



DMAUtil

Introduction.....	1
DMAUtil : Additive constante	3
DMAUtil : Binary Dissimilarity.....	5
DMAUtil : Canonical distance.....	8
DMAUtil : Edit dma file	11
DMAUtil : From fuzzy traits.....	13
DMAUtil : From taxonomic tree	16
DMAUtil : Genetic distance.....	20
DMAUtil : Lingoes correction.....	23
DMAUtil : Neighbourhood To Distance	27
DMAUtil : Proportion data	29
DMAUtil : Quantitative variables.....	33
DMAUtil : Quasi_Euclidean.....	37
DMAUtil : Read distance file.....	42
DMAUtil : Read half distances matrix.....	44
DMAUtil : ToClusters	46
DMAUtil : Triplet To Distance	49
Références	51

Introduction

En mathématiques, on appelle distance définie sur un ensemble E une fonction d de $E \times E$ dans \mathbf{R} qui vérifie pour tout x, y et z éléments de E :

- (1) $d(x, y) \geq 0$
- (2) $d(x, y) = 0 \Leftrightarrow x = y$
- (3) $d(x, y) = d(y, x)$
- (4) $d(x, y) \leq d(x, z) + d(z, y)$

En statistiques, on appelle **dissimilarité** définie sur un ensemble fini I à n éléments (numérotés $1, 2, \dots, i, \dots, n$) une fonction de $I \times I$ dans \mathbf{R} qui vérifie pour tout i et j :

- (1) $\partial_{ij} \geq 0$
- (2) $\partial_{ii} = 0$
- (3) $\partial_{ij} = \partial_{ji}$

En biologie, on utilise le terme de distance pour désigner la différence mesurée entre deux individus, deux populations, deux sites, ..., sans se préoccuper de définition. Par exemple la distance génétique de Autem & Bonhomme (1980)¹ utilisée dans Agnese (1989)² n'est pas nulle pour deux populations ayant les mêmes profils de fréquences alléliques. Pour suivre la coutume on appellera **matrice de distances** une matrice contenant une **dissimilarité observée**. Les matrices de distances sont donc des matrices carrées (n lignes et n colonnes), contenant des nombres positifs (1), symétriques (3), ayant des éléments nuls sur la diagonale (2). Le module crée des matrices de distances (DMA pour *Distance Matrix*).

On dit qu'une dissimilarité est euclidienne si elle admet une image euclidienne, c'est-à-dire si on peut associer au point i ($1 \leq i \leq n$) un point P_i d'un espace euclidien tel que :

$$\|P_i P_j\|_M = \partial_{ij}$$

Si c'est le cas on dit que la matrice de distance est euclidienne. L'intérêt de cette propriété est de pouvoir reconstruire un nuage de n points dans \mathbf{R}^p (p est la dimension de la dissimilarité) dont les distances des points deux à deux sont exactement les dissimilarités observées et éventuellement de représenter ce nuage par projection (Analyse en Coordonnées Principales, PCO).

Si la matrice de distance est calculée à partir d'un tableau de données, on sait souvent si elle est euclidienne ou si elle ne l'est pas. Gower & Legendre (1986)³ ont étudié un grand nombre de cas utilisés dans ce module. Si elle dérive directement de l'observation, on testera systématiquement si elle est ou non euclidienne. On peut toujours générer une matrice de distance euclidienne à partir d'une matrice de distances qui ne l'est pas en ajoutant une constante positive c à chacune des distances ∂_{ij} . Si la matrice de distances n'est pas euclidienne, l'option DMAUtil: Additive constante calcule cette constante, la plus petite possible, avec la solution donnée par Cailliez (1983)⁴. L'option DMAUtil: Lingoes correction propose une autre manière d'obtenir une distance euclidienne à partir d'une distance qui ne l'est pas. Cette transformation a un grand intérêt pratique.

Toute matrice de distance est dans un fichier X accompagné d'un fichier X.dma. Comme ce type de matrice est particulier (symétrie et diagonale nulle), les fichiers ne contiennent que la moitié inférieure de la matrice, soit sur une colonne :

$$\partial_{21}, \partial_{31}, \partial_{32}, \partial_{41}, \partial_{42}, \partial_{43}, \dots, \partial_{n1}, \partial_{n2}, \dots, \partial_{n(n-1)}.$$

Ceci permet de stocker dans un fichier plusieurs matrices de distances colonne par colonne. X.dma contient le nombre n de lignes (et de colonnes) des matrices de distances, le nombre de matrices de distances conservées, une chaîne de caractères donnant l'origine de cette matrice ou ces matrices et la mention TRUE ou FALSE suivant que la matrice est une matrice de distances euclidienne ou non. Quand il y en a plusieurs TRUE est utilisée si elles sont toutes euclidiennes.

On accède systématiquement à une matrice de distances par son fichier .dma (comme on accède systématiquement à un fichier de variables qualitatives par son fichier .cat, ou à un graphe de voisinage par son fichier .gpl, ...). L'option DMAUtil: Read distance file permet d'importer dans ADE-4 une matrice quelconque. Le test du caractère euclidien utilise la diagonalisation de la matrice des valeurs $-\frac{1}{2}\partial_{ij}^2$ doublement centrée (Théorème de Gower, voir la thèse de Drouet d'Aubigny⁵).

La simple lecture du listing indique dans, toutes les options, si on travaille sur une matrice de distances euclidienne. Les matrices de distances sont utilisées dans le module DMAUse dans lequel la PCO est réservée aux matrices de distances euclidiennes. Le stockage de plusieurs matrices de distances dans un même fichier est essentiellement dû à l'objectif de mesurer la biodiversité fonctionnelle (plusieurs variables biologiques définissent plusieurs distances entre espèces) et la biodiversité taxonomique (plusieurs niveaux taxonomiques induisent des distances entre espèces à chaque niveau).

On notera donc que les matrices de distances sont systématiquement stockées sur une colonne, un fichier pouvant contenir, sur une ou plusieurs colonnes, une ou plusieurs matrices de distances. L'option DMAUtil: ToClusters envoie une matrice de distances au module Clusters pour les pratiques de classification et la représentation des dendrogrammes.

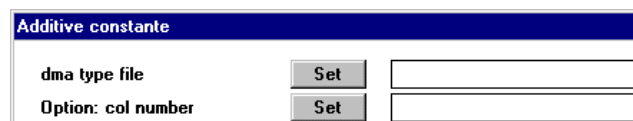
Le présent module remplace et étend l'ancien module Distances. On retrouvera l'option de calcul des arbres de longueur minimale dans NGStat: Minimal Spanning Tree.

DMAUtil : Additive constante

Type Méthode heuristique rendant euclidienne une matrice de distances.

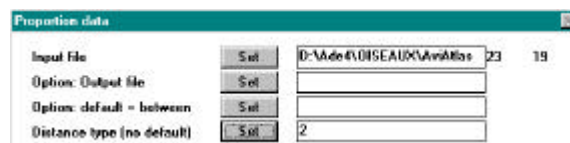
Objet Gower^{6,7} a montré l'intérêt des matrices de distances euclidiennes. On peut toujours générer une matrice de distance euclidienne à partir d'une matrice de distances qui ne l'est pas en ajoutant une constante positive c à chacun des termes hors diagonale. Si la matrice de distances n'est pas euclidienne, l'option calcule cette constante, la plus petite possible, avec la solution donnée par Cailliez (1983)⁴.

Dialogue L'option utilise une seule fenêtre de dialogue :



- 1) Nom du fichier d'entrée (.dma) associé à une matrice de distance.
- 2) Numéro de la colonne utilisée (1 par défaut).

Exemple Utiliser la carte Oiseaux de la pile de données. Calculer les distances entre espèces avec l'indice de chevauchement (DMAUtil: Proportion data) :



```
Distance amongst frequency distributions
Input file: D:\Ade4\OISEAUX\AviAtlas
It has 23 rows and 19 columns
Distances are computed among rows
```

```
d2 distances computed
Manly 1994 Multivariate statistical methods. A primer
2nd edition. Chapman & Hall 1994. formula 5.8 p. 68
Test of the euclidean property by diagonalization (theorem of GOWER)
Output file: D:\Ade4\OISEAUX\AviAtlas_Fre2
It has 253 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\OISEAUX\AviAtlas_Fre2.dma
1 -> 23
2 -> 1
3 -> d2 of MANLY on D:\Ade4\OISEAUX\AviAtlas
4 -> FALSE
```

La matrice de distance n'est pas euclidienne :



Une matrice 38×38 ($2n \times 2n$) est diagonalisée : sa première valeur propre est la constante recherchée⁽⁴⁾ :

```

Num  Eigenval. | Num.  Eigenval. | Num.  Eigenval. | Num.  Eigenval. |
001  1.854e+00 | 002  1.330e+00 | 003  1.205e+00 | 004  1.131e+00 |
005  1.099e+00 | 006  1.089e+00 | 007  1.072e+00 | 008  1.058e+00 |
009  1.043e+00 | 010  1.033e+00 | 011  1.027e+00 | 012  1.026e+00 |
013  1.019e+00 | 014  1.011e+00 | 015  1.005e+00 | 016  1.003e+00 |
017  1.001e+00 | 018  1.001e+00 | 019  1.000e+00 | 020  1.000e+00 |
021  1.000e+00 | 022  1.000e+00 | 023  1.000e+00 | 024  -5.393e-01 |
025  -7.517e-01 | 026  -8.296e-01 | 027  -8.840e-01 | 028  -9.103e-01 |
029  -9.184e-01 | 030  -9.326e-01 | 031  -9.453e-01 | 032  -9.584e-01 |
033  -9.676e-01 | 034  -9.735e-01 | 035  -9.744e-01 | 036  -9.811e-01 |
037  -9.891e-01 | 038  -9.949e-01 | 039  -9.971e-01 | 040  -9.989e-01 |
041  -9.994e-01 | 042  -9.996e-01 | 043  -1.000e+00 | 044  -1.000e+00 |
045  -1.000e+00 | 046  -1.000e+00 |

```

```

Input file: D:\Ade4\OISEAUX\AviAtlas_Fre2
additive cte: 1.854e+00
Output file: D:\Ade4\OISEAUX\AviAtlas_Fre2_c
It has 253 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\OISEAUX\AviAtlas_Fre2_c.dma
  1 -> 23
  2 -> 1
  3 -> Additive constante to D:\Ade4\OISEAUX\AviAtlas_Fre2
  4 -> TRUE
-----

```

La nouvelle matrice de distances est euclidienne :

Additive constante	
dma type file	Set ISEAUX\AviAtlas_Fre2_c.dma
Option: col number	Set

```

Input file: D:\Ade4\OISEAUX\AviAtlas_Fre2_c.dma
Distance file: D:\Ade4\OISEAUX\AviAtlas_Fre2_c
Row_Col: 23
Origin: Additive constante to D:\Ade4\OISEAUX\AviAtlas_Fre2
Euclidean distance found inside
Additive constante is not useful

```

DMAUtil : Binary Dissimilarity

Type Utilitaire de calcul de matrices de distance.

Objet L'option calcule des matrices de distance à partir d'indices de dissimilarités sur données binaires. On les utilise pour les tableaux floro-faunistiques en présence-absence.

Dialogue L'option utilise une seule fenêtre de dialogue :

- 1) Nom du fichier binaire d'entrée.
- 2) Nom générique des fichiers de sortie (création). Par défaut, il est généré avec le nom du fichier d'entrée et des chaînes de caractères qui indiquent les options choisies.
- 3) Option de calcul. Utiliser 1 pour calculer les distances entre colonnes du tableau. Par défaut, les distances sont calculées entre lignes.
- 4) Option de choix de l'indice de dissimilarité. Il y a 10 options :

Similarity coefficient (no default)

a = n(++), b = n(+), c = n(-), d = n(--)

1 = a/(a+b+c) 2 = (a+d)/(a+b+c+d)

3 = a/(a+2(b+c)) 4 = (a+d)/(a+2(b+c)+d)

5 = a/(a+.5(b+c)) 6 = (a-(b+c)+d)/(a+b+c+d)

7 = a/sqrt((a+b)(a+c))

8 = ad/sqrt((a+b)(a+c)(d+b)(d+c))

9 = ad-bc/sqrt((a+b)(a+c)(b+d)(d+c))

10 = a/(a+b+c+d) and unit self-similarity

Deux objets (en écologie, lignes ou colonnes d'un tableau floro-faunistiques) sont comparés sur une liste de valeurs. Ces valeurs sont réduites en 0-1 (1 si la valeur est strictement positive, 0 sinon). Deux relevés sont ainsi comparés par la liste des espèces présentes, deux espèces sont comparées par la liste des relevés dans lesquels elles sont présentes. Ces listes ont la forme :

01100001010010...
01010001100010...

n est le nombre d'enregistrements, a est le nombre de concordances 11, b le nombre de concordances 10, c le nombre de concordances 01 et d le nombre de concordances 00. Ainsi deux espèces sont présentes ensemble dans un même relevé a fois, deux relevés possèdent a espèces en commun. Les deux objets définissent donc la table de contingence 2-2 :

	1	0	Tot
1	a	b	$a+b$
0	c	d	$c+d$
Tot	$a+c$	$b+d$	n

Les quatre nombres de la table définissent une similarité entre les deux objets. On peut utiliser :

$S_1 = \frac{a}{a+b+c}$	Indice de communauté de Jaccard
$S_2 = \frac{a+d}{n}$	Indice de Sokal & Michener
$S_3 = \frac{a}{a+2(b+c)}$	Indice de Sokal & Sneath
$S_4 = \frac{a+d}{a+2(b+c)+d}$	Indice de Rogers et Tanimoto
$S_5 = \frac{2a}{2a+b+c}$	Indice de Sorensen
$S_6 = \frac{a-(b+c)+d}{n}$	Indice de Gower & Legendre
$S_7 = \frac{a}{\sqrt{(a+b)(a+c)}}$	Indice de Ochiai
$S_8 = \frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	Indice de Sockal & Sneath
$S_8 = \frac{ad-bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$	Phi de Pearson
$S_{10} = \frac{a}{n}$	avec l'unité si les deux objets sont identiques

On trouvera les références d'origine dans ⁸ .

