

Fuzzy correspondence analysis

Abstract

The paper deals with the multiple correspondence analysis extended to fuzzy coded disjunctive arrays. The example concerns aquatic Coleoptera species and was treated in the paper by Chevenet et al. (1994, A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* : 31, 295-309).

Contents

1 - Fuzzy coded data.....	2
1.1 - Data entry.....	2
1.2 - Reading fuzzy coded data.....	3
1.3 - Listing data	4
2 - Fuzzy correspondence analysis as an extension of MCA.....	6
2.1 - Mathematical principle.....	6
2.2 - Computation	7
2.3 - Interpretation	11
Appendix 1 - Typology of modalities	19
Appendix 2 - Typology of species	19
Références	20

S. Dolédec & F. Chevenet

1 - Fuzzy coded data

1.1 - Data entry

Create a new data folder as usual. Go to **ADE•Data** selection card and select the example «Coléoptères» (Fig. 1). Data were modified from a work of Bournaud et al. (1992)¹.

Create a text file **ES. car** with data field **A** (Fig. 1), a text file **ES_CodeEsp** with data field **B**, a text file **ES_CodeModa** with data field **C**, a text file **ES_CodeVar** with data field **D**, a text file **ES_CodeFami** with data field **E**, and a text file **ES. num** with data field **F**.

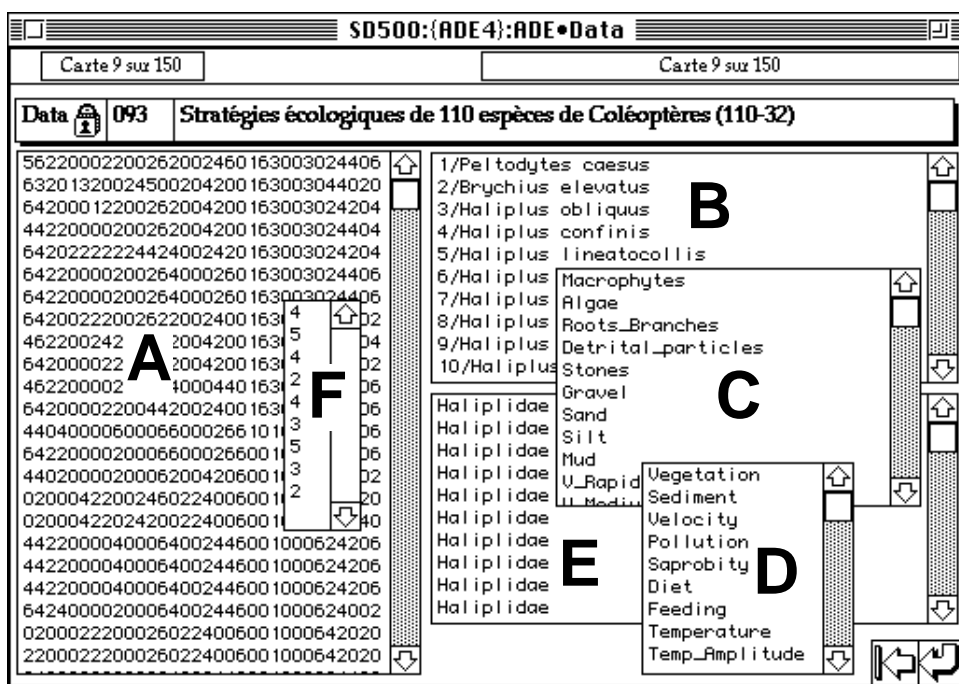


Figure 1 The «Coléoptères» example.

Transform **ES. car** into binary (ES) using **Char->Binary** option of the **TextToBin** module. ES has 110 rows-species et 32 columns-modalities. Transform **ES. num** into binary (ESnum). ESnum has nine rows et one column.

ES is a fuzzy coded table. Such a table describes the affinity of 100 aquatic Coleoptera species for modalities of ecological requirements using a positive score using positive scores (from 0: no affinity to 6: high affinity). We have modified these data to simplify graphics and calculations, by eliminating nine taxa (not determined to a specific level) and by combining the information about larvae and adults of the same species of the cited paper. The ecological-requirements table represent 32 modalities (columns) associated with nine variables. For detailed explanation about data acquisition see Bournaud et al. (op. cit.) Le fichier ESnum indicate the number of modalities for each variable.

Programs tolerate missing values for one or several rows and for one or several columns. In that case, the values of a block of columns corresponding to the same variable for a given row are equal to 0.

To test this option, some profiles were eliminated from the original data set. In such a table, one supposes that the species profile, i.e., the affinity distribution of a species

among the modalities of an ecological requirement is relevant. The original contents of such fuzzy coded tables can be expressed in other notations. For example, the notation "6,4,2,0" for the four modalities of the variable "vegetation" and the species *Haliphus obliquus* is equal to the notations "1/2,1/3,1/6,0" or "50%, 35%, 15%, 0%". If the affinity of a species for a variable is not known, it is coded by "0,0,0,0". It is thus implicitly replaced by the average profile for the corresponding variable. By contrast, the analysis does not tolerate missing data in an entire column.

The row codes (Es_CodeEsp) contains the numero and the name of species. A code for variables (Es_CodeVar) contains the name of the nine variables. A modality code (Es_CodeModa) identify the 32 modalities (4+5+4+2+4+3+5+3+2). A code (Es_CodeFami) identifies which family a species belongs to.

1.2 - Reading fuzzy coded data

Before any calculation, a fuzzy coded file should be processed by the **Read Fuzzy File** option of the module **FuzzyVar**. This operation indicates to further modules how the modalities are distributed among variables:

This results in a listing as follows:

```
Input file: ES
Row number: 110, column number: 32
Number of modalities for each variable: file ESnum
Row number: 9
Missing data: 5
```

Description of categories:

Variable number 1 has 4 categories

```
-----
[  1] Category:    1 Freq.:    0.468
[  2] Category:    2 Freq.:    0.261
[  3] Category:    3 Freq.:    0.152
[  4] Category:    4 Freq.:    0.119
Missing data Num:  2 Freq.:    0.0182
```

Variable number 2 has 5 categories

```
-----
[  5] Category:    1 Freq.:    0.0694
[  6] Category:    2 Freq.:    0.159
[  7] Category:    3 Freq.:    0.0907
[  8] Category:    4 Freq.:    0.0944
[  9] Category:    5 Freq.:    0.587
```

Variable number 3 has 4 categories

```
-----
[ 10] Category:    1 Freq.:    0.0793
[ 11] Category:    2 Freq.:    0.122
[ 12] Category:    3 Freq.:    0.179
[ 13] Category:    4 Freq.:    0.62
Missing data Num:  1 Freq.:    0.00909
```

Variable number 4 has 2 categories

```
-----  
[ 14] Category:    1  Freq.:    0.745  
[ 15] Category:    2  Freq.:    0.255  
Missing data Num:  2  Freq.:    0.0182
```

