



DMAUse

Introduction-----	1
DMAUse : BioDiv Indices -----	2
DMAUse : Classical Indices -----	7
DMAUse : Correlation matrices -----	9
DMAUse : DMA Weighths -----	15
DMAUse : Mantel-RV test -----	23
DMAUse : Principal Coordinates Analysis-----	34
Références-----	49

Introduction

Le module DMAUtil permet d'implanter des matrices de distances entre objets. Ces matrices sont utilisées dans le présent module.

Pour une matrice de distances euclidiennes, il existe un nuage de points dans un espace euclidien dont les distances deux à deux sont celles de la matrice. L'option *Principal Coordinates Analysis* calcule ce nuage de points et en donne une projection en dimensions réduites. L'analyse a été profondément modifiée pour accepter des matrices créant des problèmes numériques dans la version précédente. Elle tolère les poids nuls et les points multiples.

Les options *Correlation matrices* et *Mantel-RV test* permettent de calculer et tester la corrélation entre matrices de distances. Un test nouveau basé sur le RV d'Escoufier est introduit.

Trois options sont consacrées aux calculs d'indice de biodiversité. La première *Classical Indices* donnent quatre indices classiques (Richesse, Simpson, Shannon, Hill). La seconde *BioDiv Indices* introduit des mesures de diversité basée sur des matrices de distances entre espèces. Elles sont basées sur les travaux de Rao. La troisième *DMA Weights* donne les distributions théoriques qui maximisent ces indices de diversité.

Le module contient des innovations en cours de publications ¹.

¹ Champely, S. & Chessel, D. (2001) Measuring biological diversity using Euclidean metrics. *Environmental and Ecological Statistics* : in press. Thioulouse, J. & Chessel, D. Matching two euclidean distance matrices (soumis).

DMAUse : BioDiv Indices

Type Utilitaire de calcul d'indices de diversité basée sur une matrice de distances.

Objet Mesurer la diversité sur des bases empiriques conduit à une multitude d'indices et une énorme littérature comparative. L'ouvrage de Magurran ¹ s'achève par un chapitre «6 The empirical value of diversity measures» dont le dernier paragraphe «Why use diversity measures?» commence par «Somme ecologists reject diversity indices and use of species abundance distributions in favour of simple counts of number of species». Seul Rao ² a défini une base théorique unificatrice qui relie diversité écologique (les indices), diversité génétique (les distances) et mesures générales de la variabilité (les variances). Il a défini lui-même ces bases comme une axiomatisation dans ³. Nous avons apporté les aménagements nécessaires pour mettre à disposition cette théorie en écologie. Soit un relevé écologique contenant s espèces avec les proportions (p_1, p_2, \dots, p_s) . Soit :

$$\mathbf{D} = [d_{ij}]_{\substack{1 \leq i \leq s \\ 1 \leq j \leq s}}$$

une matrice de distances entre espèces. Les mesures de Rao sont basée sur l'hétérogénéité :

$$H_p = \sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}$$

Pour des raisons théoriques, on utilisera les carrés des distances. L'option calcule 4 types d'indices, respectivement :

$$R_s(p_1, \dots, p_s) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2 \text{ Indice de Rao simple;}$$

$$R_c(p_1, \dots, p_s) = \frac{\sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2}{\mathbf{a} \sum_{j=1}^s p_j^2} \text{ Indice de Rao canonique;}$$

$$R_m(p_1, \dots, p_s) = \frac{\sum_{i=1}^s \sum_{j=1}^s \sqrt{p_i} \sqrt{p_j} d_{ij}^2}{\mathbf{a}} \text{ Indice de Rao métrique;}$$

$$R_n(p_1, \dots, p_s) = \frac{\sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2}{\max_{(q_1, \dots, q_s)} \left(\sum_{i=1}^s \sum_{j=1}^s q_i q_j d_{ij}^2 \right)} \text{ Indice de Rao numérique;}$$

La diversité simple existe toujours. La diversité canonique n'existe que si \mathbf{D} est euclidienne. Dans ce cas, on démontre que la matrice des carrés des distances :

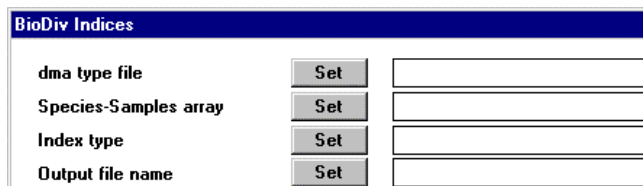
$$\mathbf{D}^{\bullet 2} = [d_{ij}^2]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

admet une seule valeur propre positive α . La diversité canonique est comprise entre 0 et 1 et vaut 1 pour la pondération canonique.

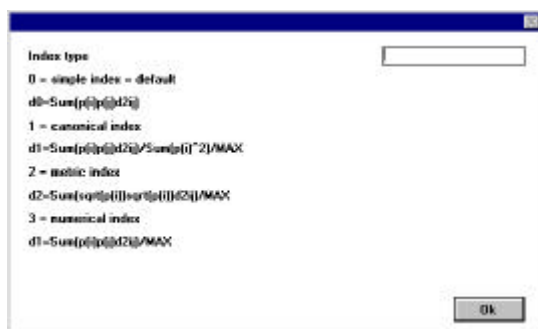
La diversité métrique n'existe que si \mathbf{D} est euclidienne. La diversité métrique est comprise entre 0 et 1 et vaut 1 pour la pondération métrique.

La diversité numérique n'existe que si \mathbf{D} est euclidienne. La diversité numérique est comprise entre 0 et 1 et vaut 1 pour la pondération numérique.

Dialogue L'option utilise une seule fenêtre de dialogue :

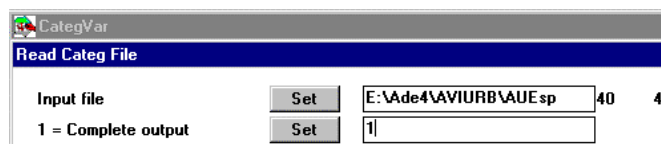


- 1) Fichier de type dma donnant accès à une famille de matrices de distances
- 1) Tableau Taxons (lignes) - Relevés (colonnes)
- 3) Type d'indices calculés 0 (défaut) pour simple, 1 pour canonique, 2 pour métrique et 3 pour numérique :



- 4) Nom du fichier de sortie.

Exemple Utiliser la carte AviUrb. Transposer le tableau faunistique AUFau en FauTR. D'autre part calculer la distances entre espèces sur leur traits biologiques. Lire le fichier de variables qualitatives :



```
Categorical variables: file E:\Ade4\AVIURB\AUEsp
Rows: 40, Variables: 4, Categories: 12, Missing data: 0
```

```
Description of categories:
-----
Variable number 1 has 3 categories
-----
[ 1]Category: 1 Num: 19 Freq.: 0.475
[ 2]Category: 2 Num: 12 Freq.: 0.3
[ 3]Category: 3 Num: 9 Freq.: 0.225
-----
Variable number 2 has 3 categories
-----
```

```
[ 4]Category: 1 Num: 27 Freq.: 0.675
[ 5]Category: 2 Num: 3 Freq.: 0.075
[ 6]Category: 3 Num: 10 Freq.: 0.25
```

Variable number 3 has 4 categories

```
-----
[ 7]Category: 1 Num: 6 Freq.: 0.15
[ 8]Category: 2 Num: 14 Freq.: 0.35
[ 9]Category: 3 Num: 12 Freq.: 0.3
[10]Category: 4 Num: 8 Freq.: 0.2
```

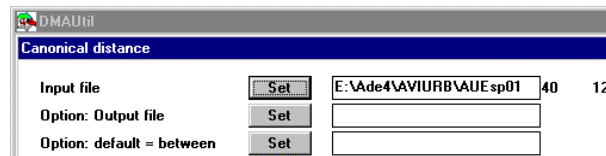
Variable number 4 has 2 categories

```
-----
[11]Category: 1 Num: 23 Freq.: 0.575
[12]Category: 2 Num: 17 Freq.: 0.425
-----
```

Le passer en disjonctif complet :



Calculer la distance canonique entre lignes sur le fichier de sortie :



Distance matrix computation

```
-----
Input file: E:\Ade4\AVIURB\AUEsp01
It has 40 rows and 12 columns
Distances are computed among rows
```

```
-----
Canonical distances computed
Output file: E:\Ade4\AVIURB\AUEsp01_EU
It has 780 rows and 1 columns
d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
Text file: E:\Ade4\AVIURB\AUEsp01_EU.dma
1 -> 40
2 -> 1
3 -> Classical metric on E:\Ade4\AVIURB\AUEsp01
4 -> TRUE
-----
```

Editer le résultat :



```
Input file E:\Ade4\AVIURB\AUEsp01_EU Col 1
max value = 2.82843e+00
Content as 1000*x/max
```

```
-----
[ 1] 0
[ 2] 500 0
[ 3] 500 500 0
[ 4] 866 707 866 0
[ 5] 866 866 866 500 0
...
-----
```

```

-----
Binary input file: E:\Ade4\AVIURB\AUEsp - 40 rows, 4 cols.
 1 | 3.0000 1.0000 4.0000 2.0000
 2 | 3.0000 1.0000 2.0000 2.0000
 3 | 3.0000 1.0000 1.0000 2.0000
 4 | 2.0000 1.0000 2.0000 1.0000
 5 | 2.0000 1.0000 3.0000 1.0000

```

Le carré de cette distance vaut exactement 2 fois le nombre de variables pour lesquelles deux espèces n'utilisent pas la même modalité. Le maximum est donc $\sqrt{8} = 2.82843$.

Les espèces 1 et 4 n'utilisent pas la même modalité pour 3 des 4 variables. Leur distance est de $\sqrt{6} = 2.44989$ soit par rapport au maximum $\frac{\sqrt{6}}{\sqrt{8}} = 0.866$.

Les indices de diversité vont mesurer pour chaque relevé la différence moyenne entre espèces présentes dans ce relevé. Cette diversité biologique sera nulle si toutes les espèces du relevés ont le même profil biologique.

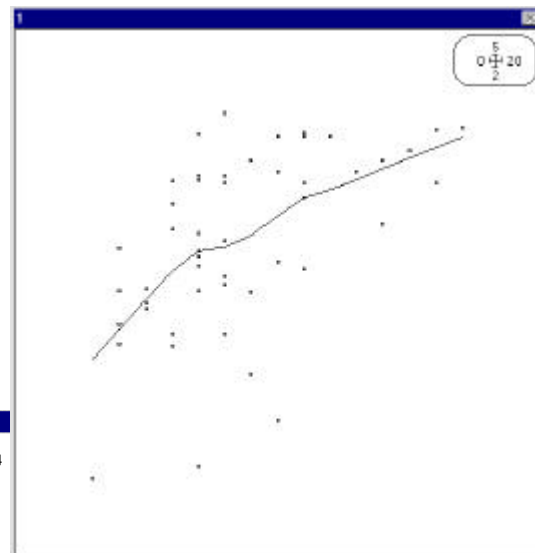
BioDiv Indices			
dma type file	Set	4\AVIURB\AUEsp01_EU.dma	
Species-Samples array	Set	E:\Ade4\AVIURB\FauTR	40 51
Index type	Set		
Output file name	Set	divbiosim	

```

Option: 0
File divbiosim contains biodiversity indices
Rows (sites): 51 Col (distances): 1
Sum(p(i)*p(j)*d2(i,j)) without rescaling

```

Le lien avec la richesse est nettement plus faible (voir DMAUse: Classical Indices):



Lowess			
X file (default = 1, 2, 3, ..., n)	Set	E:\Ade4\AVIURB\FauTR.div	51 4
X file column number (default	Set		
Y file (no default)	Set	E:\Ade4\AVIURB\divbiosim	51 1
Number of points (Min=4) ?	Set	35	

Relancer l'option pour obtenir les autres possibilités :

```

Option: 1
1.872890e+02 Valeur de alpha

```

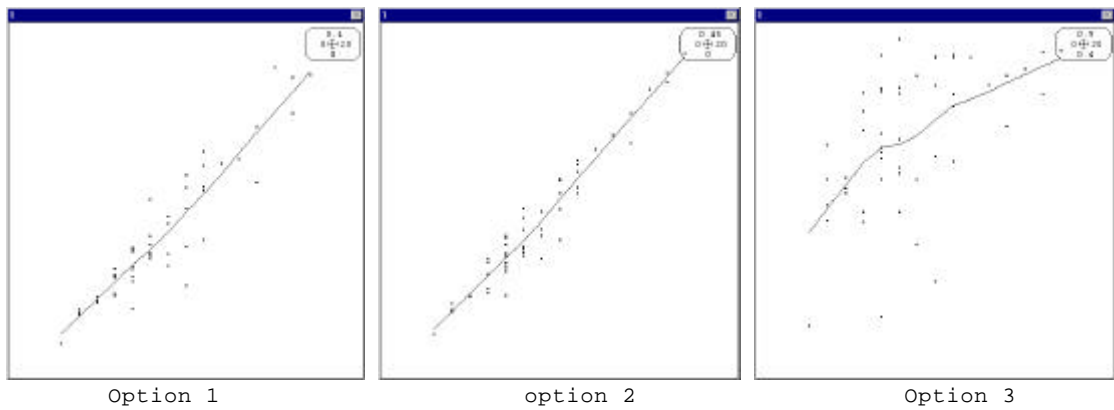
File divbiocan contains biodiversity indices
Rows (sites): 51 Col (distances): 1
Sum($p(i)*p(j)*d2(i,j)$)/Sum($p(i)*p(i)$) with rescaling on [0,1]

Option: 2
1.872890e+02

File divbiomet contains biodiversity indices
Rows (sites): 51 Col (distances): 1
Sum($\sqrt{p(i)*p(j)}*d2(i,j)$) with rescaling on [0,1]

Option: 3
Convergence : 1 45 5.157e+00 1.116415e-01
5.156794e+00

File divbionum contains biodiversity indices
Rows (sites): 51 Col (distances): 1
Sum($p(i)*p(j)*d2(i,j)$) with rescaling on [0,1]



L'indice de Rao (ou sa version ramenée sur l'intervalle 0-1) est celui qui s'éloigne le plus de la richesse. Il contient une autre information.

DMAUse : Classical Indices

Type Utilitaire de calcul d'indices de diversité classique.

Objet Les indices de diversité classiques peuvent être vus comme indicateurs de rareté⁴. C'est une idée fondamentale de Patil & Taillie. Soit un relevé écologique contenant s espèces avec les proportions (p_1, p_2, \dots, p_s) . Soit R une fonction caractéristique de la rareté de l'espèce du type :

$$R: [0, 1] \rightarrow \mathbb{R}$$

$$p \rightarrow R(p)$$

La diversité est la moyenne de la rareté des espèces du relevé. La rareté diminue quand p augmente et vaut 0 quand p vaut 1. On peut prendre :

$$R(p) = \frac{1}{p} - 1 \Rightarrow \Delta = \sum_{i=1}^s p_i \left(\frac{1}{p_i} - 1 \right) = s - 1 \quad [\text{Richesse}]$$

$$R(p) = (1 - p) \Rightarrow \Delta = \sum_{i=1}^s p_i (1 - p_i) = 1 - \sum_{i=1}^s p_i^2 \quad [\text{Simpson}]$$

$$R(p) = -\ln(p) \Rightarrow \Delta = -\sum_{i=1}^s p_i \ln(p_i) \quad [\text{Shannon}]$$

Hill⁵ unifie les indices de diversité sur une base différente. Il définit la diversité comme l'inverse d'une moyenne et note :

$$N_a = (p_1^a + p_2^a + \dots + p_s^a)^{1/(1-a)}$$

Ceci permet de voir que :

$N_{-\infty}$ est l'inverse de la proportion de l'espèce la plus rare

N_0 est le nombre d'espèces

$$N_1 = \lim_{a \rightarrow 1} (N_a) = \exp\left(-\sum_{i=1}^s p_i \ln(p_i)\right) [\text{Shannon}]$$

$$N_2 = \frac{1}{\sum_{i=1}^s p_i^2} [\text{Simpson inverse}]$$

N_{∞} est l'inverse de la proportion de l'espèce la plus abondante.