Ces indices sont tous inférieurs ou égaux à 1 et la distance associée est définie par :

$$D_k = \sqrt{1 - S_k}$$

Important Tous les indices retenus donnent des distances euclidiennes ³ .

Exemple Utiliser la carte Artificiel qui contient deux tableaux artificiels ⁹ dont les structures sont évidentes.

Pour chacun des deux, calculer la matrice de distances entre lignes :

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

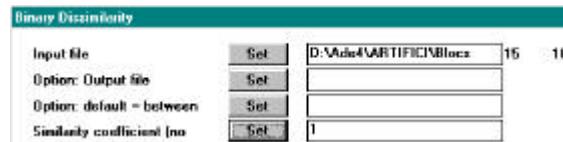
Table 2 p. 23
 Input file: D:\Ade4\ARTIFICI\Dk96
 It has 11 rows and 16 columns
 Distances are computed among rows

JACCARD index (1901)
 S3 coefficient of GOWER & LEGENDRE
 Euclidean distance
 Distances are computed by

```

s = a/(a+b+c) --> d = sqrt(1 - s)
Output file: D:\Ade4\ARTIFICI\Dk96_Sim1
It has 55 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\ARTIFICI\Dk96_Sim1.dma
1 -> 11
2 -> 1
3 -> JACCARD index on D:\Ade4\ARTIFICI\Dk96
4 -> TRUE

```



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:\Ade4\ARTIFICI\Blocs
 It has 15 rows and 16 columns
 Distances are computed among rows

```

JACCARD index (1901)
S3 coefficient of GOWER & LEGENDRE
Euclidean distance
Distances are computed by
s = a/(a+b+c) --> d = sqrt(1 - s)
Output file: D:\Ade4\ARTIFICI\Blocs_Sim1
It has 105 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\ARTIFICI\Blocs_Sim1.dma
1 -> 15
2 -> 1
3 -> JACCARD index on D:\Ade4\ARTIFICI\Blocs
4 -> TRUE

```

Voir l'usage de ces matrices par l'analyse en coordonnées principales ([DMAUse: Principal Coordinates](#)).

Difficulté Elle tient bien sûr au choix d'un indice.

DMAUtil : Canonical distance

Type Utilitaire de création de matrice de distances.

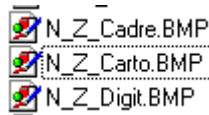
Objet L'objectif est de constituer une matrice de distances à partir d'un tableau de données (métrique canonique, voir ¹⁰ p. 57) :

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

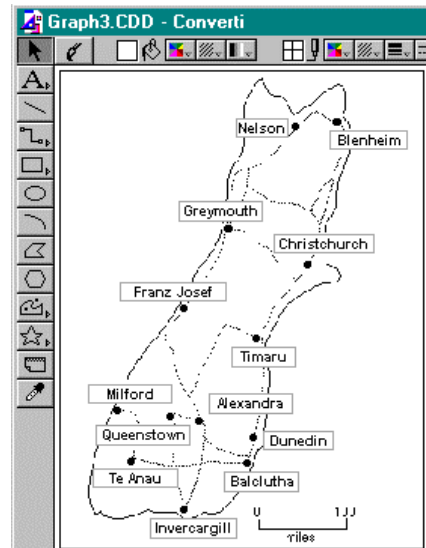
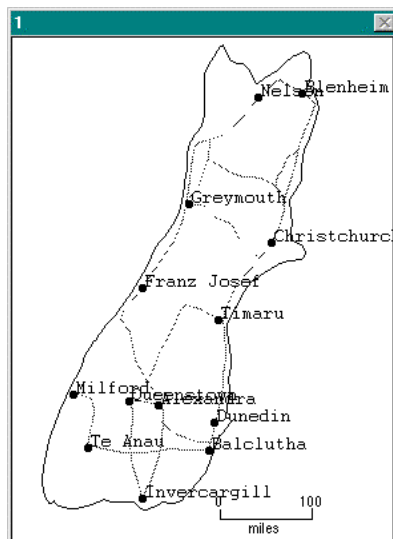
Dialogue L'option utilise une seule fenêtre de dialogue :

- 1) Tableau de données (fichier binaire) avec n lignes et p colonnes.
- 2) Nom du fichier à créer. Par défaut, il dérive du nom du fichier d'entrée.
- 3) Par défaut le calcul porte sur les distances entre lignes et donne une matrice de sortie $n-n$. Taper 1 pour obtenir une matrice de distances entre colonnes : on obtient alors une matrice de sortie $p-p$.

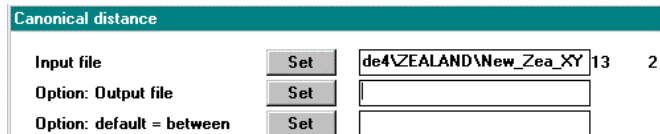
Exemple Utiliser la carte Zealand (10, p.172). Ajouter au dossier de travail les fichiers :



Vérifier la position des étiquettes sur le fond (Maps : Labels arrangé dans ClarisDraw™) :



Calculer la matrice des distances spatiales entre villes (« à vol d'oiseau ») :



```
Distance matrix computation
-----
Input file: D:\Ade4\ZEALAND\New_Zea_XY
It has 13 rows and 2 columns
Distances are computed among rows
-----
Canonical distances computed
Output file: D:\Ade4\ZEALAND\New_Zea_XY_EU
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:\Ade4\ZEALAND\New_Zea_XY_EU.dma
 1 -> 13
 2 -> 1
 3 -> Classical metric on D:\Ade4\ZEALAND\New_Zea_XY
 4 -> TRUE
-----
```

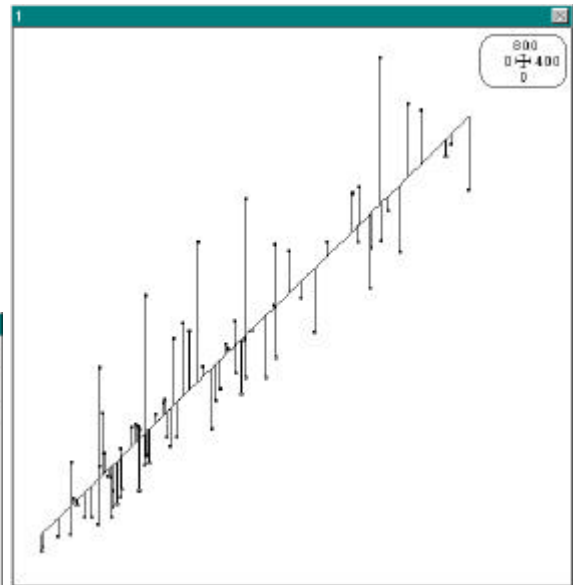
Importer la matrice des distances entre villes (« par la route ») :



```
Input file: D:\Ade4\ZEALAND\Road
D:\Ade4\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:\Ade4\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:\Ade4\ZEALAND\Road_R.dma
 1 -> 13
 2 -> 1
 3 -> Input distance file D:\Ade4\ZEALAND\Road
 4 -> FALSE
-----
```

Comparer les résultats :

Polynomials			
X file (default = 1, 2, 3, ..., n)	Set	\\ZEALAND\New_Zea_XY_EU	78 1
X file column number (default	Set		
Y file (no default)	Set	D:\Ade\ZEALAND\Foad_R	78 1
Order of polynomial (default =	Set		
Weight file (optional)	Set		
Variable label file (or #)	Set		
Draw model (1=yes, 2=no)	Set	1	
Draw observed points (1=yes,	Set	1	
Draw residual sticks (1=yes,	Set	1	



DMAUtil : Edit dma file

Type Utilitaire d'édition de fichier.

Objet A partir d'un fichier X.dma, l'option fait un listing des demi-matrices de distances contenues dans X.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Fichier du type dma permettant l'accès à un fichier de matrices de distances.

Exemple Reprendre l'exemple de DMAUtil: Canonical distance :



```
Distance matrices edition. Input file D:\Ade4W\ZEALAND\Road_R.dma
Text file: D:\Ade4W\ZEALAND\Road_R.dma
```

```
1 -> 13
2 -> 1
3 -> Input distance file D:\Ade4W\ZEALAND\Road
4 -> FALSE
```

```
Input file D:\Ade4W\ZEALAND\Road_R Col 1
max value = 7.56000e+02
Content as 1000*x/max
```

```
-----
[ 1] 0
[ 2] 132 0
[ 3] 642 632 0
[ 4] 376 365 266 0
[ 5] 167 66 565 299 0
[ 6] 308 652 433 327 468 0
[ 7] 459 532 283 209 466 151 0
[ 8] 183 118 750 483 184 503 652 0
[ 9] 328 282 914 647 348 550 734 230 0
[10] 745 710 97 353 652 397 247 836 1000 0
[11] 74 206 653 403 254 302 451 156 235 757 0
[12] 229 183 813 548 249 484 635 131 99 901 155
0
[13] 261 234 397 131 168 414 298 352 499 484 304
417 0
-----
```



```
Input file D:\Ade4W\ZEALAND\New_Zea_XY_EU Col 1
max value = 3.26827e+02
Content as 1000*x/max
```

```
-----
[ 1] 0
[ 2] 158 0
[ 3] 786 848 0
[ 4] 455 501 347 0
[ 5] 135 65 782 436 0
[ 6] 272 406 576 314 353 0
-----
```

[7]	467	571	361	209	508	221	0					
[8]	220	189	1000	661	240	486	688	0				
[9]	197	339	867	574	331	292	512	289	0			
[10]	743	822	101	335	756	512	292	962	804	0		
[11]	68	217	808	490	203	262	473	228	129	758	0	
[12]	189	279	949	633	296	388	606	173	127	895	143	
	0											
[13]	237	301	555	218	236	190	278	446	374	522	275	
	418	0										

DMAUtil : From fuzzy traits

Type Utilitaire de calcul d'une série de matrices de distances

Objet Un tableau de variables floues est une collection de tableaux. Chaque tableau est un tableau de proportions et définit une distance entre lignes. Les matrices de distances (une par tableau) sont rangées dans un même fichier.

Dialogue L'option utilise une seule fenêtre de dialogue :

- 1) Nom du fichier d'entrée de type .fuz (après FuzzyVar: Read Fuzzy File).
- 2) Type de distance utilisée. On a gardé deux options donnant des distances euclidiennes.

Dans les notations de DMAUtil: Genetic distance, ces deux options sont :

$$D_1^k(a,b) = \sqrt{\sum_{j=1}^{m(k)} (p_{aj}^k - p_{bj}^k)^2}$$

$$D_2^k(a,b) = \sqrt{1 - \sum_{j=1}^{m(k)} \sqrt{p_{aj}^k p_{bj}^k}} = \sqrt{\frac{1}{2} \sum_{j=1}^{m(k)} (\sqrt{p_{aj}^k} - \sqrt{p_{bj}^k})^2}$$

Lien Noter que dans DMAUtil: Genetic distance on somme les composantes pour définir une distance globale entre deux lignes en tenant compte de tous les traits, tandis que dans cette option on sépare chacune des distances (indice k) associée à chacun des traits.

