University of Abomey-Calavi (Benin)
Faculty of Agronomic Sciences
Laboratory of Biomathematics and Forest Estimations

Application of common components and specific weights method to analyze local perception patterns of land degradation in northern Benin (West Africa)

A thesis submitted to the Faculty of Agronomic Sciences for the degree of Master of science in Biostatistics

Presented and defended by:
Essomanda TCHANDAO MANGAMANA

On February 26^{th}, 2016.

Supervisor: **Romain GLELE KAKAÏ**

Jury Members:

Chairman: **Pr. Léonard TODJIHOUNDE**

Reporter: **Pr. Romain L. GLELE KAKAÏ**

Examiner: **Dr.(MC) Aboubacar MARCOS**

Academic year: 2015-2016

Université d'Abomey-Calavi (Bénin)
Faculté des Sciences Agronomiques
Laboratoire de Biomathematiques & d'Estimations Forestières

Application de l'analyse en composantes communes et poids spécifiques pour évaluer la perception locale de la dégradation du sol au nord Bénin (Afrique de l'Ouest)

Mémoire soumis à la Faculté des Sciences Agronomiques pour l'obtention du diplôme de Master de recherche en Biostatistiques

Presenté et soutenu par:
Essomanda TCHANDAO MANGAMANA
Le 26 Février 2016.

Superviseur: **Romain GLELE KAKAÏ**

Membres de Jury:
Président: **Pr. Léonard TODJIHOUNDE**
Rapporteur: **Pr. Romain L. GLELE KAKAÏ**
Examinateur: **Dr.(MC) Aboubacar MARCOS**

Année académique: 2015-2016

Certification

I certify that this thesis has been written by Essomanda TCHANDAO MANGAMANA under my supervision at the Faculty of Agronomic Sciences of the University of Abomey-Calavi (Bénin) to obtain his master of sciences degree in Biostatistics.

Prof. Dr. Ir. Romain GLELE KAKAÏ,
Full Professor in Biostatistics and Forest Estimations.

Abstract

Common components and specific weights analysis (CCSWA) is a relatively recent multiblock statistical method that constitutes an extension of principal components analysis (PCA) in the case where different sets of quantitative variables have been measured on the same set of individuals. We described in this thesis the principle of CCSWA and its application in R software on real data to analyze farmers' perception of land degradation and soil erosion in northern Benin (West Africa). The data considered bear on 5 sociocultural groups and variables are linked to the causes of land degradation (dataset 1), soil erosion factors (dataset 2), land use practices against soil erosion (dataset 3) and techniques of improvement of the soil fertility and crops productivity (dataset 4). On these datasets, we also applied PCA in order to show the improvement of CCSWA compared to PCA. The results of CCSWA showed that the common component q_1, opposing Djerma to Haussa farmers according to local perception of land degradation and soil erosion, expressed 60.4 %, 45.3 %, 10 % and 73.5 % of the total inertia of datasets 1, 2, 3 and 4 respectively. Djerma farmers think that land degradation is due to erosion, agricultural settlement and wildfire. Run-off and slope are the main soil erosion factors according to them. They also think that crops productivity can be enhanced by using plows and carts. Regarding Haussa farmers, deforestation is the main cause of land degradation, whereas the soil type is the main soil erosion factor. Against soil erosion, they set up stony lines and use manure and household rubbishes to improve the soil fertility and crops productivity. The common component q_2 explained 5.4 %, 30.8 %, 70 % and 9.4 % of the total inertia contained in datasets 1, 2, 3 and 4 respectively and opposed Dendi to Djerma farmers about local perception. Dendi farmers acknowledge animal stamping and soil type as the main soil erosion factors and practice fallow to improve the soil fertility and crops productivity. As regards Djerma farmers, they cover their lands and till orthogonally to the normal flow of water in order to overcome soil erosion. Globally, the results of CCSWA and PCA are almost the same but the improvement that CCSWA brings is the knowledge of how different datasets cooperate to form the common components.

Keywords: CCSWA, PCA, multivariate analysis, multiblock analysis, perception, land degradation.

Résumé

L'Analyse en composantes communes et poids spécifiques (ACCPS) est une méthode multiblock relativement récente qui constitue une extension de l'analyse en composantes principales (ACP) au cas où plusieurs groupes de variables quantitatives ont été mesurées sur le même groupe d'individus. Nous avons décris dans ce mémoire le principe de l'ACCPS et son application sur des données réelles dans le logiciel R afin d'évaluer la perception des paysans sur la dégradation du sol et les facteurs de l'érosion au nord Bénin (Afrique de l'Ouest). Les données considérées portent sur les paysans de 5 groupes ethniques dont les variables sont relatives aux causes de la dégradation du sol (tableau 1), aux facteurs de l'érosion (tableau 2), aux mesures adaptatives face à l'érosion (tableau 3) et aux techniques d'amélioration de la fertilité du sol et du rendement des cultures (tableau 4). Aussi, avons-nous appliqué l'ACP sur ces données afin de montrer l'amélioration qu'apporte l'ACCPS comparativement à l'ACP. Les résultats de l'ACCPS ont montré que la composante commune q_1 expliquant 60.4 %, 45.3 %, 10 % et 73.5 % de l'inertie totale des tableaux 1, 2, 3 et 4 respectivement, oppose les paysans Djerma aux paysans Haussa selon la perception locale de la dégradation du sol et de l'érosion. Les paysans Djerma pensent que la dégradation du sol est due à l'érosion, à l'installation des champs suite à une déforestation et aux feux de brousses. Pour eux, le ruissellement et la pente sont les principaux facteurs de l'érosion. Aussi, pensent-ils que l'usage des charrues et des charrettes pourrait accroître le rendement des cultures. A l'entendement des paysans Haussa, la déforestation constitue la principale cause de la dégradation du sol. Selon eux, le type de sol favorise l'érosion. Afin de lutter contre celle-ci, ils placent des cordons pierreux et utilisent le fumier et les ordures ménagères pour accroître la fertilité du sol et le rendement des cultures. La composante commune q_2 a expliqué 5.4 %, 30.8 %, 70 % et 9.4 % de l'inertie totale des tableaux 1, 2, 3 et 4 respectivement et a opposé les paysans Dendi aux paysans Djerma sur la perception locale. Les paysans Dendi pensent que le piétinement et le type de sol sont les principaux facteurs de l'érosion et font la jachère pour accroître la fertilité du sol et le rendement des cultures. Quant aux paysans Djerma, ils couvrent le sol et labourent perpendiculairement au flux de l'eau afin de lutter contre l'érosion. Globalement, les résultats de l'ACCPS et de l'ACP se recoupent dans une large mesure. Mais l'amélioration qu'apporte l'ACCPS réside dans le fait que nous avons une parfaite connaissance de la manière dont les différents tableaux concourent à déterminer les composantes communes.

Mots-clés: ACCPS, ACP, analyse multivariée, analyse multiblock, perception, dégradation du sol.

Acknowledgements

I am deeply grateful to my supervisor Prof. Romain GLELE KAKAÏ for his patience and helpful advice during this training. I would like to express my sincere gratitude to Prof. El Mostafa QANNARI (from the University of Nantes, ONIRIS, France) for his assistance and prompt service during this thesis writing. My thanks go to Prof. Mohamed HANAFI (from the University of Nantes, ONIRIS, France) for his help. I am grateful to my parents who tirelessly supported me during this master training and my classmates for their assistance. I thank God for giving me health to attain this training.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Problematic and objectives

Principal components analysis (PCA) is a multivariate statistical method that is used to 1) assess the proximity between individuals, 2) assess the variables on which are based these proximities and 3) describe links between variables of a sample. It constructs few principal components that are linear combination of initial variables and takes into account the maximum of inertia contained in the initial dataset.

PCA is applicable to only one dataset. However, in some situations a researcher may need to collect not only one, but at least two multivariate datasets of quantitative variables bearing on the same individuals. The number of variables of these datasets can be different. For the exploration of these datasets, several methods and approaches have been proposed, but lead sometimes to unsatisfactory results. For example, although some multidimensional methods such as PCA, factorial discriminant analysis and principal component regression are said to be highly efficient for single dataset analysis, they cannot deal simultaneously with several datasets (Kulmyrzaev and Dufour, 2010). Some authors propose to apply a multidimensional analysis on each dataset and thereafter pool the conclusions obtained from each analysis (Beuvier *et al.*, 1997, Di Cagno *et al.*, 2003, Hanafi *et al.*, 2006). Others propose to combine all the datasets into one dataset before running a multidimensional analysis (Karoui *et al.*, 2004, 2006). Although these approaches can give interesting results, they cannot assess the relationships between all the datasets (Schoojans and Massart, 2001, Mazerolles *et al.*, 2006).

The appropriate methods that can be used to analyze several sets (blocks) of variables observed on the same set of individuals are multiblock methods of data analysis. Some of them are: common components and specific weights analysis, multiple co- inertia analysis (MCoA), consensus principal components analysis (CPCA) also known as multiblock principal components analysis (MbPCA), hierarchical principal components analysis (HPCA) and multiple factor analysis (MFA). Among these alternatives, CCSWA is our main interest in this thesis. However, in situations where the same set of variables has been recorded on different sets of individuals, one could refer to dual common components and specific weights analysis, multigroup principal components analysis, Flury's common principal components analysis from the "multigroup" package and dual multiple factor analysis from the "FactoMineR" package in R software to analyze these data.

CCSWA is an improvement of the algorithm that determines the parameters of the 3^{rd} model (named "common underlying dimensions, differentially weighted") of the hierarchy of 3 models developed by Qannari *et al.* (1995) to analyze sensory data. These models were based on Flury's common principal components theory (1988) but using association matrices instead of covariance matrices. Moreover, unlike common principal components analysis, which requires the assumptions of normality and maximum likelihood estimation, these models require no distributional assumption (Qannari *et al.*, 1995). So, CCSWA was originally introduced and applied to analyze sensory data (Qannari *et al.*, 2000, 2001). But very quickly, its use was extended to analyze other kind of data in other fields. Indeed, Courcoux *et al.* (2002) used CCSWA to analyze multispectral image data. Mazerolles *et al.* (2002, 2006), Hanafi *et al.* (2006) applied CCSWA to the coupling of several measurements techniques: infrared spectroscopy, fluorescence spectroscopy, rheological analysis and chemical analysis on cheese to cope with food complexity. Karoui *et al.* (2006) coupled the front face fluorescence spectroscopy with CCSWA to study the structure of cheeses at the molecular level throughout ripening by following changes affecting proteins, fats, interactions in the matrix cheese during ripening. Blackman *et al.* (2010) applied CCSWA on analytical measurements and sensory data of wines to show the potential for using analytical measurements as a surrogate for sensory analysis. The main objective of this thesis was to describe the principle of CCSWA and use it to analyze farmers' perception of land degradation and soil erosion by considering a case study in northern Benin (West Africa).

Land degradation is one of the most serious problems currently affecting agricultural productivity in developing countries of the tropics (Akinnagbe and Umukoro, 2011). And farmers are those who can give more information about this issue. Therefore, Chizana *et al.* (2011) examined farmers' perception, understanding and interpretation of soil erosion factors and indicators and how they relate to land degradation and soil fertility decline in Zimbabwe. Avakoudjo *et al.* (2011) used farmers' perceptions to identify the main soil erosion causes and factors, to improve the knowledge on the "dongas" phenomenon in the W National Park of Benin and its surrounding areas. These aforementioned authors often use descriptive statistics, except Avakoudjo *et al.* (2011), who also used PCA to analyze their data. The following specific objectives were considered:

- describe the principle of CCSWA using sensory data of wines
- use CCSWA to analyze farmers' perception of land degradation and soil erosion
- assess the relative performance of PCA and CCSWA in describing local perception of land degradation.

CHAPTER 2

Principle

2.1 Description of sensory data of wines

The datasets used to present the principle of CCSWA are related to the evaluation of eight (8) wines appearance by four (4) expert tasters (Table 2.1). A jury made up of four judges evaluated the appearance of eight wines according to the procedure known as free profile, where each judge notes on a scale from 0 to 10 the products according to his/her own variables (Williams and Langron, 1984, Hanafi and Kiers, 2006, Hanafi and Qannari, 2008, Kissita *et al.*, 2009). For a product having a given variable, the note allotted by a judge corresponds to the intensity which he/she perceives and which he/she is able, thanks to a preliminary drive, to translate in form of a note. Each dataset is associated with one judge. The goal of the analysis is to evaluate if there is an agreement between judges or groups of judges and assess the relationships among products.

Table 2.1: Evaluation of the appearance of 8 wines by 4 judges

(a) Judge 1 (X_1)

	red	gilded	soft	plum
Wine 1	7	0	5	8
Wine 2	5	6	6	3
Wine 3	7	2	5	5
Wine 4	5	7	7	4
Wine 5	5	7	6	4
Wine 6	6	8	6	1
Wine 7	5	4	10	3
Wine 8	6	6	6	5

(b) Judge 2 (X_2)

	ruby	coloured	intensity
Wine 1	4	0	5
Wine 2	3	6	5
Wine 3	3	3	7
Wine 4	1	6	3
Wine 5	2	5	5
Wine 6	1	5	4
Wine 7	0	4	2
Wine 8	2	6	4

(c) Judge 3 (X_3)

	red	blue	gilded	intensity
Wine 1	7	4	2	6
Wine 2	2	0	6	6
Wine 3	6	3	4	7
Wine 4	2	0	6	4
Wine 5	5	1	5	6
Wine 6	3	0	5	5
Wine 7	2	0	4	3
Wine 8	4	0	4	5

(d) Judge 4 (X_4)

	deep	expenses	brilliant
Wine 1	9	7	9
Wine 2	8	6	7
Wine 3	10	6	7
Wine 4	7	7	8
Wine 5	8	7	8
Wine 6	8	8	10
Wine 7	6	5	10
Wine 8	8	9	10

2.2 Computation of common components and specific weights

Let us consider m datasets as raw data. Each dataset denoted by Y_k is a rectangular matrix nxp_k, where n is the number of individuals and p_k, the number of variables of the k^{th} dataset (k=1, 2, ..., m). From these raw data, many steps are used to perform the common components and specific weights analysis. The first one consists of preprocessing each dataset (centering each variable, then normalizing each dataset). In the second step, we compute the scalar product matrix associated with each dataset. Then, in the third step, we apply an algorithm on these scalar product matrices, in order to estimate the common components and the specific weights. The fourth step is related to the computation of a compromise matrix.

