



# Procuste

Introduction-----	1
Procuste : PROTEST -----	2
Procuste : Tow arrays -----	9
Procuste : Tow distances -----	22
Références -----	25

# Introduction

Le module met à disposition les résultats de <sup>Peres-No01</sup> :

*The Mantel test provides a means to test the association between distance matrices and has been widely used in ecology and evolutionary studies. Recently, another permutation test based on a Procrustes statistic (PROTEST) was developed to compare multivariate sets. Our study contrasts the effectiveness, in terms of power and type I error, of Mantel test and PROTEST. We illustrate the application of Procrustes superimposition to visually examine the concordance of observations for each dimension separately and how to conduct hypothesis testing in which the association between two data sets is tested while controlling for the variation related to other sources of data. Our simulation results show that PROTEST is as powerful or more powerful than the Mantel test for detecting matrix association under a variety of possible scenarios. As a result of the increased power of PROTEST and the ability to assess the match for individual observations (not available with the Mantel test), **biologists now have an additional and powerful analytical tool to study ecological and evolutionary relationships.***

Pour éviter de prendre ce commentaire pour l'annonce d'une révolution, il faut préciser que la rotation Procuste est une des plus anciennes méthodes de couplage de tableau attribuée à <sup>Green---52</sup>, <sup>Schönemmn66</sup> ou <sup>Gower---71</sup>, suivant les sources celle de <sup>Rao-----80</sup> n'étant pas contestable. Elle est introduite en écologie dans <sup>Digby&Kp87</sup> et largement utilisée en morphométrie. PROTEST de <sup>Jackson-95</sup> est l'application très simple du principe de permutation des lignes d'un tableau par rapport à l'autre comme dans CoInertia: Coinertia test - Fixed D, Projectors: Subspace Test, Discrimin: Discriminant analysis: Test, Discrimin: Between analysis: Test. La statistique de PROTEST a des liens étroits avec celle de la co-inertie entre deux ACP mais reste réservée à cette option de couplage.

Procuste: Two arrays fait l'analyse procustéenne d'un couple de tableaux de variables quantitatives. On a choisit un mode de représentation des couples de points basé sur <sup>Torre&Cs94</sup> qui permet d'ajouter une représentation des variables des deux tableaux absente à notre connaissance.

Procuste: Two distances fait l'analyse procustéenne des représentations euclidiennes de deux matrices de distances euclidiennes. Elle permet l'utilisation des nombreuses distances disponibles dans DMAUtil, en perdant toutefois les possibilités d'interprétation des variables. On pourra utiliser DMAUse: Principal Coordinates et Procuste: Two arrays pour coupler, si besoin est, un tableau et une matrice de distances.

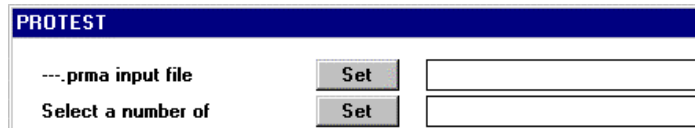
Procuste: PROTEST fait le test de <sup>Jackson-95</sup>. La fiche le replace dans la famille des tests sur les couplages de tableaux explorée par <sup>Kazi-Aol95</sup>.

# Procuste : PROTEST

**Type** Test de permutation.

**Objet** Le test porte sur la qualité d'une rotation procustéenne de <sup>Jackson-95</sup>.

**Dialogue** L'option utilise une seule fenêtre de dialogue :



1) Fichier .prma créé par l'exécution de Procuste : Two arrays ou Procuste : Two distances.

2) Nombre de permutations (100 par défaut mais utiliser plutôt 10000).

**Exemple** Continuer l'exemple de Procuste : Two distances. La situation est la suivante.  $\mathbf{D} = [d_{ij}]$  est la matrice des distances entre sites calculées sur les variables de milieu.  $\mathbf{\ddot{A}} = [\ddot{a}_{ij}]$  est la matrice des distances entre sites calculées sur la similarité des contenus faunistiques. La question est de mesurer l'accord qui existe entre ces deux matrices de distances. La plus ancienne mesure est le coefficient de congruence de Tucker <sup>Tucker--51</sup> cité dans <sup>Basalaj-01</sup>:

$$a_{\mathbf{D}\ddot{\mathbf{A}}} = \frac{\sum_{i=1}^n \sum_{j=1}^n d_{ij} \ddot{a}_{ij}}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n d_{ij}^2} \sqrt{\sum_{i=1}^n \sum_{j=1}^n \ddot{a}_{ij}^2}}$$

Il est compris entre 0 et 1 et n'est utilisé que comme statistique descriptive. Vient ensuite la statistique de Mantel (<sup>Mantel--67</sup>):

$$b_{\mathbf{D}\ddot{\mathbf{A}}} = \frac{\sum_{i < j} d_{ij0} \ddot{a}_{ij0}}{\sqrt{\sum_{i < j} d_{ij0}^2} \sqrt{\sum_{i < j} \ddot{a}_{ij0}^2}}$$

L'indice 0 indique le centrage des distances sur l'ensemble des couples  $i < j$ . C'est le coefficient de corrélation simple sur l'ensemble des valeurs des distances. Il est compris entre -1 et 1 et donne le test de Mantel (DMAUse: Mantel-RV test) si on utilise les permutations doubles (lignes et colonnes). On trouve une bonne introduction au principe de ce test dans <sup>Dietz---83</sup>. Par curiosité, on peut consulter l'article récent de <sup>Dutille&00</sup> qui compare la corrélation entre des valeurs  $x_i$  et des valeurs  $y_i$  d'une part et la corrélation entre les différences  $|x_i - x_j|$  et  $|y_i - y_j|$  !

Si les deux matrices de distance sont euclidiennes, elles définissent deux matrices de coordonnées  $\mathbf{X}_D$  et  $\mathbf{Y}_{\ddot{A}}$  qui génèrent une rotation procuste. On peut remarquer qu'on passe des distances à leur carré et que la mise à l'échelle préalable  $\mathbf{X}_D \rightarrow \mathbf{X}_D / \text{Trace}(\mathbf{X}_D' \mathbf{X}_D)$  est strictement équivalente au passage de  $\mathbf{D}^2 \rightarrow \mathbf{D}^2 / \mathbf{1}' \mathbf{D}^2 \mathbf{1}_n$  (ou le point désigne le produit

d'Hadamard terme à terme pour le distinguer du produit ordinaire ligne par colonne). La décomposition en valeurs singulières de  $\mathbf{Y}'_{\bar{A}}\mathbf{X}_D$  donne les valeurs singulières  $\mathbf{J}_1, \dots, \mathbf{J}_r$ . Leur somme définit la qualité de l'ajustement de la rotation procustéenne par :

$$\frac{1}{2} \|\mathbf{Y}_{\bar{A}} - \mathbf{X}_D \mathbf{R}^t\|^2 = \frac{1}{2} \sum_{i=1}^n \|\mathbf{y}_i - \text{Rot}(\mathbf{x}_i)\|^2 = 1 - \sum_{k=1}^r \mathbf{q}_k = 1 - m_{D\bar{A}}^2$$

Le test PROTEST porte sur  $m_{D\bar{A}}^2$  compris entre 0 et 1.

Les carrés des valeurs singulières sont les valeurs propres de la co-inertie de  $\mathbf{X}_D$  et  $\mathbf{Y}_{\bar{A}}$  après remise à l'échelle. On a :

$$\text{Trace}(\Gamma_0(\mathbf{D})\mathbf{\Delta}(\mathbf{D})) = \sum_{k=1}^r \mathbf{q}_k^2$$

où  $\Gamma_0(\mathbf{D})$  est défini par le double centrage par lignes et par colonnes (symbolisée par 00) :

$$\mathbf{D} = [d_{ij}] \Rightarrow \Gamma(\mathbf{D}) = -\frac{1}{2} [d_{ij}^2] \Rightarrow \Gamma_0(\mathbf{D}) = (\mathbf{I} - \mathbf{Q})\mathbf{\Delta}(\mathbf{D})(\mathbf{I} - \mathbf{Q}^t) = \frac{1}{2} [d_{ij00}^2]$$

ce qui conduit après normalisation à :

$$RV_{D\bar{A}} = \frac{\sum_{i<j}^n d_{ij00}^2 d_{ij00}^2}{\sqrt{\sum_{i<j}^n d_{ij00}^4} \sqrt{\sum_{i<j}^n d_{ij00}^4}}$$

Le RV est compris entre 0 et 1. Le test de Mantel, le test PROTEST et le test sur les RV sont donc distincts en entretenant des relations étroites. Le test de Mantel porte sur le coefficient de congruence entre les distances centrées et le test RV porte sur le coefficient de congruence entre les carrés des distances doublement centrées. Le test RV porte sur la somme des valeurs propres de co-inertie et le test PROTEST porte sur la somme des racines carrées de ces valeurs propres.

Mantel-RV test		
First dma type file	Set	y\Doubs\DouMil_MD0cn.dma
Matrix number	Set	
Second dma type file	Set	Try\Doubs\DouPoi_Sim1.dma
Matrix number	Set	
Permutation number	Set	10000

```
Matrix 1 of D:\Ade4\Dir_Try\Doubs\DouMil_MD0cn versus
matrix 1 of D:\Ade4\Dir_Try\Doubs\DouPoi_Sim1
r index : 4.759e-01
number of random matching: 10000   Observed: 0.476
Histogramm: minimum = -0.257616, maximum = 0.476
number of simulation X<Obs: 10000 (frequency: 1.000000)
number of simulation X>=Obs: 0 (frequency: 0.000000)
```



$$IBA = \frac{\text{Trace}(\mathbf{S}_{12}\mathbf{S}_{21})}{\text{Trace}(\mathbf{S}_{11})\text{Trace}(\mathbf{S}_{22})}$$

Car seul le numérateur varie aléatoirement.

Le test RV est repris dans **Heo&Gabe98**. Le test PROTEST est repris dans **Peres-No01**.