Les trois indices de base font donc partie d'une famille unique dans laquelle on soulignera la présence de :

$$N_{\frac{1}{2}} = \left(\sqrt{p_1} + \sqrt{p_2} + \dots + \sqrt{p_s} \right)^2 \quad [\text{Hill}_{1/2}]$$

Le programme a retenu le calcul de quatre indices Richesse, Simpson, Shannon et Hill_{1/2}.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Tableau Taxons (lignes) - Relevés (colonnes)

Exemple Utiliser la carte AviUrb. Transposer le tableau faunistique AUFau en FauTR :

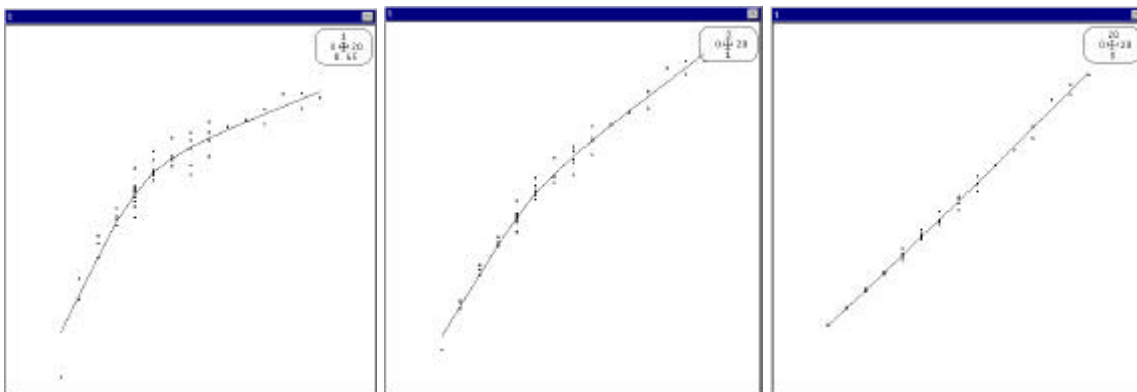
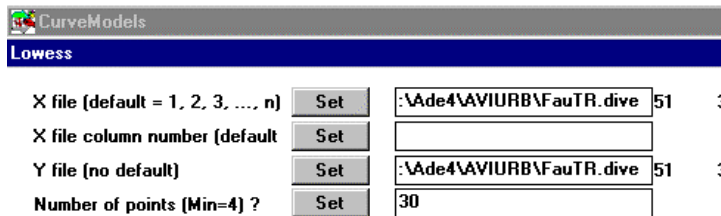


Diversity indices for species=row/Sample=column file
 Input file: E:\Ade4\AVIURB\FauTR
 Rows: 40 Col: 51

N.	Richness	Simpson	Shannon	Hill 1/2
1	10	0.867	2.243	10.286
2	11	0.875	2.288	10.935
3	11	0.908	2.441	11.755
...				
49	3	0.667	1.242	3.732
50	10	0.898	2.333	10.639
51	8	0.861	2.095	8.579

This result is stored into file E:\Ade4\AVIURB\FauTR.dive
 Rows: 51 Col: 3

Les quatre mesures sont en général très corrélées :



Simpson = f (Richesse)

Shannon = f (Richesse)

Hill = f (Richesse)

DMAUse : Corrélation

Type Utilitaire de calcul de matrices de corrélation entre matrices de distances.

Objet Soit une famille $(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_p)$ de matrices de distances calculées sur n individus statistiques. Soit une autre famille $(\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_q)$ de matrices de distances calculées sur les mêmes individus. On veut une matrice \mathbf{A} à p lignes et q colonnes dont le terme général A_{jk} donne la corrélation entre la matrice \mathbf{D}_j et la matrice \mathbf{E}_k . Il y a deux options.

La première, toujours possible, est traditionnelle et dérive du test de Mantel ⁶. Elle est décrite dans ⁷ p. 114 et consiste simplement à considérer que la demi-matrice de distances (telle qu'elle est écrite sur une colonne dans les fichiers issus de DMAUtil) est un vecteur. La corrélation entre deux matrices de distances est la corrélation ordinaire entre deux variables qu'on notera $\text{cor}(\mathbf{D}, \mathbf{E})$.

La seconde est réservée aux matrices de distances euclidiennes et est mathématiquement très différente. Si \mathbf{D} est une matrice de distances euclidienne, d'après le théorème de Gower ⁸, on considère la matrice $\Delta(\mathbf{D})$ et sa dérivée par double centrage :

$$\mathbf{D} = [d_{ij}] \Rightarrow \Delta(\mathbf{D}) = -\frac{1}{2} [d_{ij}^2] \Rightarrow \Delta_0(\mathbf{D}) = (\mathbf{I} - \mathbf{Q}) \Delta(\mathbf{D}) (\mathbf{I} - \mathbf{Q}')^t$$

\mathbf{D} est euclidienne si et seulement si $\Delta_0(\mathbf{D})$ est semi-définie positive. La diagonalisation de $\Delta_0(\mathbf{D})$ est la base de l'analyse en coordonnées principales. Si \mathbf{E} est également euclidienne, le produit scalaire d'Hilbert-Schmidt, introduit pour la première fois en analyse de données par ⁹ entre les deux opérateurs défini par :

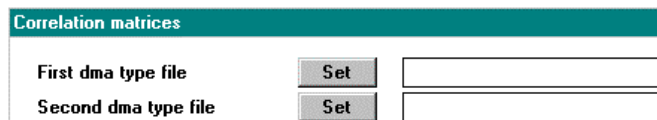
$$\langle \Delta_0(\mathbf{D}) | \Delta_0(\mathbf{E}) \rangle_{HS} = \text{trace}(\Delta_0(\mathbf{D}) \Delta_0(\mathbf{E}))$$

donne une mesure de corrélation entre les deux matrices par :

$$\text{RV}(\mathbf{D}, \mathbf{E}) = \cos(\Delta_0(\mathbf{D}), \Delta_0(\mathbf{E})) = \frac{\langle \Delta_0(\mathbf{D}) | \Delta_0(\mathbf{E}) \rangle_{HS}}{\|\Delta_0(\mathbf{D})\|_{HS} \|\Delta_0(\mathbf{E})\|_{HS}}$$

Le calcul est très simple puisqu'il reprend celui de Mantel sur les carrés des distances (en lieu des distances elles-mêmes) et *en utilisant le double centrage* de la matrice (qui implique le centrage simple sur le vecteur) pour calculer la corrélation. Le RV est toujours compris entre 0 et 1, tandis que la corrélation de Mantel est comprise entre -1 et 1.

Dialogue L'option utilise une seule fenêtre de dialogue :



- 1) Nom de fichier de type dma donnant accès à un fichier de matrices de distances.
- 2) Nom de fichier de type dma donnant accès à un fichier de matrices de distances.

Exemple Utiliser les cartes AinRho. Lire l'arbre taxonomique :



Qualitative variables file: D:\Ade4W\AINRHO\COFG
 Number of rows: 83, variables: 6, categories: 192

Description of categories:

 Variable number 1 has 1 categories

[1]Category: 1 Num: 83 Freq.: 1

Variable number 2 has 4 categories

[2]Category: 1 Num: 11 Freq.: 0.1325

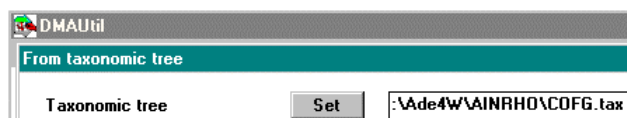
[3]Category: 2 Num: 15 Freq.: 0.1807

[4]Category: 3 Num: 4 Freq.: 0.04819

[5]Category: 4 Num: 53 Freq.: 0.6386

...

Calculer les distances associées :



Biological diversity index initialization

Input file = D:\Ade4W\AINRHO\COFG.tax

File D:\Ade4W\AINRHO\COFG_t contains distances matrices

Rows = 3403 Cols = 5

One half-matrix per column

d21, d31, d32, d41, d42, d43, ..

The matrix in the file D:\Ade4W\AINRHO\COFG_t a the property :Sum by row of squared values = 2

Text file: D:\Ade4W\AINRHO\COFG_t.dma

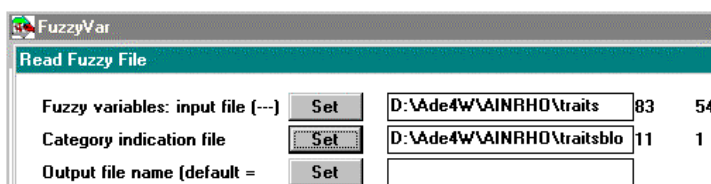
1 -> 83

2 -> 5

3 -> Taxonomic distance from file D:\Ade4W\AINRHO\COFG.tax

4 -> TRUE

Lire le tableau de traits biologiques :



Définir les distances entre espèces associées à chacun de ces traits :

```
Biological diversity index initialization
Input file = D:\Ade4W\AINRHO\traitsF.fuz
```

```
File D:\Ade4W\AINRHO\traitsF_f2 contains distance half-matrices
Rows = 3403 Cols = 11
One half-matrix per column
d21, d31, d32, d41, d42, d43, ..
Text file: D:\Ade4W\AINRHO\traitsF_f2.dma
 1 -> 83
 2 -> 11
 3 -> Fuzzy traits from file D:\Ade4W\AINRHO\traitsF.fuz option 2
 4 -> TRUE
```

Définir les distances globales entre espèces utilisant tous les traits de deux manières :

```
Distance amongst multiple frequency distributions
Input file: D:\Ade4W\AINRHO\traits
It has 83 rows and 54 columns
Bloc indicator: D:\Ade4W\AINRHO\traitsblo
Distances are computed among rows
```

```
d3 = sqrt[ Mean ( 1 - (Sum(sqrt(p(i)q(i)))) ) ]
Edwards 1971 in Hartl & Clark 1989 multi locus extension
Output file: D:\Ade4W\AINRHO\traits_Gen3
It has 3403 rows and 1 columns
d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
Text file: D:\Ade4W\AINRHO\traits_Gen3.dma
 1 -> 83
 2 -> 1
 3 -> EDWARDS on D:\Ade4W\AINRHO\traits
 4 -> TRUE
```

```
Distance amongst multiple frequency distributions
Input file: D:\Ade4W\AINRHO\traits
It has 83 rows and 54 columns
Bloc indicator: D:\Ade4W\AINRHO\traitsblo
Distances are computed among rows
```

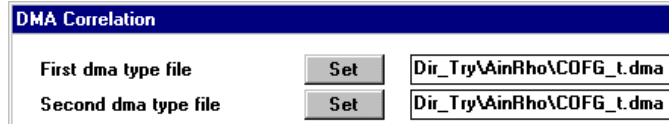
```
d2=-ln( (Sum(p(i)q(i)) / sqrt( Sum(p(i)*p(i))*Sum(q(i)*q(i)) ) ) )
Nei 1972 in Avise 1994 p. 95
Test of the euclidean property by diagonalization (theorem of GOWER)
Output file: D:\Ade4W\AINRHO\traits_Gen2
It has 3403 rows and 1 columns
```

```

d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
Text file: D:\Ade4W\AINRHO\traits_Gen2.dma
1 -> 83
2 -> 1
3 -> NEI on D:\Ade4W\AINRHO\traits
4 -> FALSE

```

Important Les distances taxonomiques ne sont pas corrélées entre niveaux :



```

First input file: D:\Ade4W\AINRHO\COFG_t.dma
Text file: D:\Ade4W\AINRHO\COFG_t.dma
1 -> 83
2 -> 5
3 -> Taxonomic distance from file D:\Ade4W\AINRHO\COFG.tax
4 -> TRUE
Second input file: D:\Ade4W\AINRHO\COFG_t.dma
Text file: D:\Ade4W\AINRHO\COFG_t.dma
1 -> 83
2 -> 5
3 -> Taxonomic distance from file D:\Ade4W\AINRHO\COFG.tax
4 -> TRUE

```

Mantel's correlation between two distance matrices

```

-----
[ 1] 1000  217 -269  -75  182
[ 2]  217 1000  114 -113 -161
[ 3] -269  114 1000 -398 -460
[ 4]  -75 -113 -398 1000 -400
[ 5]  182 -161 -460 -400 1000
-----

```

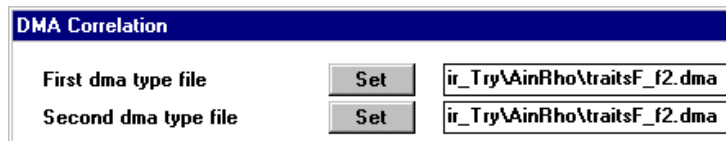
HS cosinus between two distance matrices

```

-----
[ 1] 1000   0   0   0   0
[ 2]   0 1000   0   0   0
[ 3]   0   0 1000   0   0
[ 4]   0   0   0 1000   0
[ 5]   0   0   0   0 1000
-----

```

Sur des matrices euclidiennes les deux mesures peuvent être sévèrement différentes.



```

First input file: D:\Ade4W\AINRHO\traitsF_f2.dma
Text file: D:\Ade4W\AINRHO\traitsF_f2.dma
1 -> 83
2 -> 11
3 -> Fuzzy traits from file D:\Ade4W\AINRHO\traitsF.fuz option 2
4 -> TRUE
Second input file: D:\Ade4W\AINRHO\traitsF_f2.dma
Text file: D:\Ade4W\AINRHO\traitsF_f2.dma
1 -> 83
2 -> 11
3 -> Fuzzy traits from file D:\Ade4W\AINRHO\traitsF.fuz option 2
4 -> TRUE
Mantel's correlation between two distance matrices

```

```

-----
[ 1] 1000  70  130  76  134  4  25  34  29  295  60
[ 2]  70 1000 146  84  51 255  25  72 107  75  37
[ 3] 130 146 1000 180  62  92 261 206 158 244  85
[ 4]  76  84 180 1000 169 126 238 121 222 158 101
[ 5] 134  51  62 169 1000 192 172 142 209 249 213
[ 6]  4  255  92 126 192 1000 191 224 377 103 255
[ 7]  25  25 261 238 172 191 1000 225 208 152  99
[ 8]  34  72 206 121 142 224 225 1000 331 250 399
[ 9]  29 107 158 222 209 377 208 331 1000 250 239
[10] 295  75 244 158 249 103 152 250 250 1000 324
[11]  60  37  85 101 213 255  99 399 239 324 1000
-----

```

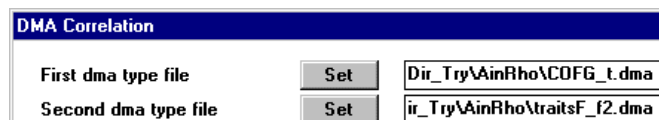
HS cosinus between two distance matrices

```

-----
[ 1] 1000  36 155  86 172  45  74  54 180 260  66
[ 2]  36 1000 119 111  32 236  16  56  91  30  26
[ 3] 155 119 1000 210 169 154 180 282 183 303 226
[ 4]  86 111 210 1000 204 230 185 200 336 171 293
[ 5] 172  32 169 204 1000 140 135 121 253 243  78
[ 6]  45 236 154 230 140 1000 247 222 338 155 228
[ 7]  74  16 180 185 135 247 1000 200 198 188 207
[ 8]  54  56 282 200 121 222 200 1000 313 246 410
[ 9] 180  91 183 336 253 338 198 313 1000 401 269
[10] 260  30 303 171 243 155 188 246 401 1000 227
[11]  66  26 226 293  78 228 207 410 269 227 1000
-----

```

Important Les domaines de variation des deux statistiques sont différents (la seconde est positive ou nulle) :



```

First input file: D:\Ade4W\AINRHO\COFG_t.dma
Text file: D:\Ade4W\AINRHO\COFG_t.dma
 1 -> 83
 2 -> 5
 3 -> Taxonomic distance from file D:\Ade4W\AINRHO\COFG.tax
 4 -> TRUE
Second input file: D:\Ade4W\AINRHO\traitsF_f2.dma
Text file: D:\Ade4W\AINRHO\traitsF_f2.dma
 1 -> 83
 2 -> 11
 3 -> Fuzzy traits from file D:\Ade4W\AINRHO\traitsF.fuz option 2
 4 -> TRUE

```

Mantel's correlation between two distance matrices

```

-----
[ 1]  -6 142  54 264 186 187 266 607 306 187 313
[ 2] 274 239 415 210 119  8 227 207  52 345 110
[ 3] 112  33 159  48 -84 -78  0  5 -120  86 -12
[ 4]  81  -6 -64 -229  37  17 -129  -9 -44  69 165
[ 5] -168  37  0 230 102 171 152  5 277 -61 -56
-----

```

HS cosinus between two distance matrices

```

-----
[ 1]  69 135 242 409 259 282 283 565 444 287 354
[ 2] 256  95 382 198 115 154 214  52 169 303 154
[ 3] 115 101  50  73  78 132  61  56  98 138 112
[ 4]  81  33  15  10  53  27  7  10  30  40  34
[ 5]  23  7  12  17  11  1  0  2  4  10  0
-----

```

DMA Correlation		
First dma type file	Set	_Try\AinRho\traits_Gen2.dma
Second dma type file	Set	Dir_Try\AinRho\COFG_t.dma

```

First input file: D:\Ade4W\AINRHO\traits_Gen2.dma
Text file: D:\Ade4W\AINRHO\traits_Gen2.dma
  1 -> 83
  2 -> 1
  3 -> NEI on D:\Ade4W\AINRHO\traits
  4 -> FALSE
Second input file: D:\Ade4W\AINRHO\COFG_t.dma
Text file: D:\Ade4W\AINRHO\COFG_t.dma
  1 -> 83
  2 -> 5
  3 -> Taxonomic distance from file D:\Ade4W\AINRHO\COFG.tax
  4 -> TRUE
Mantel's correlation between two distance matrices
-----
[ 1]  506  284  -95  -82  243
-----
Non euclidean distances. HS cosinus not available

```

DMA Correlation		
First dma type file	Set	_Try\AinRho\traits_Gen3.dma
Second dma type file	Set	Dir_Try\AinRho\COFG_t.dma

```

First input file: D:\Ade4W\AINRHO\traits_Gen3.dma
Text file: D:\Ade4W\AINRHO\traits_Gen3.dma
  1 -> 83
  2 -> 1
  3 -> EDWARDS on D:\Ade4W\AINRHO\traits
  4 -> TRUE
Second input file: D:\Ade4W\AINRHO\COFG_t.dma
Text file: D:\Ade4W\AINRHO\COFG_t.dma
  1 -> 83
  2 -> 5
  3 -> Taxonomic distance from file D:\Ade4W\AINRHO\COFG.tax
  4 -> TRUE
Mantel's correlation between two distance matrices
-----
[ 1]  455  420   76   4   74
-----
HS cosinus between two distance matrices
-----
[ 1]  603  381  175   59  16
-----

```

Important L'instabilité de ces résultats est non négligeable. On mesure ici le lien entre distance taxonomique et distance biologique. La corrélation entre distance taxonomique au niveau Classe et distance biologique (tous traits confondus) vaut-elle 0.506, 0.455 ou 0.603 ? Il faut souligner que mesurer la corrélation est la premier pas vers la prédiction d'une matrice par une autre, par plusieurs autres, vers la mesure de la corrélation partielle, ... Ainsi dans ¹⁰ plusieurs procédures de corrélation partielles sont comparées et elles sont décevantes. Ici on peut dire que la matrice *des produits scalaires* associés à la distance biologique est prédite par projection sur une base orthonormée de matrices de produits scalaires associés à des distances taxonomiques.