For our example m=4 datasets, the number of individuals (wines) in each dataset is equal to n=8 and the number of variables in the first, second, third and fourth dataset is respectively: $p_1 = 4$, $p_2 = 4$, $p_3 = 3$, $p_4 = 7$. These datasets are shown in Table 2.1. The first dataset is composed of:

$$Y_1 = \begin{bmatrix} 7 & 0 & 5 & 8 \\ 5 & 6 & 6 & 3 \\ 7 & 2 & 5 & 5 \\ 5 & 7 & 7 & 4 \\ 5 & 7 & 6 & 4 \\ 6 & 8 & 6 & 1 \\ 5 & 4 & 10 & 3 \\ 6 & 6 & 6 & 9 \end{bmatrix} \tag{2.1}$$

Step 1: Center each variable. Then if it is useful, normalize each dataset to unit norm before starting the computation of the common components and specific weights.

The data centering aims at removing the irrelevant differences among individuals and making the results interpretation easy. To center a variable, we subtract from its entries, the mean value of the considered variable. For a variable Y_{kj}, the mean is computed as:

$$\bar{Y}_{kj} = \frac{\sum_{i=1}^{n} y_{kji}}{n}$$

n is the number of individuals, k is the dataset index (k=1, 2, ..., m), i is the row index and j, the column index ($j = 1, 2, \ldots, p_k$).

For our example, the mean value of the first variable of Y_1 is:

$$\bar{Y}_{11} = \frac{\sum_{i=1}^{8} y_{11i}}{8} = 5.750$$

In the same way, the mean value of the second, third and fourth variable of Y_1 is respectively equal to: 5.000, 6.375 and 4.125.

Thus, the centered dataset of Y_1 is:

$$Y_1 = \begin{bmatrix} 1.25 & -5 & -1.375 & 3.875 \\ -0.75 & 1 & -0.375 & -1.125 \\ 1.25 & -3 & -1.375 & 0.875 \\ -0.75 & 2 & 0.625 & -0.125 \\ -0.75 & 2 & -0.375 & -0.125 \\ 0.25 & 3 & -0.375 & -3.125 \\ -0.75 & -1 & 3.625 & -1.125 \\ 0.25 & 1 & -0.375 & 0.875 \end{bmatrix} \tag{2.2}$$

The first row of Y_1 is given by: 1.25=7-5.750; -5=0-5; -1.375=5-6.375; 3.875=8-4.125.
In R software, the "scale" function of the base package can be used to center variables (columns) of a matrix: scale (Y_1, center=TRUE, scale=FALSE), where Y_1 is a numeric matrix. If the scale argument of the "scale" function is set to "TRUE", then each column is divided by its standard deviation. But since we don't need it, we set this argument to "FALSE".
The datasets normalization is specific to multiblock methods and corrects irrelevant differences that can exist between datasets (difference in size or in variance). The normalization is optional, but when this is done, we acquire an advantage for the results interpretation as we will see it later in step 3d. The norm considered is the Frobenius one (also called Euclidean norm):

$$\|Y_k\|_2 = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{p_k} y_{kij}^2}$$

.
n is number of individuals, k is the dataset index (k=1, 2, ..., m) and p_k, the number of variables in the k^{th} dataset.
For our example, the norm of the first centered dataset is:

$$\|Y_1\|_2 = \sqrt{\sum_{i=1}^{8}\sum_{j=1}^{4} y_{1ij}^2} = (1.25)^2 + (-5)^2 + (-1.375)^2 + (3.875)^2 + \ldots + (0.875)^2 = 10.308$$

Similarly, the norm of the second, third and fourth centered dataset is obtained in the same way. They are respectively: 7.599, 8.269 and 5.723. The norm of Y_1 can be computed in R software using the "norm" function of the base package: norm (Y_1 , "f"), where "f" specifies the Frobenius norm. But "F" can be used instead of "f".

Denote by $X_k \ \ (k = 1, \ 2, \ \ldots, \ m)$, the centered and normalized dataset of $Y_k \ \ (k = 1, \ 2, \ \ldots, \ m)$.
For our example, the centered and normalized dataset of Y_1 is:

$$X_1 = \begin{bmatrix} 0.121 & -0.485 & -0.133 & 0.376 \\ -0.073 & 0.097 & -0.036 & -0.109 \\ 0.121 & -0.291 & -0.133 & 0.085 \\ -0.073 & 0.194 & 0.061 & -0.012 \\ -0.073 & 0.194 & -0.036 & -0.012 \\ 0.024 & 0.291 & -0.036 & -0.303 \\ -0.073 & -0.097 & 0.351 & -0.109 \\ 0.024 & 0.097 & -0.036 & 0.085 \end{bmatrix} \tag{2.3}$$

The first row of X_1 is obtained by doing the following computations:
$0.121 = \frac{1.25}{10.308}, -0.485 = \frac{-5}{10.308}, -0.133 = \frac{-1.375}{10.308}$ and $0.376 = \frac{3.875}{10.308}$.
The remaining rows are obtained in the same way. Instead of dividing each centered value of a dataset by the norm of the considered dataset, one can directly use in R software, the function "normM" of the multigroup package on the centered datasets to normalize them.

Step 2: Start the computation of the common components and specific weights by computing for each dataset, the scalar product matrix:

$$W_k \quad = \quad X_k X_k' \tag{2.4}$$

where X_k' is the transpose matrix of X_k (the first column of X_k becomes the first row of X_k' , the second column of X_k becomes the second row of X_k' and so on); k=1, 2, ..., m. $W_k \ \ (k = 1, \ 2, \ \ldots, \ m)$ is a (n, n) square symmetric matrix, n being the number of individuals. Since variables are centered, the diagonal elements of $W_k \ \ (k = 1, \ 2, \ \ldots, \ m)$ are the squared distances of individuals from the origin and its off-diagonal elements are scalar products between individuals (which are quantities proportional to the cosine between individuals). Each scalar product matrix expresses similarities among individuals (Qannari *et al.*, 1995).
For our example, as n=8, we will obtain (8, 8) scalar product matrices. The one associated with X_1 is given by:

$$W_1 = \begin{bmatrix} 0.409 & -0.092 & 0.206 & -0.116 & -0.103 & -0.247 & -0.050 & -0.007 \\ -0.092 & 0.028 & -0.041 & 0.023 & 0.027 & 0.061 & -0.05 & 0.000 \\ 0.206 & -0.041 & 0.124 & -0.074 & -0.061 & -0.103 & -0.037 & -0.013 \\ -0.116 & 0.023 & -0.074 & 0.047 & 0.041 & 0.056 & 0.009 & 0.014 \\ -0.103 & 0.027 & -0.061 & 0.041 & 0.044 & 0.060 & -0.025 & 0.017 \\ -0.247 & 0.061 & -0.103 & 0.056 & 0.060 & 0.179 & -0.010 & 0.004 \\ -0.050 & -0.005 & -0.037 & 0.009 & -0.025 & -0.010 & 0.150 & -0.033 \\ -0.007 & 0.000 & -0.013 & 0.014 & 0.017 & 0.004 & -0.033 & 0.019 \end{bmatrix} \quad (2.5)$$

CCSWA model stipulates the existence of common components to all the datasets but the weights (saliences) of each dataset on these common components can be different. This weighting difference can be explained by the presence of information in some datasets but not in others. Thus, the model of CCSWA can be written as:

$$W_k \quad = \quad Q\Lambda_k Q' + E_k = \sum_{r=1}^{n-1} \lambda_r^k q_r q_r' + E_k \quad (2.6)$$

where n is the number of individuals, Q is an orthogonal matrix whose columns $q_1,\ q_2,\ \ldots,\ q_{n-1}$ are the common components, Λ_k is a diagonal matrix whose diagonal elements $\lambda_1^{(k)}, \lambda_2^{(k)}, \ldots, \lambda_{n-1}^{(k)}$ are the specific weights (saliences) associated with the common components and E_k, the residual matrix of the dataset X_k (k=1, 2, …, m).

It can be noticed from (2.6) that the number of common components is n-1. This is because we assumed that the variables are centered and the number of individuals is less than the total number of variables in all the datasets. But, in general, if n is the number of individuals and p, the total number of variables of all the datasets ($p = p_1 + p_2 + \ldots + p_k,\ k = 1,\ 2,\ \ldots,\ m$), then the number of dimensions (common components) is at most min (n, p). Moreover, when variables are centered, the number of independent dimensions is reduced from 1 (Kroonenberg, 2007, Hanafi and Qannari, 2008). Hence, if n < p and the variables are centered, the number of dimensions is at most n-1.

For our example, since n=8, p=4+3+4+3=14; min (n, p) = 8 and variables are centered, the number of common components is at most n-1=7.

The parameters to be estimated from the model of CCSWA (equation (2.6)) are the common components q_r (r=1, 2, …, n-1) and the specific weights $\lambda_r^{(k)}$ (k=1, 2, …, m; r=1, 2, …, n-1).

Step 3: Choose an algorithm to estimate the parameters of CCSWA.
Three algorithms can be used to estimate these parameters: the iterative algorithm of Qannari *et al.* (2000), the pseudo simultaneous and the simultaneous algorithm of Kissita *et al.* (2009). The difference between the pseudo simultaneous and the simultaneous algorithm is the objective function to be maximized. However, in a case study involving all these three algorithms, Kissita *et al.* (2009) found similar results. Thus, we present in the frame of this thesis, the iterative algorithm of Qannari *et al.* (2000).

Step 3a: Initialize the weights $\lambda_1^{(k)} (k = 1,\ 2,\ ...,\ m)$ to 1. Or choose m positive values, instead of considering unit weights.

Step 3b: Extract the vector q_1 given by the normed eigenvector of:

$$W = \sum_{k=1}^{m} \lambda_1^{(k)} W_k \qquad (k = 1,\ 2,\ ...,\ m) \tag{2.7}$$

associated with the largest eigenvalue.
Eigenvalues and eigenvectors can be extracted by means of a PCA on W. To perform the PCA in R software, two methods can be used: the spectral decomposition (also known as eigen-decomposition) and the singular value decomposition (SVD). The spectral decomposition method is defined for square matrices (matrices in which the number of rows is equal to the number of columns), whereas the SVD method works even with rectangular matrices (matrices in which the number of rows is less than the number of columns or vice versa). These two methods are simple to be implemented in R software but when there is a choice among them, the SVD method is preferred for numerical accuracy (R Development Core Team 2011).In R software, there are several functions from different packages that allow us to perform PCA but they give almost the same results. Hence, for our example, we used the SVD method with the ”svd” function of the base package. The important arguments of that function are: the matrix on which we would like to perform the SVD and the number of left and right singular vectors to be computed. The left singular vectors of a matrix W are given by the eigenvectors of W W’ whereas the right singular vectors are given by the eigenvectors of W’ W (W’ being the transpose matrix of W). The returned values of a SVD are: a matrix U that contains the left singular vectors, a diagonal matrix D that contains the singular values and a matrix V that contains the right singular vectors. The matrices U and V are orthonormal because U’ U=V’ V=I, where I is the identity matrix (a matrix that contains only ones on its diagonal and zeros on its off-diagonal).

The singular values are sorted in a decreasing order. The first singular vector is associated with the largest singular value; the second singular vector is associated with the second largest singular value and so on. Since the SVD is performed on a square symmetric matrix, we noted that the first left singular vector is always the same as the first right singular vector.
As the singular values are positive, squaring them to obtain the eigenvalues will not change the order of the associated singular vectors. Consequently, retaining the vector q_1 associated with the largest eigenvalue of W is equivalent to retain the singular vector (either the left or right singular vector because they are the same) associated with the largest singular value. The vector q_1 is called "common component". Its sign and length are arbitrary. But as the common components are constrained to be orthogonal, it is common to normalize them to unit length. This operation consists of dividing each element of q_1 by its length (norm). But we should not wonder about this because by default vectors are already normalized, when performing the SVD method in R software.

Step 3c: Update the previous weights by:

$$\lambda_1^{(k)} = q_1' W_k q_1 \tag{2.8}$$

In this formula: q_1 is the common component computed in step 3b, q_1' is its transpose, W_k is the scalar product matrix associated with the dataset X_k (k=1, 2, ..., m) and $\lambda_1^{(k)}$ is the weight of each dataset X_k (k=1, 2, ..., m) in the computation of the common component q_1.

Step 3d: Evaluate the loss function as:

$$L_1 = \sum_{k=1}^{m} \|W_k - \lambda_1^{(k)} q_1 q_1'\|^2$$

$$L_1 = \sum_{k=1}^{m} \|W_k\|^2 - 2\sum_{k=1}^{m} \lambda_1^{(k)} trace(W_k q_1 q_1') + \sum_{k=1}^{m} (\lambda_1^{(k)})^2$$

But $\lambda_1^{(k)}$= trace $(W_k q_1 q_1')$
Hence:

$$L_1 = \sum_{k=1}^{m} \|W_k\|^2 - \sum_{k=1}^{m} (\lambda_1^{(k)})^2 \tag{2.9}$$

In L_1,$\sum_{k=1}^{m} \|W_k\|^2$ is the total inertia contained in all the datasets, whereas $\sum_{k=1}^{m}(\lambda_1^{(k)})^2$ is the total inertia of all the datasets explained by q_1. So L_1 is the total inertia of all the

datasets not explained by q_1.
At this level, the user can fix a threshold ϵ in order to break the iterative loop. The default value used is $\epsilon = 10^{-10}$ (Qannari *et al.*, 2000).