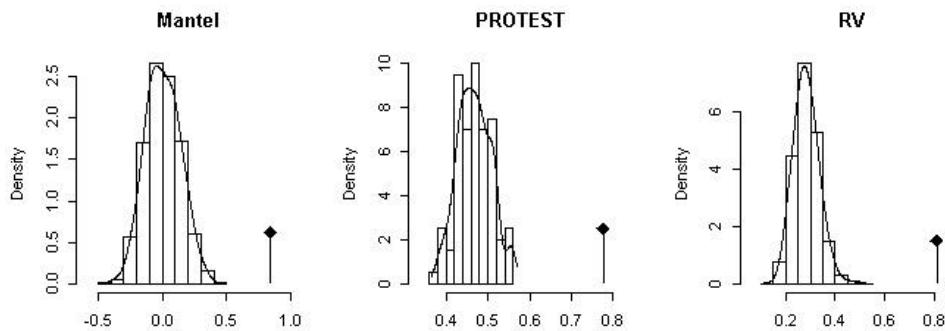
**Important** On peut se demander si toutes ces nuances sont d'une importance réelle. Est-il possible de trouver des cas où ces tests donneraient des résultats contradictoires. Ils auraient alors des fonctions différentes. Des simulations ont été faite sous R.

```

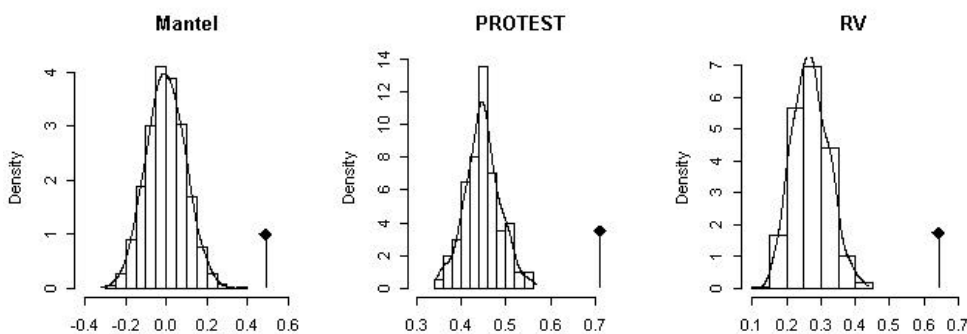
essai_function() {
  X_rmvnorm(20,m=rep(0,5))
  dx_mattod2(X)
  dy_matrix(abs(as.vector(dx)+rnorm(400,sd=1)),ncol=20)
  dy_dy+t(dy)
  for (i in 1:20) dy[i,i]_0
  Y_pcored(dy)
  return(list(A=X,B=Y))
}

```

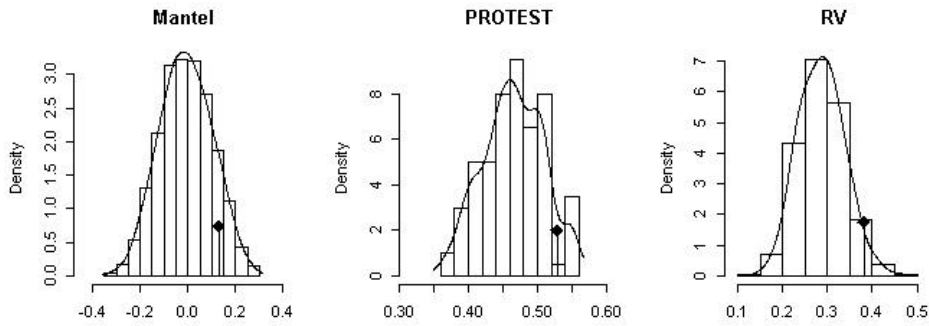
On part d'un nuage de 20 points échantillon d'une loi normale sphérique (5 dimensions) sans corrélation. On calcule la matrice des distances (métrique canonique), on introduit sur ces distances une erreur normale de variance 1 et on fait la PCO du résultat. On obtient un deuxième tableau lié au premier. Les tests de simulation donnent :



sd=1



sd=2

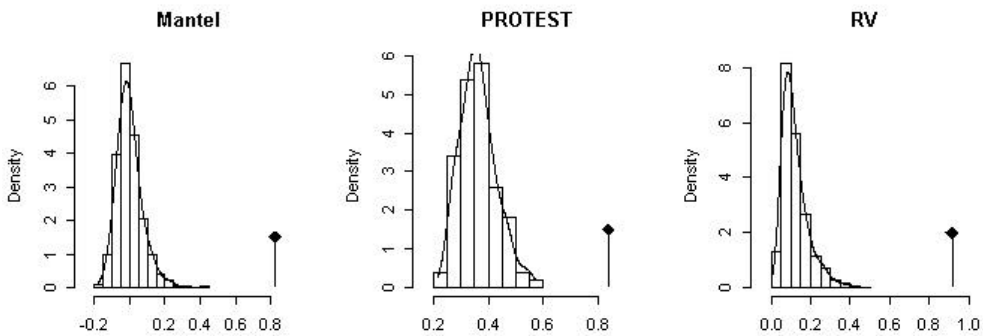


sd=5

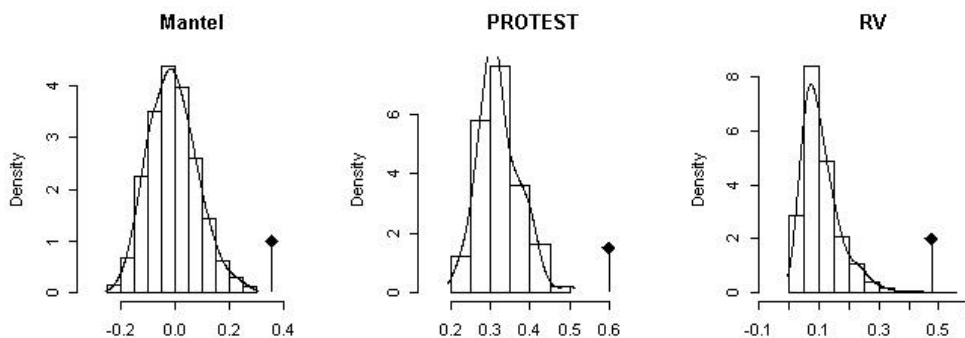
*Systématiquement quand on simule des distances corrélées, les trois tests sont de même signification.*

```
essai2_function() {
  w1_seq(0,2*pi,le=21)[1:20]
  x_3*cos(w1)
  y_sin(w1)
  xrand_x+rnorm(20,sd=4)
  yrand_y+rnorm(20,sd=4)
  X_cbind.data.frame(x,y,matrix(rnorm(60),20,3))
  Y_cbind.data.frame(xrand,yrand,matrix(rnorm(60),20,3))
  return(list(A=X,B=Y))
}
```

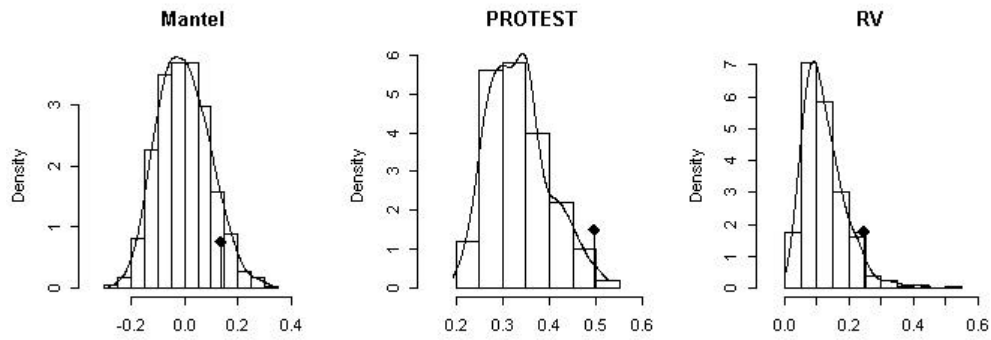
On place 20 points sur une ellipse. On reconstruit cette figure avec une petite erreur. On rajoute 3 composantes aléatoires au deux tableaux.



sd=0.25



sd=2



sd=4

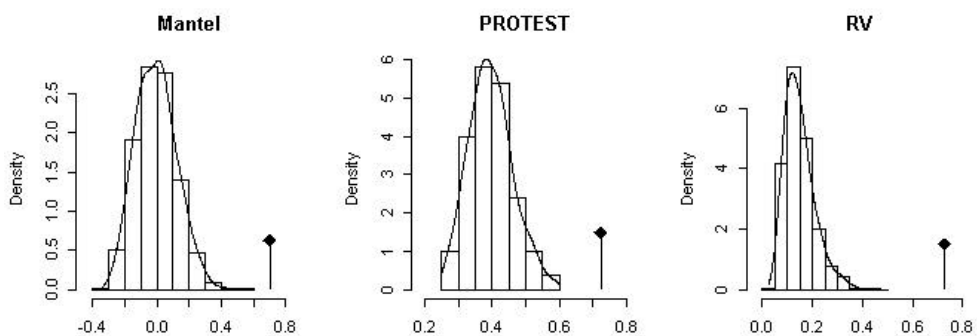
Les trois test ont des comportements voisins, le premier à perdre sa signification étant le test de Mantel.

```

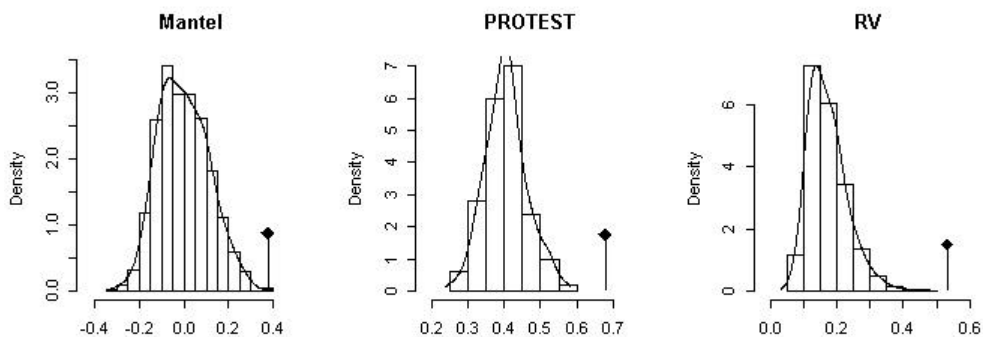
essai3_fonction(g) {
  a1_matrix(c(1,0.7,0.7,0.7,1,0.7,0.7,0.7,1),3,3)
  a2_matrix(0,3,3)
  a3_cbind(a1,a2)
  a4_cbind(a2,a1)
  cor_rbind(a3,a4)
  X_rmvnorm(20,sigma=cor)
  Y_X+rmvnorm(20,sigma=diag(g,6))
  return(list(A=X,B=Y))
}

```

On génère un échantillon d'une variable normale de dimension 6 à deux paquets de composantes corrélées. On ajoute au résultat une erreur sphérique de variance g.

