



# Mesurer l'amplitude des espèces

## Résumé

Suite à une question d'E. Castella sur Adelist, la fiche illustre les mesures de diversité dérivées des indices de Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* : 21, 24-43.) quand on s'en sert comme mesure d'amplitude d'habitat.

## Plan

INTRODUCTION.....	2
LA QUESTION DE L'ECHELLE .....	4
LE CAS DES VARIABLES QUANTITATIVES.....	5
POUR FAIRE LES CALCULS.....	6
CHOISIR SA DISTANCE .....	11
REFERENCES .....	17

D. Chessel

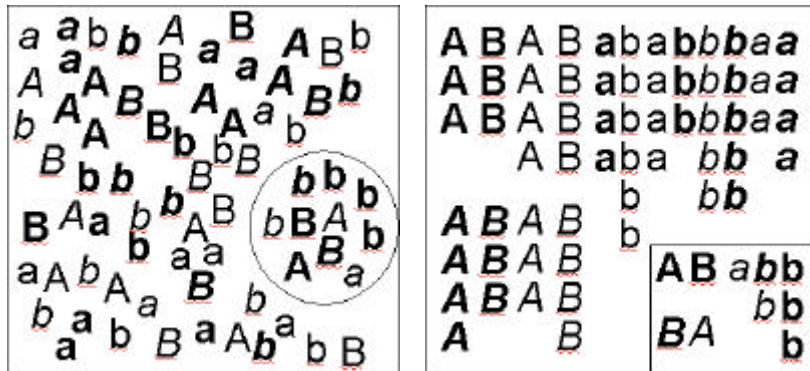
# Introduction

La question est d'Emmanuel Castella (Emmanuel.Castella@leba.unige.ch) sur Adelist.

Bonjour,

Je cherche à calculer un indice d' "amplitude " d'espèces à partir du fichier de variances/covariances à l'issue de COA/Reciprocal scaling sur un tableau espèces x relevés. Quelqu'un aurait-il déjà utilisé ce fichier pour cela ? Si l'on considère la surface de l'ellipse (si deux axes factoriels conservés) ou le volume de l'ellipsoïde (si plus de 2 axes sont conservés) comme un tel indice, les variances doivent-elles être pondérées par le pourcentage d'inertie expliquée par l'axe factoriel correspondant dans le calcul de la surface ou du volume ? La démarche a-t-elle un sens ?

La fiche donne des indications sur cette question d'intérêt très général. On partira du constat que la mesure de l'amplitude d'une espèce et la mesure de la diversité d'un relevé sont deux exemples du même problème. Il s'agit de quantifier la variabilité d'une collection.



Le carré contient une collection d'objets. Le cercle en contient une partie. Comment mesurer la variabilité de ce qui est dans le cercle ? Si les objets sont des individus, si les propriétés des objets sont des génotypes et si le cercle est une population on mesurera la *variabilité génétique*. Si les objets sont des relevés, si les propriétés des objets sont des caractéristiques environnementales et si le cercle est la partie des relevés où se trouve une espèce on mesurera l'*amplitude d'habitat*. Si les objets sont des individus, si les propriétés des objets sont des paramètres biologiques ou taxonomiques et si le cercle est un relevé on mesurera la *diversité écologique*.

La question posée associe amplitude et variance, typologie et diversité. Elle cherche une référence d'échelle. Elle contient toutes les difficultés du problème.

<b>AAAABBBBaaaaabbbb</b>															
4	3	3	4	4	3	4	3	4	4	3	4	6	3	5	5
0	1	1	0	0	1	0	1	0	0	1	0	0	3	1	1

La diversité, au sens large, d'une collection est la variabilité des objets qui forment cette collection : variabilité entre génotypes d'une population, variabilité entre biotopes où on trouve

une espèce, variabilité entre espèces présentes en un site. Pour mesurer la variabilité d'une collection, il est nécessaire de mesurer la dissemblance entre deux objets de la collection. La mesure la plus simple est celle qui attribue 0 à un couple de deux objets identiques et 1 sinon.  $A$  n'est pas  $b$ , donc  $d(A,b) = 1$ . On a l'avantage de la simplicité ! En écologie, on se gardera bien de plaisanter sur le caractère caricatural de cette mesure.

On peut dire aussi que  $A$  est plus différent de  $b$  que  $b$  ne l'est de  $b$ . Les objets ont quatre propriétés (a ou A, a ou b, a ou a, a ou a).  $A$  et  $b$  n'ont aucune propriété en commun,  $A$  et a en ont une,  $A$  et a en ont deux,  $A$  et A en ont trois. On peut définir une matrice de distances :

	A	A	A	A	B	B	B	B	a	a	a	a	b	b	b	b
A	0	1	1	2	1	2	2	3	1	2	2	3	2	3	3	4
A	1	0	2	1	2	1	3	2	2	1	3	2	3	2	4	3
A	1	2	0	1	2	3	1	2	2	3	1	2	3	4	2	3
A	2	1	1	0	3	2	2	1	3	2	2	1	4	3	3	2
B	1	2	2	3	0	1	1	2	2	3	3	4	1	2	2	3
B	2	1	3	2	1	0	2	1	3	2	4	3	2	1	3	2
B	2	3	1	2	1	2	0	1	3	4	2	3	2	3	1	2
B	3	2	2	1	2	1	1	0	4	3	3	2	3	2	2	1
a	1	2	2	3	2	3	3	4	0	1	1	2	1	2	2	3
a	2	1	3	2	3	2	4	3	1	0	2	1	2	1	3	2
a	2	3	1	2	3	4	2	3	1	2	0	1	2	3	1	2
a	3	2	2	1	4	3	3	2	2	1	1	0	3	2	2	1
b	2	3	3	4	1	2	2	3	1	2	2	3	0	1	1	2
b	3	2	4	3	2	1	3	2	2	1	3	2	1	0	2	1
b	3	4	2	3	2	3	1	2	2	3	1	2	1	2	0	1
b	4	3	3	2	3	2	2	1	3	2	2	1	2	1	1	0

	X/x	a/b	X/X	X/X
A	1	1	1	1
A	1	1	1	0
A	1	1	0	1
A	1	1	0	0
B	1	0	1	1
B	1	0	1	0
B	1	0	0	1
B	1	0	0	0
a	0	1	1	1
a	0	1	1	0
a	0	1	0	1
a	0	1	0	0
b	0	0	1	1
b	0	0	1	0
b	0	0	0	1
b	0	0	0	0

La distance entre deux objets est le nombre de critère pour lesquels les deux objets ont une valeur différente. Disons, par souci de cohérence que ce nombre définit le carré de la distance entre deux objets : c'est tout simplement le carré de la distance euclidienne calculée sur le tableau des indicatrices d'une des deux modalités ( $(1-0)^2 = 1$ ).