Exemple Utiliser les cartes Coleop et Coleop+1.

```
Biological diversity index initialization
Input file = D:\Ade4\COLEOP\EmF.fuz
```

```
File D:\Ade4\COLEOP\EmF_f1 contains distance half-matrices
Rows = 5995 Cols = 2
One half-matrix per column
d21, d31, d32, d41, d42, d43, ..
Text file: D:\Ade4\COLEOP\EmF_f1.dma
1 -> 110
2 -> 2
3 -> Fuzzy traits from file D:\Ade4\COLEOP\EmF.fuz option 1
4 -> TRUE
```

From fuzzy traits

Species-Traits array D:\Ade4\COLEOP\EmF.fuz

Distance type (no default) 2

Biological diversity index initialization
 Input file = D:\Ade4\COLEOP\EmF.fuz

File D:\Ade4\COLEOP\EmF_f2 contains distance half-matrices
 Rows = 5995 Cols = 2
 One half-matrix per column
 d21, d31, d32, d41, d42, d43, ..
 Text file: D:\Ade4\COLEOP\EmF_f2.dma
 1 -> 110
 2 -> 2
 3 -> Fuzzy traits from file D:\Ade4\COLEOP\EmF.fuz option 2
 4 -> TRUE

Correlation matrix PCA

Matrix input file 4User\Dir_Try\Coleop\Esf_f25995 9

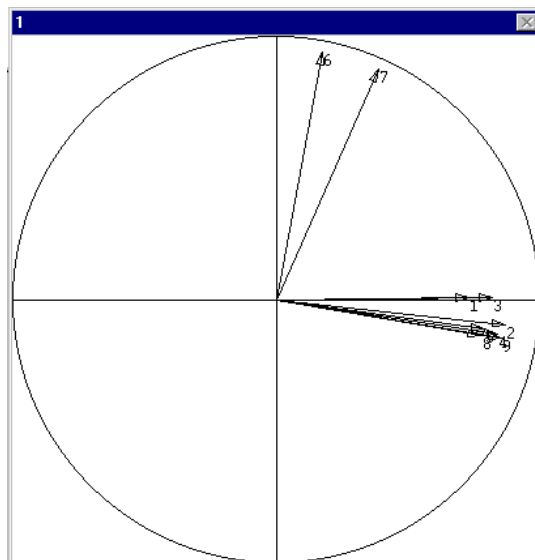
Row weights (default=1/n)

Column weights (default=1)

Option: file for row weight

Option: file for column weight

1 = Save correlation matrix 1



On notera qu'on peut utiliser une collection de matrices de distances comme un tableau de données. Les lignes sont les couples de points (ici les couples d'espèces) et les colonnes sont les différentes distances du fichier. La matrice de corrélation est celle des corrélations de Mantel :

```
Binary input file: D:\ADE4USER\DIR_TRY\COLEOP\Esf_f2.cn+r - 9 rows, 9 cols.
 1 | 1.0000 0.5691 0.5340 0.6375 0.5401 0.1118 0.2623 0.4292 0.5100
 2 | 0.5691 1.0000 0.7423 0.6572 0.6435 0.0689 0.2383 0.6225 0.7338
...
 7 | 0.2623 0.2383 0.3243 0.1988 0.2062 0.8043 1.0000 0.1998 0.2087
 8 | 0.4292 0.6225 0.5533 0.5606 0.5435 0.0313 0.1998 1.0000 0.7808
 9 | 0.5100 0.7338 0.6230 0.6954 0.5618 0.0359 0.2087 0.7808 1.0000
```

On a ici deux groupes de traits écologiques très corrélés et indépendants entre groupes. On retrouve ce résultat dans DMAUse :

DMA Correlation		
First dma type file	<input type="button" value="Set"/>	r\Dir_Try\Coleop\Esf_f2.dma
Second dma type file	<input type="button" value="Set"/>	r\Dir_Try\Coleop\Esf_f2.dma

Mantel's correlation between two distance matrices

```

-----
[ 1] 1000 569 534 637 540 112 262 429 510
[ 2] 569 1000 742 657 644 69 238 623 734
...
[ 7] 262 238 324 199 206 804 1000 200 209
[ 8] 429 623 553 561 543 31 200 1000 781
[ 9] 510 734 623 695 562 36 209 781 1000
-----

```


DMAUtil : From taxonomic tree

Type Utilitaire d'implantation de structure de données.

Objet Le but est de définir les matrices de distances entre espèces à partir d'un arbre taxonomique. Chaque niveau définit une matrice de distance.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Nom du fichier .tax d'entrée qui a été lu par CategVar: Read Taxa File.

Exemple Utiliser la carte AviUrb et le fichier Taxo.tax de la carte AviUrba+2. Lire ce fichier avec CategVar: Read Taxa File :



```
-----  
|      Description of a coded matrix      |  
-----  
Qualitative variables file: D:\Ade4\AVIURB\Taxo  
Number of rows: 40, variables: 4, categories: 89
```

```
Description of categories:  
-----  
Variable number 1 has 1 categories  
-----  
[  1]Category:      1 Num:   40 Freq.:      1
```

```
Variable number 2 has 19 categories  
-----  
[  2]Category:      1 Num:    1 Freq.:    0.025  
[  3]Category:      2 Num:    1 Freq.:    0.025  
....  
[ 19]Category:     18 Num:    1 Freq.:    0.025  
[ 20]Category:     19 Num:    3 Freq.:    0.075
```

```
Variable number 3 has 29 categories  
-----  
[ 21]Category:      1 Num:    1 Freq.:    0.025  
[ 22]Category:      2 Num:    1 Freq.:    0.025  
...  
[ 48]Category:     28 Num:    1 Freq.:    0.025  
[ 49]Category:     29 Num:    2 Freq.:    0.05
```

```
Variable number 4 has 40 categories  
-----  
[ 50]Category:      1 Num:    1 Freq.:    0.025  
[ 51]Category:      2 Num:    1 Freq.:    0.025  
...  
[ 88]Category:     39 Num:    1 Freq.:    0.025  
[ 89]Category:     40 Num:    1 Freq.:    0.025  
-----
```

```
-----  
Auxiliary ASCII output file D:\Ade4\AVIURB\Taxo.123: labels for modalities  
-----
```

Convenient file for taxonomic hierarchy

Category coordinates for graphical display if taxonomic tree in
D:\Ade4\AVIURB\Taxo.xy

Neighbouring relationship in text file: D:\Ade4\AVIURB\TaxoG
It contains graph matrix (LEBART's M) with 89 rows and columns
Neighbouring weights in binary file: D:\Ade4\AVIURB\TaxoG.gpl
It contains 89 rows and 1 column

File D:\Ade4\AVIURB\Taxo.@ob contains an orthonormal basis
It has 40 rows and 40 columns



Biological diversity index initialization
Input file = D:\Ade4\AVIURB\Taxo.tax

File D:\Ade4\AVIURB\Taxo_t contains distances matrices
Rows = 780 Cols = 3

One half-matrix per column
d21, d31, d32, d41, d42, d43, ..

The matrix in the file D:\Ade4\AVIURB\Taxo_t a the property :Sum by row of squared
values = 2

Text file: D:\Ade4\AVIURB\Taxo_t.dma

```
1 -> 40
2 -> 3
3 -> Taxonomic distance from file D:\Ade4\AVIURB\Taxo.tax
4 -> TRUE
```

Ces matrices de distances sont originales. La première définit la distance entre deux espèces au niveau de la famille. La seconde définit la distance entre deux espèces au niveau du genre. La troisième définit la distance entre deux espèces au niveau espèce. La somme des carrés de ces trois distances est toujours égale à 2 :

Binary input file: D:\ADE4\AvisUrb\Taxo_t - 780 rows, 3 cols.

```
1 | 1.4142 0.0000 0.0000
2 | 1.4142 0.0000 0.0000
3 | 1.4142 0.0000 0.0000
4 | 1.1547 0.8165 0.0000
5 | 1.1547 0.8165 0.0000
6 | 1.1547 0.8165 0.0000
7 | 1.1547 0.4082 0.7071
8 | 1.1547 0.4082 0.7071
9 | 1.1547 0.4082 0.7071
10 | 0.0000 1.2247 0.7071
11 | 1.1547 0.4082 0.7071
...
```

Cette valeur de 2 entre deux espèces quelconques est décomposée par niveaux taxonomiques. Pour comprendre la nature de cette décomposition, éditer les matrices avec :



Garder les 12 premières lignes :

```
Input file D:\Ade4\AVIURB\Taxo_t Col 1
max value = 1.41421e+00
Content as 1000*x/max
```

```

-----
[ 1] 0
[ 2] 1000 0
[ 3] 1000 1000 0
[ 4] 816 816 816 0
[ 5] 816 816 816 0 0
[ 6] 816 816 816 0 0 0
[ 7] 1000 1000 1000 816 816 816 0
[ 8] 1000 1000 1000 816 816 816 1000 0
[ 9] 1000 1000 1000 816 816 816 1000 1000 0
[ 10] 866 866 866 645 645 645 866 866 866 0
[ 11] 866 866 866 645 645 645 866 866 866 0 0
-----

```

```

-----
Input file D:\Ade4\AVIURB\Taxo_t Col 2
max value = 1.41421e+00
Content as 1000*x/max
-----

```

```

[ 1] 0
[ 2] 0 0
[ 3] 0 0 0
[ 4] 577 577 577 0
[ 5] 289 289 289 866 0
[ 6] 289 289 289 866 0 0
[ 7] 0 0 0 577 289 289 0
[ 8] 0 0 0 577 289 289 0 0
[ 9] 0 0 0 577 289 289 0 0 0
[ 10] 500 500 500 764 577 577 500 500 500 0
[ 11] 500 500 500 764 577 577 500 500 500 1000 0
-----

```

```

-----
Input file D:\Ade4\AVIURB\Taxo_t Col 3
max value = 1.41421e+00
Content as 1000*x/max
-----

```

```

[ 1] 0
[ 2] 0 0
[ 3] 0 0 0
[ 4] 0 0 0 0
[ 5] 500 500 500 500 0
[ 6] 500 500 500 500 1000 0
[ 7] 0 0 0 0 500 500 0
[ 8] 0 0 0 0 500 500 0 0
[ 9] 0 0 0 0 500 500 0 0 0
[ 10] 0 0 0 0 500 500 0 0 0 0
[ 11] 0 0 0 0 500 500 0 0 0 0 0
-----

```

La distance totale (au carré) entre deux espèces est systématiquement 2. Ce total est entièrement décomposé aux niveaux espèce-genre-famille. La part d'un niveau dans cette décomposition est nulle si à ce niveau les espèces sont dans la même classe. 5 et 6 sont du même genre, donc de la même famille et la décomposition de leur distance est du type (0, 0, x) et donc x vaut $\sqrt{2}$. 20 et 21 sont de la même famille mais pas du même genre et la décomposition de leur distance est du type (0, x, y) avec $x + y = 2$. Pour deux espèces de deux familles différentes la décomposition est du type (x, y, z) avec $x + y + z = 2$.

Ensuite le procédé est complexe. Au niveau famille les espèces de Sylviidae sont toutes à une distance nulle entres :

```

[ 20] 0 Niveau famille
[ 21] 0 0
[ 22] 0 0 0
[ 23] 0 0 0 0
[ 24] 0 0 0 0 0
-----

```

Elles sont dans trois genres (20, 21-22 et 23-24) :

```
[ 20] 0 Niveau genre
[ 21] 750 0
[ 22] 750 0 0
[ 23] 750 500 500 0
[ 24] 750 500 500 0 0
```

La distance entre deux espèces de deux genres au niveau genre dépend du nombre d'espèces par genre. Le procédé répartit la distance entre les niveaux :

```
[ 20] 0 Niveau espèce
[ 21] 250 0
[ 22] 250 1000 0
[ 23] 250 500 500 0
[ 24] 250 500 500 1000 0
```

Deux espèces de deux genres différents d'une même famille sont à une distance de 2 qui se décompose en $3/2 + 1/2$ ou $1 + 1$ en fonction du nombre d'espèces dans les genres (ici 1 ou 2). L'utilisation de genre monospécifique augmentera donc la diversité au niveau genre et l'utilisation de genres plurispécifiques augmentera la diversité au niveau espèce, comme si la diversité qu'engendre un taxon est fonction du nombre potentiel des éléments voisins. Il ne semble pas y avoir d'algorithme explicatif simple mais le support mathématique garantit que le procédé est parfaitement univoque.

DMAUtil : Genetic distance

Type Utilitaire de calcul de matrices de distances.

Objet Cette option calcule des matrices de distances à partir de tableaux de fréquences alléliques. Les lignes du tableau sont des populations. Les colonnes sont regroupées par bloc, chaque bloc étant un locus. Les colonnes d'un même bloc sont les allèles de ce locus. Les cellules du tableau contiennent des nombres positifs ou nuls indiquant la fréquence de rencontre dans la population de chaque allèle de chaque locus (structure de variables floues). Les données sont soit des effectifs (carte de données Chrysichtys ¹¹), soit des fréquences exprimées en pour cent (carte de données Sicile ¹² ou Chevaine ¹³) ou en valeur. Dans tous les cas les données sont ramenées en fréquence par blocs (somme unité).

Soit **A** un tableau de fréquences alléliques avec t lignes (populations) et m colonnes (allèles). Soit v le nombre de loci. Le locus j a $m(j)$ allèles.

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

Pour la $i^{\text{ème}}$ ligne et la $k^{\text{ème}}$ modalité de la variable j , on note la valeur a_{ij}^k ($1 \leq i \leq t$, $1 \leq j \leq v$, et $1 \leq k \leq m(j)$), la valeur du tableau des données brutes. Soit :

$$a_{ij}^+ = \sum_{k=1}^{m(j)} a_{ij}^k \quad \text{et} \quad p_{ij}^k = \frac{a_{ij}^k}{a_{ij}^+}$$

Soit le tableau $\mathbf{P} = [p_{ij}^k]$ et les paramètres :

$$p_{ij}^+ = \sum_{k=1}^{m(j)} p_{ij}^k = 1, \quad p_{i+}^+ = \sum_{j=1}^v p_{ij}^+ = v, \quad p_{++}^+ = \sum_{j=1}^v p_{i+}^+ = tv$$

L'option calcule des matrices de distances entre populations utilisant les fréquences p_{ij}^k .