Variable number 5 has 4 categories

```
-----  
[ 16] Category:    1  Freq.:    0.0867  
[ 17] Category:    2  Freq.:    0.26  
[ 18] Category:    3  Freq.:    0.291  
[ 19] Category:    4  Freq.:    0.362
```

Variable number 6 has 3 categories

```
-----  
[ 20] Category:    1  Freq.:    0.473  
[ 21] Category:    2  Freq.:    0.133  
[ 22] Category:    3  Freq.:    0.393
```

Variable number 7 has 5 categories

```
-----  
[ 23] Category:    1  Freq.:    0.352  
[ 24] Category:    2  Freq.:    0.047  
[ 25] Category:    3  Freq.:    0.139  
[ 26] Category:    4  Freq.:    0.0545  
[ 27] Category:    5  Freq.:    0.407
```

Variable number 8 has 3 categories

```
-----  
[ 28] Category:    1  Freq.:    0.332  
[ 29] Category:    2  Freq.:    0.502  
[ 30] Category:    3  Freq.:    0.165
```

Variable number 9 has 2 categories

```
-----  
[ 31] Category:    1  Freq.:    0.318  
[ 32] Category:    2  Freq.:    0.682
```

Output file: ESF

Row number: 110, column number: 32

It contains the modified table with type [L1/.../Lv]

[0, 0, ..., 0] are preserved for missing data

File ESF.fuz contains

----- number of rows: 110

----- number of variables: 9

----- number of categories: 32

----- categorie number per variable (vector of 9 values)

----- marginal frequency of modality (vector of 32 values)

Auxiliary binary output file ESFModa: Indicator vector of modalities

It contains variable number for each modality

It has 32 rows (modalities) and one column

Auxiliary ASCII output file ESF.123: labels (two characters) for 32 modalities

It contains one label for each modality

It has 32 rows (modalities) and labels 1a, 1b, ..., 2a, 2b, ...

Variable number 1, 2, ..., A, ..., Z, +, Modality number a, b, ..., z, +

This operation computes the frequency of use of each modality. The table ESF contains the row profile for each variable (The sum of values by group of modalities is equal to 1). The missing values were replaced by the average profile for the corresponding variable. The module allows the use of any row weights. This option was not used here.

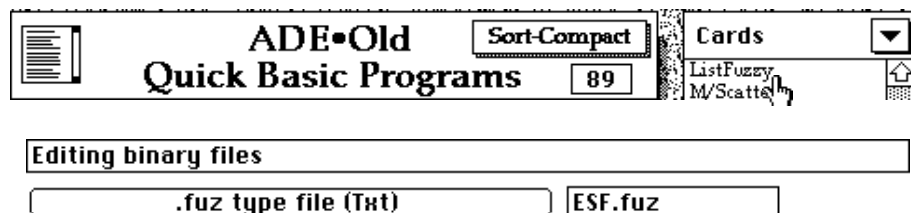
1.3 - Listing data

File ESF may be listed using the option **Edit with** and Excel™ for example as follows :

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.33	0.40	0.13	0.13				0.50	0.50			0.25	0.75
2	0.55	0.27	0.18		0.17	0.50	0.33			0.18	0.36	0.45	
3	0.50	0.33	0.17				0.20	0.40	0.40			0.25	0.75
4	0.33	0.33	0.17	0.17					1.00			0.25	0.75
5	0.50	0.33	0.17		0.20	0.20	0.20	0.20	0.20	0.17	0.33	0.33	0.17
6	0.43	0.29	0.14	0.14					1.00			0.25	0.75
7	0.43	0.29	0.14	0.14					1.00			0.25	0.75
8	0.50	0.33	0.17			0.33	0.33	0.33			0.20	0.60	0.20
9	0.29	0.43	0.14	0.14			0.25	0.50	0.25		0.11	0.22	0.67
10	0.50	0.33	0.17					0.50	0.50		0.25	0.50	0.25
11	0.29	0.43	0.14	0.14					1.00			0.25	0.75
12	0.50	0.33	0.17					0.50	0.50			0.50	0.50
13	0.33	0.33		0.33					1.00				1.00
14	0.43	0.29	0.14	0.14					1.00				1.00
15	0.40	0.40		0.20					1.00				1.00
16		1.00				0.50	0.25	0.25			0.17	0.33	0.50
17		1.00				0.50	0.25	0.25		0.25	0.50	0.25	
18	0.33	0.33	0.17	0.17					1.00				1.00
19	0.33	0.33	0.17	0.17					1.00				1.00
20	0.33	0.33	0.17	0.17					1.00				1.00
21	0.38	0.25	0.13	0.25					1.00				1.00
22		1.00				0.33	0.33	0.33				0.25	0.75
23	0.50	0.50				0.33	0.33	0.33				0.25	0.75
24	0.43	0.29	0.14	0.14					1.00			0.25	0.75
25	0.33	0.33	0.17	0.17					1.00				1.00
26	0.33	0.33	0.17	0.17					1.00			0.25	0.75
27	0.33	0.33	0.17	0.17					1.00			0.25	0.75
28					0.18	0.55	0.27						
29	0.33	0.33	0.17	0.17					1.00				1.00
30	0.50	0.25		0.25		0.50	0.50					0.25	0.75

File ESF may also be listed using the Quickbasic program **ListFuzzy**. The file with .fuz used as suffix for fuzzy coded variables is similar to a file with .cat used as suffix for categorical variables.

Go to **ADE•Old** selection card and select the program as follows :



The listing is as follows:

