

STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling

Hervé Abdi,^{1*} Lynne J. Williams,² Domininique Valentin³
and Mohammed Bennani-Dosse⁴

STATIS is an extension of principal component analysis (PCA) tailored to handle multiple data tables that measure sets of variables collected on the same observations, or, alternatively, as in a variant called dual-STATIS, multiple data tables where the same variables are measured on different sets of observations. STATIS proceeds in two steps: *First* it analyzes the between data table similarity structure and derives from this analysis an optimal set of weights that are used to compute a linear combination of the data tables called the *compromise* that best represents the information common to the different data tables; *Second*, the PCA of this compromise gives an optimal map of the observations. Each of the data tables also provides a map of the observations that is in the same space as the optimum compromise map. In this article, we present STATIS, explain the criteria that it optimizes, review the recent inferential extensions to STATIS and illustrate it with a detailed example.

We also review, and present in a common framework, the main developments of STATIS such as (1) X-STATIS or partial triadic analysis (PTA) which is used when all data tables collect the same variables measured on the same observations (e.g., at different times or locations), (2) COVSTATIS, which handles multiple covariance matrices collected on the same observations, (3) DISTATIS, which handles multiple distance matrices collected on the same observations and generalizes metric multidimensional scaling to three way distance matrices, (4) Canonical-STATIS (CANOSTATIS), which generalizes discriminant analysis and combines it with DISTATIS to analyze multitable discriminant analysis problems, (5) power-STATIS, which uses alternative criteria to find STATIS optimal weights, (6) ANISOSTATIS, which extends STATIS to give specific weights to each variable rather than to each whole table, (7) $(K + 1)$ -STATIS (or *external-STATIS*), which extends STATIS (and PLS-methods and Tucker inter battery analysis) to the analysis of the relationships of several data sets and one external data set, and (8) double-STATIS (or DO-ACT), which generalizes $(K + 1)$ -STATIS and analyzes two sets of data tables, and STATIS-4, which generalizes double-STATIS to more than two sets of data.

*Correspondence to: herve@utdallas.edu

¹School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, TX, USA

²Rotman Research Institute, Baycrest, Toronto, Ontario, Canada

³Université de Bourgogne, Dijon, France

⁴Université de Rennes 2, Rennes, France

These recent developments are illustrated by small examples. © 2012 Wiley Periodicals, Inc.

How to cite this article:

WIREs Comput Stat 2012, 4:124–167. doi: 10.1002/wics.198

Keywords: STATIS; STATIS-4; DISTATIS; COVSTATIS; power-STATIS; anisotropic STATIS; ANISOSTATIS; double-STATIS; dual-STATIS; canonical-STATIS; SUM-PCA; R_V -PCA; multiple factor analysis; multiblock correspondence analysis; multidimensional scaling; INDSCAL; multiblock barycentric discriminant analysis; co-inertia analysis; STATICO; COSTATIS; barycentric discriminant analysis; generalized singular value decomposition; principal component analysis; consensus PCA; multitable PCA; multiblock PCA; X-STATIS; partial triadic analysis; PTA

INTRODUCTION

STATIS is an acronym which stands for the French expression ‘*Structuration des Tableaux à Trois Indices de la Statistique*’ (which can, approximately, be translated as ‘structuring three way statistical tables’). This technique is also known under the acronym of ‘ACT’ which stands for the French ‘*Analyse Conjointe de Tableaux*’ (joint analysis of tables). Behind these rather obscure acronyms is a generalization of principal component analysis (PCA) whose goal is to analyze several data sets of variables collected on the same set of observations, or—as in its dual version called dual-STATIS—several sets of observations measured on the same set of variables. As such, STATIS is part of the multitable (also called multiblock or consensus analysis^{1–15}) PCA family which comprises related techniques such as multiple factor analysis (MFA), multiblock discriminant correspondence analysis (MUDICA), and SUM-PCA.

STATIS originated from the work of Escouffier^{15,16} and was first described by L’Hermier des Plantes (Ref 17; see also Refs 18–20, for an early introduction see Ref 21). A related approach is known as *Procrustes matching by congruence coefficients* in the English speaking community.²²

The goals of STATIS are (1) to compare and analyze the relationships between the different data sets, (2) to integrate these data sets into an optimum weighted average called a *compromise* (sometimes also called a *consensus*) which is then analyzed via PCA to reveal the common structure between the observations, and finally (3) to project each of the original data sets onto the compromise to analyze communalities and discrepancies.

STATIS is a popular method for analyzing multiple sets of variables measured on the same

observations and it has been recently used in various domains, such as sensory and consumer science research,^{12,13,23–29} chemometry and process monitoring,³⁰ ecology,^{31–33} computer vision,³⁴ hydrology,^{35,36} information science,³⁷ neuroimaging,^{38–42} medicine,³ statistical quality control,^{43,44} and molecular biology.^{10,45–48}

In addition to being used in several domains of applications, STATIS is also a vigorous domain of theoretical developments that are explored later in this article.

When To Use It

STATIS is used when several sets of variables have been measured on the same set of observations. The number and/or nature of the variables used to describe the observations can vary from one set of variables to the other, but the observations should be the same in all the data sets.

For example, the data sets can be measurements taken on the same observations (individuals or objects) at different occasions. In this case, the first data set corresponds to the data collected at time 1, the second one to the data collected at time 2 and so on. The goal of the analysis, then is to evaluate how the positions of the observations change over time.

In another example, the data sets can be measurements taken on the same observations by different subjects or groups of subjects. In this case, the first data set corresponds to the first subject, the second one to the second subject and so on. The goal of the analysis, then, is to evaluate if there is an agreement between the subjects or groups of subjects.

The Main Idea

The general idea behind STATIS is to analyze the structure of the individual data sets (i.e., the

relation between the individual data sets) and to derive from this structure an *optimal* set of weights for computing the best common representation of the observations called the *compromise*, or also sometimes the *consensus*. To compute this compromise, the elements of each table are multiplied by the optimal weight of this table and the compromise is obtained by the addition of these ‘weighted’ K tables (i.e., the compromise is a linear combination of the tables). These weights are chosen so that the compromise provides the best representation (in a least square sense) of the whole set of tables. The PCA of the compromise decomposes the variance of the compromise into a set of new orthogonal variables called principal components (also often called dimensions, axes, factors, or even latent variables) ordered by the amount of variance that each component explains. The coordinates of the observations on the components are called *factor scores* and these can be used to plot maps of the observations in which the observations are represented as points such that the distances in the map best reflect the similarities between the observations. The position of the observations ‘as seen by’ each data set can be also represented as points in the compromise. As the components are obtained by combining the original variables, each variable contributes a certain amount to each component. This quantity, called the *loading* of a variable on a component, reflects the

importance of that variable for this component and can also be used to plot maps of the variables that reflect their association. Finally, as a byproduct of the computation of the optimal weights, the data sets can be also represented as points in a multidimensional space. A sketch of the technique is provided in Figure 1.

NOTATIONS AND PRELIMINARIES

Matrices are denoted by boldface uppercase letters (e.g., \mathbf{X}), vectors by boldface lowercase letters (e.g., \mathbf{q}), elements of vectors and matrices are denoted by italic lower case letters with appropriate indices if needed (e.g., x_{ij} is an element of \mathbf{X}). Blocks of variables (i.e., tables) are considered as submatrices of larger matrices and are represented in brackets separated by vertical bars (e.g., a matrix \mathbf{X} made of two submatrices $\mathbf{X}_{[1]}$ and $\mathbf{X}_{[2]}$ is written $\mathbf{X} = [\mathbf{X}_{[1]}|\mathbf{X}_{[2]}]$). The identity matrix is denoted by \mathbf{I} , a vector of ones is denoted by $\mathbf{1}$ (indices may be use to specify the dimensions if the context is ambiguous). The transpose of a matrix is denoted T . The inverse of a matrix is denoted $^{-1}$. When applied to a square matrix, the **diag** operator takes the diagonal elements of this matrix and stores them into a column vector; when applied to a vector, the **diag** operator stores the elements of this vector on the diagonal elements of a diagonal matrix. The **vec** operator transforms a matrix into a vector by

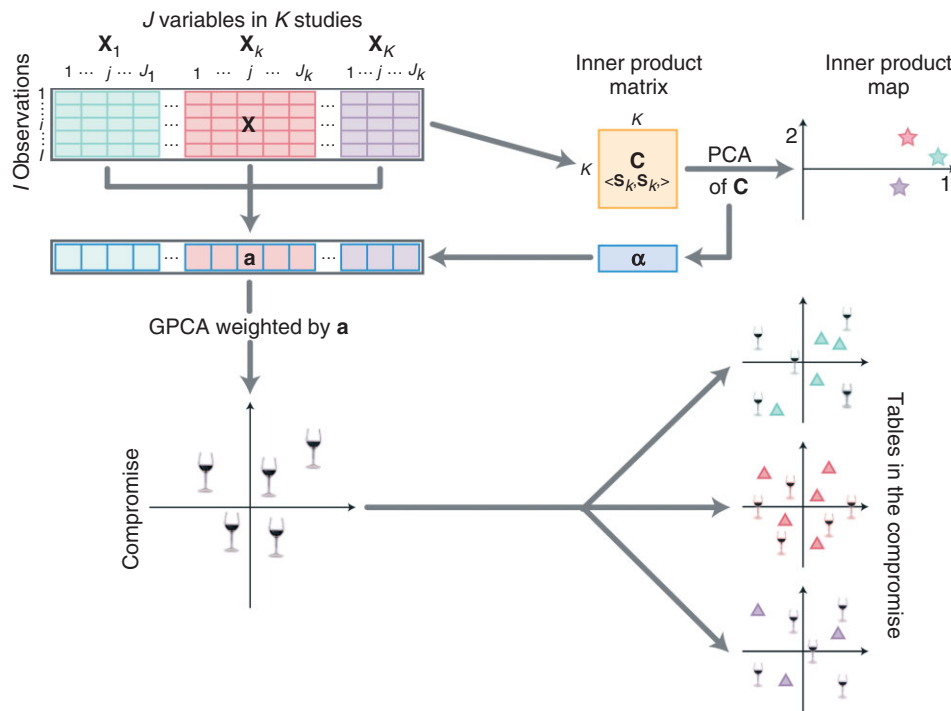


FIGURE 1 | The different steps of STATIS.

stacking the elements of a matrix into a column vector. The standard product between matrices is implicitly denoted by simple juxtaposition or by \times when it needs to be explicitly stated (e.g., $\mathbf{XY} = \mathbf{X} \times \mathbf{Y}$ is the product of matrices \mathbf{X} and \mathbf{Y}). The Hadamard or element-wise product is denoted by \circ (e.g., $\mathbf{X} \circ \mathbf{Y}$).

The raw data consist of K data sets collected on the same observations. Each data set is also called a table, a subtable, or sometimes also a block or a study (in this article, we prefer the term table or occasionally block). The data for each table are stored in an $I \times J_{[k]}$ rectangular data matrix denoted $\mathbf{Y}_{[k]}$, where I is the number of observations and $J_{[k]}$ the number of variables collected on the observations for the k th table. The total number of variables is denoted J (i.e., $J = \sum J_{[k]}$). Each data matrix is, in general, preprocessed (e.g., centered, normalized) and the preprocessed data matrices actually used in the analysis are denoted $\mathbf{X}_{[k]}$ (the preprocessing steps are detailed below in Section ‘More on Preprocessing’).

The K data matrices $\mathbf{X}_{[k]}$, each of dimensions I rows by $J_{[k]}$ columns, are concatenated into the complete I by J data matrix denoted \mathbf{X} :

$$\mathbf{X} = [\mathbf{X}_{[1]} | \cdots | \mathbf{X}_{[k]} | \cdots | \mathbf{X}_{[K]}]. \quad (1)$$

A mass, denoted m_i , is assigned to each observation. These masses are collected in the mass vector, denoted \mathbf{m} , and in the diagonal elements of the mass matrix denoted \mathbf{M} , which is obtained as

$$\mathbf{M} = \text{diag}\{\mathbf{m}\}. \quad (2)$$

Masses are positive or null elements whose sum equals one. Often, equal masses are chosen with $m_i = \frac{1}{I}$.

To each matrix $\mathbf{X}_{[k]}$, we associate its *cross-product* matrix defined as

$$\mathbf{S}_{[k]} = \mathbf{X}_{[k]} \mathbf{X}_{[k]}^T. \quad (3)$$

A cross-product matrix of a table expresses the pattern of relationships between the observations as seen by this table. Note that because of the block structure of \mathbf{X} , the cross product of \mathbf{X} can be expressed as

$$\begin{aligned} \mathbf{X}\mathbf{X}^T &= [\mathbf{X}_{[1]} | \cdots | \mathbf{X}_{[k]} | \cdots | \mathbf{X}_{[K]}] \\ &\quad \times [\mathbf{X}_{[1]} | \cdots | \mathbf{X}_{[k]} | \cdots | \mathbf{X}_{[K]}]^T \\ &= \sum_k \mathbf{X}_{[k]} \mathbf{X}_{[k]}^T \\ &= \sum_k \mathbf{S}_{[k]}. \end{aligned} \quad (4)$$

Singular Value Decomposition and Generalized Singular Value Decomposition

STATIS is part of the PCA family and therefore the main analytical tool for STATIS is the singular value decomposition (SVD) and the generalized singular value decomposition (GSVD) of a matrix (see for tutorials, e.g., Refs 49–54). We briefly describe these two methods below.

Singular Value Decomposition

Recall that the SVD of a given $I \times J$ matrix \mathbf{X} decomposes it into three matrices as:

$$\mathbf{X} = \mathbf{U}\mathbf{\Gamma}\mathbf{V}^T \quad \text{with} \quad \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}, \quad (5)$$

where \mathbf{U} is the I by L matrix of the normalized left singular vectors (with L being the rank of \mathbf{X}), \mathbf{V} the J by L matrix of the normalized right singular vectors, and $\mathbf{\Gamma}$ the L by L diagonal matrix of the L singular values; also γ_ℓ , \mathbf{u}_ℓ , and \mathbf{v}_ℓ are, respectively, the ℓ th singular value, left, and right singular vectors. Matrices \mathbf{U} and \mathbf{V} are orthonormal matrices (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$). The SVD is closely related to and generalizes the well-known *eigendecomposition* as \mathbf{U} is also the matrix of the normalized eigenvectors of $\mathbf{X}\mathbf{X}^T$, \mathbf{V} is the matrix of the normalized eigenvectors of $\mathbf{X}^T\mathbf{X}$, and the singular values are the square root of the eigenvalues of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$ (these two matrices have the same eigenvalues).

Key property: the SVD provides the best reconstitution (in a least squares sense) of the original matrix by a matrix with a lower rank.

Generalized Singular Value Decomposition

The generalized singular value decomposition (GSVD) generalizes the SVD of a matrix by incorporating two additional positive definite matrices (recall that a positive definite matrix is a square symmetric matrix whose eigenvalues are all positive) that represent ‘constraints’ to be incorporated in the decomposition (formally, these matrices are constraints on the orthogonality of the singular vectors, see Refs 50,53 for more details). Specifically let \mathbf{M} denote an I by I positive definite matrix representing the ‘constraints’ imposed on the rows of an I by J matrix \mathbf{X} , and \mathbf{A} a J by J positive definite matrix representing the ‘constraints’ imposed on the columns of \mathbf{X} . Matrix \mathbf{M} is almost always a diagonal matrix of the ‘masses’ of the observations (i.e., the rows); whereas matrix \mathbf{A} implements a metric on the variables and is often but not always diagonal. Obviously, when $\mathbf{M} = \mathbf{A} = \mathbf{I}$, the GSVD reduces to the plain SVD. The GSVD of \mathbf{X} , taken into account \mathbf{M} and \mathbf{A} , is expressed as (compare

with Eq (5)):

$$\mathbf{X} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^T \quad \text{with} \quad \mathbf{P}^T\mathbf{M}\mathbf{P} = \mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{I}, \quad (6)$$

where \mathbf{P} is the I by L matrix of the normalized left generalized singular vectors (with L being the rank of \mathbf{X}), \mathbf{Q} the J by L matrix of the normalized generalized right singular vectors, and $\mathbf{\Delta}$ the L by L diagonal matrix of the L generalized singular values. The GSVD implements the whole class of generalized PCA which includes (with a proper choice of matrices \mathbf{M} and \mathbf{A} and preprocessing of \mathbf{X}) techniques such as discriminant analysis, correspondence analysis, canonical variate analysis, etc. With the so called ‘triplet notation’, that is used as a general framework to formalize multivariate techniques, the GSVD of \mathbf{X} under the constraints imposed by \mathbf{M} and \mathbf{A} is equivalent to the statistical analysis of the triplet $(\mathbf{X}, \mathbf{A}, \mathbf{M})$.^{33,55–59}

Key property: the GSVD provides the best reconstitution (in a least squares sense) of the original matrix by a matrix with a lower rank under the constraints imposed by two positive definite matrices. The generalized singular vectors are orthonormal with respect to their respective matrix of constraints.

THE DIFFERENT STEPS OF STATIS

STATIS comprises two main steps: *First* the similarity structure of the set of tables is analyzed. This step is sometimes called the analysis of the *inter-structure* of the tables.^{21,60,61} This provides a set of optimal weights which are used in the *second* step. This step performs a generalized PCA (i.e., a GSVD) of \mathbf{X} that integrates the weights as constraints on the tables and their variables. This step is sometimes called the analysis of the *intra-structure* of the tables.

Step One: Analyzing the Between-Table Structure

The first step of STATIS analyzes the similarity structure of the set of the K tables. The similarity between two data matrices $\mathbf{X}_{[k]}$ and $\mathbf{X}_{[k']}$ is denoted $c_{k,k'}$. This similarity is evaluated by first transforming the matrices into I by I cross-product matrices denoted $\mathbf{S}_{[k]}$ and $\mathbf{S}_{[k']}$ which are computed as

$$\mathbf{S}_{[k]} = \mathbf{X}_{[k]}\mathbf{X}_{[k]}^T \quad \text{and} \quad \mathbf{S}_{[k']} = \mathbf{X}_{[k']}\mathbf{X}_{[k']}^T. \quad (7)$$

The similarity measure $c_{k,k'}$ (akin to a coefficient of correlation between square matrices^{16,62}) is called the *Hilbert–Schmidt inner product*,⁶³ or simply, here, the inner product (it is also called the *scalar*

product, the vector covariance—denoted COVV—or the *Frobenius* product) between the matrices $\mathbf{S}_{[k]}$ and $\mathbf{S}_{[k']}$. The inner product between matrices $\mathbf{S}_{[k]}$ and $\mathbf{S}_{[k']}$ is denoted $\langle \mathbf{S}_{[k]}, \mathbf{S}_{[k']} \rangle$, and so the coefficient $c_{k,k'}$ is obtained as

$$\begin{aligned} c_{k,k'} &= \langle \mathbf{S}_{[k]}, \mathbf{S}_{[k']} \rangle \\ &= \text{trace} \left\{ \mathbf{X}_{[k]}\mathbf{X}_{[k]}^T \times \mathbf{X}_{[k']}\mathbf{X}_{[k']}^T \right\} \\ &= \text{trace} \left\{ \mathbf{S}_{[k]} \times \mathbf{S}_{[k']} \right\} \\ &= \text{vec} \left\{ \mathbf{S}_{[k]} \right\}^T \times \text{vec} \left\{ \mathbf{S}_{[k']} \right\} \\ &= \sum_i \sum_j s_{i,j,k} s_{i,j,k'}. \end{aligned} \quad (8)$$

Geometrically, the inner product can also be interpreted as a *scalar product* between two positive semidefinite matrices (recall that a square matrix is positive semidefinite if its eigenvalues are all positive or null) $\mathbf{S}_{[k]}$ and $\mathbf{S}_{[k']}$ and is therefore proportional to the *cosine* between the matrices (because the set of positive semidefinite matrices is a vector space). The inner product is also used to define the norm of a matrix (e.g., $\mathbf{S}_{[k]}$) as the square root of the inner product of this matrix with itself. Formally, the norm of a cross product matrix $\mathbf{S}_{[k]}$ is denoted $\|\mathbf{S}_{[k]}\|$ and is defined as:

$$\|\mathbf{S}_{[k]}\|^2 = \langle \mathbf{S}_{[k]}, \mathbf{S}_{[k]} \rangle. \quad (9)$$

Because all matrices $\mathbf{S}_{[k]}$ are positive semidefinite, the inner product between two of these matrices is always equal or larger than zero (see Appendix for a proof). Note that when the matrices $\mathbf{S}_{[k]}$ and $\mathbf{S}_{[k']}$ are normalized such that the sum of squares of their elements is one, the inner product between these two matrices is equal to the cosine between these matrices. This cosine is also known as the R_V coefficient introduced by Escoufier as a measure of similarity between squared symmetric matrices (specifically positive semidefinite matrices,⁶⁴ see also Ref 65) and as a theoretical tool to analyze multivariate techniques (see Refs 66, 67 for reviews).

The inner products between all K matrices are collected into a K by K inner product matrix denoted \mathbf{C} where $c_{k,k'}$ gives the value of the inner product between tables k and k' . Matrix \mathbf{C} can also be computed from the K by I^2 matrix \mathbf{Z} defined as

$$\mathbf{Z} = [\text{vec} \{ \mathbf{S}_{[1]} \} | \cdots | \text{vec} \{ \mathbf{S}_{[k]} \} | \cdots | \text{vec} \{ \mathbf{S}_{[K]} \}]^T. \quad (10)$$

With this matrix Z , we compute C as:

$$C = ZZ^T. \tag{11}$$

This shows that C is a positive semidefinite matrix and that, therefore, its eigenvalues are all positive or null and its eigenvectors are real and orthogonal to each other. Consequently, C can be eigen-decomposed as

$$C = U\Theta U^T \quad \text{with} \quad U^T U = I. \tag{12}$$

So, the eigendecomposition of C provides the PCA of the similarity structure between the tables (stored in X) which can, now, be represented as points in a PCA map by using for coordinates their factor scores that are computed as:

$$G = U\Theta^{\frac{1}{2}}. \tag{13}$$

Note that the eigenvectors of C , could also have been obtained from the SVD of Z as

$$Z = UV^T. \tag{14}$$

This approach, called R_V -PCA, is sometimes used to analyze multiple correlation or covariance matrices. It provides the similarity structure of the tables and information on the values of what pairs of observations contribute to the pattern of similarity measured by $\langle S_{[k]}, S_{[k']} \rangle$.

Optimum Weights from the First Eigenvector of C

In addition to providing a visual representation of the similarity structure, the eigendecomposition of matrix C also provides optimal weights for combining the tables into a compromise.

The eigendecomposition of matrix C always gives a first eigenvector whose elements have the same sign (which is, for convenience, considered positive). This property is a consequence of the Perron–Frobenius theorem that states that positive semi-definite matrices whose elements are all positive always have a first eigenvector with all elements having the same sign (see Ref 68, p. 34ff).

In fact, the value of a table for the first eigenvector reflects its overall similarity to all the other tables (e.g., its ‘communality’). This suggests to use the values of this first eigenvector to weight the tables in order to give more importance to tables that well represent the group and less importance to idiosyncratic tables. It can be shown (see Section ‘What Does Statist Optimize?’ and Appendix), that this

procedure provides an optimal representation of the set of tables. So, if the first eigenvector of C is denoted u_1 , the set of optimal weights for the tables is stored in a K by 1 vector denoted α and computed by rescaling u_1 such that the sum of the elements of α is equal to one:

$$\alpha = u_1 \times (u_1^T \mathbf{1})^{-1}. \tag{15}$$

For convenience, the α weights can be gathered in a J by 1 vector denoted a where each variable is assigned the α weight of the matrix to which it belongs. Specifically, a is constructed as:

$$a = [\alpha_1 \mathbf{1}_{[1]}^T, \dots, \alpha_k \mathbf{1}_{[k]}^T, \dots, \alpha_K \mathbf{1}_{[K]}^T], \tag{16}$$

where $\mathbf{1}_{[k]}$ stands for a $J_{[k]}$ vector of ones. Alternatively, the weights can be stored as the diagonal elements of a diagonal matrix denoted A obtained as

$$A = \text{diag} \{a\} = \text{diag} \left\{ [\alpha_1 \mathbf{1}_{[1]}^T, \dots, \alpha_k \mathbf{1}_{[k]}^T, \dots, \alpha_K \mathbf{1}_{[K]}^T] \right\}. \tag{17}$$

Step Two: Generalized PCA of X GSVD of X

After the weights have been collected, they are used to compute the GSVD of X under the constraints provided by M (masses for the observations) and A (optimum weights for the K tables). This GSVD is expressed as:

$$X = PAQ^T \quad \text{with} \quad P^T M P = Q^T A Q = I. \tag{18}$$

This GSVD corresponds to a GPCA of matrix X and, consequently, will provide *factor scores* to describe the observations and *factor loadings* to describe the variables. Each column of P and Q refers to a *principal component* also called a *dimension* (because the numbers in these columns are often used as coordinates to plot maps, see Ref 50 for more details). In PCA, Eq. 18 is often rewritten as

$$X = FQ^T \quad \text{with} \quad F = PA, \tag{19}$$

where F stores the factor scores (describing the observations) and Q stores the loadings (describing the variables). Note, incidentally, that in the triplet notation, STATIS is equivalent to the statistical analysis of the triplet (X, A, M) .

