

ESA CNRS 5023
Écologie des Eaux Douces et des Grands Fleuves
Université Claude Bernard Lyon I
Bâtiment 401C
69622 Villeurbanne CEDEX

DIPLÔME D'ÉTUDES APPROFONDIES :
«Analyse Et Modélisation Des Systèmes Biologiques»

UTILISATION DES LISTES D'OCCURRENCES
SPÉCIFIQUES SPATIALISÉES
EN ÉCOLOGIE ET EN BIOGÉOGRAPHIE

- Rapport bibliographique -

DRAY Stéphane

Directeur de recherche : Daniel CHESSEL

Parrains : Roland ALLEMAND
Jean-Pierre PASCAL

Janvier 1999

SOMMAIRE

INTRODUCTION.....	2
I. PRESENTATION DES DONNEES	5
I.1. Origine des données.....	5
I.2. Originalité des données	7
II. ESTIMATION DE L'AIRE GEOGRAPHIQUE DE DISTRIBUTION.....	10
II.1. Distance linéaire (Fig. 7A)	10
II.2. Aire contenue dans les limites (Fig. 7B, 7C)	11
II.3. Nombre de surfaces élémentaires occupées (Fig. 7D)	11
III. ASSEMBLAGE D'ESPECES : ESTIMATION DE LA DIVERSITE	14
III.1. Un grand nombre d'indices	14
III.2. Application aux listes d'occurrences.....	15
IV. RELATIONS ESPECES-MILIEU.....	19
IV.1. Le concept de niche multidimensionnelle	19
IV.2. Richesse spécifique et facteurs environnementaux.....	20
IV.3. Composition spécifique et facteurs environnementaux.....	22
CONCLUSIONS ET PERSPECTIVES	26
BIBLIOGRAPHIE.....	28

Introduction

Pour commencer ce rapport, quelques définitions tirées du dictionnaire "Le petit Larousse 1994" :

biogéographie : n. f. Étude de la répartition des espèces vivantes (végétales et animales) et des causes de cette répartition.

écologie : n. f. (du gr. *oikos*, maison, et *logos*, science) : Science qui étudie les relations des êtres vivants entre eux et avec leur milieu.

statistique : n. f. (du lat. *status*, état) : Ensemble de méthodes mathématiques qui, à partir du recueil et de l'analyse de données réelles, permettent l'élaboration de modèles probabilistes autorisant les prévisions.

Ces différentes disciplines sont intimement liées lorsque l'on analyse des données concernant la répartition spatiale d'espèces car il existe des relations entre les espèces d'une part et entre les espèces et leur milieu d'autre part. Faire des statistiques nécessite des interrogations sur le statut des données à traiter et notamment sur leur indépendance. De par la complexité du monde vivant cette étape est primordiale en écologie et en biogéographie. Est-ce que les individus d'une espèce se répartissent de façon aléatoire ou existe-t-il des facteurs influençant cette répartition ? L'existence de multiples interactions entre les êtres vivants et leur milieu pose le problème de la part d'aléatoire et de déterminisme lors de la réalisation d'un événement. Dans la théorie du chaos, Gleick (1989) nous dit que "*Pour comprendre la complexité de la nature, il fallait soupçonner que cette complexité n'était pas seulement un hasard, un accident.*". Tout événement réalisé a donc des causes et des conséquences. Ce concept déterministe est le fondement même de l'écologie d'aujourd'hui. En effet, l'écologie a d'abord considéré les **espèces individuelles** (écologie du Chêne, écologie du Renard...) : on dit aujourd'hui qu'il s'agit **autécologie** - science des réponses des espèces aux facteurs de l'environnement, en fonction de leur physiologie et de leurs adaptations respectives. Cette écologie était, au départ, supposée expliquer la répartition des espèces à toutes les échelles d'observations (depuis les "microrépartitions" jusqu'à la répartition planétaire, en passant par les "mésorépartitions" dues au climat local, à l'altitude ...).

Rapidement, il a fallu introduire parmi les facteurs de répartition d'une espèce la présence d'autres espèces dans son voisinage. On s'est en effet rendu compte que les espèces interagissaient entre elles : relations prédateurs-proies, de compétitions ; intervention dans les phénomènes de reproduction telles que le transport des éléments sexuels de certaines plantes par certains animaux,

etc... Est alors née une écologie de ces interactions, et une écologie des **ensembles d'espèces**, dite **synécologie**.

On s'est avisé enfin que les organismes et les populations **modifient le milieu environnant**, suscitant de nouvelles interactions (indirectes cette fois) entre espèces : modification du milieu par des substances chimiques excrétées, espèces consolidant un sol et rendant possible l'installation d'autres espèces, etc. On aboutit dès lors à la vision d'un système d'interactions complexes des espèces entre elles et avec le milieu. Ce système a été nommé **l'écosystème**. L'ensemble des individus d'une même espèce, vivant à un endroit donné, forme une **population**, et l'ensemble des populations habitant ensemble un même milieu caractérise une **communauté**. En fin de compte la notion d'écosystème peut être symbolisé comme suit (Fig. 1) :

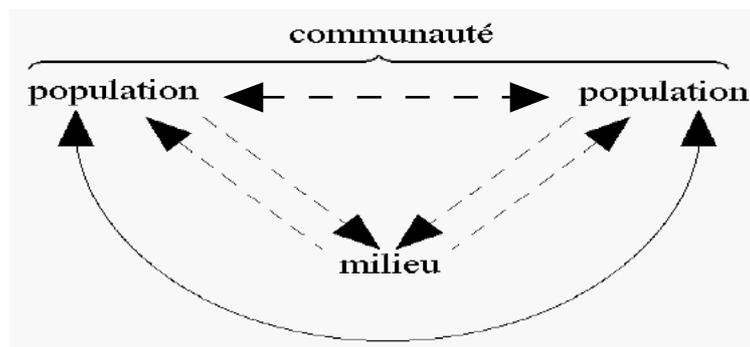


Figure 1 : Schéma symbolique d'un écosystème. Les flèches en pointillé représentent les interactions directes, celles en trait plein représentent les interactions indirectes.

L'étude des patterns spatiaux et temporels des organismes ainsi que le développement des bases de données permettant le stockage d'une grande quantité d'information, a ravivé l'exploitation de données biogéographiques existantes. Ainsi, les occurrences d'espèces tirées de la littérature, d'atlas ou d'herbier sont couramment utilisées pour analyser les patterns multispécifiques, tels que la richesse spécifique ou l'aire géographique de distribution. Ces occurrences résultent de rencontres indépendantes entre un organisme et, par exemple, un botaniste en un lieu et à une date donnée. Par conséquent, ces données sont sensiblement différentes de celles venant d'un relevé classique dans lequel, un inventaire minutieux des espèces est réalisé pour une station. Lors du premier congrès international de botanique en 1900 à Paris, Flahault définit la station comme *"une circonscription d'étendue quelconque, mais le plus souvent restreinte, représentant un ensemble complet et défini de conditions d'existence. La station résume tout ce qui est nécessaire aux espèces qui l'occupent, la combinaison des facteurs climatiques et géographiques avec les facteurs édaphiques et biologiques, c'est-à-dire les rapports de chaque espèce avec le sol et avec les espèces auxquelles elle est associée."* La présence d'une espèce A dans un relevé est donc corrélée à la présence d'une espèce B avec laquelle A entretient des interactions directes ou indirectes. La station est donc le reflet d'un écosystème ou d'une partie d'un écosystème. Par contre, dans une liste d'occurrences, toutes les citations sont indépendantes, et la rencontre entre le botaniste et la plante est le résultat de la présence croisée du botaniste et de la plante en cette occasion particulière.

