

Diversité et typologie biologiques

Résumé

La fiche illustre les mesures de diversité et de dissimilarité biologique dérivées des indices de Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* : 21, 24-43.). Les données traitées sont celles de Blondel, J., Vuilleumier, F., Marcus, L.F. & Terouanne, E. (1984) Is there ecomorphological convergence among mediterranean bird communities of Chile, California, and France. In : *Evolutionary Biology*. Hecht, M.K., Wallace, B. & MacIntyre, R.J. (Eds.) Vol. 18. Plenum Press, New York. 141-213.

Plan

INTRODUCTION.....	2
LIEU D'ALIMENTATION.....	4
REGIME ALIMENTAIRE	11
MORPHOLOGIE	15
CONCLUSION	17
REFERENCES	18

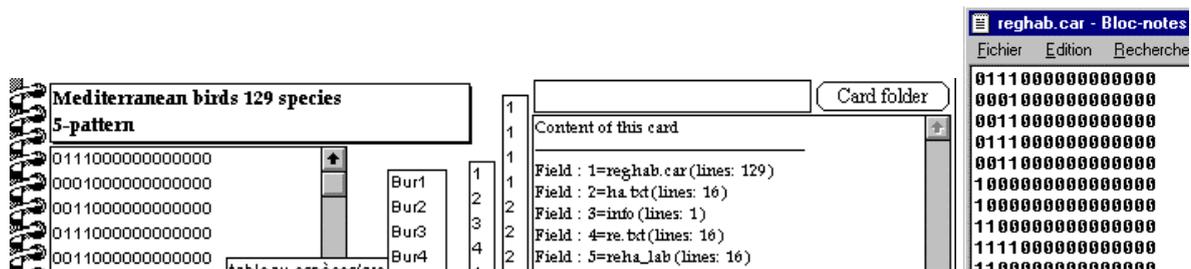
D. Chessel & S. Champely

Introduction

Mesurer la diversité d'une communauté et étudier la différence entre deux communautés sont des problèmes connexes. On utilise ici, pour illustrer des propositions méthodologiques nouvelles, les données de J. Blondel et al. ¹ reproduites dans les cartes de données Ecomor. La problématique sous-jacente est clairement définie dans l'article.

Les données sont basées sur une liste de 129 espèces d'oiseaux présentée dans le tableau 1. Nous utiliserons uniquement le numéro d'ordre (1 à 129). Ces espèces vivent respectivement sur 3 continents (Europe, Amérique du Nord et Amérique du Sud), mais sont présentes ensemble dans l'étude sur la base de leurs traits biologiques. La question est donc : «peut-on comparer des communautés sur la base de liste de taxons entièrement distinctes ? »

Les données traitées sont exactement les données publiées. Seul un recodage est nécessaire pour l'analyse statistique. La liste est classée en quatre groupes étiquetés Burgandy, California, Chile et Provence. Dans le groupe des espèces de Provence, une astérisix indique celle qu'on trouve également en Bourgogne. Ceci impose de prendre quatre indicatrices régionales (variable binaire 0 = absente, 1 = présente). Chaque espèce est dotée d'un profil d'habitats du type 0111, 0001, ..., 1111 indiquant que le taxon est présent ou non dans des formations de quatre types, respectivement 2 = forêt de 1 à 7 ans, 3 = forêt de 9 à 20 ans, 4 = forêts de 40 à 70 ans, 4 = forêt de 71 à 125 ans. Nous utiliserons simplement le code 1|2|3|4 pour désigner ces quatre types de milieu. Nous supposerons que le profil d'habitat d'une espèce présente en Provence et en Bourgogne est le même, puisqu'il n'y a pas d'indications contraires. Ceci indique qu'il y a 16 combinaisons possibles. L'espèce 31 (*Columba fasciata*) n'est présente que dans le milieu 4 de Californie alors que l'espèce 100 (*Streptopelia turtur*) est présente dans les milieux 3 et 4 de Provence et de Bourgogne. Ceci définit 16 communautés théoriques qui sont consignées dans le fichiers reghab.car de la carte ecomor+3.



La typologie des communautés sur la base de leur liste d'espèces ne ferait qu'indiquer que les cortèges sont entièrement distinctes dans trois groupes. Elle n'a pas de sens et c'est ce qui fait tout l'intérêt de ces données.

