RESEARCH ARTICLE

Journal of CHEMOMETRICS WILEY

Making sense of multiple distance matrices through common and distinct components

Lars Erik Solberg[1] | Tobias Dahl[2,3] | Tormod Næs[1,4]

[1]Nofima AS, Ås, Norway

[2]Smart Sensor Systems, SINTEF Digital, Oslo, Norway

[3]Department of Informatics, University of Oslo, Oslo, Norway

[4]Department of Food Science, University of Copenhagen, Copenhagen, Denmark

Correspondence
Lars Erik Solberg, Nofima AS, Oslovegen 1, 1430 Ås, Norway.
Email: lars.erik.solberg@nofima.no

Abstract

Multiblock analysis attacks the problem of how to combine data from various data sources for purposes such as prediction, classification, clustering, or visual data analysis. A key concept is the distinction between "common" and "distinct" parts, that is, what information repeats itself across the blocks and what is unique to an individual block.

The statistical field of multiblock analysis holds many different approaches, which leads to different treatments both of the terms *distinct* and *common* themselves and to differences in the numerical results. In this article, we extend the discussion of distinct and common in multiblock analysis to the domain of *distance matrices*, that is, the situation where data point sets, so-called *configurations*, are analyzed via relative distances either because configurations are not available directly or because a distance representation is favorable. Situations typical for chemometrics will be highlighted and illustrated in examples.

When analyzing different methods, we have focused on three key aspects. First, during the transition from the distance to configuration domains, one needs to consider how multiple distance matrices are treated. Second, when extracting common and distinct parts, one needs to manage a tradeoff between explaining variance and ensuring similarity between subspaces. Third, there is a design choice to be made as to whether the subspace containing the common parts is "shared" between blocks or if separate subspaces are associated with each individual block. The three aspects help to categorize and explain well-known methods in the field. A selection of methods was analyzed and subsequently applied to examples.

KEYWORDS

common, consensus, distances, distinct, multiblock, multidimensional scaling

---

1 | INTRODUCTION

Distance data are relevant in several domains and have been used extensively in psychology and sociology based on notions of "similar" and "dissimilar," or rankings, to quantify the distance between sets of concepts, categories, samples, and so forth.[1,2] A similar application occurs in sensory analysis where distances between products, for instance wines, are used to map these onto a sensory map and using frequencies of word descriptions to interpret the meaning of the coordinate axes.[3] In a completely different domain, for localization of "objects," it is frequently necessary to combine distance estimates relative to base stations, or transmitter/emitters, to determine the coordinates of these objects. Applications include time of arrival from access points for indoor localization in WIFI networks or signal strength to multiple beacons for localization in low energy bluetooth,[4] the latter being particularly relevant for smartphones.

Even when the original representation is not in the form of distance data, it may be convenient to use distances in some analyses. For instance, when fusing data sources of very different formats due to differing dimension or to variables being of different types such as binary and continuous.[5,6] Another example is when prior information is most easily incorporated in the form of UniFrac[7] distance matrices, such as phylogenetic information about microbial species.

In psychology, sociology, and in sensory analysis, a basic work horse for analyzing distance data to obtain a low dimensional representation is classic multidimensional scaling (MDS) or variations thereof.[1] This tool takes a distance matrix and produces a coordinate representation, hereafter called "configuration," with a predefined dimension (typically a low-dimensional representation) such that distances between samples approximate the original distance matrix. We will later briefly present the principles of MDS.

In the sensory example above, it is common to use a larger panel of tasters to arrive at a consensus, a kind of "average" interpretation of samples.[3] This process involves multiple distance matrices, one for each assessor. Also, for the other examples of data fusion, one ends up with a collection of distance data describing relationships between the same set of samples. Of course, to analyze data, an obvious approach would be to convert the distance matrices to low-dimensional configurations and then proceed from there. However, methods also exist for handling multiple distance data, "blocks," and two such methods are DISTATIS[8] and INDSCAL.[1] A key step for the former is a form of averaging in the distances domain while the latter employs a specific model on distances for the individual blocks.

Extracting information from multiple blocks of data, data fusion, is a rich field and applies to standard data tables, "configurations." An important question in that context is what the different blocks have in common and what distinguishes between them. A review of approaches for analyzing common and distinct for configurations is given in Smilde et al.[9] However, to the knowledge of the authors, little work has been done to extend this analysis to multiple blocks of distance data.

The objective of this article is to investigate the analysis of common and distinct scores from sets of distance matrices. We will present a framework for categorizing and analyzing different approaches. This framework describes three axes along which methods may be categorized and which we believe are relevant for understanding properties of the different approaches. We will also apply the framework to several existing methods, for instance, DISTATIS and INDSCAL mentioned above. We also think the framework is relevant to describe the span of possible approaches to describing common and distinct scores for distance matrices.

Three examples will be presented to illustrate the use of the framework. The first example is a simple simulation and will be used to illustrate the variance-correlation trade-off. The second example is from sensory analysis of food (Figure 1A) where the individual distance matrices represent the tasters' judgements of differences between olive oils. The focus is on obtaining a consensus among tasters. The third example is from pharmacogenomics where the blocks represent different measurement principles for tumor cell lines of cancers. The data considered are given as drug response, proteomics, gene expression, and copy number alterations. The focus here is to get insight into the common and distinct variability.

While we believe the framework is relevant for categorizing and understanding the different approaches, our ambition is not to provide a hierarchy along "better" to "worse" among them. What is recommended may be dependent both on the actual data and the aims of the analysis. Also, treatments of the above examples are more illustrative of pitfalls than of best practice.

A brief note on terminology: We follow the convention that bold face letters represent matrices (upper case) and vectors (lower case) while italics denote scalars. $R(\mathbf{A})$ denotes the range of $\mathbf{A}$ while $r(\mathbf{A})$ denotes its rank. Unless
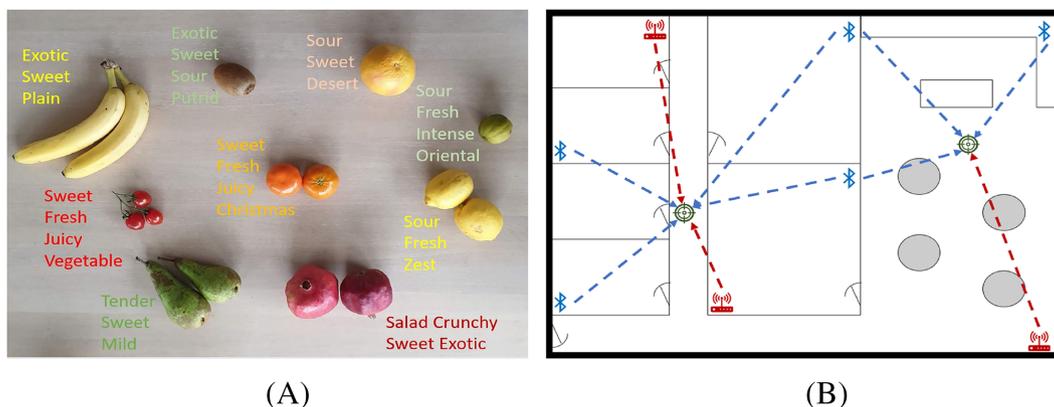
FIGURE 1   (A) In projective mapping (PM) sensory analysis, assessors are asked to place similar samples close to one another on a piece of paper, thereby creating a set of distances between the fruits. The axes are arbitrary and may vary from assessor to assessor. (B) An illustration of the concept of using distances to locate receivers in Wi-Fi and Bluetooth networks, or combinations thereof as illustrated here. Here, the distances may be proxies such as received signal strength indicators. Any three distances to known references is ideally sufficient for pin-pointing the location in the two dimensional plane

otherwise stated, the norm $\|\mathbf{A}\|$ is the Frobenius norm for matrices while $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the inner product between vectors $\mathbf{x}, \mathbf{y}$.

In Section 2, we will present some background material and introduce the framework as well as applying it to a selection of methods: MDS with generalized canonical correlations analysis (GCA), DISTATIS, and INDSCAL. These methods will be illustrated through application to examples in Section 3. A discussion is given in Section 4.

2 | FRAMEWORK

We begin with a very brief introduction to multidimensional scaling (MDS) as this is an important tool underlying the different approaches addressed later in the article, see Dokmanic[10] and the book Modern MDS[1] for more information.

2.1 | Multidimensional scaling (MDS)

The distance matrices are often Euclidean distance matrices (EDM) $\mathbf{D}$, which contain the squared distances between all pairs of samples (i,j).

$$\mathbf{D}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \|\mathbf{x}_i\|^2 - 2\langle \mathbf{x}_i, \mathbf{x}_j \rangle + \|\mathbf{x}_j\|^2 \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_N]^{\mathbf{T}}$ represents the original, perhaps unknown, configuration containing N samples and each $\mathbf{x}_i$ is an M-dimensional vector representing the sample. Without lack of generality, assume the columns are centered. It can be shown that by applying an operator $(\mathbf{J} = \mathbf{I} - \frac{\mathbf{11}^{\mathbf{T}}}{\mathbf{N}})$ to the EDM, effectively centering the configuration, and appropriately scaling the result we are left with a positive semi-definite matrix $\boldsymbol{G} = -\frac{1}{2}\boldsymbol{JDJ} = \boldsymbol{XX}^{\boldsymbol{T}}$,[10] called the Gram matrix. The positive definiteness means an eigenvector decomposition is possible with eigenvectors in $\mathbf{V}$ and positive real eigenvalues along the diagonal of Λ and hence its square-root is well defined. We can calculate a configuration using this square root:

$$\boldsymbol{Z} = \boldsymbol{V}_m \boldsymbol{\Lambda}_{\mathrm{m}}^{\mathbf{1/2}}. \tag{2}$$

where the subscript m indicates the dimension of the approximation to the original $\mathbf{X}$. This result is the core of MDS. A cause for differences between the original $\mathbf{X}$ and approximation $\mathbf{Z}$ is the invariance of distances $\mathbf{D}$ to both rotations and translations: the choice of reference system is arbitrary. Incidentally, this motivates the above assumption of column centered data. In the following, the matrix $\mathbf{Z}$ will be denoted "scores."