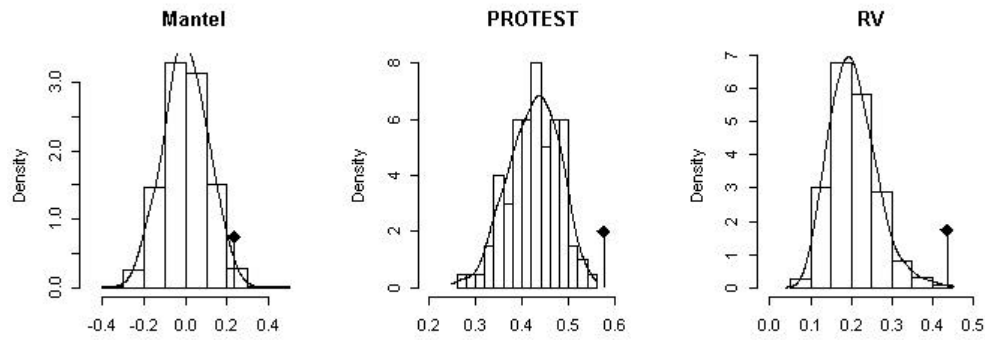


g=1



g=2





g=3

*Les trois tests ont encore un comportement voisin, le premier à perdre sa signification étant le test de Mantel.*

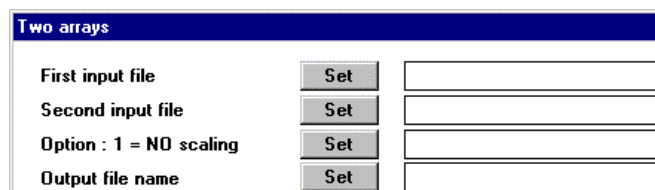
PROTEST est un test de permutation pour le couplage de deux tableaux. Dans **Kazi-Aol95** on en trouvera quatre autres avec des solutions analytiques. Il est sans doute légèrement plus puissant que le test de Mantel, mais dans les problèmes réels cela n'a guère d'importance car si on voisine le seuil de 5 % avec un couple de tableaux écologiques, il y a beaucoup de soucis en perspective. Le test de permutation sur l'analyse canonique des correspondances **Ter\_Bra87c** (Projectors: Subspace Test dans ADE-4) est disponible depuis longtemps et le test sur la co-inertie **Dolédech94** (CoInertia: Coinertia test - Fixed D) est souvent utilisé. Ces différentes approches donneront en général des résultats voisins mais les comparaisons sont intéressantes.

# Procuste : Tow arrays

**Type** Analyse multivariée d'un couple de tableaux.

**Objet** Un tableau  $X$  a  $n$  lignes et  $p$  colonnes. Typiquement c'est un tableau de variables environnementales mesurées sur  $n$  sites. Un tableau  $Y$  a  $n$  lignes et  $q$  colonnes. Typiquement c'est un tableau de variables floro-faunistiques mesurées sur les mêmes  $n$  sites. L'analyse procustéenne étudie le couple. Pour l'essentiel, l'option assure la rotation procustéenne dite *Procustes Rotation* présentée dans **Digby&Kp87** Chapitre 4 *Methods for comparing ordinations*, § 4.1 *Procustes rotation*.

**Dialogue** L'option utilise une seule fenêtre de dialogue :



- 1) Premier tableau dans un fichier binaire.
- 2) Second tableau dans un fichier binaire (il doit avoir le même nombre de lignes que le premier).
- 3) Option de mise à l'échelle (voir ci-dessous). Taper 1 pour ne pas avoir la mise en échelle qui est l'option par défaut.
- 4) Nom générique des fichiers de sortie (chaîne de caractères dans point d'extension).

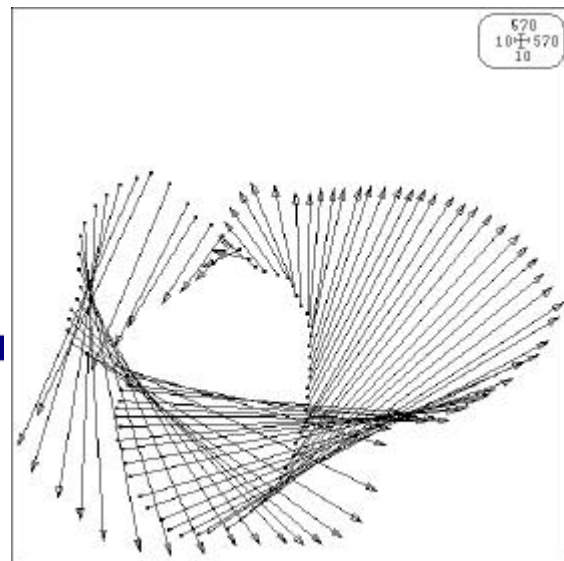
**Exemple** Pour comprendre le principe de la rotation procustéenne, utiliser la carte Macaca Olshan&182



Le fichier xy1 (à gauche) et le fichier xy2 (à droite) contiennent 72 points définis par deux coordonnées dans un plan. C'est le même individu à deux âges différents. Le dessin de la tête a été redéfini en coordonnées polaire par rapport à un point fixe (il y a 72 points associés à 360° divisés en unités de 5°). Pour cet exercice, on a ignoré le point de référence, tourner le petit d'un angle de 90° et digitaliser la série des 72 points (Digit: Digitize). L'objectif est de replacer une figure sur l'autre. Ce problème fondamental en morphométrie est l'objet d'une intense réflexion méthodologique (nombreuses citations et discussions dans l'article cité). On ne s'en sert ici que comme illustration, sans prétendre

participer au débat. Il est idiot de superposer brutalement ces deux figures sans repère commun. Faisons le (Scatters: Match two scatters) :

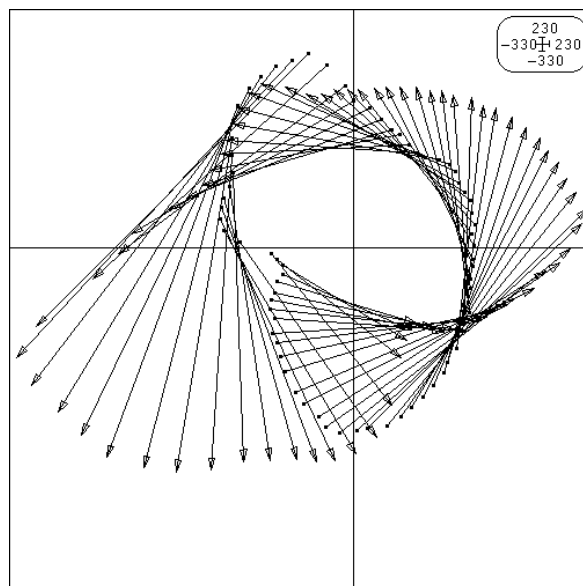
Match two scatters			
XY coordinates file	Set	:\Ade4\Dir_Try\Macaca\xy1	72 2
X-axis column number (default)	Set		
Y-axis column number (default)	Set		
Second XY coordinates file	Set	:\Ade4\Dir_Try\Macaca\xy2	72 2
Label file (or #) for items	Set		
Label file (or #) for groups	Set		



Il faut évidemment faire tourner le petit pour le recaler sur le grand (ou le grand pour le recaler sur le petit). C'est l'objectif de la rotation procuste (du nom d'une terreur de l'antiquité qui périt de la torture qu'il infligeait à ses victimes). Si  $\mathbf{X}$  contient  $n$  points en lignes sur  $p$  coordonnées en colonnes et si  $\mathbf{Y}$  contient  $n$  points en lignes sur  $p$  coordonnées en colonnes, l'accord entre les deux nuages se mesure par :

$$d^2(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|^2 = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - y_{ij})^2$$

On améliore immédiatement cette mesure en donnant aux deux nuages le même centre de gravité. Centrer les deux tableaux (Bin-Bin: Centring) et recommencer la superposition (Scatters: Match two scatters) :



Les deux nuages de points sont dans le même espace.  $\mathbf{X}$  et  $\mathbf{Y}$  sont maintenant centrés.

**Théorie** Supposons que  $\mathbf{X}$  soit la cible (le nuage en place). On veut appliquer à  $\mathbf{Y}$  une rotation  $\mathbf{R}$  pour optimiser l'ajustement, donc minimiser :

$$d^2(\mathbf{X}, \hat{\mathbf{Y}}) = \|\mathbf{X} - \mathbf{Y}\mathbf{R}^t\|^2 = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \hat{y}_{ij})^2$$

$\mathbf{R}$  est une matrice  $p \times p$  qui conserve les angles et les distances, donc les produits scalaires et vérifie  $\mathbf{R}^t\mathbf{R} = \mathbf{I}_p = \mathbf{R}\mathbf{R}^t$ .

La mesure se réécrit (noter que  $\text{Trace}(\mathbf{X}'\mathbf{X}) = \text{Trace}(\mathbf{X}\mathbf{X}')$ ):

$$d^2(\mathbf{X}, \mathbf{Y}) = \text{Trace}((\mathbf{X} - \mathbf{Y})'(\mathbf{X} - \mathbf{Y})) = \text{Trace}(\mathbf{X}'\mathbf{X}) + \text{Trace}(\mathbf{Y}'\mathbf{Y}) - 2\text{Trace}\mathbf{Y}'\mathbf{X}$$

d'où :

$$d^2(\mathbf{X}, \hat{\mathbf{Y}}) = \|\mathbf{X} - \mathbf{Y}\mathbf{R}^t\|^2 = \text{Trace}(\mathbf{X}'\mathbf{X}) + \text{Trace}(\mathbf{Y}'\mathbf{Y}) - 2\text{Trace}\mathbf{R}'\mathbf{Y}'\mathbf{X}$$

On cherche donc  $\mathbf{R}$  qui minimise :

$$\|\mathbf{Y} - \mathbf{R}\mathbf{X}'\|^2 = Cte - 2\text{Trace}(\mathbf{R}\mathbf{Y}'\mathbf{X})$$

La décomposition en valeurs singulières de  $\mathbf{Y}'\mathbf{X}$  s'écrit :

$$\mathbf{Y}'\mathbf{X} = \mathbf{V}\mathbf{\Theta}\mathbf{U}'$$

On cherche donc  $\mathbf{R}$  qui maximise :

$$\text{Trace}(\mathbf{R}\mathbf{V}\mathbf{\Theta}\mathbf{U}') = \sum_{k=1}^n \mathbf{q}_k(\mathbf{u}_k | \mathbf{R}\mathbf{v}_k) \leq \sum_{k=1}^n \mathbf{q}_k$$

La borne est atteinte pour  $\mathbf{R} = \mathbf{U}\mathbf{V}'$ . Cette isométrie envoie le nuage des lignes de  $\mathbf{Y}$  au plus près du nuage des lignes de  $\mathbf{X}$ .