Dés qu'on a attribué une distance entre éléments d'une collection, on peut définir la diversité de cette collection. C'est Rao <sup>1</sup> qui a défini une base théorique unificatrice qui relie diversité écologique (les indices), diversité génétique (les distances) et mesures générales de la variabilité (les variances). Il a défini lui-même ces bases comme une axiomatisation dans <sup>2</sup>. Nous avons apporté les aménagements nécessaires pour mettre à disposition cette théorie en écologie <sup>3</sup>. Soit un relevé écologique contenant  $s$  espèces avec les proportions  $(p_1, p_2, \dots, p_s)$ . Soit :

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

une matrice de distances entre espèces. L'indice de Rao simple est :

$$R_s(p_1, \dots, p_s) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2$$

On aurait pu dire : soit une espèce présente dans  $s$  relevés avec les proportions  $(p_1, p_2, \dots, p_s)$ . Soit :

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

une matrice de distances entre relevés. L'indice de Rao simple mesure l'amplitude d'habitat de l'espèce. La diversité de la petite collection de l'exemple choisi se réduit à :

	A	A	A	A	B	B	B	B	a	a	a	a	b	b	b	b	
A	0	1	1	2	1	2	2	3	1	2	2	3	2	3	3	4	0
A	1	0	<b>2</b>	1	2	<b>1</b>	3	<b>2</b>	2	1	<b>3</b>	2	3	<b>2</b>	<b>4</b>	<b>3</b>	0.1
A	1	2	<b>0</b>	1	2	<b>3</b>	1	<b>2</b>	2	3	<b>1</b>	2	3	<b>4</b>	<b>2</b>	<b>3</b>	0.1
A	2	1	1	0	3	2	2	1	3	2	2	1	4	3	3	2	0
B	1	2	2	3	0	1	1	2	2	3	3	4	1	2	2	3	0
B	2	1	<b>3</b>	2	1	<b>0</b>	2	<b>1</b>	3	2	<b>4</b>	3	2	<b>1</b>	<b>3</b>	<b>2</b>	0.1
B	2	3	1	2	1	2	0	1	3	4	2	3	2	3	1	2	0
B	3	2	<b>2</b>	1	2	<b>1</b>	1	<b>0</b>	4	3	<b>3</b>	2	3	<b>2</b>	<b>2</b>	<b>1</b>	0.1
a	1	2	2	3	2	3	3	4	0	1	1	2	1	2	2	3	0
a	2	1	3	2	3	2	4	3	1	0	2	1	2	1	3	2	0
a	2	3	<b>1</b>	2	3	<b>4</b>	2	<b>3</b>	1	2	<b>0</b>	1	2	<b>3</b>	<b>1</b>	<b>2</b>	0.1
a	3	2	2	1	4	3	3	2	2	1	1	0	3	2	2	1	0
b	2	3	3	4	1	2	2	3	1	2	2	3	0	1	1	2	0
b	3	2	<b>4</b>	3	2	<b>1</b>	3	2	2	1	<b>3</b>	2	1	<b>0</b>	<b>2</b>	<b>1</b>	0.3
b	3	4	<b>2</b>	3	2	<b>3</b>	1	2	2	3	<b>1</b>	2	1	<b>2</b>	<b>0</b>	<b>1</b>	0.1
b	4	3	<b>3</b>	2	3	<b>2</b>	2	1	3	2	<b>2</b>	1	2	<b>1</b>	<b>1</b>	<b>0</b>	0.1
	0	0.1	0.1	0	0	0.1	0	0.1	0	0	0.1	0	0	0.3	0.1	0.1	

- 1) définir la liste des objets de la collection ;
- 2) définir le distance entre les objets de la collection ;
- 2) définir la composition de la collection comme une distribution de fréquence ;
- 3) calculer l'indice. On trouve 1.82.

Cette définition très générale contient la réponse à la question posée. Elle en pose elle-même plusieurs autres.

## La question de l'échelle

Trouver 1.82, c'est bien, mais ça ne dit ce que ça signifie. La question s'est toujours posée dans les mesures de diversité. On peut essayer d'y répondre en comparant au maximum possible. La question est : dans quelles proportions faut-il mélanger les objets :

A	A	A	A	B	B	B	B	a	a	a	a	b	b	b	b
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

pour que la diversité de la collection soit maximale. Quel est le maximum de la quantité :

$$R_s(p_1, \dots, p_s) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2$$

lorsque la matrice de distances est donnée. C'est une question, en général difficile, qui a ici une solution particulière. Nous avons  $v$  variables dichotomiques pour définir la distance entre objets.  $k$  est l'indice de la variable. Elle prend les valeurs 0 et 1 et, pour une distribution donnée, sa moyenne est le proportion de 1 qu'on notera  $f_k = \sum_{i=1}^s p_i x_{ik}$ . La distance entre

deux objets est  $d_{ij}^2 = \sum_{k=1}^v (x_{ik} - x_{jk})^2$  et la diversité est :

$$R_s(p_1, \dots, p_s) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j \sum_{k=1}^v (x_{ik} - x_{jk})^2$$

$$R_s = \sum_{k=1}^v \sum_{i=1}^s \sum_{j=1}^s p_i p_j (x_{ik} - x_{jk})^2$$

Les valeurs des variables et leur carré sont identiques (0-1) et on trouve :

$$R_s = 2 \sum_{k=1}^v f_k (1 - f_k) \leq 2v \left( \frac{1}{4} \right) = \frac{v}{2}$$

La borne est atteinte si chaque variable présente le proportion  $\left(\frac{1}{2}, \frac{1}{2}\right)$  de valeurs 0 et 1 donc si chacune des combinaisons est équiprobable (distribution uniforme). Ceci est un cas particulier d'une propriété très générale associée à la définition de la distance :

$$\left\{ \begin{array}{l} R_d(\mathbf{p}) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2 \\ d_{ij}^2 = d_{ij1}^2 + \dots + d_{ijv}^2 \end{array} \right\} \Rightarrow R_d(\mathbf{p}) = R_{d_1}(\mathbf{p}) + \dots + R_{d_v}(\mathbf{p})$$

L'indice de diversité utilisé est en fait bien connu. Les données :

A	A	A	A	B	B	B	B	a	a	a	a	b	b	b	b
0	0.1	0.1	0	0	0.1	0	0.1	0	0	0.1	0	0	0.3	0.1	0.1

peuvent se réduire à la distribution des modalités de chacune des variables :

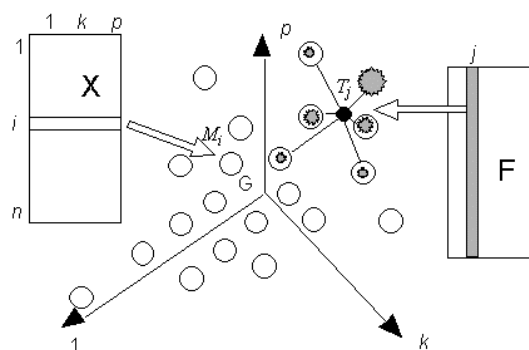
X	x	X	X	X	X	Ta	Tb
0.4	0.6	0.3	0.7	0.5	0.5	0.3	0.7

La diversité de chacun des critères se calcule par l'indice de Simpson  $(1 - \sum_{j=1}^s p_j^2)$ . Ces indices sont sommés et on trouve 1.82. L'indice est maximum pour la distribution  $\left(\frac{1}{2}, \frac{1}{2}\right)$  pour chacune des variables et le maximum de la somme est 4. L'indice de Simpson est un indice de Rao, la somme sur plusieurs indices de Simpson aussi.