2.2 | Axes for classifying methods

As stated in the introduction, the basis for analysis is multiple blocks of distances, from which common and distinct configurations shall be calculated. As a motivational example, in projective mapping (see Figure 1A) a set of tasters place samples on a table such that distance reflects dissimilarity of samples with no prior instruction regarding which qualities to judge. In this scenario, two assessors may judge "sweetness" along one axis but switch between "sourness" and "bitterness" along the second. Ideally, small differences in perceived sweetness would be found in differences within the common subspaces, and sourness and bitterness may be assigned distinct subspaces.

We will in this paper consider methods which all are constructed upon the same workflow: the use of a set of K distance matrices $\mathbf{D}_k$, where rows and columns in different blocks refer to the same set of samples, to construct a consensus $\mathbf{V}$ and a set of common scores $\mathbf{Z}_{kc}$ and distinct scores $\mathbf{Z}_{kd}$:

$$\{\boldsymbol{D}_k\} \rightarrow (\boldsymbol{V}, \{(\boldsymbol{Z}_{kc}, \boldsymbol{Z}_{kd})\}) \text{ such that } \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I} \text{ and } \forall k, \boldsymbol{Z}_{kc} \perp \boldsymbol{Z}_{kd} \tag{3}$$

where K refers to block, c to common, and d to distinct, see Figure 2. All methods aggregate information across blocks to define the consensus $\mathbf{V}$, an orthonormal basis of this subspace. The consensus can be conceptualized as the axes shared by all blocks—the "level" playing field. The common configurations for each block $\mathbf{Z}_{kc}$ should lie either within this subspace ("on the field"), or at least close to it.

How individual methods formulate the optimization problem defines how the consensus is calculated and how common scores relate to the consensus. The distinct scores are further defined to be orthogonal to the common scores.

In the following, we will denote by "common subspace" the subspace that is spanned by the scores (columns of $\mathbf{Z}_k$), and sometimes will address elements of a basis of this subspace as "common components." These "components" are basically synonymous with the individual scores (column vectors in $\mathbf{Z}_k$) in the sense that they span the same subspace. Finally, as noted earlier, a "configuration" is the collection of N (row) samples and the subspace spanned by its column vectors is sometimes implicitly implied in this article as "column space". Similar remarks apply when talking about distinct "subspaces," "configurations," and "components."

The methods discussed in this paper may be categorized along three conceptual "axes" which will be called:

- "domain shift,"
- "variance-correlation tradeoff," and finally the
- "within-between requirement."

For methods discussed in this article, we will emphasize their relations to the axes, and believe this will make it easier to see similarities and differences between them.

2.2.1 | Domain shift

This axis addresses how the methods transform the data from the distances domain to the configurations domain. The different methods may be classified into three groups: "MDS first", "averaging of distances" and "direct".

MDS first: Distances $\{\boldsymbol{D}_k\}$ to total configurations $\{\boldsymbol{Z}_k\}$
In multiblock analysis, there is significant work on the decomposition of feature blocks into common and distinct configurations. Therefore, it appears that the easiest route to a decomposition of distance matrices is to first calculate block scores $\mathbf{Z}_k$ using MDS and then analyze these using the above-mentioned framework for features. This will give both
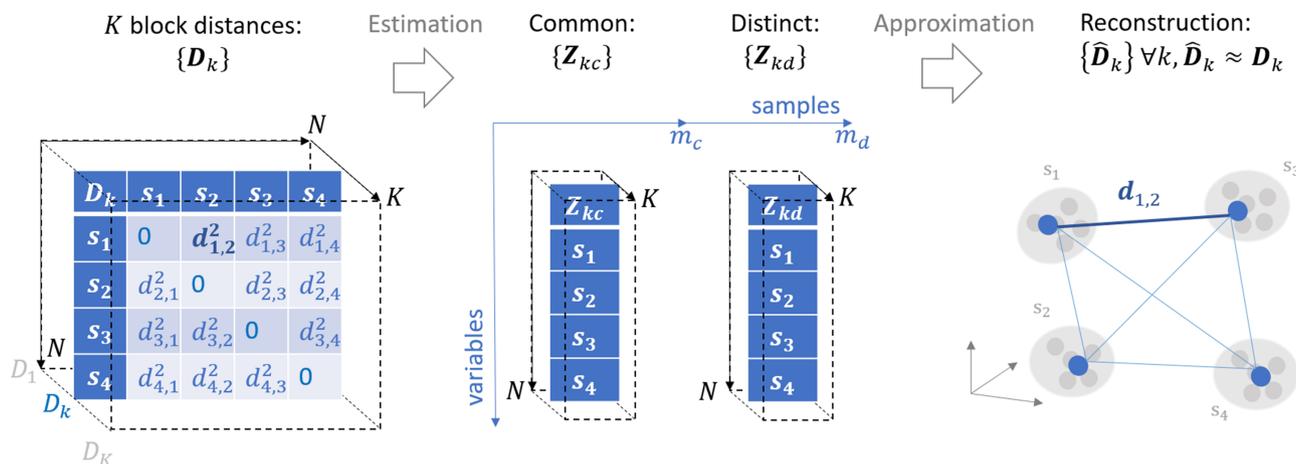
FIGURE 2 Illustrating the workflow. There are four samples, with all pairwise distances measured in K blocks (assessors, instruments, time, etc): $\{D_K\}$, see left part of figure. Here, $d_{1,2}$ represents the distance between samples 1 and 2 for block k (blue). From these blocks, common and distinct configurations are estimated and shown in center. To the right the recalculation of distances from configurations approximates the original distances. The common configurations have dimension m_c while the distinct configurations have dimension m_d; each dimension (column) represents a latent variable. The variables of the common configurations are related to the consensus while variables of the distinct configurations are orthogonal to the variables of common configurations of the same block

common and distinct scores through the use of established methods (e.g., GCA, SCA, JIVE, and DISCO, see Smilde[9]). This is called the "MDS first" group.

Averaging of distances: defining a distance-to-configuration operator T
The second group of approaches aggregates information directly in the distance domain which leads to a single distance matrix. Applying MDS to this matrix defines the consensus. Regarding the common scores, several approaches exist including the definition of a common subspace from the analysis of **D** and analyzing the blocks w.r.t. this subspace; or defining a transform T based on the eigen-decomposition and applying it to the block distances. An example of the latter is DISTATIS[8] which will briefly be discussed in Section 2.3.2.

Direct: Distances $\{D_k\}$ to common configurations $\{Z_k\}$
The last group is denoted "direct" because the original distance matrices are used directly as input to the methods without any prior processing other than a possible standardization. Common configurations are a direct result of the original distance data, such as for the well-known method INDSCAL.[1]

In the two latter groups, "direct" and "averaging of distances," a second stage estimates the distinct parts for each block. Our approach has been to apply a constrained MDS on the original distance matrices which will be addressed in Section 2.3.3.

2.2.2 | The variance-correlation tradeoff

Different viewpoints on commonness can be envisioned. One perspective aims at finding common scores with high resemblance across blocks. Another perspective emphasizes the stability of estimates as it is often desirable that the common scores "explains" a large part of the observations. The first objective favors correlation between the common scores while the second requires that reconstructed Gram matrices are good approximations to the original Gram matrices. The well-known method simultaneous components analysis (SCA[11]) belongs to the last group, while generalized canonical correlation analysis (GCA[9]) belongs to the first group.

These two objectives are not always aligned, and one is left with a trade-off, see Section 3.2 for an example. In Dahl and Næs,[12] a method for seamlessly compromising between maximizing correlation and explained variance was described. In a similar fashion, we will show that the concrete methods considered in this article can be parametrized

by the power α of the eigenvalues of the individual Gram matrices, and that this parameter implicitly adjusts the variance-correlation tradeoff.

All the methods to be considered below can be formulated as an optimization of a sum of Frobenius norms measuring the residuals between data and model. An intermediate step defines the consensus **V**—an orthonormal basis—and which may be formulated, for several of the methods, as an eigenvector problem involving an appropriate semi-definite, positive matrix **G** defined as:

$$G = \sum_k \omega_k V_k \Lambda_k^\alpha V_k^T,$$

(4)

where ω_k is a scalar scaling factor, $\mathbf{V}_k, \Lambda_\mathbf{k}$ are the eigenvectors and values of the blocks' Gram matrices $\mathbf{G}_k$. For the methods considered here, α is an integer in the set $\{0,1,2\}$. The scaling factors ω_k could be rescaling of the individual Gram matrices so they are more easily comparable. In all cases the m_c first eigenvectors define the consensus.

α = 0

This value means that all eigenvectors will have the same weight in the sum, the ones with the larger eigenvalues and those with the smaller eigenvalues. The first eigenvector can be shown to be a function of principal angles to blocks and will favor small angles, or equivalently high correlation. The story concerning the subsequent eigenvectors is more complicated. This behavior is relevant for the MDS first followed by GCA, see Section 2.3.1. It is important to note that some of the effects of high sensitivity to eigenvectors with small eigenvalues can be avoided if each of the individual MDS decompositions are truncated before being incorporated in the sum in Equation (4).

α = 1

Contrary to the previous case, for α = 1 the scale of the Gram matrices will influence the results as the "large" blocks (either with large norms or eigenvalues) will contribute most to **G** and hence pull considerably on the consensus. Examples where this is the case is DISTATIS and MDS followed by SCA.

α = 2

Finally, the case α = 2 applies when using SCA directly on distance matrices: "direct" SCA (see Appendix A.2 for details). In this case, directions corresponding to large eigenvalues are even more favored compared to the previous case.

As can be seen, this shows that α is one way of parametrizing the different problems and sets the tradeoff between correlation and explained variance. The choice of tradeoff in large part defines the interpretation of "common" between "similarity" and "best explained." For a simulated example, see Section 3.2.

2.2.3 | The within-between requirement

The third "axis" qualifying the different methods concerns the question of whether each common subspace should be contained within the span of each block, or if they may lie in the span of the set of blocks (this case is denoted "between"). Whether solutions "within" or "between" are acceptable depends on the specific application. For instance, projective mapping (PM) focuses on the consensus, which necessarily lies "between" the blocks.