- If $L_1 < \epsilon$, the computation of the common component q_1 is completed; we reach the convergence and the algorithm stops. Thus, the common component q_1 is given by the common component that was extracted in step 3b and the specific weight of each dataset X_k $(k = 1,\ 2, \ldots,\ m)$ in the computation of this common component q_1 is given by the weights computed in step 3c.

- If $L_1 \geq \epsilon$, the algorithm starts from step 3b but instead of using the unit weight for each dataset, we consider the weights that were computed in step 3c and we reiterate this algorithm until the convergence.

Since the datasets are normalized, the salience of a dataset for a given common component is the percentage of the total inertia of that dataset explained by the considered common component (here is the advantage of dataset normalization for results interpretation that we stated above in step 1). A high salience of a given dataset and for a given common component, a great importance this dataset has for that common component. The saliences are always positive or null. When the salience of a dataset in the computation of a given common component is null, this means that the considered dataset is not underlying to that common component. And therefore, the spelling «common components» can be seen as an excessive use.
For our example, the previous algorithm was reiterated seven times before reaching the convergence for the common component q_1. At the convergence, the loss function is equal to $L_1 = 6.26\ 10^{-12}$, the largest singular value is equal to 1.541 and the associated eigenvector is given by:

$$q_1 = [\ 0.714 \quad -0.184 \quad 0.445 \quad -0.332 \quad -0.022 \quad -0.263 \quad -0.263 \quad -0.097\ ] \qquad (2.10)$$

Thus, the specific weight of the dataset 1, 2, 3 and 4 in the computation of the common component q_1 is respectively: $\lambda_1^{(1)} = 0.670$, $\lambda_1^{(2)} = 0.664$, $\lambda_1^{(3)} = 0.780$ and $\lambda_1^{(4)} = 0.205$. This means that 67 %, 66.4 %, 78 % and 20.5 % of the total inertia contained respectively in the datasets 1, 2, 3 and 4 is explained by the common component q_1.

Step 3e: Apply a deflation procedure on Y_k (k=1, 2, ..., m) in order to determine the common component q_2 and the specific weights $\lambda_2^{(k)}$ $(k = 1,\ 2,\ \ldots,\ m)$. This deflation procedure consists of considering the datasets:

$$Y_k^{(2)} = Y_k - q_1 q_1' Y_k \quad (k = 1,\ 2,\ \ldots, m) \tag{2.11}$$

that contain the residuals of the orthogonal projection of the variables of Y_k (k=1, 2, ..., m) on the common component q_1. In other words, for a given dataset Y_k, the first column (variable) of $Y_k^{(2)}$ is given by the residuals obtained from a simple linear regression where the response variable is the first column (variable) of Y_k and the explanatory variable is the common component q_1. The second column of $Y_k^{(2)}$ is given by the residuals obtained from a simple linear regression where the response variable is the second column of Y_k and the explanatory variable is the common component q_1, and so on. The residuals are used here in order to take into account the information that was left for the previous common component q_1. In doing so, this deflation procedure ensures the orthogonality of the common components (avoiding thus, the redundancy of the information on the common components).

For our example, the dataset obtained after the first deflation of Y_1 is:

$$Y_1^{(2)} = \begin{bmatrix} -0.015 & -0.040 & 0.043 & 0.068 \\ -0.038 & -0.017 & -0.082 & -0.030 \\ 0.037 & -0.014 & -0.024 & -0.107 \\ -0.010 & -0.013 & -0.021 & 0.131 \\ -0.069 & 0.180 & -0.042 & -0.003 \\ 0.074 & 0.128 & -0.101 & -0.190 \\ -0.023 & -0.261 & 0.287 & 0.004 \\ 0.043 & 0.037 & -0.060 & 0.127 \end{bmatrix} \tag{2.12}$$

Thereafter, we compute the scalar product matrix:

$$W_k^{(2)} = X_k^{(2)} X_k^{(2)\prime} \tag{2.13}$$

associated with $Y_k^{(2)}$ (k=1, 2, ..., m).

For our example, we obtained the scalar product matrix $W_1^{(2)}$ associated with $Y_k^{(2)}$ as:

$$W_1^{(2)} = \begin{bmatrix} 0.008 & -0.004 & -0.008 & 0.009 & -0.008 & -0.024 & 0.023 & 0.004 \\ -0.004 & 0.009 & 0.004 & -0.002 & 0.003 & 0.009 & -0.018 & -0.001 \\ -0.008 & 0.004 & 0.014 & -0.014 & -0.004 & 0.024 & -0.004 & -0.011 \\ 0.009 & -0.002 & -0.014 & 0.018 & -0.001 & -0.025 & -0.002 & 0.017 \\ -0.008 & 0.003 & -0.004 & -0.001 & 0.039 & 0.023 & -0.057 & 0.006 \\ -0.024 & 0.009 & 0.024 & -0.025 & 0.023 & 0.068 & -0.065 & -0.010 \\ 0.023 & -0.018 & -0.004 & -0.002 & -0.057 & -0.065 & 0.151 & -0.027 \\ 0.004 & -0.001 & -0.011 & 0.017 & 0.006 & -0.010 & -0.027 & 0.023 \end{bmatrix} \quad (2.14)$$

Finally, the common component q_2 and the associated specific weights $\lambda_2^{(k)}$ (k=1, 2, . . ., m) are estimated by using the same above algorithm (from step 3a to step 3d) but by taking into account the scalar product matrices $W_k^{(2)}$ instead of W_k (k=1, 2, . . ., m). The loss function for the common component q_2 is defined as:

$$\begin{aligned} L_2 &= \sum_{k=1}^{m} \|W_k^{(2)} - \lambda_1^{(k)} q_1 q_1' - \lambda_2^{(k)} q_2 q_2'\|^2 \\ &= \sum_{k=1}^{m} \|W_k^{(2)}\|^2 - \sum_{k=1}^{m}\sum_{i=1}^{2} (\lambda_i^{(k)})^2 \\ L_2 &= L_1 - \sum_{k=1}^{m} (\lambda_2^{(k)})^2 \end{aligned} \quad (2.15)$$

For our example, the previous algorithm was reiterated sixteen times before reaching the convergence for the common component q_2. At the convergence, the loss function is equal to $L_2 = 1.3\ 10^{-11}$, 0.230 is the largest singular value and the vector associated with that singular value is:

$$q_2 = [\ -0.247 \quad 0.51 \quad 0.398 \quad 0.155 \quad 0.108 \quad -0.38 \quad 0.036 \quad -0.582\ ] \quad (2.16)$$

Thus, the specific weight of the dataset 1, 2, 3 and 4 in the computation of the common component q_2 is respectively: $\lambda_2^{(1)} = 0.012$, $\lambda_2^{(2)} = 0.041$, $\lambda_2^{(3)} = 0.061$ and $\lambda_2^{(4)} = 0.474$. So, 1.2 %, 4.1 %, 6.1 % and 47.4 % of the total inertia contained respectively in the datasets 1, 2, 3 and 4 is explained by the common component q_2.

Step 3f: Determine the common component q_3 and the specific weights $\lambda_3^{(k)}$, by considering the datasets

$$Y_k^{(3)} = Y_k - q_1 q_1' Y_k - q_2 q_2' Y_k \quad (k = 1\ , 2,\ \ldots, m) \quad (2.17)$$

that contain the residuals of the orthogonal projection of the variables of Y_k (k=1, 2, . . ., m) on the common components q_1 and q_2.

In other words, for a given dataset Y_k, the first column of $Y_k^{(3)}$ is the residuals obtained from a multiple linear regression where the response variable is the first column of Y_k and the explanatory variables are the common components q_1 and q_2. The second column of $Y_k^{(3)}$ is given by the residuals of a multiple linear regression where the response variable is the second column of Y_k and the explanatory variables are the common components q_1 and q_2. The third and subsequent columns of $Y_k^{(3)}$ are found in the same way.

For our example, the dataset obtained after the second deflation of Y_1 is:

$$Y_1^{(3)} = \begin{bmatrix} -0.030 & -0.056 & 0.046 & 0.054 \\ -0.005 & 0.016 & -0.089 & -0.001 \\ 0.062 & 0.013 & -0.029 & -0.085 \\ 0.000 & -0.002 & -0.023 & 0.139 \\ -0.062 & 0.187 & -0.043 & 0.004 \\ 0.050 & 0.102 & -0.096 & -0.211 \\ -0.021 & -0.258 & 0.286 & 0.006 \\ 0.006 & -0.002 & -0.052 & 0.094 \end{bmatrix} \qquad (2.18)$$

On the obtained datasets, we compute the scalar product matrices:

$$W_k^{(3)} = Y_k^{(3)} Y_k^{(3)\prime} \quad (k = 1\ , 2,\ \ldots, m)$$

For our example:

$$W_1^{(3)} = \begin{bmatrix} 0.009 & -0.005 & -0.009 & 0.007 & -0.011 & -0.023 & 0.029 & 0.003 \\ -0.005 & 0.008 & 0.003 & 0.002 & 0.007 & 0.010 & -0.030 & 0.004 \\ -0.009 & 0.003 & 0.012 & -0.011 & 0.000 & 0.025 & -0.013 & -0.006 \\ 0.007 & 0.002 & -0.011 & 0.020 & 0.001 & -0.027 & -0.005 & 0.014 \\ -0.011 & 0.007 & 0.000 & 0.001 & 0.041 & 0.020 & -0.060 & 0.002 \\ -0.023 & 0.010 & 0.025 & -0.027 & 0.020 & 0.067 & -0.056 & -0.015 \\ 0.029 & -0.030 & -0.013 & -0.005 & -0.060 & -0.056 & 0.149 & -0.014 \\ 0.003 & 0.004 & -0.006 & 0.014 & 0.002 & -0.015 & -0.014 & 0.012 \end{bmatrix} \qquad (2.19)$$

Finally, we repeat again the same algorithm (from step 3a to step 3d) by considering the scalar product matrices $W_k^{(3)}$ instead of W_k (k=1, 2, ..., m).

The previous algorithm was reiterated 5 times before reaching the convergence for the common component q_3. At the convergence, the loss function is equal to $L_3 = 8.8\ 10^{-11}$, the largest singular value is 0.174 and the common component q_3 is given by:

$$q_3 = [\ -0.252 \quad 0.210 \quad 0.187 \quad -0.001 \quad 0.214 \quad 0.197 \quad -0.834 \quad 0.279\ ] \tag{2.20}$$

Thus, the specific weight of the dataset 1, 2, 3 and 4 in the computation of the common component q_3 is respectively: $\lambda_3^{(1)} = 0.191$, $\lambda_3^{(2)} = 0.216$, $\lambda_3^{(3)} = 0.110$ and $\lambda_3^{(4)} = 0.280$. This means that 19.1 %, 21.6 %, 11 % and 28 % of the total inertia contained respectively in the datasets 1, 2, 3 and 4 is explained by the common component q_3.

At step r (r=n-1, n being the number of individuals), the common component q_r and the specific weights $\lambda_r^{(k)}$ are determined by considering the scalar product matrices:

$$W_k^{(r)} = X_k^{(r)} X_k^{(r)\prime} \tag{2.21}$$

where

$$X_k^{(r)} = X_k - \sum_{i<r} q_i q_i' X_k \tag{2.22}$$

k=1, 2, ..., m and r=1, 2, ..., n-1.

The loss function is evaluated as:

$$\begin{aligned} L_r &= \sum_{k=1}^{m} \|W_k^{(r)} - \sum_{i=1}^{r} \lambda_i^{(k)} q_i q_i'\|^2 \\ &= \sum_{k=1}^{m} \|W_k^{(r)}\|^2 - \sum_{k=1}^{m}\sum_{i=1}^{r} (\lambda_i^{(k)})^2 \\ L_r &= L_{r-1} - \sum_{k=1}^{m} (\lambda_r^{(k)})^2 \end{aligned} \tag{2.23}$$

For our example, n=8. So, we have r=7 common components and the dataset obtained after the sixth deflation of Y_1 is:

$$Y_1^{(7)} = \begin{bmatrix} 0.001 & 0.000 & -0.001 & 0.000 \\ -0.021 & -0.001 & 0.013 & -0.007 \\ 0.011 & 0.001 & -0.007 & 0.004 \\ 0.034 & 0.002 & -0.020 & 0.011 \\ -0.015 & -0.001 & 0.009 & -0.005 \\ 0.009 & 0.001 & -0.006 & 0.003 \\ -0.009 & -0.001 & 0.005 & -0.003 \\ -0.012 & -0.001 & 0.007 & -0.004 \end{bmatrix} \tag{2.24}$$

The scalar product matrix of $Y_1^{(7)}$ is:

$$W_1^{(7)} = \begin{bmatrix} 0 & 0.000 & 0.000 & 0.000 & 0.000 & 0 & 0 & 0.000 \\ 0 & 0.001 & 0.000 & -0.001 & 0.000 & 0 & 0 & 0.000 \\ 0 & 0.000 & 0.000 & 0.001 & 0.000 & 0 & 0 & 0.000 \\ 0 & -0.001 & 0.001 & 0.002 & -0.001 & 0 & 0 & -0.001 \\ 0 & 0.000 & 0.000 & -0.001 & 0.000 & 0 & 0 & 0.000 \\ 0 & 0.000 & 0.000 & 0.000 & 0.000 & 0 & 0 & 0.000 \\ 0 & 0.000 & 0.000 & 0.000 & 0.000 & 0 & 0 & 0.000 \\ 0 & 0.000 & 0.000 & -0.001 & 0.000 & 0 & 0 & 0.000 \end{bmatrix} \tag{2.25}$$

The same algorithm (from step 3a to step 3d) applied on $W_k^{(7)}$, was reiterated twice to reach the convergence for the common component q_7. At the convergence, the loss function was evaluated to $L_7 = 1.6\ 10^{-34}$, the largest singular value equals to 0.0004 and the common component q_7 is given by:

$$q_7 = [\ -0.031 \quad 0.446 \quad -0.239 \quad -0.717 \quad 0.307 \quad -0.198 \quad 0.184 \quad 0.248\ ] \tag{2.26}$$

The specific weight of the dataset 1, 2, 3 and 4 in the computation of the common component q_7 is respectively: $\lambda_7^{(1)} = 0.003$, $\lambda_7^{(2)} = 0.014$, $\lambda_7^{(3)} = 0.012$ and $\lambda_7^{(4)} = 0.005$.
So, 0.3 %, 1.4 %, 1.2 % and 0.5 % of the total inertia contained respectively in the datasets 1, 2, 3 and 4 is explained by the common component q_7.
Some properties of common components and specific weights are presented in Hanafi and Qannari (2008).