### Retour à l'exemple

Two arrays			
First input file	Set	:\Ade4\Dir_Try\Macaca\xy1	72 2
Second input file	Set	:\Ade4\Dir_Try\Macaca\xy2	72 2
Option : 1 = NO scaling	Set	1	
Output file name	Set	a	

Comme le problème est une question de croissance, il n'y pas de mise à l'échelle des tableaux. Conserver les deux axes.

```
Procuste analysis
X : First input array D:\Ade4\Dir_Try\Macaca\xy1
Number of rows: 72, columns: 2
Y : Second input array D:\Ade4\Dir_Try\Macaca\xy2
Number of rows: 72, columns: 2
```

Le tableau X est xy1. Le tableau Y est xy2. Ils ont mêmes dimensions (72-2).

```
File a.prX contains X after centring
--- It has 72 rows and 2 columns
File a.prY contains Y after centring
--- It has 72 rows and 2 columns
```

Les deux tableaux sont simplement centrés comme dans l'essai préliminaire.

```
File a.prrotX contains table X after rotation to Y
```

--- It has 72 rows and 2 columns

File :a.pprotX

Col.	Mini	Maxi
1	-1.859e+02	1.775e+02
2	-1.277e+02	1.163e+02

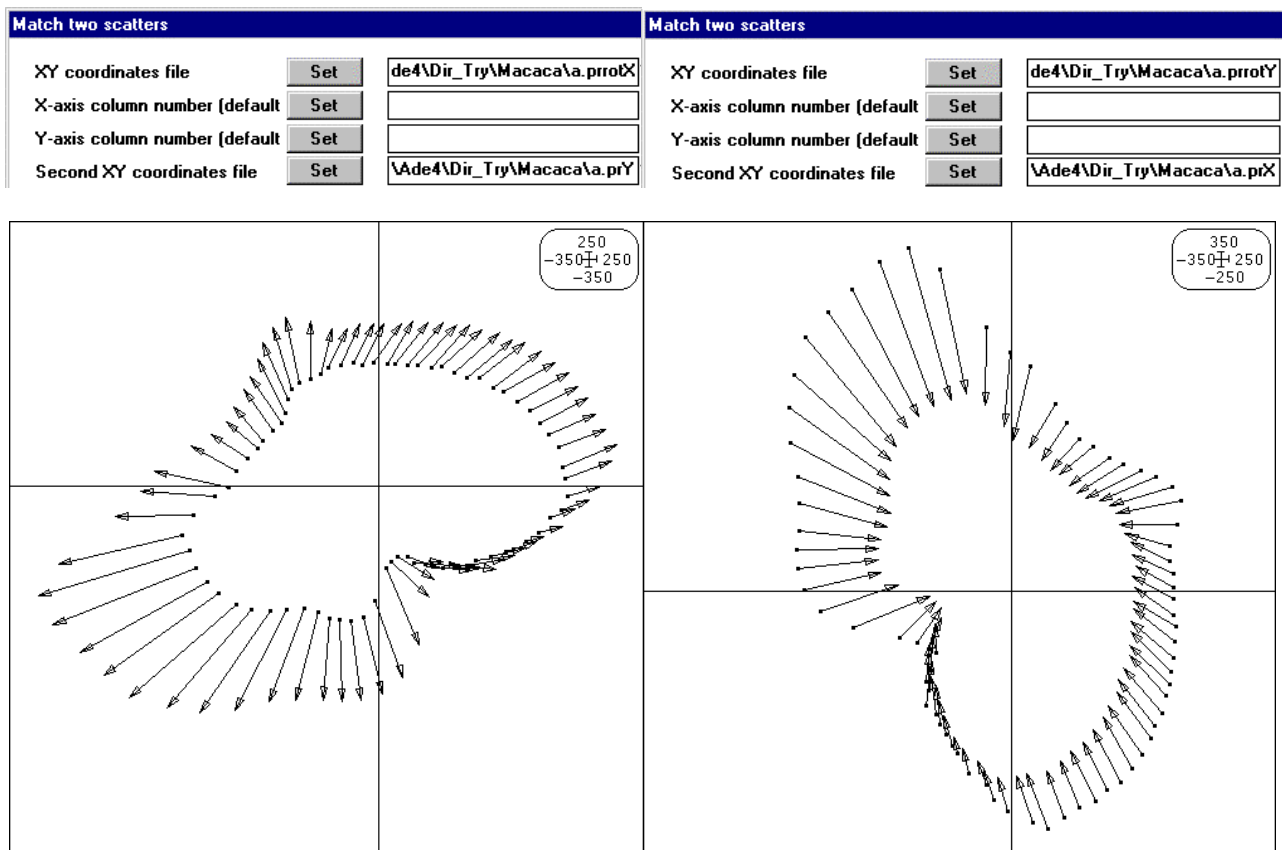
File a.pprotY contains table Y after rotation to X

--- It has 72 rows and 2 columns

File :a.pprotY

Col.	Mini	Maxi
1	-2.114e+02	1.601e+02
2	-2.252e+02	3.243e+02

Ce résultat n'est accessible que si les dimensions des deux tableaux sont identiques. On peut donc représenter X et le résultat de la rotation sur Y ou Y et le résultat de la rotation sur X :



Cette méthode a été introduite par Gower<sup>71</sup>. Une solution analytique avait été introduite par Sneath<sup>67</sup>. Ces sources sont citées dans Rohlf&S<sup>90</sup> qui décrivent cette procédure (p.42) sous le terme "Orthogonal Procrustes Analysis" avec le scaling (voir ci-après).

Si on quitte la morphométrie pour l'écologie, deux problèmes se posent immédiatement.

1) Le premier touche les échelles, le second les dimensions. Pour les échelles, la situation est assez simple. La rotation procuste ne déforme pas le nuage qui est ajusté sur la cible. Les deux dessins sont strictement équivalents car aucune déformation n'est en jeu. Ajuster par rotation X à Y ou ajuster par rotation Y à X ne change rien aux objets mais ajuste la représentation de l'un à l'autre. Si les deux objets ne sont pas à la même échelle expérimentale, on sent la nécessité de renforcer l'adéquation en multipliant un tableau par une constante, ce qui ne change pas sa forme ou à multiplier certaines composantes par une

constante, ce qui introduit des dilatations et un changement de formes. Il y a alors deux voies. L'une est foncièrement dissymétrique en disant qu'un tableau est la cible et qu'il définit les modifications nécessaires pour ajuster l'autre. Ces propositions, issues des stratégies d'analyse factorielle, sont initiées par Schönemmn70 qui cite déjà une *standard orthogonal Procrustes subroutine* (Green---52, Cliff---66, Schönemmn66). Le terme "Procrustes" est attribué par Schönemmn68 à Hurley&t62. L'autre conserve une certaine symétrie au problème en mettant à une échelle commune, avant toute autre pratique, les deux tableaux.

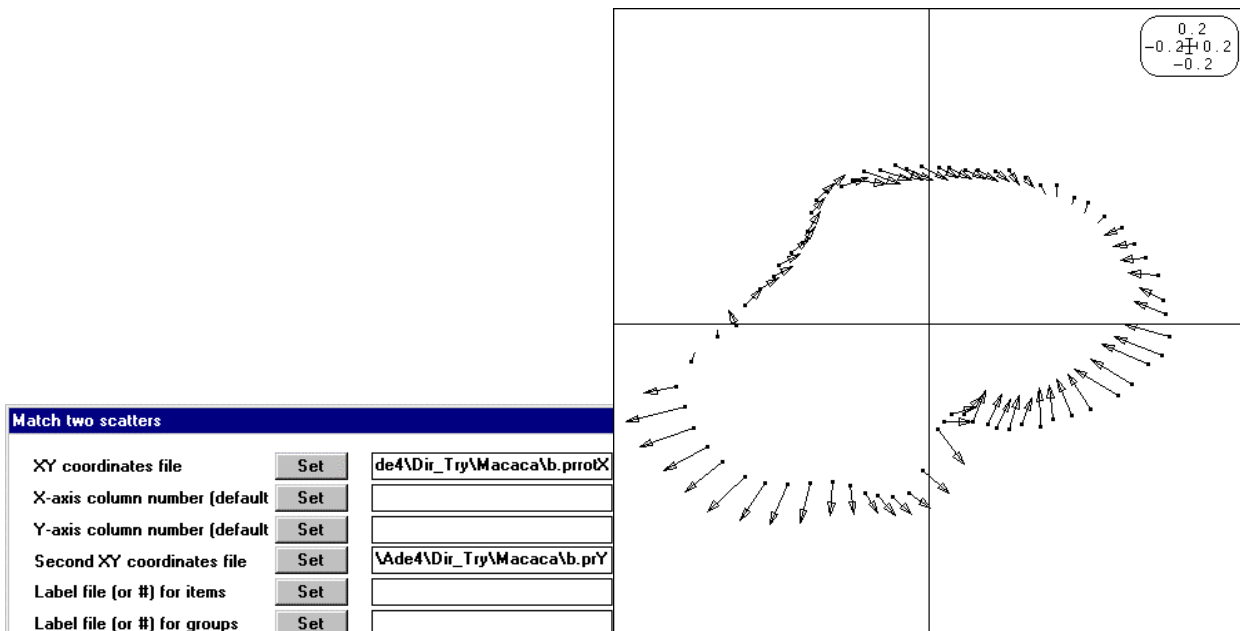
C'est le scaling de Gower (Gower---71) repris par Rohlf&Sc90. La variabilité totale des mesures est ( $\mathbf{X}$  est le tableau centré) :

$$T(\mathbf{X}) = \sum_{i=1}^n \sum_{j=1}^p x_{ij}^2 = n \times \text{Iner}(\mathbf{X}) = \text{Trace}(\mathbf{X}\mathbf{X}^t) = \text{Trace}(\mathbf{X}^t\mathbf{X})$$

Les tableaux mis à une échelle commune sont alors  $\frac{\mathbf{X}}{\sqrt{T(\mathbf{X})}}$  et  $\frac{\mathbf{Y}}{\sqrt{T(\mathbf{Y})}}$ . Pour éviter de multiplier les notations, on appellera dorénavant  $\mathbf{X}$  et  $\mathbf{Y}$  les tableaux centrés et mis à l'échelle, donc vérifiant  $T(\mathbf{X}) = T(\mathbf{Y}) = 1$ . C'est le cas normal d'utilisation de ce module.

Two arrays			
First input file	Set	:\Ade4\Dir_Try\Macaca\xy1	72 2
Second input file	Set	:\Ade4\Dir_Try\Macaca\xy2	72 2
Option : 1 = NO scaling	Set		
Output file name	Set	b	

```
File b.prX contains X/sqrt(trace1) after centring
trace1 = Trace (XtX) = 1.35E+06
--- It has 72 rows and 2 columns
File b.prY contains Y/sqrt(trace2) after centring
trace2 = trace (YtY) = 2.834E+06
--- It has 72 rows and 2 columns
```



La mise à l'échelle modifie sensiblement le résultat et illustre les changements de forme.