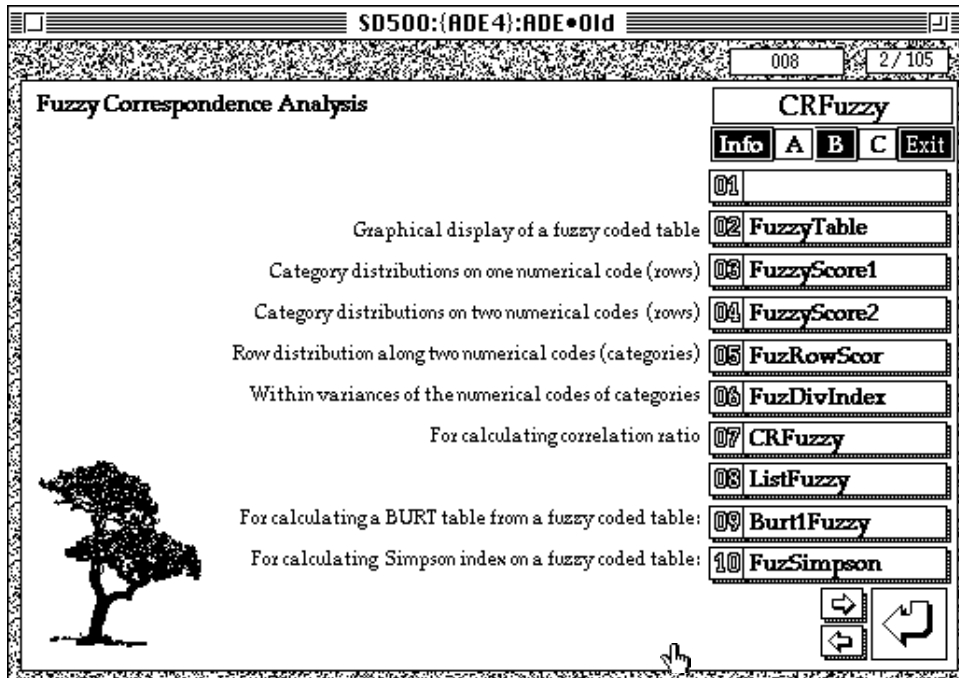
```

Li g 1 /333/400/133/133/ / - / - / - /500/500/ / - / - /250/750/
Li g 2 /545/273/182/ - / /167/500/333/ - / - / /182/364/455/ - /
Li g 3 /500/333/167/ - / / - / - /200/400/400/ / - / - /250/750/
..... etc.
Li g 108 /286/286/143/286/ / - / - / - / - /***/ / - / - /500/500/
Li g 109 /500/250/250/ - / / - / - /500/500/ - / /400/400/200/ - /
Li g 110 /857/143/ - / - / / - / - / - / - /***/ / - / - /429/571/
..... etc.
Li g 1 /200/400/400/ / - /***/
Li g 2 /500/500/ - / /***/ - /
Li g 3 /250/500/250/ / - /***/
..... etc.
Li g 108 /250/500/250/ / - /***/
Li g 109 /500/500/ - / /***/ - /
Li g 110 /250/500/250/ / - /***/

```

The character string **/***/** stands for 100 % of affinity.

All the Quickbasic programs associated with Fuzzy correspondence analysis are incorporated into the following ADE•Old card:



2 - Fuzzy correspondence analysis as an extension of MCA

2.1 - Mathematical principle

Fuzzy correspondence analysis (or homogeneity analysis, see van Rijkevorsel, 1987²) incorporates specific properties ensuing from theorems associated with the duality diagram³.

Let \mathbf{A} be a fuzzy coded table with t species as row and m modalities as columns. Let v be the variable number. Variable 1 has $m(1)$ modalities, variable 2 has $m(2)$ modalities, ..., and variable v has $m(v)$ modalities.

$$m = \sum_{j=1}^v m(j)$$

At the intersection of the i -th species and the k -th modality of the j -th variable lies the value a_{ij}^k with $1 \leq i \leq t$, $1 \leq j \leq v$, and $1 \leq k \leq m(j)$. Let a_{ij}^{\bullet} and p_{ij}^k be respectively equal to

$$a_{ij}^{\bullet} = \sum_{k=1}^{m(j)} a_{ij}^k \text{ and } p_{ij}^k = \frac{a_{ij}^k}{a_{ij}^{\bullet}}$$

Let $\mathbf{P} = [p_{ij}^k]$ be a table with t species as rows and m modalities as columns. Let $p_{i\bullet}^{\bullet}$, $p_{\bullet\bullet}^{\bullet}$ and $p_{\bullet\bullet}^{\bullet}$ be respectively equal to

$$p_{ij}^{\bullet} = \prod_{k=1}^{m(j)} p_{ij}^k = 1, p_{i\bullet}^{\bullet} = \prod_{j=1}^v p_{ij}^{\bullet} = v, \text{ and } p_{\bullet\bullet}^{\bullet} = \prod_{i=1}^t p_{i\bullet}^{\bullet} = tv$$

Row weights of table **P** are uniformly equal to $r_i = \frac{v}{tv} = \frac{1}{t}$ and column weights of table **P** are equal to

$$c_j^k = \frac{\prod_{i=1}^t p_{ij}^k}{tv} = \frac{1}{v} \left(\frac{1}{t} \prod_{i=1}^t p_{ij}^k \right) = \frac{1}{v} \overline{p_j^k}$$

If the affinity of a species for a variable is unknown, it is coded "0, 0, 0, 0". It is thus implicitly replaced by the average profile for the corresponding variable. As a consequence, a species with such a score is not taken into account in the calculation of the column weight. Table **P** can be processed by correspondence analysis. If we use the notations of the duality diagram, the correspondence analysis of a fuzzy coded table or fuzzy correspondence analysis is the principal components analysis of the statistical triplet:

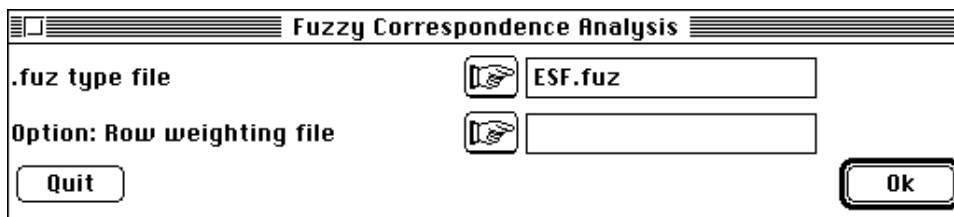
$$\left(\frac{p_{ij}^k}{\overline{p_j^k}} - 1, \text{Diag}(c_j^k), \text{Diag}(r_i) \right)$$

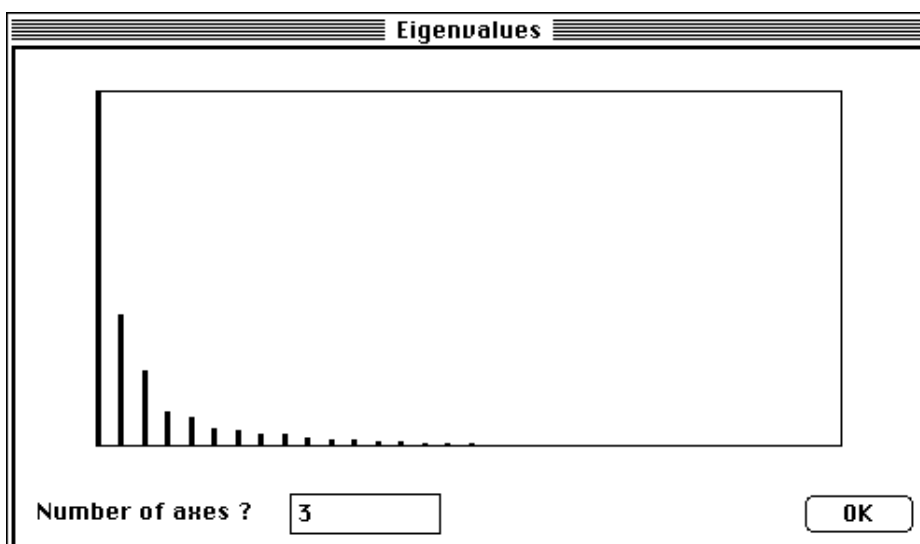
where $\text{Diag}(w_l)$ is the matrix

$$\text{Diag}(w_l) = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_L \end{pmatrix}$$