Les statuts des données issues de relevés ou des listes d'occurrences ne sont donc pas similaires ; il devrait alors en être de même pour les méthodes destinées à leur analyse. Dans ce rapport, nous discuterons donc de l'exploitation des listes d'occurrences en écologie et en biogéographie ainsi que des problèmes méthodologiques que ce type de données peut poser.

I. Présentation des données

I.1. Origine des données

Les bases de données de spécimens végétaux ou animaux existent depuis des siècles sous forme de plantes séchées dans des herbiers, spécimens préservés dans les Muséums d'Histoire Naturelle et dans des collections privées, ou information dans les registres associés aux spécimens (cahiers de terrain). Avec le développement des ordinateurs, le stockage et le traitement de grands jeux de données sont devenus beaucoup plus aisés. Ainsi, ces bases de données initialement associées aux travaux concernant la systématique des organismes, sont maintenant exploitables dans d'autres domaines tels que l'écologie ou la biogéographie. Hawksworth (1995) estime à environ 280 millions le nombre d'exemplaires contenus dans les collections des Muséums à travers le monde. Cette quantité astronomique d'information peut être le support privilégié de travaux s'intéressant à la répartition spatiale d'espèces animales ou végétales car chaque échantillon est, dans l'idéal, identifié par le nom du taxon, l'auteur de la citation, la date et le site de la collecte. D'autres renseignements relatifs au milieu, ou à la technique de capture, par exemple, peuvent également être indiqués (Fig. 2) :

N° Bordereau :	<input type="text" value="1"/>	Date de Saisie :	<input type="text" value="3 Octobre 1996"/>	N° :	<input type="text" value="1"/>
N° Boîte :	<input type="text" value="464033"/>	Rédacteur :	<input type="text" value="Bouget Christophe"/>	Nom Collection :	<input type="text" value="VIALIER"/>
Lieu de Saisie :	<input type="text" value="Muséum de Lyon"/>	Lieu Conservation :	<input type="text" value="MUSEUM DE LYON"/>	Auteur Citation :	<input type="text"/>
Valideur :	<input type="text" value="Coulon Jacques"/>	Année Citation :	<input type="text"/>	<input type="button" value="Référence"/>	
Déterminateur :	<input type="text"/>	Auteur Observation :	<input type="text"/>		
Collecteur :	<input type="text" value="Schaefer L."/>				
Espèce :	<input type="text" value="limbatum"/>	Commune :	<input type="text" value="AMELIE-LES-BAINS-PALALDA"/>	INSEE :	<input type="text" value="66003"/>
Genre :	<input type="text" value="Omophron"/>	Dépt. :	<input type="text" value="Pyrénées-Orientales"/>	N° Dépt. :	<input type="text" value="66"/>
Mode Collecte :	<input type="text"/>	Station :	<input type="text"/>	Biotope :	<input type="text"/>
Stade :	<input type="text" value="I (imago)"/>				
Statut Donnée :	<input type="text" value="C (collection)"/>				
Jour : -= 6 +	Nombre : <input type="text" value="4"/>	Altitude : -= +	Carré_Fr :	<input type="text"/>	<input type="button" value="Revoir"/>
Mois : -= 7 +	Mâles : <input type="text" value="0"/>		Carré_Rég :	<input type="text"/>	
Année : <input type="text" value="1949"/>	Femelles : <input type="text" value="0"/>		Carré_Dépt :	<input type="text"/>	

Figure 2 : Fenêtre d'entrée des occurrences du logiciel Biogeographica (L. Delaunay, A.F.C.C.F.). Le rédacteur entre le nom de l'espèce, la date, la localité, le nom du collecteur. D'autres renseignements peuvent être indiqués : le mode de collecte, le biotope, l'altitude...

La compilation de ces données conduit à des listes d'occurrences où chaque ligne correspond à une rencontre entre, par exemple, un entomologiste et un insecte en un lieu précis et à une date donnée (Fig. 3) :

Liste des Collectes							
17764 / 17764							
Bord	Fiche	Genre	Espèce	Localité	Collecteur	Date	
○	1	Ornophron	limbatum	66 AMELIE-LES-BAINS-PALALDA	Schaefer L.	06.07.1949	
○	1	Ornophron	limbatum	38 ASSIEU	Viallier J.	24.04.1980	
○	1	Ornophron	limbatum	38 ASSIEU	Viallier J.	18.04.1982	
○	1	Calosoma	sycophanta	30 NIMES	Thérond J.	10.06.1947	
○	1	Calosoma	sycophanta	30 UZES	Audras G.	06.1963	
○	1	Calosoma	sycophanta	69 TASSIN-LA-DEMI-LUNE	Viallier J.	21.06.1937	
○	1	Calosoma	sycophanta	13 VITROLLES	Viallier J.	28.05.1975	
○	1	Calosoma	sycophanta	34 GANGES	Viallier J.	25.05.1959	
○	1	Calosoma	sycophanta	34 SAINT-GUILHEM-LE-DESERT	Viallier J.	.	
○	1	Calosoma	sycophanta	30 DURFORT-ET-SAINT-MARTIN-DE-SOSSEN	Viallier J.	26.06.1988	
○	1	Calosoma	sycophanta	13 VITROLLES	Viallier J.	28.05.1975	
○	1	Calosoma	sycophanta	34 SAINT-GUILHEM-LE-DESERT	Viallier J.	25.06.1984	
○	1	Calosoma	sycophanta	34 SAINT-MARTIN-DE-LONDRES	Viallier J.	27.06.1988	

Figure 3 : Liste des occurrences (extrait de Biogeographica) de carabiques. Chaque occurrence est définie par un numéro, le nom de l'espèce, le nom du collecteur, la localité et la date de la collecte.

Pour une espèce végétale, les coordonnées spatiales peuvent identifier un individu qui est fixe.

Quand il s'agit d'espèces animales mobiles, cette évidence n'est plus vraie mais, on peut considérer que l'individu n'occupe pas un endroit au hasard : il se situe dans un milieu satisfaisant ses exigences de niche. L'utilisation des coordonnées spatiales est donc valide.

L'exploitation de listes d'occurrences en écologie ou en biogéographie nécessite tout de même de s'interroger sur les motivations d'un collecteur afin de mieux appréhender les problèmes qui peuvent se poser. Par exemple, Donald & Fuller (1998) nous précisent les buts des atlas ornithologiques :

"Most atlases are compiled using data collected by large numbers of volunteer fieldworkers and the end product must therefore be designed at least in part to fulfil their wishes and needs.(...) Although this review concentrates on the scientific uses of atlas data, it must be remembered that the educational and recreational value of atlases alone justifies their production and that the greatest use of atlas data is made by non-scientists". Les volontaires participant à ce type d'atlas sont pour la

plupart des amateurs plus ou moins confirmés. Sachant que la reconnaissance des différentes espèces d'oiseaux est basée essentiellement sur le chant, un observateur identifiera plus facilement le chant d'une espèce qui l'intéresse ou d'une espèce commune que celui d'autres espèces. Pour les entomologistes, l'attrait du spécimen rare provoque certainement un délaissement des espèces communes. En botanique, Rich & Woodruff (1992) montre que les occurrences de deux plantes rares sont étroitement liées à la présence de deux botanistes spécialistes de ces plantes (Fig. 4).

Ainsi, l'existence d'une occurrence résulte à la fois de la probabilité qu'il y ait une rencontre entre un collecteur et un spécimen (ce qui dépend de l'abondance de l'espèce), mais également des motivations et des capacités du collecteur (Fig. 5).