On the other hand, different blocks may come from qualitatively different sources, such as different instruments or a combination of instruments and questionnaires. In this case, it may be important that the common configurations represent each part faithfully and one way to do so is to require it to lie within the range of the respective blocks, $R(\mathbf{G}_k)$. This implies that the common configurations can be expressed as linear combinations of the original variables and therefore be viewed as latent variables within the individual blocks.

Another distinction applicable to "between" versus "within" has to do with the total number of dimensions spanned by common scores. For instance, when parametrizing common to m_c components for the common space "between" in

SCA, this space contains *all* common scores. With the same choice for MDS followed by GCA, each block has dimension m_c which means that they collectively likely span a larger space.

The choice for the common subspace immediately implies a choice for the distinct subspace as they have been defined to be orthogonal. When the common subspace lies within, it is natural to define the distinct subspace within, too, so it can be extracted from the projection onto the orthogonal complement (for instance, using the m_d principal components). When the common subspace lies between, the orthogonal complement also lies between. It is possible to add constraints to force the distinct subspace to lie within even in this case for instance using the method described in Section 2.3.3 or by further projections onto the space spanned by the block.

2.3 | Selected methods

The methods discussed in the next sections are intended as examples of possible approaches and have been selected to illustrate the "axes" discussed above. Furthermore, they will be applied to example analyses in Section 3. It will be explicitly stated where they belong in the general framework above, that is, how they relate to the three axes. After discussing the three methods, we will give a brief review of other possibilities in the same framework and present them in a table.

2.3.1 | Method 1—MDS first followed by GCA

This method is an example from the "MDS first" methods, which favors correlation over explained variance and which results in common and distinct scores within blocks. The method starts by extracting total scores from each distance matrix separately.

$$MDS(\boldsymbol{D}_k, m_k) = \boldsymbol{Z}_k \in \mathbb{R}^{N \times m_k} \tag{5}$$

where m_k is the dimension of the space containing the total scores. As we propose below a method for extracting the distinct part from $\boldsymbol{Z}_k$, m_k should account for both the dimensions of the common and distinct spaces: $m_k \geq m_{kc} + m_{kd}$. We will comment on this choice below.

Generalized canonical correlation analysis (GCA[9]) is then applied to the set of total scores:

$$\min_{\boldsymbol{V}, \boldsymbol{P}_k} \sum_k ||\boldsymbol{Z}_k \boldsymbol{P}_k - \boldsymbol{V}||^2, \tag{6}$$

where the consensus $\boldsymbol{V}$ is orthogonal and of rank m_c. The solution to the consensus $\boldsymbol{V}$ are the first m_c eigenvectors of $\sum_k \boldsymbol{V}_k \boldsymbol{V}_k^T$, see Smilde et al.,[9] and as such has parameter α equal to zero as discussed in Section 2.2.2. Note again that some of the effects of the directions with smaller eigenvalues may be avoided if each of the individual MDS operations are truncated prior to (6).

The product $\boldsymbol{Z}_k \boldsymbol{P}_k$ approximates the orthogonal matrix $\boldsymbol{V}$ and does not lead to a good approximation of the original Gram matrices because these are not necessarily normalized. This product effectively identifies a subspace in $R(\boldsymbol{Z}_k)$, so the common space for this block should be the $m_k c$ principal components within this subspace. Let $\boldsymbol{U}_{kc}$ be an orthonormal basis for $\boldsymbol{Z}_k \boldsymbol{P}_k$, then define $\boldsymbol{Z}_{kc}$ using the m_{kc} largest principal components of $\boldsymbol{U}_{kc} \boldsymbol{U}_{kc}^T \boldsymbol{Z}_k$. It should be stressed that this is not normally a part of GCA but has been proposed to fit within the framework of this article. This procedure leads to common configurations contained within the blocks.

The distinct scores can be defined as the m_{kd} principal components of the projection of $\boldsymbol{Z}_k$ onto the orthogonal complement of $\boldsymbol{U}_{kc}$: $(\boldsymbol{I} - \boldsymbol{U}_{kc} \boldsymbol{U}_{kc}^T) \boldsymbol{Z}_k$. Also, the distinct scores lie within blocks.

Choosing $m_k \geq m_{kc} + m_{kd}$ was based on the above way to extract the distinct part. This need not be enforced: other methods exist for extracting distinct parts and which do not need the residual of $\boldsymbol{Z}_k$, for instance the method CMDS described in Section 2.3.3. In this alternative, it would be sufficient with $m_k \geq m_{kc}$.

While the above allows for separate m_{kc}, the objective of the method is to identify a common space, and therefore, a departure from m_c should be justified, for instance, if dimensions greater than m_{kc} would be considered too noisy for some block K.

Relation to the three axes: This method is MDS first, focuses on correlation, and scores lie within the blocks' spaces.

2.3.2 | Method 2—DISTATIS

The second group of approaches applies a weighted sum in the distance domain. This approach emphasizes explained variance and results in a decomposition within blocks.

There are several ways of selecting the coefficients of the weighted sum: using equal weights amounts to the Sum-PCA approach mentioned in Kiers,[13] while in DISTATIS[8] the choice is based on emphasizing "similar" structures.

DISTATIS normalizes block Gram matrices with respect to their largest eigenvalues λ_{k+}. The RV coefficients[14] between these matrices are calculated and put into a new matrix, which, under fairly general circumstances, has only positive entries and thus admits a first eigenvector (e) with positive entries. The weights of the convex sum are then $\boldsymbol{\omega} = e/\|e\|_1$ and leads to:

$$G = \sum_k \frac{\omega_k}{\lambda_{k+}} G_k. \tag{7}$$

MDS is applied to this matrix which defines both the consensus V and eigenvalues Λ, see also Equation (4). In analogy with $Z = V\Lambda^{1/2} = G \cdot (V\Lambda^{-1/2})$, DISTATIS defines the blocks' common scores Z_{kc} based on an operator T (where we have added the factor γ_k in order for reconstructed Gram matrices to be of equal norm as the originals):

$$Z_{kc} = T(G_k) = \gamma_k \cdot G_k \cdot \left(V\Lambda^{-1/2}\right), \tag{8}$$

which means that each block is handled by the same decomposition as the average. As the transformation amounts to a linear combination of the columns of G_k, the common scores lie within the space spanned by individual blocks. The value γ_k scales the reconstructed Gram matrix to the same size as the original Gram matrix: $\left\|Z_{kc}Z_{kc}^T\right\|^2 = \|G_k\|^2$.

An interesting observation about methods averaging in distances before applying MDS is made in Kiers[13]:

$$\underset{\{V,\Lambda\}}{\mathrm{argmin}}\left\|G - V\Lambda V^T\right\|^2 = \underset{\{V,\Lambda\}}{\mathrm{argmin}}\sum_k \omega_k \left\|\frac{G_k}{\lambda_{k+}} - V\Lambda V^T\right\|^2 \tag{9}$$

where we see that the approximation is also in sum the best approximation to the individual (scaled) Gram matrices.

In terms of relative weights between eigenvalues (4), this method is parametrized by $\alpha = 1$, that is, the method is focused on "explaining variance."

With regards to extracting distinct scores, no "total scores" is available as for the MDS first group. The option proposed here is to apply the constrained MDS approach described in the next section.

Relation to the three axes: This method is based on averaging of distances, explained variance, and scores lie within the block spaces.

2.3.3 | Constrained MDS

In this case, there are no "total scores" Z_k from which to deduce the distinct scores. Instead, a constrained version of MDS (CMDS) is used to extract supplementary information from the individual distance matrices. A key step in MDS is an eigen analysis of the distance matrix, and in Rao,[15] a constrained version is defined allowing to extract eigenvectors which are orthogonal to a set of vectors, see Appendix A.1 for details of the method.

$$\boldsymbol{Z}_{kd} = CMDS(\boldsymbol{D}_k, m_d, \boldsymbol{Z}_{kc}) \tag{10}$$

were m_d is the dimension of the distinct configurations and $\boldsymbol{Z}_{kd} \perp \boldsymbol{Z}_{kc}$. In general, the distinct scores lie in a space which is not contained in any block—"between"—for reasons discussed in appendix A1.

2.3.4 | Method 3—INDSCAL with CMDS

In this last group of approaches, a "direct" solution for common components is sought for: some model $\{\boldsymbol{Y}_k\}$ approximates the block Gram matrices in the Frobenius norm.

$$\min_{\boldsymbol{Y}_k} \sum_k \|\boldsymbol{G}_k - \boldsymbol{Y}_k\|^2 \tag{11}$$

A well-known example is INDSCAL[1] which suggests a model where in the MDS stage the subspaces are identical but each block may weigh axes differently, and as such explicitly introduces a weighted distance model: $\boldsymbol{D}_{ij} = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{W}_k (\boldsymbol{x}_i - \boldsymbol{x}_j)$. The problem is formulated as follows:

$$\min_{\boldsymbol{V}, \boldsymbol{W}_k} \sum_k \|\boldsymbol{G}_k - \boldsymbol{V}\boldsymbol{W}_k\boldsymbol{V}^T\|^2 \tag{12}$$

where the consensus is $\boldsymbol{V}$ (*not necessarily orthogonal*) with rank m_c and $\boldsymbol{W}_k$ are diagonal matrices with strictly positive entries. The solution is described in chapter 22 of Borg and Groenen.[1] The common configurations are defined as $\boldsymbol{Z}_{kc} = \boldsymbol{V}\boldsymbol{W}_k^{1/2}$ and lie in the consensus subspace $R(\boldsymbol{V})$ and hence between blocks. INDSCAL is the only method discussed in this paper that does not fit into the eigenvector solution framework described in Equation (4).

INDSCAL is only concerned with the consensus, and as for DISTATIS, there is no total configuration to relate to, and the constrained MDS approach may also here be applied to calculate distinct scores.

Relation to the three axes: this method is based on the direct use of distances and explains variance, and the solution is between block spaces.

2.4 | Overview and alternatives

Table 1 summarizes many aspects of the various approaches discussed above, as well as adding two: the problem formulation defining the consensus, the solution for the consensus, the common and distinct configurations. The expressions for the common configuration indicate when they lie "within" the individual blocks or between (amount to linear combinations of the columns of the consensus). Furthermore, the exponent of the eigenvalue matrix for blocks indicates if the method favors correlation or explained variance.