At step r:

$$i : (\sum_{k=1}^{m} \lambda_r^{(k)} W_k^{(r)})\ q_r \quad = \quad \mu_r^{max}\ q_r \tag{2.27}$$

where μ_r^{max} is the largest eigenvalue of $\sum_{k=1}^{m} \lambda_r^{(k)} W_k^{(r)}$ associated with the eigenvector q_r

$$ii : \lambda_r^{(k)} = q_r' W_k q_r \tag{2.28}$$

$$
\begin{aligned}
iii: L_r &= \sum_{k=1}^{m} \|W_k - \sum_{i=1}^{r} \lambda_i^{(k)} q_i q_i'\|^2 = \sum_{k=1}^{m} \|W_k\|^2 - \sum_{k=1}^{m}\sum_{i=1}^{r} (\lambda_i^{(k)})^2 \\
L_r &= \sum_{k=1}^{m} \|W_k\|^2 - \sum_{i=1}^{r} \mu_i^{max} \qquad (2.29)
\end{aligned}
$$

From these properties, it appears that the relative importance of the common component q_r (r=1, 2, ..., n-1) can be evaluated as:

$$
V_r = \frac{\sum_{k=1}^{m} (\lambda_r^{(k)})^2}{\sum_{k=1}^{m} \|W_k\|^2} = \frac{\mu_r^{max}}{\sum_{k=1}^{m} \|W_k\|^2}, r = 1,\ 2,\ \ldots,\ n-1. \qquad (2.30)
$$

The numerator μ_r^{max} is the singular value of the common component q_r (r=1, 2, ..., n-1) and the denominator $\sum_{k=1}^{m} \|W_k\|^2$ is the total inertia of all the datasets. The singular value of the common component q_1, q_2, ..., q_7 is respectively: 1.541, 0.230, 0.174, 0.009, 0.001, 0.001 and 0.0004. And, the variance (inertia) of a given dataset X_k is computed as: $\|W_k\|^2$ (k=1, 2, ..., m).

For our example, the variance of X_1, X_2, X_3 and X_4 is respectively: 0.597, 0.549, 0.685 and 0.429. Thus, the total inertia of all the datasets is 0.597+0.549+0.685+0.429=2.26.

So, the relative importance of the common component q_1, q_2, ..., q_7 is respectively: $\frac{1.541}{2.26} = 68.197$ %; $\frac{0.230}{2.26} = 10.174$ %; $\frac{0.174}{2.26} = 7.693$ %; $\frac{0.009}{2.26} = 0.402$ %; $\frac{0.0015}{2.26} = 0.066$ %; $\frac{0.001}{2.26} = 0.044$ % and $\frac{0.0004}{2.26} = 0.016$ %.

Note that, the formula used to compute the relative importance of the common components is different from the one used to evaluate the importance of principal components or common factors in PCA and MFA respectively. Indeed, in PCA and MFA, once the eigenvalues are determined, the importance of a dimension is given by the eigenvalue of that dimension divided by the sum of all the eigenvalues.

Step 4: Compute a compromise matrix.

After estimating the common components and the specific weights, a compromise matrix C is computed as: $C = Q\sqrt{D}$, where Q is a (n, n-1) matrix of common components, D is a (n-1, n-1) diagonal matrix of the mean saliences per dimension, and n, the number of individuals.

For our example n=8, so the size of Q and D is respectively (8, 7) and (7, 7).

$$Q = \begin{bmatrix} -0.714 & -0.247 & -0.252 & -0.163 & -0.008 & -0.461 & -0.031 \\ 0.184 & 0.510 & 0.210 & -0.089 & -0.418 & -0.394 & 0.446 \\ -0.445 & 0.398 & 0.187 & 0.291 & -0.162 & 0.562 & -0.239 \\ 0.332 & 0.155 & -0.001 & -0.456 & 0.076 & -0.117 & -0.717 \\ 0.022 & 0.108 & 0.214 & -0.024 & 0.849 & 0.035 & 0.307 \\ 0.263 & -0.380 & 0.197 & 0.710 & -0.052 & -0.277 & -0.198 \\ 0.263 & 0.036 & -0.834 & 0.123 & -0.022 & 0.245 & 0.184 \\ 0.097 & -0.582 & 0.279 & -0.392 & -0.263 & 0.406 & 0.248 \end{bmatrix} \quad (2.31)$$

$$D = \begin{bmatrix} 0.58 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.00 & 0.147 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.00 & 0.000 & 0.199 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.00 & 0.000 & 0.000 & 0.037 & 0.000 & 0.000 & 0.000 \\ 0.00 & 0.000 & 0.000 & 0.000 & 0.016 & 0.000 & 0.000 \\ 0.00 & 0.000 & 0.000 & 0.000 & 0.000 & 0.012 & 0.000 \\ 0.00 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.008 \end{bmatrix} \quad (2.32)$$

Hence we obtain a (8, 7) compromise matrix C:

$$C = \begin{bmatrix} -0.544 & -0.095 & -0.113 & -0.032 & -0.001 & -0.051 & -0.003 \\ 0.140 & 0.196 & 0.094 & -0.017 & -0.053 & -0.044 & 0.041 \\ -0.339 & 0.153 & 0.094 & 0.056 & -0.021 & 0.062 & -0.022 \\ 0.253 & 0.059 & 0.000 & 0.088 & 0.010 & -0.013 & -0.065 \\ 0.017 & 0.042 & 0.096 & -0.005 & 0.108 & 0.004 & 0.028 \\ 0.200 & -0.145 & 0.088 & 0.137 & -0.007 & -0.031 & -0.018 \\ 0.200 & 0.014 & -0.372 & 0.024 & -0.003 & 0.027 & 0.017 \\ 0.074 & 0.223 & 0.124 & -0.076 & -0.033 & 0.045 & 0.023 \end{bmatrix} \quad (2.33)$$

This compromise also known as a ”consensus” is used to compute the correlation between the initial variables of different datasets and the common components on the one hand and the Escoufier R_V coefficient between each dataset and this compromise on the other hand. The Escoufier R_V coefficient is the generalization of squared Pearson correlation coefficient and takes values in the range 0 and 1. This coefficient can be used to compare matrices. But, before comparing two rectangular matrices using the Escoufier R_V coefficient, we must first of all transform them into positive semi-definite matrices (square matrices) by

multiplying each matrix by its transpose (Abdi, 2007).
For example, if we wish to compute the R_V coefficient between the (8, 4) matrix Y_1 and (8, 7) matrix C, the first step is to compute: $S = Y_1Y_1'$ and $T = CC'$. Then, use the following formula:

$$R_V = \frac{trace\{S'T\}}{\sqrt{(trace\{S'S\})(trace\{T'T\})}} = \frac{trace\{Y_1Y_1'CC'\}}{\sqrt{(trace\{Y_1Y_1'Y_1Y_1'\})(trace\{CC'CC'\})}} = 0.886$$

The trace operation is applied to square matrices and gives the sum of the diagonal elements. In R software, the Escoufier R_V coefficient between Y_1 and the compromise C can be computed using the "coeffRV" function of the FactoMineR package: coeffRV(Y_1, C)$rv. For our example, this coefficient of the second, third and fourth dataset with the compromise C is respectively 0.930, 0.926 and 0.593. The Escoufier R_V coefficient of the dataset Y_2 with the compromise is equal to 0.93. So, 0.93 is the amount of variance that the dataset Y_2 shares with the compromise C.

CHAPTER 3

Materials and methods

3.1 Description of data related to farmers' perception of land degradation and soil erosion

The data sample considered to present the application of CCSWA is related to farmers' perception of land degradation and soil erosion in northern Benin (Avakoudjo *et al.*, 2011). A total number of 136 farmers from 5 sociocultural groups were interviewed on their opinion on various causes and factors of land degradation and soil erosion in their farms and in surrounding areas. They were also interviewed on different cultivation techniques they adopt against the soil erosion and techniques that can be used to improve the soil fertility and crops productivity. Variables on which they were interviewed are presented in Table 3.1 below.

Thus, 4 datasets were formed, each bearing on the same individuals. The number of variables of these datasets are different. Dataset 1 (Table 3.2) is related to the causes of land degradation and has 4 variables (Erosion, Deforest, Agricset and Wildfire). Dataset 2 (Table 3.3) is related to soil erosion factors and has 4 variables (Stamping, Run-off, Soiltype, Slope). Dataset 3 (Table 3.4) is linked to the land use practices with 3 variables (Landcover, Orthogcult and Stonyline). And dataset 4 (Table 3.5) describes the techniques of improvement of the soil fertility and crops productivity with 7 variables (Fallow, Fertilizers, Manure, Rubbish, Penning, Plow and Cart). These 4 datasets are presented in Tables 3.2, 3.3, 3.4 and 3.5 below and were used to present the application of CCSWA. Values inside these tables are the percentage of positive responses of the considered sociocultural group to the considered variable. For example in Table 3.2, 92.9 that is in the cross of the first row (Dendi) and the first column (Erosion) means that: 92.9 % of Dendi farmers think that land degradation is due to erosion. We provide in Figure 3.1, some photos to show an example of: run-off, erosion, deforestation, wildfire, stony line, plow and cart.

Table 3.1: Variables of the application datasets

Short name	Long name/ definition
Erosion	is the process by which the surface of the earth is worn away for example by the action of water and winds
Deforest	Deforestation/ is the removal of a forest or stand of trees, where the land is thereafter converted to a non-forest use
Agricset	Agricultural settlement/ is one of the use after removing a forest
Wildfire	is a fire burning in a wild area of land that is difficult to control and sometimes spreads quickly
Stamping	Animal stamping/ is the action of bringing animal foots down onto the soil surface forcibly.
Run-off	is a phenomenon of the flow of the water on the soil surface
Soiltype	Soil type/ is the different sizes of mineral particles in a particular sample
Slope	is a typographic factor that favors the erosion. It increases the speed of the flow
Landcover	Land cover/ is the physical material at the surface of the earth
Orthogcult	Orthogonal cultivation/ is a technique that is used to decelerate the flow of the water
Stonyline	Stony line/ is a practice that consists of placing stones in one or more rows along the level curves
Fallow	is the land that has undergone plowing and harrowing and has been left unseeded for one or more growing seasons
Fertilizers	are any material of natural or synthetic origin that are applied to soils or plants to supply one or more plant nutrients essential to their growth
Manure	is the organic matter mostly derived from animal feces (for example the chicken manure and the cow dung) and which contributes to the soil fertility
Rubbish	Household rubbishes/ are solid waste comprising of garbage such as compost, disposables, food packaging, food scraps
Penning	is the action of bringing animals in a field to use the vegetal resources for their feeding
Plow	is a machine that is used to turn and break up soil, to bury crop residues and to help control weeds
Cart	is a two-wheeled vehicle drawn by an animal or individuals and used in farm work and for transporting goods

Source: https://fr.wikipedia.org/wiki/Ruissellement

Source: https://en.wikipedia.org/wiki/Erosion

Source: https://en.wikipedia.org/wiki/Deforestation

Source: http//agroecologie.cirad.fr

Source: http://www.kouminto.fr/agriculture.php

Source: https://en.wikipedia.org/wiki/Plough

Source: https://en.wikipedia.org/wiki/Cart

Figure 3.1: Photos of: a. Run-off- b. Erosion- c. Deforestation- d. Wildfire- e. Stony line- f. Plow- g. Cart

Table 3.2: Causes of land degradation

	Erosion	Deforest	Agricset	Wildfire
Dendi	92.9	76.8	91.1	37.5
Djerma	100	70	90	40
Gourmanche	90.3	74.2	90.3	25.8
Haussa	88.9	88.9	77.8	22.2
Peulh	80	70	86.7	23.3

Table 3.3: Soil erosion factors

	Stamping	Run-off	Soiltype	Slope
Dendi	46.4	82.1	39.3	80.4
Djerma	30	100	10	90
Gourmanche	25.8	80.6	45.2	90.3
Haussa	22.2	77.8	33.3	66.7
Peulh	30	80	33.3	83.3

Table 3.4: Land use practices

	Landcover	Orthogcult	Stonyline
Dendi	57.1	60.7	8.9
Djerma	60	90	20
Gourmanche	45.2	77.4	12.9
Haussa	44.4	88.9	44.4
Peulh	56.7	90	3.3

Table 3.5: Techniques of improvement of the soil fertility and crops productivity

	Fallow	Fertilizers	Manure	Rubbish	Penning	Plow	Cart
Dendi	48.2	30.4	60.7	51.8	83.9	76.8	50
Djerma	30	30	40	30	90	80	50
Gourmanche	45.2	38.7	51.6	54.8	77.4	67.7	32.3
Haussa	44.4	22.2	77.8	66.7	88.9	44.4	22.2
Peulh	30	36.7	60	56.7	86.7	70	46.7

3.2 Statistical analysis

PCA and CCSWA have been applied on the above four datasets (Tables 3.2, 3.3, 3.4 and 3.5) in order to compare the results obtained from the 2 analyses and show the improvement of CCSWA compared to PCA. For PCA, we horizontally merged these 4 datasets and used correlation matrix (each variable is centered and scaled). Variable scaling consists of dividing each entry of that variable by its standard deviation. Before applying CCSWA, each variable was centered. Then, each dataset was normalized. The normalization consisted of giving more weight to smaller datasets so that the inertia of each dataset is set up to 1 (Hanafi and Qannari, 2008). All the analyses were done in R version 3.0.2 (R Core Team, 2013). The scripts used to perform PCA and CCSWA are presented in appendix.