En morphométrie, le rescaling ne s'impose pas car les changements de taille ont un sens. Mais il peut être utilisé. Par exemple dans Klingenr98 les auteurs étudient l'asymétrie des

ails de mouches tsetse en superposant par rotation procuste une aile sur le symétrique de l'autre après *scaling to unit centroid size* pour se concentrer sur les variations de forme. En écologie, ajuster deux modèles sans mise à l'échelle n'aurait pas de sens.

**2) Les dimensions.** Le second problème est de loin le plus important. Ajuster deux objets à deux dimensions n'a pas besoin de plus de commentaires. L'objet de référence et celui qui lui est ajusté se voient intégralement dans le plan. Par contre, si le premier est dans  $\mathbb{R}^p$  et le second est dans  $\mathbb{R}^q$  avec  $p$  et  $q$  grands et différents, il faut une méthode pour voir le résultat, d'une part, la question des rotations est plus difficile, d'autre part. A priori les deux nuages de points ne sont pas dans le même espace. On peut, d'un point de vue formel, compléter le plus petit des tableaux par des colonnes de valeurs nulles pour lui donner la dimension du plus grand. La décomposition en valeur singulière définit deux bases orthonormées du même espace de dimension  $\max(p, q)$  :

$$\mathbf{Y}'\mathbf{X} = \mathbf{V}\Theta\mathbf{U}'$$

$\mathbf{X}_{rot} = \mathbf{X}\mathbf{U}\mathbf{V}'$  contient le nuage issu de  $\mathbf{X}$  qui s'ajuste à  $\mathbf{Y}$  et  $\mathbf{Y}_{rot} = \mathbf{Y}\mathbf{V}\mathbf{U}'$  contient le nuage issu de  $\mathbf{Y}$  qui s'ajuste à  $\mathbf{X}$ . La qualité de l'ajustement s'écrit :

$$\|\mathbf{Y} - \mathbf{R}\mathbf{X}'\|^2 = 2 - 2\text{Trace}(\mathbf{R}\mathbf{Y}'\mathbf{X}) = 2 - 2\sum_{k=1}^r \mathbf{q}_k$$

où  $r$  est le rang de  $\mathbf{Y}'\mathbf{X}$ . Ajouter des zéros dans un tableau ne change rien puisqu'on ajoute des valeurs singulières nulles. Pour voir l'ajustement, il faut donc projeter sur un plan les lignes de  $\mathbf{X}$  et de  $\mathbf{Y}_{rot}$  ou celles de  $\mathbf{Y}$  et de  $\mathbf{X}_{rot}$ . On obtient des approximations procustéennes. Il y a de nombreuses possibilités, en particulier les deux ACP des tableaux accolés et les deux ACP des tableaux moyens (e.g. <sup>Mouttet-81</sup>) :

$$\begin{bmatrix} \mathbf{X}_{rot} \\ \mathbf{Y} \end{bmatrix} \quad \begin{bmatrix} \mathbf{X} \\ \mathbf{Y}_{rot} \end{bmatrix} \quad \frac{1}{2}(\mathbf{X}_{rot} + \mathbf{Y}) \quad \frac{1}{2}(\mathbf{X} + \mathbf{Y}_{rot})$$

La question est clairement explicitée dans l'article fondamental de <sup>Sibson--78</sup> :

*It might be thought that (by analogy with the centroid matching used to fit X and it might be thought that (by analogy with the centroid matching used to fit X and optimally under translation) rotation/reflexion fit would have something to do with matching principal axes. This is simply not the case; all that can be said is that if X and Y can be well matched, and if the principal variances are well distinguished, then the principal axes will themselves correspond reasonably closely after fitting under rotation/reflexion (p. 237).*

Toutes ces analyses donnent des représentations voisines mais différentes. Or les rotations ne déformant pas les nuages, les configurations de  $2n$  points (les  $n$  d'un tableau et les  $n$  points de l'autre après rotation) sont strictement les mêmes. La représentation en dimensions réduites devrait être unique. Pour résoudre ce problème, l'option utilise le résultat de l'analyse de co-inertie de deux tableaux totalement appariés (<sup>Torre&Cs94</sup>).

**Théorie** Considérons deux tableaux  $\mathbf{A}$  et  $\mathbf{B}$  portant sur les mêmes  $n$  lignes et les mêmes  $s$  colonnes. Ils donnent deux nuages de  $n$  points dans  $\mathbb{R}^s$ . Si on cherche un axe d'ACP  $\mathbf{z}$  commun à ces deux tableaux, il doit avoir une propriété d'axe principal pour  $\mathbf{A}$  ( $\text{Var}(\mathbf{A}\mathbf{z})$  grand), une propriété d'axe principal pour  $\mathbf{B}$  ( $\text{Var}(\mathbf{B}\mathbf{z})$  grand), et une propriété de cohérence entre les

deux systèmes de coordonnées, donc une propriété d'analyse canonique ( $corr^2(\mathbf{Az}, \mathbf{Bz})$  grand). On cherche à maximiser :

$$f(\mathbf{z}) = \langle \mathbf{Az} | \mathbf{Bz} \rangle = Cte \times \sqrt{Var(\mathbf{Az})} \times \sqrt{Var(\mathbf{Bz})} \times corr(\mathbf{Az}, \mathbf{Bz})$$

Le maximum est atteint pour le premier vecteur propre de la matrice :

$$\mathbf{W} = \frac{1}{2}(\mathbf{A}'\mathbf{B} + \mathbf{B}'\mathbf{A})$$

Les axes successifs de co-inertie sont les vecteurs propres normés de  $\mathbf{W}$ . Pour le couple  $\mathbf{X}_{rot} = \mathbf{XUV}'$  et  $\mathbf{Y}$  on a :

$$\begin{aligned} \mathbf{W} &= \frac{1}{2}(\mathbf{A}'\mathbf{B} + \mathbf{B}'\mathbf{A}) = \frac{1}{2}(\mathbf{VU}'\mathbf{X}'\mathbf{Y} + \mathbf{Y}'\mathbf{XUV}') \\ &= \frac{1}{2}(\mathbf{VU}'\mathbf{U}\Theta\mathbf{V}' + \mathbf{V}\Theta\mathbf{U}'\mathbf{U}\mathbf{V}') = \mathbf{V}\Theta\mathbf{V}' \end{aligned}$$

Ils s'en suit immédiatement que la co-inertie entre  $\mathbf{X}_{rot} = \mathbf{XUV}'$  et  $\mathbf{Y}$ , d'une part, et entre  $\mathbf{Y}_{rot} = \mathbf{YVU}'$  et  $\mathbf{X}$  d'autre part, conduisent à la même représentation et à la même qualité de représentation. Les deux systèmes de coordonnées sont  $\mathbf{XU}$  et  $\mathbf{YV}$  et les critères associés sont les valeurs singulières de  $\mathbf{Y}'\mathbf{X}$  (ou  $\mathbf{X}'\mathbf{Y}$ ) c'est-à-dire les racines des valeurs propres de  $\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}$  (ou  $\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'$  ou  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$  ou  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'$ ).

Les fichiers LX et LY contiennent les matrices  $\mathbf{XU}$  et  $\mathbf{YV}$ .

```
File b.prLX contains row scores (data set X)
--- It has 72 rows and 2 columns
File :b.prLX
```

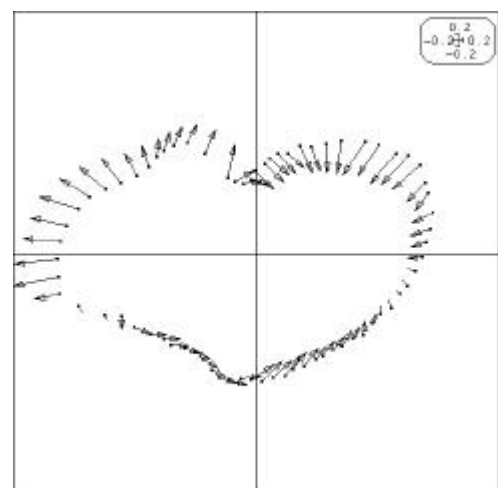
Col.	Mini	Maxi
1	-1.633e-01	1.452e-01
2	-1.046e-01	9.308e-02

```
File b.prLY contains row scores (data set Y)
--- It has 72 rows and 2 columns
File :b.prLY
```

Col.	Mini	Maxi
1	-1.997e-01	1.303e-01
2	-1.076e-01	1.066e-01

**Match two scatters**

XY coordinates file	Set	Ade4\Dir_Try\Macaca\A.prX	72	2
X-axis column number [default]	Set			
Y-axis column number [default]	Set			
Second XY coordinates file	Set	Ade4\Dir_Try\Macaca\A.prY	72	2
Label file (or #) for items	Set			
Label file (or #) for groups	Set			







On édite les valeurs propres de de  $\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y}$  (ou  $\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{X}'$  ou  $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$  ou  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'$ ). Faire l'acp normée (PCA: Correlation matrix PCA) de DouMil, l'ACP centrée (PCA: Covariance matrix PCA) de DouPoi et l'analyse de co-inertie des deux triplets (CoInertia: Matching two statistical triplets avec le nom générique coiner et CoInertia: Coinertia analysis sur coiner.iita en conservant deux facteurs). On trouve :

```
-----
Total inertia: 134.703
-----
```

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.1902E+02	+0.8836	+0.8836	02	+1.3871E+01	+0.1030	+0.9865
03	+7.5658E-01	+0.0056	+0.9922	04	+5.2783E-01	+0.0039	+0.9961
05	+2.7089E-01	+0.0020	+0.9981	06	+1.6462E-01	+0.0012	+0.9993
07	+6.6011E-02	+0.0005	+0.9998	08	+1.8209E-02	+0.0001	+0.9999
09	+4.6041E-03	+0.0000	+1.0000	10	+2.7934E-03	+0.0000	+1.0000
11	+0.0000E+00	+0.0000	+1.0000				

A un facteur multiplicatif près ces valeurs propres sont les mêmes. Ce facteur vient de la pondération qu'on utilise dans la co-inertie ( $1/n$ ) et de la mise à l'échelle qu'on utilise avant la rotation procuste. Mais dans la rotation procuste, ce sont les racines carrées de ces valeurs propres qui sont importantes alors que dans la co-inertie ce sont les valeurs propres elles-mêmes. En effet, on cherche à minimiser :

$$d^2(\mathbf{X}, \hat{\mathbf{Y}}) = \|\mathbf{X} - \mathbf{Y}\mathbf{R}'\|^2 = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \hat{y}_{ij})^2$$

A l'optimum, cette distance sera :

$$\|\mathbf{Y} - \mathbf{R}\mathbf{X}'\|^2 = 2\left(1 - \sum_{k=1}^r \mathbf{q}_k\right)$$

La quantité  $\sum_{k=1}^r \mathbf{q}_k$  est caractéristique de la qualité d'ajustement de la rotation procuste. On appelle cette quantité  $m_{\mathbf{X}\mathbf{Y}}^2 = m_{\mathbf{Y}\mathbf{X}}^2 = m^2$  quand les deux nuages ont reçu la mise à l'échelle initiale (Digby&Kp87 p. 114) :

$$m^2 = \sum_{k=1}^r \mathbf{q}_k$$

Le test PROTEST <sup>Jackson-95</sup> est le test de permutation (Procuste : PROTEST) entre les deux tableaux équivalent du test de permutation de la co-inertie (CoInertia: Coinertia test - Fixed D) qui a une constante près porte donc sur :

$$\mathbf{a}^2 = \sum_{k=1}^r \mathbf{q}_k^2$$

PROTEST porte donc sur  $\sum_{k=1}^r \text{cov}(\mathbf{X}\mathbf{u}_k, \mathbf{Y}\mathbf{v}_k)$  alors que le test de la co-inertie porte sur  $\sum_{k=1}^r \text{cov}^2(\mathbf{X}\mathbf{u}_k, \mathbf{Y}\mathbf{v}_k)$ . Les deux systèmes de vecteurs propres sont identiques en dépit de la mise à l'échelle initiale qui n'a pas lieu dans la co-inertie.