2.2 - Computation

Select the option **Fuzzy Correspondence Analysis** of the module **MCA**:





Select three axes. You should note that the first axis is quite prominent. This result in a listing as follows:

```
f1/FuzzyMCA: Multiple correspondence analysis on fuzzy table
Input file: ESF.fuz for access to file ESF
Row number: 110, column number: 32
Uniform row weights
```

```
File ESF.flpl contains the row weights
It has 110 rows and 1 column
```

```
File ESF.flta contains the table processed by MCA
It has 110 rows and 32 columns (categories)
```

```
File ESF.flma contains
----- number of rows: 110
----- number of variables: 9
----- number of categories: 32
----- variable number of each category (vector of 32 values)
```

```
File ESF.flpc contains the column weights (1/V)*DM
It has 32 rows and 1 column
```

The three first files form a statistical triplet $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$. ESF.flpl (fl for "flou" in French) incorporates the row weights which are uniformly equal to $1/110 = 0.0091$.

ESF.flpc incorporates the column weights as follows:

1	0.0520	11	0.0135	22	0.0437
2	0.0290	12	0.0199	23	0.0391
3	0.0169	13	0.0689	24	0.0052
4	0.0133	14	0.0828	25	0.0155
5	0.0077	15	0.0283	26	0.0061
6	0.0176	16	0.0096	27	0.0452
7	0.0101	17	0.0289	28	0.0369
8	0.0105	18	0.0323	29	0.0558
9	0.0652	19	0.0402	30	0.0184
10	0.0088	20	0.0526	31	0.0354
		21	0.0148	32	0.0758

The sum of the column weights by group of modalities is uniformly equal to $1/9 = 0.111$. These values result from the transformation of the fuzzy coded data into

percentages per variable. The average profile of species presenting data is computed for each variable. The average profile is divided by the number of variables (nine in our example) resulting in values of ESF. f1 pc.

In the file ESF. f1 t a values as follows are incorporated:

$$x_{ij} = \frac{\text{Affinity of species } i \text{ for modality } j}{P_i \cdot P_j} \text{ number of variables}$$

This operation is similar to the classical correspondence analysis processed on the table that contains the percentage of affinity of species for each group (variable) of modalities.

The listing continues with the following information:

Marginal distributions by variable:

Variable number 1 has 4 categories

[1]	Category:	1	Weight:	0.468
[2]	Category:	2	Weight:	0.261
[3]	Category:	3	Weight:	0.152
[4]	Category:	4	Weight:	0.119

Variable number 2 has 5 categories

[5]	Category:	1	Weight:	0.0694
[6]	Category:	2	Weight:	0.159
[7]	Category:	3	Weight:	0.0907
[8]	Category:	4	Weight:	0.0944
[9]	Category:	5	Weight:	0.587

Variable number 3 has 4 categories

[10]	Category:	1	Weight:	0.0793
[11]	Category:	2	Weight:	0.122
[12]	Category:	3	Weight:	0.179
[13]	Category:	4	Weight:	0.62

Variable number 4 has 2 categories

[14]	Category:	1	Weight:	0.745
[15]	Category:	2	Weight:	0.255

Variable number 5 has 4 categories

[16]	Category:	1	Weight:	0.0867
[17]	Category:	2	Weight:	0.26
[18]	Category:	3	Weight:	0.291
[19]	Category:	4	Weight:	0.362

Variable number 6 has 3 categories

[20]	Category:	1	Weight:	0.473
[21]	Category:	2	Weight:	0.133
[22]	Category:	3	Weight:	0.393

Variable number 7 has 5 categories

[23]	Category:	1	Weight:	0.352
[24]	Category:	2	Weight:	0.047

```
[25] Category: 3 Weight: 0.139
[26] Category: 4 Weight: 0.0545
[27] Category: 5 Weight: 0.407
```

Variable number 8 has 3 categories

```
-----
[28] Category: 1 Weight: 0.332
[29] Category: 2 Weight: 0.502
[30] Category: 3 Weight: 0.165
```

Variable number 9 has 2 categories

```
-----
[31] Category: 1 Weight: 0.318
[32] Category: 2 Weight: 0.682
-----
```

A fundamental property in fuzzy correspondence analysis is that the number of eigenvalues equal to 0 is equal to the number of variable. This may be verified in the listing below:

```
-----
DiagoRC: General program for two diagonal inner product analysis
```

```
Input file: ESF.flta
```

```
--- Number of rows: 110, columns: 32
```

```
-----
Total inertia: 1.01509
```

```
-----
Num. Eigenval. R. Iner. R. Sum | Num. Eigenval. R. Iner. R. Sum |
01 +4.9546E-01 +0.4881 +0.4881 | 02 +1.8411E-01 +0.1814 +0.6695 |
03 +1.0530E-01 +0.1037 +0.7732 | 04 +4.8528E-02 +0.0478 +0.8210 |
05 +4.1277E-02 +0.0407 +0.8617 | 06 +2.6060E-02 +0.0257 +0.8873 |
07 +2.1283E-02 +0.0210 +0.9083 | 08 +1.6103E-02 +0.0159 +0.9242 |
09 +1.5835E-02 +0.0156 +0.9398 | 10 +1.3013E-02 +0.0128 +0.9526 |
11 +1.0411E-02 +0.0103 +0.9628 | 12 +8.1362E-03 +0.0080 +0.9709 |
13 +6.5830E-03 +0.0065 +0.9773 | 14 +5.3510E-03 +0.0053 +0.9826 |
15 +4.9017E-03 +0.0048 +0.9874 | 16 +3.5256E-03 +0.0035 +0.9909 |
17 +2.8509E-03 +0.0028 +0.9937 | 18 +2.3119E-03 +0.0023 +0.9960 |
19 +1.7797E-03 +0.0018 +0.9978 | 20 +1.1714E-03 +0.0012 +0.9989 |
21 +7.5992E-04 +0.0007 +0.9997 | 22 +2.7709E-04 +0.0003 +0.9999 |
23 +6.4894E-05 +0.0001 +1.0000 | 24 +0.0000E+00 +0.0000 +1.0000 |
25 +0.0000E+00 +0.0000 +1.0000 | 26 +0.0000E+00 +0.0000 +1.0000 |
27 +0.0000E+00 +0.0000 +1.0000 | 28 +0.0000E+00 +0.0000 +1.0000 |
29 +0.0000E+00 +0.0000 +1.0000 | 30 +0.0000E+00 +0.0000 +1.0000 |
31 +0.0000E+00 +0.0000 +1.0000 | 32 +0.0000E+00 +0.0000 +1.0000 |
-----
```

```
File ESF.flvp contains the eigenvalues and relative inertia for each axis
```

```
--- It has 32 rows and 2 columns
```

```
File ESF.flco contains the column scores
```

```
--- It has 32 rows and 3 columns
```

```
File ESF.flli contains the row scores
```

```
--- It has 110 rows and 3 columns
```

To have a graphical representation of the fuzzy coded table, select the Quickbasic program **FuzzyTable** via the **ADE•Old** selection card:

Fuzzy table diagramm	
---.flma type input file (Txt)	ESF.flma
Option : Row ordination (Bin)	
Column used (default=1)	
Option : Row label (Txt)	ES_CodeEsp
Option : Modality ordination (Bin)	
Column used (default=1)	

For drawing the left part of Fig. 2 select the following size:

Rectangle window width :	300
Rectangle window height :	1200

The resulting graphic (Fig. 2) may be stored.