CHAPTER 4

Results and discussion

4.1 Results of CCSWA performed on the 4 datasets

Table 4.1 below presents the saliences of each dataset on the 4 common components of CCSWA applied on the 4 datasets. As each dataset is normalized, each salience can be seen as the percentage of the total inertia of a given dataset restituted by the considered common component. Based on these saliences, one can assess the relationships between different datasets. From this table, it appears that the common component q_1 explained 60.4 %, 45.3 %, 10 % and 73.5 % of the variability in datasets 1, 2, 3 and 4 respectively. Thus, the common structure highlighted by q_1 has been identified in datasets 1, 2 and 4. However, the common component q_2 expressed 70 % of the total inertia of dataset 3, 30.8 % of the total inertia of dataset 2 and a relatively low percentage of the total inertia of the remaining datasets (5.4 % and 9.4 % of datasets 1 and 4 respectively). So, the common structure highlighted by q_2 has been identified in datasets 2 and 3. It can be concluded that datasets 1, 2 and 4 give higher importance (weight) to the common component q_1 whereas datasets 2 and 3 give higher importance to the common component q_2. Consequently, farmers who perceive well the causes of land degradation, have a good knowledge of soil erosion factors and techniques that can be used to improve the soil fertility and crops productivity. But these farmers do not have necessarily a good knowledge of adaptation measures to overcome the soil erosion, since the dataset 3 weights heavily the common component q_2.

Table 4.1: Saliences and relative importance of the four common components of CCSWA performed on the 4 datasets

		Dim 1	Dim 2	Dim 3	Dim 4
Dataset 1	Saliences	0.604	0.054	0.307	0.035
	Cumul.	0.604	0.658	0.965	1
Dataset 2	Saliences	0.453	0.308	0.134	0.106
	Cumul.	0.453	0.761	0.895	1
Dataset 3	Saliences	0.1	0.7	0.192	0.008
	Cumul.	0.1	0.8	0.992	1
Dataset 4	Saliences	0.735	0.094	0.081	0.089
	Cumul.	0.735	0.829	0.91	1
Relative importance	Values	0.532	0.284	0.074	0.0097
	Cumul.	0.532	0.816	0.89	0.899

With the first two common components, at least 65.8 % of the total inertia contained in each dataset is explained (65.8 %, 76.1 %, 80 % and 82.9 % of the total inertia contained in datasets 1, 2, 3 and 4 respectively). Based on the (cumulative) relative importance of the first two common components, 81.6 % of the total inertia contained in the 4 datasets is explained. Thus, retaining the first two common components is sufficient for a good synthesis of the analysis. The graphical representation of the 4 datasets based on their saliences on the first two common components of CCSWA performed on the 4 datasets is presented in Figure 4.1.

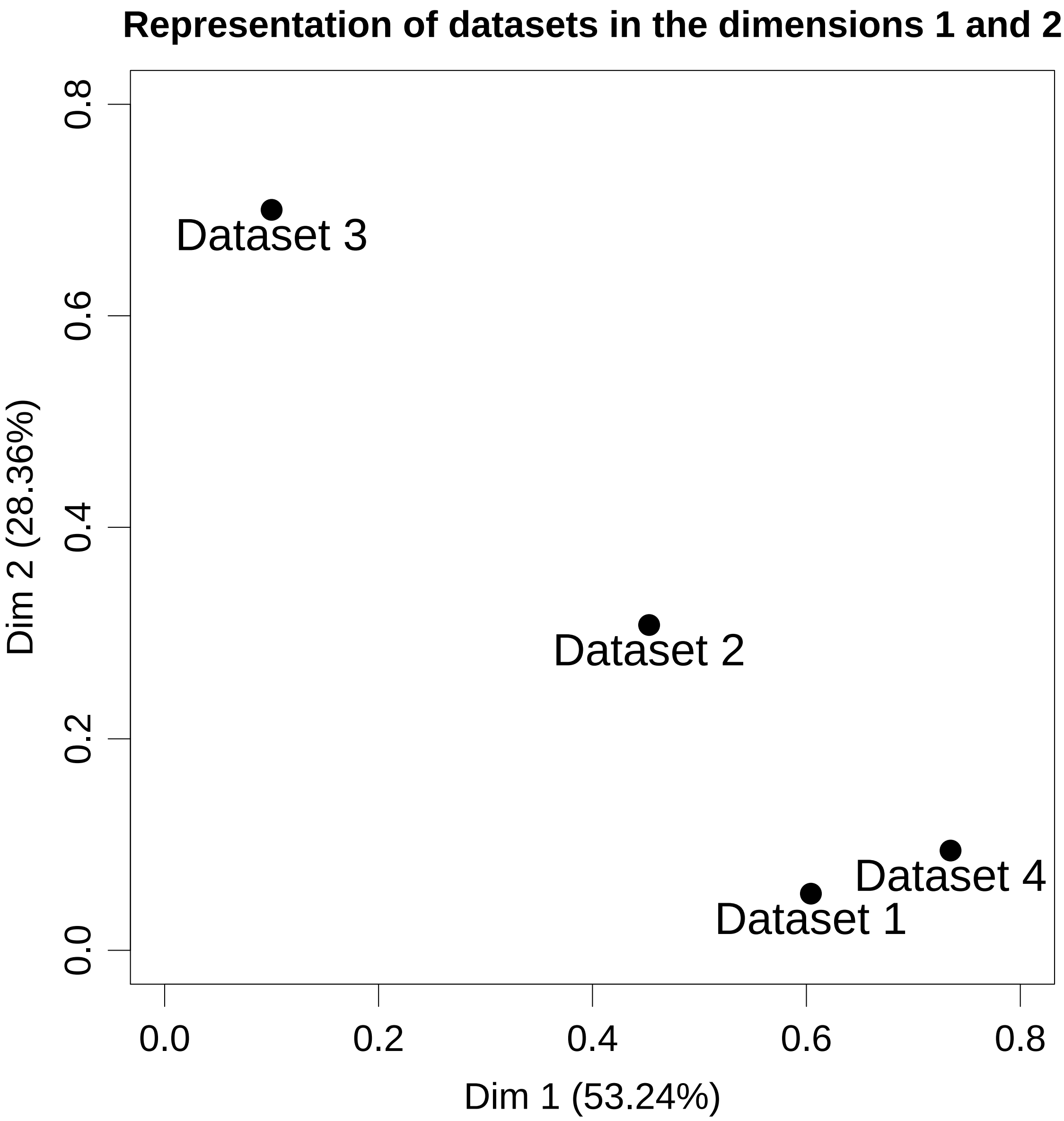

Figure 4.1: Representation of datasets based on their saliences on the first two common components of CCSWA performed on the 4 datasets.

In order to interpret the common components, it is important to examine correlations between the initial variables of each dataset and these common components. These correlations are presented in Table 4.2. Values highlighted in boldface are correlations between initial variables and common components that are deemed important (greater or equal to 0.5 in absolute value). The sign of the correlation coefficient is just an indicator of the side of the axis on which each variable should be interpreted.

Table 4.2: Correlations between initial variables and the four common components of CCSWA performed on the 4 datasets.

	Dim.1	Dim.2	Dim.3	Dim.4
Erosion	**-0.589**	-0.016	**-0.746**	0.311
Deforest	**0.820**	-0.082	**-0.566**	0.011
Agricset	**-0.820**	-0.417	0.294	0.259
Wildfire	**-0.841**	-0.286	-0.452	-0.079
Stamping	-0.442	**-0.827**	-0.040	-0.346
Run-off	**-0.857**	0.365	-0.365	-0.006
Soiltype	**0.587**	**-0.605**	0.420	0.336
Slope	**-0.808**	0.047	0.418	0.412
Landcover	-0.106	**0.962**	0.242	0.061
Orthogcult	0.107	**0.979**	0.082	-0.151
Stonyline	**0.556**	0.384	**-0.733**	0.077
Fallow	0.484	**-0.711**	-0.308	0.406
Fertilizers	-0.367	-0.109	**0.852**	0.357
Manure	**0.928**	-0.167	-0.132	-0.306
Rubbish	**0.958**	-0.196	0.206	-0.032
Penning	-0.081	0.489	-0.463	**-0.735**
Plow	**-0.937**	-0.262	0.225	-0.041
Cart	**-0.846**	-0.256	0.205	-0.421

The scores of different sociocultural groups are presented in Table 4.3. These scores are the common components (eigenvectors associated with the largest eigenvalues of the weighted association matrices, as described in the principle).

Table 4.3: Scores of different sociocultural groups (common components)

	Dim 1	Dim 2	Dim 3	Dim 4
Dendi	-0.146	-0.833	-0.187	-0.223
Djerma	-0.680	0.448	-0.371	-0.019
Gourmanche	0.058	-0.064	0.312	0.834
Haussa	0.715	0.246	-0.468	-0.096
Peuhl	0.052	0.203	0.715	-0.495

The graphical representation of different sociocultural groups in the system axes defined by the first two common components is presented in Figure 4.2. From that figure, one can assess the existing relationships between individuals.

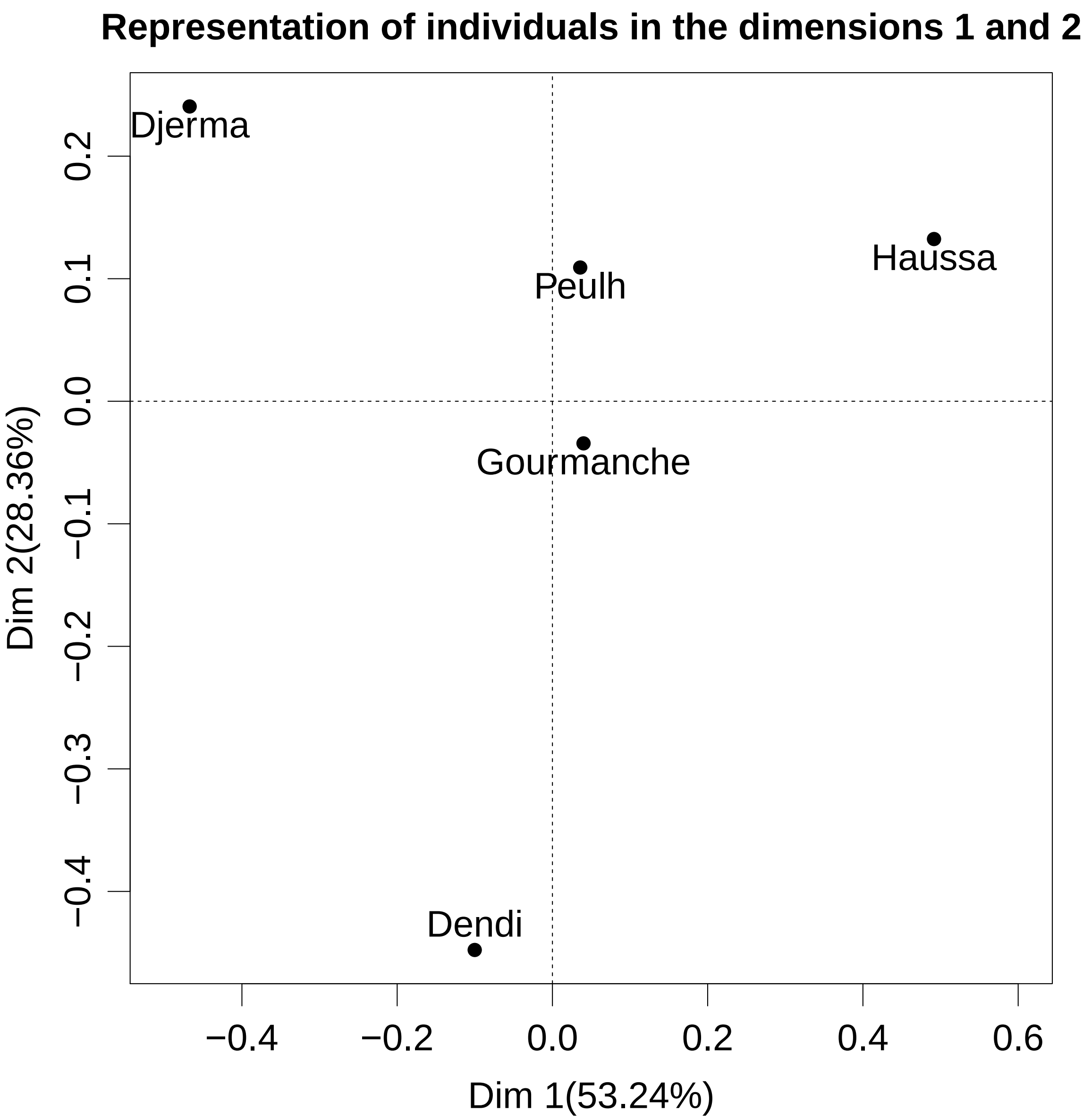

Figure 4.2: Projection of sociocultural groups in the system axis defined by the first two common components of CCSWA performed on the 4 datasets.