Cependant, pour voir le comportement de la méthode, nous pouvons calculer une distance entre espèces sur la base de leur distribution. On prend la distance d'Edwards ² entre deux distributions de fréquence, utilisée en génétique ³ :

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{1 - \sum_i \sqrt{p_i q_i}}$$

Pour des données binaires c'est exactement la distance associée à l'indice de similarité d'Ochiai⁴ :

$$S = \frac{a}{\sqrt{(a+b)(a+c)}}$$

où a est le nombre de communautés où sont présentes les deux espèces alors que b et c sont les nombre de communautés où on ne rencontre que l'une des deux. La distance est alors $D = \sqrt{1 - S}$ et elle est euclidienne⁵.

Binary Dissimilarity

Input file: 129 16

Option: Output file:

Option: default - between:

Similarity coefficient (no):

Principal Coordinates

dma type file:

Stars

XY coordinates file: 129 2

X-axis column number (default):

Y-axis column number (default):

Categories file (.cat):

Label file (or #) for variables:

Label file (or #) for items:

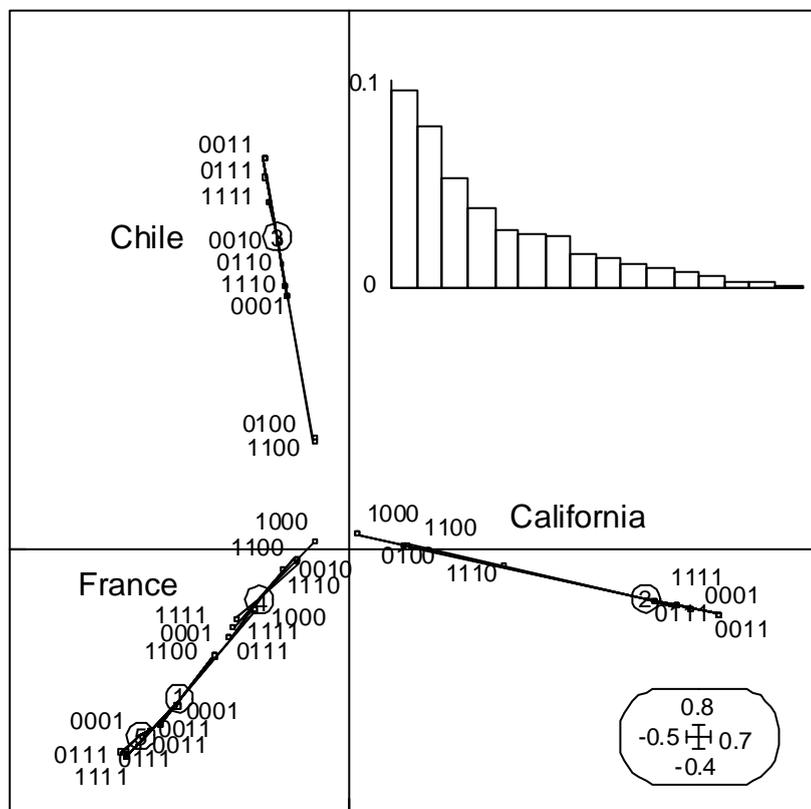


Fig 1 : Représentation euclidienne du tableau reghab contenant la liste des espèces présentes dans les 16 communautés. Classes définies par la variable groupe. Etiquettes définies par la variable habitat.

Nous ne devons utiliser que l'information biologique fournie par les auteurs. On reproduit simplement le calcul du nombre d'espèces. Les représentations graphiques utilisent deux variables. La première re donne pour chaque communauté le numéro de la région (16-1 à 4 modalités) et la seconde re donne le numéro de l'habitat (16-1 à 4 modalités). reha_lab donne les étiquettes des 16 communautés (même carte).

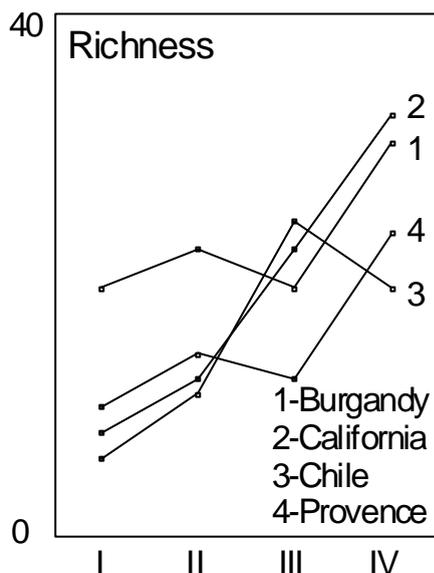


Fig 2 : Richesse spécifique dans les 16 communautés.