Lien La comparaison des deux stratégies continue dans DMAUse: Mantel-RV test.

DMAUse : DMA Weights

Type Utilitaire d'association de pondérations à une matrice de distances.

Objet Soit une matrice de distances \mathbf{D} :

$$\mathbf{D} = [d_{ij}]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

\mathbf{D} s'associe implicitement ou explicitement à des pondérations des n individus. Une pondération \mathbf{p} est un vecteur :

$$\mathbf{p} = [p_i]_{1 \leq i \leq n} \quad p_i \geq 0 \quad \sum_{i=1}^n p_i = 1$$

La PCOA de \mathbf{D} utilise une pondération. \mathbf{D} définit la \mathbf{D} -diversité des pondérations. \mathbf{D} propose une typologie des pondérations. Chaque matrice \mathbf{D} définit elle-même quatre pondérations particulières qui lui sont propres de la manière suivante.

Pondération simple : la pondération simple vaut simplement :

$$\mathbf{s}_D = [s_i]_{1 \leq i \leq n} \quad s_i = \frac{\sum_{j=1}^n d_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^n d_{ij}^2}$$

Le poids simple de la ligne i est la part de cette ligne dans la somme des carrés des distances.

Pondération canonique : la pondération canonique n'existe que si \mathbf{D} est euclidienne. Dans ce cas, on démontre que la matrice des carrés des distances :

$$\mathbf{D}^{\bullet 2} = [d_{ij}^2]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

admet une seule valeur propre positive et un seul vecteur propre associé dont toutes les composantes sont positives ou nulles. Si $\mathbf{z} = (z_1, z_2, \dots, z_n)$ est ce vecteur propre normé ($\sum_{i=1}^n z_i^2 = 1$), la pondération canonique est définie par :

$$\mathbf{c}_D = [c_i]_{1 \leq i \leq n} \quad c_i = \frac{z_i}{\sum_{i=1}^n z_i}$$

Pondération métrique : la pondération métrique n'existe que si \mathbf{D} est euclidienne. Comme dans le cas précédent, elle utilise le vecteur propre avec :

$$\mathbf{m}_D = [m_i]_{1 \leq i \leq n} \quad m_i = z_i^2$$

Pondération numérique : la pondération numérique n'existe que si \mathbf{D} est euclidienne. Elle est définie par sa propriété d'optimalité. Si la pondération numérique est notée :

$$\mathbf{n}_D = [n_i]_{1 \leq i \leq n}$$

alors pour toute autre pondération \mathbf{p} on a $\sum_{i=1}^n \sum_{j=1}^n n_i n_j d_{ij}^2 \geq \sum_{i=1}^n \sum_{j=1}^n p_i p_j d_{ij}^2$.

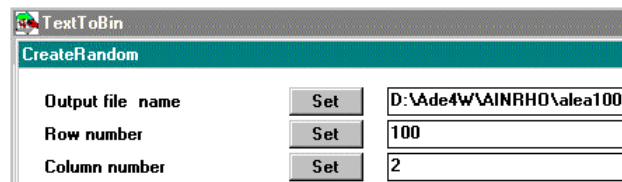
Cette pondération doit son nom au fait qu'on sait démontrer qu'elle existe mais qu'on ne peut la calculer que par une procédure numérique complexe. L'option calcule ces pondérations pour toutes les distances contenues dans un fichier issu de DMAUtil.

Dialogue L'option utilise une seule fenêtre de dialogue :



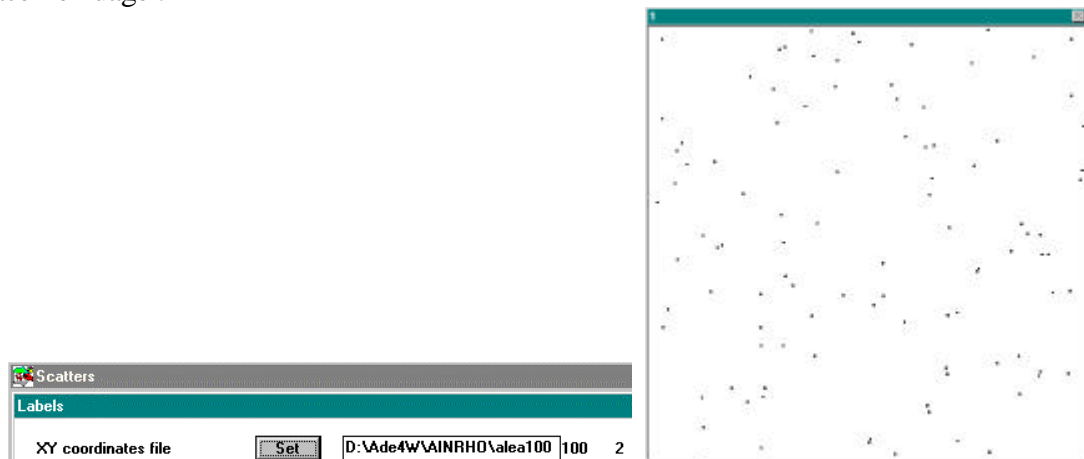
1) Nom de fichier de type dma donnant accès à un fichier de matrices de distances.

Exemple Placer 100 points dans un carrés avec :

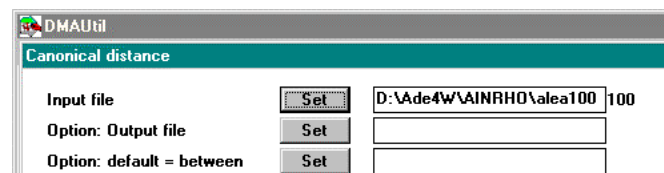


File D:\Ade4W\AINRHO\alea100 contains random numbers
Row number: 100 Column number: 2

Tracer le nuage :



Calculer les distances entre les points :



```
Distance matrix computation
-----
Input file: D:\Ade4W\AINRHO\alea100
It has 100 rows and 2 columns
Distances are computed among rows
-----
Canonical distances computed
Output file: D:\Ade4W\AINRHO\alea100_EU
It has 4950 rows and 1 columns
d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
```

```

Text file: D:\Ade4W\AINRHO\alea100_EU.dma
1 -> 100
2 -> 1
3 -> Classical metric on D:\Ade4W\AINRHO\alea100
4 -> TRUE

```



```

Input file: D:\Ade4W\AINRHO\alea100_EU.dma
Text file: D:\Ade4W\AINRHO\alea100_EU.dma
1 -> 100
2 -> 1
3 -> Classical metric on D:\Ade4W\AINRHO\alea100
4 -> TRUE

```

```

Output file: D:\Ade4W\AINRHO\alea100_EU.siwd
It contains simple weights
It as 100 rows ans 1 columns

```

```

Output file: D:\Ade4W\AINRHO\alea100_EU.cawd
It contains canonical weights
It as 100 rows ans 1 columns

```

```

Output file: D:\Ade4W\AINRHO\alea100_EU.mewd
It contains metric weights
It as 100 rows ans 1 columns
4      36      7.599e-01

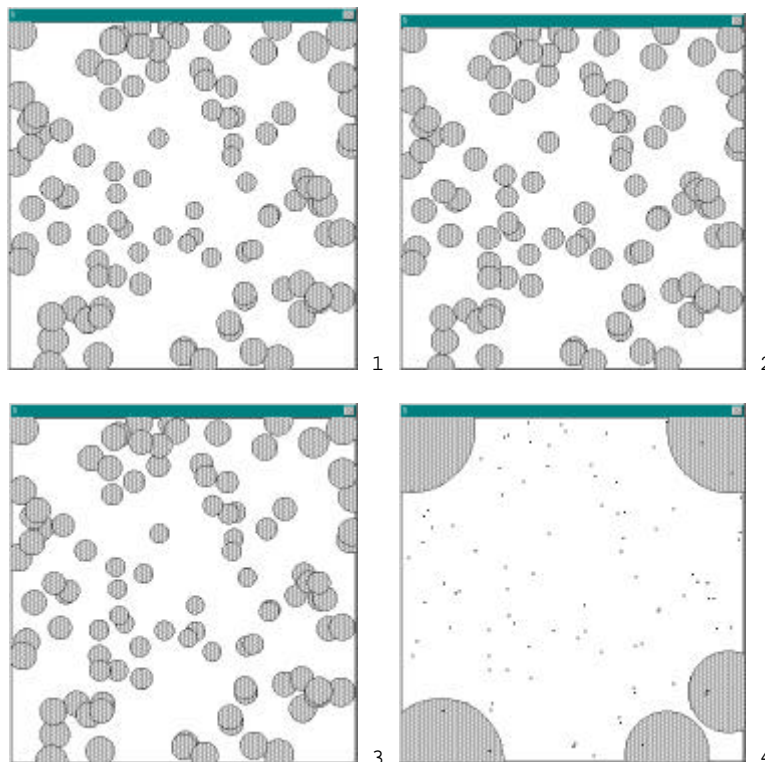
```

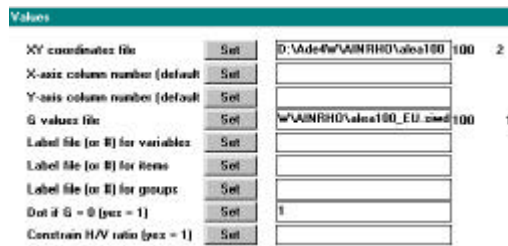
```

Output file: D:\Ade4W\AINRHO\alea100_EU.nuwd
It contains numeric weights
It as 100 rows ans 1 columns

```

Représenter le résultat (1-simple, 2-canonique, 3-métrique et 4 -numérique :

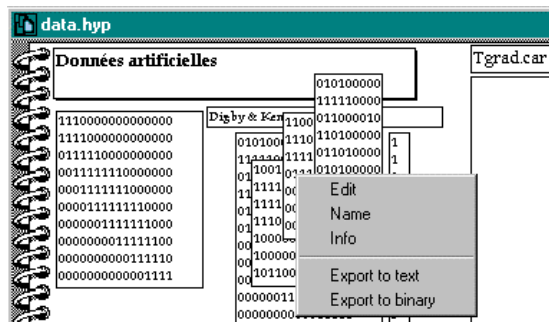




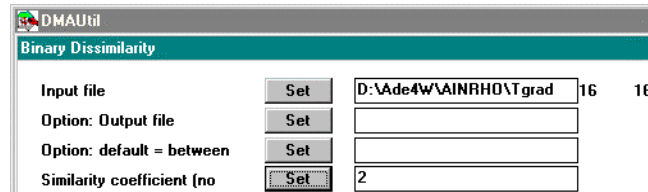
1

La pondération numérique a un statut très différent. Elle optimise la moyenne des carrés des distances (diversité spatiale).

Exemple Utiliser le tableau Tgrad¹¹ (p. 96) de la carte Artifi :



Calculer les distances entre les lignes (indice S2) :



Euclidean distance matrix computation from dissimilarity coefficients
 Gower J.C. & Legendre P. (1986)
 Metric and Euclidean properties of dissimilarity coefficients
 Journal of Classification, 3, 5-48

Table 2 p. 23
 Input file: D:\Ade4W\AINRHO\Tgrad
 It has 16 rows and 16 columns
 Distances are computed among rows

S4 coefficient of GOWER & LEGENDRE
 SOCKAL & MICHENER index (1958)
 Euclidean distance
 Distances are computed by
 $s = (a+d)/(a+b+c+d) \rightarrow d = \sqrt{1 - s}$
 Output file: D:\Ade4W\AINRHO\Tgrad_Sim2
 It has 120 rows and 1 columns
 $d(2,1), d(3,1), d(3,2), \dots, d(n,1), d(n,2), \dots, d(n,n-1)$
 Text file: D:\Ade4W\AINRHO\Tgrad_Sim2.dma
 1 -> 16
 2 -> 1
 3 -> SOCKAL & MICHENER index on D:\Ade4W\AINRHO\Tgrad
 4 -> TRUE

DMA Weights

dma type file W\AINRHO\Tgrad_Sim2.dma

Input file: D:\Ade4W\AINRHO\Tgrad_Sim2.dma
 Text file: D:\Ade4W\AINRHO\Tgrad_Sim2.dma
 1 -> 16
 2 -> 1
 3 -> SOCKAL & MICHENER index on D:\Ade4W\AINRHO\Tgrad
 4 -> TRUE

Output file: D:\Ade4W\AINRHO\Tgrad_Sim2.siwd
 It contains simple weights
 It as 16 rows ans 1 columns

Output file: D:\Ade4W\AINRHO\Tgrad_Sim2.cawd
 It contains canonical weights
 It as 16 rows ans 1 columns

Output file: D:\Ade4W\AINRHO\Tgrad_Sim2.mewd
 It contains metric weights
 It as 16 rows ans 1 columns
 13 159 6.149e-01

Output file: D:\Ade4W\AINRHO\Tgrad_Sim2.nuwd
 It contains numeric weights
 It as 16 rows ans 1 columns

Faire la PCOA de la matrice :

Principal Coordinates

dma type file W\AINRHO\Tgrad_Sim2.dma

Scatters

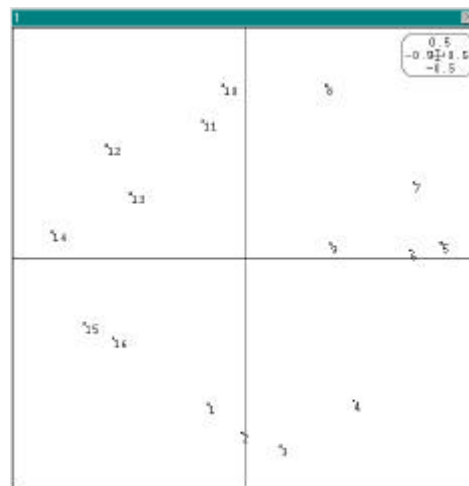
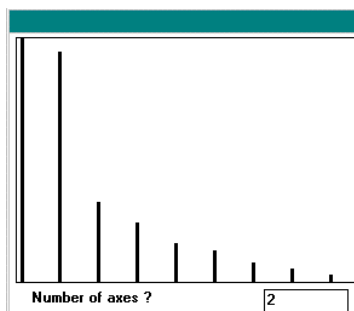
Labels