On gardera donc les mêmes axes dans les deux analyses, mais leurs utilisation sera différente. Notons cependant que la comparaison entre Procuste et Co-inertie ne tient que pour deux ACP centrées ou normées, la seconde méthode trouvant de multiples autres usages.

Le module conserve les axes et les valeurs singulières :

```
File aa.prsd contains the singular values of YtX
--- It has 11 rows and 1 column
```

```
File aa.prU contains loadings (data set X)
--- It has 11 rows and 2 columns
File :aa.prU
| Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | -4.117e-01 | 4.928e-01 |
| 2 | -4.352e-01 | 5.353e-01 |
|-----|-----|-----|
```

```
File aa.prV contains loadings (data set Y)
--- It has 27 rows and 2 columns
File :aa.prV
| Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | -2.935e-01 | 3.832e-01 |
| 2 | -4.421e-01 | 1.716e-01 |
|-----|-----|-----|
```

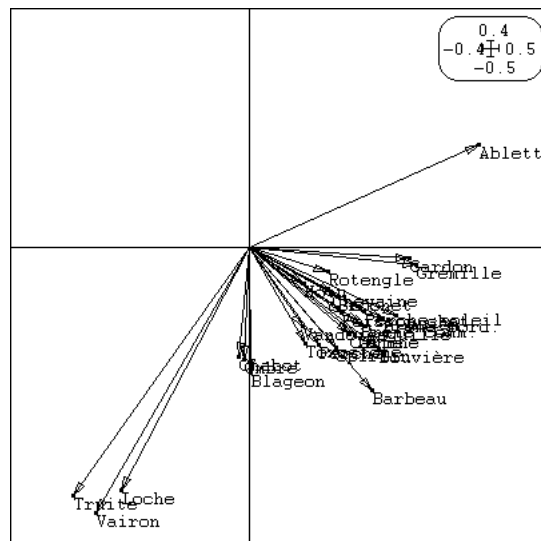
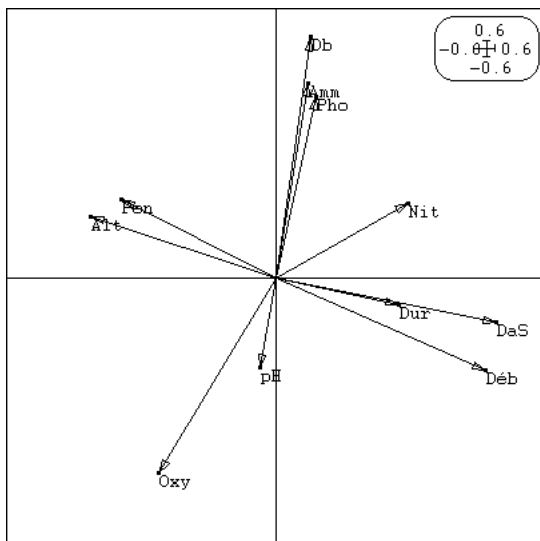
Noter l'identité prévue :

```
Binary input file: D:\Ade4\DIR_TRY\DOUBS\aa.prU Binary input file: D:\Ade4\Dir_Try\Doubs\coiner.iiv1
1 | 0.4928 -0.1004 1 | 0.4928 -0.1004
2 | -0.4117 0.1364 2 | -0.4117 0.1364
3 | -0.3425 0.1734 3 | -0.3425 0.1734
4 | 0.4683 -0.2075 4 | 0.4683 -0.2075
5 | -0.0341 -0.2013 5 | -0.0341 -0.2013
6 | 0.2748 -0.0607 6 | 0.2748 -0.0607
7 | 0.0913 0.4032 7 | 0.0913 0.4032
8 | 0.2957 0.1659 8 | 0.2957 0.1659
9 | 0.0732 0.4336 9 | 0.0732 0.4336
10 | -0.2592 -0.4352 10 | -0.2592 -0.4352
11 | 0.0777 0.5353 11 | 0.0777 0.5353
```

Utiliser ces fichiers pour représenter les variables des tableaux comme projections des vecteurs des bases canoniques sur les plans de références (observation qui ne semble pas être mentionnée dans la bibliographie) :

Labels		Labels	
XY coordinates file	Set	\Ade4\Dir_Try\Doubs\aa.prU	11 2
X-axis column number (default	Set		
Y-axis column number (default	Set		
Label file (or #) for items	Set	ir_Try\Doubs\VarMil_label.txt	
Label file (or #) for groups	Set		
Draw vectors from origin (yes	Set	1	

Le résultat est facilement interprétable.



```
File aa.prLX contains row scores (data set X)
--- It has 30 rows and 2 columns
File :aa.prLX
```

Col.	Mini	Maxi
1	-2.466e-01	2.438e-01
2	-1.164e-01	3.491e-01

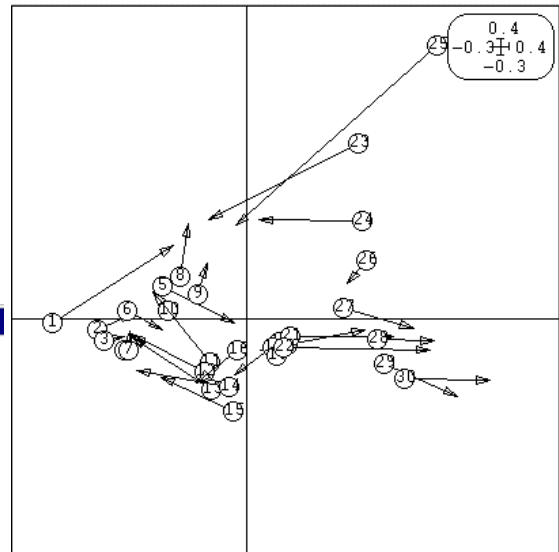
File aa.prLY contains row scores (data set Y)  
 --- It has 30 rows and 2 columns

File :aa.prLY

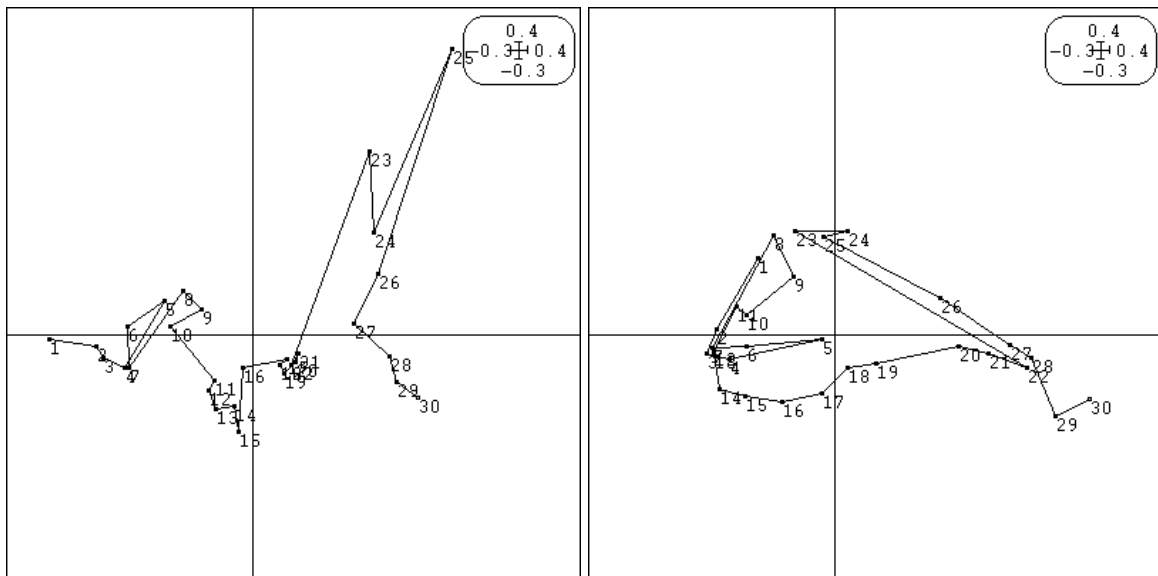
Col.	Mini	Maxi
1	-1.560e-01	3.113e-01
2	-9.766e-02	1.274e-01

On peut superposer les deux nuages sur les axes de référence :

Match two scatters		
XY coordinates file	Set	Ade4\Dir_Try\Doubs\aa.prLY 30 2
X-axis column number (default	Set	
Y-axis column number (default	Set	
Second XY coordinates file	Set	Ade4\Dir_Try\Doubs\aa.prLY 30 2
Label file (or #) for items	Set	#
Label file (or #) for groups	Set	

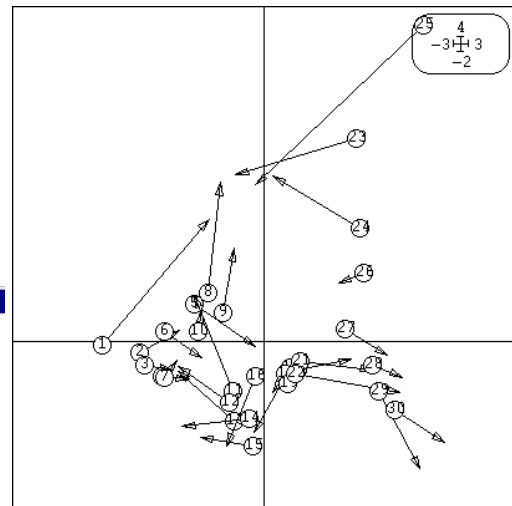


On identifie la co-structure très particulière des tableaux :