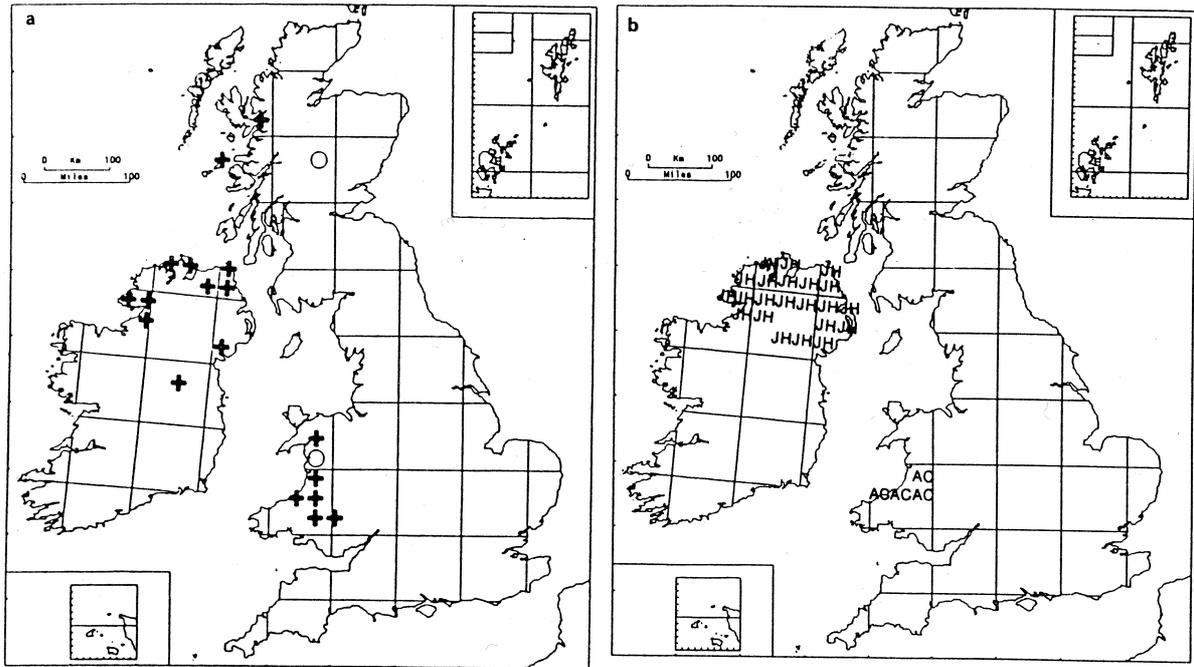


Figure 4: Corrélation entre (a) la distribution de *Carex hostiana x viridula* et (b) les aires explorés par deux botanistes (AC = A. O. Chater, JH = J. Harron) d'après Rich & Woodruff (1992)

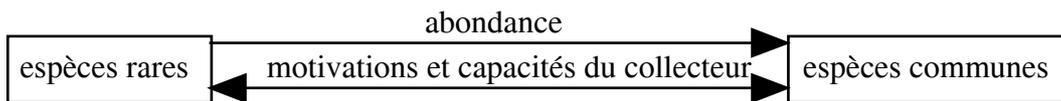


Figure 5 : Schéma représentant les facteurs influençant l'existence d'une occurrence. L'abondance d'une espèce commune augmente la probabilité que le collecteur rencontre un individu de cette espèce. Les motivations et les capacités du collecteur peuvent augmenter ou diminuer la probabilité d'existence d'une occurrence.

I.2. Originalité des données

Les listes d'occurrences diffèrent des relevés faunistiques classiques car elles synthétisent un ensemble de rencontres indépendantes les unes des autres, alors que dans un relevé on s'intéresse aux espèces trouvées dans une zone préalablement définie par un protocole expérimental. Dans de nombreuses études, les listes d'occurrences sont résumées grâce à une partition de l'espace en quadrats de taille variable par superposition d'une grille sur une carte de distribution (Tab. 1). Les critères servant à choisir la taille du maillage sont généralement flous et subjectifs. Dans Owen (1989), l'auteur justifie l'utilisation de quadrats de 63,9 km de côté de la façon suivante : "*I believe that quadrats of this size are sufficiently small to resolve the biogeographical phenomena being investigated without being so small as to misrepresent the accuracy of available distribution data*". Ces méthodes étant devenues tellement répandues, certains auteurs ne précisent même pas l'origine des données sur lesquelles ils ont réalisé cette transformation (Tab. 2). Cependant, il semble indéniable que cette étape induit une perte d'information au niveau géographique. De plus, dans le cas des listes d'occurrences, l'appartenance d'un point à un quadrat, défini *a posteriori*, est complètement subjective et diffère selon le maillage utilisé. Dans le cas des relevés, le choix du site

est fait *a priori*, et le lien entre une occurrence et un relevé est alors unique et incontestable. L'utilisation de données issues de relevés pour faire des listes d'occurrences, en dupliquant les lignes, amplifie la ressemblance "visuelle" entre ces deux types de données pourtant foncièrement très distinctes (Fig. 6). De nombreux auteurs ne font donc pas de différences et appliquent les mêmes méthodes aux relevés et aux listes d'occurrences.

De plus, en étudiant des listes d'occurrences, on ne s'intéresse qu'aux espèces effectivement présentes et rencontrées. L'absence d'une espèce dans un relevé peut avoir différentes causes, ce que Green a explicité en 1971 : "*If the species is absent, there are three possible interpretations : (i) The species cannot live there; that is, its niche does not include that point. (ii) The species can live there, but never had the opportunity for zoogeographic reasons. (iii) The species can and does live there, but the sample failed, by chance, to include a representative of that species.*". Dans le cas de relevés, la plupart du temps, une part d'absence est un aléa, une autre est une information. Pour un relevé en milieu aride, l'absence d'une espèce est certainement une information. Dans le cas d'un tableau issu de listes d'occurrences, il ne fait aucun doute qu'un zéro est synonyme d'absence d'information car les données originales ne concernent que les présences. On ne connaît rien sur l'intensité de prospection et la pauvreté observée d'une zone peut être effective ou résulter d'un faible nombre de visites. Par conséquent, tous les travaux d'écologie ou de biogéographie se doivent de n'utiliser que l'information disponible, c'est-à-dire uniquement les présences pour les listes d'occurrences.