Because matrix X comprises K tables, each of them comprising $J_{[k]}$ variables, the matrix Q of the left singular vectors can be partitioned in the same way as X . Specifically, Q can be expressed as a column block

matrix as:

$$Q = \begin{bmatrix} Q_{[1]} \\ \vdots \\ Q_{[k]} \\ \vdots \\ Q_{[K]} \end{bmatrix} = \left[Q_{[1]}^T | \cdots | Q_{[k]}^T | \cdots | Q_{[K]}^T \right]^T, \quad (20)$$

where $Q_{[k]}$ is a $J_{[k]}$ by L (with L being the rank of X) matrix storing the right singular vectors corresponding to the variables of matrix $X_{[k]}$. With this in mind, Eq. 18 can re-expressed as:

$$\begin{aligned} X &= [X_{[1]} | \cdots | X_{[k]} | \cdots | X_{[K]}] = P \Delta Q^T \\ &= P \Delta \left(\left[Q_{[1]}^T | \cdots | Q_{[k]}^T | \cdots | Q_{[K]}^T \right]^T \right)^T \\ &= P \Delta \left[Q_{[1]}^T | \cdots | Q_{[k]}^T | \cdots | Q_{[K]}^T \right] \\ &= \left[P \Delta Q_{[1]}^T | \cdots | P \Delta Q_{[k]}^T | \cdots | P \Delta Q_{[K]}^T \right]. \end{aligned} \quad (21)$$

Note, that, the pattern in Eq. 18 does not completely generalize to Eq. 21 because, if we define $A_{[k]}$ as

$$A_{[k]} = \alpha_k I, \quad (22)$$

we have, in general, $Q_{[k]}^T A_{[k]} Q_{[k]} \neq I$.

Factor Scores

The factor scores for X represent the best compromise (i.e., the best common representation) for the set of the K matrices. Recall that these factor scores, called the *compromise* factor scores, are computed (cf., Eqs (18) and (19)) as

$$F = P \Delta. \quad (23)$$

Factor scores can be used to plot the observations as done in standard PCA for which each column of F represents a dimension. Note that the variance of the factor scores of the observations is computed using their masses (stored in matrix M) and can be found as the diagonal of the matrix $F^T M F$. This variance is equal, for each dimension, to the square of the singular value of this dimension as shown by

$$F^T M F = \Delta P^T M P \Delta = \Delta^2. \quad (24)$$

As in standard PCA, F can be obtained from X by combining Eqs (18) and (23) to get:

$$F = P \Delta = X A Q. \quad (25)$$

Taking into account the block structure of X , A , and Q , Eq. 18 can also be rewritten as (cf., Eq. (22)):

$$\begin{aligned} F &= X A Q = [X_{[1]} | \cdots | X_{[k]} | \cdots | X_{[K]}] \times A \times \begin{bmatrix} Q_{[1]} \\ \vdots \\ Q_{[k]} \\ \vdots \\ Q_{[K]} \end{bmatrix} \\ &= \sum_k X_{[k]} A_{[k]} Q_{[k]} = \sum_k \alpha_k X_{[k]} Q_{[k]}. \end{aligned} \quad (26)$$

This equation suggests that the *partial factor scores* for a table can be defined as the projection of this table onto its right singular vectors (i.e., $Q_{[k]}$). Specifically, the partial factor scores for the k th study are stored in a matrix denoted $F_{[k]}$ computed as

$$F_{[k]} = X_{[k]} Q_{[k]}. \quad (27)$$

Note that the compromise factor scores matrix is the *barycenter* of the partial factor scores because it is the weighted average (where the weights are given by the α_k 's) of the partial factor scores (cf., Eq. (25)):

$$\sum_k \alpha_k F_{[k]} = \sum_k \alpha_k X_{[k]} Q_{[k]} = F. \quad (28)$$

Also as in standard PCA, the elements of Q are *loadings* and can be plotted either on their own or along with the factor scores as a biplot.^{69,70}

As the loadings come in blocks (i.e., the loadings correspond to the variables of a table), it is practical to create a biplot with the partial factor scores (i.e., $F_{[k]}$) for a block and the loadings (i.e., $Q_{[k]}$) for this block. In doing so, it is often practical to normalize the loadings such that their variance is commensurable with the variance of the factor scores. This can be achieved, for example, by normalizing, for each dimension, the loadings of a block such that their variance is equal to the square of the singular value of the dimension or even to the singular value itself (as illustrated in the example that we present in a following section). These biplots are helpful for understanding the statistical structure of each block, even though the relative positions of the factor scores and the loadings are not directly interpretable because only the projections of observations on the loading vectors can be meaningfully interpreted in a biplot.^{69,70}

An alternative pictorial representation of the variables and the components plots the correlations between the original variables of X and the factor scores. These correlations are plotted as

two-dimensional maps in which a circle of radius one (called the *circle of correlation*⁵⁰) is also plotted. The closer to the circle a variable is, the better this variable is ‘explained’ by the components used to create the plot (see Ref 71 for an example).

Contribution of Observations, Variables, and Tables to a Dimension

In STATIS, just like in standard PCA, the importance of a dimension (i.e., principal component) is reflected by its eigenvalue which indicates how much of the total inertia (i.e., variance) of the data is explained by this component.

In order to better understand the relationships between components and observations, variables, and tables and also to help interpret a component, we can evaluate how much an observation, a variable, or a whole table contribute to the inertia extracted by a component. In order to do so, we compute descriptive statistics, called *contributions*⁵⁰ (p. 437ff). The stability of these descriptive statistics can also be assessed by cross-validation techniques such as the bootstrap, and this approach can then be used to select the relevant elements for a dimension.

Contribution of an Observation to a Dimension

As stated in Eq. (24), the variance of the factor scores for a given dimension is equal to its eigenvalue (i.e., the square of the singular value) associated with this dimension. If we denote λ_ℓ , the eigenvalue of a given dimension, we can rewrite Eq. (24) as

$$\lambda_\ell = \sum_i m_i \times f_{i,\ell}^2 \tag{29}$$

(where m_i and $f_{i,\ell}$ are, respectively, the mass of the i th observation and the factor score of the i th observation for the ℓ th dimension).

As all the terms $m_i \times f_{i,\ell}^2$ are positive or null, we can evaluate the contribution of an observation to a dimension as the ratio of the squared weighted factor score to the dimension eigenvalue. Formally, the contribution of observation i to component ℓ , denoted $\text{ctr}_{i,\ell}$, is computed as

$$\text{ctr}_{i,\ell} = \frac{m_i \times f_{i,\ell}^2}{\lambda_\ell}. \tag{30}$$

Contributions take values between 0 and 1, and for a given component, the sum of the contributions of all observations is equal to 1. The larger a contribution, the more the observation contributes

to the component. A useful heuristic is to base the interpretation of a component on the observations that have contributions larger than the average contribution. Observations with high contributions and whose factor scores have different signs can then be contrasted to help interpreting the component. Alternatively (as described in a later section) we can use bootstrap ratios to keep for consideration observations with large bootstrap ratios for a given dimension.

Contributions of a Variable to a Dimension

In a manner similar to the observations, we can find the important variables for a given dimension by computing variable contributions. The variance of the loadings for the variables is equal to one when the α weights are taken into account. (cf., Eq. (18)). So if we denote by a_j the alpha weight for the j th variable (cf., Eq. (16)), we have

$$1 = \sum_j a_j \times q_{j,\ell}^2 \tag{31}$$

(where $q_{j,\ell}$ is the loading of the j th variable for the ℓ th dimension). As all terms $a_j \times q_{j,\ell}^2$ are positive or null, we can evaluate the contribution of a variable to a dimension as its squared weighted loading for this dimension. Formally, the contribution of variable j to component ℓ , denoted $\text{ctr}_{j,\ell}$, is computed as

$$\text{ctr}_{j,\ell} = a_j \times q_{j,\ell}^2. \tag{32}$$

Variable contributions take values between 0 and 1, and for a given component, the contributions of all variables sum to 1. The larger a contribution of a variable to a component the more this variable contributes to this component. Variables with high contributions and whose loadings have different signs can then be contrasted to help interpreting the component.

Contribution of a Table to a Dimension

Specific to multiblock analysis is the notion of a table contribution. As a table comprises several variables, the contribution of a table can simply be defined as the sum of the contributions of its variables (a simple consequence of the Pythagorean theorem that states that *squared* lengths are additive). So the contribution of table k to component ℓ is denoted $\text{ctr}_{k,\ell}$ and is defined as

$$\text{ctr}_{k,\ell} = \sum_j^{J[k]} \text{ctr}_{j,\ell}. \tag{33}$$

Table contributions take values between 0 and 1, and for a given component, the contributions of all tables sum to 1. The larger a contribution of a table to a component, the more this table contributes to this component. The contributions of the tables for a given dimension sum to one. An alternative approach re-scales the contributions so that the sum of the contributions for a dimension is now equal to the eigenvalue of this dimension. These re-scaled contributions are called *partial inertias* and are denoted $\mathcal{I}_{\text{partial}}$. The partial inertias are obtained from the contributions by multiplying the contributions for a dimension by the dimension eigenvalue.

STATIS with Cross-Product Matrices

The traditional approach to STATIS focuses on the cross-product matrices $S_{[k]}$ and on the computation of the factor scores for the observations. In this context, the compromise is computed as the optimal linear combination of the $S_{[k]}$ matrices with weights provided by the elements of α . Specifically, the compromise cross-product matrix is denoted $S_{[+]}$ and is computed as

$$S_{[+]} = \sum_k^K \alpha_k S_{[k]}. \tag{34}$$

Note that $S_{[+]}$ can also be directly computed from X as

$$S_{[+]} = XAX^T. \tag{35}$$

The compromise matrix being a weighted sum of cross-product matrices, is also a cross-product matrix (i.e., it is a positive semidefinite matrix) and therefore its eigendecomposition amounts to a PCA. The generalized eigendecomposition under the constraints provided by matrix M of the compromise gives:

$$S_{[+]} = PAP^T \quad \text{with} \quad P^T M P = I. \tag{36}$$

Eqs (35) and (18), together indicate that the generalized eigenvectors of the compromise are the left generalized singular vectors of X (cf., Eq. 18) and that the eigenvalues of $S_{[+]}$ are the squares of the singular values of X (i.e., $\Lambda = \Delta^2$). The loadings can be computed by rewriting Eq. (18) as

$$Q = X^T M P \Delta^{-1}. \tag{37}$$

Similarly, the compromise factor scores can be computed from $S_{[+]}$ (cf., Eq. (25)) as

$$F = S_{[+]} M P \Delta^{-1}. \tag{38}$$

In this context, the loadings for the variables from table k are obtained from Eq. (37) as

$$Q_{[k]} = X_{[k]}^T M P \Delta^{-1}. \tag{39}$$

The factor scores for table k are obtained from Eqs (39) and (27) as

$$F_{[k]} = X_{[k]} Q_{[k]} = X_{[k]} X_{[k]}^T M P \Delta^{-1} = S_{[k]} M P \Delta^{-1}. \tag{40}$$

STATIS as Simple PCA

STATIS can also be computed as the simple PCA of the set of the $X_{[k]}$ matrices, each weighted by the square root of its respective α weight (this assumes, as it is the case in general for STATIS, that matrix M is equal to $\frac{1}{I}I$). Specifically, if we define the matrix

$$\tilde{X} = [\sqrt{\alpha_1} X_{[1]} | \dots | \sqrt{\alpha_{[k]}} X_{[k]} | \dots | \sqrt{\alpha_K} X_{[K]}], \tag{41}$$

whose (simple) SVD is given by

$$\tilde{X} = \tilde{P} \tilde{\Delta} \tilde{Q}^T \quad \text{with} \quad \tilde{P}^T \tilde{P} = \tilde{Q}^T \tilde{Q} = I. \tag{42}$$

Then the factor scores for the observations can be obtained as

$$F = \tilde{P} \tilde{\Delta}. \tag{43}$$

The loadings for table k are obtained as

$$Q_{[k]} = \frac{1}{\sqrt{\alpha}} \tilde{Q}_{[k]}. \tag{44}$$

More on Preprocessing

The preprocessing step is a crucial part of the analysis and can be performed on the columns, on the rows, and on the whole matrix.

Column and Row Preprocessing

Most of the time, each variable is centered (i.e., the mean of each variable is zero) and normalized (i.e., the sum of the squared elements of each column is equal to one, I , or even $I - 1$). In some cases, the normalization affects the rows of the matrix and in this case the sum of each row can be equal to one (e.g., as in correspondence analysis,⁵³ see also Ref 72 for an explicit integration of correspondence analysis and STATIS) or the sum of squares of the elements of a given row can be equal to one (e.g., as in Hellinger/Bhattacharyya analysis⁷³⁻⁷⁷).

Whole Table Preprocessing

A STATIS analysis of a set of tables with very different scales or factorial structures will, in general, favor the tables with the largest variance. A large variance can be due to a large number of variables (compared to the other tables) or to a strong factorial structure. So, in the same way as normalizing variables (e.g., using Z-scores) allows to compare variables measured on different scales, normalizing data tables allows to compare data tables that have different structures or different scales.

There are different ways to normalize data tables. The simplest one is to divide all entries of the data table by its number of columns or, better, by the square root of the number of columns (as done in the so-called ‘Tucker-1’ model and in ‘consensus PCA’^{78–80}). Another straightforward transformation divides all entries of the data table by the square root of the total sum of squares of its elements. This sets the total variance of each table to one and will guarantee that all tables participate equally in the analysis. This normalization is used, for example, in the SUM-PCA technique which originated in the chemometrics tradition.^{14,15} In this article we will favor this normalization because of its simplicity. A closely related procedure is to normalize $\mathbf{X}_{[k]}$ by the square root of the norm of the matrix $\mathbf{Y}_{[k]}\mathbf{Y}_{[k]}^T$. Specifically, here $\mathbf{X}_{[k]}$ is obtained as

$$\mathbf{X}_{[k]} = \mathbf{Y}_{[k]} \times \|\mathbf{Y}_{[k]}\mathbf{Y}_{[k]}^T\|^{-\frac{1}{2}}. \quad (45)$$

This normalization ensures that the cross-product matrices $\mathbf{S}_{[k]}$ all have norm equal to one and therefore that the inner product between cross-product matrices is equal to the R_V -coefficient. A more sophisticated alternative to the normalization problem originated in the multiple factor analysis framework.^{81,82} Here, each table is normalized by dividing each of its entries by the first singular value of this table. This normalization guarantees that all data tables have a first component of unit variance and therefore that all tables can contribute equally to the first dimension of the common structure.

SUPPLEMENTARY ELEMENTS (ALSO KNOWN AS OUT OF SAMPLE)

As in standard PCA, we can use the results of the analysis to compute approximate statistics (e.g., factor scores, loadings, or optimum weights) for *new* elements (i.e., elements that have not been used in the analysis). These new elements are called *supplementary elements*^{50,83} or ‘*out of sample*’

elements.⁸⁴ By contrast with the supplementary elements, the elements actually used in the analysis are called *active* elements. The statistics for the supplementary elements are obtained by projecting these elements onto the active space. In the STATIS framework, we can have supplementary rows and columns (like in PCA) but also supplementary tables. Supplementary rows for which we have values for all J variables and supplementary variables for which we have measurements for all I observations are projected in the same way as for PCA (see, e.g., Ref 50, p. 436ff). Computing statistics for supplementary tables, however, is specific to STATIS.

Supplementary Rows and Columns

Because STATIS is a generalized PCA, we can add supplementary rows and columns as in standard PCA (see Ref 50 for details). Note incidentally, that this procedure assumes that the supplementary rows and columns are scaled in a manner comparable to the rows and columns of the original matrix \mathbf{X} . Specifically, from Eqs (27) and (18), we can compute the factor scores, denoted \mathbf{f}_{sup} for a supplementary observation (i.e., a supplementary row of dimensions 1 by J recording measurements on the same variables as the whole matrix \mathbf{X}). This supplementary row is represented by a 1 by J vector denoted $\mathbf{r}_{\text{sup}}^T$ (which has been preprocessed in the same way as \mathbf{X}), the supplementary factor scores are computed as

$$\mathbf{f}_{\text{sup}} = \mathbf{r}_{\text{sup}}^T \mathbf{A} \mathbf{Q}. \quad (46)$$

Loadings are denoted \mathbf{q}_{sup} for a new column which is itself denoted by an I by 1 vector \mathbf{o}_{sup} (note that \mathbf{o}_{sup} needs to have been preprocessed in a way comparable with the tables). These loadings are obtained, in a way similar to Eq. (46) as

$$\mathbf{q}_{\text{sup}} = \mathbf{o}_{\text{sup}} \mathbf{M} \mathbf{P} \mathbf{A}^{-1}. \quad (47)$$

Supplementary Partial Observation

In some cases, we have supplementary observations for only one (or some) table(s). In this case, called a *partial observation*, we can obtain the *partial* factor scores for this observation from Eq. (27). Specifically, let $\mathbf{x}_{\text{sup}[k]}^T$ denotes a 1 by $J_{[k]}$ vector of measurements collected on the $J_{[k]}$ variables of table k ($\mathbf{x}_{\text{sup}[k]}^T$ should have been pre-processed in the same way as the whole matrix $\mathbf{X}_{[k]}$). The partial factor scores for this supplementary observation from table k are obtained as:

$$\mathbf{f}_{\text{sup}[k]} = \mathbf{x}_{\text{sup}[k]}^T \times \mathbf{Q}_{[k]}. \quad (48)$$

Note that the factor scores for a supplementary observation collected on all tables can be obtained as the weighted (with the α weights) average of the partial factors scores (see Eqs (28) and (46)). Specifically, we can show that

$$\begin{aligned} \sum_k^K \alpha_k \mathbf{f}_{\text{sup}[k]} &= \sum_k^K \mathbf{x}_{\text{sup}[k]}^T (\alpha_k \times \mathbf{I}) \mathbf{Q}_{[k]} \\ &= \sum_k^K \mathbf{x}_{\text{sup}[k]}^T \mathbf{A}_{[k]} \mathbf{Q}_{[k]} = \mathbf{x}_{\text{sup}}^T \mathbf{A} \mathbf{Q} = \mathbf{f}_{\text{sup}}. \end{aligned} \quad (49)$$

In order to compute the loadings for a supplementary variable for a specific table, it suffices to preprocess this variable like the variables of this table (and this includes the whole table normalization if it was used) and then to use Eq. (47).

Supplementary tables

Because STATIS involves tables, it is of interest to be able to project a whole table as a supplementary element. This table will include new variables measured on the same observations described by the active tables. Such a table is represented by the I by J_{sup} matrix \mathbf{Y}_{sup} . The matrix \mathbf{Y}_{sup} is preprocessed in the same manner (e.g., centered, normalized) as the $\mathbf{Y}_{[k]}$ matrices to give the supplementary matrix \mathbf{X}_{sup} . This matrix will provide supplementary factor scores and loadings for the compromise solution, factor scores for the inner-product map, and one supplementary α value as described below.

Factor Scores

In order to obtain the factor scores for a new table, the first step is to obtain (from Eq. (47)) the supplementary loadings which are computed as

$$\mathbf{Q}_{\text{sup}} = \mathbf{X}_{\text{sup}}^T \mathbf{M} \mathbf{P} \mathbf{\Delta}^{-1}. \quad (50)$$

Then, using Eq. (27), we obtain the supplementary factor scores for the new table \mathbf{X}_{sup} as

$$\mathbf{F}_{\text{sup}} = \mathbf{X}_{\text{sup}} \mathbf{Q}_{\text{sup}} = \mathbf{X}_{\text{sup}} \mathbf{X}_{\text{sup}}^T \mathbf{M} \mathbf{P} \mathbf{\Delta}^{-1} = \mathbf{S}_{\text{sup}} \mathbf{M} \mathbf{P} \mathbf{\Delta}^{-1}. \quad (51)$$

Inner Product Map Projection

When evaluating a supplementary table, it is important to evaluate its similarity to the set of the active tables. This is done by computing the inner product between this supplementary table and all the other tables and then projecting the vector made of these inner products onto the eigenvectors of the inner product matrix.

The first step is to compute the cross-product matrix associated to \mathbf{X}_{sup} as

$$\mathbf{S}_{\text{sup}} = \mathbf{X}_{\text{sup}} \mathbf{X}_{\text{sup}}^T. \quad (52)$$

Then the K inner products between \mathbf{S}_{sup} and each of the K cross-product matrices $\mathbf{S}_{[k]}$ are computed (see Eq. 8) and stored into the K by 1 vector denoted \mathbf{c}_{sup} :

$$\mathbf{c}_{\text{sup}} = [(\mathbf{S}_{\text{sup}}, \mathbf{S}_{[1]}), \dots, (\mathbf{S}_{\text{sup}}, \mathbf{S}_{[k]}), \dots, (\mathbf{S}_{\text{sup}}, \mathbf{S}_{[K]})]^T. \quad (53)$$

The factor scores of the supplementary table in the inner product map are stored in the 1 by L vector denoted $\mathbf{g}_{\text{sup}}^T$ computed as (see Eq. (13)):

$$\mathbf{g}_{\text{sup}}^T = \mathbf{c}_{\text{sup}}^T \mathbf{U} \mathbf{\Gamma}^{-\frac{1}{2}}. \quad (54)$$

Alternatively, if we denote by \mathbf{z}_{sup} the I^2 by 1 vector version of \mathbf{S}_{sup} (i.e., $\mathbf{z}_{\text{sup}} = \text{vec} \{ \mathbf{S}_{\text{sup}} \}$), we can, using Eq. (14), compute \mathbf{g}_{sup} as

$$\mathbf{g}_{\text{sup}}^T = \mathbf{z}_{\text{sup}}^T \mathbf{V} = \text{vec} \{ \mathbf{S}_{\text{sup}} \}^T \mathbf{V} = \text{vec} \{ \mathbf{X}_{\text{sup}} \mathbf{X}_{\text{sup}}^T \}^T \mathbf{V}. \quad (55)$$

STATIS Weight: α_{sup}

The value of the STATIS weight for a supplementary table is denoted α_{sup} and is obtained by combining Eqs (15) and (55) to give:

$$\alpha_{\text{sup}} = \gamma_1^{-\frac{1}{2}} g_{\text{sup},1} \times (\mathbf{u}_1 \mathbf{1})^{-1}. \quad (56)$$

where $g_{\text{sup},1}$ is the value of the supplementary factor score for the first dimension of the inner product matrix and γ_1 is the first eigenvalue of the inner product matrix.

INFERENCE ASPECTS

STATIS is a descriptive multivariate technique, but it is often important to be able to complement the descriptive conclusions of an analysis by assessing if its results are reliable and replicable. For example, a standard question, in the PCA framework, is to find the number of reliable components and most of the approaches used in PCA will work also for STATIS. For example, we can use of the informal ‘scree’ test (also known as ‘elbow’) and the more formal tests RESS, PRESS, and Q^2 statistics (see, for details, e.g., Ref 50, p. 440ff).

In addition to standard PCA methodology, two problems are specific to STATIS. The first one is to

decide if the set of tables is sufficiently homogenous to be considered a random sample from a single population. If so, integrating these tables into a compromise is relevant. If not then maybe the set of tables needs to be partitioned into homogenous sets of tables and a new analysis can be run on each of these homogenous sets. This approach has been developed analytically recently^{85–88} and can, under standard (but somewhat stringent) assumptions (e.g., normality, random sampling of the tables, etc.), provide point estimates, statistical tests, and confidence intervals. This recent important line of work is likely to be expanded to more general cases in the near future.

A promising nonparametric alternative to this analytical approach is to use cross-validation resampling techniques such as permutation tests,⁸⁹ the jackknife,⁹⁰ or the bootstrap.^{91–93} For example, a null hypothesis test for the significance of the first or second eigenvalue of the inner product matrix could be obtained by permuting the elements of the columns of the matrices $X_{[k]}$ (the permutation needs to be done per column in order to keep the column normalization) and to compute the related matrices $S_{[k]}$ which are then used to compute a ‘permuted matrix’ Z (see Eq. (10)). This permuted Z is, in turn, used to create a ‘permuted’ version of matrix C whose eigenvalues will correspond to eigenvalues obtained when the data are random (permuting the entries of C will not lead to the correct test because C will *not*, in general, be positive definite under permutation of its elements). Repeating this procedure a large number of times will provide a sampling distribution of the eigenvalues of C under the null hypothesis and an observed rare value for the analysis under consideration can then be judged significant (see Ref 42 for an example).

Another interesting problem is to evaluate if STATIS improves over a simple multitable analysis which has all weights equal to $\frac{1}{I}$. This question can be addressed with techniques such as the jackknife or the bootstrap as illustrated below.

Jackknife for the Optimum Weights

Recall that the optimal weights (i.e., the α ’s stored in vector α) are used to compute the compromise as a linear combination of the original tables. These weights are computed to optimize the similarity between the tables and the compromise. In some cases we can consider that these tables as fixed (i.e., they constitute the population of interest) and that the observations are random (specifically, independent and identically distributed random variables, or ‘i.i.d.’

in statistical jargon). In such cases, replications would keep the same tables but will sample new (i.e., different) observations. An obvious question, then, is: How different would the new α values be? If we assume that the observations are i.i.d., we can use estimation techniques such as the jackknife⁹⁰ or the bootstrap.⁹¹ As an illustration, we present how the jackknife could be used to estimate the values of α when the tables are considered as a fixed factor and the observations are a random factor.

For the jackknife, we leave out each observation in turn and compute an estimate of α without this observation. We then integrate all these estimates to derive an unbiased estimate of α and of its variability (we follow here the notations and approach detailed in Ref 90).