La question de l'échelle de la mesure peut donc s'aborder soit à l'intérieur d'un même jeu de données soit dans l'absolu. Dans le premier cas deux indices de Rao calculés sur la même matrice de distance sont comparables. Dans le second on compare au maximum possible et l'échelle est [0,1].

## Le cas des variables quantitatives

La cas le plus général tient dans le schéma (fiche Thema47) :



$n$  relevés sont dans un espace euclidien. Une espèce est une distribution de fréquences sur l'ensemble de ces relevés (en fait la majorité des fréquences sont nulles). Son amplitude est l'étendue de l'espace écologique qu'elle occupe. C'est la variabilité totale de la distribution de

fréquence de l'espèce. L'indice de Rao définit cette variabilité dès qu'une mesure de distances est en place entre les relevés. Remarquons encore une fois qu'on pourrait tenir le discours suivant sur cette figure.  $n$  espèces sont dans un espace euclidien. Un relevé est une distribution de fréquences sur l'ensemble de ces espèces (en fait la majorité des fréquences sont nulles). La diversité du relevé est l'étendue de l'espace biologique qu'il utilise. C'est la variabilité totale de la distribution de fréquence du relevé. L'indice de Rao définit cette variabilité dès qu'une mesure de distances est en place entre les espèces.

On peut même dire qu'il n'est pas nécessaire que les relevés soient dans un espace euclidien. Il suffit de connaître la distance des relevés deux à deux pour définir l'amplitude de l'espèce :

$$R_s(p_1, \dots, p_s) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j d_{ij}^2$$

La différence entre relevés implique l'amplitude des espèces et *vice versa*. Utiliser l'analyse des correspondances, c'est affirmer qu'on veut les deux <sup>4</sup>.

Pour un gradient formé d'une seule variable, la distance s'impose et :

$$R_s(p_1, \dots, p_s) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j (x_i - x_j)^2$$

$$R_s(\mathbf{p}) = 2 \sum_{j=1}^s p_j x_j^2 - 2 \sum_{i=1}^s \sum_{j=1}^s p_i p_j x_i x_j = 2 \text{Var}_{\mathbf{p}}(\mathbf{x})$$

Pour un nuage de relevé dans un espace de dimension  $p$  :

$$R_s(\mathbf{p}) = \sum_{i=1}^s \sum_{j=1}^s p_i p_j \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{Q}}^2 = 2 \text{Iner}_{\mathbf{X}, \mathbf{Q}}(\mathbf{p})$$

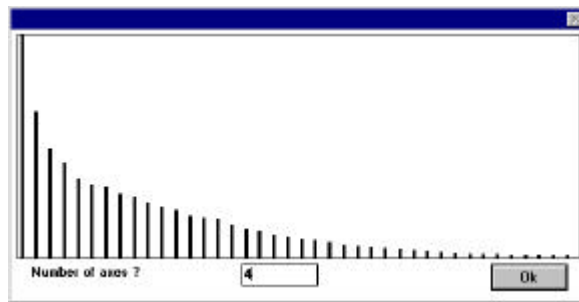
$\mathbf{x}_i$  est la ligne  $i$  du tableau  $\mathbf{X}$ ,  $\mathbf{Q}$  est une métrique mesurant dans l'ensemble des lignes, *Iner* est l'inertie du nuage des lignes de  $\mathbf{X}$  pour la métrique  $\mathbf{Q}$  et la pondération  $\mathbf{p}$ .

On aurait pu prendre la moitié de ces quantités sans changer l'essentiel. La question de l'échelle se repose.

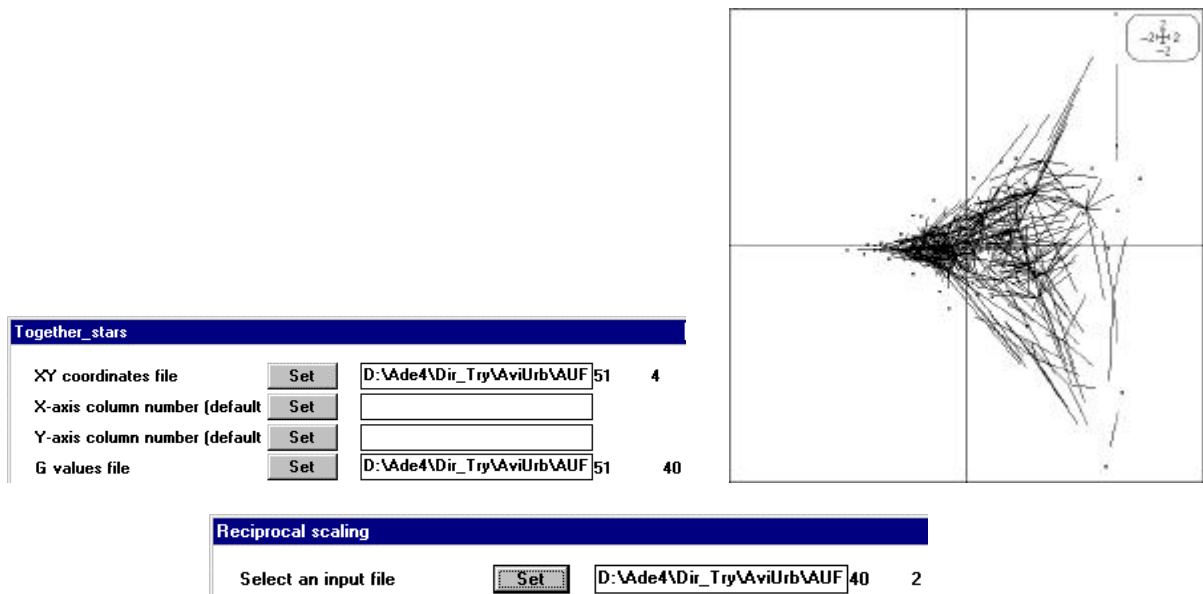
## Pour faire les calculs

Venons en aux choses sérieuses, se dit l'utilisateur d'ADE-4. Tout ça, c'est bien joli mais comment on s'en sert ! Le but est d'observer que ce qui se passe dans COA: Reciprocal scaling qui donne des variances est un cas très particulier qui peut maintenant s'utiliser dans toutes les circonstances. Prenons un exemple bien connu (carte AviUrb, 51 relevés et 40 espèces d'oiseaux).





Dans le plan 1-2 des relevés, on peut représenter la dispersion des espèces (ScatterDistri: Together\_stars) :



DualScalCOA: Reciprocal scaling after a CA  
 Title of the analysis: D:\Ade4\Dir\_Try\AviUrb\AUFau.  
 Number of rows: 51, columns: 40

File D:\Ade4\Dir\_Try\AviUrb\AUFau\_score contains the coordinates of non empty table cells. It has 492 rows and 4 columns

File D:\Ade4\Dir\_Try\AviUrb\AUFau\_weigh contains the value, in the initial table, of these cells. It has 492 rows and 1 columns

File D:\Ade4\Dir\_Try\AviUrb\AUFau\_class contains the row and column number of these cells. It has 492 rows and 2 columns

...