In addition to the mathematical details of the different methods, the table also places these methods along the "axes" defined previously: domain shift, the variance-correlation tradeoff and the within-between requirement.

Many other methods could be inserted into the space spanned by the axis, JIVE[16] and DISCO[17] to mention a couple for the MDS first approach, IDIOSCAL[1] as a slight alternative to INDSCAL which in addition allows for individual rotations; and more.

2.5 | Explained variance

Once a decomposition has been obtained for each of the distance matrices, one could attempt to qualify the results. However, as the separation into common and distinct configurations is not known a priori, there is in general no

TABLE 1 Methods and properties

Preprocessing	Method	Problem	G	Z_{kc}	Z_{kd}	Variance/Correlation	Within/between
MDS first	MDS + GCA	$\sum_k \|Z_k P_k - V\|^2$	$\sum_k V_k V_k^T$	$P(U_{kc} U_{kc}^T Z_k)$	$P(Z_k - U_{kc} U_{kc}^T Z_k)$	Correlation	Within
MDS first	MDS + SCA	$\sum_k \|Z_k - VP_k^T\|^2$	$\sum_k V_k \Lambda_k V_k^T$	$P(VV^T Z_k)$	$P(Z_k - VV^T Z_k)$	Variance	Between
Average in distances	DISTATIS	$\sum_k \omega_k \|G_k/\lambda_k{}_+ - V\Lambda V^T\|^2$	$\sum_k \frac{\omega_k}{\lambda_{k+}} V_k \Lambda_k V_k^T$	$\gamma_k G_k V \Lambda^{-1/2}$	$CMDS(D_k, m_d, Z_{kc})$	Variance	Within
Direct	INDSCAL	$\sum_k \|G_k - VW_k V^T\|^2$	-	$VW_k^{1/2}$	$CMDS(D_k, m_d, Z_{kc})$	Variance	Between
Direct	DIRECT-SCA	$\sum_k \|G_k - VP_k^T\|^2$	$\sum_k V_k \Lambda_k^2 V_k^T$	$\widetilde{V}_k \widetilde{\Lambda}_k^{1/2}$	$CMDS(D_k, m_d, Z_{kc})$	Variance	Within

Note: Two new methods are included: MDS followed by SCA where SCA is used to define the consensus, and Direct-SCA where Gram matrices are directly analyzed using SCA to define the consensus. Here, Z_k is the "total configuration" discussed in MDS first followed by GCA, $\widetilde{V}_k$ and $\widetilde{\Lambda}_k$ are the left singular vectors and values of SVD of P_k. U_{kc} is an orthonormal basis of $R(Z_k P_k)$. The operation $P(Z)$ represents the multiplication by the right singular vectors U of Z: $P(Z) = ZU$ so the columns have full rank. The parameter γ_k for DISTATIS is mentioned in Section 2.3.2 and adjusts norms of configurations to best match Gram matrices.

reference to measure how well these configurations have been identified. The only reference one has is the set of distance matrices, or equivalently the set of Gram matrices.

Although Gram matrices do not in general decompose into common and distinct Gram matrices (in the sense that if $X_k = X_{kc} + X_{kd}$, then $G_k = G_{kc} + G_{kd}$), if the common and distinct scores are combined by concatenating their matrices, a decomposition is possible:

$$Z_k = (Z_{kc}, Z_{kd}) \rightarrow \widehat{G}_k = G_{kc} + G_{kd}. \tag{13}$$

This choice will allow for expressing "explained variance" in terms of the common or distinct Gram matrices separately, see below.

The notion of "explained variance" (EV) is often used to summarize the accuracy of some model and one way of calculating it is:

$$EV = 1 - \frac{\sum_k \left\| G_k - \widehat{G}_k \right\|^2}{\sum_k \|G_k\|^2}. \tag{14}$$

This expression can be decomposed as a convex sum of explained variance per block:

$$EV = \sum_k \frac{\|G_k\|^2}{\sum_n \|G_n\|^2} EV_k = \sum_k w_k EV_k, \ EV_k = 1 - \frac{\left\| G_k - \widehat{G}_k \right\|^2}{\|G_k\|^2}. \tag{15}$$

Using the common scores to estimate the Gram matrices ($\widehat{G}_k = \widehat{G}_{kc}$), the above expression may also be used to quantify their contributions (EV_c, EV_{ck})—and similarly for the distinct parts (EV_d, EV_{dk}).

2.6 | Measuring overlapping dimensions

In this paper, our main emphasis has been on common (shared by all) and distinct, orthogonal to common. In practice, however, it may both happen that the common subspaces share little and that distinct subspace are not orthogonal to all other blocks. To quantify these cases, we propose below a useful index which, in a sense, measures the number of overlapping dimensions. This index is similar to other matrix similarity indices, in particular, the similarity of matrices index (SMI)[18] and Yanai's generalized coefficient of determination (GCD),[19] but avoids normalization for the index to have an easy interpretation.

The index "overlap" between two matrices (X, Y) is based on their orthonormal column bases (U_X, U_Y) and is a function of the m principal angles θ_i between these bases, where m is the minimum of the size of the respective bases, see Hamm and Lee[20]:

$$overlap(X, Y) = \left\| U_X^T U_Y \right\|^2 = \sum_{i=1}^M cos^2(\theta_i) \tag{16}$$

To illustrate the index, assume the bases contain a single component each: u_{A1}, u_{B1}. The index lies in the range $[0, 1]$ indicating anything between zero overlap with no common subspace and full overlap of one with parallel bases. Extending the example to two components in each basis, the first pair of "closest vectors" in respective subspaces defines the first principal angle θ_1. Orthogonalizing the bases with respect to this pair leads to a second pair of vectors and which define the second principal angle θ_2. Here, the index lies in the range $[0, 2]$ between zero overlap (orthogonal subspaces) and a complete intersection between the two planes.

2.7 | Selecting dimensions

When applying principal components analysis, a key question concerns the number of latent variables to include in the analysis. This number may be chosen based on how much of the variance is "explained". Because this value is directly related to eigenvalues of the analysis, Scree plots—where eigenvalues are plotted in decaying size—may be used to the same effect (see Figure 3 below for an example).

In the context of distance matrices, the basic ideas could be used in the same way and applied to the Gram matrices. The interpretations of "number of latent variables" is equivalent to the dimension of the embedding space and how well distances are represented in any given number of dimensions.

While this may provide indications of how well individual blocks may be compressed with a given number of dimensions, this is not enough because these subspaces may not overlap as shown in the pharmacogenomics example in Section 3.4. The key objective is after all to identify subspaces that are common in some sense.

A natural way of providing a scree plot for Gram matrices is to plot EV as a function of the number of components along with that for the individual blocks (EV_k) and the common configurations (EV_c), see Section 3.4 for an example. In this example, monitoring EV as the number of components increases may reveal properties relevant to commonness, as it did in the INDSCAL with CMDS solution which clearly showed that subsequent components essentially belonged to individual blocks. In this example, the measure of overlap was used to qualify the solution post hoc: given a solution to common scores, how well do the subspaces overlap?

Incidentally, it may also be relevant to monitor overlap between distinct subspaces, too. This may be surprising given the objective of the analysis of common subspaces to contain basically all that is common. As shown in the simulated example in Section 3.2, having extracted a common configuration does not preclude distinct configurations with high degrees of overlap between associated subspaces. In this example, this behavior was related to dominant distinct configurations and very small common configurations.

A final remark concerns the shift between common and distinct parts: the more dimensions are used to describe common, the less is left over for distinct subspace. While in some cases the distinction may be clear, it is expected that often subsequent increases in dimension of the common spaces will only to some degree overlap. Setting the cutoff between common and distinct is hence left to the discretion of the analyst.

3 | EXAMPLES

3.1 | Pitfalls and success stories

The selected examples in this section illustrate some of the concepts introduced in Section 2.2 and methods introduced in Section 2.3. In Section 3.2, the first example illustrates a potential pitfall in the variance-correlation tradeoff, namely, the case when there exists a small common configuration and large, orthogonal distinct configuration. It shows how a correlation-based method may prove superior and reveals possible misinterpretations when using a method maximizing explained variance. In Section 3.3, a projective mapping (PM) study is analyzed and also it illustrates the role of the consensus. It is also an example of noisy data with numerous blocks and where the distinct parts are not primarily the focus. Finally, in Section 3.4, a study concerning the effect of drugs on cancers is analyzed.[5] The different blocks represent different kinds of information: gene expression, copy number aberration, proteomics, and drug response; and both common and distinct are of interest. The example shows another pitfall where GCA, which handled the first example elegantly, now suggests common subspaces with little overlap between them in spite of the method optimizing for correlation. It also shows the application of INDSCAL to a problem and as such contrasts two methods along the within-between requirement.