Specifically, if we consider the i th observation, we denote α_{-i} the vector of optimum weights computed without this observation. We then compute the *pseudo-value* of α , denoted α_i^* as:

$$\alpha_i^* = I\alpha - (I - 1)\alpha_{-i}. \tag{57}$$

The jackknife estimator of α is denoted α^* and is computed as the mean of the pseudovalues:

$$\alpha^* = \frac{1}{I} \sum_i \alpha_i^* = \alpha - \frac{I - 1}{I} \sum_i \alpha_{-i} \tag{58}$$

(note that some values of α_i^* could be negative, in such unlikely cases, negative values are set to zero).

The vector variance of the α_i^* is computed as

$$\hat{\sigma}_{\alpha_i^*}^2 = \frac{1}{I - 1} \sum_i (\alpha_i^* - \alpha^*)^2. \tag{59}$$

The standard error of the mean is obtained as

$$\hat{\sigma}_{\alpha^*} = \sqrt{\frac{\hat{\sigma}_{\alpha_i^*}^2}{I}}. \tag{60}$$

When the observations are i.i.d., this standard error can be used to compute confidence intervals as the sampling distribution of the α ’s (i.e., the elements of α^*) converges to a normal distribution. The standard error can also be used to compute pseudo- t values called *jackknife ratios* in order to test null-hypothesis. In the case of STATIS, one interesting null hypothesis for an α_k would be to test if its value differs from the chance value of $\frac{1}{I}$.

Bootstrap for the Factor Scores

If we assume that the tables constitute a random factor (i.e., the tables are i.i.d, and sampled from a potentially infinite set of tables) and if we consider the observations as a fixed factor, we may want to estimate the stability of the compromise factor scores.

Such an evaluation could be done, for example, by using a jackknife or a bootstrap approach.

We briefly sketch here a possible bootstrap approach for the factor scores (see Ref 93 for the problem of bootstrapping in the context of the SVD, and Ref 94 for a review of the bootstrap in the PCA context). The main idea is to use the properties of Eq. (28) which indicate that the compromise factor scores are the weighted (by the α 's) average of the factor scores of the tables. Therefore we can obtain bootstrap confidence intervals by repeatedly sampling with replacement from the set of tables and compute for each of these samples an estimate of the α weights and the compromise factor scores (this approach corresponds to the *partial bootstrap* of Ref 94, see Ref 42 for an alternative approach using split-half resampling). From these estimates we can also compute *bootstrap ratios* for each dimension by dividing the mean of the bootstrap estimates by their standard deviation. These bootstrap ratios are akin to t statistics and can be used to detect observations that reliably contribute to a given dimension. So, for example, for a given dimension and a given observation a value of the bootstrap ratio larger than 2 will be considered reliable (by analogy with a t larger than 2 which would be 'significant' at $p < .05$). When evaluating bootstrap ratios, the multiple comparisons problem needs to be taken into account by using, for example, a Bonferroni-type correction⁹⁵ and instead of using critical values corresponding to say $p < 0.05$ we would use values corresponding to $p < \frac{0.05}{T}$.

More formally, in order to compute a bootstrap estimate, we need to generate a bootstrap sample. To do so, we first take a sample of integers with replacement from the set of integers from 1 to K . Recall that, when sampling with replacement, any element from the set can be sampled zero, one, or more than one times. We call this set \mathcal{B} (for *bootstrap*). For example, with 5 elements, a possible bootstrap set could be $\mathcal{B} = \{1, 5, 1, 3, 3\}$. We then generate a new data set (i.e., a new \mathbf{X} matrix comprising K tables) using matrices $\mathbf{X}_{[k]}$ with these indices. So with $K = 5$, this would give the following bootstrap set

$$\{\mathbf{X}_{[1]}, \mathbf{X}_{[5]}, \mathbf{X}_{[1]}, \mathbf{X}_{[3]}, \mathbf{X}_{[3]}\}. \quad (61)$$

We will then use this set of matrices to compute a matrix of inner products (denoted \mathbf{C}_1^* , see Eq. (8)) from

which we would derive a set of α weights (denoted α_1^* , see Eq. (15)). From these weights we will then compute a compromise cross product matrix (denoted $\mathbf{S}_{[+1]}^*$, see Eq. (34)). Finally we will compute the *bootstrapped factor scores* (denoted \mathbf{F}_1^*) from this compromise by projecting the bootstrapped compromise on the compromise as a supplementary element (see Eqs (40) and (51)). Interestingly, as a consequence of the barycentric properties of the factor scores (see Eq. (28)), this last step can also directly obtained by computing \mathbf{F}_1^* as the weighted (by the α^* 's) average of the corresponding partial factor scores. We then repeat the procedure for a large number of times (e.g., $L = 1000$) and generate L bootstrapped matrices of factor scores \mathbf{F}_ℓ^* . From these bootstrapped matrices of factor scores, we can derive confidence intervals and estimate the mean factor scores as the mean of the bootstrapped factor scores. Formally, $\bar{\mathbf{F}}^*$ denotes the bootstrap estimated factor scores, and is computed as

$$\bar{\mathbf{F}}^* = \frac{1}{L} \sum_{\ell} \mathbf{F}_\ell^*. \quad (62)$$

In a similar way, the bootstrapped estimate of the variance is obtained as the variance of the \mathbf{F}_ℓ^* matrices. Formally, $\hat{\sigma}_{\mathbf{F}^*}^2$ denoted the bootstrapped estimate of the variance and is computed as

$$\hat{\sigma}_{\mathbf{F}^*}^2 = \frac{1}{L} \left(\sum_{\ell} (\mathbf{F}_\ell^* - \bar{\mathbf{F}}^*) \circ (\mathbf{F}_\ell^* - \bar{\mathbf{F}}^*) \right) \quad (63)$$

(where \circ denotes the Hadamard product). Bootstrap ratios are computed by dividing the bootstrapped mean by the bootstrapped standard deviation (denoted $\hat{\sigma}_{\mathbf{F}^*}$ and is the square root of the bootstrapped estimate of the variance).

Bootstrapped Confidence Intervals

The bootstrap factor scores (i.e., the \mathbf{F}_ℓ^* 's) can also be used to compute confidence intervals (CI) for the observations. For a given dimension, the bootstrapped factor scores of an observation can be ordered from the smallest to the largest and a confidence interval for a given p value can be obtained by trimming the upper and lower $\frac{1}{p}$ proportion of the distribution. In general, for a given dimension, the bootstrap ratios, and the CIs will agree to detect the relevant observations. In addition, CIs can also be plotted directly on the factor scores map as confidence ellipsoids or confidence convex-hulls which comprise a $1 - p$ proportion of the bootstrapped factor scores (see Ref 42 and our example below for an illustration). When the ellipsoids or convex-hulls of two observations do not

overlap, these two observations can be considered as reliably different. Like for the bootstrap ratios, in order to correct for the potential problem of multiple comparisons, a Bonferroni type of correction can also be implemented when plotting hulls or ellipsoids (see Ref 42 for details).

WHAT DOES STATIS OPTIMIZE?

Intuitively, the goal of the STATIS method is to find a set of weights (i.e., the α_k 's) such that the factor scores of matrix \mathbf{X} are as close as possible to the set of the partial factor scores of the K tables comprising \mathbf{X} . This intuition can be formalized by two main types of criteria (which are equivalent as they correspond to maximizing the same quantities). We express formally these criteria in this section and give the proofs in Appendix.

Largest Variance of the Compromise

The first criterion can be expressed in several equivalent ways. The simplest one indicates that the variance of the compromise is maximal (and therefore that the factor scores of \mathbf{X} have maximal variance as they are obtained from its GSVD). Formally, this amounts to maximizing the following criterion:

$$\mathcal{V} = \|\mathbf{S}_{[+]} \|^2 = \langle \mathbf{S}_{[+]}, \mathbf{S}_{[+]} \rangle \text{ with } \boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1. \quad (64)$$

This criterion can be re-expressed as a minimization problem, for which STATIS seeks the *minimum* distance between the individual cross-product matrices and the compromise. Namely, STATIS minimizes:

$$\mathcal{D} = \sum_k^K \|\mathbf{S}_{[k]} - \alpha_k \mathbf{S}_{[+]} \|^2 \text{ with } \boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1. \quad (65)$$

Compromise Most Similar to Tables

An alternative criterion maximized by STATIS is the overall similarity between the tables and the compromise. Specifically, STATIS maximizes²¹:

$$\mathcal{S} = \sum_k^K \langle \mathbf{S}_{[k]}, \mathbf{S}_{[+]} \rangle^2 \text{ with } \boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1. \quad (66)$$

Note that, for the STATIS criterion, the similarity between two cross-product matrices is evaluated by the *squared* inner product of these two matrices.

DUAL-STATIS

In dual-STATIS, the data consist of K sets of observations measured on the same set of variables. Here, instead of computing K cross-product matrices between the observations, we compute K covariance matrices between the variables (one per set of observations). The dual-STATIS approach then follows the same steps as standard STATIS and will provide a compromise map for the variables (instead of the observations in STATIS), and partial loadings for each table.

AN EXAMPLE

A typical example of using STATIS is the description of a set of products by a group of expert assessors. This type of data could be analyzed using a standard PCA, but this approach obviously neglects the inter-assessor differences. STATIS has the advantages of providing a compromise space for the products as well as evaluating differences between the assessors. We illustrate STATIS with a fictitious example from wine tasting.

For our example, we selected twelve wines made from Sauvignon Blanc grapes coming from three wine regions (four wines from each region): New Zealand, France, and Canada. We then asked 10 expert assessors to evaluate these wines. The assessors were asked (1) to evaluate the wines on 9-point rating scales, using four variables considered as standard for the evaluation of these wines (cat-pee, passion-fruit, green pepper, and mineral) and, (2) if they felt the need, to add some variables of their own (some assessors choose none, some choose one or two more variables). The raw data are presented in Table 1. The goals of the analysis were twofold: (1) to obtain a typology of the wines, and (2) to discover agreement (if any) between the assessors.

For the example, the data consist of $K = 10$ tables (one for each assessor) shown in Table 1. For example, the first table denoted $\mathbf{Y}_{[1]}$ is equal to

$$\mathbf{Y}_{[1]} = \begin{bmatrix} 8 & 6 & 7 & 4 & 1 & 6 \\ 7 & 5 & 8 & 1 & 2 & 8 \\ 6 & 5 & 6 & 5 & 3 & 4 \\ 9 & 6 & 8 & 4 & 3 & 5 \\ 2 & 2 & 2 & 8 & 7 & 3 \\ 3 & 4 & 4 & 9 & 6 & 1 \\ 5 & 3 & 5 & 4 & 8 & 3 \\ 5 & 2 & 4 & 8 & 7 & 4 \\ 8 & 6 & 8 & 4 & 4 & 7 \\ 4 & 6 & 2 & 5 & 3 & 4 \\ 8 & 4 & 8 & 1 & 3 & 3 \\ 5 & 3 & 6 & 4 & 4 & 2 \end{bmatrix}. \quad (67)$$

TABLE 1 | Raw Data (Tables $Y_{(1)}$ through $Y_{(10)}$)

	Assessor 1					Assessor 2					Assessor 3					Assessor 4					Assessor 5											
	V1	V2	V3	V4	V5	V1	V2	V3	V4	V7	V8	V1	V2	V3	V4	V9	V10	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8	V1	V2	V3	V4	V11
NZ ₁	8	6	7	4	1	6	8	6	8	3	7	5	8	6	8	3	7	2	9	5	8	2	6	9	6	9	3	8	2			
NZ ₂	7	5	8	1	2	8	6	5	6	3	7	7	8	7	7	2	8	2	8	7	7	3	5	7	7	7	1	9	2			
NZ ₃	6	5	6	5	3	4	6	6	6	5	8	7	8	7	7	6	9	1	8	8	9	2	7	7	7	7	1	7	2			
NZ ₄	9	6	8	4	3	5	8	6	8	4	6	6	8	2	8	3	9	3	8	8	9	4	7	8	9	7	5	6	1			
FR ₁	2	2	2	8	7	3	2	3	1	7	4	3	3	4	3	6	4	6	4	2	2	4	3	4	4	4	2	4	4			
FR ₂	3	4	4	9	6	1	4	3	4	9	3	5	4	3	4	8	3	9	3	2	2	6	2	4	5	5	6	1	5			
FR ₃	5	3	5	4	8	3	3	2	7	4	4	5	4	5	4	5	2	3	6	4	4	6	4	6	5	7	2	3	1			
FR ₄	5	2	4	8	7	4	4	3	5	5	3	3	6	3	7	7	1	7	5	2	2	9	4	6	6	5	8	4	5			
CA ₁	8	6	8	4	4	7	8	6	9	5	5	6	8	5	9	1	5	2	7	5	6	3	2	8	6	8	2	5	4			
CA ₂	4	6	2	5	3	4	5	5	5	6	5	8	5	5	4	6	5	1	5	6	6	4	4	6	6	6	4	6	3			
CA ₃	8	4	8	1	3	3	8	4	8	3	7	7	8	3	7	3	5	4	7	3	6	1	6	7	4	8	4	5	1			
CA ₄	5	3	6	4	4	2	5	3	7	4	8	5	5	4	4	5	4	3	5	2	2	6	6	5	5	5	5	6	1			

	Assessor 6					Assessor 7					Assessor 8					Assessor 9					Assessor 10										
	V1	V2	V3	V4	V5	V1	V2	V3	V4	V7	V8	V1	V2	V3	V4	V14	V15	V1	V2	V3	V4	V8	V1	V2	V3	V4	V15	V1	V2	V3	V4
NZ ₁	8	5	6	2	9	8	5	8	4	7	6	7	4	9	2	8	9	1	7	8	6	9	1	7	8	6	7	5			
NZ ₂	6	6	6	2	4	7	6	8	4	6	5	6	2	7	2	8	7	9	1	6	7	9	1	6	7	5	7	3			
NZ ₃	7	7	7	2	7	6	7	6	3	6	6	6	4	9	2	7	8	4	7	7	8	4	7	7	6	6	2				
NZ ₄	8	7	8	2	8	7	8	6	1	8	7	8	2	8	2	8	9	3	9	8	9	9	3	9	8	7	7	4			
FR ₁	3	2	2	7	2	4	2	3	6	3	3	4	4	4	4	3	4	4	3	4	4	4	5	4	2	3	1	7			
FR ₂	3	3	3	3	4	4	4	4	4	4	4	4	4	7	3	6	5	5	7	2	3	5	7	2	3	3	9				
FR ₃	4	2	3	3	3	4	3	4	4	5	3	5	3	3	3	3	3	3	5	5	5	5	6	3	4	2	5	8			
FR ₄	5	3	5	9	3	5	3	5	7	6	4	6	3	2	4	4	5	6	5	5	6	5	3	3	4	2	8	8			
CA ₁	7	7	7	1	4	8	4	9	4	8	6	5	4	5	4	4	5	4	8	7	8	4	7	8	6	7	7	7			
CA ₂	4	6	2	4	6	4	7	5	2	5	5	4	6	1	5	5	4	6	4	6	4	5	5	6	4	2	8	8			
CA ₃	7	4	8	2	3	8	5	7	3	7	4	8	2	6	2	6	2	8	4	4	4	7	4	5	7	4	8	5			
CA ₄	4	5	3	3	7	4	3	5	2	5	4	6	2	4	4	3	4	5	4	4	4	5	4	4	5	4	6	6			

V1, Cat Pee; V2, Passion Fruit; V3, Green Pepper; V4, Mineral; V5, Smoky; V6, Citrus; V7, Tropical; V8, Leafy; V9, Grassy; V10, Flinty; V11, Vegetal; V12, Hay; V13, Melon; V14, Grass; V15, Peach.

Each table was then preprocessed by first centering and normalizing each column and then by normalizing the whole table such that the sum of the square values of all elements of this table was equal to one. For example, $X_{[1]}$, the preprocessed matrix for Assessor 1, is equal to:

$$X_{[1]} = \begin{bmatrix} 0.12 & 0.13 & 0.07 & -0.04 & -0.18 & 0.11 \\ 0.07 & 0.05 & 0.13 & -0.18 & -0.12 & 0.23 \\ 0.01 & 0.05 & 0.02 & 0.01 & -0.07 & -0.01 \\ 0.18 & 0.13 & 0.13 & -0.04 & -0.07 & 0.05 \\ -0.21 & -0.18 & -0.20 & 0.16 & 0.15 & -0.07 \\ -0.16 & -0.03 & -0.09 & 0.21 & 0.10 & -0.19 \\ -0.05 & -0.11 & -0.04 & -0.04 & 0.21 & -0.07 \\ -0.05 & -0.18 & -0.09 & 0.16 & 0.15 & -0.01 \\ 0.12 & 0.13 & 0.13 & -0.04 & -0.01 & 0.17 \\ -0.10 & 0.13 & -0.20 & 0.01 & -0.07 & -0.01 \\ 0.12 & -0.03 & 0.13 & -0.18 & -0.07 & -0.07 \\ -0.05 & -0.11 & 0.02 & -0.04 & -0.01 & -0.13 \end{bmatrix} \quad (68)$$

The eigendecomposition of C is then obtained as (cf., Eq. (12)):

$$C = U\Theta U^T = \begin{bmatrix} 0.325 & -0.212 \\ 0.326 & -0.359 \\ 0.289 & 0.212 \\ 0.326 & 0.561 \\ 0.253 & 0.443 \\ 0.311 & -0.237 \\ 0.294 & -0.391 \\ 0.279 & 0.124 \\ 0.349 & 0.137 \\ 0.389 & -0.165 \end{bmatrix} \times \text{diag} \left\{ \begin{matrix} 4.135 \\ 0.224 \end{matrix} \right\} \times \begin{bmatrix} 0.325 & -0.212 \\ 0.326 & -0.359 \\ 0.289 & 0.212 \\ 0.326 & 0.561 \\ 0.253 & 0.443 \\ 0.311 & -0.237 \\ 0.294 & -0.391 \\ 0.279 & 0.124 \\ 0.349 & 0.137 \\ 0.389 & -0.165 \end{bmatrix}^T \quad (71)$$

Analyzing the Between-Table Structure

To each $X_{[k]}$ matrix, we associated its $I \times I$ cross-product matrix, denoted $S_{[k]}$. For example, the $S_{[k]}$ matrix for Assessor 1 is obtained as (cf., Eq. (3)):

$$S_{[1]} = X_{[1]}X_{[1]}^T = \begin{bmatrix} 0.08 & 0.08 & 0.02 & 0.07 & -0.11 & -0.08 & -0.07 & -0.07 & 0.06 & 0.00 & 0.03 & -0.03 \\ 0.08 & 0.13 & 0.01 & 0.06 & -0.11 & -0.12 & -0.05 & -0.07 & 0.08 & -0.02 & 0.05 & -0.03 \\ 0.02 & 0.01 & 0.01 & 0.01 & -0.02 & -0.01 & -0.02 & -0.02 & 0.01 & 0.01 & 0.01 & -0.00 \\ 0.07 & 0.06 & 0.01 & 0.07 & -0.11 & -0.07 & -0.04 & -0.06 & 0.07 & -0.02 & 0.04 & -0.02 \\ -0.11 & -0.11 & -0.02 & -0.11 & 0.17 & 0.12 & 0.07 & 0.11 & -0.10 & 0.03 & -0.08 & 0.03 \\ -0.08 & -0.12 & -0.01 & -0.07 & 0.12 & 0.12 & 0.04 & 0.07 & -0.08 & 0.03 & -0.06 & 0.02 \\ -0.07 & -0.05 & -0.02 & -0.04 & 0.07 & 0.04 & 0.06 & 0.05 & -0.04 & -0.02 & -0.01 & 0.02 \\ -0.07 & -0.07 & -0.02 & -0.06 & 0.11 & 0.07 & 0.05 & 0.09 & -0.05 & -0.01 & -0.05 & 0.01 \\ 0.06 & 0.08 & 0.01 & 0.07 & -0.10 & -0.08 & -0.04 & -0.05 & 0.08 & -0.02 & 0.02 & -0.04 \\ 0.00 & -0.02 & 0.01 & -0.02 & 0.03 & 0.03 & -0.02 & -0.01 & -0.02 & 0.07 & -0.04 & -0.01 \\ 0.03 & 0.05 & 0.01 & 0.04 & -0.08 & -0.06 & -0.01 & -0.05 & 0.02 & -0.04 & 0.07 & 0.02 \\ -0.03 & -0.03 & -0.00 & -0.02 & 0.03 & 0.02 & 0.02 & 0.01 & -0.04 & -0.01 & 0.02 & 0.03 \end{bmatrix} \quad (69)$$

Next, the inner-products between all pairs of $S_{[k]}$ tables (i.e., the coefficients $c_{k,k'}$) are computed and stored in matrix C . Using Eqs (8), (10), and (11), we get the following C matrix:

$$C = \begin{bmatrix} 0.51 & 0.44 & 0.40 & 0.38 & 0.34 & 0.40 & 0.41 & 0.35 & 0.48 & 0.52 \\ 0.44 & 0.52 & 0.36 & 0.41 & 0.30 & 0.42 & 0.40 & 0.39 & 0.43 & 0.54 \\ 0.40 & 0.36 & 0.42 & 0.39 & 0.33 & 0.35 & 0.33 & 0.31 & 0.42 & 0.46 \\ 0.38 & 0.41 & 0.39 & 0.56 & 0.37 & 0.40 & 0.37 & 0.39 & 0.46 & 0.51 \\ 0.34 & 0.30 & 0.33 & 0.37 & 0.35 & 0.30 & 0.27 & 0.29 & 0.37 & 0.39 \\ 0.40 & 0.42 & 0.35 & 0.40 & 0.30 & 0.48 & 0.39 & 0.34 & 0.45 & 0.51 \\ 0.41 & 0.40 & 0.33 & 0.37 & 0.27 & 0.39 & 0.46 & 0.32 & 0.41 & 0.46 \\ 0.35 & 0.39 & 0.31 & 0.39 & 0.29 & 0.34 & 0.32 & 0.43 & 0.38 & 0.44 \\ 0.48 & 0.43 & 0.42 & 0.46 & 0.37 & 0.45 & 0.41 & 0.38 & 0.60 & 0.54 \\ 0.52 & 0.54 & 0.46 & 0.51 & 0.39 & 0.51 & 0.46 & 0.44 & 0.54 & 0.68 \end{bmatrix} \quad (70)$$

for the first two components.

From the eigendecomposition of C , we obtain the factor scores (for the first two components) of the tables for the inner product matrix as

(cf., Eq. 13):

$$G = U\Theta^{\frac{1}{2}} = \begin{bmatrix} 0.662 & -0.100 \\ 0.662 & -0.170 \\ 0.588 & 0.100 \\ 0.662 & 0.265 \\ 0.515 & 0.209 \\ 0.633 & -0.112 \\ 0.598 & -0.185 \\ 0.567 & 0.059 \\ 0.710 & 0.065 \\ 0.791 & -0.078 \end{bmatrix}$$

The projections of the assessors onto the first and second components are shown in Figure 2 (note that the elements of the first column of U , denoted u_1 , are all positive). From this visual display, we can see that the 10 tables of the assessors should all receive similar weights in the compromise.

Permutation Test: Significance of the First Two Eigenvalues

When using STATIS two questions are important: (1) Is the first dimension of the C matrix significant? In practice, it seems to always be the case, however it is useful to confirm that it indeed is so for a specific analysis; but also (2) is the second dimension larger than chance? This last question is particularly relevant for the STATIS model, because a positive answer suggests that there are clusters of tables and that separate STATIS analyses (i.e., one per cluster) would be more appropriate^{85–88} than an overall analysis.