To interpret this Figure 4.2, we evaluate the contribution of each sociocultural group in the computation of each common component. If we denote by s_{ij} the score of the sociocultural group i on the common component j and c_{ij} the contribution of the sociocultural group i in the computation of the common component j (i=1, 2, ..., 5 and j=1, 2, 3, 4), then

$$c_{ij} = \frac{s_{ij}^2}{\sum_{i=1}^{n} s_{ij}^2}$$

.

But as the length of each common component is equal to 1, the sum of the squared scores is equal to 1. Hence, we merely square the scores of sociocultural groups on each common component (Table 4.3). The mean contribution of each sociocultural group in the computation of each common component is: $c = \frac{1}{5} = 0.2$.
Thus, the contribution of a sociocultural group i in the computation of a common component j is deemed important if and only if $c_{ij} \geq c$.
Table 4.4 presents the contribution of each sociocultural group in the computation of each common component. For a given common component, only sociocultural groups for which the contributions are highlighted in boldface will be interpreted. For a given common component, the sign of scores of the retained sociocultural groups (Table 4.3) is an indicator of the side of that common component on which they will be interpreted.

Table 4.4: Contribution of sociocultural groups in the computation of each common component of CCSWA performed on the 4 datasets.

	Dim 1	Dim 2	Dim 3	Dim 4
Dendi	0.021	**0.694**	0.035	0.050
Djerma	**0.462**	**0.200**	0.138	0.000
Gourmanche	0.003	0.004	0.097	**0.695**
Haussa	**0.511**	0.061	**0.219**	0.009
Peuhl	0.003	0.041	**0.511**	**0.245**

The biplot of CCSWA is presented in Figure 4.3. This figure, shows how different sociocultural groups express their perception of land degradation and soil erosion by means of different variables. Only variables and individuals that are surrounded in each side of the 2 axes will be interpreted.

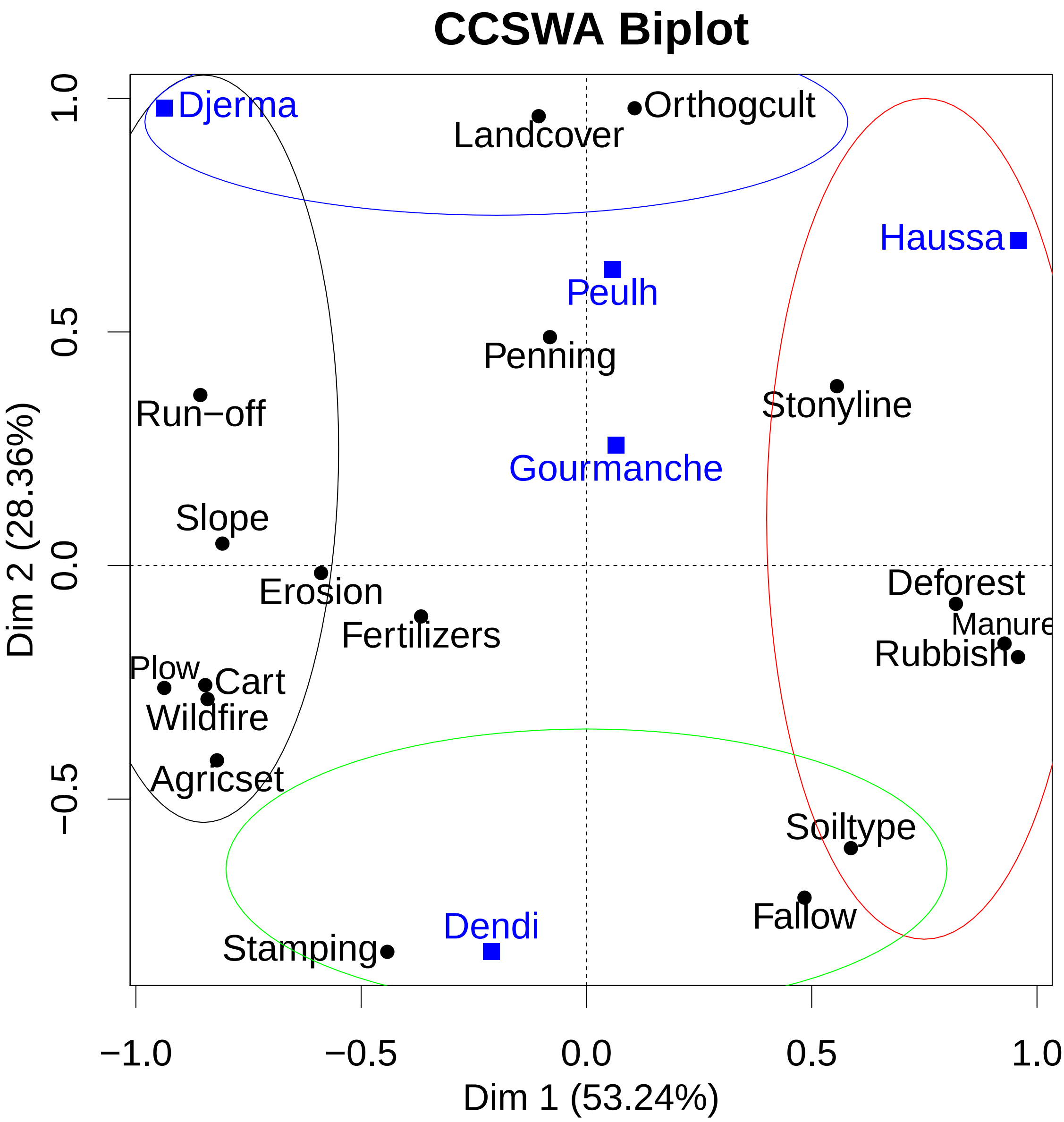

Figure 4.3: Biplot of CCSWA performed on the 4 datasets

The common component q_1 sets apart Djerma from Haussa farmers. Farmers from the two sociocultural groups are those who perceive well the causes of land degradation,the soil erosion factors and the techniques to be used in order to enhance the soil fertility and crops productivity. Djerma farmers think that land degradation is due to the erosion, agricultural settlement and wildfire. Run-off and slope are the main soil erosion factors according to them. They also think that crops productivity can be enhanced by using plows and carts. Regarding Haussa farmers, deforestation is the main cause of land degradation whereas soil type is the main soil erosion factor. To overcome this, they set up stony lines and use manure and household rubbishes to improve the soil fertility and crops productivity. The common component q_2 is mainly related to the soil erosion factors and land use practices. Only Dendi and Djerma farmers account for this common component, but they are opposed. Dendi farmers acknowledge animal stamping and soil type as main soil erosion factors and practice fallow to improve the soil fertility and crops productivity. As regards Djerma farmers, they cover their lands and till orthogonally to the normal flow of water in order to overcome soil erosion.

4.2 Results of PCA performed on the 4 datasets

Table 4.5 presents correlations between initial variables and the first four axes of PCA performed on the 4 datasets. The biplot, showing associations between sociocultural groups and initial variables is presented in Figure 4.4. In this figure, only variables and individuals that are surrounded in each side of the 2 axes will be interpreted. In Table 4.5, values highlighted in boldface are correlations between initial variables and principal components that are deemed important (greater or equal to 0.5 in absolute value).

Table 4.5: Correlations between initial variables and the first four axes of PCA performed on the 4 datasets

	Dim 1	Dim 2	Dim 3	Dim 4
Erosion	0.467	-0.251	**0.678**	**0.509**
Deforest	**-0.894**	-0.022	0.427	0.131
Agricset	**0.890**	0.435	-0.032	0.130
Wildfire	**0.769**	-0.028	**0.637**	0.027
Stamping	0.471	**0.590**	**0.500**	-0.423
Run-off	**0.759**	**-0.565**	0.279	0.164
Soiltype	-0.459	**0.844**	-0.257	0.109
Slope	**0.873**	0.100	-0.365	0.307
Landcover	0.083	**-0.776**	**-0.611**	0.130
Orthogcult	-0.161	**-0.871**	-0.463	-0.032
Stonyline	**-0.690**	**-0.516**	0.402	0.310
Fallow	-0.468	**0.685**	0.428	0.359
Fertilizers	**0.519**	0.437	**-0.727**	0.105
Manure	**-0.933**	0.148	0.128	-0.302
Rubbish	**-0.897**	0.361	-0.214	-0.137
Penning	-0.050	**-0.785**	0.355	**-0.506**
Plow	**0.975**	0.187	0.036	-0.113
Cart	**0.870**	0.097	0.113	-0.471

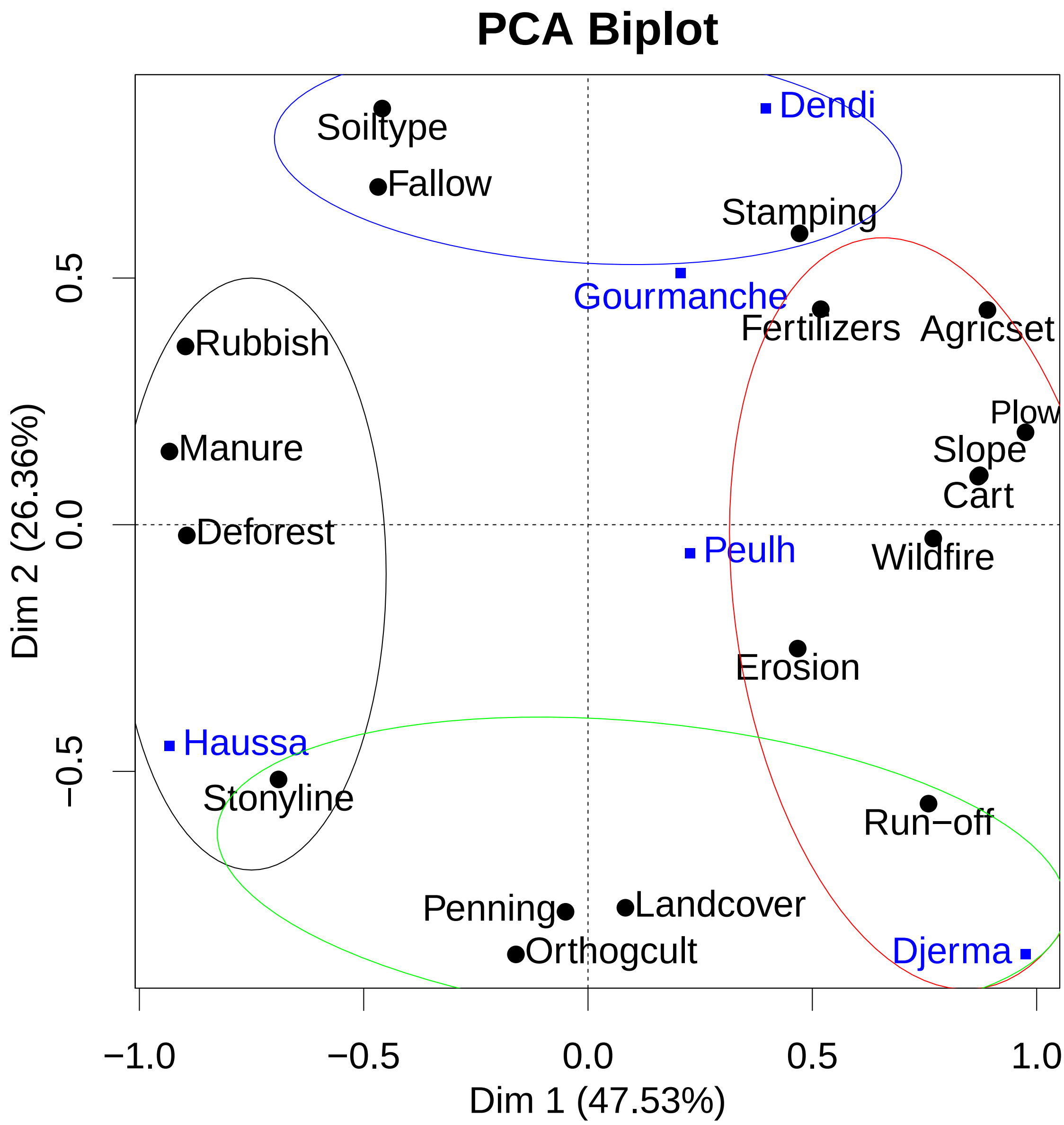

Figure 4.4: Biplot of PCA performed on the 4 datasets

PCA performed on the 4 datasets showed that the first two principal components explained 73.89 % of the total inertia contained in the initial dataset. So, with the first two axes, we have a good synthesis of the analysis. The first axis contrasts Haussa to Djerma farmers. Haussa farmers consider deforestation as the main cause of land degradation; they set up stony lines to overcome the soil erosion and use manure and household rubbishes to improve the soil fertility and crops productivity. However, according to Djerma farmers, the main causes of land degradation are agricultural settlement and wildfire. They consider the run-off and slope as the main soil erosion factors and use fertilizers, plow and cart to improve the soil fertility and crops productivity. The second axis contrasts Djerma to Dendi farmers. Djerma farmers recognize the run-off as the main soil erosion factor. Against soil erosion, they cover their lands, till orthogonally to the normal flow of the water and set up stony lines. They practice the penning in order to improve the soil fertility and crops productivity. Dendi farmers find animal stamping and soil type as the main soil erosion factors and practice the fallow to improve the soil fertility and crops productivity.

4.3 Comparison of PCA and CCSWA based on their outputs

Table 4.6 presents the percentage and cumulative percentage of total inertia explained by the dimensions of PCA and CCSWA performed on the 4 datasets.