Lieu d'alimentation

Trois groupes de variables décrivent la biologie des espèces. Le premier (Foraging substrate) définit le lieu d'alimentation. Il est donné par un code littéral F-G, Tr-F, ..., G-B-F avec les codes 1=F=foliage feeders, 2=G=ground feeders, 3=T=twig feeders, 4=B=bush feeders, 5=Tr=trunk feeders et 6=A=aerial feeders. Le fichier forsub donne cette information avec 6 indicatrices binaires (carte Ecomor) :

Mediterranean birds 129 species						F	
1-foraging substrate						G	
						T	
						B	
						Tr	
						A	
1	1	0	0	0	0		
1	0	0	0	1		forsub.txt	
0	0	1	0	1		foraging substrates	
1	0	0	0	1		1=F=foliage feeders	
0	1	0	0	1		2=G=ground feeders	
0	0	0	0	1		3=T=twig feeders	
0	0	0	0	0	1	4=B=bush feeders	
0	0	0	0	0	1	5=Tr=trunk feeders	
0	0	0	0	0	1	6=A=aerial feeders	

La variable groupe code les taxons en 5 classes (1=Bourgogne, 2=Californie, 3=Chili, 4=Provence et 5=Mixte Bourgogne+Provence).

Les résultats sont totalement confirmés. Il y a variation significative du profil de comportement alimentaire des espèces entre communautés. Entre régions ? Entre habitats ? Entre zone témoin (Bourgogne) et zone méditerranéenne (Provence_Chili-Corse) ? Toutes les questions posées dans l'article étudié sont ouvertes.

La distance biologique (entre profils des lieux d'alimentation) et le profil en espèces par communauté génèrent une nouvelle notion de diversité pour les communautés (variation dans une communauté des propriétés biologiques des espèces) et une nouvelle typologie des communautés (variation entre communautés des propriétés biologiques des espèces) . Le premier point donne :

BioDiv Indices		
dma type file	Set	OR\CalADE\forsub_Sim7.dma
Species-Samples array	Set	e4\ECOMOR\CalADE\reghab 129 16
Index type	Set	0
Output file name	Set	divforsub0

Distances - Input file: E:\Ade4\ECOMOR\CalADE\forsub_Sim7.dma
 Text file: E:\Ade4\ECOMOR\CalADE\forsub_Sim7.dma
 1 -> 129
 2 -> 1
 3 -> S12 index of GOWER & LEGENDRE on E:\Ade4\ECOMOR\CalADE\forsub
 4 -> TRUE

Species-Sites array - Input file: E:\Ade4\ECOMOR\CalADE\reghab
 Species: 129 Sites: 16

Option: 0
 File divforsub0 contains biodiversity indices
 Rows (sites): 16 Col (distances): 1
 Sum(p(i)*p(j)*d2(i,j)) without rescaling

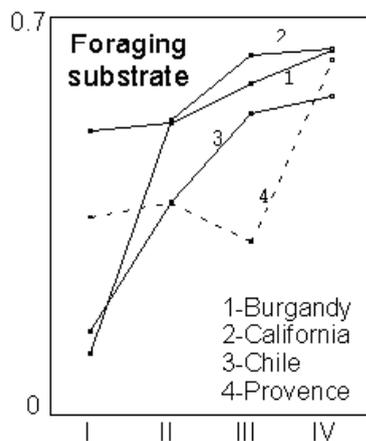


Fig 3 : Diversité biologique (lieu d'alimentation) dans les 16 communautés.

La diversité biologique ne se réduit donc pas au nombre d'espèces. On voit la variation de diversité en représentant les profils des communautés sur le plan de la représentation euclidienne de la distance biologique.