3.2 | Simulated example—When correlation matters

We have created a scenario that is simple to grasp and nevertheless illustrates some of the issues with the variance-correlation tradeoff. Two assessors agree on all samples along one dimension z_c and disagree for *some* samples (50%) along a second direction. We construct this as:
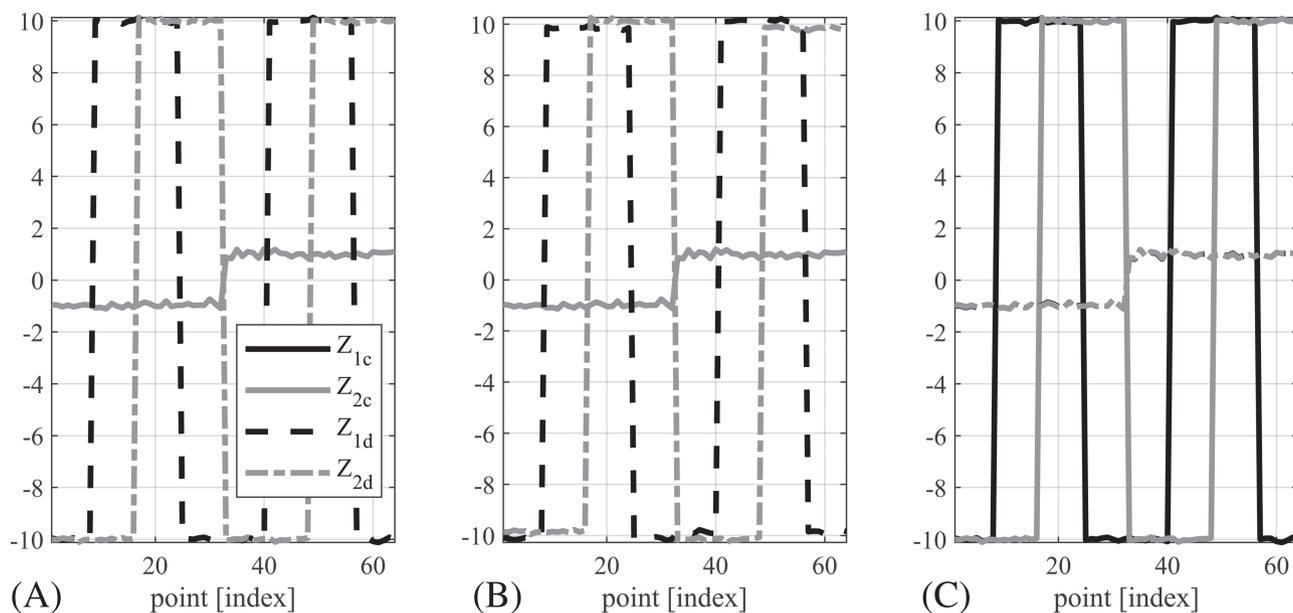
FIGURE 3 (A) True common (lines) and distinct (dashes) components, where both common are identical ($\mathbf{z}_{1c} = \mathbf{z}_{2c}$). (B) Estimated common and distinct components using MDS first and GCA. The estimated are good approximations of true components. (C) Estimated common and distinct components using DISTATIS. Each block common corresponds to true distinct while each block distinct corresponds to the true common: This represents a reversal between common and distinct components. This reversal happens even though the combined estimated components are good approximations of the combined true components

$$\mathbf{Z}_k = (\mathbf{z}_c, \mathbf{z}_{kd}) + \mathbf{E}_k, \ k \in \{1,2\}$$

$$\mathbf{z}_c = \left[-\mathbf{b}^T, -\mathbf{b}^T, \mathbf{b}^T, \mathbf{b}^T \right]^T,$$

$$\mathbf{z}_{1d} = 10 \cdot \left[-\mathbf{a}^T, \mathbf{b}^T, -\mathbf{b}^T, \mathbf{b}^T, -\mathbf{a}^T \right]^T, \ \mathbf{z}_{2d} = 10 \cdot \left[-\mathbf{b}^T, \mathbf{b}^T, -\mathbf{b}^T, \mathbf{b}^T \right]^T$$

where a is a vector of N ones, $\mathbf{b}$ is a vector of $2N$ ones and $\mathbf{E}_k$ is an i.i.d. normally distributed noise term and all $\mathbf{z}_c, \mathbf{z}_{id}$ are column vectors of length $8N$. With the above definitions, both $\mathbf{z}_{kd}^T \mathbf{z}_c = 0$, i.e. distinct are orthogonal to common. Incidentally, the two distinct $\mathbf{z}_{1d}$ and $\mathbf{z}_{2d}$ are also orthogonal. The important point, however, is that the distinct components are much larger than the common components by a factor 10 between norms and this is expected to affect methods that aim at explaining variance. The data are depicted in Figure 3A.

The MDS first with GCA method presented in Section 2.3.1 successfully identifies both common and distinct components, despite the small relative length of the common components, see Figure 3B.

On the other hand, DISTATIS is displayed in Figure 3C, and this method focuses on explaining variance. It confounds common with distinct and thereby ends up with common components for each block that are orthogonal, a result which runs counter to the idea of "common." Also, the two distinct components are highly correlated, again running counter to the idea of "distinct." The root cause is that by approximating original distance matrices with a single component, it is more efficient to approximate the larger, distinct components.

In summary, this example illustrates the difference between similarity and explained variance and shows a possible pitfall when selecting the latter group of approaches. They may end up with common components which have low correlation between them: the spaces they span are not necessarily "close." This pitfall may be detected using the overlapping dimensions index, see Table 2.

TABLE 2 Overlap for methods

MDS first followed by GCA					DISTATIS with CMDS				
	$\widehat{z}_{1c}$	$\widehat{z}_{2c}$	$\widehat{z}_{1d}$	$\widehat{z}_{2d}$		$\widehat{z}_{1c}$	$\widehat{z}_{2c}$	$\widehat{z}_{1d}$	$\widehat{z}_{2d}$
$\widehat{z}_{1c}$	1	1			$\widehat{z}_{1c}$	1			
$\widehat{z}_{2c}$	1	1			$\widehat{z}_{2c}$		1		
$\widehat{z}_{1d}$			1		$\widehat{z}_{1d}$			1	1
$\widehat{z}_{2d}$				1	$\widehat{z}_{2d}$			1	1

Note: Zeros in the table are omitted for clarity. It appears clearly in the DISTATIS with CMDS method that the two estimated common components are orthogonal while the distinct parts are parallel.

3.3 | Sensory—Olive oil data

Our sensory example concerns an olive oil tasting experiment using projective mapping (PM) with so-called ultra-flash profiling where assessors assign descriptive words to different olive oils. A typical aim of this type of experiment is to understand significant dimensions in describing products, and then using these dimensions to place each product in this map. As PM data are often noisy, it is important to include a fair number of assessors and the analysis is tuned towards the description of group preferences. This is an example where the main objective is to study the consensus. The common components are, however, also sometimes of interest in order to study variability among assessors and therefore the validity of the consensus.[3]

PM data are essentially two-dimensional at an individual level. Still, it is technically possible to extract more components when the assessors' data are aggregated. This phenomenon was discussed thoroughly in Næs et al.[21] Typically, there may be one dominating sensory phenomenon which is perceived more or less similarly for all assessors, while the second may be related to different sensory dimensions for the different assessors. This means that going beyond one common component may induce errors or fail to represent qualities for at least some of the different assessors. The consensus for all components will in any case be dominated by the assessors with a similar distance matrix, a kind of majority vote.

The olive oil data set contains 11 products, each judged by 10 assessors. The coordinates for each assessor's projective map X_k are used to calculate squared distances, D_k, and then the consensus is extracted using DISTATIS.
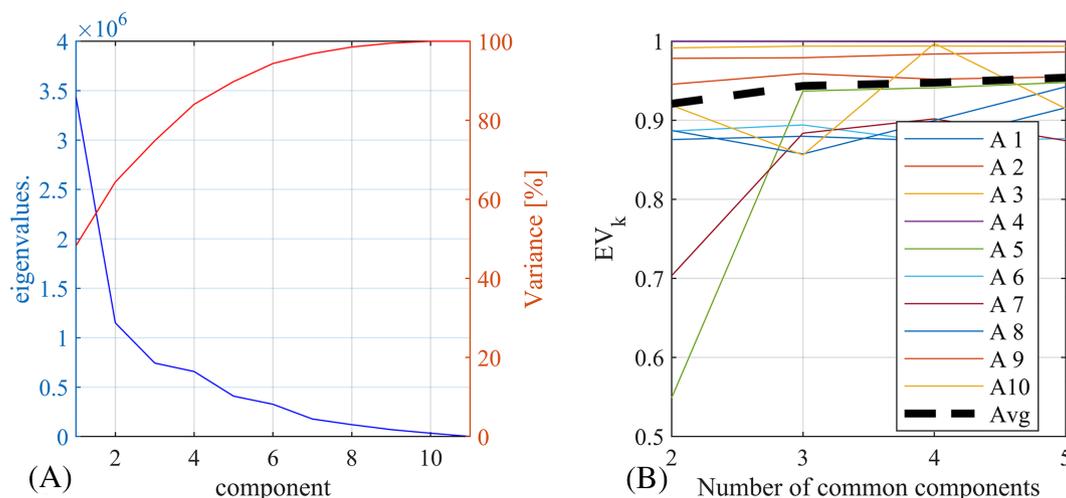

FIGURE 4 Olive oil data consensus by DISTATIS for a multiple of runs while excluding two of 11 assessors in order to visualize variability in estimates. The labels "PC_1" and "PC_2" correspond to the first and second dimensions of the consensus after aligning all simulations using generalized Procrustes analysis, and the percentages are the average explained variance of the original projective maps by projections onto the consensus dimensions

While this provides a single analysis, here, we wished to illustrate the sensitivity to noise. A series of analyses were applied to subsets of 8 assessors, leading to a set of $R = 90$ consensuses: $\{V_r\}$, where r indexes the selection of assessors. This set was subject to generalized Procrustes analysis[22] producing aligned consensuses: $\{V_r'\}$ and which are depicted in gray in Figure 4. An ellipse is constructed for each product so that, under hypothesis of a multivariate normal distribution, 90% of the subsets of assessors fit inside.

Finally, the average consensus: $V_{avg} = \frac{1}{R}\sum_r V_r'$, is used to create a biplot with the most frequent words used in the ultra-flash profiling. A vector f_w is constructed based on frequencies of the word for each product. This vector is projected into the space spanned by the average consensus: $X_w = \left(V_{avg}^T V_{avg}\right)^{-1} V_{avg}^T f_w$, and are collectively scaled so that their lengths are comparable to the consensus. These projections are added to the biplot in Figure 4. The final length of a word vector in the biplot depends hence on both sum frequency and correlations with the average consensus' dimensions.

Considering the word vector projections X_w, the axis between "mild" and "pungent" is most significant and aligns well with the first dimension of the average consensus. There is no natural pair of word vectors that align well with the second dimension, and the projections are generally short. Consider for instance "fruity-berry" which has a higher sum frequency than "mild," but whose word vector projection is much shorter—this is due to low correlations with the average consensus dimensions, so while the word is often used, assessors tend to disagree on the products associated with them. Considering the second consensus dimension, it appears nevertheless to be associated with positive words, "green," "fruity-berry," in the positive direction, and more negative words: "machine oil," "pungent," in the negative direction (Figure 5).