Because the data at hand do not seem to follow the assumptions of the analytical approach, we decided to evaluate the question of the significance of the eigenvalues with a permutation test. To do so, we

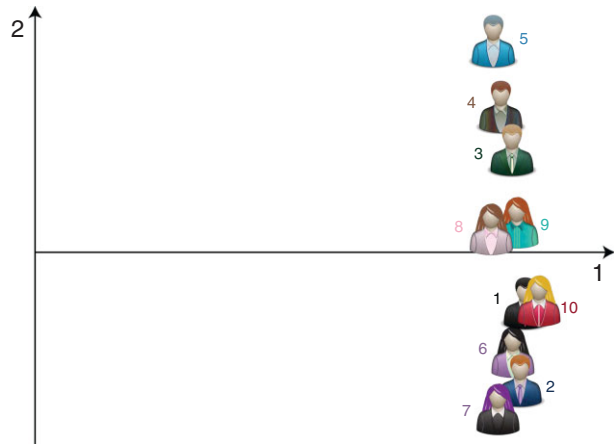


FIGURE 2 | Inner product map. Projection of the assessors onto the first and second components. Note that the first dimension is positive because all the inner products are positive.

started by permuting the entries within each column of the matrices $X_{[k]}$. For example, the first permuted matrix denoted $X_{[1]}^*$ would be:

$$X_{[1]}^* = \begin{bmatrix} -0.16 & -0.11 & 0.13 & 0.16 & -0.07 & 0.17 \\ -0.05 & 0.05 & -0.09 & 0.21 & 0.15 & -0.19 \\ -0.05 & 0.13 & -0.20 & -0.04 & -0.01 & -0.01 \\ 0.12 & -0.11 & -0.20 & 0.01 & -0.07 & 0.23 \\ 0.07 & 0.05 & 0.07 & -0.04 & 0.15 & -0.07 \\ 0.12 & 0.13 & 0.02 & -0.04 & -0.07 & -0.07 \\ -0.05 & 0.13 & -0.04 & 0.01 & 0.21 & -0.13 \\ -0.10 & -0.18 & 0.13 & -0.04 & -0.12 & -0.07 \\ 0.12 & 0.13 & 0.02 & -0.18 & -0.07 & -0.01 \\ 0.18 & -0.03 & 0.13 & -0.04 & -0.01 & 0.05 \\ 0.01 & -0.18 & 0.13 & 0.16 & -0.18 & -0.01 \\ -0.21 & -0.03 & -0.09 & -0.18 & 0.10 & 0.11 \end{bmatrix} \tag{72}$$

We then computed all the corresponding permuted cross product matrices (denoted $S_{[k]}^*$). For example $S_{[1]}^*$ is equal to (see Eq. (69)):

$$S_{[1]}^* = X_{[1]}^* X_{[1]}^{*T} = \begin{bmatrix} 0.11 & -0.02 & -0.04 & 0.01 & -0.04 & -0.04 & -0.05 & 0.04 & -0.06 & -0.01 & 0.07 & 0.01 \\ -0.02 & 0.12 & 0.02 & -0.05 & 0.02 & -0.00 & 0.07 & -0.03 & -0.05 & -0.04 & -0.01 & -0.03 \\ -0.04 & 0.02 & 0.06 & 0.02 & -0.01 & 0.01 & 0.02 & -0.04 & 0.02 & -0.04 & -0.05 & 0.03 \\ 0.01 & -0.05 & 0.02 & 0.12 & -0.04 & -0.01 & -0.06 & -0.03 & -0.00 & 0.01 & 0.01 & 0.01 \\ -0.04 & 0.02 & -0.01 & -0.04 & 0.04 & 0.01 & 0.04 & -0.02 & 0.01 & 0.01 & -0.03 & -0.01 \\ -0.04 & -0.00 & 0.01 & -0.01 & 0.01 & 0.04 & 0.01 & -0.02 & 0.04 & 0.02 & -0.01 & -0.04 \\ -0.05 & 0.07 & 0.02 & -0.06 & 0.04 & 0.01 & 0.08 & -0.04 & -0.00 & -0.03 & -0.06 & 0.01 \\ 0.04 & -0.03 & -0.04 & -0.03 & -0.02 & -0.02 & -0.04 & 0.08 & -0.02 & 0.00 & 0.07 & 0.00 \\ -0.06 & -0.05 & 0.02 & -0.00 & 0.01 & 0.04 & -0.00 & -0.02 & 0.07 & 0.03 & -0.04 & -0.01 \\ -0.01 & -0.04 & -0.04 & 0.01 & 0.01 & 0.02 & -0.03 & 0.00 & 0.03 & 0.05 & 0.02 & -0.04 \\ 0.07 & -0.01 & -0.05 & 0.01 & -0.03 & -0.01 & -0.06 & 0.07 & -0.04 & 0.02 & 0.11 & -0.06 \\ 0.01 & -0.03 & 0.03 & 0.01 & -0.01 & -0.04 & 0.01 & 0.00 & -0.01 & -0.04 & -0.06 & 0.11 \end{bmatrix}$$

With these permuted cross-product matrices, we computed a permuted inner product matrix denoted C^* equal to (cf., Eq. (70)):

$J = 53 \times 1$ vector \mathbf{a} , which can itself be used to fill in the diagonal elements of the 53×53 diagonal matrix \mathbf{A} :

$$C^* = \begin{bmatrix} 0.24 & 0.09 & 0.08 & 0.10 & 0.07 & 0.12 & 0.11 & 0.10 & 0.09 & 0.05 \\ 0.09 & 0.22 & 0.08 & 0.11 & 0.12 & 0.11 & 0.10 & 0.12 & 0.06 & 0.05 \\ 0.08 & 0.08 & 0.22 & 0.08 & 0.12 & 0.05 & 0.08 & 0.10 & 0.07 & 0.08 \\ 0.10 & 0.11 & 0.08 & 0.26 & 0.07 & 0.11 & 0.11 & 0.08 & 0.09 & 0.12 \\ 0.07 & 0.12 & 0.12 & 0.07 & 0.25 & 0.05 & 0.06 & 0.12 & 0.06 & 0.08 \\ 0.12 & 0.11 & 0.05 & 0.11 & 0.05 & 0.30 & 0.13 & 0.09 & 0.11 & 0.08 \\ 0.11 & 0.10 & 0.08 & 0.11 & 0.06 & 0.13 & 0.32 & 0.10 & 0.10 & 0.05 \\ 0.10 & 0.12 & 0.10 & 0.08 & 0.12 & 0.09 & 0.10 & 0.24 & 0.10 & 0.06 \\ 0.09 & 0.06 & 0.07 & 0.09 & 0.06 & 0.11 & 0.10 & 0.10 & 0.23 & 0.11 \\ 0.05 & 0.05 & 0.08 & 0.12 & 0.08 & 0.08 & 0.05 & 0.06 & 0.11 & 0.30 \end{bmatrix}. \quad (73)$$

The eigendecomposition of C^* gave the following eigenvalues:

$$\Theta^* = \text{diag} \{ [1.070 \ 0.298 \ 0.282 \ 0.192 \ 0.178 \ 0.143 \ 0.129 \ 0.109 \ 0.097 \ 0.084] \}. \quad (74)$$

We repeated this procedure 10,000 times and found that the first eigenvalue of matrix C was larger than the first eigenvalue of all 10,000 permuted C^* matrices. From that pattern we can conclude that the first eigenvalue is highly significant (i.e., $p < 0.0001$). By contrast, all 10,000, but four, second eigenvalues turned out to be larger than the second eigenvalue of C (i.e., $p = 0.9996$) and this pattern strongly suggests that there is no cluster (or subpopulation or ‘segment’) in the sample of assessors used for this example.

$$\mathbf{a} = \begin{bmatrix} \mathbf{1}_{[1]} \times 0.104 \\ \mathbf{1}_{[2]} \times 0.104 \\ \mathbf{1}_{[3]} \times 0.092 \\ \mathbf{1}_{[4]} \times 0.104 \\ \mathbf{1}_{[5]} \times 0.081 \\ \mathbf{1}_{[6]} \times 0.099 \\ \mathbf{1}_{[7]} \times 0.094 \\ \mathbf{1}_{[8]} \times 0.089 \\ \mathbf{1}_{[9]} \times 0.111 \\ \mathbf{1}_{[10]} \times 0.124 \end{bmatrix} \text{ and}$$

Finding the Optimal Weights from the First Eigenvector of C

From \mathbf{u}_1 (see Eq. (71)), we find the optimum set of weights, denoted α , by rescaling \mathbf{u}_1 so that the sum of its elements is equal to 1 (cf., Eq. (15)):

$$\alpha = \mathbf{u}_1 \times (\mathbf{u}_1 \mathbf{1})^{-1} = \begin{bmatrix} 0.325 \\ 0.326 \\ 0.289 \\ 0.326 \\ 0.253 \\ 0.311 \\ 0.294 \\ 0.279 \\ 0.349 \\ 0.389 \end{bmatrix} \times 3.141^{-1} = \begin{bmatrix} 0.104 \\ 0.104 \\ 0.092 \\ 0.104 \\ 0.081 \\ 0.099 \\ 0.094 \\ 0.089 \\ 0.111 \\ 0.124 \end{bmatrix}. \quad (75)$$

$$\mathbf{A} = \text{diag} \{ \mathbf{a} \} = \text{diag} \left\{ \begin{bmatrix} \mathbf{1}_{[1]} \times 0.104 \\ \mathbf{1}_{[2]} \times 0.104 \\ \mathbf{1}_{[3]} \times 0.092 \\ \mathbf{1}_{[4]} \times 0.104 \\ \mathbf{1}_{[5]} \times 0.081 \\ \mathbf{1}_{[6]} \times 0.099 \\ \mathbf{1}_{[7]} \times 0.094 \\ \mathbf{1}_{[8]} \times 0.089 \\ \mathbf{1}_{[9]} \times 0.111 \\ \mathbf{1}_{[10]} \times 0.124 \end{bmatrix} \right\}, \quad (76)$$

where $\mathbf{1}_{[k]}$ is an $J_{[k]}$ vector of ones.

Generalized PCA of X

There are $K = 10$ values in α (see Eqs (75), (2) and (15)–(17)). These values are stored in the

The 12×12 diagonal mass matrix \mathbf{M} for the observations (with equal masses set to $m_i = \frac{1}{7} = 0.08$)

is given as

$$M = \text{diag}\{m\} = \text{diag}\left\{\frac{1}{12} \times 1\right\} = \text{diag}\left\{\begin{matrix} 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \\ 0.08 \end{matrix}\right\}. \quad (77)$$

The GSVD of X with matrices M (Eq. (77)) and A (Eq. (76)) gives $X = PAQ^T$. The eigenvalues (denoted λ) are equal to the squares of the singular values and are often used to informally decide upon the number of components to keep for further inspection. The eigenvalues and the percentage of the inertia that they explain are given in Table 2.

The pattern of the eigenvalue distribution suggests to keep two or three dimensions for future examination and we decided (somewhat arbitrarily) to keep only the first two dimensions for this example. So, in the remainder of this example we show only these first two dimensions.

The matrix Q of the right singular vectors (loadings for the variables) is given in Table 3, the matrix P of the right singular vectors and the matrix Δ of the singular values are given below:

$$P = \begin{bmatrix} -1.098 & 0.460 \\ -0.921 & 0.185 \\ -0.876 & -1.297 \\ -1.275 & -0.507 \\ 1.557 & -0.405 \\ 1.415 & -0.351 \\ 0.958 & 0.663 \\ 1.068 & 0.968 \\ -0.781 & 0.810 \\ 0.083 & -2.212 \\ -0.535 & 1.612 \\ 0.406 & 0.073 \end{bmatrix} \quad \Delta = \text{diag}\left\{\begin{bmatrix} 0.229 \\ 0.089 \end{bmatrix}\right\}. \quad (78)$$

Factor Scores

The factor scores for X show the best two-dimensional representation of the compromise of the K tables. Using Eqs (23), (25), and (78), we obtain:

$$F = PA = XAQ = \begin{bmatrix} -0.252 & 0.041 \\ -0.211 & 0.016 \\ -0.201 & -0.115 \\ -0.292 & -0.045 \\ 0.357 & -0.036 \\ 0.324 & -0.031 \\ 0.220 & 0.059 \\ 0.245 & 0.086 \\ -0.179 & 0.072 \\ 0.019 & -0.196 \\ -0.123 & 0.143 \\ 0.093 & 0.006 \end{bmatrix}. \quad (79)$$

In the F matrix, each row represents a wine and each column is a component. Figure 3a shows the wines in the space created by the first two components. The first component (with an eigenvalue equal to $\lambda_1 = 0.229^2 = 0.053$) explains 63% of the inertia, and contrasts the French and New Zealand wines. The second component (with an eigenvalue of $\lambda_2 = 0.089^2 = 0.008$) explains 9% of the inertia and is more delicate to interpret from the wines alone (its interpretation will become clearer after looking at the loadings).

Partial Factor Scores

The partial factor scores (which are the projection of a table onto the compromise) are computed from Eq. (27). For example, for the first assessor, the partial factor scores of the 12 wines are obtained as:

$$F_{[1]} = X_{[1]}Q_{[1]} = \begin{bmatrix} -0.275 & 0.044 \\ -0.312 & 0.173 \\ -0.057 & -0.028 \\ -0.250 & 0.125 \\ 0.409 & -0.201 \\ 0.311 & -0.217 \\ 0.185 & 0.048 \\ 0.266 & -0.033 \\ -0.244 & 0.126 \\ 0.050 & -0.266 \\ -0.170 & 0.203 \\ 0.086 & 0.026 \end{bmatrix}, \quad (80)$$

TABLE 2 | Eigenvalues and Percentage of Explained Inertia

	Component											
	1	2	3	4	5	6	7	8	9	10	11	12
Eigenvalue (λ)	0.053	0.008	0.006	0.004	0.004	0.002	0.002	0.002	0.001	0.001	0.001	0
cumulative	0.053	0.060	0.066	0.071	0.074	0.077	0.079	0.081	0.082	0.083	0.083	0.083
% Inertia (τ)	63	9	7	5	5	3	2	2	1	1	1	0
cumulative	63	72	79	84	89	92	94	96	97	98	100	100

TABLE 3 | α Weights, Loadings, and Squared Loadings

	Assessor 1					Assessor 2					Assessor 3																		
	V1	V2	V3	V4	V5	V6	V5	V4	V3	V2	V1	V2	V3	V4	V5	V6	V7	V8	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	
α -weights	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.092	
Loadings (Q)																													
Dim 1	-0.459	-0.418	-0.406	0.374	0.449	-0.363	-0.465	-0.464	-0.419	0.398	-0.373	-0.349	-0.475	-0.212	-0.403	0.315	-0.434	0.418											
Dim 2	-0.514	0.436	-0.674	0.385	-0.223	-0.149	-0.290	0.339	-0.323	0.441	0.073	0.507	-0.365	0.359	-0.547	0.609	0.453	-0.434											
Squared Loadings																													
Dim 1	0.210	0.175	0.165	0.140	0.201	0.132	0.216	0.215	0.176	0.159	0.139	0.122	0.226	0.045	0.162	0.099	0.189	0.174											
Dim 2	0.265	0.190	0.455	0.148	0.050	0.022	0.084	0.115	0.104	0.195	0.005	0.257	0.133	0.129	0.299	0.371	0.205	0.188											
	Assessor 4					Assessor 5					Assessor 6					Assessor 7													
	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8	V1	V2	V3	V4
α -weights	0.104	0.104	0.104	0.104	0.104	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.094	0.094	0.094	0.094
Loadings (Q)																													
Dim 1	-0.535	-0.446	-0.518	0.395	-0.349	-0.460	-0.333	-0.417	0.192	-0.403	0.273	-0.516	-0.478	-0.453	0.396	-0.351	-0.526	-0.473	-0.530	0.347									
Dim 2	-0.146	0.645	0.310	-0.082	0.103	-0.290	0.448	-0.414	-0.187	0.191	0.103	-0.356	0.505	-0.556	0.005	0.566	-0.718	0.815	-0.543	-0.643									
Squared Loadings																													
Dim 1	0.287	0.199	0.269	0.156	0.122	0.212	0.111	0.174	0.037	0.162	0.074	0.266	0.229	0.205	0.157	0.123	0.276	0.223	0.280	0.120									
Dim 2	0.021	0.416	0.096	0.007	0.011	0.084	0.200	0.171	0.035	0.036	0.011	0.126	0.256	0.309	0.000	0.320	0.515	0.664	0.295	0.413									
	Assessor 8					Assessor 9					Assessor 10																		
	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8	V1	V2	V3	V4	V8									
α -weights	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.111	0.111	0.111	0.111	0.111	0.124	0.124	0.124	0.124	0.124									
Loadings (Q)																													
Dim 1	-0.430	-0.388	-0.365	0.213	-0.447	0.373	-0.519	-0.403	-0.491	0.428	-0.508	-0.617	-0.527	-0.547	0.542														
Dim 2	-0.460	0.664	-0.426	0.433	0.401	-0.432	-0.424	0.464	-0.379	0.201	0.362	-0.169	0.581	-0.447	-0.528														
Squared Loadings																													
Dim 1	0.185	0.151	0.133	0.045	0.200	0.139	0.270	0.162	0.241	0.183	0.258	0.381	0.278	0.299	0.294														
Dim 2	0.211	0.440	0.181	0.188	0.161	0.186	0.179	0.216	0.143	0.041	0.131	0.028	0.337	0.200	0.279														

V1, Cat Pee; V2, Passion Fruit; V3, Green Pepper; V4, Mineral; V5, Smoky; V6, Citrus; V7, Tropical; V8, Leafy; V9, Grassy; V10, Flinty; V11, Vegetal; V12, Hay; V13, Melon; V14, Cut Grass; V15, Peach.

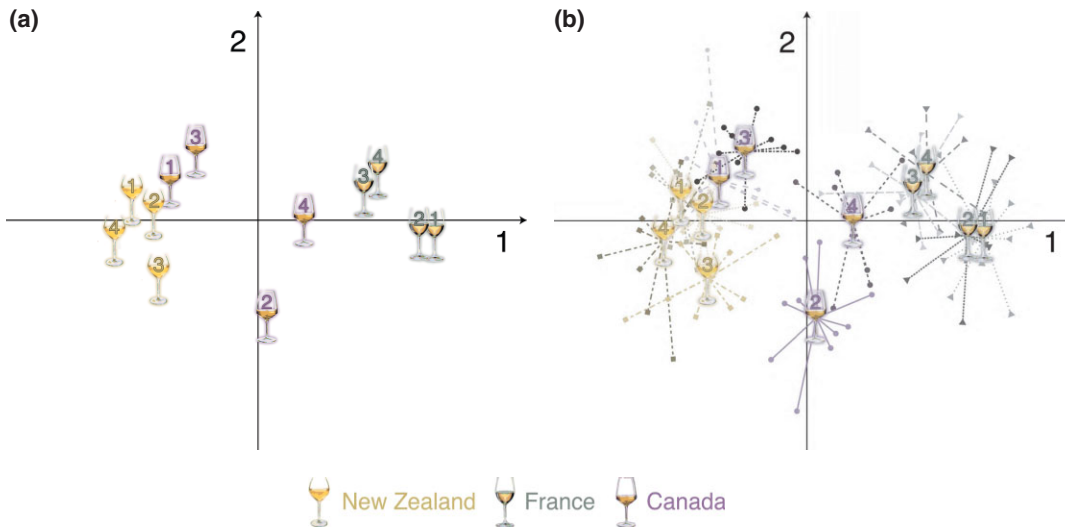


FIGURE 3 | Compromise of the 10 tables. (a) Factor scores (wines). (b) Assessors' partial factor scores projected into the compromise as supplementary elements. Each assessor is represented by a dot, and for each wine a line connects the wine factor scores to the partial factors scores of a given assessor for this wine. ($\lambda_1 = 0.053$, $\tau_1 = 63\%$; $\lambda_2 = 0.008$, $\tau_2 = 9\%$).

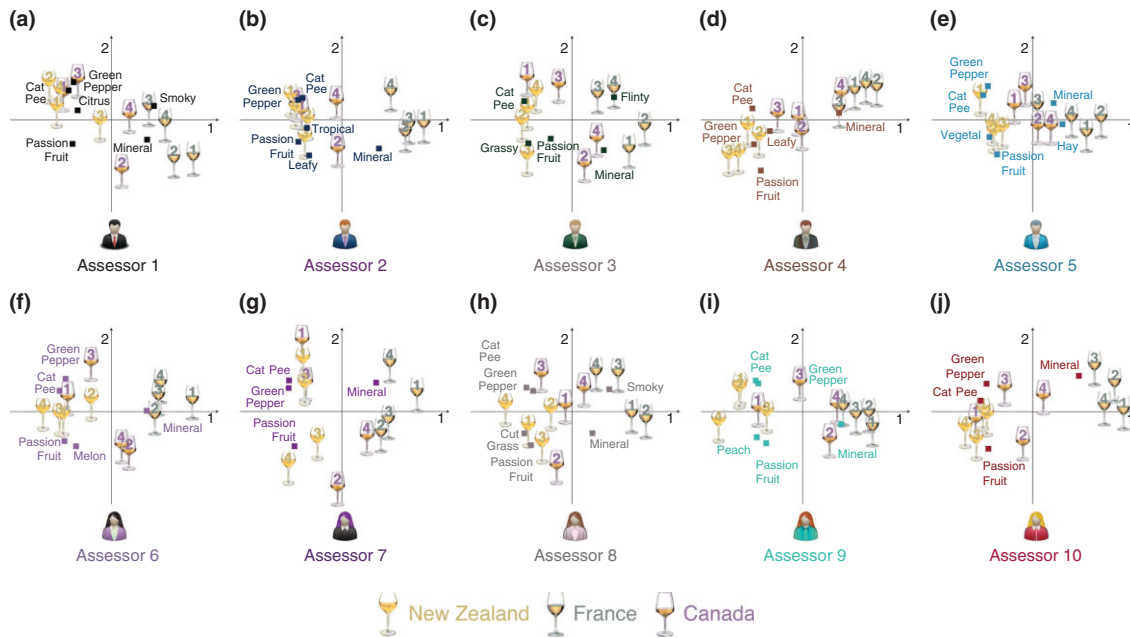


FIGURE 4 | Partial factor scores and variable loadings for the first two dimensions of the compromise space. The loadings have been re-scaled to have a variance equal the singular values of the compromise analysis.

and are displayed in Figure 4a. From Figure 4, we can see that, for all the assessors, Component 1 separates the New Zealand from the French Sauvignon Blancs, a pattern replicating the pattern seen in the compromise (Figure 3a). However, the assessors show a large inter-individual differences in how they rated the Canadian wines.

The original variables are analyzed, as in standard PCA, by computing loadings which are given in Table 3 for the first two dimensions. The loadings are also plotted in a biplot fashion in Figure 4. Here we show the partial factor scores (of the wines) along with the loadings for each assessor (which we have re-scaled so that their variance, for each dimension,

is equal to the singular value of the compromise). From these plots, we can see that for all the assessors, the New Zealand wines are rated as having a more cat-pee aroma, with some green pepper and passion fruit, while the French wines are rated as being more mineral, smoky, or hay-like.

Determining the Importance of the tables in the Compromise

There are two ways to determine which tables play the largest role in the compromise: contributions and partial inertias. The contributions of a table reflect the proportion of the variance of a dimension that can be attributed to this table (see Eq. (33)). The larger the contribution of a table to a component, the more important this table is for this component. The contributions for the tables are:

$$ctr_{k,l} = \begin{bmatrix} 0.106 & 0.117 \\ 0.106 & 0.079 \\ 0.082 & 0.122 \\ 0.107 & 0.057 \\ 0.062 & 0.043 \\ 0.097 & 0.100 \\ 0.084 & 0.177 \\ 0.076 & 0.121 \\ 0.124 & 0.079 \\ 0.155 & 0.105 \end{bmatrix} \quad (81)$$

The contributions can also be plotted to obtain a visual representation of the importance of the studies. Figure 5 shows the relative contributions of each of the tables to Components 1 and 2. From this figure, we can see that Assessor 10 contributes the most

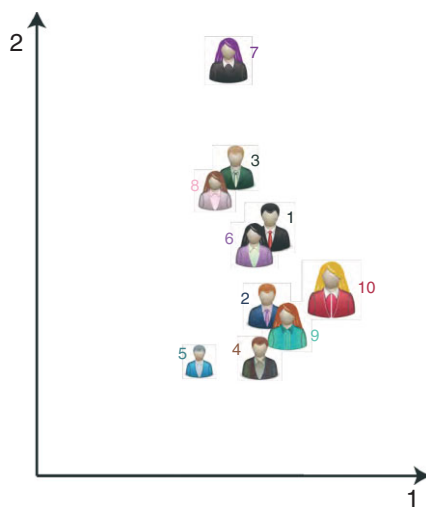


FIGURE 5 | Contributions of the tables to the compromise. The sizes of the assessors' icons are proportional to their contribution to components 1 and 2.

to the first component of the compromise, while Assessor 7 contributes most to the second component. Assessor 5, by contrast, contributes the least to both Components 1 and 2.

In addition to using contributions, we can determine a table's importance with the partial inertia that gives the proportion of the compromise variance (i.e., inertia) explained by the table. This is obtained by multiplying, for each dimension, the table contribution by the eigenvalue of the dimension. For our example, the partial inertias denoted $\mathcal{I}_{\text{partial}}$ are:

$$\mathcal{I}_{\text{partial}} = \begin{bmatrix} 0.0056 & 0.0009 \\ 0.0056 & 0.0006 \\ 0.0043 & 0.0010 \\ 0.0056 & 0.0004 \\ 0.0033 & 0.0003 \\ 0.0051 & 0.0008 \\ 0.0044 & 0.0014 \\ 0.0040 & 0.0010 \\ 0.0065 & 0.0006 \\ 0.0082 & 0.0008 \end{bmatrix} \quad (82)$$

Like the contributions, the partial inertia can be plotted to get a visual representation. From Figure 6, we see that Assessor 10 accounts for the most inertia on the first dimension, while Assessor 5 accounts for the lowest proportion of the total variance.

Supplementary Table

Because we would like to know what qualities of the wines are associated with the assessors' ratings, we wanted to include some chemical components of the wines as variables, namely, titratable acidity, pH, alcohol, and residual sugar. The values for these variables are shown in Table 4. However, because these properties are qualitatively different from the assessors' ratings (i.e., did not come from the same population), we did not want to include them as active elements in the analysis and therefore projected them as a supplementary table.

To obtain the factor scores, we first compute the supplementary loadings, which are obtained as

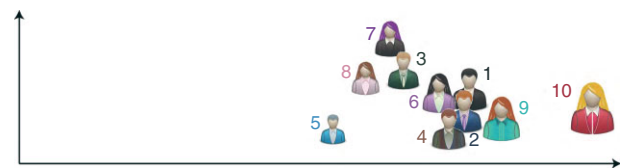


FIGURE 6 | Partial inertias of the studies. The sizes of the assessors' icons are proportional to their explained inertia for Components 1 and 2.