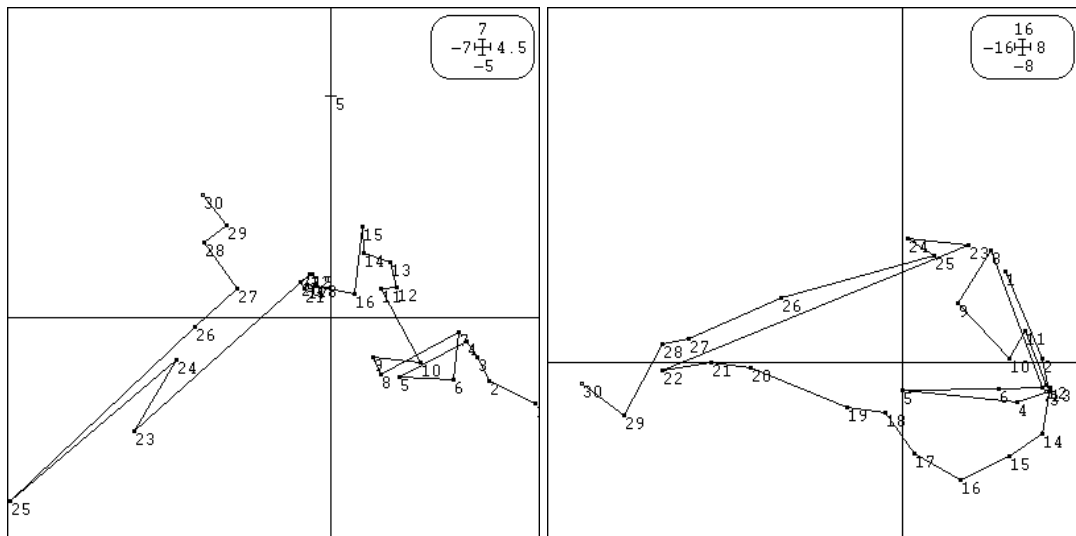
Le gradient amont-aval est perturbé par la pollution (à gauche), il est replié dans la structure faunistique (à droite) : tête de bassin et zone polluée se retrouvent dans la même diminution de richesse faunistique. On retrouve un graphique voisin en co-inertie :

Match two scatters			
XY coordinates file	Set	e4\Dir_Try\Doubs\coiner.iim1	30 2
X-axis column number (default	Set		
Y-axis column number (default	Set		
Second XY coordinates file	Set	e4\Dir_Try\Doubs\coiner.iim2	30 2
Label file (or #) for items	Set	#	
Label file (or #) for groups	Set		

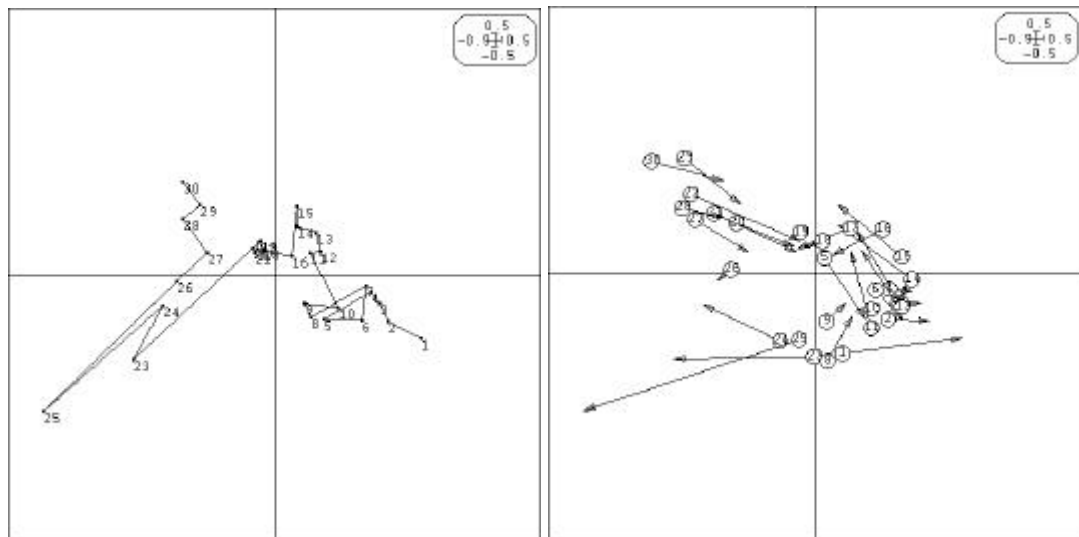


Il y a donc une unité de fond entre l'analyse Procuste classique introduite en écologie par Digby & Kp87 qui l'attribuent à Gower--71 et l'analyse de co-inertie introduite dans Chessele93 et qui trouve son origine dans Tucker--58. La grande différence vient de la mise à l'échelle préalable dans le premier cas, suivie d'une rotation non déformante, suivie d'une projection alors que la seconde remet à l'échelle la projection sur chacun des axes.

La rotation Procuste trouve son plein intérêt quand on veut superposer deux analyses. Si on fait l'ACP normée de DouMil et l'ACP centrée de DouPoi en gardant deux facteurs on a :



Si on veut conserver exactement le premier pour y superposer le second, la rotation Procuste s'impose.



C'est ce type d'exemple qui est proposé par **Digby&Kp87** (p. 116). Le chapitre 4 de cet ouvrage est fondamental. C'est le plus court et il ne contient pas de citations de la littérature écologique. **Jackson-95** dit que :

*Procrustean methods are used infrequently in ecology. This lack of use likely reflects the previously limited availability of the procedure.*

Mais ce chapitre 4 indique clairement que la comparaison de typologie sera fondamentale en écologie des communautés. Les analyses procustéennes généralisées y sont citées. Elles sont initiées par **Gower---75**. Elles ont été l'objet d'une abondante bibliographie et de débats méthodologiques (**ten\_Ber-77**, **ten\_Ber93b**). Elles sont en concurrence avec STATIS, AFM et ACOM disponibles dans ADE-4).

# Procuste : Tow distances

**Type** Analyse multivariée d'un couple de deux matrices de distances euclidiennes.

**Objet** L'option assure la rotation procustéenne (Procuste : Tow arrays) entre les représentations euclidiennes de deux matrices de distances euclidiennes (voir DMAUtil et DMAUse). Elle équivaut strictement avec deux usages de DMAUse: Principal Coordinates sur deux matrices de distances et Procuste : Tow arrays sur les coordonnées.

**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) Nom générique des fichiers de sortie.

**Exemple** La seule contrainte impose que les distances utilisées soient euclidiennes. Reprendre l'exemple de Procuste : Tow arrays sur la carte Doubs. Faire l'acp normée de DouMil :

puis calculer la distance canonique entre stations (DMAUtil: Triplet To Distance) :

```

Computed distances use the diagonal metric and the centered table of the triplet
Output file: D:\Ade4\Dir_Try\Doubs\DouMil_MD0cn
It has 435 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\Dir_Try\Doubs\DouMil_MD0cn.dma
1 -> 30
2 -> 1
3 -> Euclidean distance among rows from triplet D:\Ade4\Dir_Try\Doubs\DouMil.cnta
4 -> TRUE
    
```

Calculer la distance associée à la dissimilarité de Jaccard (entre contenu faunistique) :

Table 2 p. 23

Input file: D:\Ade4\Dir\_Try\Doubs\DouPoi  
 It has 30 rows and 27 columns  
 Distances are computed among rows

JACCARD index (1901)  
 S3 coefficient of GOWER & LEGENDRE  
 Euclidean distance  
 Distances are computed by  
 $s = a/(a+b+c) \rightarrow d = \sqrt{1 - s}$   
 Output file: D:\Ade4\Dir\_Try\Doubs\DouPoi\_Sim1  
 It has 435 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\Dir\_Try\Doubs\DouPoi\_Sim1.dma  
 1 -> 30  
 2 -> 1  
 3 -> JACCARD index on D:\Ade4\Dir\_Try\Doubs\DouPoi  
 4 -> TRUE

Associer ces deux matrices de distances :

Two distances		
First dma type file	Set	y\Doubs\DouMil_MD0cn.dma
Matrix number	Set	
Second dma type file	Set	Try\Doubs\DouPoi_Sim1.dma
Matrix number	Set	
Output file name	Set	dd

Input file: D:\Ade4\Dir\_Try\Doubs\DouMil\_MD0cn.dma  
 Distance file: D:\Ade4\Dir\_Try\Doubs\DouMil\_MD0cn  
 Row: 30 Col: 1 Col used: 1  
 Origin: Euclidean distance among rows from triplet D:\Ade4\Dir\_Try\Doubs\DouMil.cnta  
 Euclidean distance / Uniform row weights

*Première analyse en coordonnées principales, (PCO) :*

Rank	Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
11	01	+6.3216E+00	+0.5747	+0.5747	02	+2.2316E+00	+0.2029	+0.7776
	03	+1.0042E+00	+0.0913	+0.8689	04	+5.0068E-01	+0.0455	+0.9144
	05	+3.7518E-01	+0.0341	+0.9485	06	+2.4797E-01	+0.0225	+0.9710
	07	+1.6647E-01	+0.0151	+0.9862	08	+1.0716E-01	+0.0097	+0.9959
	09	+2.3544E-02	+0.0021	+0.9980	10	+1.7259E-02	+0.0016	+0.9996
	11	+4.3137E-03	+0.0004	+1.0000				

*Seconde analyse en coordonnées principales, (PCO) :*

File D:\Ade4\Dir\_Try\Doubs\DouMil\_MD0cn\_PCO contains the principal coordinates  
 (norm=sqrt(lambda))  
 --- It has 30 rows and 11 columns

Input file: D:\Ade4\Dir\_Try\Doubs\DouPoi\_Sim1.dma  
 Distance file: D:\Ade4\Dir\_Try\Doubs\DouPoi\_Sim1  
 Row: 30 Col: 1 Col used: 1  
 Origin: JACCARD index on D:\Ade4\Dir\_Try\Doubs\DouPoi  
 Euclidean distance / Uniform row weights

Rank	Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
	01	+1.0228E-01	+0.3135	+0.3135	02	+3.5472E-02	+0.1087	+0.4223
	03	+3.0343E-02	+0.0930	+0.5153	04	+2.5033E-02	+0.0767	+0.5920
	05	+2.0422E-02	+0.0626	+0.6546	06	+1.6410E-02	+0.0503	+0.7049
	07	+1.4028E-02	+0.0430	+0.7479	08	+1.2564E-02	+0.0385	+0.7864
	09	+1.1713E-02	+0.0359	+0.8223	10	+9.2675E-03	+0.0284	+0.8507
	11	+8.2476E-03	+0.0253	+0.8760	12	+7.5126E-03	+0.0230	+0.8990

...