3.4 | Pharmacogenomics

In this case, we have selected a subset of a larger dataset which relates drug response to various characteristics of a set of tumor cell lines. The data subset we will be considering consists of four blocks: drug response (DR), proteomics (PR), gene expression (GE), and copy number alterations (CNA). The 276 samples are tumor cell lines of various cancers. For details on the data set and selection of samples as well as for missing data imputations, see Aben et al.,[5(sec 2.6)] hereafter called the "iTop" article. Distances are Euclidean except for block CAN, which is based on the Jaccard distance between two Boolean vectors $(d(x,y) = 1 - (\sum_i x_i \text{ and } y_i)/(\sum_i x_i \text{ or } y_i))$. See Table 3 for a summary of the data set.

After calculation of squared distance matrices, $\widetilde{D}_k$, the norms of these matrices are very different. Therefore, these matrices have been normalized by the largest eigenvalue of their respective Gram matrices G_k, $D_k = \widetilde{D}_k / \lambda_k^+$.
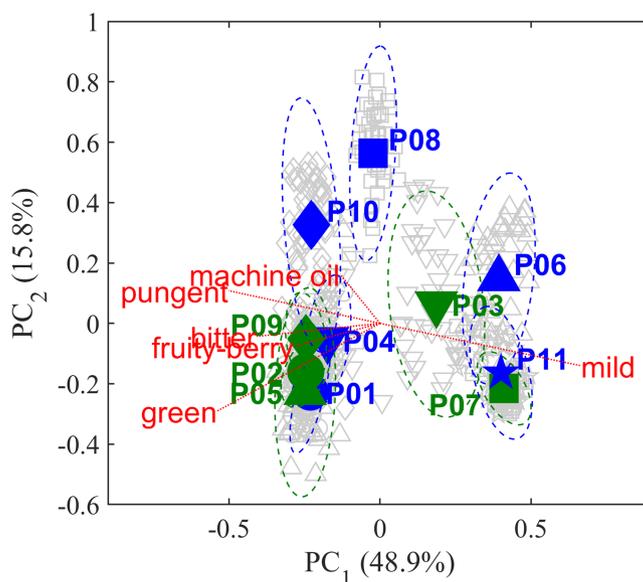
**FIGURE 5** (A) Scree plot of eigenvalues of the concatenated, centered projective maps X_k, and associated part of the total variance $(\sum \|X_k\|^2)$ that is "explained" by the principal components. (B) Explained variance as a function of the number of dimensions used for the common subspaces. Adding a third component increases the EV_k for two assessors. Further increases do not appear on average to improve explained variance

Block	Raw data type	Width	Distance metric
Copy number alterations (CNA)	Boolean: $x_{ij} \in \{0,1\}$	427	$d(i,j) = 1 - \frac{x_i \cap x_j}{x_i \cup x_j}$
Gene expression (GE)	Continuous: $x_{ij} \in \mathbb{R}$	17739	$d(i,j) = \|x_i - x_j\|_2$
Proteomics (PR)	Continuous: $x_{ij} \in \mathbb{R}$	187	$d(i,j) = \|x_i - x_j\|_2$
Drug response (DR)	Continuous: $x_{ij} \in \mathbb{R}$	218	$d(i,j) = \|x_i - x_j\|_2$

Note: For every block, there are 276 samples from tumor cell lines.
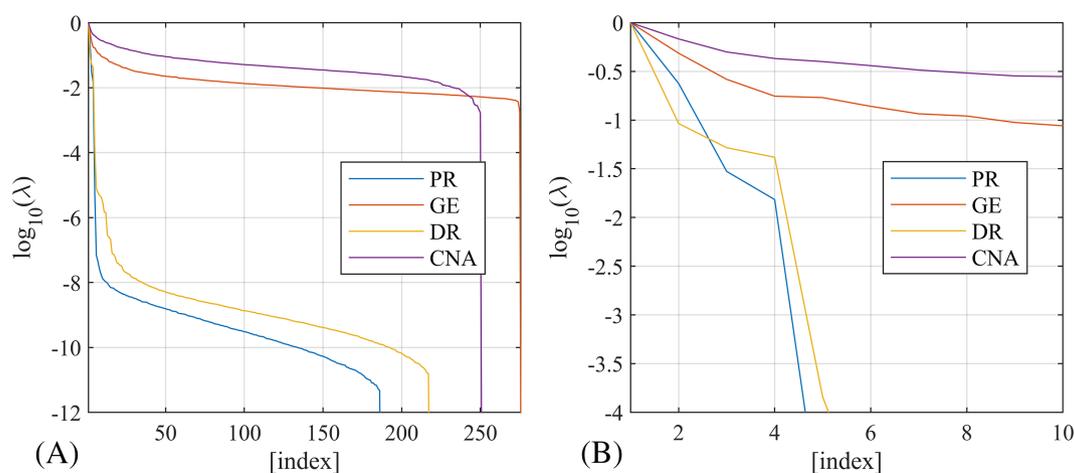
FIGURE 6 Sorted, eigenvalues ($\log_{10}(\lambda)$) of the blocks' Gram matrices. (A) Full set, (B) zoom in on first few components. The blocks gene expression and copy number alteration have slowly decaying eigenvalues, and this is an issue when attempting to synthesize information in a small selection of directions

One of the basic assumptions in all methods considered is that the data blocks can be compressed into a smaller set of components, which we group in the common and distinct parts. However, we see in Figure 6 that the eigenvalue structure is different between blocks, with GE and CNA both having distributed eigenvalues, while PR and DR can be significantly compressed with only a few components.

For this dataset, we have focused on two methods: INDSCAL with CMDS; and MDS first followed by GCA. The first method was selected due to a different model of distances, and the latter to illustrate a pitfall in the GCA method. We will see that in the former, components appear to be selected in part per block while in the latter there is little "in common" between the common blocks $\boldsymbol{Z}_{kc}$.

Explained variance, dimensions, and common parts

In Figure 7, we have plotted the explained variance as a function of the number of components. For simplicity, this number is equal for both common and distinct subspaces. Also, to ease comparison, the number of components in the MDS stage is fixed at 20 (MDS first followed by GCA). Regarding the INDSCAL with CMDS method, we observe jumps in single blocks for each added component in common. In the MDS first followed by GCA method, the common subspaces pick up first on the GE and CNA blocks (with distributed eigenvalues), and significant increases only occurs for the PR block after 6 components have been identified.

To continue, for the methods, five components were selected for both the common and distinct subspaces. This choice is loosely motivated by the number of significant eigenvalues for the PR and DR blocks. Also, while this could have been chosen differently for each block, for simplicity the selection is equal for all blocks. Table 4 shows how well the two methods approximate the Gram matrices expressed through EV. As the MDS first approaches are limited by the MDS stage, the EV of this stage is given at the end of Table 4. While the difference in total explained variance between them is not large, how the explained variance is distributed across blocks is qualitatively different: there is a
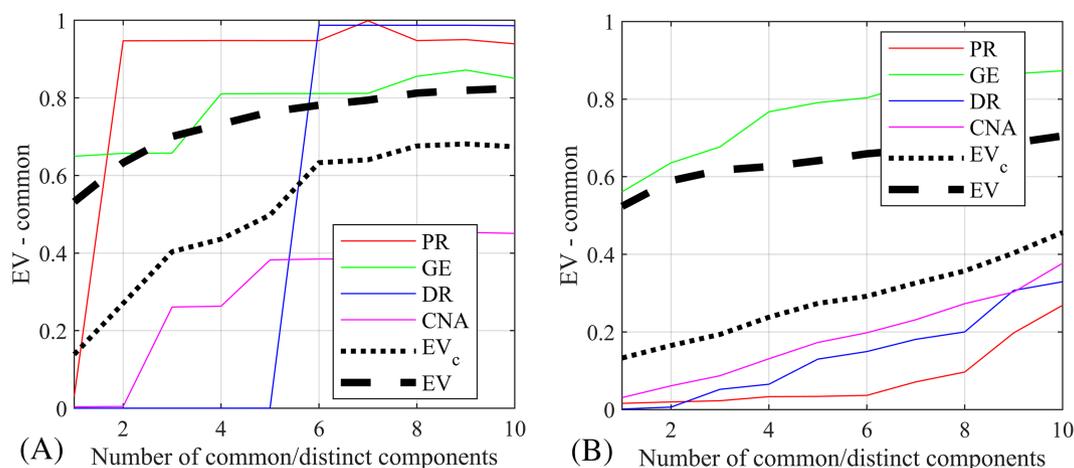
FIGURE 7 Explained variance of common configurations for the different blocks as well as the total contribution from common (EV_c—dotted) and total (EV—dashed). (A) INDSCAL, (B) MDS first followed by GCA where 20 components were chosen for the MDS stage for all cases

TABLE 4 Explained variance in example with pharmacogenomics data

INDSCAL with CMDS	Proteomics	Gene expression	Drug response	CNA
$EV_c = 50\%$	95%	81%	0%	38%
$EV_d = 27\%$	4%	10%	97%	21%
Weighted total	$EV = 76\%$			
MDS first followed by GCA	Proteomics	Gene expression	Drug response	CNA
$EV_c = 41\%$	7%	85%	62%	28%
$EV_d = 26\%$	56%	6%	5%	32%
Weighted total	$EV = 67\%$			
Norms $\|G_k\|^2$	1.1	1.5	1	3.9
Weights	14%	20%	14%	52%
MDS (10 components)	Proteomics	Gene expression	Drug response	CNA
$EV = 81\%$	100%	94%	100%	66%

Note: This table explained variance for each method, and a decomposition along blocks. The largest contribution (see weights at the bottom—these are w_k in 15) comes from the explained variance of the CNA block due to the large spread of eigenvalues of this block. This same cause limits the possible EV with a limited number of common and distinct components. The bottom part of the table gives the explained variance of the MDS stage alone, which limits the values for the MDS followed by GCA method.

strong complementarity in how PR and DR data are incorporated into the common configurations. While common *EV* for GE and CNA are similar for the two methods, INDSCAL favors PR and MDS followed by GCA favors DR.