TABLE 4 | Supplementary Table: Chemical Properties of the Wines

	Chemical Properties			
	Titrateable Acidity	pH	Alcohol	Residual Sugar
NZ ₁	5.60	3.38	14.00	3.00
NZ ₂	5.30	3.53	13.50	3.60
NZ ₃	6.20	3.27	14.00	3.00
NZ ₄	8.50	3.19	13.50	3.90
F ₁	5.00	3.60	12.50	1.50
F ₂	5.88	3.00	12.50	2.00
F ₃	4.50	3.33	13.00	0.80
F ₄	5.60	3.40	12.00	2.10
CA ₁	7.60	3.30	13.00	2.80
CA ₂	5.70	3.43	13.50	2.10
CA ₃	6.20	3.30	12.50	2.50
CA ₄	6.90	2.20	13.00	2.00

(cf., Eq. (50)):

$$Q_{sup} = X_{sup}^T M P \Delta^{-1} = \begin{bmatrix} -0.337 & 0.054 \\ -0.059 & 0.036 \\ -0.460 & 0.782 \\ -0.537 & 0.124 \end{bmatrix}. \quad (83)$$

From these loadings, we can see that, on Component 1, the New Zealand Sauvignon Blancs

(cf., Eq. (51)):

$$F_{sup} = X_{sup} Q_{sup} = X_{sup} X_{sup}^T M P \Delta^{-1} = \begin{bmatrix} -0.133 & -0.181 \\ -0.124 & -0.101 \\ -0.158 & -0.184 \\ -0.289 & -0.125 \\ 0.191 & 0.131 \\ 0.120 & 0.123 \\ 0.231 & 0.061 \\ 0.168 & 0.210 \\ -0.094 & -0.004 \\ -0.001 & -0.070 \\ 0.052 & 0.106 \\ 0.038 & 0.035 \end{bmatrix}. \quad (84)$$

In a manner similar to the biplot approach we used for the assessors (see Figure 4) we plotted together, in a biplot way, the supplementary partial factor scores and loadings for the chemical properties of the wines (see Figure 7). This figure confirms the interpretation that we reached from the numerical values.

Inner Product Map Projection

To evaluate the similarity between the supplementary and the active tables, we first compute the cross-product matrix associated to X_{sup} (cf., Eq. (52)):

$$S_{sup} = X_{sup} X_{sup}^T = \begin{bmatrix} 0.06 & 0.05 & 0.06 & 0.02 & -0.03 & -0.04 & -0.02 & -0.05 & -0.01 & 0.02 & -0.03 & -0.04 \\ 0.05 & 0.07 & 0.04 & 0.02 & -0.01 & -0.04 & -0.03 & -0.02 & -0.01 & 0.01 & -0.01 & -0.08 \\ 0.06 & 0.04 & 0.06 & 0.05 & -0.05 & -0.04 & -0.03 & -0.06 & 0.01 & 0.02 & -0.03 & -0.01 \\ 0.02 & 0.02 & 0.05 & 0.18 & -0.11 & -0.04 & -0.14 & -0.06 & 0.08 & -0.02 & -0.01 & 0.02 \\ -0.03 & -0.01 & -0.05 & -0.11 & 0.09 & 0.02 & 0.08 & 0.06 & -0.03 & 0.01 & 0.02 & -0.06 \\ -0.04 & -0.04 & -0.04 & -0.04 & 0.02 & 0.04 & 0.03 & 0.04 & -0.01 & -0.02 & 0.02 & 0.05 \\ -0.02 & -0.03 & -0.03 & -0.14 & 0.08 & 0.03 & 0.13 & 0.04 & -0.06 & 0.03 & -0.00 & -0.02 \\ -0.05 & -0.02 & -0.06 & -0.06 & 0.06 & 0.04 & 0.04 & 0.08 & -0.01 & -0.01 & 0.04 & -0.03 \\ -0.01 & -0.01 & 0.01 & 0.08 & -0.03 & -0.01 & -0.06 & -0.01 & 0.05 & -0.01 & 0.01 & 0.01 \\ 0.02 & 0.01 & 0.02 & -0.02 & 0.01 & -0.02 & 0.03 & -0.01 & -0.01 & 0.02 & -0.01 & -0.04 \\ -0.03 & -0.01 & -0.03 & -0.01 & 0.02 & 0.02 & -0.00 & 0.04 & 0.01 & -0.01 & 0.02 & -0.01 \\ -0.04 & -0.08 & -0.01 & 0.02 & -0.06 & 0.05 & -0.02 & -0.03 & 0.01 & -0.04 & -0.01 & 0.21 \end{bmatrix}. \quad (85)$$

are more acidic, have a greater alcohol content, and more residual sugar than the French wines.

Next, we compute the supplementary factor scores for the first two components as

Then, we compute the K inner products between S_{sup} and each of the $S_{[k]}$ matrices (cf., Eq. (8)):

$$c_{sup,k} = \langle S_{sup}, S_{[k]} \rangle = \text{trace} \{ S_{sup} S_{[k]} \}, \quad (86)$$

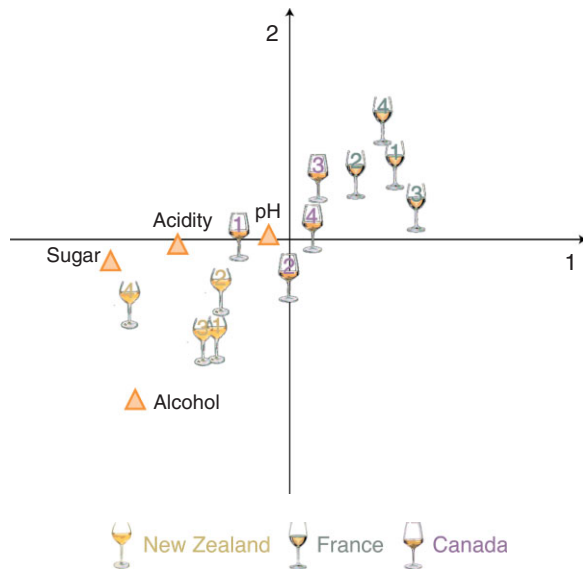


FIGURE 7 | Supplementary table: Chemical components of the wines. Supplementary partial scores and loadings. (cf., Figure 3a).

which gives:

$$\mathbf{c}_{\text{sup}} = [0.247 \quad 0.269 \quad 0.233 \quad 0.281 \quad 0.205 \quad 0.297 \quad 0.254 \quad 0.230 \quad 0.317 \quad 0.326]. \quad (87)$$

The factor scores of the supplementary table in the inner product map are obtained as (cf., Eqs (54) and (55)):

$$\begin{aligned} \mathbf{g}_{\text{sup}} &= \mathbf{c}_{\text{sup}} \mathbf{U} \mathbf{\Gamma}^{\frac{1}{2}} \\ &= [0.417 \quad -0.006 \quad -0.008 \quad 0.136 \quad -0.038 \quad -0.031 \quad -0.051 \quad 0.013 \quad -0.038 \quad 0.058]. \end{aligned} \quad (88)$$

When \mathbf{g}_{sup} is projected onto the inner product map (Figure 8), we can see that the chemical properties of the wines are most similar to the ratings of Assessors 1, 8, 9, and 10. This suggests that for the sample of Sauvignon Blanc wines, these four assessors may be quite sensitive to the chemical properties of the wines.

$$\begin{aligned} \boldsymbol{\alpha}^* &= \frac{1}{I} \sum_i^I \boldsymbol{\alpha}_i^* = \boldsymbol{\alpha} - \frac{I-1}{I} \sum_i^I \boldsymbol{\alpha}_{-i} \\ &= [0.105 \quad 0.105 \quad 0.089 \quad 0.103 \quad 0.076 \quad 0.100 \quad 0.092 \quad 0.087 \quad 0.113 \quad 0.129]. \end{aligned} \quad (90)$$

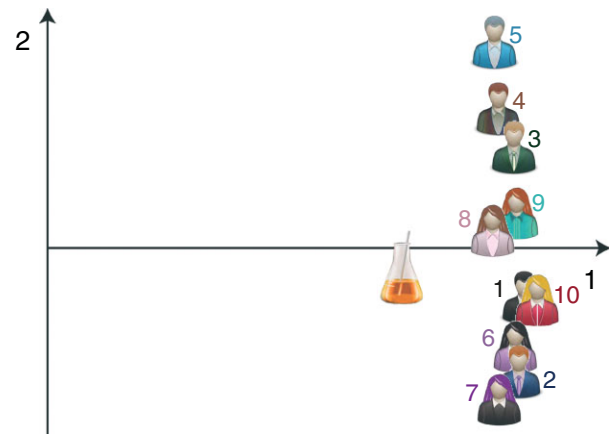


FIGURE 8 | Projection of the supplementary table into the inner product map.

To get the α weight of the supplementary table, we compute α_{sup} (from Eqs (56), (71), and (75)) as:

$$\begin{aligned} \alpha_{\text{sup}} &= \gamma_1^{-\frac{1}{2}} \times g_{\text{sup},1} \times (\mathbf{u}_1^T \mathbf{1})^{-1} = 4.135^{-\frac{1}{2}} \times 0.417 \\ &\times 3.141^{-1} = \frac{0.417}{2.034 \times 3.141} = 0.065. \end{aligned} \quad (89)$$

Inferential Steps

Jackknife

To estimate the reliability of the α -weights, we use the jackknife procedure as described in Eqs 57–60.

This gives the following estimates of the elements of the vector $\boldsymbol{\alpha}$:

Then, the standard error of the vector α_i^* (cf., Eq. (60)) is obtained as:

$$\hat{\sigma}_{\alpha^*}^2 = [0.016 \quad 0.014 \quad 0.016 \quad 0.023 \quad 0.017 \quad 0.019 \quad 0.016 \quad 0.018 \quad 0.016 \quad 0.010]. \quad (91)$$

Finally, using a critical value of 2 (for an approximate value of $p < 0.05$), we find, according to the jackknife estimator, the following 95% confidence intervals for the α -weights for each table:

$$\begin{bmatrix} 0.137 & 0.134 & 0.121 & 0.150 & 0.109 & 0.138 & 0.124 & 0.122 & 0.146 & 0.149 \\ 0.073 & 0.077 & 0.057 & 0.057 & 0.043 & 0.061 & 0.059 & 0.052 & 0.081 & 0.109 \end{bmatrix}. \quad (92)$$

With one (slight) exception, the value of 0.100 is contained within all confidence intervals. This value corresponds to an equal weighting scheme of $\alpha_i = \frac{1}{7}$. This suggests that the STATIS solution differs little from a simple multitable approach. Also, only the 10th assessor can be considered as having a ‘commonality’ higher than the group average.

Bootstrap

To estimate the stability of the compromise factor scores, we used a bootstrap approach. We generated 1000 bootstrap samples that gave 1000 estimated bootstrapped factor scores. For example, for the first bootstrapped sample, we sampled with replacement in the set of the integers from 1 to 10 and obtained the following bootstrap set

$$B = \{6 \quad 4 \quad 4 \quad 2 \quad 9 \quad 3 \quad 1 \quad 1 \quad 2 \quad 8\}. \quad (93)$$

We then computed the C_1^* matrix using the following inner product matrices:

$$\{S_{[6]} \quad S_{[4]} \quad S_{[4]} \quad S_{[2]} \quad S_{[9]} \quad S_{[3]} \quad S_{[1]} \quad S_{[1]} \quad S_{[2]} \quad S_{[8]}\}. \quad (94)$$

From the eigendecomposition of C_1^* , we obtained the following bootstrapped values for the α weights:

$$\alpha_1^* = [0.096 \quad 0.103 \quad 0.103 \quad 0.103 \quad 0.109 \quad 0.090 \quad 0.102 \quad 0.102 \quad 0.103 \quad 0.088]. \quad (95)$$

With these weights, we then computed the bootstrapped estimate of the factor scores as

$$\begin{aligned} F_1^* &= \sum_{k \in B} F_{[k]} = 0.096 \times F_{[6]} + 0.103 \times F_{[4]} \\ &+ 0.103 \times F_{[4]} + 0.103 \times F_{[2]} + 0.109 \times F_{[9]} \\ &+ 0.090 \times F_{[3]} + 0.102 \times F_{[1]} + 0.102 \times F_{[1]} \\ &+ 0.103 \times F_{[2]} + 0.088 \times F_{[8]} \\ &= \begin{bmatrix} -0.262 & -0.020 \\ -0.219 & -0.022 \\ -0.203 & 0.108 \\ -0.289 & 0.005 \\ 0.352 & 0.045 \\ 0.328 & 0.050 \\ 0.228 & -0.035 \\ 0.246 & -0.091 \\ -0.163 & -0.055 \\ 0.027 & 0.185 \\ -0.137 & -0.133 \\ 0.092 & -0.038 \end{bmatrix}. \quad (96) \end{aligned}$$

From the bootstrap estimates, we can also compute bootstrap ratios, which, like t statistics, can be used to find the observations that reliably contribute to a given component. To get the bootstrap ratios, we first computed the mean of the bootstrap

samples which are equal to (see Eq. (62))

$$\bar{F}^* = \frac{1}{L} \sum_{\ell} F_{\ell}^* = \begin{bmatrix} -0.252 & -0.041 \\ -0.211 & -0.016 \\ -0.201 & 0.116 \\ -0.292 & 0.045 \\ 0.356 & 0.036 \\ 0.325 & 0.031 \\ 0.219 & -0.060 \\ 0.245 & -0.086 \\ -0.179 & -0.073 \\ 0.019 & 0.196 \\ -0.122 & -0.143 \\ 0.094 & -0.006 \end{bmatrix}, \quad (97)$$

and the standard deviations as (see Eq. (63)):

$$\hat{\sigma}_{F^*} = \begin{bmatrix} 0.010 & 0.029 \\ 0.023 & 0.028 \\ 0.030 & 0.022 \\ 0.023 & 0.037 \\ 0.022 & 0.032 \\ 0.022 & 0.037 \\ 0.026 & 0.031 \\ 0.022 & 0.028 \\ 0.030 & 0.036 \\ 0.021 & 0.030 \\ 0.021 & 0.018 \\ 0.021 & 0.033 \end{bmatrix}. \quad (98)$$

Then, the bootstrap ratios are computed by dividing the bootstrap mean by its standard deviation. This gives the following bootstrap ratios:

$$\text{Boot}_{\text{ratio}} = \begin{bmatrix} -24.204 & -1.385 \\ -9.258 & -0.561 \\ -6.820 & 5.347 \\ -12.721 & 1.208 \\ 16.041 & 1.121 \\ 15.079 & 0.833 \\ 8.328 & -1.951 \\ 11.155 & -3.084 \\ -6.035 & -2.004 \\ 0.898 & 6.478 \\ -5.955 & -7.769 \\ 4.510 & -0.172 \end{bmatrix}. \quad (99)$$

These bootstrap ratios can be plotted as a bar chart to give a visual representation of which observations most reliably contribute to a component (see Figure 9). To be conservative and to take into account the multiple comparisons problem, we chose a bootstrap ratio critical value of ± 3 . This value correspond roughly to a Bonferroni corrected p -value for $J = 53$ comparisons (i.e., p -corrected is equal to $\frac{0.05}{53} \approx 0.001$ which approximately corresponds to a t value of 3).

From the bootstrap ratios shown in Figure 9 (see also Figure 10), we can see that all of the wines—except the second Canadian Sauvignon Blanc—contribute reliably to the first component, with the New Zealand wines separated from the French wines. This confirms our previous interpretation of Figure 3a. However, only four wines contribute reliably to the second component, with the third New Zealand Sauvignon Blanc grouping with the second Canadian wine and the fourth French wine grouping together with the third Canadian wine.

We also used the set of bootstrapped factors scores to obtain the 95% confidence intervals (CI) around the factor scores (see Figure 10). Here, around each wine, we fitted an ellipsoid that comprises 95% of the bootstrapped factor scores. This ellipsoid represents the possible positions of a wine for replications of the analysis (assuming that the assessors were randomly sampled from a population of assessors). When the ellipses of two wines do not overlap, these two wines can be considered as reliably differentiated by the assessors.

RECENT DEVELOPMENTS AND VARIATIONS OVER STATIS

Recently, several papers have been dedicated to exploring the developments of various aspects of the STATIS approach. We briefly review some of the most relevant ones.

X-STATIS or Partial Triadic Analysis

Partial triadic analysis (PTA), also called X-STATIS, was first named and described by Jaffrenou.⁹⁶ This simple variation over STATIS can be used when all the $X_{[k]}$ tables are measuring the same variables for the same observations^{32,33,97–101} and can be seen also as a particular simplified case of Tucker three-mode factor analysis,^{102,103} hence the name partial (for simplified) triadic (for three mode) analysis (the X-STATIS name is also used because it is a STATIS analysis on the $X_{[k]}$ matrices, see below). The procedure for PTA mirrors the STATIS procedure with two small differences: (1) the inner product matrix used for the computation of the α weights is obtained from the $X_{[k]}$ matrices rather than from the $S_{[k]}$ matrices; and (2) the compromise is obtained as the weighted average of the $X_{[k]}$ matrices (instead of the weighted average of the $S_{[k]}$ matrices).

In PTA, the alpha weights are obtained from the eigendecomposition of the inner product matrix

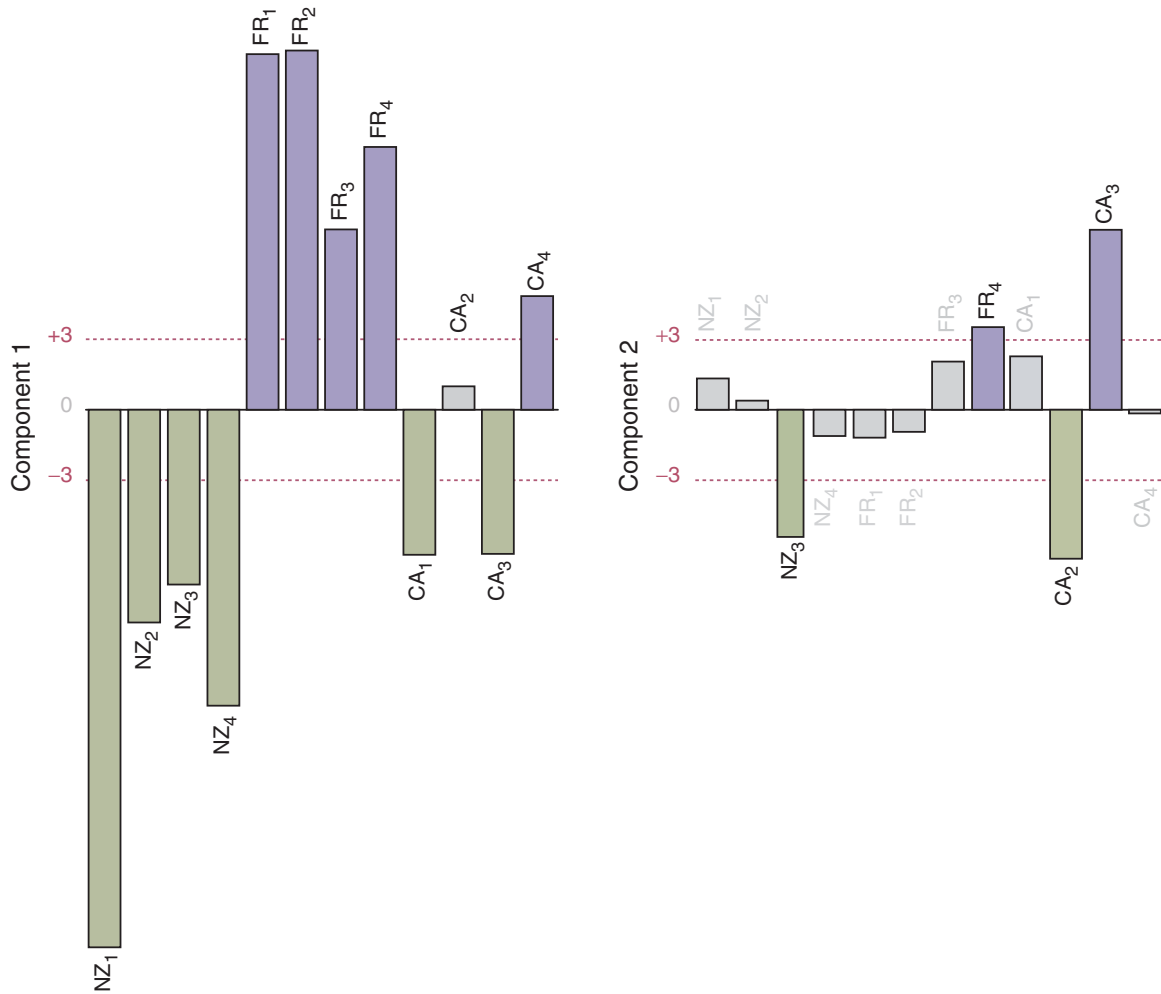


FIGURE 9 | Bootstrap ratio plot for Components 1 and 2.

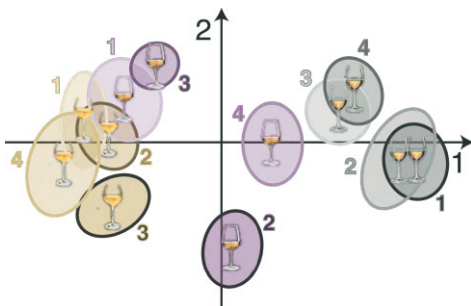


FIGURE 10 | Bootstrap confidence ellipses plotted on Components 1 and 2.

C' whose general term is obtained as (compare with Eq. (8))

$$c'_{k,\ell} = \text{vec} \{X_{[k]}\}^T \times \text{vec} \{X_{[\ell]}\} \quad (100)$$

(note that, contrary to standard STATIS, the terms $c'_{k,\ell}$ are not guaranteed to be non-negative but PTA

requires them to be non-negative; a negative $c'_{k,\ell}$, indicates that, e.g. STATIS should be used rather than PTA, but also that the reason of a negative value needs to be explored). The first eigenvector for C' , gives after re-scaling (see Eq. (15)), α' , the vector of the optimum weights. These weights are then used to compute the compromise denoted $X_{[+]}$ which is obtained as the weighted average of the $X_{[k]}$ matrices (instead of the weighted average of the $S_{[k]}$ matrices). Specifically, $X_{[+]}$ is computed as

$$X_{[+]} = \sum_k^K \alpha'_k X_{[k]}. \quad (101)$$

The SVD of $X_{[+]}$ provides factors scores for the observations and loadings for the variables. Note that the loadings are computed for the *common* variables, (i.e., not for the K sets of variables). The loadings for the variables of an $X_{[k]}$ table can be recovered by projecting this table as a supplementary table (from Eq. (50)).

PTA maximizes a criterion akin to (but different from) the criterion of STATIS. Specifically, PTA maximizes: (see Ref 100, compare with Eq. (66))

$$S_{pta} = \sum_k^K \left(\text{vec} \{ \alpha'_k X_{[k]} \}^T \times \text{vec} \{ X_{[+]} \} \right)^2 \quad \text{with}$$

$$\alpha'^T \alpha' = 1. \quad (102)$$

Recent models called STATICO^{32,33,98} and COSTATIS³³ generalizes PTA to two sets of tables by combining co-inertia analysis (which relates one set of tables to the other set) with PTA (which integrates the information within one set of tables). These models are closely related to double-STATIS which is described later in this article. These models can also be seen as a generalization of Tucker inter battery analysis,¹⁰⁴ partial least square correlation,^{105,106} and STATIS.

Example X-STATIS

As an illustration, the results of a PTA on the four variables common to all the wines are shown in Figure 11. Here the results are very close to those of the STATIS analysis with these four variables (not shown here, but the two figures would be almost identical). In practice, this similarity of the results between PTA and STATIS is often observed, but the computational load of PTA is smaller (a feature of interest for large data sets such as brain imaging or ‘-omics’ data). In PTA, the K tables are often considered as i.i.d. observations and, in this case, a bootstrap resampling scheme on the K matrices $X_{[k]}$ can be used to estimate the variability (and derive confidence interval, and bootstrap ratios) of factor scores and loadings. Here, because the variables are put in correspondence, the variability of the estimates is likely to be smaller for PTA than in plain STATIS because of the larger number of measurements per variable. However, in PTA as in STATIS, when using the bootstrap to estimate the variability of the factor scores or the loadings, the estimation of the α vector needs to be performed for each bootstrap sample and this can add to the computational load of the procedure.

COVSTATIS and DISTATIS

COVSTATIS and DISTATIS^{21,27,34,42} are straightforward adaptations of the STATIS methodology to the problem of integrating, respectively, covariance (or correlation), and distance matrices. Here the raw data consists of K covariance or distance matrices collected on the same observations. As such, COVSTATIS and DISTATIS can be considered as three-way extensions of metric multidimensional scaling (see e.g., Ref 107 for an earlier closely related approach).

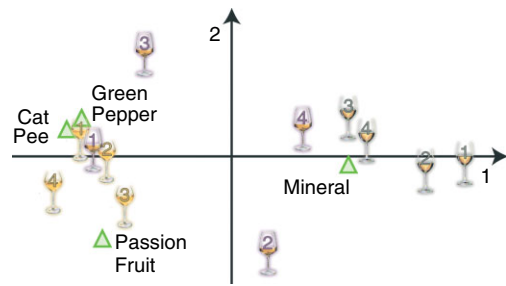


FIGURE 11 | X-STATIS or partial triadic analysis. Biplot of the wines and the four common variables.

COVSTATIS

COVSTATIS is used to analyze a set of covariance (or correlation) matrices. When dealing with a set of covariance matrices, the problem is similar to the cross-product approach for STATIS (we have covariance matrices instead of cross-products), and a first idea could be to use directly the covariance or correlation matrices in lieu of the $S_{[k]}$ in Eqs (34)–(40). However, when dealing with covariance matrices, the problem of having different units for the different covariance matrices needs to be eliminated by some normalization scheme (e.g., by ensuring that all the covariance matrices have a norm of 1, or have a first eigenvalue equal to 1).