Principal Coordinates		
dma type file	Set	OR\CalADE\forsub_Sim7.dma

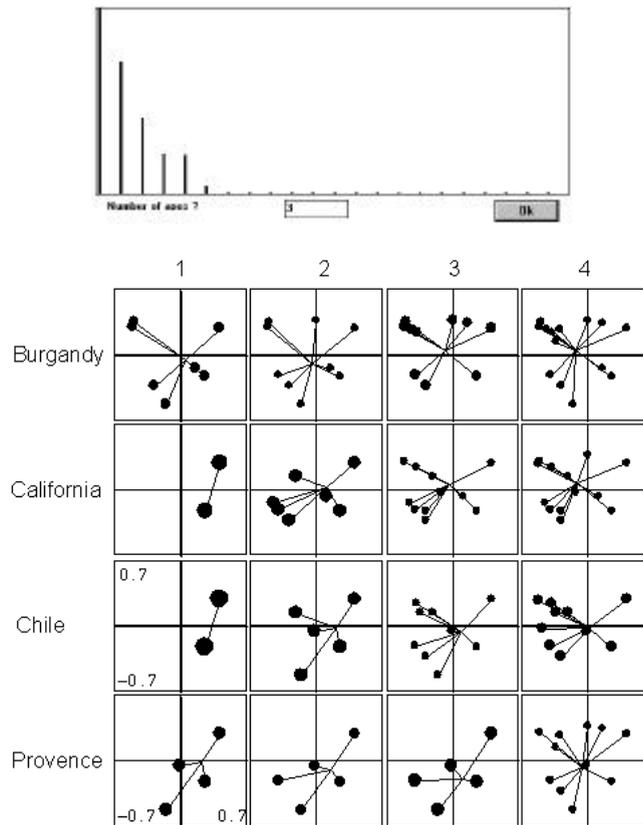


Fig 4 : Profils spécifiques des 16 communautés dans le plan 1-2 de la représentation euclidienne des espèces pour leur profil de lieu d'alimentation. Il y a variation d'amplitude (diversité) et de position (contenu) d'abord entre habitats et non entre régions.

La figure est obtenue dans Scatterdistri :

Stars		
XY coordinates file	Set	R\CalADE\forsub_Sim71.coo
X-axis column number (default	Set	
Y-axis column number (default	Set	
G values file	Set	e4\ECOMOR\CalADE\reghab

Le fond de carte se comprend par (Scatters, figure 5) :

Labels		
XY coordinates file	Set	R\CalADE\forsub_Sim71.coo
X-axis column number (default	Set	
Y-axis column number (default	Set	
Label file (or #) for items	Set	#

La distance entre communautés basée sur la distance entre espèces est alors calculée par :

Euclidean Dissimilarity Coefficient		
Species-Samples array	Set	e4\ECOMOR\CalADE\reghab 129 16
dma type file	Set	OR\CalADE\forsub_Sim7.dma
Column number (default=1)	Set	
Output file name	Set	edcforsub

