_b\mathbf{t}_{\mathrm{sup}}\left(\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{S}_b^2\mathbf{t}_{\mathrm{sup}}\right)^{-1/2} = d_b^{-1/2}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} \end{aligned} \tag{32}$$

where $d_b = \mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{S}_b^2\mathbf{t}_{\mathrm{sup}}$. Inserting Equation (32) into (21) yields the eigenvector problem

$$\begin{aligned} \mathbf{t}_{\mathrm{sup}} \propto \sum_{b=1}^{B}\mathbf{t}_b\mathbf{t}_b^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} &= \sum_{b=1}^{B}d_b^{-1}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}}\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} \\ &= \sum_{b=1}^{B}c_bd_b^{-1}\mathbf{S}_b\mathbf{t}_{\mathrm{sup}} \end{aligned} \tag{33}$$

Equation (33) clearly shows of which expression $\mathbf{t}_{\mathrm{sup}}$ is an eigenvector. The corresponding eigenvalue can be found by premultiplying Equation (33) by $\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}$:

$$\begin{aligned} \lambda\mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\mathbf{t}_{\mathrm{sup}} = \mathbf{t}_{\mathrm{sup}}^{\mathrm{T}}\left(\sum_{b=1}^{B}c_bd_b^{-1}\mathbf{S}_b\right)\mathbf{t}_{\mathrm{sup}} \\ \Rightarrow \lambda = \sum_{b=1}^{B}c_b^2d_b^{-1} \end{aligned} \tag{34}$$

AII. 5. Proof that PCA maximizes summed covariances (or correlations)

Let $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_J]$ $(I \times J)$ be column-centered and $\mathbf{t}_1$ be the first principal component of $\mathbf{Z}$. Then $\mathrm{cov}(\mathbf{z}_j, \mathbf{t}_1) = (1/I)\mathbf{t}_1^{T}\mathbf{z}_j$. Hence

$$\begin{aligned} \sum_{j=1}^{J}\mathrm{cov}^2(\mathbf{t}_1, \mathbf{z}_j) &= \left(\frac{1}{I}\right)^2\sum_{j=1}^{J}\left(\mathbf{t}_1^{\mathrm{T}}\mathbf{z}_j\right)^2 \\ &= \left(\frac{1}{I}\right)^2\sum_{j=1}^{J}\mathbf{t}_1^{\mathrm{T}}\mathbf{z}_j\mathbf{z}_j^{\mathrm{T}}\mathbf{t}_1 = \left(\frac{1}{I}\right)^2\mathbf{t}_1^{\mathrm{T}}\left(\sum_{j=1}^{J}\mathbf{z}_j\mathbf{z}_j^{\mathrm{T}}\right)\mathbf{t}_1 \\ &= \left(\frac{1}{I}\right)^2\mathbf{t}_1^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\mathbf{t}_1 \end{aligned} \tag{35}$$

and the last form of (35) is exactly what PCA maximizes. For autoscaled $\mathbf{Z}$ and $\mathrm{var}(\mathbf{t}_1) = 1$ the covariances are correlations. The second principal component $\mathbf{t}_2$ is found orthogonal to $\mathbf{t}_1$ and using a similar expression.

REFERENCES

1. Coppi R, Bolasco S. *Multiway Data Analysis*. North-Holland: Amsterdam, 1989.
2. Law HG, Snyder CW, Hattie JA, McDonald RP. *Research Methods for Multimode Data Analysis*. Praeger: New York, 1984.
3. Van der Geer JP. Linear relations among k sets of variables. *Psychometrika* 1984; **49**: 79–94.
4. Smilde AK, Westerhuis JA, Boqué R. Multiway multiblock component and covariates regression models. *J. Chemometrics* 2000; **14**: 301–331.