Table 4.6: Percentage and cumulative percentage of total variance explained by the dimensions of PCA and CCSWA performed on the 4 datasets.

	PCA		CCSWA	
	% of inertia explained	Cumul.	% of inertia explained	Cumul.
Dim 1	47.530	47.530	53.243	53.243
Dim 2	26.356	73.886	28.365	81.608
Dim 3	18.002	91.888	7.388	88.996
Dim 4	8.112	100.000	0.973	89.969

It can be seen from this table that the cumulative percentage of inertia explained by the first two dimensions is larger in CCSWA than in PCA.
Common components and specific weights method can be used to perform the analysis in individual level and the analysis in datasets level. The analysis in individual level consists of looking at the relationships between individuals. The examination of the associations between the individuals and variables through the PCA biplot and CCSWA biplot, shows that the results of the two biplots are almost the same. The analysis in datasets level consists of investigating the relationships between different datasets involved in the analysis. This can be achieved by looking at the way different datasets contribute to form the common components. For that purpose, the saliences of each dataset in the computation of common components are of paramount importance. Thus, the improvement of CCSWA compared to PCA is the analysis in the datasets level.

4.4 Discussion

Djerma farmers are those who perceive very well the causes of land degradation, the soil erosion factors, the adaptation measures against the soil erosion and the techniques that can be used to improve the soil fertility and crops productivity. This can be explained by the fact that only few farmers of this sociocultural group were interviewed (10 over 136). It is also possible that the interviewed farmers of this sociocultural group are more experienced in working with land. This idea is supported by Akinnagbe and Umukoro (2011), Avakoudjo *et al.* (2011), who found that when the respondents are experienced farmers; they have acquired enough farming experience needed to perceive the effect of degradation on farming activities in their area, over the years. According to local perception of land degradation and soil erosion, Djerma farmers are followed by Haussa farmers. It is somewhat surprising that although Dendi people are farmers in the study area and the most interviewed people in the sample (56 over 136), they do not perceive well the causes of land degradation, the land use practices and the techniques that can be used to enhance the soil fertility and crops productivity. They only have a good knowledge of the factors favoring soil erosion. This can be due to the fact that, those who were interviewed work only with the land; they do not pay attention to the changes affecting their lands; they are not experienced in working with land. Unlike the findings of Avakoudjo *et al.* (2011) who stated that globally, all the sociocultural groups have the same perception of land degradation and soil erosion, we find that the perception of land degradation and soil erosion depends on sociocultural groups.

Finding almost the same results for CCSWA and PCA when assessing the relationships between individuals is not surprising. It can be explained by the fact that the hierarchy of 3 models developed by Qannari *et al.* (1995) are based on association matrices, which show similarities between individuals. It is also worth noting that PCA on the horizontal merged datasets is the solution of the first model whereas STATIS method is the solution led by the second model of this hierarchy. Moreover, it has been shown by Hanafi and Qannari (2008) that the results of STATIS and CCSWA are almost the same, when assessing the relationships between individuals.

4.5 Conclusion

This analysis revealed that each sociocultural group has its way of understanding the causes of land degradation, the soil erosion factors, the adaptation measures against soil erosion and the techniques that can be used to improve the soil fertility and crops productivity. From CCSWA, datasets linked to the causes of land degradation, the soil erosion factors, and the techniques to improve the soil fertility and crops productivity are related whereas they are independent to the dataset linked to the land use practices. PCA cannot assess the relationships between different datasets. But, when investigating the relationships between individuals, CCSWA and PCA give almost the same results.

Bibliography

[1] Blackman J., Rutledge D. N., Tesic D., Saliba A. and Scollary G. R. (2010). Examination of the potential for using chemical analysis as a surrogate for sensory analysis, *Analytica Chimica Acta*, **660**, 2-7.

[2] Hanafi M. and Kiers H. A. (2006). Analysis of K sets of data, with differential emphasis on agreement between and within sets, *Computational Statistics and Data Analysis*, **51**, 1491-1508.

[3] Hanafi M. and Qannari E. M. (2008). Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques, *Journal de la Société Française de Statistique*, tome 149, **2**, 75-97.

[4] Hanafi M., Mazerolles G., Dufour E. and Qannari E. M. (2006). Common components and specific weight analysis and multiple co-inertia analysis applied to the coupling of several measurement techniques, *J. Chemometrics*, **20**, 172-183. doi:10.1002/cem.988.

[5] Karoui R., Dufour E., Pillonel L., Cattenoz T. and Bosset J. O.(2004). Fluorescence and infrared spectroscopies: a tool for the determination of the geographic origin of Emmental cheeses manufactured during summer, *Dairy Science & Technology*, **84**, 359-374. DOI:10.1051/lait:2004010.

[6] Karoui R., Dufour E. and De Baerdemaeker J. (2006). Common components and specific weights analysis: A tool for monitoring the molecular structure of semi-hard cheese throughout ripening, *Analytica Chimica Acta*, **572**, 125-133.

[7] Kissita G., Ambapour S., Makany R. A. and Mizere D. (2009). Two methods of simultaneous analysis in common components and unit weights, *ICASTOR Journal of Mathematical Sciences*, **3**, 149-160.

[8] Kulmyrzaev A. A. and Dufour E. (2010). Relations between spectral and physicochemical properties of cheese, milk, and whey examined using multidimensional analysis, *Food Bioprocess Technol*, **3**, 247-256. DOI 10.1007/s11947-008-0074-x

[9] Mazerolles G., Devaux M. F., Dufour E., Qannari E. M. and Courcoux Ph. (2002). Chemometrics methods for the coupling of spectroscopic techniques and the extraction of relevant information contained in spectral data, *Chemometrics Intell. Lab. Syst*, **63**, 57-68.

[10] Mazerolles G., Hanafi M., Dufour E., Bertrand D. and Qannari E. M. (2006). Common components and specific weights analysis: A chemometric method for dealing with complexity of food products, *Chemometrics and Intelligent Laboratory Systems*, **81**, 41-49.

[11] Qannari E. M., Wakeling I. and MacFie H.J.H. (1995). A hierarchy of models for analysing sensory data, *Food Quality and Preference*, **6**, 309-314.

[12] Qannari E. M., Wakeling I., Courcoux Ph. and MacFie J. M. (2000). Defining the underlying sensory dimensions, *Food Qual. Preference*, **11**, 151-154.

[13] Qannari E. M., Courcoux Ph. and Vigneau E. (2001). Common Components and Specific Weights Analysis performed on preference data, *Food Quality and Preference*, **12**, 365-368.

[14] Schoojans V. and Massart D. L. (2001). Combining spectroscopic data (MS, IR): exploratory chemometric analysis for characterising similarity diversity of chemical structures, *Journal of Pharmaceutical and Biomedical Analysis*, **26**, 225-239.

[15] Williams A. and Langron S.P. (1984). The use of free-choice profiling for the evaluation of commercial ports, *J. Sci. Food Agric*, **35**, 558-568.

[16] Kroonenberg P. M. (2007). Applied multiway data analysis, DOI: 10.1002/9780470238004.app2. John Wiley & Sons, Inc.

[17] Avakoudjo J., Kindomihou V. and Sinsin B. (2011). Farmers' perception and response to soil erosion while abiotic factors are the driving forces in sudanian zone of Benin, *Agric. Engineering Res. J.*, **1**, 20-30.

[18] Chizana C. T., Mapfumo P., Albrecht A., Van Wijk M. and Giller K. (2011). Smallholder farmers' perceptions on land degradation and soil erosion in Zimbabwe, *African crop Science Conference Proceedings*, **8**, 1485-1490.

[19] Akinnagbe O. M. and Umukoro E. (2011). Farmers' Perception of the Effects of Land Degradation on Agricultural Activities in Ethiope East Local Government Area of Delta State, Nigeria, *Agriculturae Conspectus Scientificus*, **76**, 135-141.

[20] Courcoux Ph., Devaux F. and Bouchet B. (2002). Simultaneous decomposition of multivariate image using three-way data analysis: Application to the comparison of cereal grains by confocal Chemometrics methods for the coupling of spectroscopic techniques and for the extraction of relevant information contained in the spectra data tables, *Chemometrics and Intelligent Laboratory Systems*, **63**, 57-68.

[21] Beuvier E., Berthaud K., Cegarra S., Dasen A., Pochet S., Buchin S. and Duboz G. (1997). Ripening and quality of Swiss-type cheese made from raw pasteurised or micro filtered milk, *Int. Dairy J.*, **7**, 311-323.

[22] Di Cagno R., Banks J., Sheehan L., Fox P. F., Brechany E. Y., Corsetti A. and Gobbetti M. (2003). Comparison of the microbiological, compositional, biochemical, volatile profile and sensory characteristics of three Italian PDO ewe's milk cheeses, *Int. Dairy J.*, **13**, 961-972.

[23] Flury B. (1988). Common Principal Component and Related Multivariate Models. Wiley, Chichester.

[24] Husson F., Josse J., Le S. and Mazet J. (2014). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. R package version 1.26. http://CRAN.R-project.org/package=FactoMineR

[25] Eslami A., Qannari E. M., Bougeard S. and Sanchez G. Questions, comments go to Aida Eslami and Stephanie Bougeard (2014). multigroup: methods for multigroup data analysis. R package version 0.4.2. http://CRAN.R-project.org/package=multigroup

[26] Dray S., Dufour A. B. and Chessel D. (2007): The ade4 package-II: Two-table and K-table methods, *R News*, **7**, 47-52.

[27] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[28] R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[29] https://fr.wikipedia.org/wiki/Ruissellement, December 6, 2015