File D:\Ade4\Dir\_Try\Doubs\DouPoi\_Sim1\_PCO contains the principal coordinates  
 (norm=sqrt(lambda))  
 --- It has 30 rows and 26 columns

File :D:\Ade4\Dir\_Try\Doubs\DouPoi\_Sim1\_PCO

Procuste analysis  
 X : First input array D:\Ade4\Dir\_Try\Doubs\DouMil\_MD0cn\_PCO



Number of rows: 30, columns: 11  
 Y : Second input array D:\Ade4\Dir\_Try\Doubs\DouPoi\_Sim1\_PCO  
 Number of rows: 30, columns: 26

File dd.prma contains the name of the input arrays  
 X ---> D:\Ade4\Dir\_Try\Doubs\DouMil\_MD0cn\_PCO  
 Y ---> D:\Ade4\Dir\_Try\Doubs\DouPoi\_Sim1\_PCO

File dd.prX contains X/sqrt(trace1) after centring  
 trace1 = Trace (XtX) = 330  
 --- It has 30 rows and 11 columns  
 File dd.prY contains Y/sqrt(trace2) after centring  
 trace2 = trace (YtY) = 9.787  
 --- It has 30 rows and 26 columns

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.0199E-01	+0.7910	+0.7910	02	+1.9854E-02	+0.1540	+0.9449
03	+3.1178E-03	+0.0242	+0.9691	04	+1.7367E-03	+0.0135	+0.9826
05	+1.2114E-03	+0.0094	+0.9920	06	+6.9938E-04	+0.0054	+0.9974
07	+1.4918E-04	+0.0012	+0.9986	08	+1.4263E-04	+0.0011	+0.9997
09	+2.3338E-05	+0.0002	+0.9998	10	+1.5322E-05	+0.0001	+1.0000
11	+4.1459E-06	+0.0000	+1.0000				

Garder deux axes.

File dd.prLX contains row scores (data set X)  
 --- It has 30 rows and 2 columns

File :dd.prLX

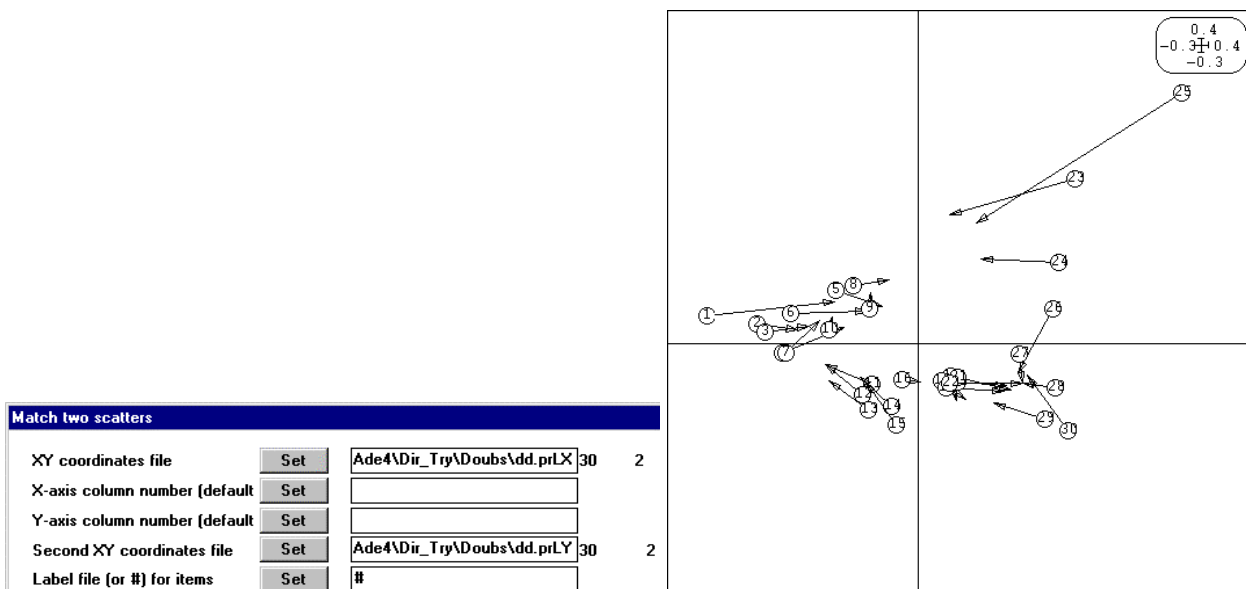
Col.	Mini	Maxi
1	-2.523e-01	3.158e-01
2	-1.034e-01	3.001e-01

File dd.prLY contains row scores (data set Y)  
 --- It has 30 rows and 2 columns

File :dd.prLY

Col.	Mini	Maxi
1	-1.435e-01	1.311e-01
2	-7.020e-02	1.545e-01

Représenter la paire d'ordinations :



Cette figure est très satisfaisante. Elle permet de découper la rivière en zones continues assez simplement (1-10, 11-15, 16-30 sauf 23-25). PCO et rotation procustéenne sont cohérentes au niveau des principes.

# Références

- Basalaj-01** Basalaj, W. (2001) *Proximity visualization of abstract data*. PhD thesis, Trinity College Cambridge. 1-117. (<http://www.pavis.org/essay/pavis.pdf>)
- Chessele93** Chessel, D. & Mercier, P. (1993) Couplage de triplets statistiques et liaisons espèces-environnement. In : *Biométrie et Environnement*. Lebreton, J.D. & Asselain, B. (Eds.) Masson, Paris. 15-44.
- Cliff---66** Cliff, N. (1966) Orthogonal rotation to congruence. *Psychometrika* : 31, 33-42.
- Digby&Kp87** Digby, P. G. N. & Kempton, R. A. . (1987) *Multivariate Analysis of Ecological Communities*. Chapman and Hall, Population and Community Biology Series, London. 1-205.
- Dietz---83** Dietz, E.J. (1983) Permutation tests for association between two distance matrices. *Systematic Zoology* : 32, 21-26.
- Dolédech94** Dolédec, S. & Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.
- Dutille&00** Dutilleul, P., Stockwell, J.D., Frigon, D. & Legendre, P. (2000) The Mantel test versus Pearson's correlation analysis: assesment of the differences for biological and environmental studies. *Journal of Agricultural, Biological, and Environmental Statistics* : 5, 131-150.
- Gower---71** Gower, J.C. (1971) Statistical methods of comparing different multivariate analyses of the same data. In : *Mathematics in the archaeological and historical sciences*. Hodson, F.R, Kendall, D.G. & Tautu, P. (Eds.) University Press, Edinburgh. 138-149.
- Gower---75** Gower, J.C. (1975) Generalized procustes analysis. *Psychometrika* : 40, 33-51.
- Green---52** Green, B.F. (1952) The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika* : 17, 429-440.
- Heo&Gabe98** Heo, M. & Gabriel, K.R. (1997) A permutation test of association between configurations by means of the RV coefficient. *Communications in Statistics - Simulation and Computation* : 27, 843-856.
- Hurley&t62** Hurley, J.R. & Cattell, R.B. (1962) Producing direct rotation to test a hypothesized factor structure. *Behavioural Science* : 7, 258-262.
- Jackson-95** Jackson, D.A. (1995) PROTEST: a PROcustean randomization TEST of community environment concordance. *Ecosciences* : 2, 297-303.
- Kazi-Aol95** Kazi-Aoual, F., Hitier, S., Sabatier, R. & Lebreton, J.D. (1995) Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis* : 20, 643-656.
- Klingentr98** Klingenberg, C.P. & McIntyre, G.S. (1998) Geometric morphometrics of developmental instability: Analyzing patterns of fluctuating asymmetry with procrustes. *Evolution* : 52, 1363-1375.

- Mantel--67** Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* : 27, 209-220.
- Mouttet-81** Mouttet, F. (1981) *Comparaison de tableaux par la méthode Procruste*. Thèse de 3<sup>e</sup> cycle, Université Paris VI. 1-98.
- Olshan&182** Olshan, A.F., Siegel, A.F. & Swindler, D.R. (1982) Robust and least-squares orthogonal mapping: Methods for the study of cephalofacial form and growth. *American Journal of Physical Anthropology* : 59, 131-137.
- Peres-No01** Peres-Neto, P.R. & Jackson, D.A. (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Æologia* : (in press).
- Rao-----80** Rao, C.R. (1980) Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In : *Multivariate Analysis - V. Krishnaiah, P.R. (Ed.)* North-Holland Publishing Company. 3-22.
- Rohlf&Sc90** Rohlf, F.J. & Slice, D. (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* : 39, 40-59.
- Schönemmn66** Schönemann, P.H. (1966) A generalized solution solution of the orthogonal Procrustes problem. *Psychometrika* : 31, 1-10.
- Schönemmn68** Schönemann, P.H. (1968) On two-sided Procrustes problems. *Psychometrika* : 33, 19-34.
- Schönemmn70** Schönemann, P.H. & Carrol, R.M. (1970) Fitting one matrix to another under choice of a central dilatation and a rigid motion. *Psychometrika* : 35, 245-256.
- Sibson--78** Sibson, R. (1978) Studies in the robustness of multidimensional scaling. *Journal of the Royal Statistical Society, B* : 40, 234-238.
- Sneath--67** Sneath, P.H.A. (1967) Trend-surface analysis of transformation grids. *Journal of Zoology* : 151, 65-122.
- ten\_Ber-77** Ten Berge, J.M.F. (1977) Orthogonal Procrustes rotation for two or more matrices. *Psychometrika* : 42, 267-276.
- ten\_Ber93b** Ten Berge, J.M.F. & Bekker, P.A. (1993b) The isotropic scaling problem in Generalized Procrustes Analysis. *Computational Statistics and Data Analysis* : 16, 201-204.
- Ter\_Bra87c** Ter Braak, C.J.F. (1987c) CANOCO - a FORTRAN program for Canonical community ordination by [partial][detrended][canonical] correspondence analysis and redundancy analysis. Software documentation. Version 2.1, TNO Institute of Applied Computer Science, Wageningen.
- Torre&Cs94** Torre, F. & Chessel, D. (1994) Co-structure de deux tableaux totalement appariés. *Revue de Statistique Appliquée* : 43, 109-121.
- Tucker--51** Tucker, L.R. (1951) *A method for the synthesis of factor analysis studies*. Technical Report 984, Department of the Army, Washington, DC.
- Tucker--58** Tucker, L.R. . (1958) An inter-battery method of factor analysis. *Psychometrika* : 23, 111-136.

**Verneau-73** Verneaux, J. (1973) *Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie.* Thèse d'état, Besançon. 1-257.

---