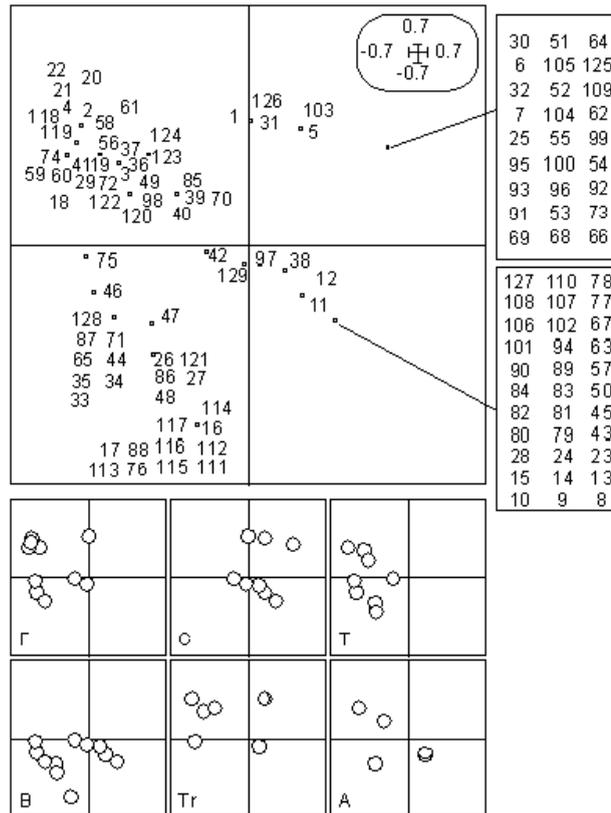


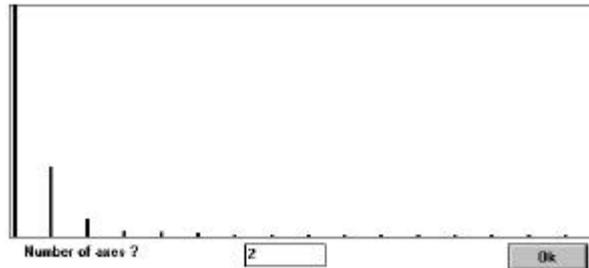
Fig 5 : Plan 1-2 de la représentation euclidienne des espèces pour leur profil de lieu d'alimentation. Dans les forêts très jeunes, les espèces qui s'alimentent au sol ou en vol sont largement sur-représentées.

```
Distances - Input file: E:\Ade4\ECOMOR\CalADE\forSub_Sim7.dma
Column used: 1
Text file: E:\Ade4\ECOMOR\CalADE\forSub_Sim7.dma
1 -> 129
2 -> 1
3 -> S12 index of GOWER & LEGENDRE on E:\Ade4\ECOMOR\CalADE\forSub
4 -> TRUE
```

```
Species-Sites array - Input file: E:\Ade4\ECOMOR\CalADE\reghab
Species: 129 Sites: 16
Output file: edcforsub
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: edcforsub.dma
1 -> 16
2 -> 1
3 -> edc from E:\Ade4\ECOMOR\CalADE\reghab
and E:\Ade4\ECOMOR\CalADE\forSub_Sim7.dma
4 -> TRUE
```

La nouvelle distance est euclidienne et définit ne nouvelle représentation inter-communautés :





Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.3370E-04	+0.9142	+0.9142	02	+1.1821E-05	+0.0808	+0.9950
03	+6.6570E-07	+0.0046	+0.9996	04	+3.1805E-08	+0.0002	+0.9998
05	+2.7558E-08	+0.0002	+1.0000	06	+4.9006E-09	+0.0000	+1.0000

ERREUR DE COPIE 71+21

Le premier axe récupère 91 % de la variabilité et sera seul interprété par :

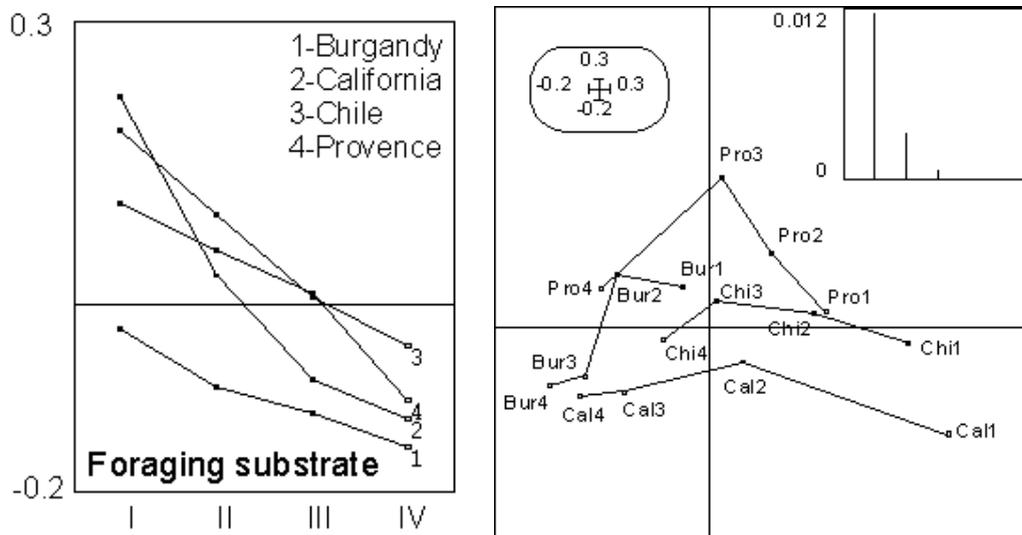


Fig 6 : Typologie biologique (lieu d'alimentation) des 16 communautés.

Le résultat est extraordinaire de précision. La reproductibilité du lien formation forestière / trait biologique entre les zones méditerranéennes est indéniable et le témoin est franchement à part. La convergence inter-continentale se fait donc sur le lien espèces-environnement.