Dialogue L'option utilise une seule fenêtre de dialogue :

Genetic distance		
Input file	Set	
Category indication file	Set	
Option: Output file	Set	
Distance type (no default)	Set	

- 1) Nom du fichier binaire d'entrée.
- 2) Nom du fichier binaire d'indicateur de blocs.
- 3) Nom du fichier binaire de sortie. Par défaut, il dérive du nom de fichier d'entrée et du type de distance utilisée.
- 4) Type de distance utilisée.

Il y a 3 options :

1 — Distance de Rogers ¹⁴ (Voir ¹⁵) :

$$D_1(a,b) = \frac{1}{v} \sum_{k=1}^v \sqrt{\frac{1}{2} \sum_{j=1}^{m(k)} (p_{aj}^k - p_{bj}^k)^2}$$

2 — Distance de Nei ¹⁶ (Voir ¹⁵) :

$$D_2(a,b) = -Ln \left(\frac{\sum_{k=1}^v \sum_{j=1}^{m(k)} p_{aj}^k p_{bj}^k}{\sqrt{\sum_{k=1}^v \sum_{j=1}^{m(k)} (p_{aj}^k)^2} \sqrt{\sum_{k=1}^v \sum_{j=1}^{m(k)} (p_{bj}^k)^2}} \right)$$

3 — Distance de Edwards ¹⁷ (Voir ¹⁸) :

$$D_3(a,b) = \sqrt{\frac{1}{v} \sum_{k=1}^v \left(1 - \sum_{j=1}^{m(k)} \sqrt{p_{aj}^k p_{bj}^k} \right)}$$

Exemple Utiliser la carte Chevaïne :

Genetic distance			
Input file	Set	D:\Ade4\CHEVAINE\Freq	27 9
Category indication file	Set	D:\Ade4\CHEVAINE\BloVar	4 1
Option: Output file	Set		
Distance type (no default)	Set	1	

```
Distance amongst multiple frequency distributions
Input file: D:\Ade4\CHEVAINE\Freq
It has 27 rows and 9 columns
Bloc indicator: D:\Ade4\CHEVAINE\BloVar
Distances are computed among rows
```

```
d1=mean(sqrt(0.5*Sum(p(i)-q(i)^2)))
Rogers 1972 in Avise 1994 p. 95
Output file: D:\Ade4\CHEVAINE\Freq_Gen1
It has 351 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\CHEVAINE\Freq_Gen1.dma
1 -> 27
2 -> 1
3 -> ROGERS on D:\Ade4\CHEVAINE\Freq
4 -> TRUE
```

Genetic distance			
Input file	Set	D:\Ade4\CHEVAINE\Freq	27 9
Category indication file	Set	D:\Ade4\CHEVAINE\BloVar	4 1
Option: Output file	Set		
Distance type (no default)	Set	2	

```
Distance amongst multiple frequency distributions
Input file: D:\Ade4\CHEVAINE\Freq
It has 27 rows and 9 columns
Bloc indicator: D:\Ade4\CHEVAINE\BloVar
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:\Ade4\CHEVAINE\Freq_Gen2
It has 351 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\CHEVAINE\Freq_Gen2.dma
1 -> 27
2 -> 1
3 -> NEI on D:\Ade4\CHEVAINE\Freq
4 -> FALSE

```

Genetic distance		
Input file	Set	D:\Ade4\CHEVAINE\Freq 27 9
Category indication file	Set	D:\Ade4\CHEVAINE\BloVar 4 1
Option: Output file	Set	
Distance type (no default)	Set	3

Distance amongst multiple frequency distributions

Input file: D:\Ade4\CHEVAINE\Freq

It has 27 rows and 9 columns

Bloc indicator: D:\Ade4\CHEVAINE\BloVar

Distances are computed among rows

$$d3 = \sqrt{\text{Mean} (1 - (\text{Sum}(\sqrt{p(i)q(i)})))}$$

Edwards 1971 in Hartl & Clark 1989 multi locus extension

Output file: D:\Ade4\CHEVAINE\Freq_Gen3

It has 351 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:\Ade4\CHEVAINE\Freq_Gen3.dma

```

1 -> 27
2 -> 1
3 -> EDWARDS on D:\Ade4\CHEVAINE\Freq
4 -> TRUE

```

Voir dans DMAUtil: ToClusters l'utilisation de ces matrices.

DMAUtil : Lingoès correction

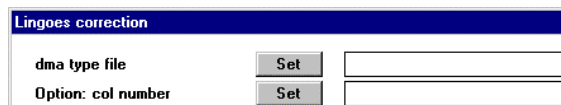
Type Méthode heuristique rendant euclidienne une matrice de distances.

Objet Gower⁶ et⁷ a montré l'intérêt des matrices de distances euclidiennes. On peut toujours générer une matrice de distance euclidienne à partir d'une matrice de distances qui ne l'est pas en ajoutant une constante positive c à chacun des carrés des termes hors diagonale, c'est-à-dire en utilisant le changement de variable :

$$d_{ij} = \sqrt{d_{ij}^2 + c}$$

Si la matrice de distances n'est pas euclidienne, l'option calcule cette constante, la plus petite possible, avec la solution donnée par Lingoès (1971)¹⁹. Cette transformation est de grand intérêt et est conseillée par²⁰.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Nom du fichier dma permettant l'accès à un fichier de matrice de distances

2) Numéro de la colonne à transformer. Par défaut, c'est la première. Si le fichier d'entrée s'appelle X.dma et renvoie au fichier X, le fichier de sortie contenant la distance transformée de la colonne n s'appellera X_1_n et sera accessible par le fichier X_1_n.dma.

Exemple Utiliser la carte Zealand (voir DMAUtil : Canonical distance):



```
Input file: E:\Ade4\ZEALAND\Road
E:\Ade4\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: E:\Ade4\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: E:\Ade4\ZEALAND\Road_R.dma
1 -> 13
2 -> 1
3 -> Input distance file E:\Ade4\ZEALAND\Road
4 -> FALSE
```

La matrice des distances routière n'est pas euclidienne :



```
Input file: E:\Ade4\ZEALAND\Road_R
additive cte: 1.970e+03
```



```

Output file: E:\Ade4\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: E:\Ade4\ZEALAND\Road_R_c.dma
1 -> 13
2 -> 1
3 -> Additive constante to E:\Ade4\ZEALAND\Road_R
4 -> TRUE

```



```

Num Eigenval. | Num. Eigenval. | Num. Eigenval. | Num. Eigenval. |
001 6.701e+05 | 002 1.282e+05 | 003 2.423e+04 | 004 1.500e+04 |
005 6.347e+03 | 006 3.861e+03 | 007 1.371e+03 | 008 -1.100e-11 |
009 -6.650e+02 | 010 -2.772e+03 | 011 -8.324e+03 | 012 -2.233e+04 |
013 -7.773e+04 |

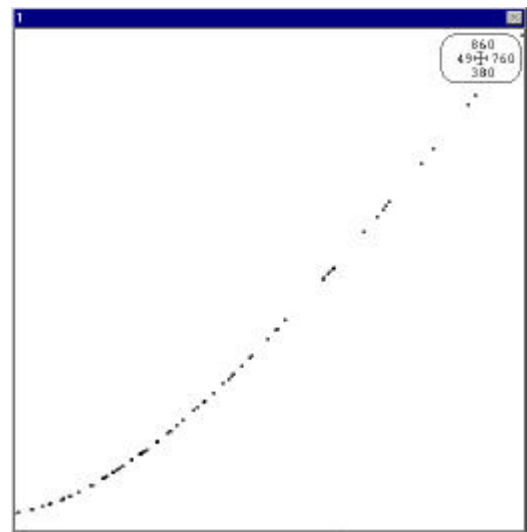
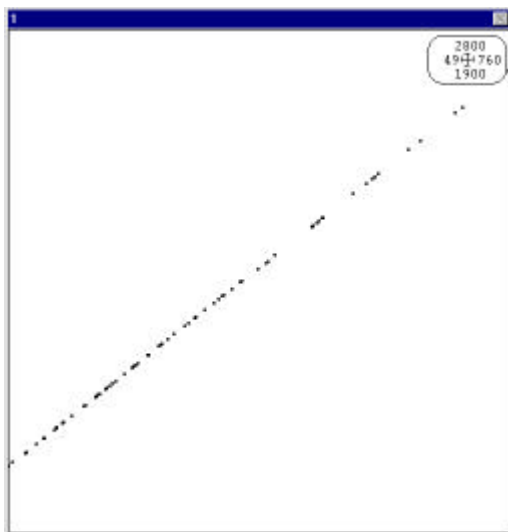
```

```

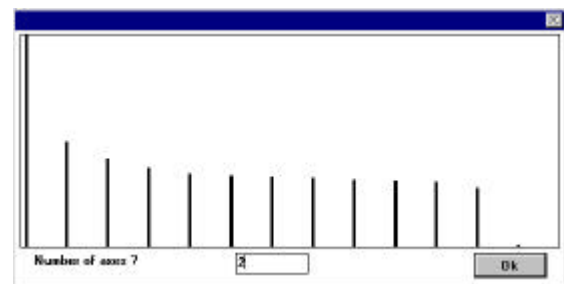
Input file: E:\Ade4\ZEALAND\Road_R
Lingoes correction 2*c1 : 1.555e+05
Output file: E:\Ade4\ZEALAND\Road_R_1
It has 78 rows and 1 columns
d(2,1), d(3,1), d(3,2),..., d(n,1), d(n,2), ... d(n,n-1)
Text file: E:\Ade4\ZEALAND\Road_R_1.dma
1 -> 13
2 -> 1
3 -> Lingoes correction on E:\Ade4\ZEALAND\Road_R
4 -> TRUE

```

La nouvelle matrice de distances est euclidienne.

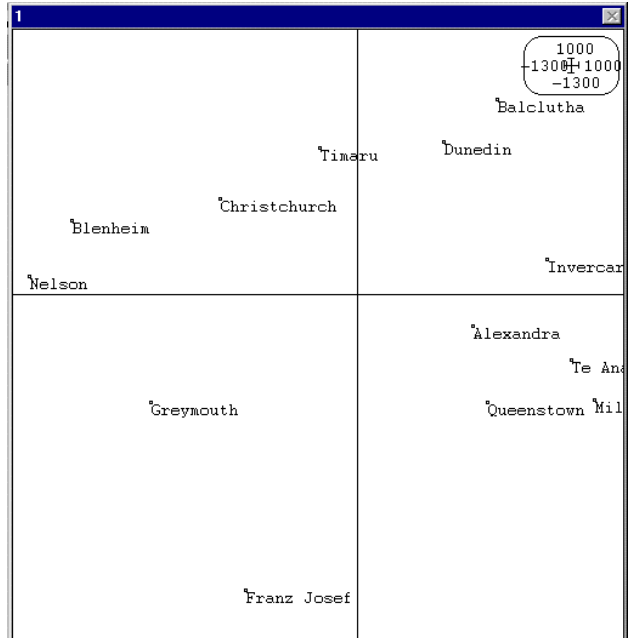
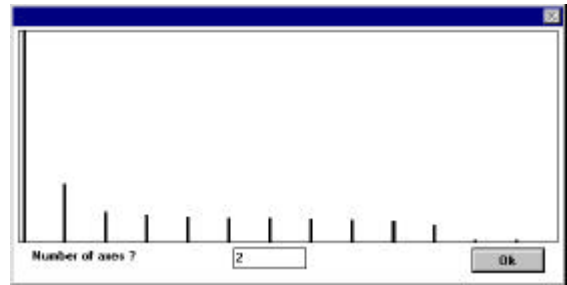


A gauche, transformation par la constante additive, à droite, transformation par la procédure de Lingoes.



Principal Coordinates

dma type file e4\ZEALAND\Road_R_I.dma



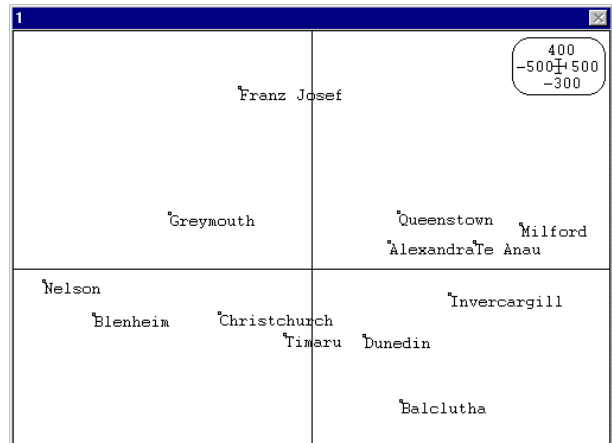
Labels

XY coordinates file e4\ZEALAND\Road_R_I.coo 13 2

X-axis column number (default)

Y-axis column number (default)

Label file (or #) for items Adb\ZEALAND\Code_Town



Labels

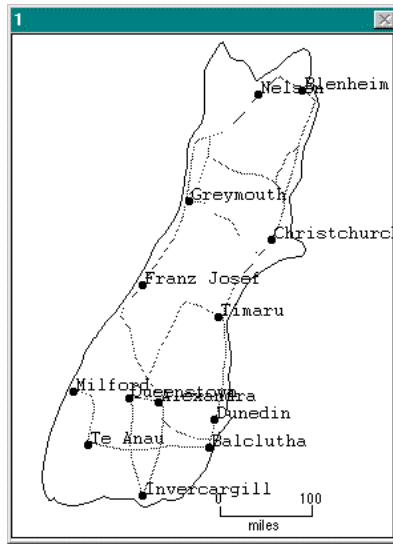
XY coordinates file e4\ZEALAND\Road_R_I.coo 13 2

X-axis column number (default)

Y-axis column number (default)

Label file (or #) for items Adb\ZEALAND\Code_Town

Comparer avec la « carte routière ».