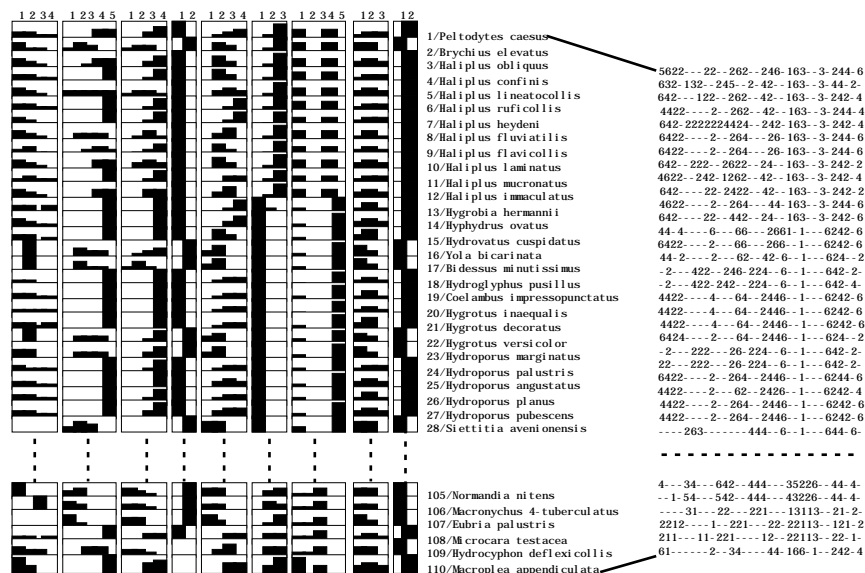


Figure 2 On the left edition of the table using **FuzzyTable**. On the right edition of the raw data. The image was reduced (50%).

2.3 - Interpretation

2.3.2 - Typology of modalities

Mathematical support is given in Appendix 1.

2.3.2.1 - Correlation ratios

This is the first aid for interpreting fuzzy coded data. Fuzzy correspondence analysis aims to compute numerical codes for species to separate the modalities of variables or it aims to compute numerical codes for variables to discriminate the species affinity distribution among modalities. The calculation of correlation ratio enables a fast check of which variables are involved in the separation of species profiles.

Select the Quickbasic program **CRFuzzy** via the **ADE•Old** selection card and fill in the dialog box as follows:

Fuzzy coding variable analysis

File ---.flta type(Bin)	ESF.flta	110	32
-------------------------	----------	-----	----

```

Variable number 1
> Categ= 1 Freq=0.468 -0.248 -0.108 -0.132
> Categ= 2 Freq=0.261 0.265 0.184 0.208
> Categ= 3 Freq=0.152 -0.045 -0.061 0.074
> Categ= 4 Freq=0.119 0.450 0.097 -0.032
-----> r=0.072 r=0.016 r=0.020

Variable number 2
> Categ= 1 Freq=0.069 -1.853 -0.300 -0.935
> Categ= 2 Freq=0.159 -1.301 0.264 0.103
> Categ= 3 Freq=0.091 -0.654 0.514 1.093
> Categ= 4 Freq=0.094 -0.024 0.057 1.312
> Categ= 5 Freq=0.587 0.676 -0.124 -0.297
-----> r=0.814 r=0.051 r=0.385

Variable number 3
> Categ= 1 Freq=0.079 -1.778 -0.232 -0.625
> Categ= 2 Freq=0.122 -1.247 0.062 0.258
> Categ= 3 Freq=0.179 -0.348 0.123 0.761
> Categ= 4 Freq=0.620 0.573 -0.018 -0.191
-----> r=0.666 r=0.008 r=0.165

Variable number 4
> Categ= 1 Freq=0.745 0.519 -0.096 0.064
> Categ= 2 Freq=0.255 -1.520 0.282 -0.187
-----> r=0.789 r=0.027 r=0.012

Variable number 5
> Categ= 1 Freq=0.087 -1.737 0.102 -0.732
> Categ= 2 Freq=0.260 -0.688 0.280 0.387
> Categ= 3 Freq=0.291 0.322 0.009 0.280
> Categ= 4 Freq=0.362 0.652 -0.233 -0.328
-----> r=0.569 r=0.041 r=0.147

Variable number 6
> Categ= 1 Freq=0.473 0.420 0.945 -0.087
> Categ= 2 Freq=0.133 -0.627 -0.765 -0.216
> Categ= 3 Freq=0.393 -0.293 -0.878 0.179
-----> r=0.170 r=0.804 r=0.022

Variable number 7
> Categ= 1 Freq=0.352 0.233 -0.686 0.013
> Categ= 2 Freq=0.047 -1.413 -0.448 -0.322
> Categ= 3 Freq=0.139 -1.410 -0.456 -0.354
> Categ= 4 Freq=0.055 0.210 -1.131 1.752
> Categ= 5 Freq=0.407 0.417 0.953 -0.087
-----> r=0.464 r=0.644 r=0.193

Variable number 8
> Categ= 1 Freq=0.332 -0.340 0.097 -0.021
> Categ= 2 Freq=0.502 0.019 0.023 0.024
> Categ= 3 Freq=0.165 0.626 -0.264 -0.031
-----> r=0.104 r=0.015 r=0.001

Variable number 9
> Categ= 1 Freq=0.318 -1.320 0.334 0.061
> Categ= 2 Freq=0.682 0.616 -0.156 -0.028
-----> r=0.813 r=0.052 r=0.002

```

```

File ESF.flrc is binary
Number of rows = 9 Number of columns = 3
It contains correlation ratios

```

The highest values of correlation ratios are in bold in the above listing. These correlation ratio values represent percentage of variance explained by a given axis. Each species is given a score for each axis (factorial score). The variance of these scores on one axis is equal to the corresponding eigenvalue.