Common versus similar

A part of the story that Table 4 does *not* tell is how similar the different common components are. To address this question, the overlaps between the blocks' common subspaces was calculated for each of the methods. For the common subspaces of the INDSCAL with CMDS method, the overlap was basically complete between all blocks, as expected because all are in the consensus subspace. However, as can be seen in Table 5 the same is not true for the MDS first followed by GCA method where there was only significant overlap between PR and GE. This difference is a consequence of defining the common subspaces "between" blocks as opposed to "within" each of the blocks. So, while GCA maximizes correlation in its search for a consensus V, the individual common Z_{kc} end up almost orthogonal for several blocks.

TABLE 5 Overlap in pharmacogenomics example

MDS first followed by GCA	Proteomics	Gene expression	Drug response	CNA
Proteomics	5	2.8	0.4	0.4
Gene expression		5	0.3	0.3
Drug response			5	0.2
CNA				5

Note: Due to symmetry, only the upper-triangular part is filled in. The table shows that there is little overlap between blocks, except for between Proteomics and Gene expression. Recall that 5 common components are extracted for each block, hence the identical number 5 on the diagonal.

To conclude that INDSCAL produces "more common" components in terms of correlation is hardly justified, too, because as mentioned before, successive components are attributed entirely to individual blocks, see Figure 6.

Local common

The focus in the iTop article was on establishing links between data sets using a concept of "partial correlation". A link between any two blocks was strong if significant correlation remained after enforcing conditional independence w.r.t. some other blocks. The aim with such analyses is to come closer to describing causal pathways and the analysis was beautifully synthesized in a concise graph, see Figure 7 in the iTop article. For our selection of 4 blocks, a strong link was found between PR and GE; both were linked to DR; and only GE was linked to CNA. The article further provided a biological interpretation of the importance and foundation of links between PR, GE, and DR.

We saw that when using the INDSCAL with CMDS method, the common subspace in effect disregards DR while representing a fair amount of all the other blocks. Even restricting the selection of blocks to PR, GE, and DR only captures a small part of DR using the same parameters as above. While the iTop article showed that there are links between these three data sets, the axes that connect them are not the same as those identified by extracting common components with this method. Also, while the overlap in this method is high—the method finds common "between" blocks and so are shared, which promotes overlap, see Table 1—the way *EV* jumps for a single block with the inclusion of new components suggests that in effect each new component is mainly associated with one and not several blocks. For example, at six common components, the *EV* for DR increases abruptly while those of the other blocks remain essentially constant—see Figure 7. This point is not revealed by the overlap between *blocks*, which was high for this method. While one could suspect that the focus of this method on explaining variance, see Table 1, were the culprit, we will see that by using the MDS first followed by SCA, also focusing on *EV*, there are large components explaining a significant proportion of PR, GE and DR. A remaining hypothesis explaining the above is the specific distance model underpinning the INDSCAL method and which may place too strong restrictions.

Optimizing EV is not the culprit

It is interesting to note that MDS first followed by SCA only needed two components to create significant associations, see Table 6. Furthermore, the overlaps between all blocks were full, which follows from the method. A similar question is how much the consensus overlaps with each block, and this is given at the bottom of Table 6; this emphasizes how similar the consensus is to individual blocks. As a method that emphasizes variance, see Table 1, this result suggests significant components for at least three of the blocks.

Different stories

Tables 4 and 6 appear to tell different stories about the underlying data: the INDSCAL with CMDS method tends to suggest that there is no strong subspace linking GE and PR with DR, while both MDS first with GCA and SCA both tell a different story. Furthermore, in the latter, with only two components, the DR is the block that is explained the most.

These stories are further complicated by the fact that MDS first with GCA has low overlap between GE and DR, suggesting little in common, even though both blocks have large *EV*, which in turn suggests common information. On the other hand, INDSCAL with CMDS has high overlap, suggesting "common," but still effectively associates components with blocks, which is closer to a notion of "distinct." It would appear that MDS first followed by SCA has the most consistent story in the sense that all blocks have high overlap, that the consensus overlaps with the individual blocks *and* three have significant *EV*.

TABLE 6 Overlap in pharmacogenomics data, MDS followed by SCA

MDS first followed by SCA	Proteomics	Gene expression	Drug response	CNA
$EV_c = 22\%$	38%	28%	69%	3%
$EV_d = 29\%$	20%	32%	4%	36%
Weighted total	$EV = 51\%$			
Overlap with consensus (2 components):	1.0	2.0	1.4	1.9

Note: This table explained variance for MDS first followed by SCA with two components for both common and distinct. Overlap between common for all pairs of blocks was two (full overlap).

A final word concerns the overall aim of the analysis of the pharmacogenomics data set. If one wishes to support the assumption that there exist links between three or more blocks, using the concept of common among these blocks posits the extra assumption that such a link exists in a common subspace. If data were such that blocks A, B, C only taken pairwise share subspaces, then it may be so that they collectively share nothing: no vector v exists such that it is not orthogonal to at least one in the triplet. Such a situation would be better addressed through the notion of local common, where common subspaces are searched among the set of blocks. As mentioned earlier, this approach is outside the scope of the present article.

4 | DISCUSSION

In the discussion of various methods for analyzing distance matrices for common and distinct configurations, three "axes" were identified and placed under the headings of "domain shift," "variance-correlation tradeoff," and the "within-between requirement." Also, some tools were suggested in Section 2. In the following, the different examples will be evaluated in terms of these concepts.

4.1 | The notion of common

As noted previously, the consensus is central to the notion of "common" as defined here. We have also seen that by requiring common subspaces to lie within the spaces of the individual blocks, each block's common subspace may end up quite dissimilar.

An intuitive definition of "common" is one that places emphasis on correlation in the variance-correlation tradeoff: "common" is what behaves similarly. This showed benefits in the simulated example in Section 3.2 where a small, correlated part was identified even though each block contained a much larger distinct part. In the same example, a method skewed towards variance failed to provide a simple story for the example, by confusing "common" with "distinct."

However, the same method which was so powerful in Section 3.2 did not show the same consistency across blocks in the pharmacogenomics example of Section 3.4 where the common subspaces were virtually orthogonal between several of the blocks. The application of INDSCAL to this example explains about as much as the GCA based approach and it does so in a well-defined subspace (defined by the consensus). However, here too there is a catch: each added component only improved the approximation of one block at a time, meaning that the components can hardly be considered common.

The more information of respective blocks is contained in the common configuration, at least when using few components, the more compact the representation. This is generally a good thing and represents the main motivation for PCA. As SCA can be considered a generalization of PCA, using it instead of GCA should lead to a more compact representation, albeit attention must be paid to possible asymmetries in scale least a few blocks dominate. In effect, in the pharmacogenomics example, this approach seemed to provide the simplest story: with two components, significant parts were explained in all blocks and the consensus had a high degree of overlap with the individual blocks.

In the sensory example of Section 3.3 an important consideration is noise reduction as it is well known that such experiments generate noisy data. The focus for these cases is often on methods that "explain" the largest part of the variance—equivalently minimize residuals. The idea is that the values for each sample are stabilized by noisy contributions from each block.

4.2 | Within or between?

As discussed in Section 2.2.3, there are two broad approaches to defining the individual common subspaces: either in a single, common subspace or as a set of distinct subspaces which lie in the range of respective blocks.

The blocks may be replications of an experiment as in the Olive Oil example: a small set of components are assumed to be noisily expressed by a "large" set of assessors, and this noise may explain the departure from a common subspace. Hence in such cases, there seems to be little reason for needing individual common subspaces. Actually, in this example the individual common subspaces were not even considered as these lie within each block and with two components 100% of every block was "explained". In other words, there was no compression—no synthesis. Hence, the focus on the consensus.

When blocks come from qualitatively different sources, as in the pharmacogenomics example, the reason for choosing "common" between or within blocks is less clear-cut. We did however see that the MDS first followed by GCA method, which finds individual common subspaces "most correlated" with each other, extracted subspaces that were mostly dissimilar.

In Section 2.2.3, we discussed the within-between requirement and argued that the "within" approaches lead collectively to a larger number of dimensions spanning the common scores compared with the "between" methods, at least when a single parameter m_c is used. This effect explained in part the simulated example in Section 3.2 where DISTATIS assigns two separate common scores (within) rather than being forced to use a single component (between). This distinction may also be what contributes to the small overlap between common subspaces when using MDS followed by GCA in the pharmacogenomics example in Section 3.4.

4.3 | Domain shift

The main aim of this article has been to discuss the analysis of common and distinct configurations with regards to distance matrices. As mentioned, this rests on the body of work already applied to feature matrices and the most direct manner of exploiting known tools is to first use classical MDS to convert to the feature matrix domain, and then apply known methods from multiblock analysis. This is exemplified by the MDS first followed by GCA approach.

There are other approaches that have been investigated and which attempt in varying degrees to perform part of the analysis directly in the distance domain. Some form of averaging in the distance domain is applied in DISTATIS, and INDSCAL uses an individually weighted distance model.

While details are different between approaches there are striking similarities between them. The way the consensus is defined unifies most approaches we have considered as seen in Equation (4) where the eigenvalues of different blocks are weighted according to method. The convex sum of for instance DISTATIS could have been applied to SCA/GCA formulations. Also, given the limitation to Euclidean distances, all methods resemble a principal component analysis (PCA) because presented methods attempt a decomposition which should approximate the original Gram matrices. For the case were a concatenation of the common and distinct configurations, $\boldsymbol{Z}_k = (\boldsymbol{Z}_{kc}, \boldsymbol{Z}_{kd})$, we aim for a good approximation:

$$\gamma \circ \kappa \left(\boldsymbol{Z}_{kc}\right) + \gamma \circ \kappa \left(\boldsymbol{Z}_{kd}\right) \approx \boldsymbol{G}_k.$$

where κ calculates squared distances from configurations and γ calculates Gram matrices from these. In the same way, the principal components $\boldsymbol{Z}'_k$ of a feature matrix X_k may be viewed as such an approximation:

$$\gamma \circ \kappa\left(\boldsymbol{Z}_k'\right) \approx \boldsymbol{G}_k = \boldsymbol{X}_k \boldsymbol{X}_k^T.$$

The limitation to Euclidean distances excludes a broad set of approaches[1] where the analogy with PCA is less relevant. The non-Euclidean case is addressed indirectly in the pharmacogenomics example. There, both the Gene expression and CNA blocks do not lend themselves to concise compression in the sense of PCA, and the CNA block does not rely on Euclidean distances given that the components are binary vectors.