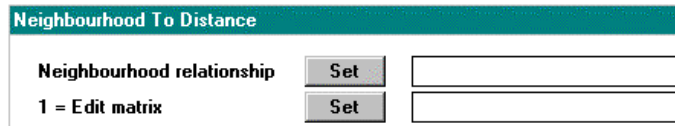


DMAUtil : Neighbourhood To Distance

Type Utilitaire de calcul de la distance associée à un graphe de voisinages.

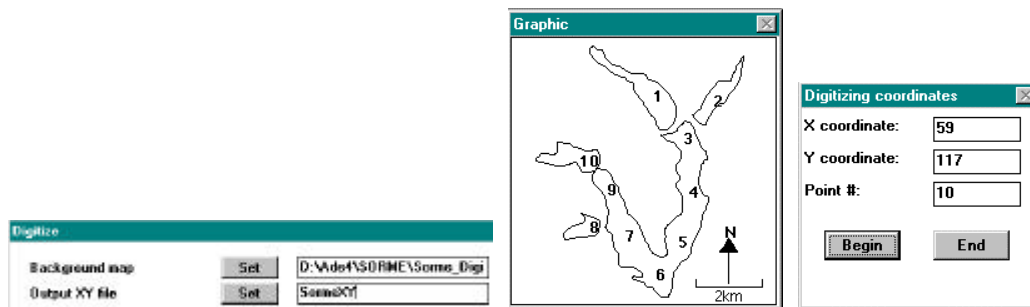
Objet Si M est la matrice d'un graphe connexe, on peut simplement définir la distance entre deux sommets comme le nombre d'arêtes minimum d'un chemin reliant ces deux sommets ²¹ (p. 191).

Dialogue L'option utilise une seule fenêtre de dialogue :

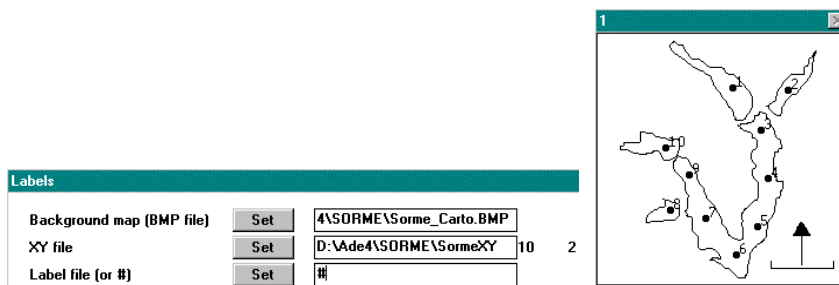


- 1) Nom du fichier .gpl d'accès au graphe de voisinage.
- 2) Option d'édition : taper 1 pour éditer la matrice de distances.

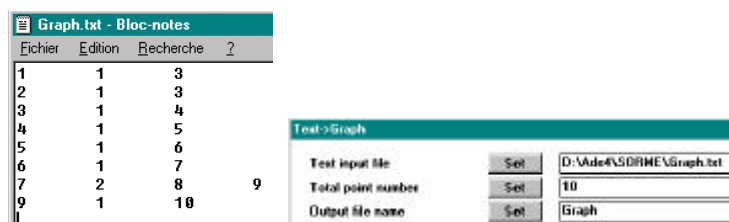
Exemple Utiliser la carte Sorme. Digitaliser les 10 stations sur le fond de carte (Digit : Digitize) :



Vérifier les positions (Maps : Labels) :



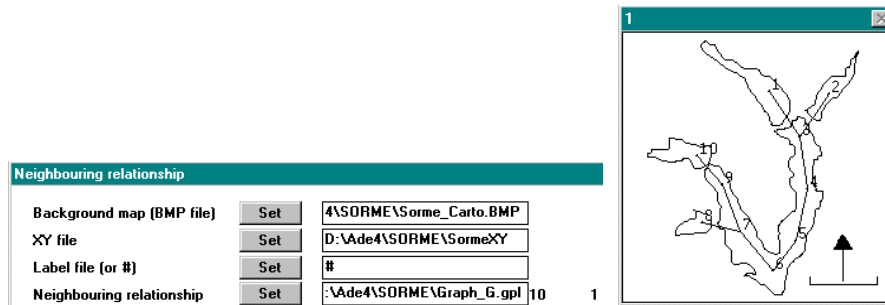
Implanter le graphe de voisinage pour NGUtil: Text->Graph :



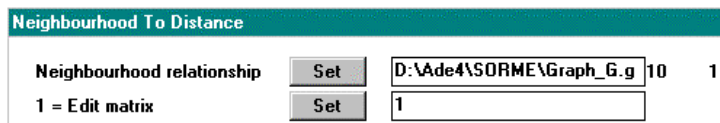
```
Neighbouring relationship
Input file: D:\Ade4\SORME\Graph.txt
Neighbouring relationship in text file: Graph_G
It contains graph matrix (LEBART's M) with 10 rows and columns
```

Neighbouring weights in binary file: Graph_G.gpl
 It contains 10 rows and 1 column

Vérifier (Maps : Neighbouring Relationship) :



Calculer la matrice de distances :



```
Multiscale neighborhood relation from neighborhood graph
Input file (neighborhood graph): D:\Ade4\SORME\Graph_G.gpl
Multiscale neighborhood graph in binary file: D:\Ade4\SORME\Graph_G_gms
Aij = k <=> arcs of shortest length between i and j have k edges
Test of the euclidean property by diagonalization (theorem of GOWER)
Output file: D:\Ade4\SORME\Graph_G_gms
It has 45 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\SORME\Graph_G_gms.dma
1 -> 10
2 -> 1
3 -> Shortest arc length distance from D:\Ade4\SORME\Graph_G.gpl
4 -> FALSE
```

```
-----
- - - - -
* 2 1 2 3 4 5 6 6 7
2 * 1 2 3 4 5 6 6 7
1 1 * 1 2 3 4 5 5 6
2 2 1 * 1 2 3 4 4 5
3 3 2 1 * 1 2 3 3 4
4 4 3 2 1 * 1 2 2 3
5 5 4 3 2 1 * 1 1 2
6 6 5 4 3 2 1 * 2 3
6 6 5 4 3 2 1 2 * 1
7 7 6 5 4 3 2 3 1 *
- - - - -
```

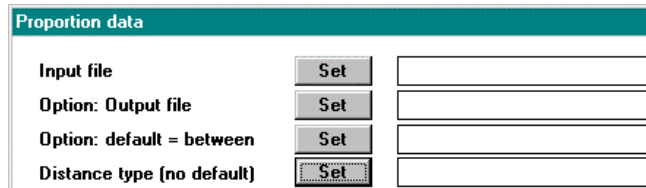
Observer qu'il faut bien 7 arcs pour aller de la station 1 à la station 10 en utilisant ce graphe.

DMAUtil : Proportion data

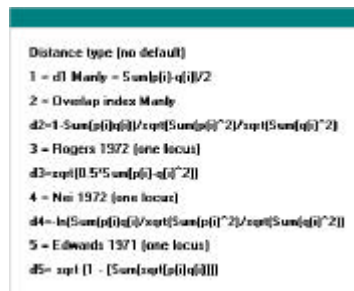
Type Utilitaire de calcul de matrices de distances.

Objet Un tableau **X** ne contient que des nombres positifs ou nuls et est considéré comme définissant des distributions de fréquences par ligne ou par colonne. L'option calcule des matrices de distances entre ces profils.

Dialogue L'option utilise une seule fenêtre de dialogue :



- 1) Nom du fichier binaire d'entrée.
- 2) Nom générique des fichiers de sortie de sortie (création). Par défaut, il dérive du nom du fichier d'entrée et de l'option de calcul choisi.
- 3) Option de calcul. Utiliser 1 pour calculer les distances entre colonnes du tableau. Par défaut, les distances sont calculées entre lignes.
- 4) Option de choix de l'indice de distance. Il y a 5 options :



1 — Distance d_1 (voir ¹⁰ formule (5.7) p. 68).

$$d_1 = \frac{1}{2} \sum_i |p_i - q_i|$$

2 — Distance d_2 (indice de chevauchement de niche, voir ¹⁰ formule (5.8) p. 68).

$$d_2 = 1 - \frac{\sum_i p_i q_i}{\sqrt{\sum_i p_i^2} \sqrt{\sum_i q_i^2}}$$

3 — Distance de Rogers (Voir DMAUtil : Genetic distance) :

$$d_3 = \frac{1}{2} \sqrt{\sum_i (p_i - q_i)^2}$$

4 — Distance de Nei (Voir DMAUtil : Genetic distance) :

$$d_4 = -\ln \left(\frac{\sum_i p_i q_i}{\sqrt{\sum_i p_i^2} \sqrt{\sum_i q_i^2}} \right)$$

5 — Distance de Edwards (Voir DMAUtil : Genetic distance) :

$$d_5 = \sqrt{1 - \sum_i \sqrt{p_i q_i}}$$

Exemple Utiliser la carte Butterfly (^{22, 10} p. 29).

Proportion data	
Input file	Set : \Ade4\BUTTERFL\But_Biol 16 6
Option: Output file	Set :
Option: default = between	Set :
Distance type (no default)	Set : 1

Distance amongst frequency distributions

Input file: D:\Ade4\BUTTERFL\But_Biol

It has 16 rows and 6 columns

Distances are computed among rows

d1 distances computed

Manly 1994 Multivariate statistical methods. A primer

2nd edition. Chapman & Hall 1994. formula 5.7 p. 68

Test of the euclidean property by diagonalization (theorem of GOWER)

Output file: D:\Ade4\BUTTERFL\But_Biol_Fre1

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_Biol_Fre1.dma

1 -> 16

2 -> 1

3 -> d1 of MANLY on D:\Ade4\BUTTERFL\But_Biol

4 -> **FALSE**

De même, avec les options suivantes :

Distance amongst frequency distributions

Input file: D:\Ade4\BUTTERFL\But_Biol

It has 16 rows and 6 columns

Distances are computed among rows

d2 distances computed

Manly 1994 Multivariate statistical methods. A primer

2nd edition. Chapman & Hall 1994. formula 5.8 p. 68

Test of the euclidean property by diagonalization (theorem of GOWER)

Output file: D:\Ade4\BUTTERFL\But_Biol_Fre2

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_Biol_Fre2.dma

1 -> 16

2 -> 1

3 -> d2 of MANLY on D:\Ade4\BUTTERFL\But_Biol

4 -> **FALSE**

Distance amongst frequency distributions

Input file: D:\Ade4\BUTTERFL\But_Biol

It has 16 rows and 6 columns

Distances are computed among rows

d3=sqrt(0.5*Sum(p(i)-q(i)^2))

Rogers 1972 in Avise 1994 p. 95

Output file: D:\Ade4\BUTTERFL\But_Biol_Fre3

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_Biol_Fre3.dma
  1 -> 16
  2 -> 1
  3 -> ROGERS on D:\Ade4\BUTTERFL\But_Biol
  4 -> TRUE
-----
Distance amongst frequency distributions
Input file: D:\Ade4\BUTTERFL\But_Biol
It has 16 rows and 6 columns
Distances are computed among rows

d4=-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:\Ade4\BUTTERFL\But_Biol_Fre4
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_Biol_Fre4.dma
  1 -> 16
  2 -> 1
  3 -> NEI on D:\Ade4\BUTTERFL\But_Biol
  4 -> FALSE

Distance amongst frequency distributions
Input file: D:\Ade4\BUTTERFL\But_Biol
It has 16 rows and 6 columns
Distances are computed among rows

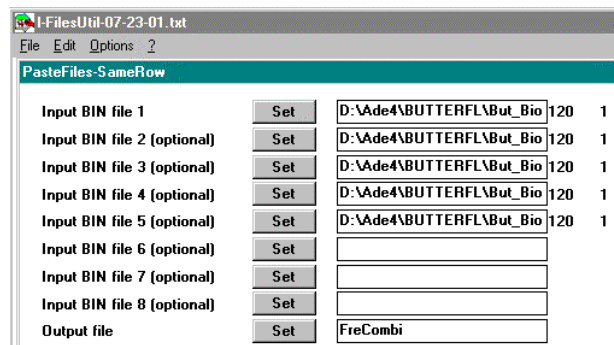
```

```

d5 = sqrt (1 - (Sum(sqrt(p(i)q(i))))
Edwards 1971 in Hartl & Clark 1989
Output file: D:\Ade4\BUTTERFL\But_Biol_Fre5
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_Biol_Fre5.dma
  1 -> 16
  2 -> 1
  3 -> EDWARDS on D:\Ade4\BUTTERFL\But_Biol
  4 -> TRUE
-----

```

Assembler les cinq fichiers :