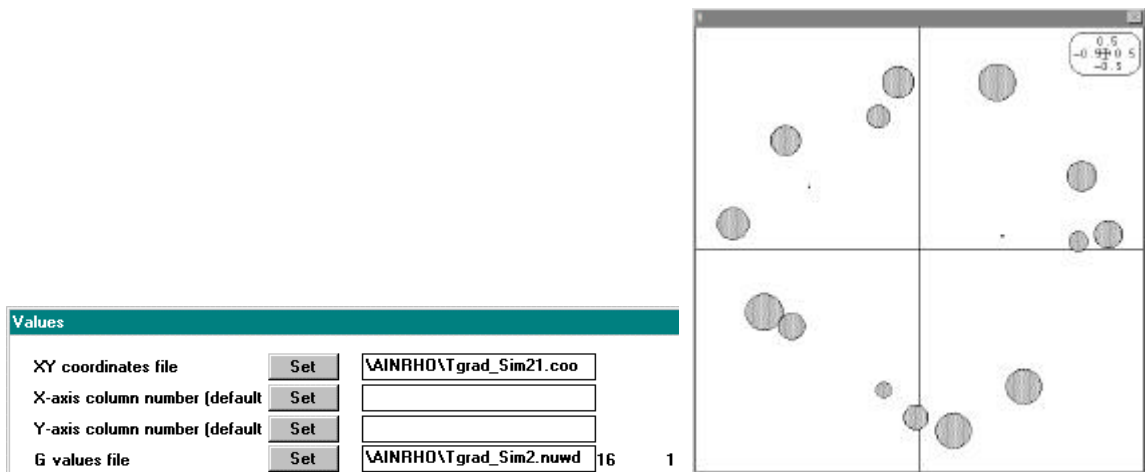
XY coordinates file W\AINRHO\Tgrad_Sim21.coo 16 2

X-axis column number (default)

Y-axis column number (default)

Label file (or #) for items #

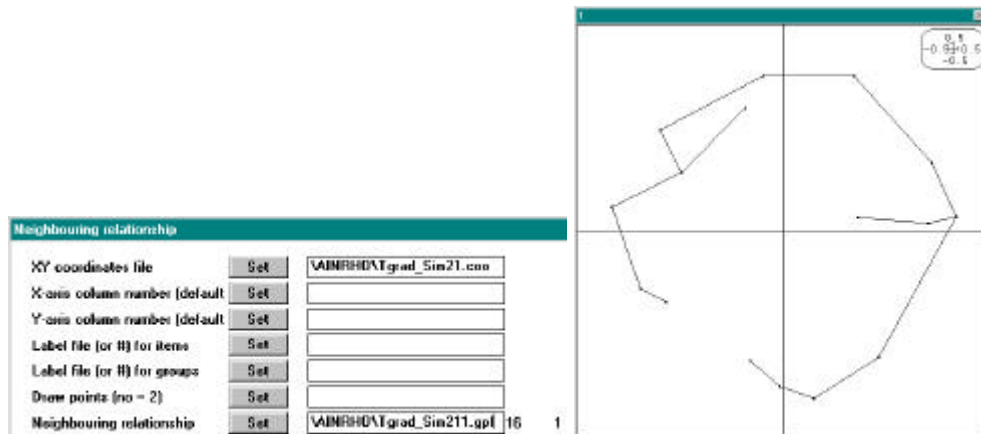




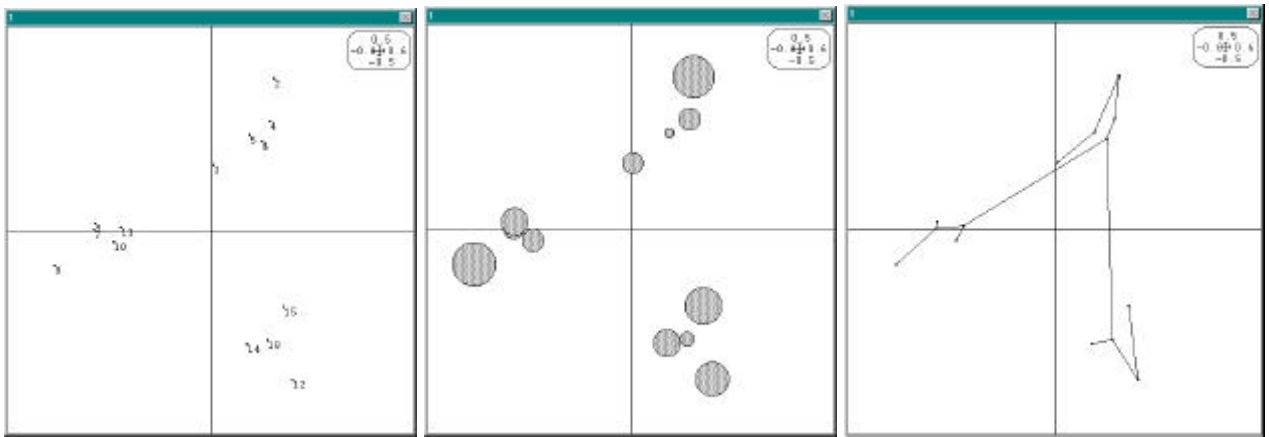
Tracer sur la carte l'arbre de longueur minimale :



Neighbouring relationship from Minimal Spanning Tree
 Input file (distance matrix): D:\Ade4W\AINRHO\Tgrad_Sim2
 Rank: 1
 Neighbouring relationship in text file: D:\Ade4W\AINRHO\Tgrad_Sim211
 It contains graph matrix (LEBART's M) with 16 rows and columns
 Neighbouring weights in binary file: D:\Ade4W\AINRHO\Tgrad_Sim211.gpl
 It contains 16 rows and 1 column



Refaire les mêmes opérations sur le tableau Tparti de la même carte (simulation d'une partition :



Ces pratiques aident à l'interprétation des structures des matrices de distances. On notera surtout l'idée suivante. Une matrice de distances euclidienne définit une représentation euclidienne : les distances entre points sont celles d'un nuage de points dans un espace euclidien. La forme du nuage est entièrement déterminé par la matrice mais la représentation en petite dimension (PCoA) doit utiliser une pondération des points. Cette pondération est libre : la plupart des programmes ne prennent que la pondération uniforme.

En introduisant des poids des lignes on accède à la notion d'inertie. Cette inertie est maximale avec la pondération numérique. Par exemple :

Principal Coordinates		
dma type file	Set	W\AINRH0\Tgrad_Sim2.dma
Column number (default=1)	Set	
Row weight (default 1/n)	Set	
1 = Complete output	Set	1

Rank : 13 Inertia 2.043e-01 Utilisation de la pondération uniforme

Principal Coordinates		
dma type file	Set	W\AINRH0\Tgrad_Sim2.dma
Column number (default=1)	Set	
Row weight (default 1/n)	Set	W\AINRH0\Tgrad_Sim2.cawd 16 1
1 = Complete output	Set	1
Option output file	Set	

Rank : 13 Inertia 2.054e-01 Utilisation de la pondération canonique

Principal Coordinates		
dma type file	Set	W\AINRH0\Tgrad_Sim2.dma
Column number (default=1)	Set	
Row weight (default 1/n)	Set	W\AINRH0\Tgrad_Sim2.surd 16 1

Rank : 13 Inertia 2.060e-01 Utilisation de la pondération simple

Principal Coordinates		
dma type file	Set	W\AINRH0\Tgrad_Sim2.dma
Column number (default=1)	Set	
Row weight (default 1/n)	Set	W\AINRH0\Tgrad_Sim2.mewd 16 1

Rank : 13 Inertia 2.064e-01 Utilisation de la pondération métrique

Principal Coordinates		
dma type file	Set	W\AINRH0\Tgrad_Sim2.dma
Column number (default=1)	Set	
Row weight (default 1/n)	Set	W\AINRH0\Tgrad_Sim2.nurd 16 1

Rank : 13 Inertia 2.173e-01 Utilisation de la pondération numérique

Ici les valeurs sont proches, car la pondération numérique est relativement bien répartie. Ceci indique que la participation des points à la définition de la structure est importante. Par contre, dans l'exemple qui précède :

```
Input file: D:\Ade4W\AINRHO\alea100_EU.dma
Distance file: D:\Ade4W\AINRHO\alea100_EU
Row: 100 Col: 1 Col used: 1
Origin: Classical metric on D:\Ade4W\AINRHO\alea100
Euclidean distance / Uniform row weights
```

Rank : 2 Inertia 1.661e-01

Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum
01 +8.3935E-02	+0.5052	+0.5052	02 +8.2197E-02	+0.4948	+1.0000

```
Input file: D:\Ade4W\AINRHO\alea100_EU.dma
Distance file: D:\Ade4W\AINRHO\alea100_EU
Row: 100 Col: 1 Col used: 1
Origin: Classical metric on D:\Ade4W\AINRHO\alea100
Euclidean distance / Row weights from file D:\Ade4W\AINRHO\alea100_EU.nuwd
```

Rank : 2 Inertia 3.934e-01

Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum
01 +2.2084E-01	+0.5614	+0.5614	02 +1.7254E-01	+0.4386	+1.0000

Important Les pondérations définissent des familles d'indice de diversité et ont été introduites pour cet objectif. Voir dans ce module BioDiv Indices.

DMAUse : Euclidean Dissimilarity Coefficient

Type Utilitaire de calcul des coefficients de dissimilarité de Rao.

Objet Soit une matrice de distances **D** euclidienne entre s espèces et un tableau faunistique **F** portant sur n sites et ces mêmes s espèces. Les distances entre espèces définissent la diversité d'un relevé :

$$H_p = \sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2$$

Si **p** et **q** sont deux relevés, la quantité :

$$H_{pq} = \sum_{i=1}^s \sum_{j=1}^s p_i q_j d_{ij}^2$$

définit ²(modifié) une distance entre deux relevés sous la forme :

$$\Delta_{pq} = \sqrt{H_{pq} - \frac{1}{2}(H_p + H_q)}$$

On démontre ⁴ que si **D**, la distance entre espèces, est euclidienne, \ddot{A} , la distance entre sites, l'est aussi. On peut ainsi mesurer la distance biologique entre deux relevés du point de vue exprimé dans la matrice de distance entre espèces.

Dialogue L'option utilise une seule fenêtre de dialogue :

- 1) Nom du fichier espèces-sites.
- 2) Nom de fichier de type dma donnant accès à un fichier de matrices de distances entre espèces.
- 3) Numéro de la colonne de la matrice de distances utilisée (1 par défaut).
- 4) Nom générique des fichiers de sortie.

Exemple Utiliser les cartes Ecomor ¹² de la pile de données. Le tableau Reghab contient la distribution de 129 espèces dans 16 types d'habitat. Faire l'AFC de ce tableau. Elle a deux valeurs propres égales à 1 et est complètement dégénérée parce que les espèces se répartissent en trois groupes (Chili, Californie, Bourgogne+Provence). La typologie des sites par les espèces ne fait qu'exprimer que les listes faunistiques sont complètement disjointes entre les trois continents.

Ce qu'on veut faire, c'est une typologie sur des critères biologiques. Le tableau Forsub contient le profil de chaque espèce pour la variable "lieu d'alimentation". Calculer les distances entre espèces pour ce critère (DMAUtil: Binary Dissimilarity) :

Binary Dissimilarity		
Input file	Set	D:\Ade4User\Dir_Try\Ecomor 129 6
Option: Output file	Set	
Option: default = between	Set	
Similarity coefficient (no	Set	7

Euclidean distance matrix computation from dissimilarity coefficients
 Gower J.C. & Legendre P. (1986)
 Metric and Euclidean properties of dissimilarity coefficients
 Journal of Classification, 3, 5-48

Table 2 p. 23
 Input file: D:\Ade4User\Dir_Try\Ecomor\forsub
 It has 129 rows and 6 columns
 Distances are computed among rows

S12 coefficient of GOWER & LEGENDRE
 OCHIAI (1957)
 Euclidean distance
 Distances are computed by
 $s = a/\sqrt{(a+b)(a+c)}$ --> $d = \sqrt{1 - s}$
 Output file: D:\Ade4User\Dir_Try\Ecomor\forsub_Sim7
 It has 8256 rows and 1 columns
 $d(2,1), d(3,1), d(3,2), \dots, d(n,1), d(n,2), \dots, d(n,n-1)$
 Text file: D:\Ade4User\Dir_Try\Ecomor\forsub_Sim7.dma
 1 -> 129
 2 -> 1
 3 -> S12 index of GOWER & LEGENDRE on D:\Ade4User\Dir_Try\Ecomor\forsub
 4 -> TRUE

Calculer alors la distance entre habitats qui en découle :

Euclidean Dissimilarity Coefficient		
Species-Samples array	Set	User\Dir_Try\Ecomor\reghab 129 16
dma type file	Set	Try\Ecomor\forsub_Sim7.dma
Column number (default=1)	Set	
Output file name	Set	A

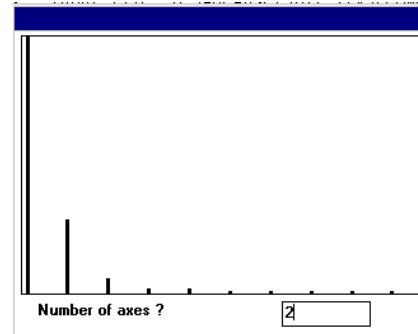
Distances - Input file: D:\Ade4User\Dir_Try\Ecomor\forsub_Sim7.dma
 Column used: 1
 Text file: D:\Ade4User\Dir_Try\Ecomor\forsub_Sim7.dma
 1 -> 129
 2 -> 1
 3 -> S12 index of GOWER & LEGENDRE on D:\Ade4User\Dir_Try\Ecomor\forsub
 4 -> TRUE

Species-Sites array - Input file: D:\Ade4User\Dir_Try\Ecomor\reghab
 Species: 129 Sites: 16
 Output file: A
 It has 120 rows and 1 columns
 $d(2,1), d(3,1), d(3,2), \dots, d(n,1), d(n,2), \dots, d(n,n-1)$
 Text file: A.dma
 1 -> 16
 2 -> 1
 3 -> edc from D:\Ade4User\Dir_Try\Ecomor\reghab and
 D:\Ade4User\Dir_Try\Ecomor\forsub_Sim7.dma
 4 -> TRUE

Nous pouvons faire l'analyse en coordonnées principales de la nouvelle matrice :

Principal Coordinates

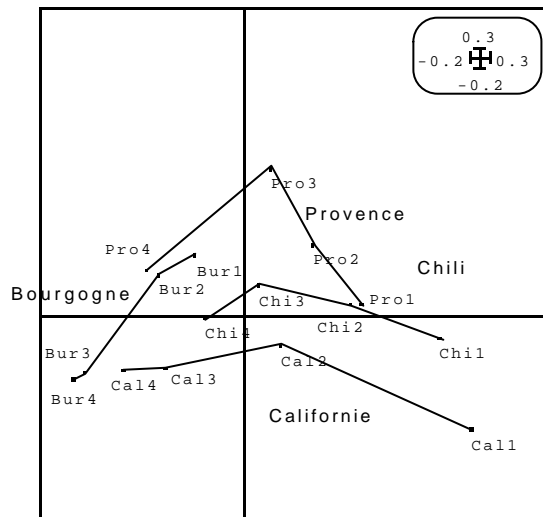
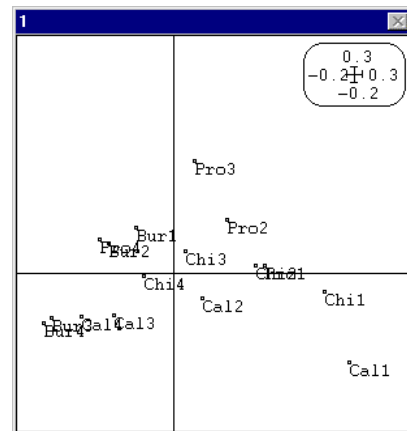
dma type file	Set	4User\Dir_Try\Ecomor\A.dma
Column number (default=1)	Set	
Row weight (default 1/n)	Set	
1 = Complete output	Set	1
Option: output file	Set	



Ceci donne une représentation de la variation entre les habitats du contenu en stratégies d'alimentation :

Labels

XY coordinates file	Set	D:\Ade4User\Dir_Try\Ecomor\16	2
X-axis column number (default	Set		
Y-axis column number (default	Set		
Label file (or #) for items	Set	D:\Ade4User\Dir_Try\Ecomor	



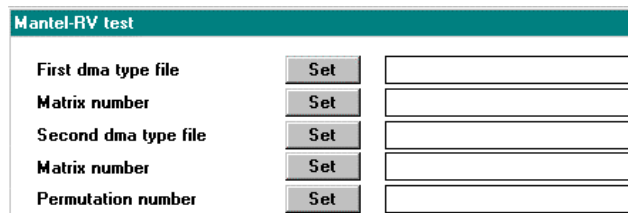
On a clairement une évolution similaire des trois gradients d'architecture de la végétation en zone méditerranéenne et l'isolement de la région témoin. La convergence recherchée par les auteurs est accessible par cette méthode.