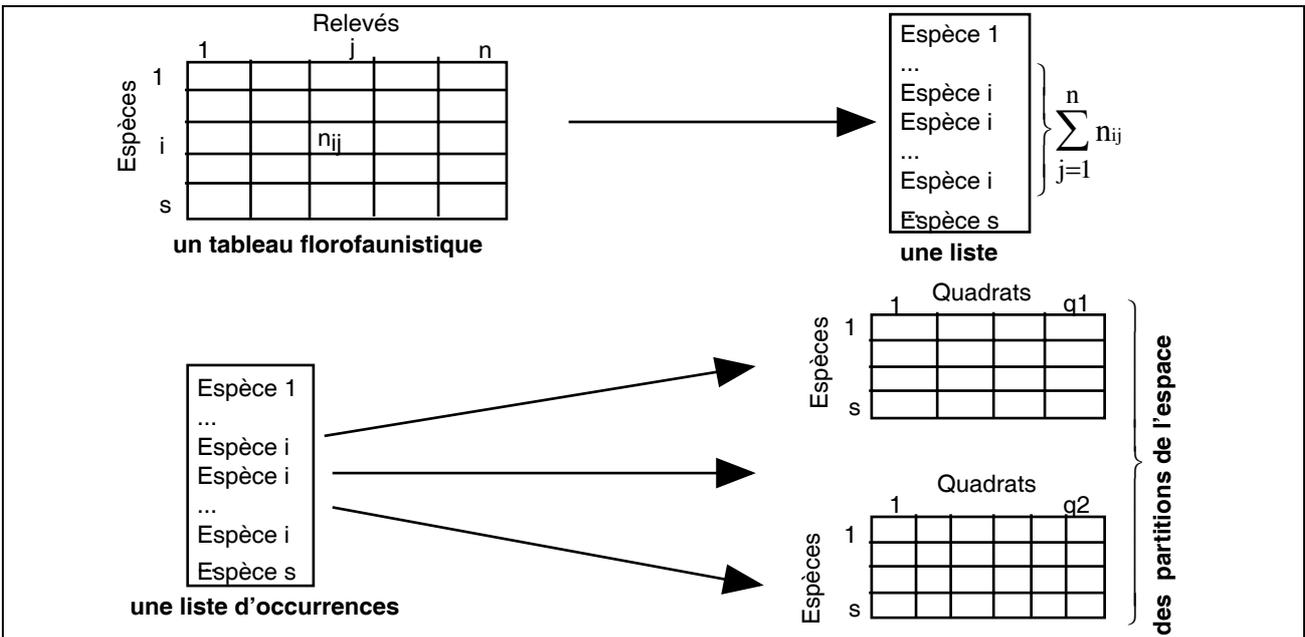


Figure 6 : Schéma résumant les transformations couramment réalisées sur les listes d'occurrences et les tableaux florafaunistiques.

Tableau 1 : Références de quelques articles où l'auteur transforme des listes d'occurrences grâce à une partition de l'espace. Origine : données issues de collections privées ou de Muséums (C), d'herbiers(H), de la littérature (L) ou récoltées par l'auteur (T). Le "?" signifie que l'auteur n'a pas précisé le renseignement recherché.

Source	Taxon	Origine	Nb. Taxa	Nb. Occ.	Surface étudiée en km ²	Transformation
Brown et Opler (1990)	papillons	C + L	176	> 10000	≈ 20000	10 bandes de 60 km
Cronk (1989)	plantes	T + H	58	?	121,7	quadrats (1x1 km)
Ferrari et al. (1993)	plantes rares	T + L	170	> 20000	22124	quadrats (3lat.x5'long.)
Gimaret-Carpentier et al. (1998)	arbres	T + H + L	269	3947	≈ 19800	198 quadrats (5'x5')
Kadmon et Danin (1997)	plantes vasculaires	T + H + L	917	71062	≈ 23525	941 quadrats (5x5 km)
Legakis et Kyriotakis (1994)	animaux et plantes endémiques	L	452	?	8250	quadrats (18x22 km)
Linder (1991)	plantes	H + L	5	?	≈ 50000	55 quadrats (1/4°x1/4°)
Mourelle et Ezcurra (1996)	cactus	H + L	223	3395	?	318 quadrats (1°x1°)
Owen (1989)	amphibiens et reptiles	C + L	200	> 100000	692000	189 quadrats (63,9x63,9 km)
Pearson et Ghorpade (1989)	cicindèles	T + C	150	?	≈ 4000000	quadrats (275x275 km)
Ranjit Daniels (1992)	amphibiens	T + L	117	?	1600 km de long	bande de 1°
Reaka (1980)	stomatopode coralien	C + L	62	?	?	quadrats (5°x5°)

Tableau 2 : Quelques articles où l'auteur indique les transformations qu'il a réalisées, sans préciser l'origine des données utilisées

source	Taxon	
Williams (1996)	bourdons	"The spatial distributions of the 241 species of bumble bees recognized at present have been mapped provisionally as presence-absence data using a cylindrical equal area grid cells based on intervals of 10° Longitude."
Prendergast et al. (1993)	odonates et hepaticae	"BRC (Biological Records Centre) species distribution data are mapped in 10x10 km squares of the Ordnance Survey, n=2876 in the British Isles."
Kerr (1996)	Mammalia, Plusiinae, Lasioglossum, Papilionidae	"I collected data on the distributions of terrestrial Mammalia (D. J. Currie, personal communication), Plusiinae (...) to produce species richness totals on a 336-quadrats map covering North America north of Mexico."
Gaston et al. (1996)	papillons	"As is usual for data on species distributions in Britain, these were summarised on a grid of 10x10 km squares, as defined by the British National Grid"
Blackburn et Gaston (1996b)	oiseaux	"Each species' distribution was transcribed onto a cylindrical equal area (e.g. Peter's) projection map of the New World."

II. Estimation de l'aire géographique de distribution

L'aire géographique de distribution d'une unité taxonomique telle qu'une espèce, un genre, une famille est la partie de la surface étudiée où elle est représentée. Cette aire est constituée en réalité de surfaces élémentaires plus ou moins nombreuses et rapprochées où l'unité considérée est effectivement présente. De nombreuses études en écologie s'intéressent aux patterns des tailles des aires de distribution en relation avec des variables telles que l'abondance locale (Gaston 1996, Gaston & Lawton 1990), la latitude (Pagel et al. 1991), la taille corporelle (Reaka 1980, Gaston & Blackburn 1996). Pour les programmes de conservation, les tailles des aires de distribution et leur répartition selon les espèces sont des critères prépondérants afin de définir les sites et les espèces à protéger en priorité (Gaston 1994). Cependant, malgré les diverses applications des mesures d'aire de distribution, les méthodes d'estimation utilisées sont généralement simples (bien qu'il en existe de plus sophistiquées) et, pour certaines, assez grossières.

Gaston (1991) distingue deux grands types de mesures des aires de distribution :

- l'étendue des occurrences (*extent of occurrence*) : ces mesures estiment une zone ou une distance entre les occurrences spécifiques les plus éloignées géographiquement. De ce fait, il peut y avoir inclusion de régions qui ne sont pas actuellement occupées.

- la surface d'occupation (*area of occupancy*) : dans ce cas, on s'intéresse uniquement aux zones où l'espèce est actuellement présente. La surface d'occupation donne donc généralement des valeurs plus faibles car une espèce n'occupe pas tout l'espace contenu à l'intérieur des limites géographiques de ses occurrences (certaines zones sont inhospitalières ou non colonisées à l'heure actuelle).

Le choix d'une méthode de mesure se fait selon la qualité des données disponibles et leur utilisation future. Il en existe donc un grand nombre que l'on peut diviser en trois classes majeures (Gaston 1994) :

II.1. Distance linéaire (Fig. 7A)

Ces méthodes évaluent l'étendue des occurrences. La démarche la plus simple est de mesurer la distance entre les deux points aux limites des occurrences. Le plus souvent, ces limites sont les sites séparés par la plus grande distance latitudinale (Reaka 1980, Macpherson & Duarte 1994, Letcher & Harvey 1994). D'autres mesures ont été utilisées telles que la distance longitudinale (Reaka 1980), ou la distance diagonale $[(NS^2 + EO^2)^{1/2}]$ avec NS et EO respectivement la distance latitudinale et longitudinale, Reaka 1980]. Ces méthodes étant fortement sensibles aux occurrences éloignées, il

est possible de réduire l'influence des sites isolés en utilisant une distance linéaire incluant 90 ou 95% des localités (Gaston et al. 1996). Pour les espèces présentes en un seul site, il n'est pas possible de mesurer de distance. Il est donc nécessaire de les exclure de l'analyse ou de leur attribuer une distance minimale.

II.2. Aire contenue dans les limites (Fig. 7B, 7C)

Les méthodes basées sur une distance linéaire sont des mesures dans un espace à une seule dimension. Il est donc impossible de distinguer deux aires de distribution dont l'étendue est la même sur l'axe de mesure mais qui diffèrent sur un autre axe. Pour cette raison, des procédures intégrant plus d'un axe de mesure ont été utilisées. Ainsi, Stevens (1986) calcule l'aire d'un rectangle défini par les deux axes majeurs de la distribution (avec la contrainte d'orthogonalité des axes).