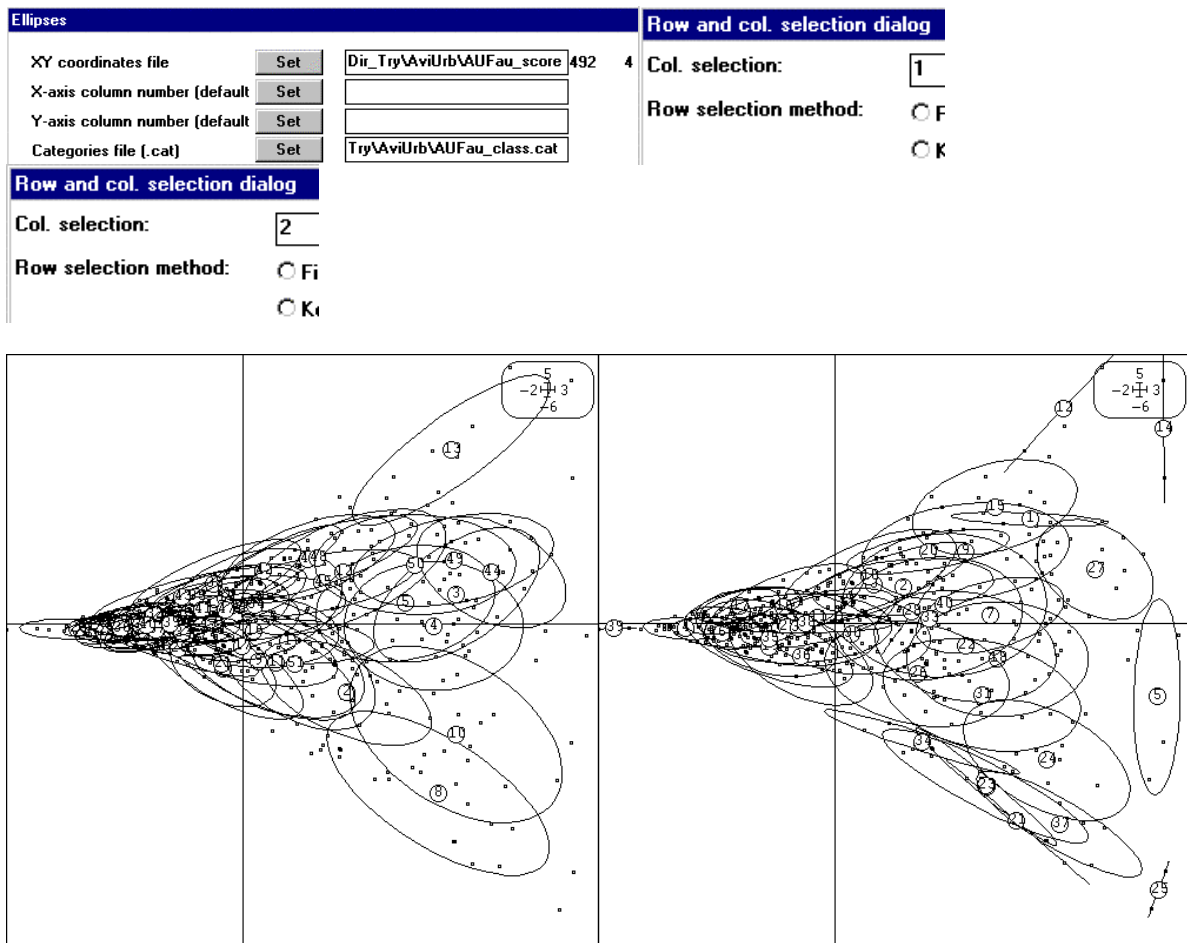
File D:\Ade4\Dir\_Try\AviUrb\AUFau\_mvli contains the within-rows mean-var.-covar. It has 51 rows and 14 columns

Col 1	Mean	1.642	1.929	1.593	-1.394	Var	0.194	0.007	0.612	0.404
Col 2	Mean	0.705	0.680	-1.761	-0.749	Var	0.333	0.109	0.600	0.088
Col 3	Mean	0.324	0.716	1.894	-0.415	Var	0.309	0.278	0.367	0.021
Col 4	Mean	-1.277	-0.105	-0.550	-1.511	Var	0.060	0.012	0.050	0.166
Col 5	Mean	2.694	-1.307	2.381	-1.319	Var	0.016	1.343	0.472	0.174
Col 6	Mean	-0.951	-0.176	0.003	-0.099	Var	0.082	0.035	0.115	0.328
Col 7	Mean	1.305	0.141	-1.061	-0.226	Var	0.172	0.222	0.717	0.145

...

Il faut identifier la contradiction qui existe entre le point de vue de la figure et celui du calcul. Sur la figure les points sont les relevés et les étoiles sont les espèces. Dans le calcul, les points sont les correspondances (occurrences d'une espèce dans un relevé, cases non vides du

tableau). Ce module mesure l'amplitude d'une espèce sur un ensemble d'occurrences pour avoir une échelle commune avec la mesure de la diversité d'un relevé vue sur le même ensemble d'occurrences. La figure associée au calcul est (ScatterClass: Ellipses) :

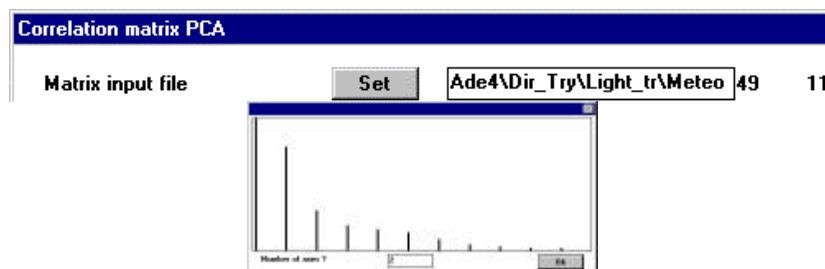


*Horizontalement, gradient urbain-rural, verticalement gradient ouvert-fermé, à gauche les relevés, à droite les espèces. En superposant, on voit un écotone.*

Les calculs de COA: Reciprocal scaling sont donc très particuliers et ne se justifient que pour une discussion double amplitude-diversité.

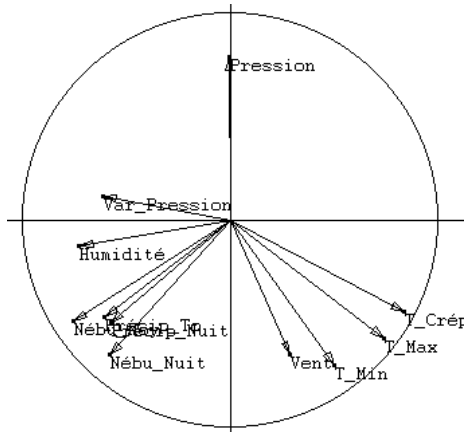
En général, on veut calculer l'amplitude de la distribution des espèces en utilisant des scores numériques arbitraires, que ce soit des variables mesurées, des scores de l'analyse du tableau lui-même ou des scores de l'analyse d'un autre tableau, en particulier d'un tableau de milieu. La procédure est générale.