An additional problem remains, however, because each covariance (or correlation) matrix can be interpreted as a space in which points are scattered from their origin, but each matrix has its own origin (see Ref 108 for a detailed account of this problem and also Refs 20, 21, 34). Covariance matrices with different origins cannot be meaningfully compared for the same reason that variables with different means cannot be compared. In order to guarantee that all covariance matrices have a common origin, each matrix is *double centered* (this is equivalent to subtracting the mean for simple variables). This is obtained by first computing an I by I centering matrix, denoted Ξ , and defined as

$$\Xi = (I - \mathbf{1m}^T). \quad (103)$$

So, if $R_{[k]}$ is a (normalized) covariance or correlation matrix, the corresponding double centered cross-product matrix (i.e., $S_{[k]}$) will be obtained as

$$S_{[k]} = \frac{1}{2} \Xi R_{[k]} \Xi. \quad (104)$$

These $S_{[k]}$ matrices are then used as in the standard cross-product approach for the STATIS procedure.²¹

With COVSTATIS, the distances on the map between observations are approximately inversely

proportional to their covariance or correlation in the original data matrices.

DISTATIS

DISTATIS is used when we have K distance matrices collected on the same set of I observations. The main idea behind DISTATIS is to transform the distance matrices into cross-product matrices and then use the cross-product approach to STATIS. This idea introduced in Ref 21 was developed later in Refs 27, 34, 42 and subsequently used in different domains such as sensory evaluation,^{27–29,109–112} brain imaging,^{42,113} survey analysis,³ computer science¹¹⁴ or molecular biology.⁴⁶

Distance matrices cannot be directly analyzed with STATIS because distance, having a diagonal of zero values, are not positive semidefinite (i.e., are not cross-product matrices). To alleviate this problem, a traditional approach is to first transform the distance matrices into cross-product matrices. This transformation (originally proposed by Young and Householder,¹¹⁵ see also Ref 116), constitutes the basis of metric multidimensional scaling,¹¹⁷ and is particularly appropriate for squared Euclidean distance matrices because this transformation is completely reversible in the sense that a squared Euclidean distance matrix can be perfectly reconstituted from its corresponding cross-product matrix and vice-versa.¹¹⁵ When this transformation is applied to a non-Euclidean distance, it gives a symmetric matrix that has both positive and negative eigenvalues (the larger the proportion of negative eigenvalues, the more ‘non-Euclidean’ the distance). For a non-Euclidean matrix, a cross-product matrix can be obtained by keeping only the eigenvectors associated with positive eigenvalues and then reconstituting the cross-product matrix with these eigenvectors and eigenvalues (transforming this cross-product matrix back into a distance matrix will provide a squared Euclidean distance matrix).

In order to transform a distance matrix into its corresponding cross-product matrix, we use a double centering similar to the one described by Eq. 104. Specifically a squared distance matrix $\mathbf{D}_{[k]}$ is transformed into a cross-product matrix $\mathbf{R}_{[k]}$ by the following transformation called ‘double-centering’ and computed as (compare to Eq. (104), note the minus sign in front of the right term of the equation):

$$\mathbf{R}_{[k]} = -\frac{1}{2} \mathbf{\Xi} \mathbf{D}_{[k]} \mathbf{\Xi}. \quad (105)$$

The $\mathbf{R}_{[k]}$ matrices are often normalized prior to the analysis such that, for example, the sum of their squared elements is equal to one (à la SUM-PCA) or

that they have a first eigenvalue equal to one (à la MFA).

The normalized cross-product matrices are denoted $\mathbf{S}_{[k]}$ and are then used as the entries for the cross-product approach of STATIS. Detailed computational examples of DISTATIS can be found in Refs 27, 28 and a short example can be found in this article in the section on canonical-STATIS.

With DISTATIS, the distances on the map between two observations are approximately proportional to their distances in the original data matrices.

So, in summary, with DISTATIS, we start with K distance matrices $\mathbf{D}_{[k]}$ of dimensions I by I . These distance matrices are then transformed into K cross-product matrices $\mathbf{S}_{[k]}$ which are then plugged into the cross-product approach of STATIS.

CANOSTATIS: Canonical STATIS

Recently, a new method called canonical-STATIS (CANOSTATIS¹¹⁸) has been proposed that integrates features of DISTATIS and discriminant analysis (as well as multiblock barycentric discriminant analysis^{5,119}). In this framework, the data consist in K tables, denoted $\mathbf{X}_{[k]}$ each of dimensions I by $J_{[k]}$ collecting measurements on the same I observations (the $\mathbf{X}_{[k]}$ matrices are supposed to have been adequately normalized). As an additional requirement of CANOSTATIS, each of the $\mathbf{X}_{[k]}$ matrix is required to be full-rank on its columns (which implies that $J_{[k]} < I$ for all K , this requirement comes from discriminant analysis which is part of CANOSTATIS, see below for more explanations). For convenience, the columns of $\mathbf{X}_{[k]}$ are supposed to be centered and normalized (and the whole matrix itself can also be normalized).

In CANOSTATIS, the I observations come from N *a-priori* known groups. The problem, here, is to explore the relationship between these groups in order for example to assess if these groups are differentiated and also to assess if the observations can reliably (or not) be assigned to their respective groups.

If the data were structured into only one table (and if the data matrix were full rank on its columns), the problem would correspond to a classic problem of discriminant analysis.¹²⁰ An equivalent way of performing discriminant analysis is to compute the so-called Mahalanobis distance⁷⁷ between the N groups and to compute a multidimensional scaling¹¹⁷ analysis on this distance matrix followed by a projection as supplementary elements of the I observations. Canonical-STATIS generalizes this discriminant analysis approach by computing for each of the K tables an N by N Mahalanobis distance matrix (see, Ref 77 for a

detailed example of this computation). In order to do so, the first step is to compute, for each table two matrices: (1) the N by $J_{[k]}$ matrix of the groups means denoted $\bar{\mathbf{X}}_{[k]}$ (i.e., an entry of $\bar{\mathbf{X}}_{[k]}$ collects the mean of one column of $\mathbf{X}_{[k]}$ for one of the N groups) and (2) the $J_{[k]}$ by $J_{[k]}$ matrix denoted $\hat{\Sigma}_{[k]}$ of the covariances between the variables of $\mathbf{X}_{[k]}$ [i.e., the (j, j') entry of $\hat{\Sigma}_{[k]}$ contains the covariance between the j th and the (j') th variables of $\mathbf{X}_{[k]}$]. The computation of the Mahalanobis distance matrix requires the computation of the inverse of the covariance matrix. This inverse is denoted $\hat{\Sigma}_{[k]}^{-1}$, and can be computed only when $\hat{\Sigma}_{[k]}$ is full rank (hence the requirement that the $\mathbf{X}_{[k]}$ matrices should be full rank on their columns, because these two statements are equivalent). With these notations, the distance matrix for the k th group is denoted $\mathbf{D}_{[k]}$ and is computed as (see next section for an example of this computation)

$$\mathbf{D}_{[k]} = \text{diag} \left\{ \bar{\mathbf{X}}_{[k]} \hat{\Sigma}_{[k]}^{-1} \bar{\mathbf{X}}_{[k]}^T \right\} \times \mathbf{1}^T + \mathbf{1} \times \text{diag} \left\{ \bar{\mathbf{X}}_{[k]} \hat{\Sigma}_{[k]}^{-1} \bar{\mathbf{X}}_{[k]}^T \right\}^T - 2 \times \bar{\mathbf{X}}_{[k]} \hat{\Sigma}_{[k]}^{-1} \bar{\mathbf{X}}_{[k]}^T. \quad (106)$$

These K distance matrices are then used as the input for a DISTATIS analysis. This provides an inner product map that shows how the tables relate to each other in terms of their discrimination and a compromise that gives the best common discrimination pattern. On this map, the discrimination pattern of each table can also be projected. As with DISTATIS (and in metric multidimensional scaling^{42,117,121}), the original observations can also be projected

onto the group map to assess the quality of the discrimination.

The constraint that each $\mathbf{X}_{[k]}$ matrix be full rank is necessary only if the (full) inverse of $\hat{\Sigma}_{[k]}$ is computed. Using a pseudo-inverse will provide the same distance matrix (cf., Ref 122, Lemma 2.2.4.iii) and does not require the full rank assumption. Alternatively, when an $\mathbf{X}_{[k]}$ is not full rank, some form of regularization or ridge could also work (see Ref 137). These approaches are likely to constitute future developments of CANOSTATIS.

So, in brief, in CANOSTATIS we have groups of observations described by several tables. For each table we compute a Mahalanobis distance matrix—a step which is equivalent to performing a linear discriminant analysis. We then use these distance matrices as the input to a DISTATIS analysis in order to integrate them and represent them as a compromise map.

Example Canonical-STATIS

As an illustration, we use the same data as in the STATIS example, with the $\mathbf{X}_{[k]}$ matrices being now the $I = 12$ by $J_{[k]}$ tables and the countries of origins of the wines as defining the groups of wines to be classified. We have $N = 3$ groups corresponding to the three countries of origin of the wines (i.e., New Zealand, France, and Canada). The goal of the analysis is to evaluate if the assessors differentiate the wines according to the country of origin.

If we look at the first assessor, the data matrix $\mathbf{X}_{[1]}$ is given by Eq. (68). The $N = 3$ by $J_{[1]} = 6$ matrix of means $\bar{\mathbf{X}}_{[1]}$ is equal to

$$\bar{\mathbf{X}}_{[1]} = \begin{bmatrix} 0.0929 & 0.0922 & 0.0859 & -0.0609 & -0.1109 & 0.095 \\ -0.1161 & -0.1252 & -0.1039 & 0.1218 & 0.1524 & -0.0856 \\ 0.0232 & 0.0329 & 0.0181 & -0.0609 & -0.0416 & -0.0101 \end{bmatrix}. \quad (107)$$

The variance/covariance matrix is equal to

$$\hat{\Sigma}_{[1]} = \begin{bmatrix} 0.0085 & 0.0010 & 0.0078 & -0.0040 & 0.0007 & 0.0035 \\ 0.0010 & 0.0073 & -0.0008 & 0.0020 & -0.0010 & 0.0023 \\ 0.0078 & -0.0008 & 0.0103 & -0.0054 & 0.0009 & 0.0025 \\ -0.0040 & 0.0020 & -0.0054 & 0.0086 & -0.0007 & -0.0027 \\ 0.0007 & -0.0010 & 0.0009 & -0.0007 & 0.0020 & 0.0001 \\ 0.0035 & 0.0023 & 0.0025 & -0.0027 & 0.0001 & 0.0112 \end{bmatrix}. \quad (108)$$

The Mahalanobis distance matrix between the three countries of origin, for the first assessor is equal to

$$D_{[1]} = \begin{bmatrix} 0 & 51.0810 & 4.3609 \\ 51.0810 & 0 & 29.9891 \\ 4.3609 & 29.9891 & 0 \end{bmatrix}, \quad (109)$$

which is then transformed into a cross-product matrix (that, for convenience, has been normalized, à la MFA, to have a first eigenvalue of 1) equal to:

$$S_{[1]} = \begin{bmatrix} 0.3239 & -0.4426 & 0.1187 \\ -0.4426 & 0.6318 & -0.1892 \\ 0.1187 & -0.1892 & 0.0706 \end{bmatrix}. \quad (110)$$

All 10 assessors are processed in a similar fashion. From their cross-product matrices, we obtain a matrix of inner products whose eigendecomposition shows the relationship between assessors (see Figure 12). From this eigendecomposition, we also obtain the compromise whose analysis in turn gives the compromise factor scores map shown in Figure 13a.



FIGURE 12 | Canonical-statistics. PCA of the inner product map: the assessors.

In order to assess the quality of the discrimination, we have performed bootstrap resampling with 1000 iterations and projected on the compromise map the 95% confidence intervals (see Figure 13b). The figure indicates that the regions are well discriminated and that this discrimination is reliable.

Power STATIS

Recently, Benasséni and Bennani Dosse¹²³ looked for alternative criteria for STATIS-like techniques. Specifically, they explored and generalized Criterion 2 (see Eq. (66) and Appendix) and, instead of maximizing $\mathcal{S} = \sum \langle S_{[k]}, S_{[+]} \rangle^2$ (see Eq. (66)), they decided to explore the solution of the more general problem

$$\mathcal{S}' = \sum_k^K \langle S_{[k]}, S_{[+]} \rangle^p \text{ with } \alpha_k \geq 0 \text{ and } \alpha^T \alpha = 1, \quad (111)$$

where p can be any non-negative real number.

Of particular interest is the case of $p = 1$, which we call power-1 STATIS (because it maximizes the criterion with $p = 1$). In this case the quantity to maximize corresponds to the sum of the inner-products between the tables and the compromise. Formally, in Power-1 STATIS, the criterion to be maximized reduces to

$$\mathcal{S}_1 = \sum_k^K \langle S_{[k]}, S_{[+]} \rangle \text{ with } \alpha_k \geq 0 \text{ and } \alpha^T \alpha = 1. \quad (112)$$

This expression seems to capture better the intuition of a least squares minimization than the \mathcal{S} criterion because the inner product already behaves like a ‘squared quantity’ and squaring it again seems superfluous. Also, this minimization problem has a

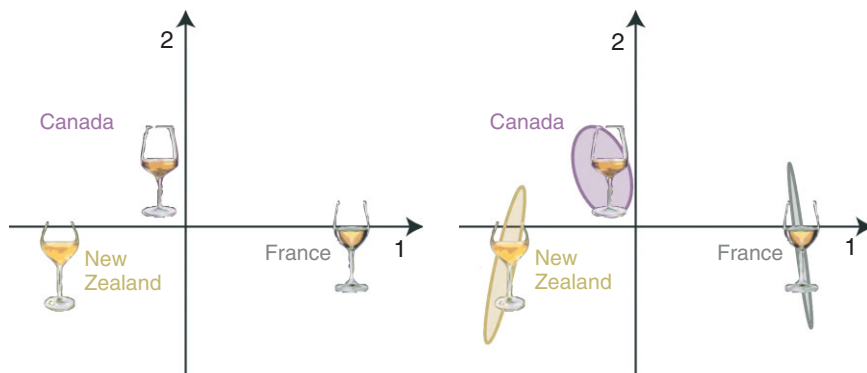


FIGURE 13 | Canonical-STATIS. Left panel: Compromise factor scores for the three wine regions. Right panel: Compromise factor scores for the three wine regions with bootstrapped 95% confidence intervals.

very simple and elegant solution because then the vector of optimal weights, can be obtained as

$$\alpha^{[1]} = C\mathbf{1} \times \|C\mathbf{1}\|^{-1} \tag{113}$$

(where $\mathbf{1}$ is a K by 1 vector of ones). For our example, α computed with standard STATIS and $\alpha^{[1]}$ computed from power-1 STATIS were almost identical (an effect also observed by Benasséni and Bennani Dosse¹²³ with several examples), and this suggests that the specific choice of the criterion—which is a matter of important theoretical interest—may not be a crucial problem for practical applications.

Power-1 STATIS can also be straightforwardly adapted to handle X-STATIS by replacing matrix C in Eq. (113) by C' from Eq. (100).

ANISOSTATIS

STATIS gives the same weight to each variable of a given table (namely α_k , see, e.g., Eqs (16) and (17)). A more formal way of expressing this condition is to say that STATIS is an *isotropic* technique. An interesting extension of STATIS is to relax the isotropic condition and to assign one weight per variable (whatever the table) instead of having the same weight for all variables of a table. This extension of STATIS is *anisotropic* and is called ANISOSTATIS.¹²⁴ So for ANISOSTATIS, the goal is to find values for the vector \mathbf{a} (cf., Eq. (16)) that will make the map of the compromise closer to the set of tables. Interestingly, when generalizing STATIS, the two equivalent criteria (i.e., Eqs (64) and (66)) maximized by STATIS give rise to two different solutions. In addition, it is also possible to derive a power-1 criterion analogous to the criterion used by power-1 STATIS.¹²³ We present these three different criteria below. All three criteria can also be directly adapted to work with PTA.

ANISOSTATIS Criterion 1

The first criterion maximized by ANISOSTATIS—equivalent to Eq. (64)—is expressed as:

$$\begin{aligned} \mathcal{V}_1 &= \|\mathbf{S}_{[+]}\|^2 = \langle \mathbf{S}_{[+]}, \mathbf{S}_{[+]} \rangle \\ &= \left\langle \sum_k \mathbf{S}_{[k]} \mathbf{A}_{[k]} \mathbf{S}_{[k]}^T, \sum_k \mathbf{S}_{[k]} \mathbf{A}_{[k]} \mathbf{S}_{[k]}^T \right\rangle \end{aligned} \tag{114}$$

with the constraint that $\mathbf{a}^T \mathbf{a} = 1$.

The solution of this problem involves the *covariance* matrix of \mathbf{X} with all of its elements squared. This J by J covariance matrix is denoted $\hat{\Sigma}$ and computed as:

$$\hat{\Sigma} = \mathbf{X}^T \mathbf{M} \mathbf{X}. \tag{115}$$

Each entry of $\hat{\Sigma}$ stores the covariance between two variables of \mathbf{X} . This matrix is then used to compute a matrix denoted C which is called the *Hadamard squared covariance matrix* and whose elements are the squares of the elements of $\hat{\Sigma}$. Specifically, C is computed as

$$C = \hat{\Sigma} \circ \hat{\Sigma} = (\mathbf{X}^T \mathbf{M} \mathbf{X}) \circ (\mathbf{X}^T \mathbf{M} \mathbf{X}) \tag{116}$$

(where \circ denotes the Hadamard—i.e., the element-wise—product between two matrices). This matrix C is positive semi-definite (because $\hat{\Sigma}$ is positive semi-definite) and all its elements, being squared numbers, are positive.

The solution of the first criterion is obtained directly from the eigendecomposition of C which is expressed as

$$C = \mathbf{U} \mathbf{\Omega} \mathbf{U}^T \text{ with } \mathbf{U}^T \mathbf{U} = \mathbf{I}. \tag{117}$$

As stated by the Perron–Frobenius theorem, the elements of the first eigenvector of C can be chosen to be all positive and so the optimum weight vector is obtained by rescaling this first eigenvector such that the sum of its elements is equal to one. All the other steps of ANISOSTATIS can then be performed in the same way as for the standard version of STATIS.

As an illustration, the set of optimum weights obtained with Criterion \mathcal{V}_1 is given in Table 5, and the factor score solution is shown in Figure 14.

ANISOSTATIS Criterion 2

The second criterion of ANISOSTATIS generalizes the criterion expressed in Eq. (66) and is expressed

TABLE 5 | Numerical Values of the \mathbf{a} Vector of the Optimum Weights for Criterion 1 of ANISOSTATIS

	Variables					
	V1	V2	V3	V4	V5	V6
Assessor 1	0.0215	0.0178	0.0174	0.0139	0.0196	0.0138
Assessor 2	0.0220	0.0216	0.0181	0.0160	0.0135	0.0124
Assessor 3	0.0229	0.0052	0.0179	0.0107	0.0184	0.0169
Assessor 4	0.0283	0.0200	0.0263	0.0158	0.0122	
Assessor 5	0.0215	0.0118	0.0179	0.0044	0.0158	0.0076
Assessor 6	0.0271	0.0227	0.0218	0.0159	0.0126	
Assessor 7	0.0295	0.0229	0.0293	0.0128		
Assessor 8	0.0195	0.0158	0.0138	0.0051	0.0196	0.0137
Assessor 9	0.0274	0.0172	0.0247	0.0180	0.0254	
Assessor 10	0.0375	0.0283	0.0297	0.0287		

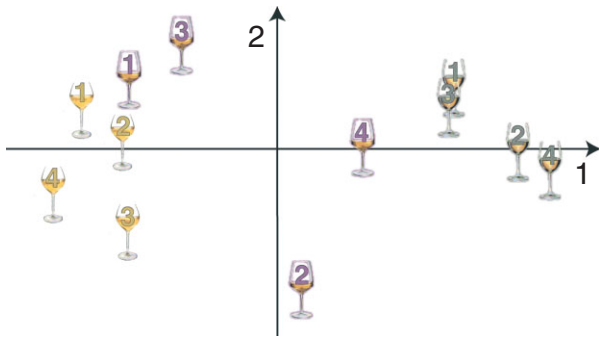


FIGURE 14 | ANISOSTATIS. Map of the factor scores for Dimensions 1 and 2 using Criterion 1 of ANISOSTATIS (because the α weights are very similar for all three criteria of ANISOSTATIS, this map also corresponds to the solutions obtained with Criteria 2 and 3).

similarly as

$$S' = \sum_k^K \langle S_{[k]}, S_{[+]} \rangle^2 \quad \text{with} \quad \mathbf{a}^T \mathbf{a} = 1. \quad (118)$$

The solution of this problem involves a new J by K matrix denoted \mathbf{B} in which a 1 in row j and column k indicates that the j th variable belongs to the k th table and a value of 0 that it does not. The solution of this problem is obtained by first computing the (positive semidefinite) matrix

$$(\mathbf{CB}) \times (\mathbf{CB})^T \quad (119)$$

which is then eigen-decomposed as

$$(\mathbf{CB}) \times (\mathbf{CB})^T = \mathbf{V} \mathbf{\Psi} \mathbf{V}^T \quad \text{with} \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}. \quad (120)$$

Note, that this eigendecomposition can also be more efficiently obtained from the singular value decomposition of the matrix \mathbf{CB} .

As stated by the Perron–Frobenius theorem, the elements of the first eigenvector of $(\mathbf{CB}) \times (\mathbf{CB})^T$ can be chosen to be all positive and the optimum weight vector is obtained by rescaling this first eigenvector to a sum of 1. All the other steps of STATIS can then be performed in the same way as for the standard version.

As an additional benefit of using the singular value decomposition of \mathbf{CB} , we can plot the table factor scores in a manner similar to the plot of matrix \mathbf{C} for standard STATIS (although in this case, because we are using the squared values of a covariance matrix, the first singular value is in general very much larger than the second and it is therefore preferable to plot Dimensions 2 versus 3 in order to explore the similarity structure of the tables).

TABLE 6 | Numerical Values of the \mathbf{a} Vector of the Optimum Weights for Criterion 2 of ANISOSTATIS

	Variables					
	V1	V2	V3	V4	V5	V6
Assessor 1	0.0209	0.0177	0.0170	0.0146	0.0198	0.0137
Assessor 2	0.0213	0.0212	0.0176	0.0163	0.0145	0.0128
Assessor 3	0.0221	0.0059	0.0168	0.0109	0.0188	0.0175
Assessor 4	0.0278	0.0203	0.0262	0.0160	0.0133	
Assessor 5	0.0207	0.0120	0.0173	0.0053	0.0164	0.0088
Assessor 6	0.0261	0.0228	0.0208	0.0162	0.0132	
Assessor 7	0.0280	0.0233	0.0281	0.0142		
Assessor 8	0.0187	0.0158	0.0140	0.0061	0.0198	0.0144
Assessor 9	0.0265	0.0171	0.0238	0.0185	0.0253	
Assessor 10	0.0369	0.0280	0.0298	0.0292		

The set of optimum weights obtained with Criterion S' is given in Table 6. As can be seen by comparing Tables 5 and 6, the optimum weights are almost identical for these two versions of ANISOSTATIS, and the factor scores are also almost identical (and so Figure 14 also shows the plot of the factor scores obtained with Criterion S').

ANISOSTATIS Criterion 3. Power-STATIS

In a manner equivalent to power-1 STATIS, the second criterion can be modified in order to involve the minimization of the sum of the inner products of the cross-product matrices to the compromise rather than the squared inner products. Specifically, the criterion to maximize is then (compare with Eqs 111 and 118):

$$S'_1 = \sum_k^K \langle S_{[k]}, S_{[+]} \rangle \quad \text{with} \quad \mathbf{a}^T \mathbf{a} = 1. \quad (121)$$

Like power-1 STATIS, the power-1 version of ANISOSTATIS does not require the computation of an eigendecomposition (a important feature when dealing with very large matrices because the eigendecomposition computational time depends on the cube of the size of the matrix to be decomposed). Specifically, the vector \mathbf{a} of optimal weights is obtained as (compare with Eq. (113))

$$\mathbf{a} = \mathbf{C} \mathbf{1} \times \|\mathbf{C} \mathbf{1}\|^{-1} \quad (122)$$

(where $\mathbf{1}$ is a J by 1 vector of ones). This vector of optimal weights is then used as in standard-STATIS to compute factor scores and loadings.

The set of optimum weights obtained with Criterion S'_1 is given in Table 7. As can be seen

TABLE 7 | Numerical Values of the \mathbf{a} Vector of the Optimum Weights for Criterion 3 (Power 1) of ANISOSTATIS

	Variables					
	V1	V2	V3	V4	V5	V6
Assessor 1	0.0210	0.0176	0.0169	0.0147	0.0197	0.0137
Assessor 2	0.0212	0.0212	0.0175	0.0164	0.0145	0.0127
Assessor 3	0.0221	0.0061	0.0169	0.0110	0.0188	0.0175
Assessor 4	0.0278	0.0203	0.0262	0.0160	0.0135	
Assessor 5	0.0208	0.0120	0.0174	0.0054	0.0165	0.0090
Assessor 6	0.0261	0.0226	0.0208	0.0161	0.0133	
Assessor 7	0.0280	0.0233	0.0281	0.0142		
Assessor 8	0.0187	0.0158	0.0142	0.0063	0.0198	0.0145
Assessor 9	0.0264	0.0170	0.0238	0.0185	0.0252	
Assessor 10	0.0367	0.0278	0.0297	0.0291		

by comparing Tables 5–7, the optimum weights are almost identical for all three versions of ANISOSTATIS, and, in fact, the factor scores are also almost identical (and so Figure 14 also shows the plot of the factor scores obtained with Criterion S_3).