DMAUse : Mantel-RV test

Type Test de permutation pour mesurer la corrélation entre deux matrices de distances.

Objet La statistique utilisée pour une matrice quelconque est celle de Mantel ⁶ mais la randomisation est faite à la machine ¹³. Utiliser ⁷ p. 70-75 pour une approche aisée du principe de ce test statistique. Pour une matrice euclidienne on a le même test sur la valeur du coefficient RV (voir DMAUse: Correlation matrices).

Dialogue L'option utilise une seule fenêtre de dialogue :



- 1) Nom du fichier X.dma donnant accès à la première matrice de distances.
- 2) Numéro de la matrice utilisée (colonne de X, 1 par défaut).
- 3) Nom du fichier Y.dma donnant accès à la seconde matrice de distances.
- 2) Numéro de la matrice utilisée (colonne de Y, 1 par défaut).
- 3) Nombre de permutations aléatoires (100 par défaut).

Exemple Obtenir les fichiers Mat1 (5-5) et Mat2 (5-5) sur la carte Mantel. C'est l'exemple proposé dans ⁷ p. 73. Lire ces matrices pour les mettre au format :



```

Input file: D:\Ade4\MANTEL\Mat1
D:\Ade4\MANTEL\Mat1 is a binary file with 5 rows and 5 columns
Squared matrix: Ok
Non negative value: Ok
Dii = 0 for all i: Ok
Symetric matrix: Ok
Test of the euclidean property by diagonalization (theorem of GOWER)
Output file: D:\Ade4\MANTEL\Mat1_R
It has 10 rows and 1 columns
d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
Text file: D:\Ade4\MANTEL\Mat1_R.dma
  1 -> 5
  2 -> 1
  3 -> Input distance file D:\Ade4\MANTEL\Mat1
  4 -> FALSE
-----

Input file: D:\Ade4\MANTEL\Mat2
D:\Ade4\MANTEL\Mat2 is a binary file with 5 rows and 5 columns
Squared matrix: Ok
Non negative value: Ok

```

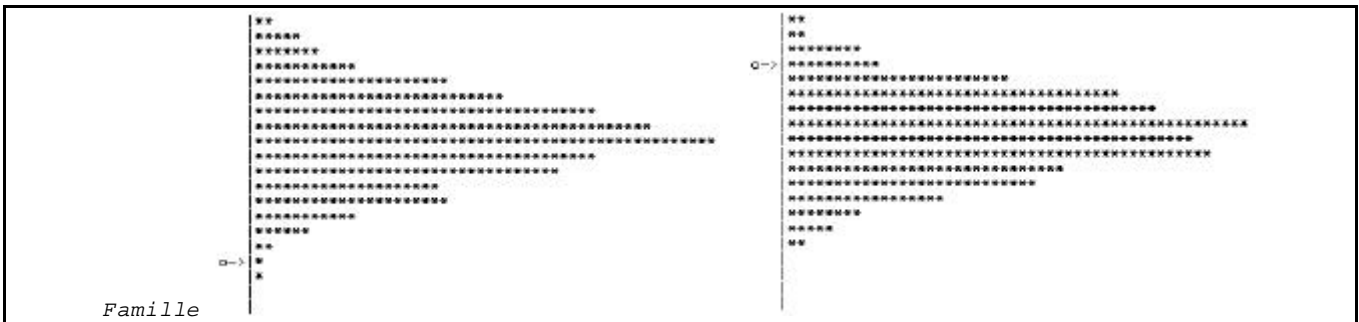
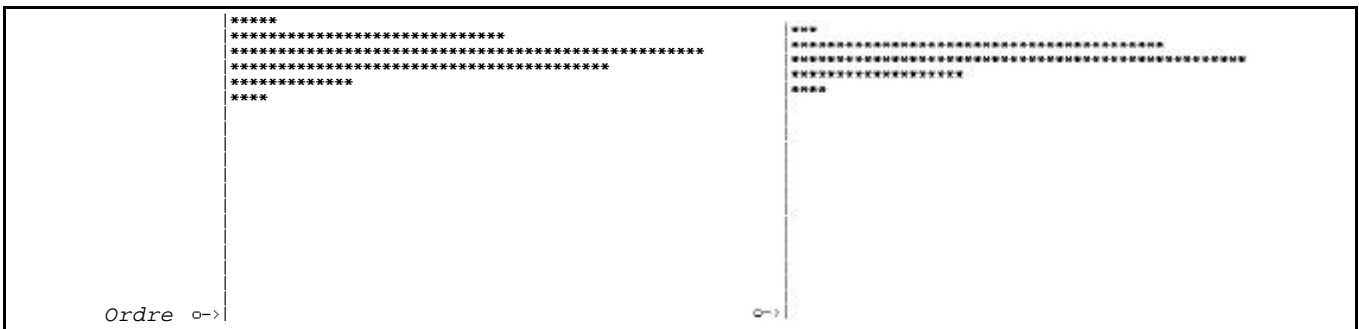
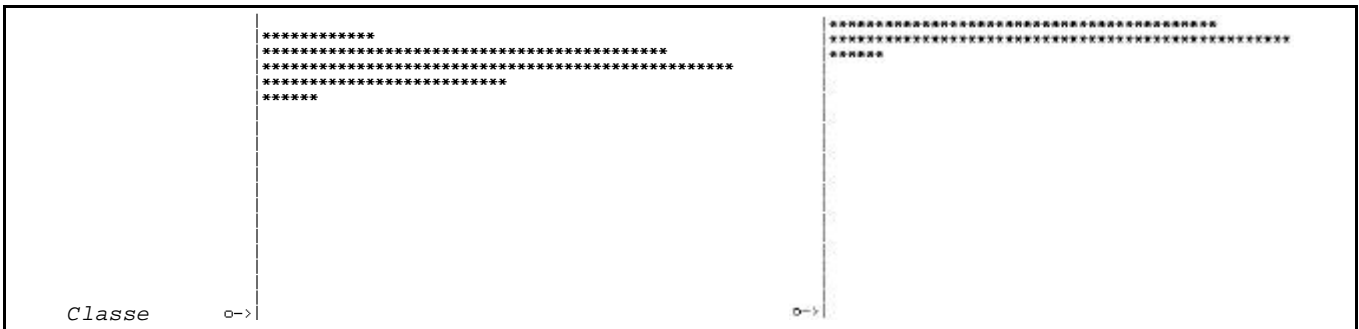

Mantel-RV test		
First dma type file	Set	_Try\AinRho\traits_Gen2.dma
Matrix number	Set	
Second dma type file	Set	D:\Ade4User\Dir_Try\AinRho
Matrix number	Set	1
Permutation number	Set	1000

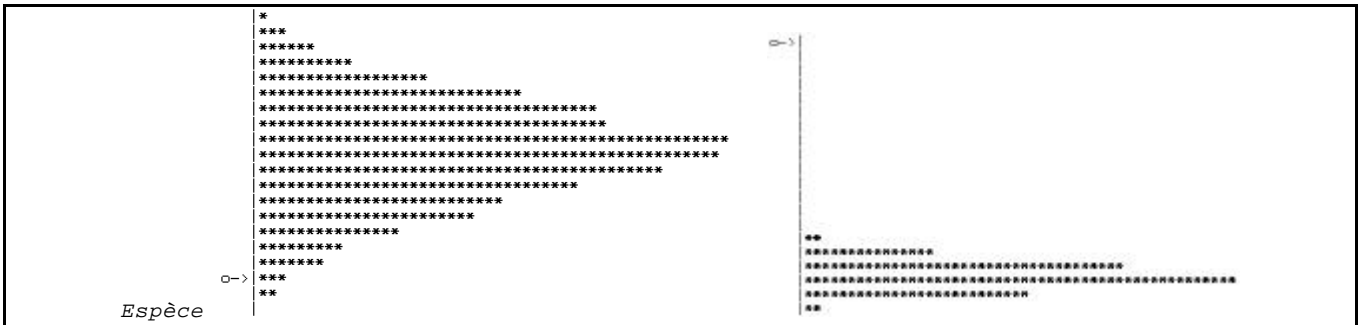
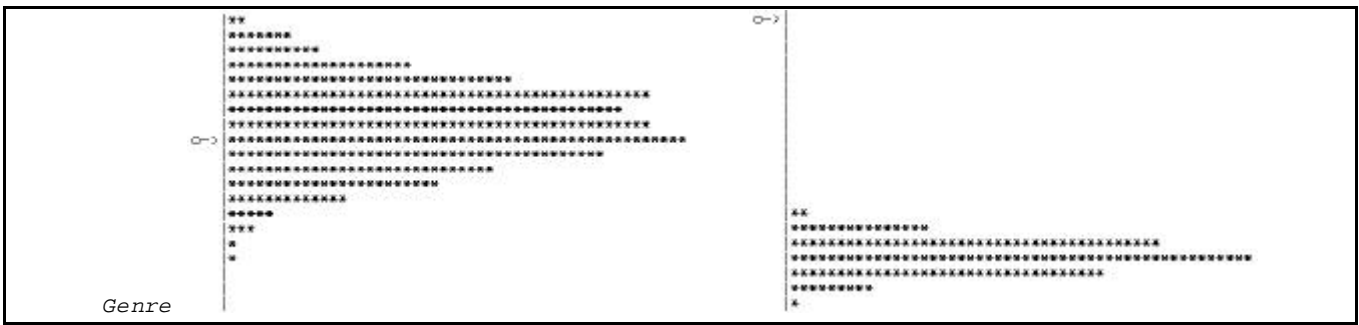
Matrix 1 of D:\Ade4W\AINRHO\traits_Gen3 versus matrix 1 of D:\Ade4W\AINRHO\COFG_tr index : 4.550e-01
 number of random matching: 1000 Observed: 0.454990
 Histogramm: minimum = -0.069046, maximum = 0.454990
 number of simulation X<Obs: 1000 (frequency: 1.000000)
 number of simulation X>=Obs: 0 (frequency: 0.000000)

ci-dessous à gauche

Matrix 1 of D:\Ade4W\AINRHO\traits_Gen3 versus matrix 1 of D:\Ade4W\AINRHO\COFG_t RV index : 6.033e-01
 number of random matching: 1000 Observed: 0.603311
 Histogramm: minimum = 0.016033, maximum = 0.603311
 number of simulation X<Obs: 1000 (frequency: 1.000000)
 number of simulation X>=Obs: 0 (frequency: 0.000000)

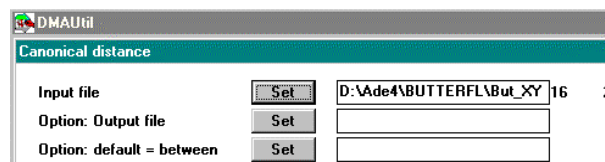
ci-dessous à droite





La contradiction entre les résultats est très forte. La cohérence est clairement supérieure avec le RV. Le test est toujours très net, la signification bascule totalement entre ordre (corrélation significativement grande) et genre (corrélation significativement petite). La validité du test de Mantel est ici fortement mise en cause.

Exemple Reprendre un exemple proposé dans ⁷. Les données ¹⁴ sont dans la carte Butterfl. Définir la distance spatiale entre les stations :



```

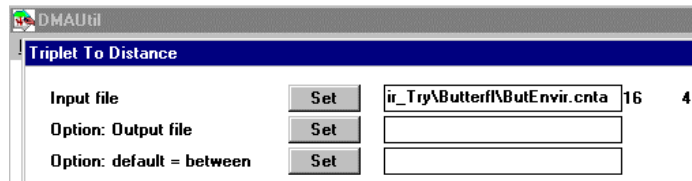
Input file: D:\Ade4\BUTTERFL\But_XY
It has 16 rows and 2 columns
Distances are computed among rows
-----
Canonical distances computed
Output file: D:\Ade4\BUTTERFL\But_XY_EU
It has 120 rows and 1 columns
d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
Text file: D:\Ade4\BUTTERFL\But_XY_EU.dma
 1 -> 16
 2 -> 1
 3 -> Classical metric on D:\Ade4\BUTTERFL\But_XY
 4 -> TRUE
-----

```

Faire l'ACP normée du tableau ButEnvir :



Calculer la matrice de distances environnementales :



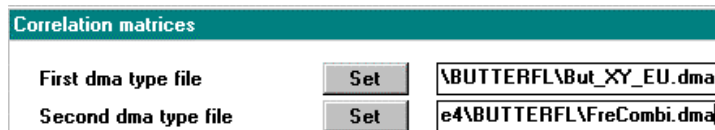
```

Input file: D:\Ade4\BUTTERFL\ButEnvir.cnta
It has 16 rows and 4 columns
Distances are computed among rows
-----
Computed distances use the diagonal metric and the centered table of the triplet
Output file: D:\Ade4\BUTTERFL\ButEnvir_MDcn
It has 120 rows and 1 columns
d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
Text file: D:\Ade4\BUTTERFL\ButEnvir_MDcn.dma
1 -> 16
2 -> 1
3 -> Euclidean distance from triplet D:\Ade4\BUTTERFL\ButEnvir.cnta
4 -> TRUE
-----

```

Calculer les distances génétiques entre populations (voir DMAUtil: Proportion data) en utilisant les cinq options.

Calculer les corrélations biologique-spatial :

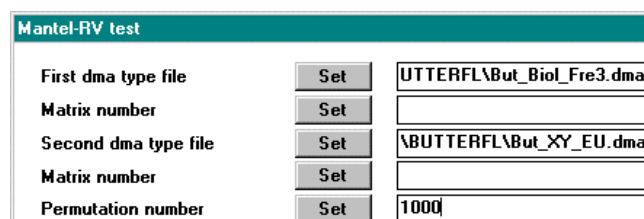


```

First input file: D:\Ade4\BUTTERFL\But_XY_EU.dma
Second input file: D:\Ade4\BUTTERFL\FreCombi.dma
Mantel's correlation between two distance matrices
-----
[ 1] 468 553 344 548 525
-----
Non euclidean distances. HS cosinus not available

```

Tester la troisième et la cinquième valeur :



```

Correlation between two distance matrices
First input file: D:\Ade4User\Dir_Try\Butterfl\But_Biol_Fre3.dma
Text file: D:\Ade4User\Dir_Try\Butterfl\But_Biol_Fre3.dma
1 -> 16
2 -> 1
3 -> ROGERS on D:\Ade4User\Dir_Try\Butterfl\But_Biol
4 -> TRUE
Matrix used : 1

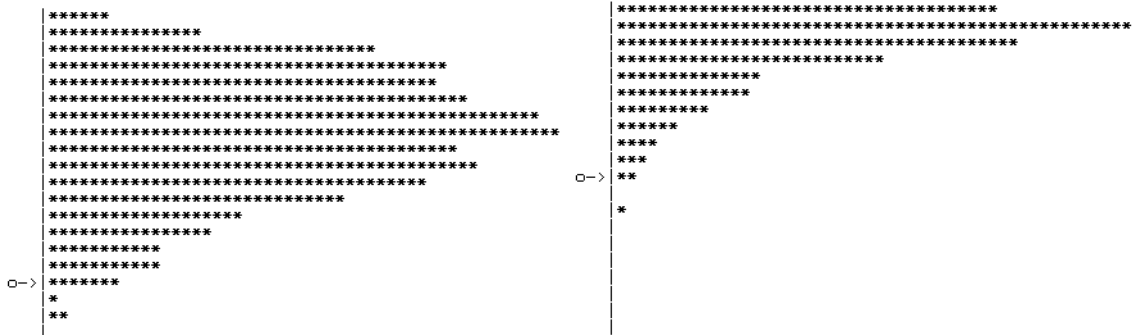
```

```

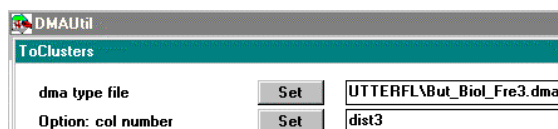
Second input file: D:\Ade4User\Dir_Try\Butterfl\But_XY_EU.dma
Text file: D:\Ade4User\Dir_Try\Butterfl\But_XY_EU.dma
  1 -> 16
  2 -> 1
  3 -> Classical metric on D:\Ade4User\Dir_Try\Butterfl\But_XY
  4 -> TRUE
Matrix used : 1
Permutation test on r value (Manly 1991 p. 114)
-----
Matrix 1 of D:\Ade4User\Dir_Try\Butterfl\But_Biol_Fre3 versus matrix 1 of
D:\Ade4User\Dir_Try\Butterfl\But_XY_EU
r index : 3.437e-01
number of random matching: 1000 Observed: 0.343670
Histogramm: minimum = -0.320404, maximum = 0.498638
number of simulation X<Obs: 978 (frequency: 0.978000)
number of simulation X>=Obs: 22 (frequency: 0.022000)
number of simulation X>=Obs: 29 (frequency: 0.029000) ci-dessous à gauche

Matrix 1 of D:\Ade4User\Dir_Try\Butterfl\But_Biol_Fre3 versus matrix 1 of
D:\Ade4User\Dir_Try\Butterfl\But_XY_EU
RV index : 2.936e-01
number of random matching: 1000 Observed: 0.293646
Histogramm: minimum = 0.003656, maximum = 0.548207
number of simulation X<Obs: 975 (frequency: 0.975000)
number of simulation X>=Obs: 25 (frequency: 0.025000) ci-dessous à droite

```

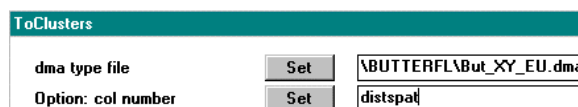


La différence est peu sensible. La distribution dissymétrique de la statistique est liée à la forme des matrices de distances. Récupérer ces matrices entières et les représenter :



Input file (distance matrix): D:\Ade4\BUTTERFL\But_Biol_Fre3.dma

Output file : D:\Ade4\BUTTERFL\But_Biol_Fre3l.dis Row: 16 Col: 16
Transformation: rescaling on [0,1] by $y=(x-\min)/(\max-\min)$



Input file (distance matrix): D:\Ade4\BUTTERFL\But_XY_EU.dma

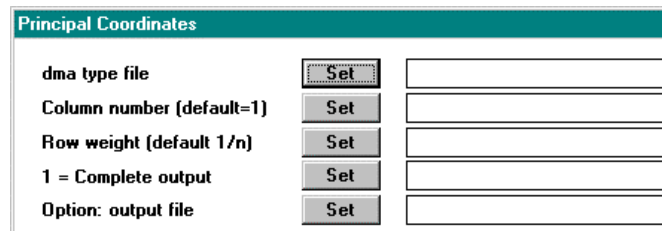
Output file : D:\Ade4\BUTTERFL\But_XY_EU1.dis Row: 16 Col: 16
Transformation: rescaling on [0,1] by $y=(x-\min)/(\max-\min)$

DMAUse : Principal Coordinates Analysis

Type Méthode d'ordination utilisant une matrice de distance ¹⁵.