Utiliser la carte Light-Trap <sup>5</sup> (discussion dans <sup>6</sup>). Faire l'ACP normée des variables météorologiques :

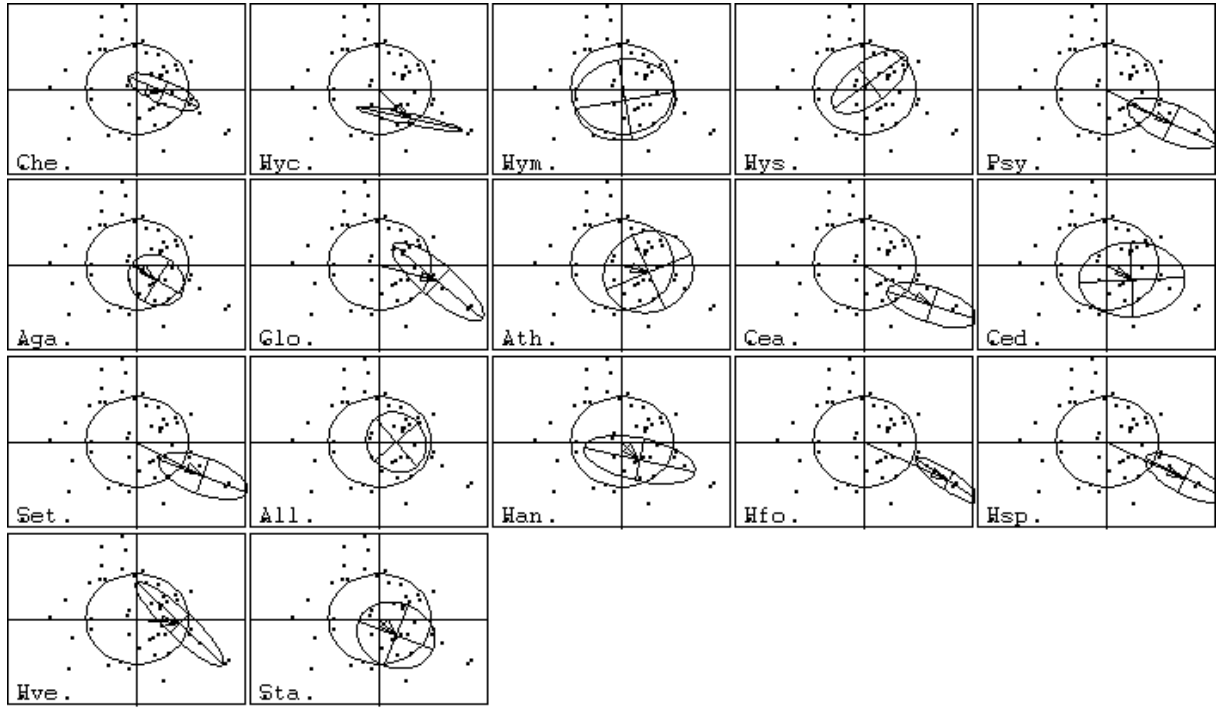




Représenter la niche des taxa :



Ellipses		
XY coordinates file	Set	4\Dir_Try\Light_tr\Meteo.cnli
X-axis column number (default	Set	
Y-axis column number (default	Set	
G values file	Set	:\Ade4\Dir_Try\Light_tr\Fau
Constrain H/V ratio (yes = 1)	Set	1
Draw background (yes = 1)	Set	1
Label file (or #) for G columns	Set	ir_Try\Light_tr\Label_Esp.txt
Add uniform distribution (yes =	Set	1



$$\begin{matrix} 4 \\ -6 \pm 5 \\ -4 \end{matrix}$$

Sur la figure le cercle définit l'amplitude de l'espèce qui aurait une distribution uniforme. Pour mesurer l'amplitude de niche d'une espèce réelle, qui est ici un indice d'influence de la météorologie sur l'émergence des larves, il suffit de définir la distances entre les relevés. Dans DMAUtil :

Canonical distance		
Input file	Set	4\Dir_Try\Light_tr\Meteo.cnli 49 2
Option: Output file	Set	A

Dans DMAUse :

BioDiv Indices		
dma type file	Set	4\Dir_Try\Light_tr\A_EU.dma
Species-Samples array	Set	:\Ade4\Dir_Try\Light_tr\Fau 49 17
Index type	Set	0
Output file name	Set	a0

Distances - Input file: D:\Ade4\Dir\_Try\Light\_tr\A\_EU.dma  
 Text file: D:\Ade4\Dir\_Try\Light\_tr\A\_EU.dma  
 1 -> 49  
 2 -> 1  
 3 -> Classical metric on D:\Ade4\Dir\_Try\Light\_tr\Meteo.cnli  
 4 -> TRUE

Species-Sites array - Input file: D:\Ade4\Dir\_Try\Light\_tr\Fau  
 Species: 49 Sites: 17

Option: 0  
 File a0 contains biodiversity indices  
 Rows (sites): 17 Col (distances): 1  
 Sum(p(i)\*p(j)\*d2(i,j)) without rescaling

Remarquer que le programme a été écrit pour un tableau espèces-relevés dont on connaît une distance entre relevés ce qui donnera un indice de diversité par relevé. On s'en sert ici pour un tableau relevés-espèces dont on connaît une distance entre relevés ce qui donnera un indice d'amplitude par espèce

-----  
 Binary input file: D:\Ade4\Dir\_Try\Light\_tr\a0 - 17 rows, 1 cols.

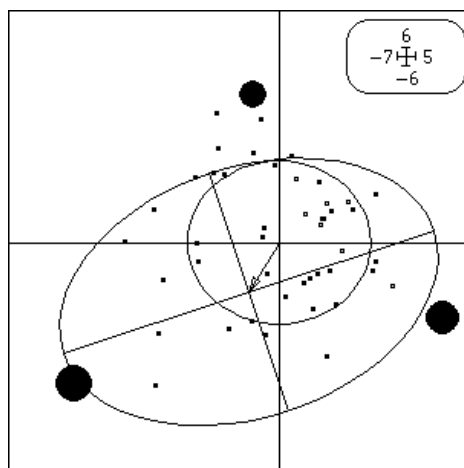
1	4.8415	6	4.4401	11	7.7235	16	11.3892
2	8.8812	7	11.2050	12	5.6116	17	8.1424
3	12.9739	8	11.4227	13	11.2074		
4	7.4901	9	7.7602	14	4.6689		
5	8.0687	10	12.9525	15	6.6673		

On peut prendre comme référence l'amplitude de "l'espèce" uniforme. C'est deux fois l'inertie du nuage donc deux fois la somme des deux valeurs propres soit  $4.0494+3.1881=14.475$ . Pour avoir la distribution qui maximise cette amplitude :