Régime alimentaire

Trois groupes de variables décrivent la biologie des espèces. Le second (Diet category) définit le type d'alimentation. Il est donné par un code littéral Gr, ..., In-Fr-Li, ..., Fr-Ca-Ch avec les codes Gr = seeds (granivorous), Fr = berries, acorns, drupes (frugivorous), Ne = nectar (nectarivorous), Fo = leaves (folivorous), In = insects, spiders, myriapods, isopods, snails, worms (invertebrate feeder), Ca = flesh of small vertebrates (carnivorous), Li = invertebrates in fresh water (limnivorous) et Ch = carrion feeder. Le fichier diet donne cette information avec 8 indicatrices binaires (carte Ecomor+2) :

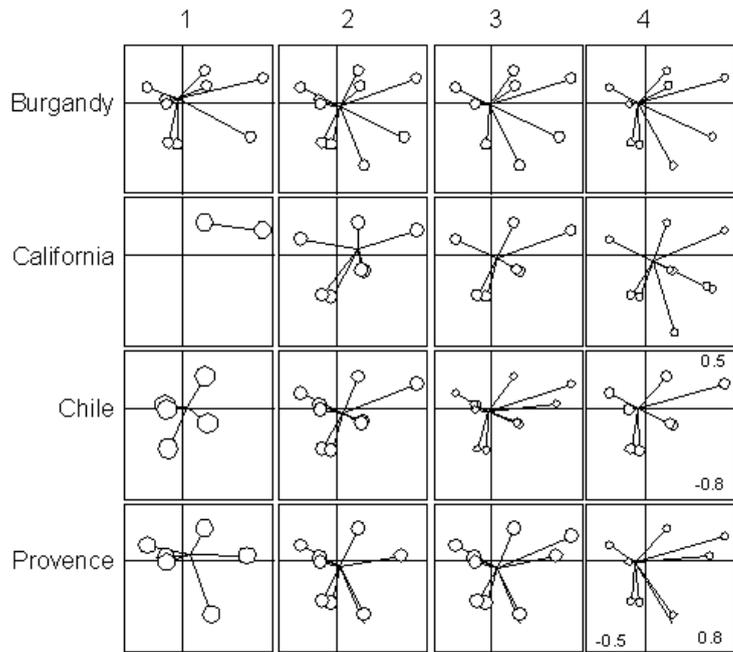


Fig 8 : Profils spécifiques des 16 communautés dans le plan 1-2 de la représentation euclidienne des espèces pour leur profil detype d'alimentation. L'originalité du premier habitat de Californie est marquée.

Les variations d'amplitude sont moins sensibles, comme les variations de position sont moins significatives.

Euclidean Dissimilarity Coefficient

Species-Samples array e4\ECOMOR\CalADE\reghab | 29 16

dma type file MOR\CalADE\diet_Sim7.dma

Column number (default=1) _____

Output file name edcdiet

Principal Coordinates

dma type file COMOR\CalADE\edcdiet.dma

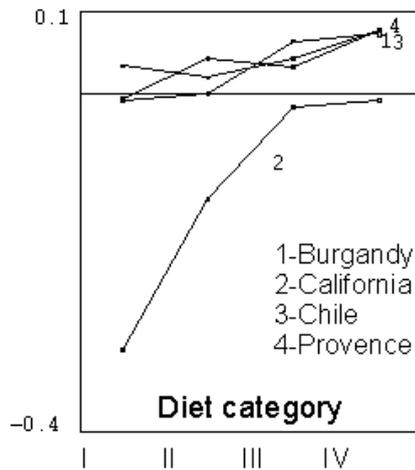
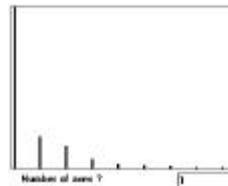


Fig 9 : Typologie biologique (lieu d'alimentation) des 16 communautés.

L'essentiel se résume à l'originalité d'une région. Les deux types de variables très proches dans leur forme donnent des résultats très différents et la méthode utilisée offre un grand nombre de possibilités.

Morphologie

Trois groupes de variables décrivent la biologie des espèces. Le troisième définit la morphologie des espèces par un tableau de 8 variables quantitatives (carte Ecomor+1) :

The screenshot shows the DATA STA software interface. The main window displays a data table with the following content:

Mediterranean birds 129 species					
2-habitats		3-morphology			
0	1-7	255	173	24.7	11
0	9-20	11	26.2	47	500
0	40-70	89.5	51.5	4.8	
0	71-125	6.9	12.8	21	

Labels 'Wing' and 'Tail' are positioned below the 4th and 5th columns respectively. To the right, the 'Content of this card' panel lists:

- Field : 1=habitat.txt (lines: 129)
- Field : 2=morpho.txt (lines: 129)
- Field : 3=habi_lab (lines: 4)
- Field : 4=morpho_lab (lines: 8)

On ne discutera que de la forme en utilisant un changement de variables en log (multiplié par 1/3 pour le poids, comme dans l'article cité). Les variables sont Wing, Tail, Culmen, Bill_height., Bill_width, Tarsus, Midtoe et Weight.