Objet L'analyse permet d'obtenir des représentations euclidiennes (cartes factorielles représentant des configurations d'objets) à partir d'une matrice de distances entre ces objets. Elle est présentée dans ¹¹ (§ 3.5, p. 83 et suivantes) ou ⁷ (§ 12.3, p. 190 et suivantes). L'abréviation la plus utilisée est PCOa. La matrice de distances entre objets est transformée par $d(i, j) \mapsto -d^2(i, j)/2$, doublement centrée et diagonalisée. Les composantes des vecteurs propres sont les coordonnées des objets. L'opération est totalement valide si et seulement si les valeurs propres sont toutes positives ou nulles ¹⁶. Elle est donc réservée aux matrices de distances euclidiennes. La présentation et la discussion de Digby & Kempton est recommandée avant l'emploi de cette technique.

Dialogue L'option utilise une seule fenêtre de dialogue :



- 1) Nom de fichier contenant dma permettant l'accès à une matrice de distances ($n-n$) entre n objets.
- 2) Numéro de la matrice utilisée (1 par défaut).
- 3) Nom de fichier contenant une pondération des n objets (par défaut, pondération uniforme).
- 4) Option d'édition des résultats (taper 1 pour une sortie complète, par défaut on obtient une version simplifiée).
- 5) Option : choix d'un nom générique des fichiers de sortie (par défaut il dérive du nom du fichier d'entrée).

Exemple Utiliser la carte Zealand de la pile ADE-4•Data pour obtenir le fichier Road (13-13) contenant une matrice des distances routières entre 13 villes (⁷, tableau 11.1) et le code des villes. Lire la matrice de distances :



```
Input file: D:\Ade4W\ZEALAND\Road
D:\Ade4W\ZEALAND\Road is a binary file with 13 rows and 13 columns
Squared matrix: Ok
Non negative value: Ok
Dii = 0 for all i: Ok
Symetric matrix: Ok
Test of the euclidean property by diagonalization (theorem of GOWER)
Output file: D:\Ade4W\ZEALAND\Road_R
```

```

It has 78 rows and 1 columns
d(2,1), d(3,1), d(3,2),..., d(n,1), d(n,2), ... d(n,n-1)
Text file: D:\Ade4W\ZEALAND\Road_R.dma
1 -> 13
2 -> 1
3 -> Input distance file D:\Ade4W\ZEALAND\Road
4 -> FALSE

```

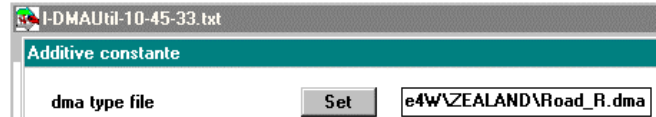


On obtient :

```

Input file: D:\Ade4W\ZEALAND\Road_R.dma
Distance file: D:\Ade4W\ZEALAND\Road_R
Row_Col: 13
Origin: Input distance file D:\Ade4W\ZEALAND\Road
Not euclidean distance found inside
PCO is not a convenient method
You can try additive constant

```

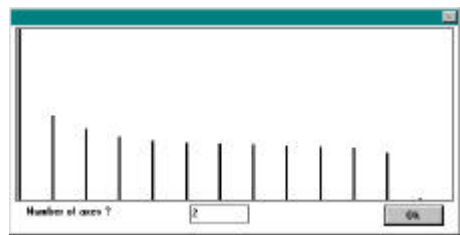
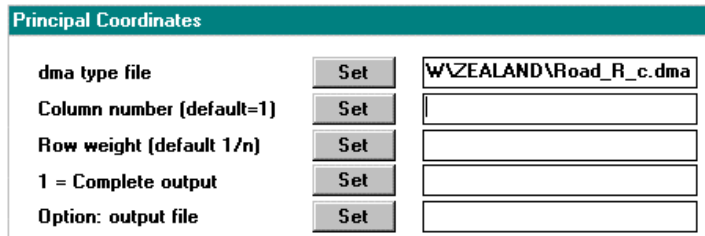


Num	Eigenval.	Num	Eigenval.	Num	Eigenval.	Num	Eigenval.
001	1.970e+03	002	5.793e+02	003	3.833e+02	004	2.320e+02
005	1.487e+02	006	1.157e+02	007	9.971e+01	008	7.467e+01
009	5.707e+01	010	4.218e+01	011	2.612e+01	012	1.000e+00
013	1.949e-02	014	-5.077e-04	015	-1.726e-03	016	-2.609e-03
017	-4.311e-03	018	-6.725e-03	019	-8.641e-03	020	-1.003e-02
021	-1.339e-02	022	-1.752e-02	023	-2.371e-02	024	-3.828e-02
025	-1.000e+00	026	-5.131e+01				

```

Input file: D:\Ade4W\ZEALAND\Road_R
additive cte: 1.970e+03
Output file: D:\Ade4W\ZEALAND\Road_R_c
It has 78 rows and 1 columns
d(2,1), d(3,1), d(3,2),..., d(n,1), d(n,2), ... d(n,n-1)
Text file: D:\Ade4W\ZEALAND\Road_R_c.dma
1 -> 13
2 -> 1
3 -> Additive constante to D:\Ade4W\ZEALAND\Road_R
4 -> TRUE

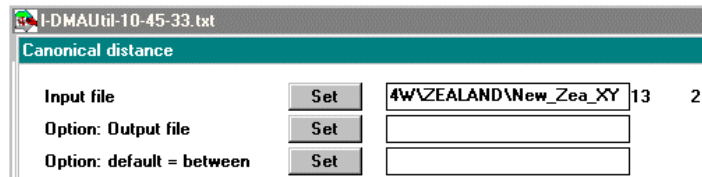
```



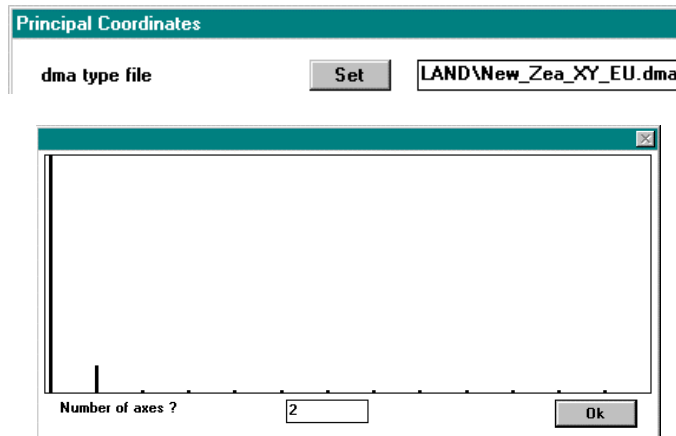
On a, entre autre :

```
File D:\Ade4W\ZEALAND\Road_R_c1.coo contains the principal coordinates
(norm=sqrt(lambda))
--- It has 13 rows and 2 columns
File :D:\Ade4W\ZEALAND\Road_R_c1.coo
|Col.| Mini | Maxi |
|----|-----|-----|
| 1 | -1.233e+03 | 8.890e+02 |
| 2 | -1.114e+03 | 7.328e+02 |
|----|-----|-----|
```

Digitaliser le fond de carte New_Zealand_Digit (dossier ADE/Files) ou utiliser le fichier New_Zea_XY et calculer les distances ordinaires entre les villes (“à vol d’oiseau”) :



Exécuter le même analyse sur la nouvelle matrice de distance :

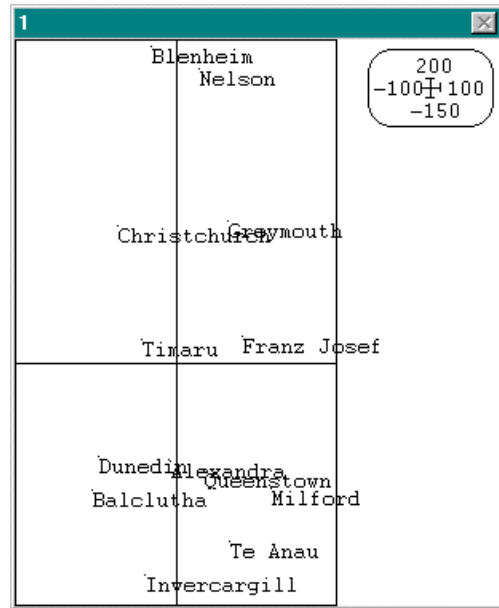


```
File D:\Ade4W\ZEALAND\New_Zea_XY_EU1.coo contains the principal coordinates
(norm=sqrt(lambda))
--- It has 13 rows and 2 columns
File :D:\Ade4W\ZEALAND\New_Zea_XY_EU1.coo
|Col.| Mini | Maxi |
|----|-----|-----|
| 1 | -1.308e+02 | 1.960e+02 |
| 2 | -5.220e+01 | 5.849e+01 |
|----|-----|-----|
```

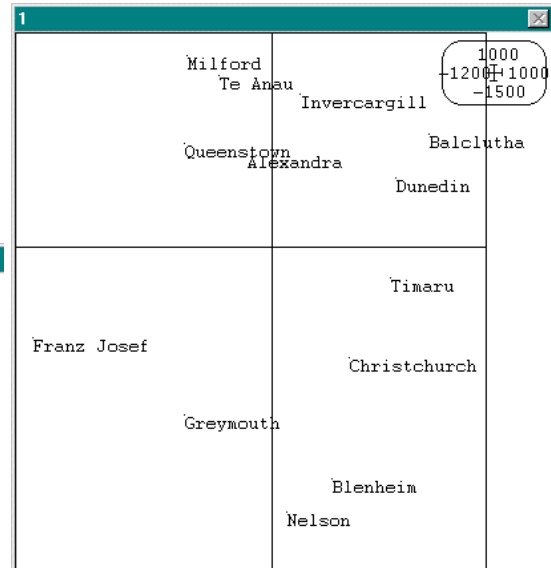
```
Rank : 2 Inertia 1.190e+04
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +1.0735E+04 +0.9020 +0.9020 | 02 +1.1664E+03 +0.0980 +1.0000 |
```

Noter qu’il n’y a que deux valeurs propres non nulles et positives (distance euclidienne dans un espace de dimension deux). Représenter les cartes des individus des deux analyses (Scatters) et comparer avec la carte initiale :

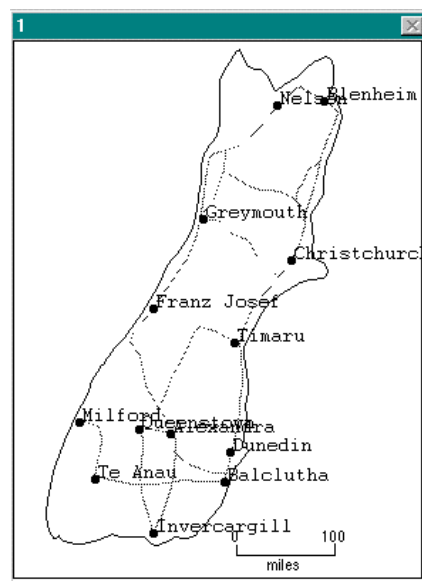
Labels		
XY coordinates file	Set	AND\New_Zea_XY_EU1.coo 13 2
X-axis column number (default	Set	2
Y-axis column number (default	Set	1
Label file (or #) for items	Set	de4w\ZEALAND\Code_Town
Label file (or #) for groups	Set	
Draw vectors from origin (yes	Set	
Draw unit circle (yes = 1)	Set	
Draw points (no = 2)	Set	
Constrain H/V ratio (yes = 1)	Set	1



Labels		
XY coordinates file	Set	\ZEALAND\Road_R_c1.coo 13 2
X-axis column number (default	Set	2
Y-axis column number (default	Set	1
Label file (or #) for items	Set	de4w\ZEALAND\Code_Town
Label file (or #) for groups	Set	
Draw vectors from origin (yes	Set	
Draw unit circle (yes = 1)	Set	
Draw points (no = 2)	Set	
Constrain H/V ratio (yes = 1)	Set	1



Maps		
Labels		
Background map (BMP file)	Set	\ZEALAND\N_Z_Carto.BMP
XY file	Set	4w\ZEALAND\New_Zea_XY 13 2
Label file (or #)	Set	de4w\ZEALAND\Code_Town



On n'oubliera pas que le signe d'un vecteur propre n'a aucune signification (si \mathbf{v} est un vecteur propre normé $-\mathbf{v}$ l'est aussi, et on trouve l'un ou l'autre au hasard). Dans le fichier .coo on a les vecteurs propres normés à la valeur propre (coordonnées de variance lambda).

Sur l'exemple qui précède, on voit que l'addition d'une constante déforme le nuage en limitant la différence qui existe entre les variances sur les deux axes. L'objectif de l'analyse n'est évidemment pas de faire des cartes routières avec des distances routières et le défaut, ici très apparent, est lié à la situation numérique. On a choisi d'imposer le caractère euclidien de la matrice pour exécuter l'analyse. On peut être d'un avis contraire (la précédente version tolérait les valeurs propres négatives). L'option choisie souligne l'importance de cette propriété.

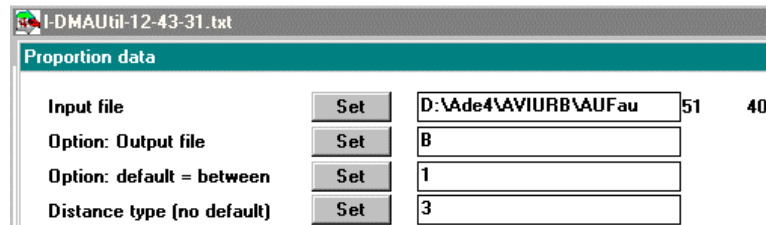
Lien L'analyse en coordonnées principales est une forme élémentaire des méthodes de positionnement multiple ("*Multidimensional scaling*"). Appliquée à une matrice de distances euclidiennes (DMAUtil : Triplet To Distance Matrix), elle donne les cartes des analyses ordinaires.

Appliquée à une matrice de distances spatiales (à deux dimensions DMAUtil : Canonical distance), elle redonne la carte à une rotation près.

Quand on dispose d'un vrai tableau de données, on peut calculer deux matrices de distances entre lignes et entre colonnes, ce qui conduit à deux typologies par ce procédé. Seule une analyse linéaire de la première couche donne les deux approches simultanées et coordonnées.

Il faut bien identifier le fait que, en partant de la matrice de distance (PCO), l'origine des cartes factorielles obtenues fait partie des résultats de la démarche ¹⁷ et non des données de départ comme dans le cas d'une ACP.