Each modality is positioned at the weighted average scores of species that are presenting that modality (i.e., using the affinity percentage of the species for the modality). The variance of these positions is obviously lower than the total variance. Its value is high if the species ordination is relevant for discriminating modalities. The result is expressed as a percentage of explained variance. Fuzzy correspondence analysis aims to maximize the average of these percentages. Thereby, this average correlation ratio is equal to the corresponding eigenvalue. Consequently, correlation ratios permit to order the relevance of ecological requirements for the various axes. Hence, on the axis F1 of the fuzzy correspondence analysis, the modalities of the variables "sediment", "pollution", "temperature amplitude" with nearly 80% of explained variance are slightly more separated than those of variables "velocity" and "saprobity". The modalities of the variables "diet" and "feeding" are much more separated on axis F2 with a percentage of explained variance, respectively, equal to 80% and 65%.

2.3.2.1 - One-axis presentation

Select the Quickbasic program **FuzzyScore1** via **ADE•Old** selection card to display the results along axis 1. This means that the ordination of ES (analysis type fl) will be displayed according to the score ES.fl1i (factorial scores of species). For this operation, one can use any quantitative variable as row scoring file (row weights for example). ES_CodeVar enables a labelling of graphics:

Fuzzy scoring univariate diagram	
---.flma type input file (Txt)	ESF.flma
Option : Row scoring file (Bin)	ESF.flii
Selected column (default=1)	
Option : variable labels (Txt)	ES_CodeVar

A new dialog window occurs. The figure (10 by default) indicates the number of classe selected for each histogram:

Class number for histogramm	10
-----------------------------	----

Follow the procedure by selecting the limits of the graphics, the limit of the graphical window and the number of graphics as follows:

Window width :	400
Window height :	600
Horizontal window number :	5
Vertical window number :	12

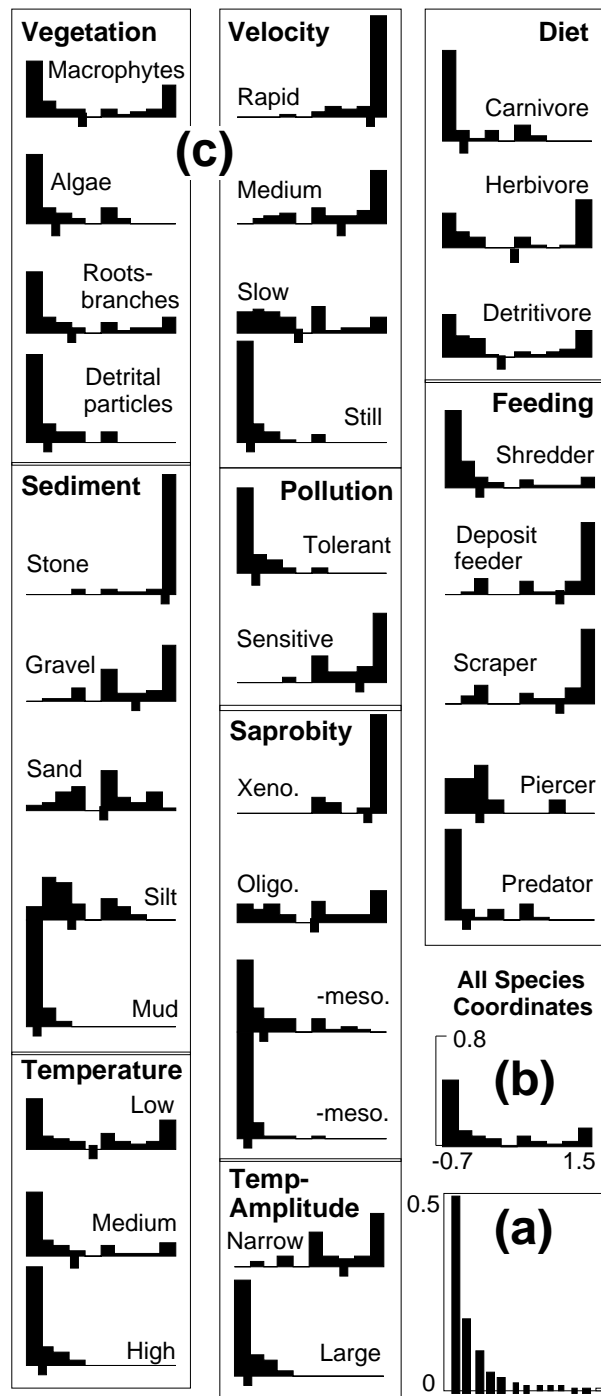
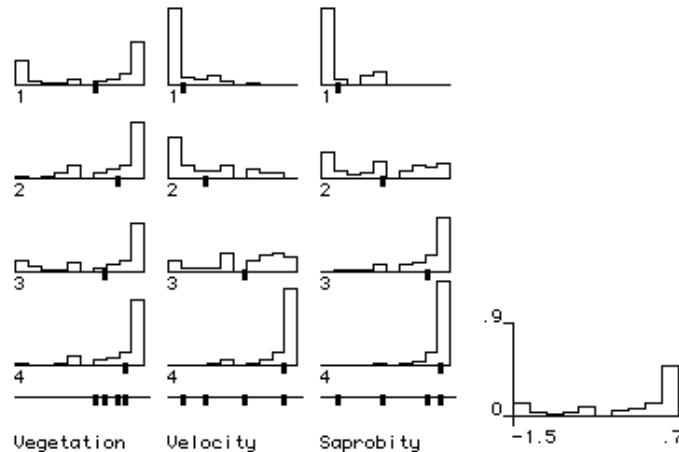


Figure 3 Presentation of ecological requirements on one factorial axis using the frequency distributions of coordinates after fuzzy correspondence analysis. (a) Histogram of eigenvalues. (b) Distribution in ten classes of all species coordinates (such species coordinates have a variance corresponding to the λ_1 -eigenvalue; multiplication of these coordinates by $1/\sqrt{\lambda_1}$ makes the variance equal to 1). (c) Each modality of each variable has a frequency distribution of its use by species, and the mark below each distribution stands for the average of that distribution.

Mini X-axis:	-1.5
Maxi X-axis:	.7
Mini Y-axis:	0
Maxi Y-axis:	.9

This results in the following graphics for which only three variables are presented (Vegetation, Velocity, Saprobity):



In this representation, the row of the original raw table are positioned according to axis 1. The distribution of all the species scores is represented by the histogram on the right. Each modality of each variable has its own distribution of species score (histogram) according to species presenting that modality. The average of that distribution is represented by a black mark. Because the row (species) distribution is common to all the variables, it allows to visualize simultaneously the separation among modalities of a variable and the association of modalities among variables. The concentration of plots at the extremity of the gradient are rarely found in such an analysis.

After some trial and modifications using MacDraw™, Fig. 3 can be obtained. This one-axis presentation (Fig. 3) demonstrates a correlation between "velocity" (rapid to still) and "sediment" (stone to mud) with each showing a similar gradient. "Saprobity" (modalities -mesosaprobitic and -mesosaprobitic) is expectedly correlated with "pollution" tolerance. The two modalities of "temperature amplitude" (narrow, large) are clearly separated on axis 1. "Feeding" types are divided into two groups: shredders, piercers, and predators; as opposed to deposit feeders and scrapers. Finally, modalities of variables such as "vegetation", "temperature" and "diet" are not clearly separated.