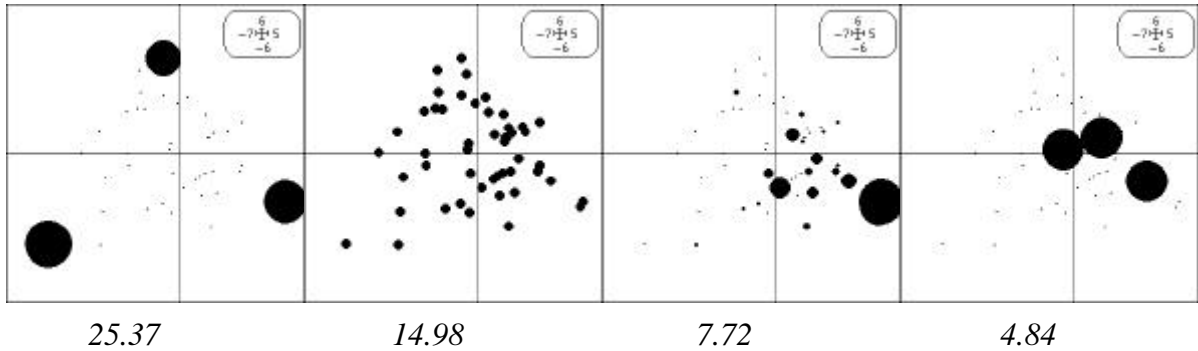


Convergence : 1 65 5.163e+01 25.37  
 Output file: D:\Ade4\Dir\_Try\Light\_tr\A\_EU.nuwd  
 It contains numeric weights. It as 49 rows ans 1 columns

Le maximum possible est 25.37 (contre 14.48 pour la distribution uniforme). La nature de cette distribution est pittoresque :



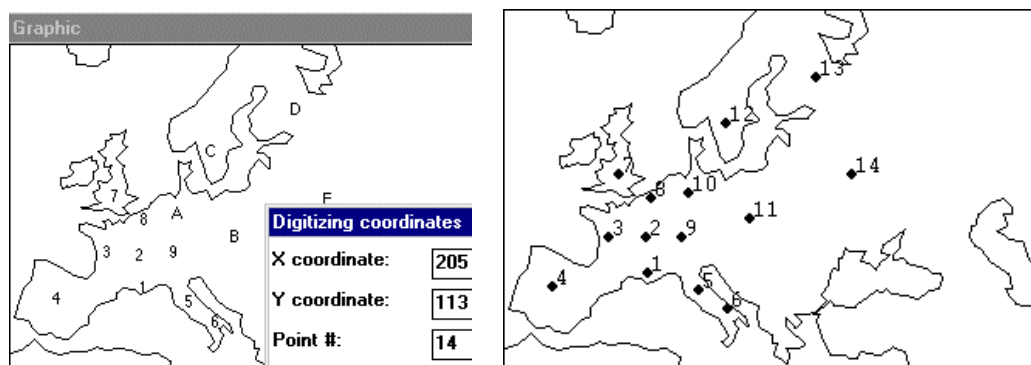
La représentation par ellipses est souvent abusive. On peut s'en tenir à la distribution exacte :



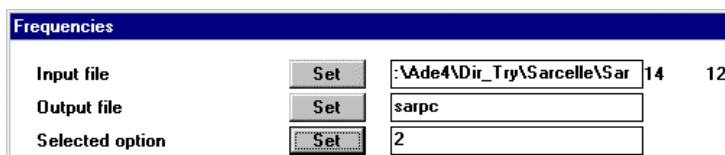
Dans une AFC, cette pratique prendrait comme référence, non pas la pondération uniforme, mais la pondération du fichier .fcpl qui représente l'espèce "totale". La pondération numérique a la propriété mathématique de définir le maximum possible de la mesure mais son interprétation écologique est un clin d'œil. D'autres points de vue sont développés dans DMAUse mais sont pour le moment des propositions théoriques. L'indice de Rao simple (ou sa moitié si on veut exactement une variance ou une inertie) est une mesure naturelle de la variabilité. On peut la ramener à une valeur de référence pour une distribution connue.

## Choisir sa distance

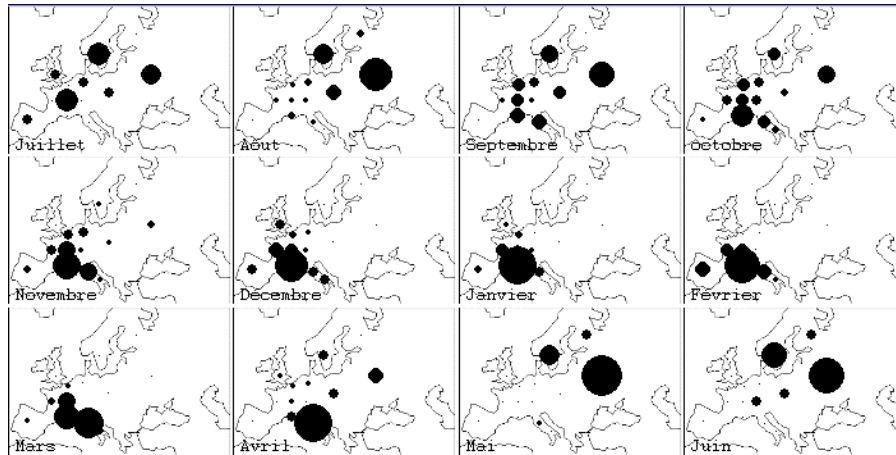
Pour calculer des variances, on pourrait dire que le détour est compliqué. L'intérêt de cette structure de calcul est dans le choix de la distance. Prenons la carte Sarcelles <sup>7</sup>. Le tableau Sar contient une table de contingence à 14 lignes (régions) et 12 colonnes (mois). On veut mesurer la dispersion des canards au cours du temps ce qui est une mesure de variabilité de la position des canards, une mesure d'amplitude. Digitaliser (Digit: Digitize) la position des régions (lignes) avec le fichier Sarc\_Carto.BMP et vérifier (Maps: Labels) :



Passer en fréquence par colonnes pour avoir des distributions mensuelles (Bin-Bin: Frequencies) :



et représenter ces distributions (Maps: Values) :

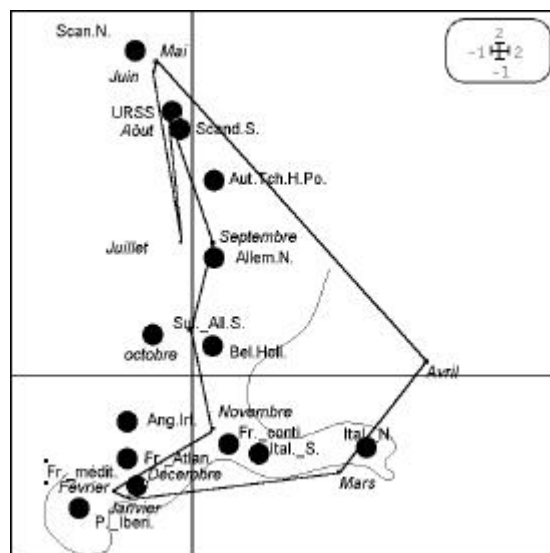


La distance euclidienne (à vol d'oiseaux) s'impose !



```
Canonical distances computed
Output file: D:\Ade4\Dir_Try\Sarcelle\sarxy_EU
It has 91 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\Sarcelle\sarxy_EU.dma
1 -> 14
2 -> 1
3 -> Classical metric on D:\Ade4\Dir_Try\Sarcelle\sarxy
4 -> TRUE
```