The screenshot shows the 'Mantel-RV test' configuration window with the following settings:

- First dma type file: R:\CalADE\reghab_Sim7.dma
- Matrix number: [empty]
- Second dma type file: a\ADE\morpholog_MDcc.dma
- Matrix number: [empty]
- Permutation number: 1000

```
Correlation between two distance matrices
First input file: E:\Ade4\ECOMOR\CalADE\reghab_Sim7.dma
Text file: E:\Ade4\ECOMOR\CalADE\reghab_Sim7.dma
 1 -> 129
 2 -> 1
 3 -> S12 index of GOWER & LEGENDRE on E:\Ade4\ECOMOR\CalADE\reghab
 4 -> TRUE
Matrix used : 1
Second input file: E:\Ade4\ECOMOR\CalADE\morpholog_MDcc.dma
Text file: E:\Ade4\ECOMOR\CalADE\morpholog_MDcc.dma
 1 -> 129
 2 -> 1
 3 -> Euclidean distance from triplet E:\Ade4\ECOMOR\CalADE\morpholog.ccta
 4 -> TRUE
Matrix used : 1
Permutation test on r value (Manly 1991 p. 114)
r index : 1.562e-02
number of random matching: 1000   Observed: 0.015618
Histogramm: minimum = -0.045857, maximum = 0.058290
number of simulation X<Obs: 809 (frequency: 0.809000)
number of simulation X>=Obs: 191 (frequency: 0.191000)
```


On retrouve une valeur exceptionnelle dans les habitats jeunes de Californie.

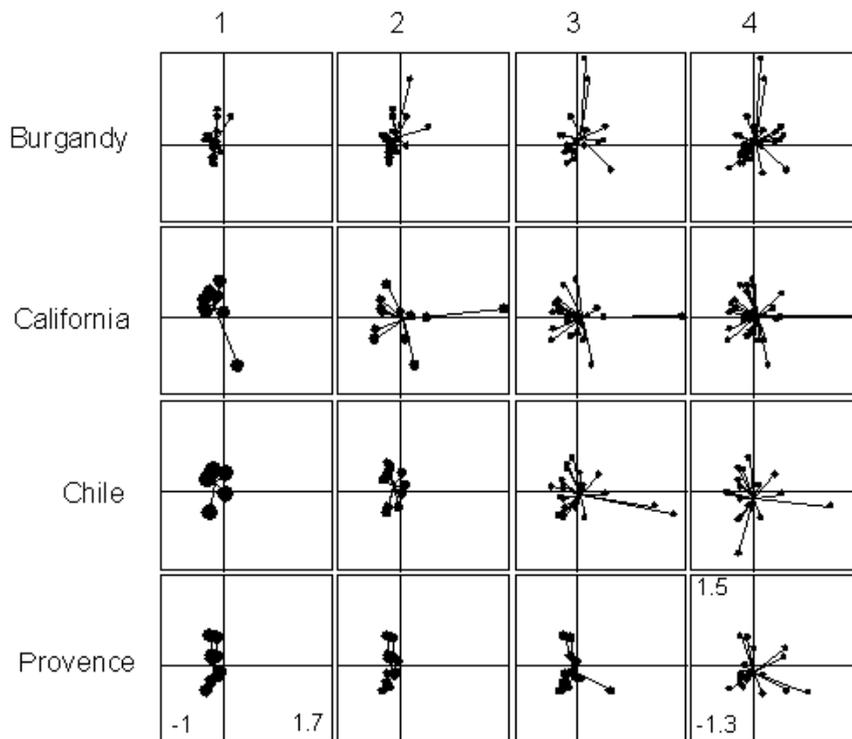


Fig 11 : Profils spécifiques des 16 communautés dans le plan 1-2 de la représentation euclidienne des espèces pour leur profil morphologique.

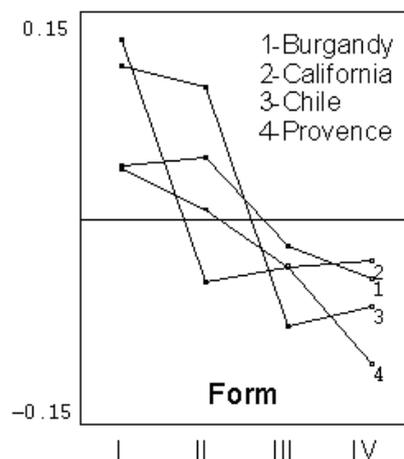


Fig 9 : Typologie biologique (lieu d'alimentation) des 16 communautés.