```

pastecol: Paste files with same row number
Input file: D:\Ade4\BUTTERFL\But_Biol_Fre1
--- Number of rows: 120, columns: 1
Input file: D:\Ade4\BUTTERFL\But_Biol_Fre2
...
Output file: FreCombi
--- Number of rows: 120, columns: 5

```

On se retrouve avec cinq matrices de distances dans un même fichier. Pour travailler simultanément sur ces cinq points de vue, il faut relire ce fichier pour lui associer un fichier .dma :


```
Read half distances matrix
Input file  120 5
```

```
Input file: D:\Ade4\BUTTERFL\FreCombi
D:\Ade4\BUTTERFL\FreCombi is a binary file with 120 rows and 5 columns
120 is 16*(16-1)/2
Test of the euclidean property by diagonalization (theorem of GOWER)
matrix number 1 ----> FALSE
matrix number 2 ----> FALSE
matrix number 3 ----> TRUE
matrix number 4 ----> FALSE
matrix number 5 ----> TRUE
Text file: D:\Ade4\BUTTERFL\FreCombi_R.dma
1 -> 16
2 -> 5
3 -> Input half distance matrix file D:\Ade4\BUTTERFL\FreCombi
4 -> FALSE
```

On a choisi de dire que la famille de distances est euclidienne si elles le sont toutes, ce qui n'est pas le cas ici.

DMAUtil : Quantitative variables

Type Utilitaire de calcul de matrices de distances pour variable quantitative.

Objet Un tableau \mathbf{X} , avec n lignes et p colonnes, contient des valeurs quantitatives $\mathbf{X} = [x_{ij}]$. L'option calcule une matrice des distances entre lignes ou entre colonnes avec un critère au choix parmi 7 possibilités choisies dans ³.

Dialogue L'option utilise une seule fenêtre de dialogue :

- 1) Nom du fichier binaire d'entrée.
- 2) Nom générique des fichiers de sortie (création). Par défaut ce nom dérive de celui du fichier d'entrée et de l'option choisie.
- 3) Option de calcul. Utiliser 1 pour calculer les distances entre colonnes du tableau. Par défaut, les distances sont calculées entre lignes.
- 4) Option de choix de l'indice de dissimilarité. Il y a 7 options :

1 — Distance de Manhattan, ou city block, ou de Gower 1971a (référence p. 20 dans ⁹), ou D3 de Gower & Legendre ³ (distance non euclidienne) :

$$d_1(i, j) = \frac{1}{p} \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{r_k} \text{ avec } r_k = \max_{i=1}^n (x_{ik}) - \min_{i=1}^n (x_{ik})$$

2 — Distance de Manhattan, ou de Cain & Harrison (référence p. 20 dans ⁹), ou D3 de Gower & Legendre ³ (distance non euclidienne) :

$$d_2(i, j) = \frac{1}{p} \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{r_k} \text{ avec } r_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - m_k)^2} \text{ et } m_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$$

3 — Distance de Canberra, ou de Lance & Williams (référence p. 20 dans ⁹), ou D7 de Gower & Legendre ³ (distance non euclidienne) :

$$d_3(i, j) = \frac{1}{p} \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|}$$

4 — Distance de Bray-Curtis ou de Odum (références p. 20 dans ⁹), ou D8 de Gower & Legendre ³ (distance non euclidienne) :

$$d_4(i, j) = \frac{\frac{1}{p} \sum_{k=1}^p |x_{ik} - x_{jk}|}{\sum_{k=1}^p x_{ik} + x_{jk}}$$

5 — Distance D₅ de Gower & Legendre ³ pour données positives seulement (distance euclidienne) :

$$d_5(i, j) = \frac{1}{p} \sum_{k=1}^p \frac{(x_{ik} - x_{jk})^2}{(x_{ik} + x_{jk})^2}$$

6 — Distance D₆ de Gower & Legendre ³ pour données positives seulement (distance non euclidienne) :

$$d_6(i, j) = \frac{\sum_{k=1}^p |x_{ik} - x_{jk}|}{\sum_{k=1}^p \max(x_{ik}, x_{jk})}$$

7 — Distance D₁₀ de Gower & Legendre ³ pour données positives seulement (distance non euclidienne) :

$$d_7(i, j) = \frac{1}{p} \sum_{k=1}^p \left(1 - \frac{\min(x_{ik}, x_{jk})}{\max(x_{ik}, x_{jk})} \right)$$

Exemple Utiliser la carte AviUrba :

Quantitative variables			
Input file	Set	D:\Ade4\AVIURB\AUFau	51 40
Option: Output file	Set		
Option: default = between	Set		
Distance type (no default)	Set	4	

```
Distance matrix computation from dissimilarity coefficients
Dissimilarity coefficients amongst quantitative variables
Gower J.C. & Legendre P. (1986)
Metric and Euclidean properties of dissimilarity coefficients
Journal of Classification, 3, 5-48
Table 3 p. 27
```

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

```
Bray-Curtis
D8 coefficient of GOWER & LEGENDRE
Non Euclidean distance
Distances are computed by
dij = (1/p)Sum|xik-xjk|/(Sum(xik+xjk) 1<=k<=p
Output file: D:\Ade4\AVIURB\AUFau_Dqv4
It has 1275 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\AVIURB\AUFau_Dqv4.dma
1 -> 51
2 -> 1
3 -> D8 coefficient of GOWER & LEGENDRE on D:\Ade4\AVIURB\AUFau
4 -> FALSE
```

ToClusters		
dma type file	Set	4\AVIURB\AUFau_Dqv4.dma
Option: col number	Set	

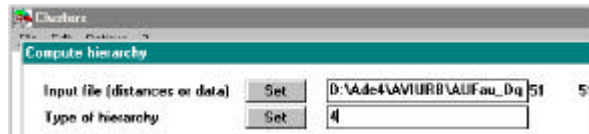
```
Input file (distance matrix): D:\Ade4\AVIURB\AUFau_Dqv4.dma
Text file: D:\Ade4\AVIURB\AUFau_Dqv4.dma
1 -> 51
```

```

2 -> 1
3 -> D8 coefficient of GOWER & LEGENDRE on D:\Ade4\AVIURB\AUFau
4 -> FALSE

```

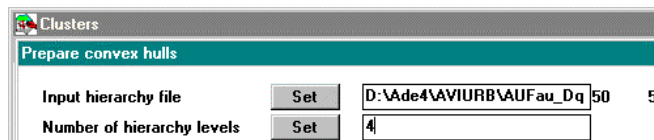
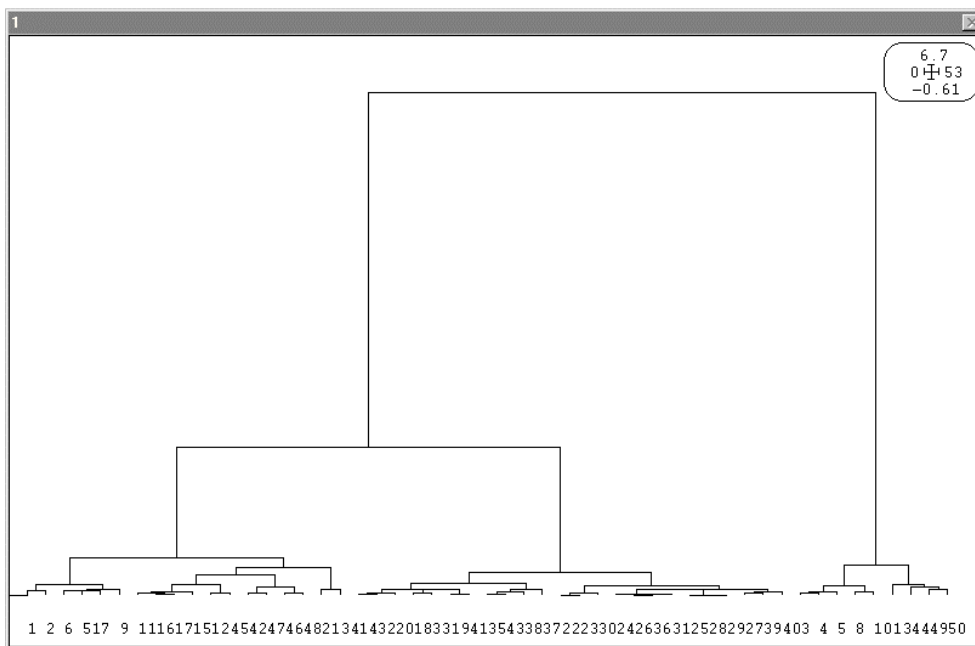
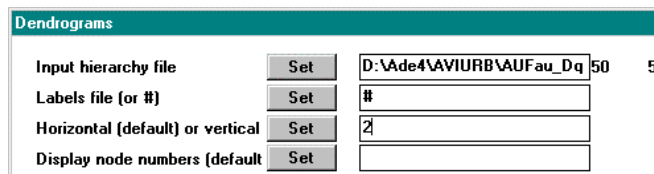
Output file : D:\Ade4\AVIURB\AUFau_Dqv41.dis Row: 51 Col: 51
 Transformation: rescaling on [0,1] by $y=(x-\min)/(\max-\min)$



```

Clusters: Compute hierarchy
Data file: D:\Ade4\AVIURB\AUFau_Dqv41.dis
Number of rows: 51, columns: 51
Output file: D:\Ade4\AVIURB\AUFau_Dqv41.2mha
Number of rows: 50, columns: 5
Hierarchy algorithm used : second order moment (Ward's method)

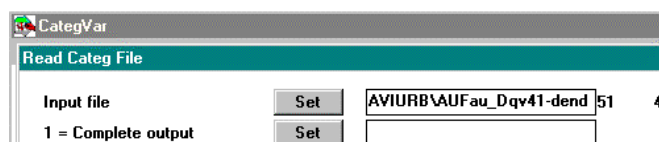
```



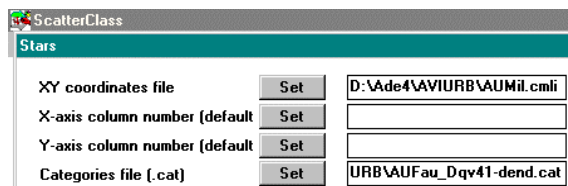
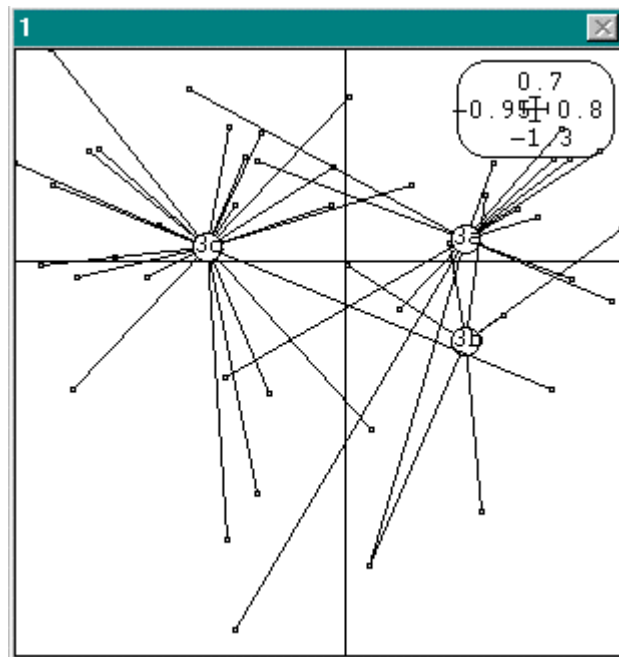
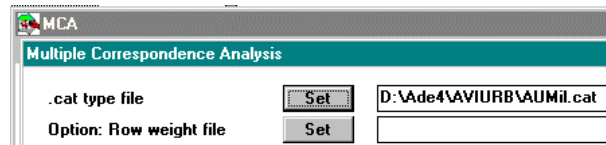
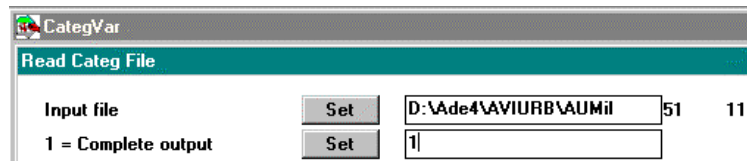
```