Dans la plupart des atlas, la réalisation de cartes de distribution se fait par définition d'une zone d'occupation en traçant, généralement à l'oeil, une frontière qui englobe l'ensemble des occurrences d'une espèce. L'estimation de la surface de ces zones (grâce à un quadrillage par exemple) peut être utilisée pour déterminer l'aire de distribution d'une espèce (Gaston & Lawton 1990, Pagel et al. 1991, Letcher & Harvey 1994, Blackburn & Gaston 1996b). Ces mesures sont des estimations de l'étendue des occurrences car elles ne tiennent pas compte de la fragmentation possible d'une aire de distribution.

Il est à déplorer que d'autres méthodes plus élaborées telles que le polygone convexe minimum (polygone de surface minimum contenant tous les sites et dont les angles n'excèdent pas 180°) ne sont que très rarement employées (Gaston et al. 1996, Swihart et al. 1988) malgré leur plus grande précision.

II.3. Nombre de surfaces élémentaires occupées (Fig. 7D)

Les techniques précédentes nécessitent le positionnement de frontières de l'aire de distribution. Une approche fondamentalement différente est de diviser l'espace d'étude en unités géographiques et de calculer le nombre d'unités où l'espèce est présente. Ces unités peuvent être définies sur des critères géopolitiques (communes, départements par exemple, Gaston et al. 1996) mais, dans la majorité des cas, elles sont basées sur l'utilisation d'une grille régulière et l'aire de distribution par la somme des quadrats où l'espèce a été enregistrée (Reaka 1980, Schoener 1987, Gaston & Lawton 1990). La dimension des cellules varie selon les études : 5° de longitude x 5° de latitude pour Reaka

(1980), 10 x 10 km pour Gaston & Lawton (1990). Les méthodes basées sur l'utilisation d'une grille mettent largement en évidence la dépendance entre la taille de l'aire de distribution d'une espèce et l'échelle choisie : pour une grille de 100 x 100 km, la surface minimum occupée par une espèce ne pourra être inférieure à 10000 km². Avec une grille de 10 x 10 km, elle n'est que de 100 km². Quand l'unité géographique de mesure est grande, ces méthodes tendent à estimer l'étendue des occurrences. Par contre, si l'unité est petite, elles tendent à estimer la surface d'occupation.

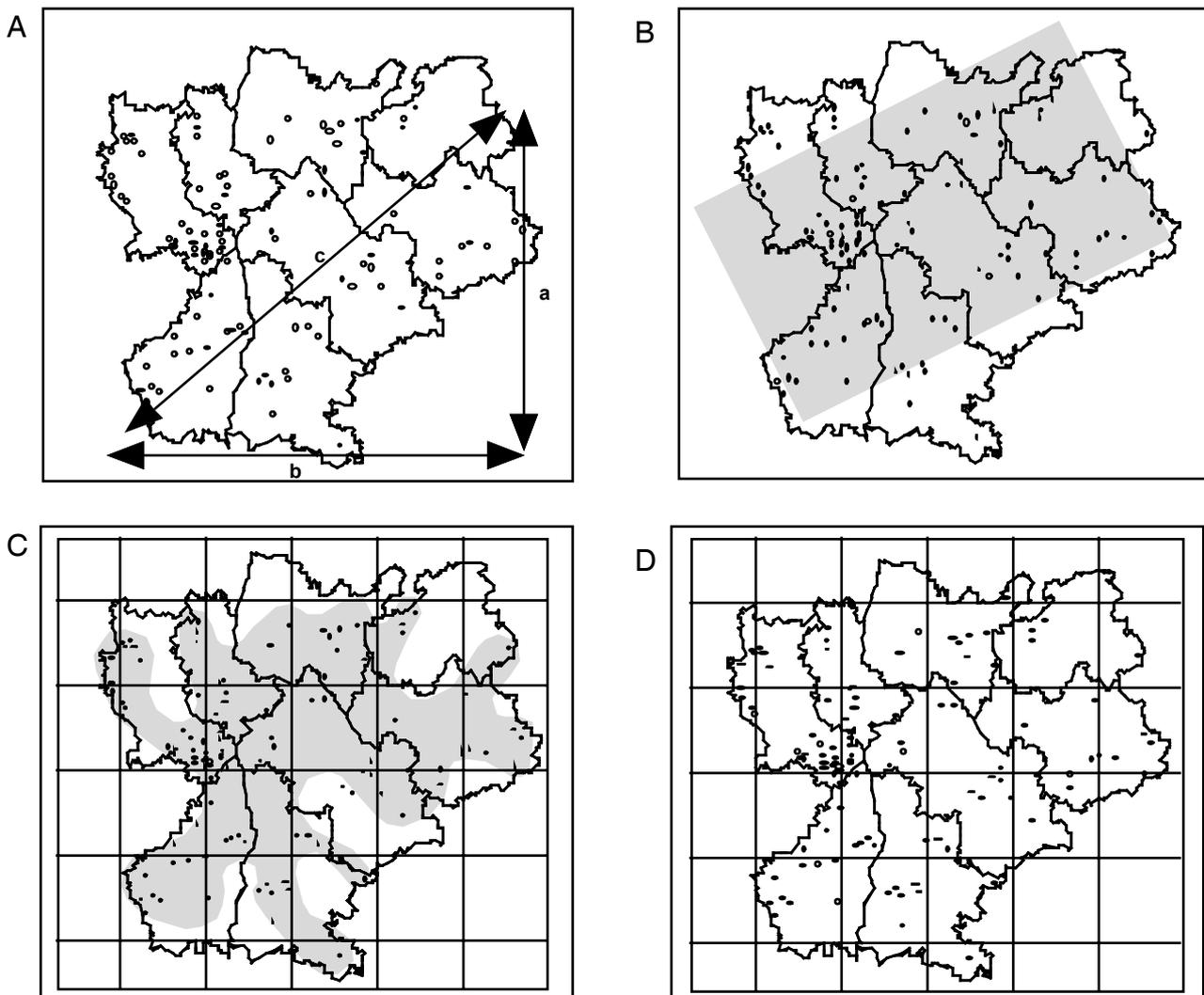


Figure 7 : Illustration de quelques méthodes d'estimation de l'aire de distribution de *Mesocarabus problematicus* (257 occurrences) dans la région Rhône-Alpes. A : distance latitudinale (a), distance longitudinale (b), distance diagonale (c). B : définition d'un rectangle à partir des axes majeurs de la distribution. C : délimitation de l'aire à l'oeil et estimation à l'aide d'un quadrillage. D : dénombrement des quadrats où l'espèce a été enregistrée.

Toutes les techniques présentées, même si elles ne mesurent pas exactement la même chose, sont étroitement liées (Gaston 1994). La mesure la plus fine de l'aire de distribution d'une espèce est probablement la somme des territoires de tous les individus (dont on soustrait les zones où il y a chevauchement). En pratique, cette valeur est impossible à calculer dans la majorité des cas et, il faut donc avoir recours à une méthode s'accommodant des données disponibles. Mais dans la

plupart des études, c'est souvent la simplicité et la facilité de mise en oeuvre d'une méthode, et non pas sa qualité, qui détermine le choix du chercheur.