Exemple Utiliser la carte AviUrb :



On a :

```
Distance amongst frequency distributions
Input file: D:\Ade4\AVIURB\AUFau
It has 51 rows and 40 columns
Distances are computed among columns

d3=sqrt(0.5*Sum(p(i)-q(i)^2))
Rogers 1972 in Avise 1994 p. 95
Output file: B_Fre3
It has 780 rows and 1 columns
d(2,1), d(3,1), d(3,2), ..., d(n,1), d(n,2), ... d(n,n-1)
Text file: B_Fre3.dma
1 -> 40
2 -> 1
3 -> ROGERS on D:\Ade4\AVIURB\AUFau
4 -> TRUE
```

DMA Weights

dma type file E:\Ade4\AVIURB\B_Fre3.dma

Input file: E:\Ade4\AVIURB\B_Fre3.dma
 Text file: E:\Ade4\AVIURB\B_Fre3.dma
 1 -> 40
 2 -> 1
 3 -> ROGERS on D:\Ade4\AVIURB\AUFau
 4 -> TRUE

Output file: E:\Ade4\AVIURB\B_Fre3.siwd
 It contains simple weights
 It as 40 rows ans 1 columns
 ...
 Output file: E:\Ade4\AVIURB\B_Fre3.nuwd
 It contains numeric weights
 It as 40 rows ans 1 columns

On peut utiliser les poids numériques :

Principal Coordinates

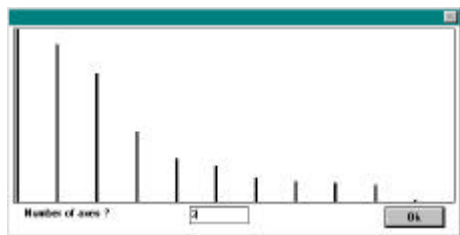
dma type file E:\Ade4\AVIURB\B_Fre3.dma

Column number (default=1) 1

Row weight (default 1/n) E:\Ade4\AVIURB\B_Fre3.nuwd 40 1

1 = Complete output 1

Option: output file



La sortie complète est assez différente de la sortie simple. Dans le cas simplifié, on se contente de conserver les coordonnées des objets sur les plans de projection du nuage euclidien, ce qui est l'objectif traditionnel. Dans le cas complet, on transforme globalement la matrice de distances en un schéma de dualité accompagné de son analyse.

Input file: E:\Ade4\AVIURB\B_Fre3.dma
 Distance file: E:\Ade4\AVIURB\B_Fre3
 Row: 40 Col: 1 Col used: 1
 Origin: ROGERS on E:\Ade4\AVIURB\AUFau
 Euclidean distance / Row weights from file E:\Ade4\AVIURB\B_Fre3.nuwd

Rank : 6 Inertia 1.923e-01							
Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+5.7151E-02	+0.2971	+0.2971	02	+5.6887E-02	+0.2958	+0.5929
03	+5.6192E-02	+0.2921	+0.8850	04	+1.9163E-02	+0.0996	+0.9847
05	+2.4916E-03	+0.0130	+0.9976	06	+4.5802E-04	+0.0024	+1.0000

File E:\Ade4\AVIURB\B_Fre31.pcta contains the principal coordinates (norm=sqrt(lambda))
 --- It has 40 rows and 6 columns
 File :E:\Ade4\AVIURB\B_Fre31.pcta

Col.	Mini	Maxi
1	-4.294e-01	2.138e-01


```

| 2|-3.201e-01| 3.311e-01|
| 3|-1.923e-01| 4.276e-01|
| 4|-2.439e-01| 2.555e-01|
| 5|-1.003e-01| 3.993e-01|
| 6|-1.386e-01| 3.374e-01|
|----|-----|-----|

```

File E:\Ade4\AVIURB\B_Fre31.pcpl contains the row weights

--- It has 40 rows and 1 columns

File :E:\Ade4\AVIURB\B_Fre31.pcpl

```

|Col.| Mini | Maxi |
|----|-----|-----|
| 1| 0.000e+00| 2.288e-01|
|----|-----|-----|

```

File E:\Ade4\AVIURB\B_Fre31.pcpcc contains the col weights

--- It has 1 rows and 1 columns

File :E:\Ade4\AVIURB\B_Fre31.pcpcc

```

|Col.| Mini | Maxi |
|----|-----|-----|
| 1| 1.000e+00| 1.000e+00|
|----|-----|-----|

```

File E:\Ade4\AVIURB\B_Fre31.pcvp contains the eigenvalues and relative inertia for each axis

--- It has 6 rows and 2 columns

File E:\Ade4\AVIURB\B_Fre31.pcli contains the principal coordinates (norm=sqrt(lambda))

--- It has 40 rows and 3 columns

File :E:\Ade4\AVIURB\B_Fre31.pcli

```

|Col.| Mini | Maxi |
|----|-----|-----|
| 1|-4.294e-01| 2.138e-01|
| 2|-3.201e-01| 3.311e-01|
| 3|-1.923e-01| 4.276e-01|
|----|-----|-----|

```

File E:\Ade4\AVIURB\B_Fre31.pcco contains the column coordinates (for complete duality diagram)

--- It has 6 rows and 3 columns

File :E:\Ade4\AVIURB\B_Fre31.pcco

```

|Col.| Mini | Maxi |
|----|-----|-----|
| 1| 0.000e+00| 5.715e-02|
| 2| 0.000e+00| 5.689e-02|
| 3| 0.000e+00| 5.619e-02|
|----|-----|-----|

```

Inner product reconstitution : quality index

Drouet d'Aubigny 1989 p. 130.

HS norm squared 1.003e-02

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+3.2662E-03	+0.3255	+0.3255	02	+3.2362E-03	+0.3225	+0.6481
03	+3.1575E-03	+0.3147	+0.9628	04	+3.6722E-04	+0.0366	+0.9994
05	+6.2082E-06	+0.0006	+1.0000	06	+2.0978E-07	+0.0000	+1.0000

Difficulté L'option diagonalise la matrice \mathbf{WD}_p (notations de Drouet (1989 p. 167) avec :

$$\mathbf{W} = -\frac{1}{2} \mathbf{Q}_1 \mathbf{D} * \mathbf{DQ}_1^t = \left[-\frac{1}{2} d_{ij}^2 \right]_{..}$$

Nous avons suivi l'auteur en introduisant une pondération arbitraire sur les objets (lignes et colonnes) de la matrice de distances. Le double centrage se fait donc avec cette pondération et

les codes numériques sont centrés, normés et non corrélés pour cette pondération. On peut hésiter entre la diagonalisation de WD_p et celle de $WD_p WD_p$ préférée par Drouet (op. cit. p. 168). Le module utilise la première, l'introduction de la constante additive permettant de se débarrasser des valeurs propres négatives si cela semble utile. La comparaison des résultats entre AFC des matrices de voisinages¹⁸ et diagonalisation de l'opérateur de Moran (NGStat : Moran EigenVectors)¹⁹ milite plutôt pour la diagonalisation de WD_p et l'utilisation des opérateurs non positifs^{20 21}.

Il y a donc deux possibilités. Soit on accepte de faire une PCOA avec des matrices non euclidiennes sans tenir compte des valeurs propres négatives, soit on impose que la PCOA ne porte que sur les matrices euclidiennes. *C'est ce second point de vue qui a été utilisé.* Pour clarifier le débat : si la matrice de distances admet une représentation euclidienne, il y a un nuage de points dont les distances sont celle de la matrice. Ce nuage de point existe et supporte une représentation en dimensions réduites par projection sur ses axes principaux. La projection existe toujours et sa valeur se juge par le graphe des valeurs propres comme pour toute analyse d'inertie. Il n'y a alors pas de contestation.

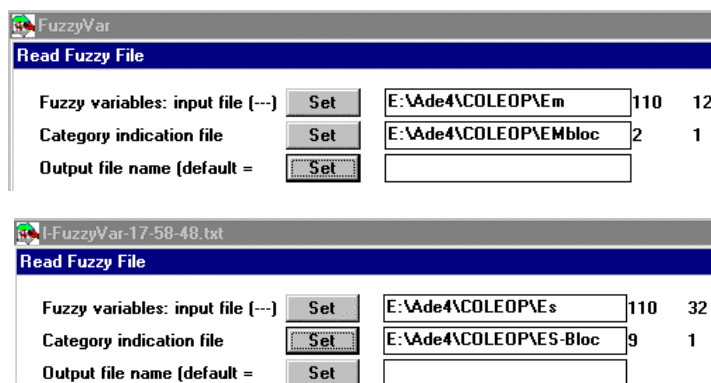
Lien

L'analyse en coordonnées principales (PCO) est un moyen simple et efficace de ramener une matrice de distance dans la logique des tableaux de données comme NGStat : Moran EigenVectors ramène un graphe de voisinage dans la logique des tableaux de données.

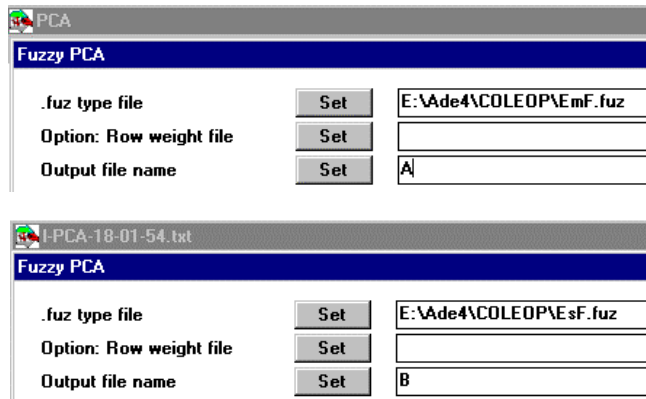
Toute analyse de tableaux définit des matrices de distances (DMAUtil : Triplet To Distance) et il y a dans ce module de nombreuses options qui définissent des matrices de distances à partir de tableaux de données.

On peut enfin passer de graphe de voisinages à matrice de distance (DMAUtil : Neighbourhood To Distance) et de matrices de distance à graphes de voisinages (Distances : Minimal Spanning Tree). Il y a donc un ensemble très vaste de possibilités d'interaction entre tableaux, graphes de voisinages et matrices de distances. Il ne faut pas confondre l'intérêt fondamental des matrices de distances dans les disciplines qui les mesurent (comme dans le contrôle des produits alimentaires par des jurys de dégustation) et le jeu qui consiste à les créer avec des tableaux de données comme c'est souvent le cas en écologie.

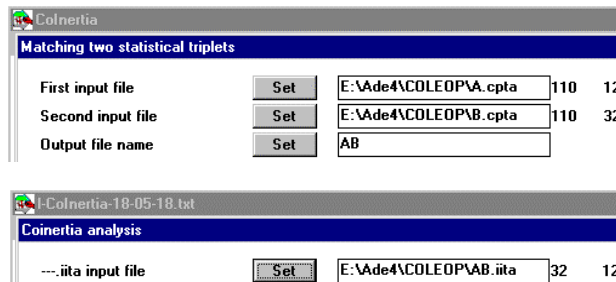
Exemple Utiliser les cartes Coleop. Lire les fichiers de variables floues :



Faire les deux ACP de ces tableaux :



Coupler et faire l'analyse de co-inertie :



Co-inertia analysis between two statistical triplets

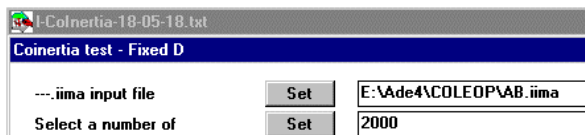
```
1 ---> E:\Ade4\COLEOP\A.cpta (rows: 110, col: 12, axes: 2, inertia: 0.076322)
2 ---> E:\Ade4\COLEOP\B.cpta (rows: 110, col: 32, axes: 2, inertia: 0.844183)
```

Co-inertia: 0.0241, RV coefficient: 0.7099

Entre stratégie spatiale (A) et stratégie écologique (B) il y a une forte liaison.

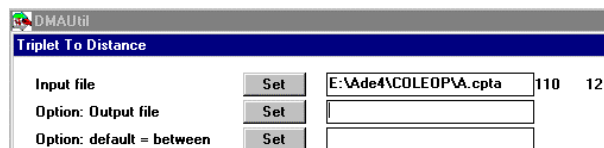
```
number of random matching: 2000   Observed: 0.024100
Histogram: minimum = 0.000074, maximum = 0.024100
number of simulation X<Obs: 2000 (frequency: 1.000000)
number of simulation X>Obs: 0 (frequency: 0.000000)
```

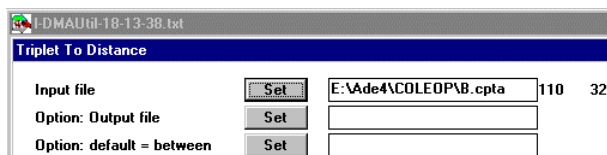
```
*****
***
```



ρ->

Définir maintenant les matrices de distances euclidiennes associées à ces deux tableaux :



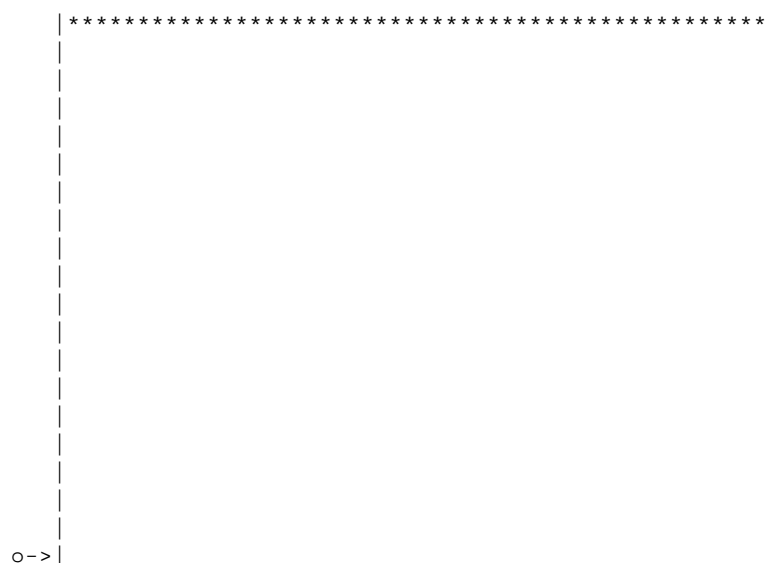


Calculer la corrélation entre les deux matrices :

```

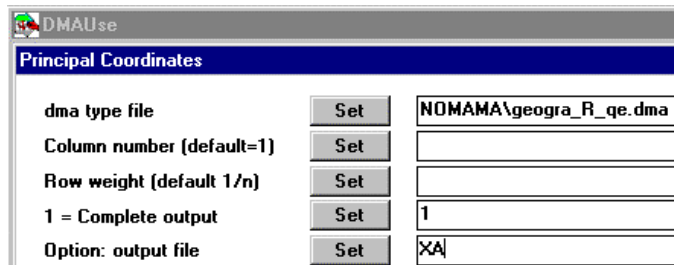
First input file: E:\Ade4\COLEOP\A_MDcp.dma
Second input file: E:\Ade4\COLEOP\B_MDcp.dma
Permutation test on r value (Manly 1991 p. 114)
-----
Matrix 1 of E:\Ade4\COLEOP\A_MDcp versus matrix 1 of E:\Ade4\COLEOP\B_MDcp
RV index : 7.099e-01
number of random matching: 2000   Observed: 0.709896
Histogramm: minimum = 0.001331, maximum = 0.709896
number of simulation X<Obs: 2000 (frequency: 1.000000)
number of simulation X>=Obs: 0 (frequency: 0.000000)

```



La corrélation (RV) entre les matrices de distances est la corrélation (RV) entre tableaux.