On peut aussi faire l'analyse des correspondances du tableau qui reproduit dans son plan 1-2 la migration <sup>7</sup> :



L'analyse induit une distance entre région vue sur cette carte (distance euclidienne sur les coordonnées) :

Canonical distance	
Input file	Set de4\Dir_Try\Sarcelle\Sar.fcli 14 2

```

Canonical distances computed
Output file: D:\Ade4\Dir_Try\Sarcelle\Sar.fcli_EU
It has 91 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\Sarcelle\Sar.fcli_EU.dma
1 -> 14
2 -> 1
3 -> Classical metric on D:\Ade4\Dir_Try\Sarcelle\Sar.fcli
4 -> TRUE

```

L'analyse induit aussi une distance entre région vue dans les propriétés de l'analyse (distance du Chi2) :

Triplet To Distance	
Input file	Set e4\Dir_Try\Sarcelle\Sar.fcta 14 12

```

Computed distances use the diagonal metric and the centered table of the triplet
Output file: D:\Ade4\Dir_Try\Sarcelle\Sar_MD0fc
It has 91 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\Sarcelle\Sar_MD0fc.dma
1 -> 14
2 -> 1
3 -> Euclidean distance among rows from triplet
D:\Ade4\Dir_Try\Sarcelle\Sar.fcta
4 -> TRUE

```

On peut ajouter une distance non euclidienne, par exemple l'indice de séparation de niche :

Proportion data	
Input file	Set :Ade4\Dir_Try\Sarcelle\Sar 14 12
Option: Output file	Set
Option: default = between	Set
Distance type (no default)	Set 2

Distance type (no default)	2
1 = d1 Manly = $\text{Sum}(p(i)-q(i))/2$	
2 = Overlap index Manly	
$d2=1-\text{Sum}(p(i)q(i))/\sqrt{\text{Sum}(p(i)^2)}/\sqrt{\text{Sum}(q(i)^2)}$	

```

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\Dir_Try\Sarcelle\Sar_Fre2
It has 91 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\Sarcelle\Sar_Fre2.dma
1 -> 14
2 -> 1
3 -> d2 of MANLY on D:\Ade4\Dir_Try\Sarcelle\Sar
4 -> FALSE

```

Assembler ces quatre distances dans un même fichier :

PasteFiles-SameRow			
Input BIN file 1	Set	4\Dir_Try\Sarcelle\sarxy_EU	91 1
Input BIN file 2 (optional)	Set	\Dir_Try\Sarcelle\Sar.fcli_EU	91 1
Input BIN file 3 (optional)	Set	\Dir_Try\Sarcelle\Sar_MD0fc	91 1
Input BIN file 4 (optional)	Set	e4\Dir_Try\Sarcelle\Sar_Fre2	91 1
Input BIN file 5 (optional)	Set		
Input BIN file 6 (optional)	Set		
Input BIN file 7 (optional)	Set		
Input BIN file 8 (optional)	Set		
Output file	Set	provi	

Output file: provi  
 --- Number of rows: 91, columns: 4

Relire cet assemblage comme un ensemble de quatre matrices de distances :

Read half distance matrix		
Input file	Set	\Ade4\Dir_Try\Sarcelle\provi

```
Test of the euclidean property by diagonalization (theorem of GOWER)
matrix number 1 ---> TRUE
matrix number 2 ---> TRUE
matrix number 3 ---> TRUE
matrix number 4 ---> FALSE
Text file: D:\Ade4\Dir_Try\Sarcelle\provi.dma
1 -> 14
2 -> 4
3 -> Input half distance matrix file D:\Ade4\Dir_Try\Sarcelle\provi
4 -> FALSE
```

Transpose		
Input file	Set	: \Ade4\Dir_Try\Sarcelle\Sar
Output file	Set	SarTR

BioDiv Indices		
dma type file	Set	4\Dir_Try\Sarcelle\provi.dma
Species-Samples array	Set	Ade4\Dir_Try\Sarcelle\SarTR
Index type	Set	0
Output file name	Set	ampli0

File ampli0 contains biodiversity indices  
 Rows (sites): 12 Col (distances): 4  
 Sum(p(i)\*p(j)\*d2(i,j)) without rescaling

Le fichier ampli0 contient la mesure de la dispersion des canards pour chaque mois de quatre points de vue différents. Repasser en texte (AdeTrans), ouvrir avec Excel :

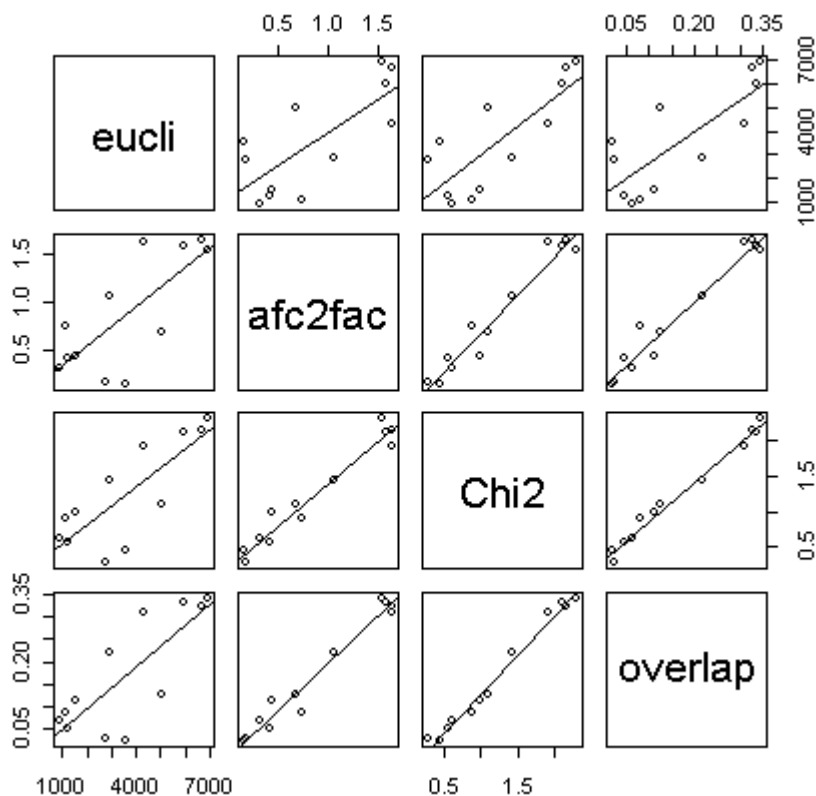
ampli0.txt					
	A	B	C	D	E
1		eucli	afc2fac	Chi2	overlap
2	Juillet	6933	1.53	2.29	0.345
3	Août	5014	0.68	1.09	0.124
4	Septembre	6648	1.64	2.14	0.327
5	octobre	5968	1.57	2.11	0.334
6	Novembre	2900	1.05	1.43	0.219
7	Décembre	1555	0.43	0.99	0.112
8	Janvier	886	0.32	0.61	0.065
9	Février	1255	0.41	0.57	0.047
10	Mars	1113	0.74	0.89	0.083
11	Avril	4310	1.63	1.91	0.310
12	Mai	2771	0.17	0.29	0.027
13	Juin	3566	0.15	0.45	0.020