4.4 | Measuring overlap

Precisely because the similarity across blocks of common and distinct is such an open issue, an index for assessing the connections proved helpful. For instance, this was used in the pharmacogenomics example to show how little of common subspaces were shared across blocks in Table 5, and for the simulated example provided a concise summary of overlap within and between common and distinct subspaces. In the simulated examples in Section 3.2, it likewise showed the good performance of the GCA approach and the correspondingly bad performance of the DISTATIS approach.

4.5 | Selecting approach

As has been shown in this article, the different approaches to common and distinct presented and exemplified lead to different solutions with potentially significant differences. That such differences exist is natural and follows from different problem formulations and the properties of solutions. These differences in solutions posed some challenges in the pharmacogenomics data set. A major issue in practice is therefore to choose the most appropriate method for the actual problem which is not always a simple task. At the present stage of development, it is hard to give any fully general advice and this article has investigated the methodological choices that would guide this choice. Furthermore, it is always possible to use different methods to illuminate the problem from different angles.

5 | CONCLUDING REMARKS

In the field of multiblock analysis, there is a large literature on the extraction of common and distinct components. In this article, we investigated possible extensions of this work to the domain of distance matrices.

We have analyzed various methods for extracting common and distinct configurations from distance matrices. To do so, we have proposed a framework for categorizing methods along three axes and which define key design choices: how the transition from the distance domain to the configuration domain is done; how the methods manage a tradeoff between explaining variance and emphasizing correlation between subspaces; and finally, whether the common space is the same for all blocks or separate for each. Several methods have been analyzed within this framework including DISTATIS and INDSCAL, and these have been summarized in table form.

The issues related to the design choices are illustrated with examples. A simulated example provides a case where the power of methods emphasizing correlation is demonstrated. A sensory example demonstrates a projective mapping case where the original data contains distances, and which emphasizes the role of the consensus. Finally, a pharmacogenomics data set shows a case where the inputs are general matrix data.

We do not suggest a final "best practice" in part because we believe this will depend on the data itself and on the objectives of the analysis. However, we do believe the analyses and framework are relevant for understanding some key properties of methods. The examples were included to illustrate these properties and demonstrated possible pitfalls.

and Paula Varela at Nofima for providing the Olive Oil data set within the project FoodSMaCK—Spectroscopy, Modeling & Consumer Knowledge (The Norwegian Agricultural Food Research Foundation project number 262308/F40). M. Carlehøg and M.E. Pedersen (Nofima) are acknowledged for the olive oil data collection.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1002/cem.3372.

DATA AVAILABILITY STATEMENT
There are three data sets involved: the pharmacogenomics data does not represent new data; the simulated data set has been completely specified; the data on the Olive Oils has not yet been published and is not currently available for sharing.

ORCID
Lars Erik Solberg https://orcid.org/0000-0003-0246-8064
Tormod Næs https://orcid.org/0000-0001-5610-3955

REFERENCES
1. Borg I, Groenen PJ. *Modern Multidimensional Scaling: Theory and Applications*. Springer Science & Business Media; 2005.
2. Jaworska N, Chupetlovska-Anastasova A. A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutor Quant Methods Psychol*. 2009;5(1):1-10. https://doi.org/10.20982/tqmp.05.1.p001
3. Næs T, Varela P, Berget I. *Individual Differences in Sensory and Consumer Science: Experimentation, Analysis and Interpretation*. Woodhead Publishing; 2018.
4. Kriz P, Maly F, Kozel T. Improving indoor localization using bluetooth low energy beacons. *Mob Inf Syst*. 2016;2016:2083094. https://doi.org/10.1155/2016/2083094
5. Aben N, Westerhuis JA, Song Y, et al. ITOP: Inferring the topology of omics data. *Bioinformatics*. 2018;34(17):i988-i996. https://doi.org/10.1093/bioinformatics/bty636
6. Smilde AK, Song Y, Westerhuis JA, Kiers HAL, Aben N, Wessels LFA. Heterofusion: fusing genomics data of different measurement scales. *J Chemometr*. May 2019;2020(2):1-19. https://doi.org/10.1002/cem.3200
7. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *ISME j*. 2011;5(2):169-172. https://doi.org/10.1038/ismej.2010.133
8. Abdi H, O'Toole AJ, Valentin D, Edelman B. DISTATIS: the analysis of multiple distance matrices. *Int Conf Comput Vis Pattern Recognit*. Published online. 2005;42-42. https://doi.org/10.1109/cvpr.2005.445
9. Smilde AK, Måge I, Næs T, et al. Common and distinct components in data fusion. *J Chemometr*. 2017;31(7):1-20. https://doi.org/10.1002/cem.2900
10. Dokmanic I, Parhizkar R, Ranieri J, Vetterli M. Euclidean distance matrices essential theory, algorithms and applications. *Signal Process Mag*. 2015;(November);12-30. https://doi.org/10.1109/MSP.2015.2398954
11. ten Berge JMF, Kiers HAL, Van der Stel V. Simultaneous components analysis. *Stat Appl*. 1992;4(4):377-392.
12. Dahl T, Næs T. A bridge between Tucker-1 and Carroll's generalized canonical analysis. *Comput Stat Data Anal*. 2006;50(11):3086-3098. https://doi.org/10.1016/j.csda.2005.06.016
13. Kiers HAL. Hierarchical relations among three-way methods. *Psychometrika*. 1991;56(3):449-470. https://doi.org/10.1007/BF02294485
14. Robert P, Escoufier Y. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *J R Stat Soc Ser C*. 1976;25(3):257-265. https://www.jstor.org/stable/2347233
15. Rao RC. The use and interpretation of principal component analysis in applied research. *Sankhya—Ser A*. 1964;26(4):329-358.
16. Lock EF, Hoadley KA, Marron JS, Nobel AB. Joint and individual variation explained (JIVE) for an integrated analysis of multiple data types. *Ann Appl Stat*. 2013;7(1):523-542. https://doi.org/10.1214/12-AOAS597.JOINT
17. van Deun K, van Mechelen I, Thorrez L, et al. DISCO-SCA and properly applied GSVD as swinging methods to find common and distinctive processes. *PLoS One*. 2012;7(5):1-13. https://doi.org/10.1371/journal.pone.0037840
18. Indahl UG, Næs T, Liland KH. A similarity index for comparing coupled matrices. *J Chemometr*. 2018;32(10):1-18. https://doi.org/10.1002/cem.3049
19. Yanai H. Unification of various techniques of multivariate analysis by means of generalized coefficient of determination (GCD). *J Behav*. 1974;1(1):45-54.
20. Hamm J, Lee DD. Grassmann discriminant analysis: a unifying view on subspace-based learning. *Proc 25th Int Conf Mach Learn*. 2008;(July);376-383.
21. Næs T, Berget I, Liland KH, Ares G, Varela P. Estimating and interpreting more than two consensus components in projective mapping: INDSCAL vs. multiple factor analysis (MFA). *Food Qual Prefer*. 2017;58:45-60. https://doi.org/10.1016/j.foodqual.2016.11.012
22. Gower J. Generalized procrustes analysis. *Psychometrika*. 1975;40(1):33-51.

APPENDIX A: APPENDICES

A.1 | Orthogonalized principal components

The Gram matrix G is a positive semidefinite matrix. As shown in Rao,[15] it is possible to find eigenvectors and eigenvalues of G which are subject to orthogonality constraints:

$$V^T V = I \wedge V^T C = 0,$$

$$\max_V \sum_i v_i^T G v_i = \sum_i \lambda_i,$$

where v_i is the ith column of V. These eigenvectors and values are those of the following matrix:

$$\left(I - C(C^T C)^{-1} C\right) G$$

and may be calculated using a symmetric formulation. The resulting eigenvectors are transformed back to those of the above expression:

$$U \Lambda U^T = G^{1/2} \left[I - C(C^T C)^{-1} C^T\right] G^{1/2}$$

$$V = \Gamma G^{-1/2} U$$

where Γ is diagonal with values $\gamma_{ii} = 1 / \sqrt{u_i^T G^{-1} u_i}$. With respect to CMDS where G is the outer product of the configuration ($G = XX^T$), V forms an orthonormal sequence which maximizes the following sum under constraints of rank (q) and orthogonality with C:

$$\sum_{i=1}^{q} v_i^T X X^T v_i = \sum_{j=1}^{N} \sum_{i=1}^{q} \left(v_i^T x_j\right)^2 = \sum_{j=1}^{N} \left\|V^T x_j\right\|^2 = \sum_{j=1}^{N} \left\|V V^T x_j\right\|^2 = \left\|V V^T X\right\|^2.$$

This means that V maximizes the norm of the projection of X. Let $Y = V V^T X$, one consequence of the above is that Y is not necessarily contained in the space spanned by X. As an example, consider the $R(X)$ a slanted plane in 3D space and $R(C)$ the z-axis. Then the principal components span the x-y plane, and Y is the projection of X onto this plane.

The above can be used for a constrained MDS approach (CMDS):

$$Z_d = CMDS(D, m_d, Z_c) = V \Lambda^{1/2},$$

where $G = \gamma(D)$, m_d is the dimension of the distinct subspace, Λ is the diagonal matrix with the above eigenvalues and Z_c are the common scores and used for C.

A.2 | Direct SCA

A "direct" version of SCA can be formulated as

$$\min_{\boldsymbol{V},\boldsymbol{P}_k}\sum_k\left|\left|\boldsymbol{G}_k-\boldsymbol{V}\boldsymbol{P}_k^T\right|\right|^2.$$

The solution to the consensus $\boldsymbol{V}$ is here given by the eigenvalue problem:

$$\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T=\sum_k\boldsymbol{G}_k\boldsymbol{G}_k^T=\sum_k\left(\boldsymbol{V}_k\boldsymbol{\Lambda}_k\boldsymbol{V}_k^T\right)\left(\boldsymbol{V}_k\boldsymbol{\Lambda}_k\boldsymbol{V}_k^T\right)^T=\sum_k\boldsymbol{V}_k\boldsymbol{\Lambda}_k^2\boldsymbol{V}_k^T.$$

We observe that the value of α is in this case 2.