Les méthodes présentées (et donc utilisées) impliquent une exploration relativement exhaustive de la région d'étude. Or, dans le cas des listes d'occurrences, aucune information ne nous permet de l'affirmer. Il est même assez vraisemblable qu'il en soit tout autrement. Ces méthodes font l'hypothèse implicite que la prospection est telle que l'absence d'une espèce est synonyme d'information (notamment pour les méthodes qui dénombre les surfaces élémentaires occupées). Pour les listes d'occurrences, il s'avère donc difficile de décrire la zone occupée par une espèce autrement que par une carte où les localités de collecte sont indiquées. De plus, la présence d'une espèce étant intimement liée aux paramètres environnementaux (altitude, type de sol, etc...), il apparaît nécessaire de corréler les localités à ces paramètres. Ainsi, il est possible d'étudier la répartition spatiale d'une espèce en définissant des districts écogéographiques pour lesquels les caractéristiques mésologiques sont homogènes.

III. Assemblage d'espèces : Estimation de la diversité

III.1. Un grand nombre d'indices

La diversité biologique (ou biodiversité), un thème central de l'écologie théorique et pratique, est un concept intuitivement simple et compréhensible mais, elle a le désavantage d'être dépendant de la taille de l'échantillon. En effet, le nombre d'espèces trouvées est relié positivement avec la taille de la zone étudiée et le nombre d'organismes collectés (Elphick 1997). De plus, cette notion est intimement liée à celle de communauté dont il n'existe pas de définition sans équivoque (Magurran 1988). Pour cette raison, la richesse spécifique est souvent mesurée pour une surface définie à l'aide d'une grille (Blackburn & Gaston 1996a, Mourelle & Ezcurra 1996, Böhning-Gaese 1997). Pourquoi la littérature scientifique contient-elle un très grand nombre d'indices pour estimer la diversité? La réponse vient certainement du fait que la diversité a deux composantes : le nombre d'espèces et leur abondance relative. Cette dualité du concept peut être illustré à l'aide de la figure 8 :

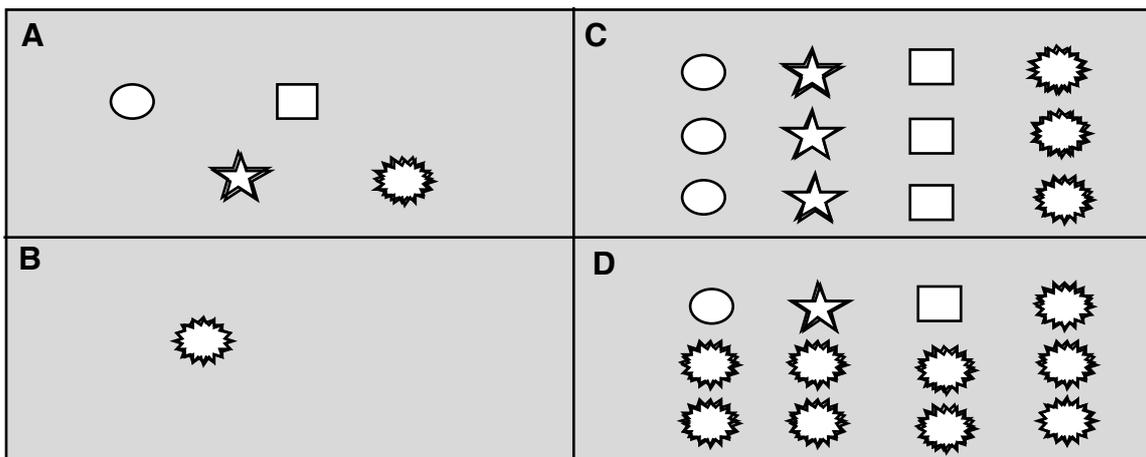


Figure 8 : Illustration du concept de diversité biologique. ◊ ★ □ ♀ représente quatre espèces. Le quadrat A est plus riche que le B car il contient quatre espèces contre une. Mais, il n'y a pas de différences en terme de richesse entre les quadrats C et D. Ces deux quadrats ont le même nombre d'espèces (4) et d'individus (12) mais dans D une espèce est très abondante alors que les autres sont plus rares. Le site C a donc une plus grande diversité.

Cet exemple très simple montre bien les deux concepts liés à la notion de diversité. La plupart des indices utilisés en écologie se distinguent par le poids relatif donné au nombre d'espèces et à leur abondance relative. En d'autres termes, c'est l'importance que l'on donne aux espèces rares qui varie selon les indices. Par exemple, le nombre d'espèces, qui est certainement la mesure la plus utilisée (Brown & Opler 1990, Blackburn & Gaston 1996a), ne tient pas du tout compte de

l'abondance : toutes les espèces ont le même poids. L'indice de Simpson ($D = \sum_{i=1}^n \frac{n_i(n_i-1)}{N(N-1)}$, pour n espèces avec n_i le nombre d'individus de l'espèce i et N le nombre total d'individus) et l'indice de

Shannon ($H = -\sum_{i=1}^n p_i * \ln(p_i)$, avec p_i l'abondance relative de l'espèce i) sont les deux autres

mesures les plus utilisées. Le tableau 3 résume les caractéristiques et les performances de quelques indices :

Tableau 3 : Caractéristiques de quelques mesures de la diversité (d'après Magurran 1988, p. 79). Le but de ce tableau n'est pas de donner une classification définitive mais simplement de montrer les qualités et les défauts des différentes mesures de diversité. La simplicité ou la complexité du calcul est jugée du point de vue d'un étudiant ayant les bases en mathématiques et possédant une calculatrice de poche. La colonne Richesse ou abondance indique si un indice est biaisé plutôt vers la richesse ou l'abondance.

	Capacité discriminante	Sensibilité à la taille de l'échantillon	Richesse ou abondance	Calcul	Couramment utilisé ?
α (log series)	Bonne	Faible	Richesse	Simple	Oui
λ (log normal)	Bonne	Modérée	Richesse	Complexe	Non
Q statistic	Bonne	Faible	Richesse	Complexe	Non
S (species richness)	Bonne	Forte	Richesse	Simple	Oui
Margalef index	Bonne	Forte	Richesse	Simple	Non
Shannon index	Modérée	Modérée	Richesse	Intermédiaire	Oui
Brillouin index	Modérée	Modérée	Richesse	Complexe	Non
McIntosh U index	Bonne	Modérée	Richesse	Intermédiaire	Non
Simpson index	Modérée	Faible	Abondance	Intermédiaire	Oui
Berger-Parker index	Faible	Faible	Abondance	Simple	Non
Shannon evenness	Faible	Modérée	Abondance	Simple	Non
Brillouin evenness	Faible	Modérée	Abondance	Complexe	Non
McIntosh D index	Faible	Modérée	Abondance	Simple	Non

III.2. Application aux listes d'occurrences

Hurlbert (1971) qualifie la diversité de "non-concept" et montre, à l'aide d'un exemple simple, que différents indices peuvent classer des communautés de différentes façons. Pour Patil & Taillie (1979), cité dans Tóthmérész (1995), il faut s'attendre à de telles inconsistances lorsque l'on essaie de réduire un concept multidimensionnel à un simple nombre. Ces problèmes inhérents au concept de diversité ne sont pas les seuls lorsque l'on traite des listes d'occurrences. En effet, la nature de ce type de données engendre d'autres problèmes dont les conséquences se font surtout ressentir pour des estimations de la diversité. Tout d'abord, la qualité de l'information spatiale des occurrences est rarement constante sur toute l'aire de distribution d'une espèce : certaines régions sont moins visitées à cause de leur inaccessibilité, d'autres le sont plus grâce à leur notoriété. Par exemple, si une espèce rare est trouvée sur un site, d'autres collecteurs iront visiter ce site pour trouver cette espèce. De plus, les occurrences ne donnent des informations que sur la rencontre entre un collecteur et un spécimen. Une absence d'information sur une zone peut donc refléter l'absence d'espèce ou tout simplement un faible niveau de visites. Par conséquent, il est impossible de connaître l'intensité avec laquelle une zone a été visitée et donc, toutes les méthodes destinées à pondérer une mesure de diversité par l'effort d'échantillonnage (Elphick 1997, Keating & Quinn

1998) ne peuvent s'appliquer. En outre, la récolte d'un spécimen en un site pour une collection ne donne aucune information sur l'abondance de l'espèce à cet endroit, il n'est donc pas cohérent d'utiliser les indices tenant compte de l'abondance. Dans le cas des listes d'occurrences, le nombre d'espèces apparaît donc comme la valeur la plus appropriée pour mesurer la diversité.