5. Kettenring JR. Canonical analysis of several sets of variables. *Biometrika* 1971; **58**: 433–451.
6. Levin J. Simultaneous factor analysis of several gramina matrices. *Psychometrika* 1966; **31**: 413–419.
7. Kiers HAL, Ten Berge JMF. Alternating least squares algorithms for simultaneous components analysis with equal components weight matrices in two or more populations. *Psychometrika* 1989; **54**: 467–473.
8. Kiers HAL, Ten Berge JMF. Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *Br. J. Math. Statist. Psychol.* 1994; **47**: 109–126.
9. Nierop AFM. Multidimensional analysis of grouped variables: an integrated approach. *PhD Thesis*, University of Leiden, 1993.
10. Ten Berge JMF, Kiers HAL, Van der Stel V. Simultaneous component analysis. *Statist. Appl.* 1992; **4**: 277–392.
11. Timmerman ME. Component analysis of multisubject multivariate longitudinal data. *PhD Thesis*, University of Groningen, 2001.
12. Casin Ph. A generalization of principal component analysis to *K* sets of variables. *Comput. Statist. Data Anal.* 2001; **35**: 417–428.
13. Frank IE, Kowalski BR. A multivariate method for relating groups of measurements connected by a causal pathway. *Anal. Chim. Acta* 1985; **167**: 51–63.
14. Wangen LE, Kowalski BR. A multiblock partial least squares algorithm for investigating complex chemical systems. *J. Chemometrics* 1988; **3**: 3–20.
15. Wold S, Hellberg S, Lundstedt T, Sjostrom M, Wold H. PLS modeling with latent variables in two or more dimensions. *Proceedings PLS-meeting*, Frankfurt, Germany, 1987; 1–21.
16. Wold S, Kettaneh-Wold N, Tjessem K. Hierarchical multiblock PLS and PC models, for easier model interpretation, and as an alternative to variable selection. *J. Chemometrics* 1996; **10**: 463–482.
17. Qin SJ, Valle S, Piovoso MJ. On unifying multiblock analysis with application to decentralized process monitoring. *J. Chemometrics* 2001; **15**: 715–742.
18. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemometrics* 1998; **12**: 301–321.
19. Kourti T, Nomikos P, MacGregor JF. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *J. Process Control* 1995; **5**: 277–284.
20. Rännar S, MacGregor JF, Wold S. Adaptive batch monitoring using hierarchical PCA. *Chemometrics Intell. Lab. Syst.* 1998; **41**: 73–81.
21. Gower JC. Generalized Procrustes analysis. *Psychometrika* 1975; **40**: 33–51.
22. Peay ER. Multidimensional rotation and scaling of configurations to optimal agreement. *Psychometrika* 1988; **53**: 199–208.
23. Tauler R, Kowalski BR, Flemming S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal. Chem.* 1993; **65**: 2040–2047.
24. Tauler R, Smilde AK, Kowalski BR. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemometrics* 1995; **9**: 31–58.
25. Wold S, Kettaneh-Wold N, Friden H, Holmberg A. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics Intel. Lab. Syst.* 1998; **44**: 331–340.
26. Geladi P, Martens H, Martens M, Kalvenes S, Esbensen K. Multivariate comparison of laboratory measurements. *Symp. on Applied Statistics*, Copenhagen, 1988; 15–30.
27. Höskuldsson A. PLS regression methods. *J. Chemometrics* 1988; **2**: 211–228.
28. Claessens HA, Van Straten MA, Cramers CA, Jezierska M, Buszewski B. Comparative study of test methods for reversed-phase columns for high-performance liquid chromatography. *J. Chromatogr.* 1998; **826**: 135–156.
29. Daldrup T, Kardel B. A simple test to characterise C8/C18-bonded stationary phases, particularly for toxicological screening. *Chromatographia* 1984; **18**: 81–83.
30. Engelhardt H, Arangio M, Lobert T. A chromatographic test procedure for reversed-phase HPLC column evaluation. *LC GC Int.* 1997 December (1997); 803–812.
31. Claessens HA, Van Straten MA, Boelens HFM, Smilde AK. Multiple column tests for reversed-phase liquid chromatographic columns. Part 1: A comparative chromatographic study. *J. Chromatogr.* 2003; in press.
32. Schoenmakers PJ. *Optimization of Chromatographic Selectivity*. Elsevier: Amsterdam, 1986.
33. Atkinson AC. *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Clarendon: Oxford, 1985.
34. Ramsay JO, Ten Berge JMF, Styan GPH. Matrix correlation. *Psychometrika* 1984; **49**: 403–423.
35. Robert P, Escoufier. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl. Statist.* 1976; **25**: 257–265.
36. Yanai H. Unification of various techniques of multivariate analysis by means of generalized coefficient of determination (GCD). *J. Behaviormetrics* 1974; **1**: 45–54.