On peut faire l'expérience inverse. Partons de deux matrices de distances dans la carte Yanomama (publiées par ²² et reprises dans ⁷). Utiliser DMAUtil: Quasi_Euclidean pour obtenir les approximations euclidiennes des trois matrices de distances (voir la fiche de l'option citée). Faire la PCOA de deux d'entre elles :



```

Input file: E:\Ade4\YANOMAMA\geogra_R_qe.dma
Distance file: E:\Ade4\YANOMAMA\geogra_R_qe
Row: 19 Col: 1 Col used: 1
Origin: Euclidean matrix by positive eigenvalues from E:\Ade4\YANOMAMA\geogra_R
Euclidean distance / Uniform row weights

```

Rank : 2 Inertia 9.871e+03

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+6.6843E+03	+0.6772	+0.6772	02	+3.1865E+03	+0.3228	+1.0000

File XA.pcta contains the principal coordinates (norm=sqrt(lambda))
 --- It has 19 rows and 2 columns

```
File :XA.pcta
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-1.099e+02| 2.146e+02|
|  2|-9.690e+01| 7.825e+01|
|----|-----|-----|
```

File XA.pcpl contains the row weights
 --- It has 19 rows and 1 columns

```
File :XA.pcpl
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1| 5.263e-02| 5.263e-02|
|----|-----|-----|
```

File XA.pcpc contains the col weights
 --- It has 1 rows and 1 columns

```
File :XA.pcpc
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1| 1.000e+00| 1.000e+00|
|----|-----|-----|
```

File XA.pcvp contains the eigenvalues and relative inertia for each axis
 --- It has 2 rows and 2 columns

File XA.pcli contains the principal coordinates (norm=sqrt(lambda))
 --- It has 19 rows and 2 columns

```
File :XA.pcli
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-1.099e+02| 2.146e+02|
|  2|-9.690e+01| 7.825e+01|
|----|-----|-----|
```

File XA.pcco contains the column coordinates (for complete duality diagram)
 --- It has 2 rows and 2 columns

```
File :XA.pcco
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1| 0.000e+00| 6.684e+03|
|  2| 0.000e+00| 3.187e+03|
|----|-----|-----|
```

Inner product reconstitution : quality index

Drouet d'Aubigny 1989 p. 130.

HS norm squared 5.483e+07

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+4.4680E+07	+0.8148	+0.8148	02	+1.0154E+07	+0.1852	+1.0000

La sortie complète transforme la matrice de distances en une analyse complète. La matrice étant euclidienne, il existe un nuage de points dont les distances deux à deux sont celles de la matrice. Le fichier .cpta contient les coordonnées complètes de ce nuage de points sans réduction de dimension. Le fichier .cppl contient les poids des lignes introduits en option (par défaut les poids uniformes). Le fichier .cpcp contient les poids des colonnes (unitaires par définition). Dans la PCOa on retient les premières coordonnées (ici 2, la totalité). Cela donne le .cpli. Le .cpcp est obtenu par cohérence (la coordonnée de la k ème variable sur la base des composantes principales est $\sqrt{I_k}$ sur la k ème et 0 ailleurs).

Principal Coordinates		
dma type file	Set	OMAMA\anthropo_R_qe.dma
Column number (default=1)	Set	
Row weight (default 1/n)	Set	
1 = Complete output	Set	1
Option: output file	Set	XB

Input file: E:\Ade4\YANOMAMA\anthropo_R_qe.dma
Distance file: E:\Ade4\YANOMAMA\anthropo_R_qe
Row: 19 Col: 1 Col used: 1
Origin: Euclidean matrix by positive eigenvalues from E:\Ade4\YANOMAMA\anthropo_R
Euclidean distance / Uniform row weights

Rank : 15 Inertia 5.119e+04

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+2.5527E+04	+0.4987	+0.4987	02	+9.1138E+03	+0.1780	+0.6767
03	+7.1882E+03	+0.1404	+0.8171	04	+3.1012E+03	+0.0606	+0.8777
05	+2.2461E+03	+0.0439	+0.9216	06	+1.3688E+03	+0.0267	+0.9483
07	+8.5244E+02	+0.0167	+0.9650	08	+6.7333E+02	+0.0132	+0.9781
09	+5.5853E+02	+0.0109	+0.9891	10	+3.5210E+02	+0.0069	+0.9959
11	+1.3383E+02	+0.0026	+0.9986	12	+4.5160E+01	+0.0009	+0.9994
13	+1.6974E+01	+0.0003	+0.9998	14	+8.6089E+00	+0.0002	+0.9999
15	+3.4134E+00	+0.0001	+1.0000				

File XB.pcta contains the principal coordinates (norm=sqrt(lambda))
--- It has 19 rows and 15 columns

File :XB.pcta

Col.	Mini	Maxi
1	-1.182e+02	4.492e+02
2	-1.998e+02	1.858e+02
...		
13	-7.738e+00	7.001e+00
14	-5.641e+00	5.693e+00
15	-3.177e+00	3.781e+00

File XB.pcpl contains the row weights
--- It has 19 rows and 1 columns

File :XB.pcpl

Col.	Mini	Maxi
1	5.263e-02	5.263e-02

File XB.ppcp contains the col weights
--- It has 1 rows and 1 columns

File :XB.ppcp

Col.	Mini	Maxi
1	1.000e+00	1.000e+00

File XB.pcvp contains the eigenvalues and relative inertia for each axis
--- It has 15 rows and 2 columns

File XB.pcli contains the principal coordinates (norm=sqrt(lambda))
--- It has 19 rows and 3 columns

File :XB.pcli

Col.	Mini	Maxi
1	-1.182e+02	4.492e+02
2	-1.998e+02	1.858e+02
3	-1.426e+02	1.440e+02

File XB.pcco contains the column coordinates (for complete duality diagram)
 --- It has 15 rows and 3 columns

```
File :XB.pcco
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1| 0.000e+00| 2.553e+04|
|  2| 0.000e+00| 9.114e+03|
|  3| 0.000e+00| 7.188e+03|
|----|-----|-----|
```

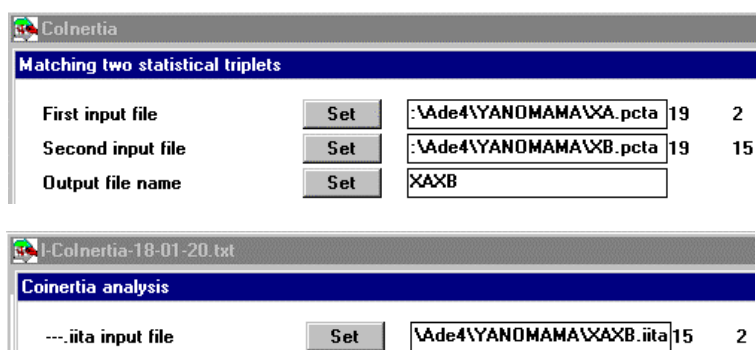
Inner product reconstitution : quality index

Drouet d'Aubigny 1989 p. 130.

HS norm squared 8.046e+08

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+6.5165E+08	+0.8100	+0.8100	02	+8.3062E+07	+0.1032	+0.9132
03	+5.1670E+07	+0.0642	+0.9774	04	+9.6176E+06	+0.0120	+0.9894
05	+5.0451E+06	+0.0063	+0.9956	06	+1.8736E+06	+0.0023	+0.9980
07	+7.2665E+05	+0.0009	+0.9989	08	+4.5338E+05	+0.0006	+0.9994
09	+3.1196E+05	+0.0004	+0.9998	10	+1.2397E+05	+0.0002	+1.0000
11	+1.7910E+04	+0.0000	+1.0000	12	+2.0394E+03	+0.0000	+1.0000
13	+2.8810E+02	+0.0000	+1.0000	14	+7.4113E+01	+0.0000	+1.0000
15	+1.1651E+01	+0.0000	+1.0000				

La seconde matrice de distances donnent de même un nuage de points, des poids et des coordonnées. Deux analyses complètes permettent une analyse de co-inertie :

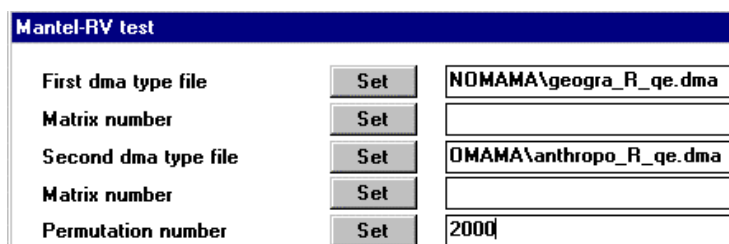


Co-inertia analysis between two statistical triplets

1 ---> E:\Ade4\YANOMAMA\XA.pcta (rows: 19, col: 2, axes: 2, inertia: 9870.849609)

2 ---> E:\Ade4\YANOMAMA\XB.pcta (rows: 19, col: 15, axes: 3, inertia: 51189.898438)

Co-inertia: 1.4307E+08, RV coefficient: 0.68117



Correlation between two distance matrices

First input file: E:\Ade4\YANOMAMA\geogra_R_qe.dma

Text file: E:\Ade4\YANOMAMA\geogra_R_qe.dma

1 -> 19

2 -> 1

3 -> Euclidean matrix by positive eigenvalues from E:\Ade4\YANOMAMA\geogra_R

4 -> TRUE

Matrix used : 1

Second input file: E:\Ade4\YANOMAMA\anthropo_R_qe.dma

Text file: E:\Ade4\YANOMAMA\anthropo_R_qe.dma

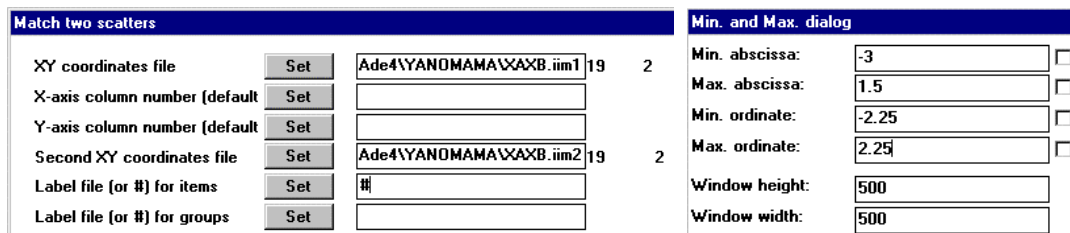
```

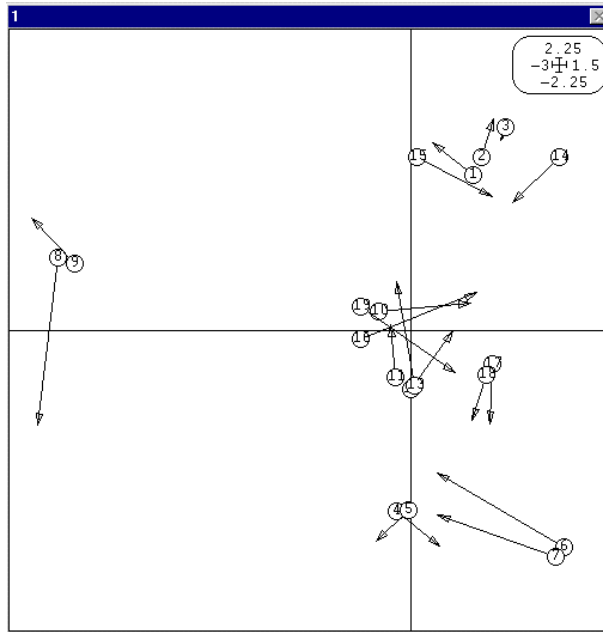
1 -> 19
2 -> 1
3 -> Euclidean matrix by positive eigenvalues from E:\Ade4\YANOMAMA\anthropo_R
4 -> TRUE
Matrix used : 1
Permutation test on r value (Manly 1991 p. 114)
-----
Matrix 1 of E:\Ade4\YANOMAMA\geogra_R_qe versus matrix 1 of
E:\Ade4\YANOMAMA\anthropo_R_qe
RV index : 6.812e-01
number of random matching: 2000 Observed: 0.681174
Histogram: minimum = 0.013127, maximum = 0.681174
number of simulation X<Obs: 2000 (frequency: 1.000000)
number of simulation X>=Obs: 0 (frequency: 0.000000)

*****
*****
*****
*****
**
o->

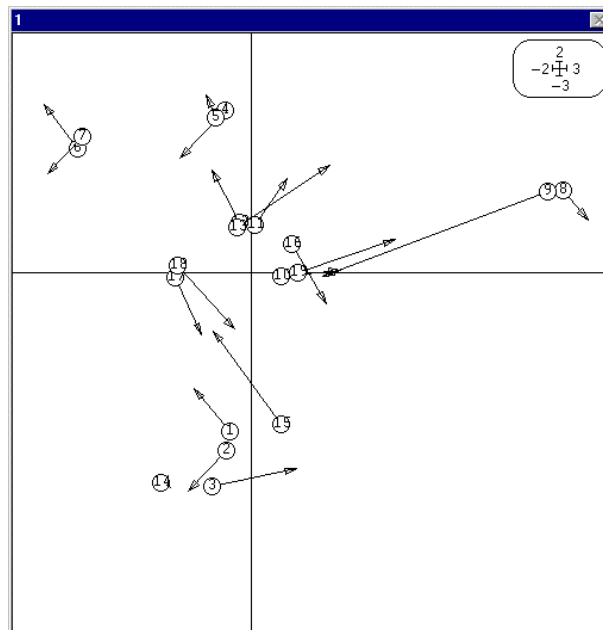
```

Le RV entre les deux matrices de distances euclidiennes est le RV de la co-inertie entre les deux représentations euclidiennes. Ceci n'est pas une observation expérimentale mais une illustration de la définition mathématique. On a alors :





On voit comment distances spatiales et distances morphométriques sont reliées : on peut définir quatre groupes qui font concorder les deux structures. Le RV sur les représentations euclidiennes est accompagné de son interprétation. En ce sens, le test de Mantel est considérablement enrichi. La même opération entre distances spatiales et distances génétiques donne une co-structure nettement moins nette qui correspond à un RV nettement plus faible (RV coefficient: 0.5127) :



L'analyse en coordonnées principales trouve donc dans le module Coinertia sa généralisation à deux tableaux. L'analyse en coordonnées principales simultanée de K matrices de distances euclidiennes est donc en perspective.

Références

- ¹ Magurran, A.E. (1988) *Ecological diversity and its measurement*. Croom Helm Limited, London. 1-179.
- ² Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* : 21, 24-43.
- ³ Rao, C.R. (1986) Rao's axiomatization of diversity measures. In : *Encyclopedia of Statistical Sciences*, Vol. 7. Kotz, S. & Johnson, N.L. (Eds.) Wiley & Sons, New York. 614-617.
- ⁴ Patil, G.P. & Taillie, C. (1979) An overview of diversity. In : *Ecological diversity in theory and practice*. Grassle, J.F., Patil, G.P., Smith, W. & Taillie, C. (Eds.) International Co-operative Publishing House, Fairland, Maryland. 3-27.
- ⁵ Hill, M.O. (1973b) Diversity and evenness: a unifying notation and its consequences. *Ecology* : 54: 427-432.
- ⁶ Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* : 27, 209-220.
- ⁷ Manly, B.F. (1994) *Multivariate Statistical Methods. A primer*. Second edition. Chapman & Hall, London. 1-215.
- ⁸ Gower, J.C. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* : 3, 5-48.
- ⁹ Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.
- ¹⁰ Oden, N.L. & Sokal, R.R. (1992) An investigation of three-matrix permutation tests. *Journal of Classification* : 9, 275-290.
- ¹¹ Digby, P. G. N. & Kempton, R. A. . (1987) *Multivariate Analysis of Ecological Communities*. Chapman and Hall, Population and Community Biology Series, London. 1-205.
- ¹² Blondel, J., Vuilleumier, F., Marcus, L.F. & Terouanne, E. (1984) Is there ecomorphological convergence among mediterranean bird communities of Chile, California, and France. In : *Evolutionary Biology*. Hecht, M.K., Wallace, B. & MacIntyre, R.J. (Eds.) Vol. 18. Plenum Press, New York. 141-213.
- ¹³ Manly, B.F.J. (1991) *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London. 1-281.
- ¹⁴ McKechnie, S.W., Ehrlich, P.R. & White, R.R. (1975) Population genetics of *Euphydryas* butterflies. I. Genetic variation and the neutrality hypothesis. *Genetics* : 81, 571-594.
- ¹⁵ Gower, J.C., 1966, Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* : 53, 325-338

-
- ¹⁶ Young, G. & Householder, A.S. (1938) Discussions of a set of points in terms of their mutual distances. *Psychometrika* : 3, 19-22.
- ¹⁷ Drouet d'Aubigny, G. (1989) *L'analyse multidimensionnelle des données de dissimilarité*. Thèse de doctorat, Université Grenoble 1. 1-485.
- ¹⁸ Lebart, L. (1984) Correspondence analysis of graph structure. *Bulletin technique du CESIA*, Paris : 2, 1-2, 5-19.
- ¹⁹ Thioulouse, J., Chessel, D. & Champely, S. (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics* : 2, 1-14.
- ²⁰ Torre, F. & Chessel, D. (1994) Co-structure de deux tableaux totalement appariés. *Revue de Statistique Appliquée* : 43, 109-121.
- ²¹ Chessel, D. & Sabatier, R. (1993) Couplage de triplets statistiques et graphes de voisinage. In : *Biométrie et Données spatio-temporelles*. Asselain, B. & Coll. (Ed.) Société Française de Biométrie, ENSA, Rennes. 28-37.
- ²² Spielman, R.S. (1973) Differences among Yanomama Indian villages: do the patterns of allele frequencies, anthropometrics and map locations correspond?. *American Journal of Physical Anthropology* : 39, 461-480.