Certains auteurs ont tout de même essayé de considérer l'effort d'échantillonnage dans la mesure de la diversité à partir de listes d'occurrences. Mourelle & Ezcurra (1996) prennent en compte l'intensité de la collecte en étudiant des occurrences de cactus provenant d'herbiers. Le nombre d'espèces détectées dans une aire donnée est relié de façon non linéaire au nombre de spécimens collectés dans les herbiers (Soberón & Llorente 1993). Au début, les nouveaux spécimens seront souvent synonymes de nouvelles espèces, mais une fois la collection plus grande, la plupart des nouveaux spécimens appartiennent à des espèces déjà collectées (Fig. 9). L'équation mathématique prédisant cette relation est souvent appelée fonction d'accumulation. Soberón & Llorente (1993) présente un éventail détaillé de différentes fonctions d'accumulation. Leur étude montre que le nombre d'espèces et de spécimens présente une relation presque linéaire sur une échelle log-log (le comportement n'est pas tout à fait linéaire car le nombre d'espèces tend vers une asymptote, alors que le nombre de spécimens peut être infini).

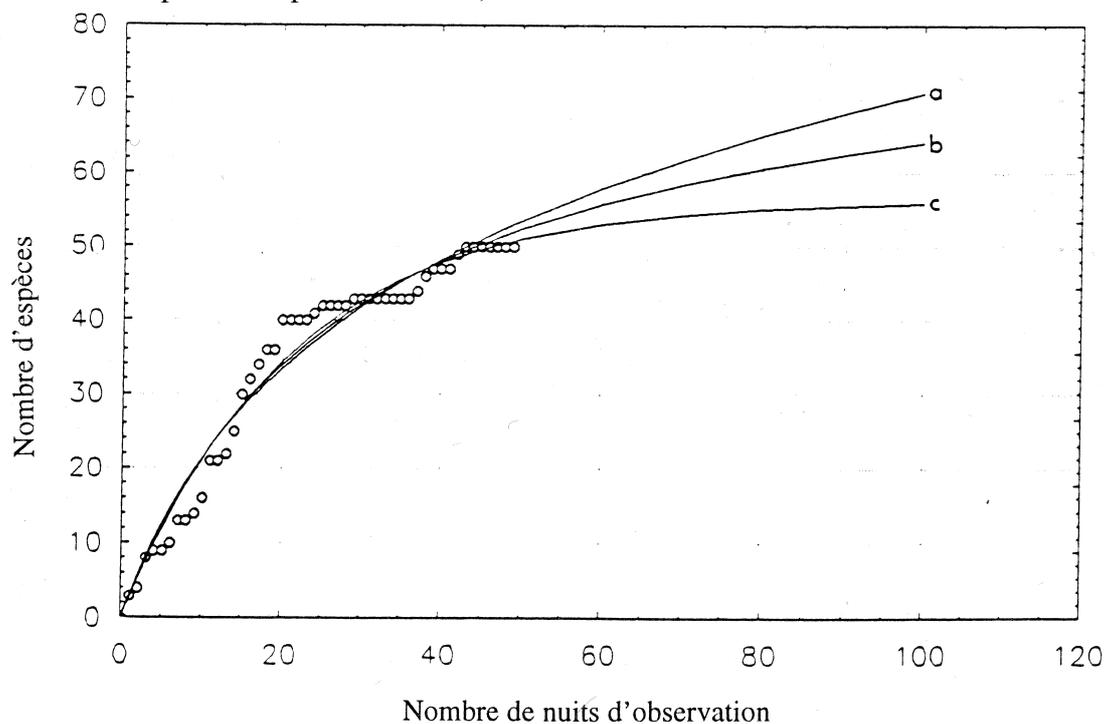


Figure 9: Courbe d'accumulation pour des occurrences de chauves-souris à Chajul, Mexique (d'après Soberon & Llorente (1993)). a correspond à un modèle logarithmique, b à l'équation de Clench et c à un modèle exponentiel.

D'après cette relation, Mourelle & Ezcurra (1996) ont incorporé à leur modèle le logarithme du nombre de spécimens collectés par quadrat comme variable explicative de la richesse spécifique afin d'évaluer l'effet potentiel de l'effort de collecte. Cependant, ce modèle ne tient pas compte du fait qu'un collecteur ne ramassera ou n'identifiera pas tous les spécimens qu'il rencontrera. Cette

démarche reste néanmoins très intéressante et assez novatrice dans le traitement des listes d'occurrences.

Mais est-il légitime d'essayer de mesurer la diversité à partir d'une liste d'occurrence ? Pour répondre à cette question, revenons sur la notion de communauté. En cherchant à ordonner et à classer leurs inventaires, les observateurs de terrain constatent que le travail leur était simplifié par l'association régulière d'espèces. Un spécialiste d'une catégorie de peuplement peut affirmer, à la vue d'une liste, si telle ou telle espèce est "normale" ou "aberrante" dans le site où elle a été observée. Ces associations (ou biocénoses, ou communautés) sont généralement caractéristiques d'une association de caractères du biotope. Les associations ont surtout été inventoriées dans deux domaines : la végétation terrestre et le macrobenthos marin. Ce n'est évidemment pas un hasard. Ce matériel se prête facilement à ce genre d'approche pour deux raisons : d'une part il est bien visible sur le terrain, et les relations entre association et biotope sont plus faciles à mettre en évidence ; d'autre part il est fixe : on peut revenir sur le terrain peu de temps après et retrouver la même association - ce qui se révèle beaucoup plus difficile dans le domaine du plancton par exemple, ou pour les animaux de grande taille se déplaçant facilement.

La notion de communauté a reçu un regain d'intérêt avec la mise en oeuvre de méthodes statistiques permettant la mise en évidence et la classification des groupements à partir d'inventaires. Ces méthodes comprennent des analyses factorielles et des classifications numériques (par exemple, TWINSpan, AFC, etc...). Il est assez remarquable que ces méthodes "objectives" aient confirmé largement les groupements d'espèces et les classifications d'associations établis par les méthodes plus intuitives des précurseurs. Les méthodes numériques récentes corroborent le travail antérieur, et fournissent un outil pour le poursuivre et, peut-être plus encore, pour l'appliquer à des cas plus complexes comme les peuplements des régions tropicales, où le très grand nombre d'espèces et d'associations impose des limites à l'analyse intuitive. Pour Viera da Silva (1979), la diversité a un sens écologique quand elle mesure les différentes solutions prises par différents types d'organismes pour occuper l'écosystème. Par exemple, dans une communauté déserticole, les stratégies adoptées par les différentes espèces de plantes sont très variées : il n'y a pas une seule solution au problème posé par le manque d'eau. Une mesure de diversité quantifie cette richesse de solutions. Le nombre de composants différents (nombre d'espèces) peut être une mesure de diversité, mais il en est une approximation grossière car l'importance de chacune des espèces peut être différente en nombre d'individus (abondance) et dans leur taille surtout chez les plantes. A la lumière de toutes ces précisions, des interrogations peuvent se poser sur le sens d'une mesure de diversité pour un quadrat. En effet, découper une région à l'aide d'une grille et affecter à chacun des quadrats les occurrences d'une liste pour mesurer la diversité biologique n'a aucun sens car un quadrat n'a pas le même statut qu'un relevé. Un relevé pour une station identifie des associations d'espèces observées dans un écosystème, alors que le quadrat peut regrouper plusieurs écosystèmes surtout dans le cas d'études à grande échelle. Analyser la diversité d'une communauté avec des listes d'occurrences ne peut donc être possible. Pour Whittaker (1972), il existe trois grands types de diversités appelées alpha, bêta, et gamma diversité. La diversité alpha est celle qui caractérise une