2.3.2.2 - Two-axes presentation

The two-axes presentation uses the same logic. Select the basic **FuzzyScore2** via the **ADE•Old** selection card:

Fuzzy scoring bivariate diagram		
---.flma type input file (Txt)	ESF.flma	
HY-file (Bin)	ESF.flli	110 3
Selected X-column (default=1)		
Selected Y-column (default=2)		
Option : variable labels (Txt)	ES_CodeVar	

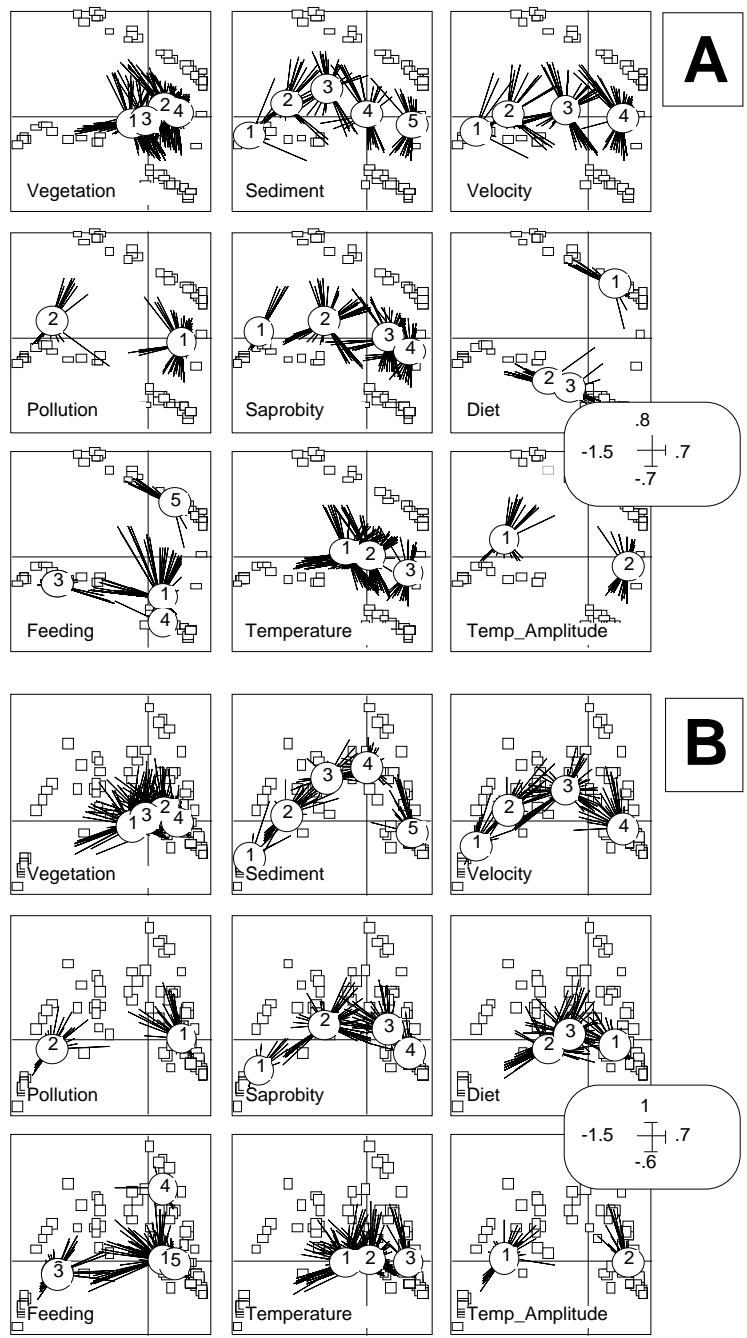


Figure 4 Ordination of ecological requirements. A - F1x2 factorial plane resulting from fuzzy correspondence analysis. Small squares represent the species positions. Each modality (number in a circle) of each variable is positioned at the weighted average of species representing that modality. Lines link species to their modality but are only 50% of their total length for readability. B - F1x3 factorial plane.

Select the limits of graphics:

Mini X-axis:	-1.5
Maxi X-axis:	.7
Mini Y-axis:	-.7
Maxi Y-axis:	.8

Select the dimension of the graphical window and the number of graphics per window:

Window width :	300
Window height :	300
Horizontal window number :	3
Vertical window number :	3

Select the length of lines that link species to their modalities. For readability, type in "0.5" indicating that these lines should be 50% of their total length:

Size coefficient	.5
------------------	----

The last coefficient allows to take into account species that contribute more than a given threshold to the modality distribution:

Minimal percentage	0
--------------------	---

A two-axes presentation shows the presence of two independent structures (Fig. 4A) in ecological requirements of aquatic Coleoptera species: (i) the affinity for "sediment", "velocity", the sensitivity to "pollution", "saprobity", and "temperature amplitude" on axis 1; and (ii) the trophic status (i.e. "feeding" and "diet") on axis 2. As a result, variables with a high correlation ratio on axis 1 show a clear arrangement of modalities along a gradient whereas variables having a high correlation ratio on axis 2 separate two different groups of modalities. A Guttman effect is demonstrated on the F1xF3 factorial plane (Fig. 4B).

2.3.3 - Typology of species

Species may be simply plotted using **Scatters** on ESF.fl1i as in a classical correspondence analysis. However, The relation between modalities (columns of the initial table) and species (rows of the initial table) may be illustrated in a symmetrical way. Select **FuzRowScor** via the ADE•Old selection card:

Fuzzy Row Scoring			
File ---.flup type (Bin)	ESF.flup	32	2
X-axis column (default=1)			
Y-axis column (default=2)			
1=Row label (default=no)			
Row label file (Txt)			
1=Categ label (default=no)	1		
Categ label file (Txt)	ES_CodeModa		