[30] https://en.wikipedia.org/wiki/Erosion, January 7, 2016

[31] https://en.wikipedia.org/wiki/Deforestation, January 8, 2016

[32] http //agroecologie.cirad.fr, December 8, 2015

[33] http://www.kouminto.fr/agriculture.php, December 8, 2015

[34] https://en.wikipedia.org/wiki/Plough, December 24, 2015

[35] https://en.wikipedia.org/wiki/Cart, December 3, 2015

Appendix

```
##########################################################################
# #
#    R Script to perform CCSWA: adapted from Qannari et al. (2000) #
#     #
##########################################################################

# This function accps performs the CCSWA.
# The inputs of this function are:
# tab which is the dataset constituted of the number of individuals
# and the total number of variables of each dataset
# group contains the number of variables per dataset
# ndimension is the number of common components to print.

accps <- function (tab,group, ndimension)  {
ntab=length(group);                     # Number of datasets
nind=nrow(tab)                          # Number of individuals
p=ncol(tab)                             # Total number of variables
ndim=min(nind,p)-1                      # Number of dimensions
W=array(0,dim=c(nind,nind,ntab+1)); # Association matrices
LAMBDA=matrix(0,ntab,ndim);             # Will contain the saliences
Q=matrix(0,nrow=nind,ncol=ndim);        # Will contain the common components

Y<-scale(tab,center=TRUE, scale=F)      # Each variable is centered
# J indicates to which block each variable belongs to.
J=rep(1:ntab , times =  group )

# To set up the inertia of each dataset to 1
library(multigroup)                     # Load the package "multigroup"
for (j in 1:ntab) { Y[,J==j]=normM(Y[,J==j]) }
```

```
# Computation of association matrices
for(j in 1:ntab) { Yj=as.matrix(Y[,J==j]); W[,,j]=Yj%*%t(Yj);}
Itot=0;

# Computation of the total variance of all dataset: sum(trace(Wj*Wj)
for (j in 1:ntab) { Itot=Itot+sum(as.matrix(W[,,j])^2) }

# Computation of the common components(Q)and specific weights(lambda)
explained<-matrix(0,ndim)
Res=NULL
for (dimension in 1:ndim)  { previousfit=100000;
lambda=matrix(1, nrow=ntab); deltafit=1000000;
threshold=1e-10   # Default threshold in CCSWA
while(deltafit>threshold) { W[,,ntab+1]=matrix(0,nrow=nind,ncol=nind);
for (ta in 1:ntab){ W[,,ntab+1]=W[,,ntab+1]+lambda[ta]*W[,,ta]  }
Ws=as.matrix(W[,,ntab+1])

# Perform the PCA
Svdw=svd(Ws)
# q extracts the eigenvector associated to the largest eigeivalue of Ws
q=as.matrix(Svdw$u[,1])
fit=0;
for (ta in 1:ntab) {

# Estimation of the residuals
lambda[ta]=t(q)%*%as.matrix(W[,,ta])%*%q;
aux=as.matrix(W[,,ta])-lambda[ta]*q%*%t(q);
fit=fit+sum(aux^2);
}
deltafit=previousfit-fit; previousfit=fit; } #deltafit>threshold

# Computation of the relative importance of each common component
explained[dimension,1]=100*sum(lambda^2)/Itot
LAMBDA[,dimension]=lambda;   Q[,dimension]=q;
```

```
# Deflation procedure: updating association matrices
aux=diag(1,nind)-q%*%t(q);
for (j in 1:ntab)   {  Y=as.matrix(Y)
Y[,J==j]=aux%*%Y[,J==j]; W[,,j]=Y[,J==j]%*%t(Y[,J==j]);}}

Res$Q=Q[,1:ndimension]; expl<-matrix(0,nrow=ndimension,ncol=2)
expl[,1]=explained[1:ndimension]; expl[,2]=cumsum(expl[,1])
Res$saliences=LAMBDA[,1:ndimension]
rownames(Res$Q)<-rownames(tab)
colnames(Res$Q)<-paste("Dim.",1:ndimension,sep="")
colnames(Res$saliences)<-paste("Dim.",1:ndimension,sep="")
rownames(Res$saliences)<-paste("Dataset ",1:ntab,sep="")
Res$expl<-expl
rownames(Res$expl)<-paste("Dim.",1:ndimension,sep="")
colnames(Res$expl)<-c("%Total Var expl", "Cumul % total Var")


# Computation of the compromise (overall agreement)
library(FactoMineR)              # Load the package "FactoMineR"
LambdaMoyen<-apply(LAMBDA,2,mean)
D=diag(LambdaMoyen)
C=Q%*%sqrt(D)                              # Compromise

# Computation of the Escouffier RV coefficient
Rv<-matrix(0,nrow=ntab,ncol=1)
rownames(Rv)<-paste("Dataset ",1:ntab,sep="")
for(j in 1:ntab) {  Yj=as.matrix(tab[,J==j]); resRV<-coeffRV(Yj,C)
Rv[j,1]<-resRV$rv }
Res$RV<-Rv
return(Res)
}                # End of the program

# To call this funcion accps, one needs to provide its inputs:
# Data
```

```
Y=matrix(c(92.857,76.7857,91.0714,37.5,46.4286,82.143   ,39.2857,
80.3571,7.1429,60.7143,8.9286,48.2143,30.3571,60.7143,51.7857,
83.9286,76.7857,50.0,100.0,70.0,90.0,40.0,30.0,100.0,10.0,90.0,
60.0,90.0,20.0,30.0,30.0,40.0,30.0,90.0,80.0,50.0,90.323,74.1935,
90.3226,25.8065,25.8065,80.645,45.1613,90.3226,45.1613,77.4194,
12.9032,45.1613,38.7097,51.6129,54.8387,77.4194,67.7419,32.2581,
88.889,88.8889,77.7778,22.2222,22.2222,77.778,33.3333,66.6667,
44.4444,88.8889,44.4444,44.4444,22.2222,77.7778,66.6667,88.8889,
44.4444,22.2222,80.0,70.0,86.6667,23.3333,30.0,80.0,33.3333,
83.3333,56.6667,90.0,3.3333,30.0,36.6667,60.0,56.6667,86.6667,
70.0,46.6667),nr=5,byrow=T,dimnames=list(c("Dendi","Djerma",
"Gourmanche","Haussa","Peulh"),c("Erosion","Deforest","Agricset",
"Wildfire","Stamping","Run-off","Soiltype","Slope","Landcover",
"Orthogcult","Stonyline","Fallow","Fertilizers","Manure","Rubbish",
"Penning","Plow","Cart")))

group=c(4,4,3,7)                  # Number of variables per dataset
ntab=length(group)                # Number of datasets
res<-accps(Y,group, 4)            # Call of the function
res                                       # Print the results
#round(res£saliences,3)

# Computation of the contribution of individuals:
ci=res$Q^2; seuil_ind=1/nrow(ci)

# Individuals to be interpreted on each side of a given axis.
which(ci[, 1]>=seuil_ind & res$Q[, 1]<0)
which(ci[, 1]>=seuil_ind & res$Q[, 1]>0)
which(ci[, 2]>=seuil_ind & res$Q[, 2]<0)
which(ci[, 2]>=seuil_ind & res$Q[, 2]>0)

# Scree plot
plot(res$expl[,1], type="o", cex=2, pch=19, xaxt="n", xlab="",
 ylab="", ylim=c(0,100))
```

```
par(new=T)
plot(res$expl[,2], type="o", col="red", cex=2, pch=19, xaxt="n",
 xlab="Dimension", ylab="Total variance restituted",
ylim=c(0,100),main="Scree plot"); axis(side=1, at=1:4)
legend("topleft", legend = c("Percentage of total variance explained",
"Cumulative percentage of total variance explained "),
lty=c(1,1), lwd=c(2.5,2.5),col=c("black","red"),bty = "n")

dev.new()  # To avoid erasing the existing figure
#Saliences
S=res$saliences
plot(S[,1],S[,2], cex=1.25, pch=19, main="Representation of datasets in
the dimensions 1 and 2",
xlab=paste0("Dim 1 (", round(res$expl[1,1], 2), "%)"),
ylab=paste0("Dim 2 (", round(res$expl[2,1], 2), "%)"),xlim=c(0,0.8),
ylim=c(0,0.8))
text(S[,1],S[,2],rownames(S),cex=1 ,pos=1,offset=0.1)

# Computation of the compromise
LambdaMoyen<-apply(res$saliences,2,mean)
D=diag(LambdaMoyen); C=res$Q%*%sqrt(D)

x11()    # To avoid erasing the existing figure
# Graphical representation of individuals in the dimensions 1 and 2
c1=C[,1]; c2=C[,2]
plot(c1,c2, cex=.8, pch=19, main="Representation of individuals in
the dimensions 1 and 2",
xlab=paste0("Dim 1(", round(res$expl[1,1], 2), "%)"),
ylab=paste0("Dim 2(", round(res$expl[2,1], 2), "%)"),xlim=c(-0.5,0.6))
#text(c1,c2,rownames(Y),cex=1,pos=3,offset=0.2)
#abline(h=0,v=0)
text(c1[1],c2[1],"Dendi",cex=1,pos=3,offset=0.3)
text(c1[2],c2[2],"Djerma",cex=1,pos=1,offset=0.2)
text(c1[3],c2[3],"Gourmanche",cex=1,pos=1,offset=0.2)
```

```
text(c1[4],c2[4],"Haussa",cex=1,pos=1,offset=0.2)
text(c1[5],c2[5],"Peulh",cex=1,pos=1,offset=0.2)
abline(h=0,v=0,lty=2)

# Computation of the correlations between the initial variables
#of each dataset and the common components
i1=0; i2=0
for(i in 1:ntab){ i1=i1+1; i2=i2+group[i]; Y.i<-Y[ ,(i1:i2)]
co<-round(cor(Y.i,C),3);print(co); i1=i2  }

# or
coo=round(cor(Y,C),3)
seuil_var=0.5

#Variables to be interpreted on each side of a given axis.
which(abs(coo[, 1])>=seuil_var & coo[, 1]<0)
which(abs(coo[, 1])>=seuil_var & coo[, 1]>0)
which(abs(coo[, 2])>=seuil_var & coo[, 2]<0)
which(abs(coo[, 2])>=seuil_var & coo[, 2]>0)

dev.new()
# Graphical representation of the correlation circle
library(ade4)                              # Load the package "ade4"
s.corcircle(coo[,1:2], xax = 1, yax = 2,label = row.names(coo),
sub = "Correlation of the original variables with the dimensions 1 and 2",
csub = 1.15, possub = "topleft", fullcircle = TRUE, box = TRUE,
add.plot = FALSE)

x11()
# Graphical representation of the biplot
plot(c1,c2,cex=1, pch=15,xaxt="n", yaxt="n",
xlab=paste0("Dim 1 (", round(res$expl[1,1], 2), "%)"),
ylab=paste0("Dim 2 (", round(res$expl[2,1], 2), "%)"),
main="CCSWA Biplot",col="blue")
```

```
#text(c1,c2,rownames(Y),cex=0.8,pos=4,offset=0.2)
text(c1[1],c2[1],"Dendi",cex=1,pos=3,offset=0.3,col="blue")
text(c1[2],c2[2],"Djerma",cex=1,pos=4,offset=0.3,col="blue")
text(c1[3],c2[3],"Gourmanche",cex=1,pos=1,offset=0.3,col="blue")
text(c1[4],c2[4],"Haussa",cex=1,pos=2,offset=0.3,col="blue")
text(c1[5],c2[5],"Peulh",cex=1,pos=1,offset=0.3,col="blue")
par(new=T)
plot(coo[,1],coo[,2],cex=0.8, pch=19,xlab="",ylab="")
#text(coo[,1],coo[,2],rownames(coo),cex=1,pos=1,offset=0.2)
text(coo[,1][1],coo[,2][1],"Erosion",cex=1,pos=1,offset=0.2)
text(coo[,1][2],coo[,2][2],"Deforest",cex=1,pos=3,offset=0.2)
text(coo[,1][3],coo[,2][3],"Agricset",cex=1,pos=1,offset=0.2)
text(coo[,1][4],coo[,2][4],"Wildfire",cex=1,pos=1,offset=0.2)
text(coo[,1][5],coo[,2][5],"Stamping",cex=1,pos=2,offset=0.2)
text(coo[,1][6],coo[,2][6],"Run-off",cex=1,pos=1,offset=0.2)
text(coo[,1][7],coo[,2][7],"Typesol",cex=1,pos=3,offset=0.2)
text(coo[,1][8],coo[,2][8],"Slope",cex=1,pos=3,offset=0.3)
text(coo[,1][9],coo[,2][9],"Landcover",cex=1,pos=1,offset=0.2)
text(coo[,1][10],coo[,2][10],"Orthogcult",cex=1,pos=4,offset=0.2)
text(coo[,1][11],coo[,2][11],"Stonyline",cex=1,pos=1,offset=0.2)
text(coo[,1][12],coo[,2][12],"Fallow",cex=1,pos=1,offset=0.2)
text(coo[,1][13],coo[,2][13],"Fertilizers",cex=1,pos=1,offset=0.2)
text(coo[,1][14],coo[,2][14],"Manure",cex=0.87,pos=3,offset=0.2)
text(coo[,1][15],coo[,2][15],"Rubbish",cex=1,pos=2,offset=0.2)
text(coo[,1][16],coo[,2][16],"Penning",cex=1,pos=1,offset=0.2)
text(coo[,1][17],coo[,2][17],"Plow",cex=0.9,pos=3,offset=0.2)
text(coo[,1][18],coo[,2][18],"Cart",cex=1,pos=4,offset=0.2)
abline(h=0,v=0,lty = 2)
draw.ellipse(-0.85,0.25,-0.3 , 0.8, border = 'black', lwd = 1,angle = 0)
draw.ellipse(-0.2,0.95,0.2 , .78, border = 'blue', lwd = 1,angle = 90)
draw.ellipse(0.75,0.1,0.35 , 0.9, border = 'red', lwd = 1,angle = 0)
draw.ellipse(0,-0.65,0.3 , .8, border = 'green', lwd = 1,angle = 90)
```

```
###################################################################
#
#   R script to perform PCA
#
###################################################################

library(FactoMineR)
pc<-PCA(Y, scale.unit=TRUE, ncp=ncol(Y), graph=F)
a=pc$ind$coord; b=pc$var$cor
dev.new()
plot(a[,1],a[,2],cex=.6, pch=15,xaxt="n", yaxt="n",
xlab=paste0("Dim 1 (", round(pc$eig[1,2], 2), "%)"),
ylab=paste0("Dim 2 (", round(pc$eig[2,2], 2), "%)"),
main="PCA Biplot",col="blue")
#text(a[,1],a[,2],rownames(Y),cex=0.8,pos=3,offset=0.3)
text(a[,1][1],a[,2][1],"Dendi",cex=1,pos=4,offset=0.3,col="blue")
text(a[,1][2],a[,2][2],"Djerma",cex=1,pos=2,offset=0.3,col="blue")
text(a[,1][3],a[,2][3],"Gourmanche",cex=1,pos=1,offset=0.3,col="blue")
text(a[,1][4],a[,2][4],"Haussa",cex=1,pos=4,offset=0.3,col="blue")
text(a[,1][5],a[,2][5],"Peulh",cex=1,pos=4,offset=0.3,col="blue")
par(new=T)
plot(b[,1],b[,2],cex=1, pch=19,xlab="",ylab="")
#text(b[,1],b[,2],rownames(b),cex=0.8,pos=3,offset=0.3)
text(b[,1][1],b[,2][1],"Erosion",cex=1,pos=1,offset=0.2)
text(b[,1][2],b[,2][2],"Deforest",cex=1,pos=4,offset=0.2)
text(b[,1][3],b[,2][3],"Agricset",cex=1,pos=1,offset=0.2)
text(b[,1][4],b[,2][4],"Wildfire",cex=1,pos=1,offset=0.2)
text(b[,1][5],b[,2][5],"Stamping",cex=1,pos=3,offset=0.2)
text(b[,1][6],b[,2][6],"Run-off",cex=1,pos=1,offset=0.2)
text(b[,1][7],b[,2][7],"Typesol",cex=1,pos=1,offset=0.2)
text(b[,1][8],b[,2][8],"Slope",cex=1,pos=3,offset=0.3)
text(b[,1][9],b[,2][9],"Landcover",cex=1,pos=4,offset=0.2)
text(b[,1][10],b[,2][10],"Orthogcult",cex=1,pos=4,offset=0.2)
text(b[,1][11],b[,2][11],"Stonyline",cex=1,pos=1,offset=0.2)
```

```
text(b[,1][12],b[,2][12],"Fallow",cex=1,pos=4,offset=0.2)
text(b[,1][13],b[,2][13],"Fertilizers",cex=1,pos=1,offset=0.2)
text(b[,1][14],b[,2][14],"Manure",cex=1,pos=4,offset=0.2)
text(b[,1][15],b[,2][15],"Rubbish",cex=1,pos=4,offset=0.2)
text(b[,1][16],b[,2][16],"Penning",cex=1,pos=2,offset=0.2)
text(b[,1][17],b[,2][17],"Plow",cex=0.9,pos=3,offset=0.2)
text(b[,1][18],b[,2][18],"Cart",cex=1,pos=1,offset=0.2)
abline(h=0,v=0,lty = 2)
draw.ellipse(-0.75,-0.1,-0.3 , 0.6, border = 'black', lwd = 1,angle = 0)
draw.ellipse(0,0.75,0.22 , .7, border = 'blue', lwd = 1,angle = 87)
draw.ellipse(0.75,-0.18,0.42 , 0.77, border = 'red', lwd = 1,angle = 10)
draw.ellipse(0.12,-0.7,0.3 , .95, border = 'green', lwd = 1,angle = 85)
```