Clusters: Compute dendrogram
Hierarchy file: D:\Ade4\AVIURB\AUFau_Dqv41.2mha
Number of rows: 50, columns: 5
Output file: D:\Ade4\AVIURB\AUFau_Dqv41-dend
Number of rows: 51, columns: 4

```



Ajouter les fichiers de la carte AviUrba+1 :



On a reporté sur le plan d'une analyse d'un tableau milieu la partition obtenue sur le tableau faunistique. On peut aussi faire l'inverse ou faire deux classifications et croiser les partitions obtenues.

DMAUtil : Quasi_Euclidean

Type Utilitaire d'approximation d'une matrice de distances par une matrice de distances euclidienne.

Objet Quand on importe une matrice de distance, on le fait avec une précision numérique fonction de la source. La matrice importée peut être euclidienne en théorie mais la précision numérique insuffisante peut laisser croire le contraire. L'option permet de remédier à ce problème.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Nom de fichier de type dma.

Exemple Manly (1991 p. 126)²³ propose une matrice de distances géographiques entre 19 villages publiée par Spilman (1973)²⁴. La matrice est dans la carte Yanomama :

	1	2	3	4	5	6	...
1	0						
2	9	0					
3	28	20	0				
4	152	161	178	0			
5	149	158	175	7	0		
6	169	175	184	102	95	0	
...							

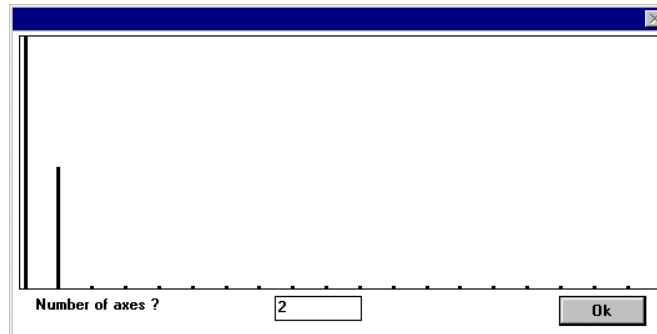
Cette matrice est vraisemblablement euclidienne. Cependant à la lecture, on obtient :



```
Input file: E:\Ade4\YANOMAMA\geogra
E:\Ade4\YANOMAMA\geogra is a binary file with 19 rows and 19 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: E:\Ade4\YANOMAMA\geogra_R
It has 171 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\YANOMAMA\geogra_R.dma
1 -> 19
2 -> 1
3 -> Input distance file E:\Ade4\YANOMAMA\geogra
4 -> FALSE
```



Num.	Eigenval.	Num.	Eigenval.	Num.	Eigenval.	Num.	Eigenval.
001	6.684e+03	002	3.187e+03	003	1.367e+01	004	1.274e+01
005	6.799e+00	006	5.000e+00	007	3.300e+00	008	1.780e+00
009	1.061e+00	010	-1.422e-12	011	-6.938e-01	012	-1.141e+00
013	-1.549e+00	014	-2.534e+00	015	-4.177e+00	016	-6.205e+00
017	-7.842e+00	018	-9.242e+00	019	-1.269e+01		



Column: 1

Num	Eigenval.	Num.	Eigenval.	Num.	Eigenval.	Num.	Eigenval.
001	6.684e+03	002	3.187e+03	003	1.367e+01	004	1.274e+01
005	6.799e+00	006	5.000e+00	007	3.300e+00	008	1.780e+00
009	1.061e+00	010	-1.422e-12	011	-6.938e-01	012	-1.141e+00
013	-1.549e+00	014	-2.534e+00	015	-4.177e+00	016	-6.205e+00
017	-7.842e+00	018	-9.242e+00	019	-1.269e+01		

File E:\Ade4\YANOMAMA\geogra_R_qe contains distance half-matrices

Rows = 171 Cols = 1

One half-matrix per column

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

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

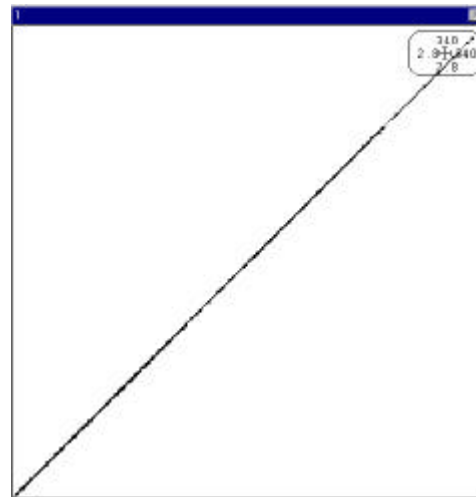
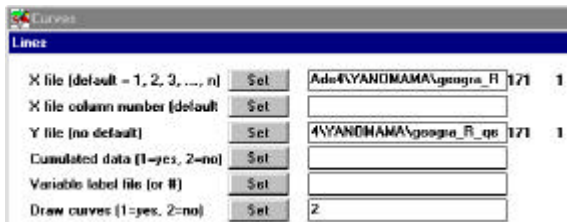
4 -> TRUE

File E:\Ade4\YANOMAMA\geogra_R_qe.eig contains eigenvalues

Rows = 19 Cols = 1

On fait la PCOA de cette matrice (diagonalisation de $-\frac{1}{2}\mathbf{D}^2 = \left[-\frac{1}{2}d_{ij}^2\right]$ doublement centrée) qui

donne deux valeurs propres 6684 et 3187 puis 17 valeurs propres comprises entre 1.37 et -1.27. Au départ on dispose de trois chiffres significatifs : les 17 valeurs propres sont théoriquement nulles et ne le sont pas uniquement par le format d'édition imposé. On reconstitue la matrice des distances en utilisant les deux premières valeurs propres. Les valeurs sont quasiment identiques :

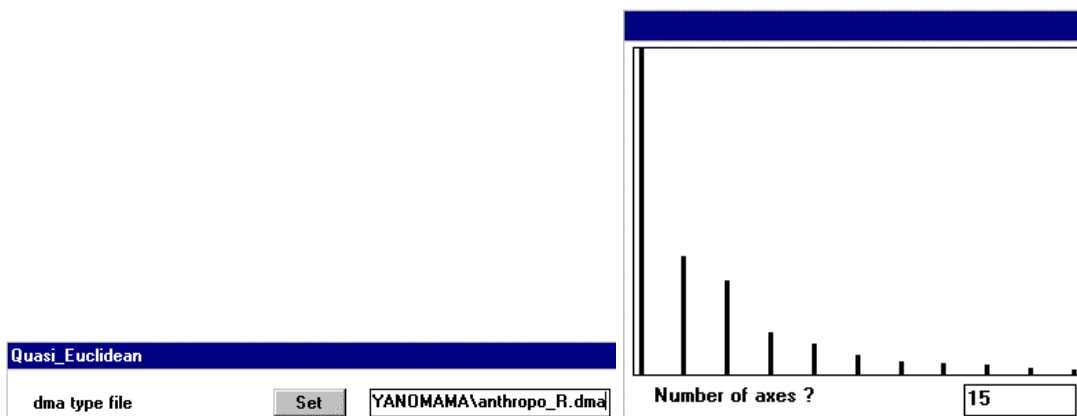


Important Il est clair que la distance d'origine est euclidienne et de dimension 2.

 Binary input file: E:\Ade4\YANOMAMA\geogra_R - 171 rows, 1 cols. à gauche
 Binary input file: E:\Ade4\YANOMAMA\geogra_R_qe - 171 rows, 1 cols. à droite

1	9.0000	1	9.0708
2	28.0000	2	28.3985
3	20.0000	3	19.6227
4	152.0000	4	151.9983
5	161.0000	5	160.8107
6	178.0000	6	177.9156
7	149.0000	7	149.3679
8	158.0000	8	158.0803
9	175.0000	9	174.7980
10	7.0000	10	7.1521
11	169.0000	11	169.2367
12	175.0000	12	175.1220
13	184.0000	13	183.9479
14	102.0000	14	102.1812
15	95.0000	15	95.2494
16	172.0000	16	171.8425
17	178.0000	17	177.9668
18	187.0000	18	187.3594
19	98.0000	19	97.9670
20	91.0000	20	91.1407
...		...	

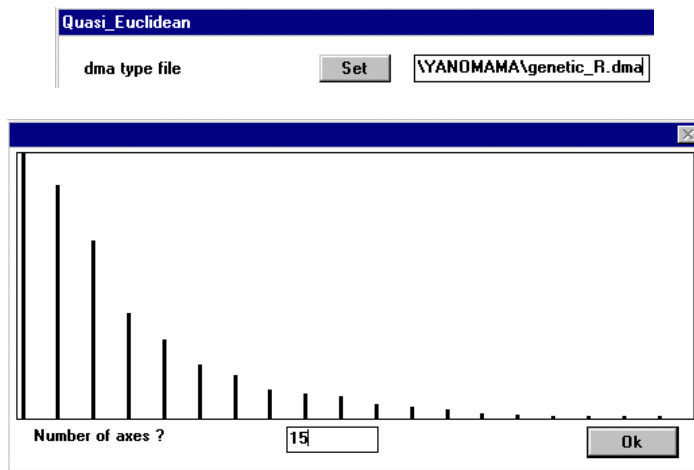
La distance anthropométrique (p. 124) est de même nature :



 Binary input file: E:\Ade4\YANOMAMA\anthropo_R - 171 rows, 1 cols.
 Binary input file: E:\Ade4\YANOMAMA\anthropo_R_qe - 171 rows, 1 cols.

1	96.0000	1	96.2543
2	147.0000	2	147.6266
3	88.0000	3	89.2215
4	295.0000	4	295.0168
5	318.0000	5	318.0458
6	309.0000	6	309.2054
7	284.0000	7	284.1272
8	339.0000	8	339.1909
9	348.0000	9	348.0422
10	158.0000	10	158.1457
11	253.0000	11	253.0680
12	258.0000	12	258.2718
13	260.0000	13	260.2217
14	235.0000	14	235.0905
15	253.0000	15	253.0599
16	289.0000	16	289.0730
17	301.0000	17	301.0679
18	321.0000	18	321.1179
19	277.0000	19	277.0242
20	270.0000	20	270.0523
...		...	

La distance génétique (p. 125) est aussi de même nature :



Binary input file: E:\Ade4\YANOMAMA\genetic_R - 171 rows, 1 cols.
 Binary input file: E:\Ade4\YANOMAMA\genetic_R_qe - 171 rows, 1 cols.

1	35.0000	1	35.1185
2	44.0000	2	43.9979
3	38.0000	3	38.0814
4	47.0000	4	47.0496
5	56.0000	5	56.0592
6	59.0000	6	59.0357
7	52.0000	7	52.0036
8	65.0000	8	65.0594
9	67.0000	9	67.0054
10	30.0000	10	30.1392
11	57.0000	11	57.0282
12	76.0000	12	76.0098
13	80.0000	13	80.0118
14	50.0000	14	50.0571
15	43.0000	15	43.0268
16	65.0000	16	65.0113
17	69.0000	17	69.0199
18	69.0000	18	69.0048
19	52.0000	19	52.0388
20	60.0000	20	60.0122
...		...	

Difficulté Cette procédure suppose que la matrice à approximer est euclidienne. Elle ne s'applique pas sans justification à une matrice quelconque.

DMAUtil : Read distance file

Type Utilitaire de lecture d'une matrice de distance.

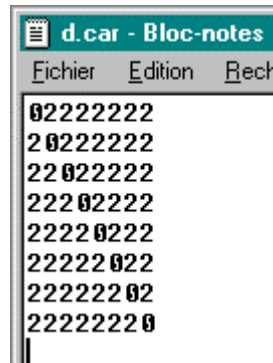
Objet Un fichier contenant un tableau est considéré comme une matrice de distance si la matrice est carrée, symétrique (la valeur de la ligne i et la colonne j est celle de la ligne j et la colonne i) et ne contient que des valeurs nulles sur la diagonale principale. L'option réécrit alors la matrice sur une colonne (demi-matrice inférieure), teste le caractère euclidien de la distance par le théorème de Gower³ et ajoute un fichier .dma.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Nom du fichier d'entrée binaire;

Exemple Implanter la matrice :



Passer en binaire par AdeTrans :

```
-----  
Binary input file: D:\Ade4\COLEOP\d - 8 rows, 8 cols.  
  1 | 0.0000  2.0000  2.0000  2.0000  2.0000  2.0000  2.0000  2.0000  
  ...  
  8 | 2.0000  2.0000  2.0000  2.0000  2.0000  2.0000  2.0000  0.0000
```



```
Input file: D:\Ade4\COLEOP\d  
D:\Ade4\COLEOP\d is a binary file with 8 rows and 8 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\COLEOP\d_R  
It has 28 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\COLEOP\d_R.dma  
  1 -> 8  
  2 -> 1
```

```
3 -> Input distance file D:\Ade4\COLEOP\d  
4 -> TRUE
```

Important La matrice de distances entre espèces du type qui précède est la plus simple qui soit et définit la diversité ordinaire. C'est elle qui est décomposée dans DMAUtil: From taxonomic tree.

DMAUtil : Read half distances matrix

Type Utilitaire de contrôle.

Objet Un fichier X contenant des matrices de distances sous forme de demi-matrices en colonnes est normalement accompagné d'un fichier X.dma contenant son descriptif. La présente option refait ce descriptif. Elle peut servir à documenter un fichier de ce type construit par ailleurs, par mélange, transformation, importation de données de ce type.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Nom du fichier d'entrée (binaire).

Important Le seul test porte sur le nombre de lignes du tableau d'entrée. Si n est ce nombre, il doit exister un entier p tel que $n = p(p-1)/2$ pour qu'on puisse considérer que le fichier d'entrée contient des demi-matrices inférieures.

Exemple Reprendre la fiche de DMAUtil: From fuzzy traits.