Sauvegarder en texte et importer dans R :

```
> ampli0_read.table("ampli0.txt",h=T,row.names=1)
> ampli0
```

```
      eucli afc2fac Chi2 overlap
Juillet  6933   1.53 2.29   0.345
Août     5014   0.68 1.09   0.124
Septembre 6648   1.64 2.14   0.327
octobre  5968   1.57 2.11   0.334
Novembre 2900   1.05 1.43   0.219
Décembre 1555   0.43 0.99   0.112
Janvier   886   0.32 0.61   0.065
Février  1255   0.41 0.57   0.047
Mars     1113   0.74 0.89   0.083
Avril    4310   1.63 1.91   0.310
Mai      2771   0.17 0.29   0.027
Juin     3566   0.15 0.45   0.020
```

```
> pairs(ampli0,panel=function(x,y) {points(x,y);abline(lm(y~x))})
```

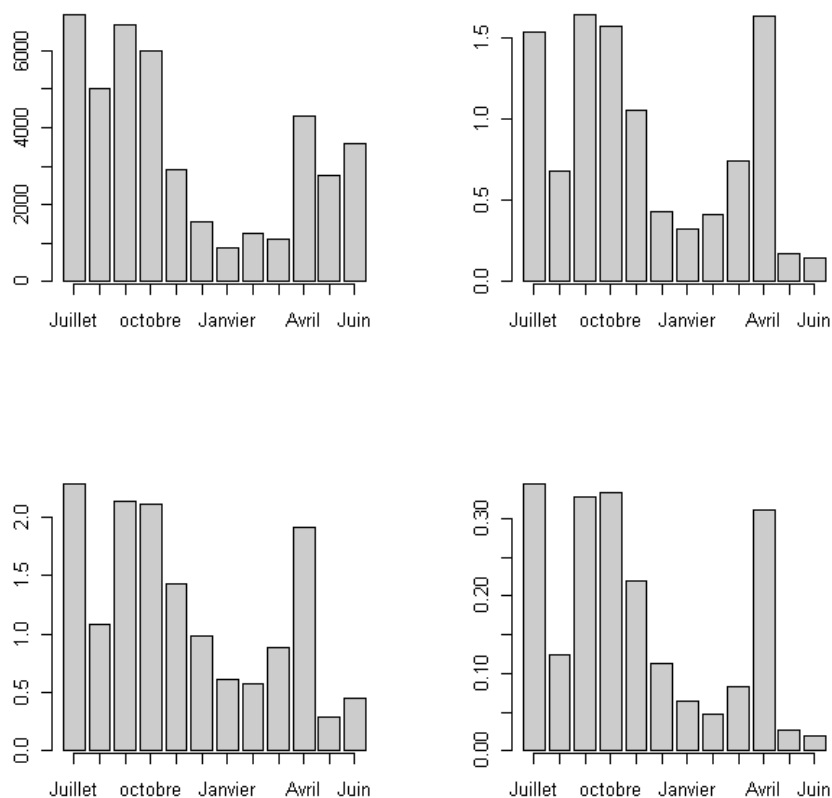


L'utilisateur d'ADE-4 trouvera à <http://pbil.univ-lyon1.fr/R/fichestd/> un paquet de fiches de TD pour apprendre R dans les meilleures conditions.

```
> cor(ampli0)
      eucli afc2fac  Chi2 overlap
eucli  1.0000  0.7484  0.7985  0.7905
afc2fac 0.7484  1.0000  0.9753 0.9791
Chi2    0.7985  0.9753  1.0000  0.9922
overlap 0.7905  0.9791  0.9922  1.0000
```

L'utilisation du tableau de plusieurs manières donne des résultats très voisins. L'espace concret semble le plus cohérent (dispersion maximale à la migration d'automne, faible sur les lieux d'hivernage où s'opère le baguage d'origine et moyenne à la migration de printemps).

```
> par(mfrow=c(2,2))
> apply(ampli0,2,barplot,col=grey(0.8))
```



On peut donc dire que l'idée de mesurer la diversité dans une AFC est légitime. C'est un cas particulier d'une approche très générale mise en œuvre dans ADE-4 dans une structure de calcul ouverte. Avec un facteur, ce cas particulier était considéré dans <sup>8</sup>. Si on conserve  $k$  facteurs on pourra diviser le résultat par 2 fois la somme des valeurs propres des facteurs conservés. Elle peut s'étendre à toute mesure de distances et le cas étudié indique que le choix n'est pas neutre. Plus largement, l'idée fondamentale à retenir est que l'association moyenne-variance n'est pas essentielle. On peut mesurer la variabilité (amplitude, diversité) sans définir de moyenne.



# Références

- <sup>1</sup> Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* : 21, 24-43.
  - <sup>2</sup> Rao, C.R. (1986) Rao's axiomatization of diversity measures. In : *Encyclopedia of Statistical Sciences*, Vol. 7. Kotz, S. & Johnson, N.L. (Eds.) Wiley & Sons, New York. 614-617.
  - <sup>3</sup> Champely, S. & Chessel, D. (2001) Measuring biological diversity using Euclidean metrics. *Environmental and Ecological Statistics* : in press.
  - <sup>4</sup> Thioulouse, J. & Chessel, D. (1992) A method for reciprocal scaling of species tolerance and sample diversity. *Ecology* : 73, 670-680.
  - <sup>5</sup> Usseglio-Polatera, P. & Auda, Y. (1987) Influence des facteurs météorologiques sur les résultats de piégeage lumineux. *Annales de Limnologie* : 23, 1, 65-79.
  - <sup>6</sup> Dolédec, S., Chessel, D. & Gimaret, C. (2000) Niche separation in community analysis: a new method. *Ecology* : 81, 2914-1927.
  - <sup>7</sup> Hoffmann, L. (1960) Untersuchungen an Enten in der Camargue. *Ornithologischer Beobachter* : 57, 35-50.
- Lebreton, J.D. (1973) Etude des déplacements saisonniers des Sarcelles d'hiver, *Anas c. crecca* L., hivernant en Camargue à l'aide de l'analyse factorielle des correspondances. *Compte rendu hebdomadaire des séances de l'Académie des sciences*. Paris, D : III, 277, 2417-2420.
- Auda, Y., Chessel, D. & Tamisier, A. (1983) La dispersion spatiale des Oiseaux au cours du cycle annuel : deux méthodes de description graphique. *Compte rendu hebdomadaire des séances de l'Académie des sciences*. Paris, D : III, 297, 387-392.
- <sup>8</sup> Chessel, D., Lebreton, J.D. & Prodon, R. (1982) Mesures symétriques d'amplitude d'habitat et de diversité intra-échantillon dans un tableau espèces-relevés: cas d'un gradient simple. *Compte rendu hebdomadaire des séances de l'Académie des sciences*. Paris, D : III, 295, 83-88.