(K + 1)-STATIS

An interesting extension to STATIS explores the relationship of K tables with one external table. As an illustration, in our example we have 10 tables representing the assessors and one table representing the chemical properties of the wines. An important question is to explicitly evaluate the similarities among the assessors relative to the way they relate to the chemical properties of the wines. This problem is called a $(K + 1)$ table problem. Here, we have a set of K tables (i.e., the K assessors), each of dimensions I by $J_{[k]}$ and one external table [i.e., the $(K + 1)$ th table of the chemical properties of the wines] represented by an I by M matrix denoted \mathbf{H} . A recent extension of STATIS called $(K + 1)$ -STATIS or *external-STATIS* has been proposed to handle this type of question.¹²⁵ The goal of the method is to analyze the pattern of similarity between the K tables in relation to the $(K + 1)$ th table. In addition to generalizing STATIS, this method can also be seen as an extension of Tucker inter-battery,¹⁰⁴ co-inertia,^{32,33} or partial-least square correlation.^{105,106,126} The basic idea is to compute the optimal α weights using correlation or covariance matrices between each of the K tables and the $(K + 1)$ th table (instead of the cross-product matrices). The optimal weights thus computed are then used in lieu of the α weights of STATIS to compute a standard solution for

STATIS (based on the K matrices). The $(K + 1)$ th matrix is projected as a supplementary element in the solution.

Formally, we have K matrices $\mathbf{X}_{[k]}$ (each of order I by $J_{[k]}$) and one I by M ‘external matrix’ \mathbf{H} , with all these tables measuring variables on the same set of I observations. These matrices are supposed to have been preprocessed in an appropriate way. The first step of the analysis is to compute a correlation or covariance matrix (depending upon the type of preprocessing) between each of the $\mathbf{X}_{[k]}$ matrices and \mathbf{H} . These matrices, denoted $\mathbf{W}_{[k]}$ (each of order M by $J_{[k]}$) are computed as

$$\mathbf{W}_{[k]} = \mathbf{H}^T \mathbf{X}_{[k]}. \tag{123}$$

For the first step of STATIS (i.e., finding the optimum weights) we use the cross-product matrices computed from the $\mathbf{W}_{[k]}$ matrices as

$$\mathbf{S}_{[k]}^* = \mathbf{W}_{[k]} \mathbf{W}_{[k]}^T = \mathbf{H}^T \mathbf{X}_{[k]} \mathbf{X}_{[k]}^T \mathbf{H}. \tag{124}$$

As the term $\mathbf{X}_{[k]} \mathbf{X}_{[k]}^T$ defines the standard cross-product (see Eq. (3)), this last equation can be rewritten as:

$$\mathbf{S}_{[k]}^* = \mathbf{H}^T \mathbf{S}_{[k]} \mathbf{H}. \tag{125}$$

This makes explicit the relationships between $(K + 1)$ -STATIS and standard STATIS.

From the cross-product matrices $\mathbf{S}_{[k]}^*$, we then compute the inner product matrix (cf., Eq. 11) denoted \mathbf{C}^* whose generic term is

$$c_{k,k'}^* = \left\langle \mathbf{S}_{[k]}^*, \mathbf{S}_{[k']}^* \right\rangle = \left\langle \mathbf{H}^T \mathbf{X}_{[k]} \mathbf{X}_{[k]}^T \mathbf{H}, \mathbf{H}^T \mathbf{X}_{[k']} \mathbf{X}_{[k']}^T \mathbf{H} \right\rangle. \tag{126}$$

The eigendecomposition of matrix \mathbf{C}^* gives (cf., Eq. (120)):

$$\mathbf{C}^* = \mathbf{U}^* \mathbf{\Gamma}^* \mathbf{U}^{*T} \text{ with } \mathbf{U}^{*T} \mathbf{U}^* = \mathbf{I}. \tag{127}$$

From the first eigenvector of \mathbf{C}^* , denoted \mathbf{u}_1^* , we obtain the set of optimal weights for $(K + 1)$ -STATIS in the same way as for STATIS (cf., Eq. (15)) as:

$$\boldsymbol{\alpha}^* = \mathbf{u}_1^* \times (\mathbf{u}_1^{*T} \mathbf{1})^{-1}. \tag{128}$$

The elements of $\boldsymbol{\alpha}^*$ are then used in lieu of the elements of $\boldsymbol{\alpha}$ and the rest of the STATIS procedure is applied with these weights.

As an illustration, we took the matrix of chemical properties of the wines (e.g., \mathbf{X}_{sup} , see

Table 4) as our H matrix. The first covariance matrix $W_{[1]}$ is

$$\begin{aligned}
 W_{[1]} = H^T X_{[1]} &= \begin{bmatrix} -0.065 & 0.056 & 0.218 & 0.096 \\ -0.105 & 0.119 & 0.099 & 0.200 \\ 0.016 & 0.011 & 0.218 & 0.096 \\ 0.326 & -0.023 & 0.099 & 0.251 \\ -0.146 & 0.148 & -0.139 & -0.162 \\ -0.027 & -0.101 & -0.139 & -0.076 \\ -0.213 & 0.036 & -0.020 & -0.283 \\ -0.065 & 0.065 & -0.258 & -0.059 \\ 0.205 & 0.023 & -0.020 & 0.062 \\ -0.051 & 0.077 & 0.099 & -0.059 \\ 0.016 & 0.023 & -0.139 & 0.010 \\ 0.110 & -0.433 & -0.020 & -0.076 \end{bmatrix}^T \times \begin{bmatrix} 0.12 & 0.13 & 0.07 & -0.04 & -0.18 & 0.11 \\ 0.07 & 0.05 & 0.13 & -0.18 & -0.12 & 0.23 \\ 0.01 & 0.05 & 0.02 & 0.01 & -0.07 & -0.01 \\ 0.18 & 0.13 & 0.13 & -0.04 & -0.07 & 0.05 \\ -0.21 & -0.18 & -0.20 & 0.16 & 0.15 & -0.07 \\ -0.16 & -0.03 & -0.09 & 0.21 & 0.10 & -0.19 \\ -0.05 & -0.11 & -0.04 & -0.04 & 0.21 & -0.07 \\ -0.05 & -0.18 & -0.09 & 0.16 & 0.15 & -0.01 \\ 0.12 & 0.13 & 0.13 & -0.04 & -0.01 & 0.17 \\ -0.10 & 0.13 & -0.20 & 0.01 & -0.07 & -0.01 \\ 0.12 & -0.03 & 0.13 & -0.18 & -0.07 & -0.07 \\ -0.05 & -0.11 & 0.02 & -0.04 & -0.01 & -0.13 \end{bmatrix} \\
 &= \begin{bmatrix} 0.118 & 0.100 & 0.109 & -0.036 & -0.080 & 0.036 \\ 0.008 & 0.029 & -0.028 & 0.000 & 0.005 & 0.097 \\ 0.088 & 0.154 & 0.069 & -0.090 & -0.148 & 0.098 \\ 0.151 & 0.142 & 0.140 & -0.090 & -0.163 & 0.135 \end{bmatrix}. \tag{129}
 \end{aligned}$$

From $W_{[1]}$, we can compute $S_{[1]}^*$ as

$$\begin{aligned}
 S_{[1]}^* &= W_{[1]} W_{[1]}^T \\
 &= \begin{bmatrix} 0.0037 & 0.0003 & 0.0043 & 0.0057 \\ 0.0003 & 0.0009 & 0.0010 & 0.0011 \\ 0.0043 & 0.0010 & 0.0063 & 0.0075 \\ 0.0057 & 0.0011 & 0.0075 & 0.0096 \end{bmatrix}. \tag{130}
 \end{aligned}$$

This cross-product matrix (and the nine other ones) are then used to compute the inner product matrix C^* (whose first element, for example, was equal to $c_{1,2}^* = 0.00039$, note that the elements of C^* are likely to be small and the use of the R_V coefficient rather than the inner product will eliminate this scale problem).

From the eigendecomposition of C^* , we obtain the first eigenvector, which, in turn, is used to compute the value of α^* (compare with Eq. (75)) as:

This vector α^* is then plugged in the standard STATIS algorithm (in lieu of α) and this will provide factor scores and loadings. As an illustration, we show in Figure 15, the factor scores for the compromise and the H table. The comparison of these figures with Figures 3a and 7 indicates that, here, taking into account the similarity of the assessor with the chemical properties of the wines did not change much the overall solution. This is probably because of the fact that the (fictitious) assessors were all roughly equally sensitive to the chemical properties of the wines.

DOUBLE-STATIS or DO-ACT

A new method, called DO-ACT (for DOUBLE-ACT) or double-STATIS, further extends the $(K + 1)$ -STATIS methodology and generalizes Tucker interbattery,¹⁰⁴ coinertia,^{32,33} and partial-least square correlation.^{105,106,126} In the double-STATIS approach,

$$\begin{aligned}
 \alpha^* = u_1^* \times (1^T u_1^*)^{-1} &= \begin{bmatrix} 0.2910 \\ 0.3108 \\ 0.2596 \\ 0.3322 \\ 0.2216 \\ 0.3426 \\ 0.2867 \\ 0.2742 \\ 0.3916 \\ 0.4047 \end{bmatrix} \times 3.1149^{-1} = \begin{bmatrix} 0.0934 \\ 0.0998 \\ 0.0833 \\ 0.1066 \\ 0.0711 \\ 0.1100 \\ 0.0920 \\ 0.0880 \\ 0.1257 \\ 0.1299 \end{bmatrix}. \tag{131}
 \end{aligned}$$

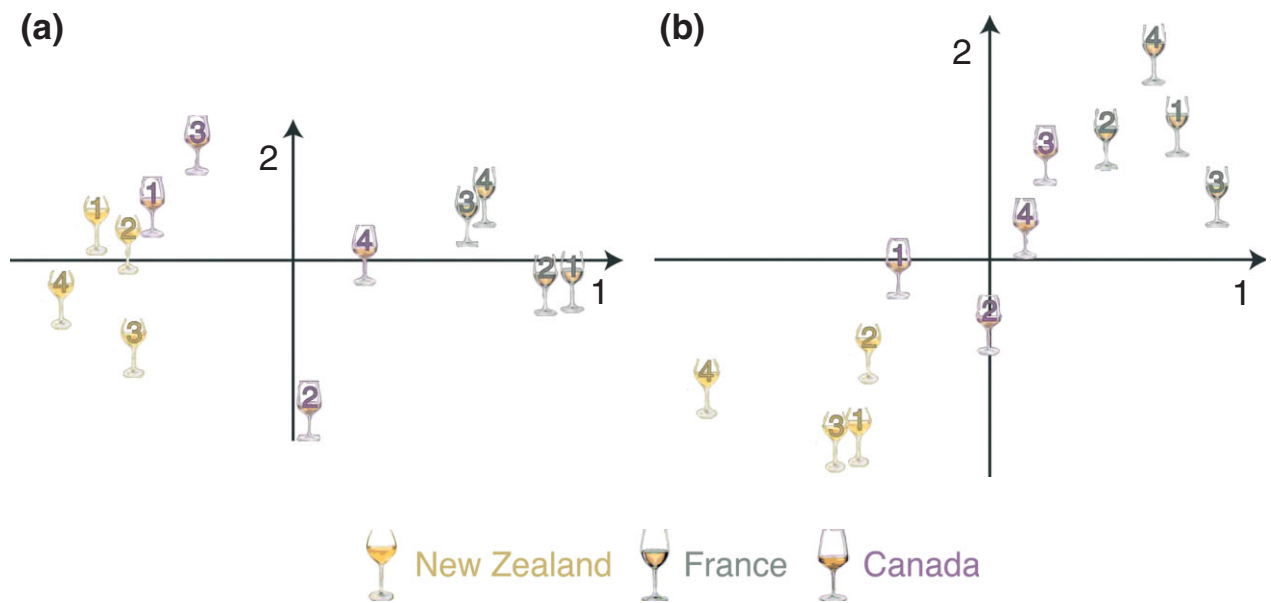


FIGURE 15 | $(K + 1)$ -STATIS. (a) Compromise factor scores. (b) Factor scores for the \mathbf{H} matrix: Chemical components of the wines (compare with Figures 3a and 7).

we have one series of K matrices $\mathbf{X}_{[k]}$ (each of order I by $J_{[k]}$) and one series of L matrices $\mathbf{H}_{[\ell]}$ (each of order I by $M_{[\ell]}$) with all these matrices collecting measures on the same I observations. The goal of the analysis is to find *two* compromises—one for the $\mathbf{X}_{[k]}$'s and one for the $\mathbf{H}_{[\ell]}$'s—such that these two compromises are as similar as possible (with the similarity being measured by the inner product between these two compromises). So, more formally, if we denote, respectively, $\mathbf{S}_{[k]}$ and $\mathbf{T}_{[\ell]}$ the I by I cross-product matrices of, respectively, $\mathbf{X}_{[k]}$ and $\mathbf{H}_{[\ell]}$ and by, respectively, $\mathbf{S}_{[+]}$, and $\mathbf{T}_{[+]}$ the I by I compromise cross-product matrices for, respectively, the $\mathbf{X}_{[k]}$'s and $\mathbf{H}_{[\ell]}$'s, we are looking for two sets of weights ($\boldsymbol{\alpha}$ for the $\mathbf{S}_{[k]}$'s and $\boldsymbol{\beta}$ for the $\mathbf{T}_{[\ell]}$'s) such that the two compromise matrices are computed as (cf., Eq. 34)

$$\mathbf{S}_{[+]} = \sum_k \alpha_k \mathbf{S}_{[k]} \text{ and } \mathbf{T}_{[+]} = \sum_\ell \beta_\ell \mathbf{T}_{[\ell]} \text{ with } \boldsymbol{\alpha}^\top \boldsymbol{\alpha} = \boldsymbol{\beta}^\top \boldsymbol{\beta} = 1. \tag{132}$$

In a manner analogous to standard STATIS, the solution of this problem requires the computation of a K by L inner product matrix. But, by contrast, with standard STATIS, this matrix (being K by L), is obviously not, in general a square symmetric matrix anymore, as its generic term gives the inner product between an $\mathbf{S}_{[k]}$ matrix and a $\mathbf{T}_{[\ell]}$ matrix (instead of between two $\mathbf{S}_{[k]}$ matrices). Specifically, the value of the element $c_{k,\ell}$ from the K by L inner product matrix

\mathbf{C} is computed as (compare with Eq. (8)):

$$\begin{aligned} c_{k,\ell} &= \langle \mathbf{S}_{[k]}, \mathbf{T}_{[\ell]} \rangle \\ &= \text{trace} \{ \mathbf{X}_{[k]} \mathbf{X}_{[k]}^\top \times \mathbf{H}_{[\ell]} \mathbf{H}_{[\ell]}^\top \} \\ &= \text{trace} \{ \mathbf{S}_{[k]} \times \mathbf{H}_{[\ell]} \} \\ &= \text{vec} \{ \mathbf{S}_{[k]} \}^\top \times \text{vec} \{ \mathbf{H}_{[\ell]} \} \\ &= \sum_i \sum_j s_{i,j,k} t_{i,j,\ell}. \end{aligned} \tag{133}$$

The *singular value decomposition* (as opposed to the eigenvalue decomposition, for standard STATIS) of \mathbf{C} provides a set of right singular vectors (corresponding to the $\mathbf{X}_{[k]}$'s), a set of left singular vectors (corresponding to the $\mathbf{H}_{[\ell]}$'s) and a set of singular values as:

$$\mathbf{C} = \mathbf{U} \mathbf{\Gamma} \mathbf{V}^\top \text{ with } \mathbf{U}^\top \mathbf{U} = \mathbf{V}^\top \mathbf{V} = \mathbf{I}. \tag{134}$$

The first left singular vector (denoted \mathbf{u}_1) and the first right singular vector (denoted \mathbf{v}_1) provide—after re-scaling such that their respective sum is one—the value of the two sets of optimal weights as (compare with Eq. (15)):

$$\boldsymbol{\alpha} = \mathbf{u}_1 \times (\mathbf{u}_1^\top \mathbf{1})^{-1} \text{ and } \boldsymbol{\beta} = \mathbf{v}_1 \times (\mathbf{v}_1^\top \mathbf{1})^{-1}. \tag{135}$$

Each of the compromises in double-STATIS can then be analyzed as in standard-STATIS. A detailed

example is provided in Ref 127. The relationships between the different extensions of STATIS for two sets of tables such as STATICO, COSTATIS, and double-STATIS remain to be explored.

The double-STATIS approach has also been, recently, extended to multiple sets of data matrices.^{60,61} This extension, called STATIS-4, uses an iterative process to compute a compromise for each of the set of tables and a global compromise (i.e., the compromise of the compromises).

RELATED METHODS

STATIS is part of the multitable family and also of the Procrustes family. A thorough evaluation of the complex relationships between all these techniques is beyond the scope of this article, but some directions for comparisons could be of interest. STATIS is also closely related to generalized canonical correlation analysis GCCA. We will look first at the relationship between STATIS and GCCA and then at the relationships between STATIS and the other techniques.

STATIS and Generalized Canonical Correlation Analysis

An important anisotropic technique is generalized canonical correlation analysis (GCCA). In GCCA,^{128–130} the goal is to maximize the inner-product of the factor scores of the compromise under a constraint that makes it equivalent to maximizing the correlations between the compromise factor scores and the partial factor scores. Specifically, in GCCA, we seek a matrix of J weights by L dimensions (with L being the number of dimensions for the solution) $\mathbf{W}_{\text{gccca}}$ such that

$$\mathbf{F}_{\text{gccca}} = \mathbf{X}\mathbf{W}_{\text{gccca}} \text{ with } \text{trace} \left\{ \mathbf{F}_{\text{gccca}}^T \mathbf{F}_{\text{gccca}} \right\} = \max. \quad (136)$$

under the constraints that

$$\mathbf{W}_{\text{gccca}}^T \tilde{\Sigma} \mathbf{W}_{\text{gccca}} = \mathbf{I}, \quad (137)$$

where $\tilde{\Sigma}$ is a J by J block diagonal matrix with each $J_{[k]}$ by $J_{[k]}$ diagonal block equal to $\frac{1}{7} \mathbf{X}_{[k]}^T \mathbf{X}_{[k]}$ (note that, here, as in CANOSTATIS, the inverse can be replaced by a pseudoinverse^{129,130}).

The set of $\mathbf{W}_{\text{gccca}}$ weights is obtained from the following GSVD of \mathbf{X} :

$$\begin{aligned} \mathbf{X} &= \mathbf{P}_{\text{gccca}} \Delta_{\text{gccca}} \mathbf{Q}_{\text{gccca}}^T \text{ with} \\ \mathbf{P}_{\text{gccca}}^T \mathbf{M} \mathbf{P}_{\text{gccca}} &= \mathbf{Q}_{\text{gccca}}^T \tilde{\Sigma}^{-1} \mathbf{Q}_{\text{gccca}} = \mathbf{I}. \end{aligned} \quad (138)$$

The factors scores $\mathbf{F}_{\text{gccca}}$ are obtained as

$$\begin{aligned} \mathbf{F}_{\text{gccca}} &= \mathbf{P}_{\text{gccca}} \Delta_{\text{gccca}} = \mathbf{X} \tilde{\Sigma}^{-1} \mathbf{Q}_{\text{gccca}} \\ &= \mathbf{X} \mathbf{W}_{\text{gccca}} \text{ with } \mathbf{W}_{\text{gccca}} = \tilde{\Sigma}^{-1} \mathbf{Q}_{\text{gccca}}. \end{aligned} \quad (139)$$

So, in brief, GCCA and ANISOSTATIS are both anisotropic techniques, but the criteria they maximize are quite different: maximum correlation for GCCA and maximum squared inner product (or inner product, depending upon the version) for ANISOSTATIS. Further work is needed to further develop the comparison between GCCA, STATIS, and their variants.

STATIS and GPA

The relationship of STATIS and general Procrustes analysis (GPA) is examined in Refs 12,131. An obvious advantage of STATIS over GPA is to be an eigendecomposition technique and therefore STATIS does not require multiple iterations to reach a consensus and is also guaranteed to converge. A main point of difference between the two techniques is the space they try to fit: STATIS considers the whole space (i.e., the whole cross product matrices) in which the K tables are, whereas GPA considers the specific dimensions (i.e., the factor scores) and, is general, restricted to a subset of these dimensions.

STATIS and Multiblock Analyses

STATIS is also part of the multitable or multiblock family of PCA extensions.^{2–15} The well-known members of this family—such as multiple factor analysis, SUM-PCA, consensus PCA, and multiblock correspondence analysis—reduce to the PCA of a matrix \mathbf{X} in which each $\mathbf{X}_{[k]}$ has been normalized in a specific way for each technique (see section on table normalization). STATIS differs essentially by the choice of the optimum weights which is an additional step after table normalization has been performed (see, e.g., Ref 132 for a comparison between STATIS and multiple factor analysis). In this sense, STATIS encompasses all these methods as particular cases (i.e., when all the weights in STATIS are equal and when the adequate table normalization has been performed).

STATIS (and in particular DISTATIS) can be seen also as a constrained simplified version of INDSCAL (see Ref 13 for a comparison between STATIS and INDSCAL, and Ref 133 for a geometrical interpretation of INDSCAL and multitable analyses) because STATIS uses isotropic weights on the tables,

whereas INDSCAL uses anisotropic weights on the *dimensions* (in addition STATIS, being an eigenvector base technique is guaranteed to converge, whereas INDSCAL is an iterative technique whose convergence is not always certain). DISTATIS, compared to INDSCAL, is obviously also an isotropic technique tailored for Euclidean distance matrices whereas INDSCAL can handle non Euclidean similarity matrices. ANISOSTATIS and INDSCAL are both anisotropic but ANISOSTATIS weights the *variables* of the tables whereas INDSCAL weights the *dimensions* of the solutions.

COMPUTER PACKAGES

Several statistical packages implement STATIS and some of their variants. A very comprehensive package (oriented for ecologists) is the R-package *ade4*.^{56,134} STATIS-4 has been programmed in S (and is likely to run with R) and is available from Ref 60. Also all the techniques described in this article (with the exception of STATIS-4) have been programmed in MATLAB and are available from the senior author's home page at www.utdallas.edu/~herve. In addition, R-programs—implementing the techniques described in this article—written by Derek Beaton are available from this address. A simple DISTATIS program written by Guillaume Blancher is also available at this address.

CONCLUSION

STATIS is a simple, elegant, and robust technique that can be used to integrate multiple data tables collected on the same set of observations. In addition of being of very practical importance, STATIS is also a domain of recent vigorous theoretical developments; and its simplicity (both theoretical and computational) makes it an ideal tool for the very large data sets of modern science.

APPENDIX

THE INNER-PRODUCT BETWEEN POSITIVE SEMI-DEFINITE MATRICES IS POSITIVE OR NULL

Here we show that the inner-product between positive semidefinite matrices is always larger or equal to zero. Recall that a matrix **S** is positive semidefinite if there is a matrix **X** such that **S** can be written as the product **XX^T**. Let **S** and **T** be two positive

semidefinite matrices. The inner product between these two matrices is equal to

$$\langle \mathbf{S}, \mathbf{T} \rangle = \text{trace} \{ \mathbf{ST} \}. \tag{140}$$

Because **T** is positive semidefinite, the left part of Eq. (140) can be rewritten as

$$\text{trace} \left\{ \mathbf{T}^{\frac{1}{2}} \mathbf{ST}^{\frac{1}{2}} \right\}. \tag{141}$$

Because **S** is positive semidefinite it can be expressed as $\mathbf{S} = (\mathbf{S}^{\frac{1}{2}})(\mathbf{S}^{\frac{1}{2}})^T$, therefore, the matrix $\mathbf{T}^{\frac{1}{2}} \mathbf{ST}^{\frac{1}{2}} = (\mathbf{T}^{\frac{1}{2}} \mathbf{S}^{\frac{1}{2}}) (\mathbf{T}^{\frac{1}{2}} \mathbf{S}^{\frac{1}{2}})^T$ is also positive semidefinite, and therefore all its eigenvalues are positive or null. Thus, its trace being equal to the sum of its eigenvalues is also positive or null and this completes the proof.

STATIS. CRITERION 1

Preliminaries

First we note that the quadratic expression $\alpha^T \mathbf{C} \alpha$ can be rewritten as:

$$\alpha^T \mathbf{C} \alpha = \sum_k^K \sum_m^K \alpha_k \alpha_m c_{k,m}. \tag{142}$$

This can be obtained by developing $\alpha^T \mathbf{C} \alpha$ to express it as a sum.

Some properties of the trace operator will be useful (see, for more details, e.g., Refs 135,136):

- **Property P1:** $\text{trace} \{ \alpha \times \mathbf{S} \} = \alpha \times \text{trace} \{ \mathbf{S} \}$ for any scalar α .
- **Property P2:** $\text{trace} \{ \mathbf{S} + \mathbf{T} \} = \text{trace} \{ \mathbf{S} \} + \text{trace} \{ \mathbf{T} \}$
- **Property P3:** Combining Properties 2 and 3, we have:

$$\text{trace} \left\{ \sum_k \alpha_k \mathbf{S}_{[k]} \right\} = \sum_t \alpha_k \text{trace} \{ \mathbf{S}_{[k]} \} .$$

- **Property P4:** $\text{trace} \{ \mathbf{ST} \} = \text{trace} \{ \mathbf{TS} \} = \text{trace} \{ \mathbf{T}^T \mathbf{S}^T \}$.