```
File D:\Ade4\COLEOP\EmF_f2 contains distance half-matrices
Rows = 5995 Cols = 2
One half-matrix per column
d21, d31, d32, d41, d42, d43, ..
Text file: D:\Ade4\COLEOP\EmF_f2.dma
1 -> 110
2 -> 2
3 -> Fuzzy traits from file D:\Ade4\COLEOP\EmF.fuz option 2
4 -> TRUE
```



```
Input file: D:\Ade4\COLEOP\EmF_f2
D:\Ade4\COLEOP\EmF_f2 is a binary file with 5995 rows and 2 columns
5995 is 110*(110-1)/2
Test of the euclidean property by diagonalization (theorem of GOWER)
matrix number 1 ---> TRUE
matrix number 2 ---> TRUE
Text file: D:\Ade4\COLEOP\EmF_f2.dma
1 -> 110
2 -> 2
3 -> Input half distance matrix file D:\Ade4\COLEOP\EmF_f2
4 -> TRUE
```

Le .dma a été refait et deux changements sont intervenus. Dans le premier cas le fichier est étiqueté comme provenant directement de D:\Ade4\COLEOP\EmF.fuz (option 2) alors que maintenant l'information mentionne le passage par la présente option. L'autre différence est moins visible. Dans le premier cas la mention TRUE (toutes les distances sont euclidiennes) a été inscrite directement car c'est une propriété mathématique générale. Dans le second cas, la même mention est obtenue par procédure numérique (théorème de Gower).

Exemple Voir aussi DMAUtil: Proportion data.

DMAUtil : ToClusters

Type Utilitaire d'interface entre modules.

Objet Le module permet de calculer des distances entre lignes ou colonnes des tableaux ou des triplets. Certaines option du module Clusters utilise des matrices de distances dans des fichiers du type --
-.dist. L'option crée ces fichiers en assurant la transformation de valeurs :

$$d_{ij} \mapsto \frac{d_{ij} - \min_{ij}(d_{ij})}{\max_{ij}(d_{ij}) - \min_{ij}(d_{ij})}$$

Les distances sont simplement ramenées dans l'intervalle [0,1] pour des questions numériques.

Dialogue L'option utilise une seule fenêtre de dialogue :



1) Nom de fichier du type dma (création dans ce module).

2) Numéro de la colonne utilisée (par défaut 1). Un fichier X.dma renvoie au fichier X qui contient des demi-matrices de distances. Quand il y a plusieurs colonnes dans le fichier X, chaque colonne contient une matrice de distances et on doit choisir ici laquelle est reconstituées sous forme de matrice ordinaire dans un fichier X.dist.

Lien Toute matrice de distances calculées par les options :

1 — DMAUtil : Binary Dissimilarity (distances basées sur une dissimilarité binaire, 10 options)

2 — DMAUtil : Quantitative variables (distances basées sur des variables quantitatives, 7 options)

3 — DMAUtil : Proportion data (distances basée sur des dissimilarités entre distributions de fréquences, 5 options)

4 — DMAUtil : Genetic distance (distances génétiques pour fréquence alléliques, 3 options)

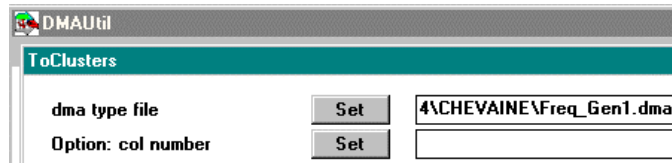
5 — DMAUtil : Triplet To Distance (distances euclidiennes associées à tout triplet statistique, sur les lignes et/ou les colonnes , x options)

6 — DMAUtil : Neighbourhood To Distance (distance basée sur un graphe de voisinage, 1 option)

7 — DMAUtil : Canonical distance (distance de la géométrie ordinaire, 1 option)

fournit une partition ou une hiérarchie et un dendrogramme en passant par la présente option.

Exemple Utiliser la carte Chevaîne et obtenir les distances décrites dans la fiche de DMAUtil : Genetic distance. Pour chacune des trois distances enchaîner :



```

Input file (distance matrix): D:\Ade4\CHEVAINE\Freq_Gen1.dma
Text file: D:\Ade4\CHEVAINE\Freq_Gen1.dma
1 -> 27
2 -> 1
3 -> ROGERS on D:\Ade4\CHEVAINE\Freq
4 -> TRUE

```

```

Output file : D:\Ade4\CHEVAINE\Freq_Gen11.dis Row: 27 Col: 27
Transformation: rescaling on [0,1] by  $y=(x-\min)/(\max-\min)$ 
-----

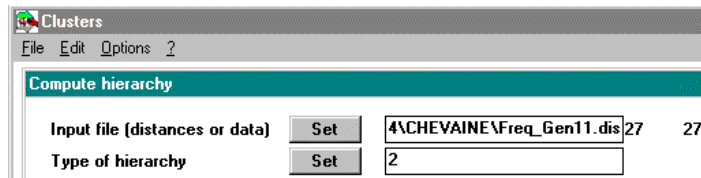
```

Freq_Gen1.dma signifie que le fichier de départ est Freq, les distances sont du type distances génétiques (Gen) avec l'option 1 (1). Le fichier de sortie Freq_Gen11.dis signifie qu'il contient une matrice de distance (dis) provenant du fichier Freq_Gen1 en utilisant sa première colonne (1). De même :

```

Output file : D:\Ade4\CHEVAINE\Freq_Gen21.dis Row: 27 Col: 27
Output file : D:\Ade4\CHEVAINE\Freq_Gen31.dis Row: 27 Col: 27

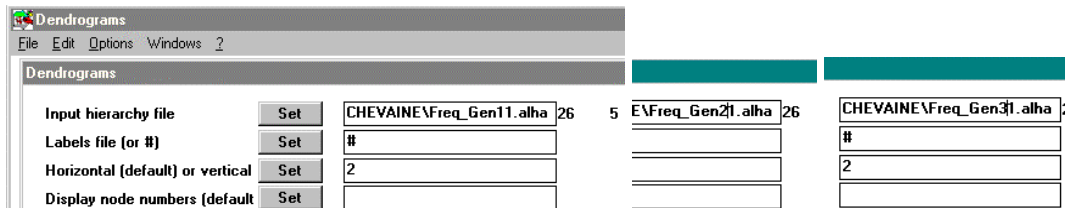
```



```

Clusters: Compute hierarchy
Distance file: D:\Ade4\CHEVAINE\Freq_Gen11.dis
Number of rows: 27, columns: 27
Output file: D:\Ade4\CHEVAINE\Freq_Gen11.alha
Number of rows: 26, columns: 5
Hierarchy algorithm used : average link (UPGMA)

```



Les trois points de vue sont voisins, le plus claire semblant être celui de la distance de Nei (au centre). Ceci prouve que le caractère non euclidien n'est pas un handicap pour la classification alors qu'il est un défaut majeur pour l'ordination.

DMAUtil : Triplet To Distance

Type Utilitaire de création de matrices de distances à partir de triplets statistiques quelconques.

Objet Un triplet statistique définit deux métriques diagonales qui donnent deux matrices de distances euclidiennes, respectivement entre les lignes et entre les colonnes du tableau transformé.

Dialogue L'option utilise une seule fenêtre de dialogue :

1) Fichier binaire du type `---.##ta`, `##` définissant le type de l'analyse. n est le nombre de lignes et p est le nombre de colonnes. Tous les types d'analyses de la première couche conviennent.

2) Nom du fichier de sortie. Par défaut, il est défini à partir du fichier d'entrée.

3) Par défaut le calcul porte sur les distances entre lignes et donne une matrice de sortie $n-n$. Taper 1 pour obtenir une matrice de distances entre colonnes : on obtient alors une matrice de sortie $p-p$.

Exemple Utiliser la carte Butterfly (^{20,10} p. 29). Faire l'ACP normée du tableau ButEnvir (PCA) :

Calculer la matrice de distances entre lignes :

```
Distance matrix computation from a statistical triplet
```

```
-----
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
```

Dupliquer le fichier normalisé (pour éliminer le point d'extension). La pondération des colonnes étant unitaire dans une ACP normée vérifier qu'on obtient le même résultat avec l'option :

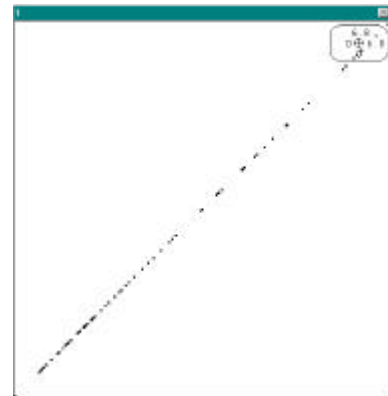
Creation of distance

Input file	Set	e4\BUTTERFL\ButEnvir_cnta16	4
Option: Output file	Set	Provi	
Option: default = between	Set		

Curves

Lines

X file (default = 1, 2, 3, ..., n)	Set	\BUTTERFL\ButEnvir_MDcn	120	1
X file column number (default	Set			
Y file (no default)	Set	:\Ade4\BUTTERFL\Provi_EU	120	1



Les fichiers Provi_EU (distances euclidiennes classiques sur tableau normalisé) et ButEnvir_MDcn (distances entre lignes associées au triplet de l'ACP normée) ont le même contenu.

Références

- ¹ Autem, M. & Bonhome, F. (1980) Eléments de systématique biochimique chez les Mugilidés (Poissons, Téléostéens) de Méditerranée. *Biochemical Systematics and Ecology* : 8, 305-308.
- ² Agnese, J.F. (1989) *Différenciation génétique de plusieurs espèces de Siluriformes ouest-africains ayant un intérêt pour la pêche et l'aquaculture*. Thèse de Doctorat, Université des Sciences et Techniques du Languedoc, Montpellier. 1-194.
- ³ Gower, J.C. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* : 3, 5-48.
- ⁴ Cailliez, F. (1983) The analytical solution of the additive constant problem. *Psychometrika* : 48, 305-310.
- ⁵ Drouet d'Aubigny, G. (1989) *L'analyse multidimensionnelle des données de dissimilarité*. Thèse de doctorat, Université Grenoble 1. 1-485.
- ⁶ Gower, J.C. (1982) Euclidean Distance Geometry. *Mathematical Scientist* : 7, 1-14.
- ⁷ Gower, J.C. (1984) Distance matrices and their approximation. In : *Data Analysis and Informatics III*. Diday, E. (Ed.) Elsevier, North-Holland. 3-21.
- ⁸ Legendre, L. & Legendre, P. (1984b) *Ecologie numérique*. Tome 2 - La structure des données écologiques. Masson, Paris. 2ème édition revue et augmentée : 1-344. Voir p. 6 et suivantes.
- ⁹ Digby, P. G. N. & Kempton, R. A. . (1987) *Multivariate Analysis of Ecological Communities*. Chapman and Hall, Population and Community Biology Series, London. 1-205. (p.96).
- ¹⁰ Manly, B.F. (1994) *Multivariate Statistical Methods. A primer*. Second edition. Chapman & Hall, London. 1-215.
- ¹¹ Agnese, J.F. (1989) *Différenciation génétique de plusieurs espèces de Siluriformes ouest-africains ayant un intérêt pour la pêche et l'aquaculture*. Thèse de Doctorat, Université des Sciences et Techniques du Languedoc, Montpellier. 1-194
- ¹² Pigliucci, M. & Barbujani, G. (1991) Geographical patterns of gene frequencies in Italian populations of *Ornithogalum montanum* (Liliaceae). *Genetical research*, Cambridge : 58, 95-104.
- ¹³ Guinand, B., Bouvet, Y. & Brohon, B. (1996) Spatial aspects of genetic differentiation of the European chub in the Rhone River basin. *Journal of Fish Biology* : 49, 714-726.
- ¹⁴ Rogers, J.S. (1972) Measures of genetic similarity and genetic distances. *Studies in Genetics*, Univ. Texas Publ. 7213: 145-153.
- ¹⁵ Avise, J.C. (1994) *Molecular markers, natural history and evolution*. Chapman & Hall, London. 1-511 (p. 95).
- ¹⁶ Nei, M. (1972) Genetic distances between populations. *American Naturalist* : 106. 283-292.

- ¹⁷ Edwards, A.W.F. (1971) Distance between populations on the basis of gene frequencies. *Biometrics* : 27, 873-881.
- ¹⁸ Hartl, D.L. & Clark, A.G. (1989) *Principles of population genetics*. Sinauer Associates, Sunderland, Massachusetts. 1-682 (p. 303).
- ¹⁹ Lingoes, J.C. (1971) Somme boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika* : 36, 195-203.
- ²⁰ Legendre, P. & Anderson, M.J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* : 69, 1-24.
- ²¹ Berge, C. (1967) *Théorie des graphes et ses applications*. Dunod, Paris. 1-269.
- ²² 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.
- ²³ Manly, B.F.J. (1991) *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London. 1-281.
- ²⁴ 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.