Le premier axe de l'analyse de la distance morphologique entre communautés exprime la stabilité de la relation forme / habitat dans les quatre régions et , dans ce cas, il n'y a pas de convergence méditerranéenne.

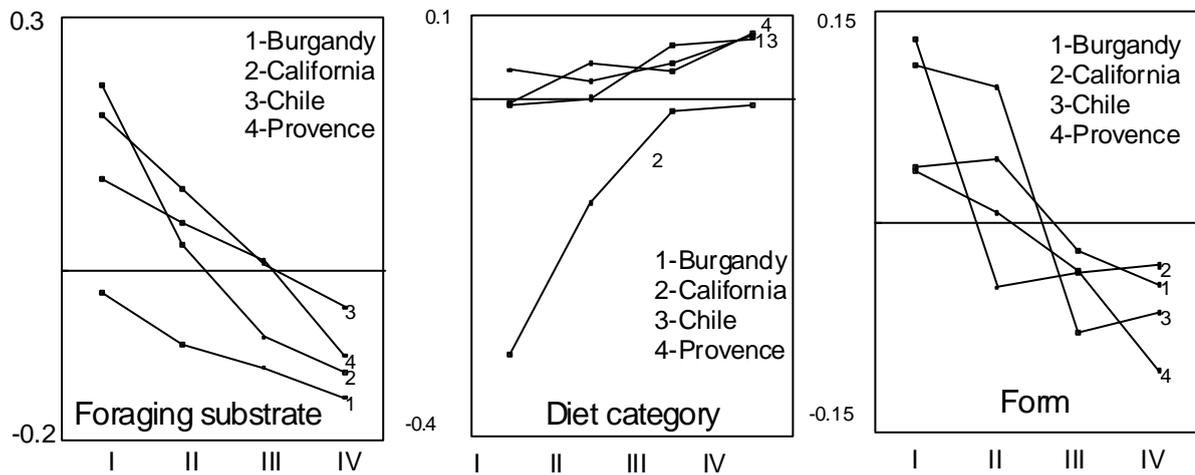
Conclusion

La première conclusion nette est que la diversité biologique calculée par application de l'indice général de Rao ⁷ n'est pas simplement la diversité taxonomique, c'est-à-dire globalement le nombre d'espèces.



Un même niveau de diversité est obtenu pour des richesses différentes, un même niveau de richesse peut correspondre à des diversités très différentes.

La seconde conclusion touche au test de Mantel trois fois pris en défaut par le test RV. L'interprétabilité des variations inter communauté est remarquable tant par la simplicité que par la diversité des résultats et conforte grandement l'existence de la signification statistique. Les distances de Rao, totalement complémentaire des mesures de diversité sont fonctionnelles. La plus originale des quatre régions est la Californie mais les convergences entre cortèges faunistiques totalement ou partiellement distincts peuvent être mises en évidence. Ceci valide l'expérience et pourrait permettre de compléter l'interprétation du ressort des spécialistes, évidemment.



Une partie de ces illustrations sont dans ⁸ dans lequel on trouvera les justificatifs théoriques.

Références

¹ Blondel, J., Vuilleumier, F., Marcus, L.F. & Terouanne, E. (1984) Is there ecomorphological convergence among mediterranean bird communities of Chile, California, and France. In :

Evolutionary Biology. Hecht, M.K., Wallace, B. & MacIntyre, R.J. (Eds.) Vol. 18. Plenum Press, New York. 141-213.

² Edwards, A.W.F. (1971) Distance between populations on the basis of gene frequencies. *Biometrics* : 27, 873-881.

³ Hartl, D.L. & Clark, A.G. (1989) *Principles of population genetics*. Sinauer Associates, Sunderland, Massachusetts. 1-682 (p. 303).

⁴ Ochiai, A. (1957) Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. *Bulletin of the Japanese Society of Scientific Fisheries* : 22, 526-530.

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

⁶ Yoccoz, N. G. (1993) Morphométrie et analyses multidimensionnelles. Une revue des méthodes séparant taille et forme. In : *Biométrie et Environment*. Lebreton, J.D. & Asselain, B. (Eds.) Masson, Paris. 73-99.

⁷ Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* : 21, 24-43.

⁸ Champely, S. & Chessel, D. (2001) Measuring biological diversity using Euclidean metrics. *Environmental and Ecological Statistics* : in press.