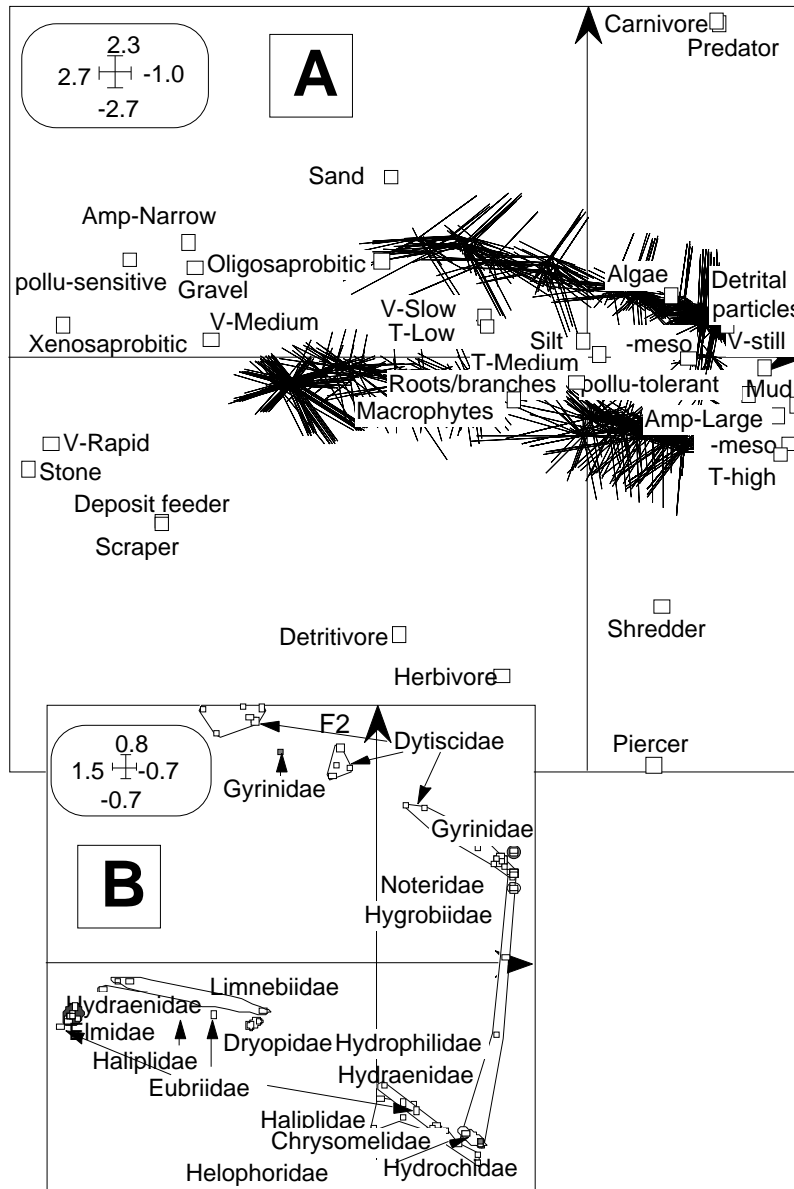


Figure 5 A - Typology of species on the F1xF2 factorial plane by fuzzy correspondence analysis. Small squares represent the modality positions (V for "velocity"; T for "temperature"; Amp for "temperature amplitude"). Each species profile is at the weighted average of modalities that are represented by the species. Lines link species to their modality but are only 33.3% of their total length for readability. B - Typology of species grouped by family; contrasting grey areas identify the Coleoptera families.

This results in Fig. 5A. In addition, a graphic of the species coordinates grouped by family (projection of ESF.fili using a family categorical variable) may be realized (Fig. 5B). The information in Figure 4B is completed. Such a typology informs about the diversity of ecological requirements within each taxonomic group if compared to Fig. 5A.

Although the ecological requirements of Coleoptera larvae and adults have been pooled by us, the typology of the Coleoptera at the family level looks similar to that of Bournaud *et al.* (op.cit.), where adults and larvae were not pooled. The family Dytiscidae groups species that are either oligosaprobic and rather rheophilic or still-water species that are much more tolerant to pollution. All Dytiscidae are predators. Gyrinidae, Noteridae, and Hygrobiidae also are predators, the majority of which live in

still waters. Non-predaceous species include: (i) rheophilic species sensitive to pollution (e.g. Elmidae and some Hydraenidae) that are deposit-feeders or scrapers; or (ii) still-water species (e.g. Helophoridae, Haliplidae, and some Hydraenidae) that are shredders or piercers tolerant to pollution. The family Hydrophilidae groups predaceous and non-predaceous species that live in still waters having a large temperature amplitude.

Appendix 1 - Typology of modalities

Let each species be positioned by a score $\mathbf{x}=(x_1, \dots, x_j, \dots, x_t)$, with the mean equal to 0 and the variance equal to $\text{var}(\mathbf{x})$. A modality of a variable is positioned by averaging as follows:

$$\bar{x}_j^k = \frac{\sum_{i=1}^t p_{ij}^k x_i}{\sum_{i=1}^t p_{ij}^k} \quad \text{so that} \quad \sum_{k=1}^{m(j)} \bar{x}_j^k = 0$$

The variance of the modality scores of one variable is equal to

$$\text{var}_j = \sum_{k=1}^{m(j)} p_j^k \bar{x}_j^k^2$$

The correlation ratio r_j of variable j for the score \mathbf{x} quantifies the discrimination of modalities of variable j as follows:

$$r_j = \frac{\text{var}_j}{\text{var}(\mathbf{x})}$$

The average of these correlation ratios quantifies the relevance of the score \mathbf{x} for separating modalities among variables:

$$r(\mathbf{x}) = \frac{1}{v} \sum_{j=1}^v r_j$$

The first coordinate of correspondence analysis is the best row score for maximizing $r(\mathbf{x})$, with the maximum r_1 being equal to the first eigenvalue. The same is true for a new score not correlated with the first one, such as the second coordinate of correspondence analysis, and so on.

Appendix 2 - Typology of species

Let each modality be positioned by a score \mathbf{y} :

$$\mathbf{y} = \left(y_j^k \right)_{j=1, \dots, v, k=1, \dots, m(j)}$$

We suppose that these scores are centred (mean equal to 0) for each variable. Let the variance of these scores be equal to var_j . Let $\text{var}(\mathbf{y})$ be equal to

$$\text{var}(\mathbf{y}) = \frac{1}{v} \sum_{j=1}^v \text{var}_j$$

For each row, the mean score of modalities for one variable is equal to

$$\bar{y}_{ij} = \frac{1}{m(j)} \sum_{k=1}^m p_{ij}^k y_j^k$$

and the overall mean is equal to:

$$\bar{y}_i = \frac{1}{v} \sum_{j=1}^v \bar{y}_{ij}$$

The new row score, resulting from that double averaging, has a mean equal to 0 and a variance equal to:

$$w(\mathbf{y}) = \frac{1}{t} \sum_{i=1}^t y_i^2$$

and the capacity of score \mathbf{y} (positioning modalities) for separating rows (species) is equal to

$$R(\mathbf{y}) = \frac{w(\mathbf{y})}{\text{var}(\mathbf{y})} - 1$$

The best modality score \mathbf{y} is given by the first coordinate of the correspondence analysis of \mathbf{P} .

Références

- ¹ Bournaud, M., Richoux, P. & Usseglio-Polatera, P. (1992) An approach to the synthesis of qualitative ecological information from aquatic coleoptera communities. *Regulated rivers: Research and Management* : 7, 165-180.
- ² Rijckevorsel, J. van. (1987) The application of fuzzy coding and horseshoes in multiple correspondence analysis. DSWO Press, Leiden. 1-272.
- ³ Chevenet, F., Dolédec, S. & Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* : 31, 295-309.