Maximum Variance of Compromise

Here we show that Criterion 1 of STATIS maximizes:

$$\mathcal{V} = \|\mathbf{S}_{[+]}\|^2 = \langle \mathbf{S}_{[+]}, \mathbf{S}_{[+]} \rangle \text{ with } \alpha^T \alpha = 1. \tag{143}$$

First we show that $\mathcal{V} = \alpha^T C \alpha$. This is shown by developing Eq. (143) as

$$\begin{aligned} \langle S_{[+]}, S_{[+]} \rangle &= \text{trace} \{ S_{[+]} S_{[+]} \} \\ &= \text{trace} \left\{ \left(\sum_k^K \alpha_k S_{[k]} \right) \left(\sum_m^K \alpha_m S_{[m]} \right) \right\} \\ &= \text{trace} \left\{ \sum_k^K \sum_m^K \alpha_k \alpha_m S_{[k]} S_{[m]} \right\} \\ &= \sum_k^K \sum_m^K \alpha_k \alpha_m \text{trace} \{ S_{[k]} S_{[m]} \} \quad \text{from P1} \\ &= \sum_k^K \sum_m^K \alpha_k \alpha_m \langle S_{[k]}, S_{[m]} \rangle \\ &= \sum_k^K \sum_m^K \alpha_k \alpha_m c_{k,m} \\ &= \alpha^T C \alpha \quad (\text{From Eq. 142}). \end{aligned} \tag{144}$$

Second, we show that $\alpha^T C \alpha$ is maximum when α is first eigenvector of C under the constraint that $\alpha^T \alpha = 1$ (see Ref 136, p. 161 or Ref 50, p. 455, for more details).

To do so, we start by writing the Lagrangian

$$\mathcal{L} = \frac{1}{2} \left\{ \alpha^T C \alpha - \lambda (\alpha^T \alpha - 1) \right\}. \tag{145}$$

Differentiating \mathcal{L} with respect to α gives

$$\frac{\partial \mathcal{L}}{\partial \alpha} = C \alpha - \lambda \alpha. \tag{146}$$

The expression \mathcal{L} is maximum when its derivative vanishes and this implies that

$$C \alpha = \lambda \alpha \tag{147}$$

which shows that α is an eigenvector of C and so $\alpha^T C \alpha = \lambda$. Therefore the maximum of \mathcal{L} is reached when α is the first eigenvector of C . Note, also that, because of the Perron–Frobenius theorem, the elements of α can always be chosen to be positive (see e.g., Ref 68 p. 34ff).

Alternative Version: Minimizing Dissimilarity

Criterion \mathcal{V} from Eq. (143) can be rewritten as a *minimization* problem for which we minimize the sum of the square distances between the cross-product matrices and the compromise. Specifically, we want to minimize:

$$\mathcal{D} = \sum_k^K \| S_{[k]} - \alpha_k S_{[+]} \|^2 \text{ with } \alpha^T \alpha = 1. \tag{148}$$

Developing and simplifying Eq. (148) gives

$$\mathcal{D} = \sum_k^K \| S_{[k]} \|^2 - \alpha^T C \alpha, \tag{149}$$

which, obviously, reaches its minimum when $\alpha^T C \alpha$ reaches its maximum.

STATIS. CRITERION 2

Here we show that STATIS maximizes the following criterion²¹

$$\mathcal{S} = \sum_k^K \langle S_{[k]}, S_{[+]} \rangle^2 \text{ with } \alpha^T \alpha = 1. \tag{150}$$

This is obtained by first developing and simplifying Eq. (150) to obtain

$$\begin{aligned} \sum_k^K \langle S_{[k]}, S_{[+]} \rangle^2 &= \sum_k^K \sum_m^K \alpha_k \alpha_m \sum_n^K \text{trace} \{ S_{[k]} S_{[n]} \} \\ &\quad \times \text{trace} \{ S_{[n]} S_{[m]} \} \\ &= \sum_k^K \sum_m^K \alpha_k \alpha_m \sum_n^K c_{k,n} \times c_{n,m}. \end{aligned} \tag{151}$$

The term $\sum_n^K c_{k,n} \times c_{n,m}$ corresponds to the element at row k and column m of C^2 , and therefore Eq. (151) can be rewritten as (cf., Eq. (142)):

$$\mathcal{S} = \alpha^T C^2 \alpha. \tag{152}$$

Because C^2 has the same eigenvectors as C (and its eigenvalues are the squared of the eigenvalues of C), the approach use previously shows \mathcal{S} reaches its maximum when α is the first eigenvector of C^2 which is also the first eigenvector of C .

ACKNOWLEDGMENT

We would like to thank two anonymous reviewers and also particularly Yoshio Takane for their very thorough and helpful comments on a draft of this article.

REFERENCES

1. Abdi H. Multivariate analysis. In: Lewis-Beck M, Bryman A, Futing T, eds. *Encyclopedia for Research Methods for the Social Sciences*. Thousand Oaks (CA): Sage; 2003, 699–702.
2. Acar E, Yener B. Unsupervised multiway data analysis: a literature survey. *IEEE Trans Knowledge Data Eng* 2009, 21:6–19.
3. Arcidiacono C, Sarnacchiaro P, Velleman R. Testing fidelity to a new psychological intervention for family members of substance misusers during implementation in Italy. *J Subst Abuse* 2008, 13:361–381.
4. Stanimirova I, Boucon C, Walczak B. Relating gas chromatographic profiles to sensory measurements describing the end products of the Maillard reaction. *Talanta* 2011, 83:1239–1246.
5. Williams LJ, Abdi H, French R, Orange JB. A tutorial on multi-block discriminant correspondence analysis (MUDICA): A new method for analyzing discourse data from clinical populations. *J Speech Lang Hearing Res* 2010, 53:1372–1393.
6. Daszykowski M, Walczak B. Methods for the exploratory analysis of two-dimensional chromatographic data. *Talanta* 2011, 83:1088–1097.
7. Carlier A, Lavit C, Pagès M, Pernin M, Turlot J. A comparative review of methods which handle a set of indexed data tables. In: Coppi R, Bollasco S, eds. *Multiway Data Analysis*. Amsterdam: North Holland; 1989, 85–101.
8. Derks EPPA, Westerhuis JA, Smilde AK, King BM. An introduction to multi-blocks component analysis by means of a flavor language case study. *Food Quality Prefer* 2003, 14:497–506.
9. Escofier B, Pagès J. Multiple factor analysis. *Comp Stat Data Anal* 1990, 18:121–140.
10. Guebel DV, Cánovas M, Torres N. 2009 Model Identification in presence of incomplete information by generalized principal component analysis: application to the common and differential responses of *Escherichia coli* to multiple pulse perturbation in continuous high-biomass density culture. *Biotechnol Bioeng* 104:785–795
11. Hassani S, Martens M, Qannari EM, Hanafi M. Analysis of -omic data: Graphical interpretation and validation tools in multi-block methods. *Chem Intell Lab Syst* 2010, 104:140–153.
12. Meyners M, Kunert J, Qannari EM. Comparing generalized Procrustes analysis and STATIS. *Food Quality Prefer* 2000, 11:77–83.
13. Qannari EM, Wakeling I, MacFie JH. A hierarchy of models for analyzing sensory data. *Food Quality Prefer* 1995, 6:309–314.
14. Smilde AK, Westerhuis JA, de Jong S. A framework for sequential multiblock component methods. *J Chem* 2003, 17:323–337.
15. Van Deun K, Smilde AK, van der Werf MJ, Kiers HAL, Van Mechelen IV. A structured overview of simultaneous component based data integration. *BMC Bioinform* 2009, 10:246–261.
16. Escoufier Y. L'analyse conjointe de plusieurs matrices de données. In: Jolivet M, ed. *Biométrie et Temps*. Paris: Société Française de Biométrie; 1980, 59–76.
17. L'Hermier des Plantes H. Structuration des Tableaux à Trois Indices de la Statistique, Thèse de troisième cycle. Université de Montpellier, 1976.
18. L'Hermier des Plantes H, Thiébaud E. Étude de la pluviosité au moyen de la méthode STATIS. *Revue Stat Appl* 1977, 25:57–81.
19. Lavit C. Application de la méthode STATIS. *Stat Anal Données* 1985, 10:103–116.
20. Lavit C. *Analyse Conjointe de Tableaux Quantitatifs*. Paris: Masson; 1988.
21. Lavit C, Escoufier Y, Sabatier R, Traissac P. The ACT (STATIS) method. *Comp Stat Data Anal* 1994, 18:97–119.
22. Korth B, Tucker LR. Procrustes matching by congruence coefficients. *Psychometrika* 1976, 41:531–535.
23. Génard M, Souty M, Holmes S, Reich M, Breuils L. Correlations among quality parameters of peach fruits. *J Sci Food Agric* 1994, 66:241–245.
24. Qannari EM, Wakeling I, Courcous P, MacFie JH. Defining the underlying dimensions. *Food Quality Prefer* 2000, 11:151–154.
25. Schlich P. Defining and validating assessor compromises about product distances and attribute correlations. In: Næs T, Risvik E, eds. *Multivariate Analysis of Data in Sensory Sciences*. New York: Elsevier; 1996, 229–306.
26. Chaya C, Perez-Hugalde C, Judez L, Wee CS, Guinard JX. Use of the STATIS method to analyze time-intensity profiling data. *Food Quality Prefer* 2003, 15:3–12.
27. Abdi H, Valentin D, Chollet S, Chrea C. Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications. *Food Quality Prefer* 2007, 18:627–640.

28. Abdi H, Valentin D. Some new and easy ways to describe, compare, and evaluate products and assessors. In: Valentin D, Nguyen DZ, Pelletier L, eds. *New Trends in Sensory Evaluation of Food and Non-Food Products*. Ho Chi Minh (Vietnam): Vietnam National University-Ho Chi Minh City Publishing House; 2007, 5–18.
29. Blancher G, Clavier B, Egoroff C, Duineveld K, Parcon J. A method to investigate the stability of a sorting map. *Food Quality Prefer* 2011, in press.
30. Gourvénec S, Stanimirova I, Saby CA, Airiau CY, Massart DL. Monitoring batch processes with the STATIS approach. *J Chem* 2005, 19:288–300.
31. Stanimirova I, Walczak B, Massart DL, Simeonov V, Sabyd CA, & di Crescenzo E. STATIS, a three-way method for data analysis: application to environmental data. *Chem Intell Lab Syst* 2004, 73:219–233.
32. Thioulouse J, Simier M, Chessel D. Simultaneous analysis of a sequence of paired ecological tables. *Ecology* 2004, 85:272–283.
33. Thioulouse J. Simultaneous analysis of a sequence of paired ecological tables: a comparison of several methods. *Ann Appl Stat* 2011, in press.
34. Abdi H, Valentin D, O'Toole AJ, Edelman B. DISTATIS: The analysis of multiple distance matrices, Proceedings of the IEEE Computer Society: International Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, 42–47.
35. Fournier M, Motelay-Massei A, Massei N, Aubert M, Bakalowiicz M, Dupond JP. Investigation of transport processes inside Karst aquifer by means of STATIS. *Ground Water* 2009, 47:391–400.
36. Gudmundsson L, Lena M, Tallaksen LM, Stahl K. Spatial cross-correlation patterns of European low, mean and high flows. *Hydrol Process* 2011, 25:1034–1045.
37. Enachescu C, Postelnicu T. Patterns in journal citation data revealed by exploratory multivariate analysis. *Scientometrics* 2003, 56:43–59.
38. Kherif F, Poline J-P, Mériaux S, Benali H, Flandin G, Brett M. Group analysis in functional neuroimaging: selecting subjects using similarity measures. *NeuroImage* 2003, 20:2197–2208.
39. Shinkareva SV, Ombao HC, Sutton BP, Mohanty A, Miller GA. Classification of functional brain images with a spatio-temporal dissimilarity map. *NeuroImage* 2006, 33:63–71.
40. Shinkareva SV, Mason RA, Malave VL, Wang W, Mitchell TM. Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE* 2008, 3:e1394. doi:10.1371/journal.pone.0001394.
41. Shinkareva SV, Malave SV, Just MA, Mitchell TM. Exploring commonalities across participants in the neural representation of objects. *Human Brain Map* 2011, in press.
42. Abdi H, Dunlop JP, Williams LJ. How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). *NeuroImage* 2009, 45:89–95.
43. Scepi G. Parametric and non parametric multivariate quality control charts. In: Lauro C, Antoch J, Esposito Vinzi V, Saporta G, eds. *Multivariate Total Quality Control*. Berlin: Physica-Verlag; 2002, 163–189.
44. Niang N, Fogliatto F, Saporta G. Contrôle multivarié de procédés par lots à l'aide de STATIS, Proceedings of the 41th “Journées de Statistique.” Bordeaux (France), 2009.
45. Coquet R, Troxler L, Wipff G. The STATIS method: Characterization of conformational states of flexible molecules from molecular dynamics simulation in solution. *J Mol Graph* 1996, 14:206–212.
46. Márquez EJ, Knowles LL. Correlated evolution of multivariate traits: detecting co-divergence across multiple dimensions. *J Evol Biol* 2007, 20:2334–2348.
47. Pavoine S, Bailly X. New analysis for consistency among markers in the study of genetic diversity: development and application to the description of bacterial diversity. *BMC Evol Biol* 2007, 7:156–172.
48. Raymond O, Fiasson JL, Jay M. Synthetic taxonomy of Rosa races using ACT-STATIS. *Zeitschrift Naturforschung C* 2000, 55:399–409.
49. Yanai H, Takeuchi K, Takane Y. *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. New York: Springer-Verlag; 2011.
50. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdiscip Rev: Comp Stat* 2010, 2:433–459.
51. Abdi H. Singular value decomposition (SVD) and generalized singular value decomposition (GSVD). In: Salkind N, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks CA: Sage; 2007, 907–912.
52. de Leeuw J. Derivatives of generalized eigensystems with applications. *UCLA Depart Stat Papers* 2007, 1–28.
53. Greenacre M. *Theory and Applications of Correspondence Analysis*. London: Academic Press; 1984.
54. Takane Y. Relationships among various kinds of eigenvalue and singular value decompositions. In: Yanai H, Okada A, Shigemasu K, Kano Y, Meulman J, eds. *New Developments in Psychometrics*. Tokyo: Springer-Verlag; 2002, 45–46.
55. Caillez F, Pagès JP. *Introduction à l'Analyse des Données*. Paris: SMASH; 1976.
56. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 2007, 22:1–20.
57. Escoufier Y. Operators related to a data matrix: a survey. *COMPSTAT: Proceedings in Computational Statistics; 17th Symposium Held in Rome, Italy, 2006*. New York: Physica Verlag; 2007, 285–297.

58. Holmes S. Multivariate analysis: the French way. In: Nolan D, Speed T, eds. *Festschrift for David Freedman*. Beachwood: IMS; 2006, 1–14.
59. de la Cruz O, Holmes S. The duality diagram in data analysis: examples of modern applications. *Ann Appl Stat* 2011, in press.
60. Sabatier R, Vivien M. A new linear method for analyzing four-way multiblocks tables: STATIS-4. *J Chem* 2008, 22:299–407.
61. Vivien M, Sune F. Two four-way multiblock methods used for comparing two consumer panels of children. *Food Quality Prefer* 2009, 20:472–481.
62. Rényi A. On measures of dependence. *Acta Math Hungar* 1959, 10:441–451.
63. Horn RA, Johnson, CR. *Matrix Analysis*. Cambridge: Cambridge University Press; 2006.
64. Escoufier Y. Le traitement des variables vectorielles. *Biometrics* 1973, 29:751–760.
65. Robert P, Escoufier Y. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl Stat* 1976, 25:257–265.
66. Abdi H. RV coefficient and congruence coefficient. In: Salkind NJ, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage; 2007, 849–853.
67. Abdi H. Congruence: Congruence coefficient, RV coefficient, and Mantel Coefficient. In: Salkind NJ, ed. *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage; 2010, 222–229.
68. Rencher AC. *Methods of Multivariate Analysis*. New York: John Wiley & Sons; 2002.
69. Greenacre M. *Biplots in Practice*. Barcelona: Fundación BBVA; 2010.
70. Gower JC, Lubbe S, le Roux N. *Understanding Biplots*. New York: John Wiley & Sons; 2011.
71. Abdi H, Valentin D. STATIS. In: Salkind NJ, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage; 2007, 955–962.
72. Gaertner JC, Chessel D, Bertrand J. Stability of spatial structures of dermersal assemblages: a multitable approach. *Aquatic Living Resour* 1998, 11:75–85.
73. Bhattacharyya A. On a measure of divergence between two multinomial populations. *Sankhya* 1941, 7:401–406.
74. Escoufier B. Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. *Revue Stat Appl* 1978, 26:29–37.
75. Domenges D, Volle M. Analyse factorielle sphérique: une exploration. *Ann l'INSEE* 1979, 35:3–83.
76. Rao CR. Use of Hellinger distance in graphical displays. In: Tiit E-M, Kollo T, Niemi H, eds. *Multivariate Statistics and Matrices in Statistics*. Leiden: Brill Academic Publisher; 1995, 143–161.
77. Abdi H. Distance. In: Salkind N, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage; 2007, 280–284.
78. Tucker LR. The extension of factor analysis to three-dimensional matrices. In: Frederiksen N, Gulliken H, eds. *Contributions to Mathematical Psychology*. New York: Holt; 1964, 110–182.
79. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J Chem* 1998, 12:301–321.
80. Næs T, Brockhoff PB, Tomic O. *Statistics for Sensory and Consumer Science*. London: John Wiley & Sons; 2010.
81. Escoufier B, Pagès J. *Analyses Factorielles Simples et Multiples: Objectifs, Méthodes, Interprétation*. Paris: Dunod; 1990.
82. Abdi H, Valentin D. Multiple factor analysis. In: Salkind NJ, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage; 2007, 657–663.
83. Lebart L, Piron M, Morineau A. *Statistique Exploratoire Multidimensionnelle: Visualisations et Inférences en Fouille de Données*. Paris: Dunod; 2006.
84. Gower JC. Adding a point to vector diagrams in multivariate analysis. *Biometrika* 1968, 55:582–585.
85. Lazraq A, Hanafi M, Cléroux R, Allaire J, Lepage Y. Une approche inférentielle pour la validation du compromis de la méthode STATIS. *JSoc Francaise Stat* 2008, 149:97–109.
86. Areia A, Oliveira MM, Mexia JT. Models for a series of studies based on geometrical representation. *Stat Methodol* 2008, 5:277–288.
87. Oliveira MM, Mexia JT. Modelling series of studies with a common structure. *Comp Stat Data Anal* 2007, 51:5876–5885.
88. Oliveira MM, Mexia JT. ANOVA-like analysis of matched series of studies with a common structure. *J Stat Plan Infer* 2007, 137:1862–1870.
89. Good P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. New York: Springer-Verlag; 2005.
90. Abdi H, Williams LJ. Jackknife. In: Salkind NJ, ed. *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage; 2010, 655–660.
91. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman and Hall; 1993.
92. Chernick MR. *Bootstrap Methods: A Guide for Practitioners and Researchers*. New York: John Wiley & Sons; 2008.
93. Milan M. Applications of the parametric bootstrap to models that incorporate a singular value decomposition. *Appl Stat* 1995, 44:31–49.
94. Lebart L. Which Bootstrap for principal axes methods? *Selected Contributions in Data Analysis and Classification. Studies in Classification, Data Analysis, and Knowledge Organization*; 2007, 581–588.
95. Abdi H. The Bonferonni and Sidak corrections for multiple comparisons. In: Salkind N, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage; 2007, 104–107.

96. Jaffrenou PA. *Sur l'Analyse des Familles Finies de Variables Vectorielles: Bases Algébriques et Applications à la Description Statistique*, Thèse de Troisième Cycle. Université de Lyon, 1978.
97. Thioulouse J, Chessel D. Les analyses multitableaux en écologie factorielle. I. De la théorie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Oecol, Oecol General* 1987, 8:463–480.
98. Simier M, Blanc L, Pellegrin F, Nandris D. Approche simultanée de K couples de tableaux: application à l'étude ds relations pathologie végétale-environnement. *Revue Stat Appl* 1999, 47:31–46.
99. Rolland A, Bertrand F, Maumy M, Jacquet S. Assessing phytoplankton structure and spatio-temporal dynamics in a freshwater ecosystem using a powerful multiway statistical analysis. *Water Res* 2009, 43:3155–3168.
100. Bertrand F, Maumy M. Using partial triadic analysis for depicting the temporal evolution of spatial structures: assessing phytoplankton structure and succession in a water reservoir. *Case Studies Business, Industry Govern Stat* 2010, 4:23–43.
101. Mendes S, Gómez JF, Pereira MJ, Azeiteiro UM, Galindo-Villardón MP. The efficiency of the partial triadic analysis methods: an ecological application. *Biometr Lett* 2010, 47:83–106.
102. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966, 31:279–311.
103. Kiers HAL. Hierarchical relations among three-way methods. *Psychometrika* 1991, 56:449–470.
104. Tucker LR. An inter-battery method of factor analysis. *Psychometrika* 1958, 23:111–136.
105. Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage* 2011, 56:455–475.
106. McIntosh AR, Lobaugh NJ. Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage* 2004, 23:250–263.
107. Tucker LR, Messick S. An individual difference model for multidimensional scaling. *Psychometrika* 1963, 28:333–367.
108. Horst P. *Factor Analysis of Data Matrices*. New York: Holt; 1965.
109. Ballester J, Abdi H, Langlois J, Peyron D, Valentin D. The odor of colors: can wine experts and novices distinguish the odors of white, red, and rosé wines? *Chem Percep* 2009, 2:203–213.
110. Chollet S, Lelièvre M, Abdi H, Valentin D. Sort and Beer: Everything you wanted to know about the sorting task but did not dare to ask. *Food Quality Prefer* 2011, 22:507–520.
111. Santosa M, Abdi H, Guinard JX. A modified sorting task to investigate consumer perceptions of extra virgin olive oils. *Food Quality Prefer* 2010, 21:881–892.
112. Lelièvre M, Chollet S, Abdi H, Valentin D. Beer trained and untrained assessors rely more on vision than on taste when they categorize beers. *Chem Percep* 2009, 2:143–153.
113. Churchill N, Oder A, Abdi H, Tam F, Lee W, Thomas C, Graham S, Strother S. Optimizing correction of head motion and physiological noise in single-subject fMRI analyses: 1. Standard temporal-based methods. *Human Brain Mapp* 2011, in press.
114. Caspi Y, Axelrod A, Matsushita Y, Gamliel A. Dynamic stills and clip trailers. *Visual Comp* 2006, 22:642–652.
115. Young G, Householder AS. Discussion of a set of points in terms of their mutual distances. *Psychometrika* 1938, 3:19–22.
116. Torgerson W. *Theory and Methods of Scaling*. New York: John Wiley & Sons; 1958.
117. Abdi H. Metric multidimensional scaling. In: Salkind N, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage; 2007, 598–605.
118. Vallejo-Arboleda A, Vicente-Villardón JL, Galindo-Villardón MP. Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology. *Comp Stat Data Anal* 2007, 51:4193–4205.
119. Abdi H, Williams LJ. Barycentric discriminant analysis (BADIA). In: Salkind NJ, ed. *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage; 2010, 64–75.
120. Lachenbruch PA. *Discriminant Analysis*. New York: Macmillan; 1975.
121. Pelé J, Abdi H, Moreau M, Thybert D, Chabbert M. Multidimensional scaling reveals the main evolutionary pathways of class A G-protein-coupled receptors. *PLoS One* 2011, 6:1–10. doi:10.1371/journal.pone.0019094.
122. Rao CR, Mitra SK. *Generalized Inverse of Matrices and its Applications*. New York: John Wiley & Sons; 1971.
123. Bénasséni J, Bennani Dosse M. Analyzing multiset data by the power STATIS-ACT method. *Adv Data Anal Classif* 2011, 5, in press.
124. Bennani Dosse M, Groenen P, Abdi H. Anisotropic STATIS, submitted.
125. Sauzay L, Hanafi M, Qannari EM, Schlich P. Analyse de $K + 1$ tableaux à l'aide de la méthode STATIS: application en évaluation sensorielle, 9^{ième} Journées Européennes Agro-industrie et Méthodes Statistiques. Montpellier (France). 2006, 1–23.
126. Abdi H. Partial least square regression, projection on latent structure regression, PLS-Regression. *Wiley Interdiscip Rev: Comp Stat* 2010, 2:97–106.
127. Vivien M, Sabatier R. A generalization of STATIS-ACT strategy: DO-ACT for two multiblocks tables. *Comp Stat Data Anal* 2004, 46:155–171.

128. Horst P. Generalized canonical correlations and their applications to experimental data. *J Clin Psychol* 1961, 17:331–347.
129. Takane Y, Yanai M, Hwang H. An improved method for generalized constrained canonical correlation analysis. *Comp Stat Data Anal* 2006, 50:221–241.
130. Takane Y, Hwang H, Abdi H. Regularized multiple set canonical correlation analysis. *Psychometrika* 2008, 73:753–775.
131. Gower JC, Dijksterhuis GB. *Procrustes Problems*. Oxford: Oxford University Press; 2004.
132. Pagès J. Éléments de comparaison entre l'analyse factorielle multiple et la méthode STATIS. *Revue Stat Appl* 1996, 44:81–95.
133. Husson F, Pagès J. INDSCAL model: geometrical interpretation and methodology. *Comp Stat Data Anal* 2006, 50:358–378.
134. Dray S, Dufour AB, Chessel D. The ade4 package-II: two-table and K-table methods. *R News* 2007, 7: 47–52.
135. Harville DA. *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag; 2008.
136. Gentle JE. *Matrix Algebra: Theory, Computation and Applications in Statistics*. New York: Springer-Verlag; 2007.
137. Mkhadri A, Celeux G, Nasroallah A. Regularization in discriminant analysis: an overview. *Comp Stat Data Anal* 1997, 23:403–423.