communauté naturelle (diversité intrabiotope), la diversité gamma correspond à une grande région hétérogène et mesure cette hétérogénéité (diversité sectorielle). Enfin, la diversité bêta mesure la différence entre peuplements de deux biotopes voisins (indice de similitude interbiotope). L'appartenance d'une espèce à une communauté ne pouvant être vérifiée avec les listes d'occurrences, une mesure de diversité à partir de ce type de données doit s'apparenter à la diversité gamma de Whittaker, sinon la signification écologique de travaux utilisant ce genre de méthodologie ne peut être validée.

IV. Relations espèces-milieu

Les naturalistes du XVIIIème signalaient souvent qu'une espèce se trouve plutôt dans les forêts, sur les bords des chemins, dans les rivières, etc. Ils prenaient même souvent en compte ce caractère "écologique" pour choisir le deuxième terme du binôme linnéen, ainsi qu'en témoignent les nombreuses espèces qualifiées de *silvatica*, *ruderalis*, *aquatica*, *montanum*, etc. Cependant, ils se souciaient peu des limites écologiques qui bornent l'extension naturelle d'une espèce. L'étude de la répartition spatiale d'espèces ou d'assemblage d'espèces comporte deux questions :

- (i) Où se distribue l'espèce ou les espèces ?
- (ii) Quels sont les causes de cette distribution ?

Pour répondre à cette deuxième question, il est nécessaire de rechercher des hypothèses explicatives. L'histoire en est une : une espèce ne peut occuper actuellement une zone si elle n'a pas eu l'occasion dans le passé de s'y introduire. Mais, l'hypothèse la plus simple est d'expliquer la présence d'une espèce par des facteurs du milieu.

IV.1. Le concept de niche multidimensionnelle

En 1957, Hutchinson développe le concept de niche écologique multidimensionnelle afin d'expliquer comment les conditions environnementales limitent l'abondance et la distribution d'une espèce. Il représente donc l'environnement d'une espèce par un espace multidimensionnel (ou hypervolume) où chaque dimension (ou axe) correspond à une variable environnementale. La niche d'une espèce est alors définie par la combinaison de ces variables qui permet aux individus de survivre et de se reproduire. Ce concept peut être illustré par l'effet de deux variables environnementales (la salinité et la température) sur une espèce aquatique (Fig. 10). Si toutes les autres conditions sont favorables, la croissance des effectifs des populations est limitée par les effets de ces deux variables.

En réalité, il est évident que notre organisme aquatique serait limité par d'autres variables telles que la concentration d'oxygène dissout, un prédateur, etc. Chacune de ces variables représenterait un nouvel axe et transformerait la niche en un volume multidimensionnel.

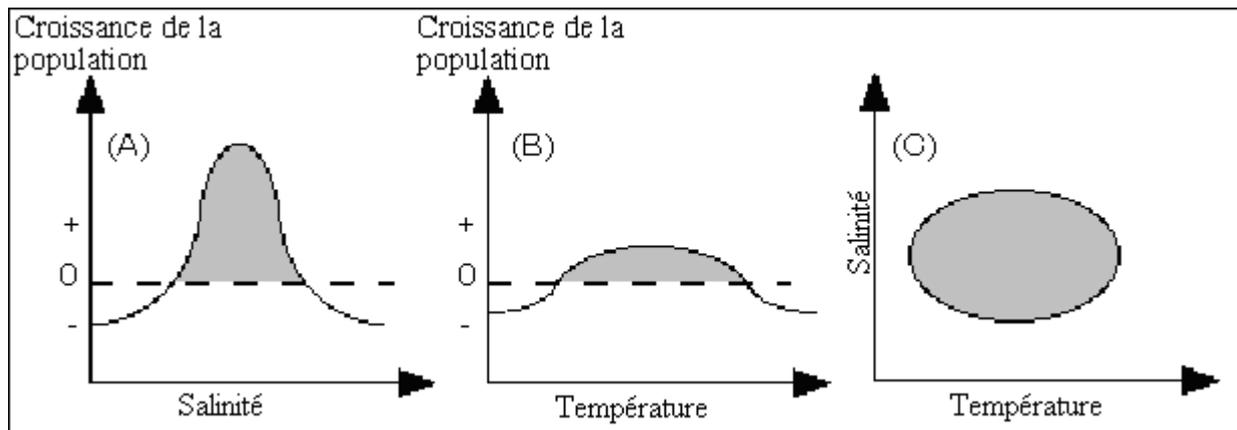


Figure 10 : Illustration du concept de niche multidimensionnelle (exemple à deux dimensions). En (A) et (B), le taux de croissance de la population est indiqué en fonction de chacune des deux variables. La ligne en pointillée indique la valeur nulle nécessaire pour représenter la niche en (C). La niche de cette hypothétique espèce aquatique est représentée par une ellipse dans l'espace à deux dimensions. L'aire contenue dans cette ellipse représente l'ensemble des combinaisons des deux variables sous lesquelles les individus peuvent survivre et se reproduire (où les effectifs des populations peuvent se maintenir).

Le concept de niche permet d'appréhender facilement la distribution spatiale d'une espèce ou d'un assemblage d'espèces au travers des relations espèces-milieu. L'aire de distribution d'une espèce peut être assimilée à une traduction spatiale de la niche : une espèce est présente quand les conditions environnementales lui conviennent. L'assemblage d'espèces peut être vu comme un groupe d'espèces vivant ensemble car leurs niches se chevauchent. Par conséquent, une espèce peut être caractérisée par le type de milieu qu'elle occupe ; et un milieu peut être caractérisé par les espèces qu'il abrite.

IV.2. Richesse spécifique et facteurs environnementaux

De nombreuses études d'écologie s'intéressent aux patterns de diversité afin de définir quels facteurs environnementaux en sont la cause. L'approche la plus commune est d'établir un modèle dans lequel le nombre d'espèces est expliqué par des variables mésologiques (altitude, température, précipitation, etc...). Lors d'un relevé, les données concernant le milieu et le peuplement sont souvent enregistrées en même temps. Par contre, pour les listes d'occurrences, la définition de surfaces élémentaires, qui se substituent aux relevés, est nécessaire (Owen 1989, Pearson & Ghorpade 1989, Mourelle & Ezcurra 1996). Ainsi, on attribue à chaque quadrat, le nombre d'espèces et les valeurs moyennes des paramètres environnementaux. Le choix de l'échelle est généralement très peu discuté, mais la taille du quadrat peut influencer fortement sur le modèle utilisé. Böhning-Gaese (1997) s'est intéressé aux facteurs déterminant la richesse spécifique aviaire à partir de données tirées d'atlas. En faisant varier la taille des quadrats de 4 à 36 km², sans changer la zone d'étude, elle a constaté que les variables expliquant la richesse n'étaient pas les mêmes : la richesse spécifique d'un quadrat est toujours liée à l'hétérogénéité spatiale du quadrat mais les variables les

plus corrélées à la richesse spécifique à petite échelle, dont la surface forestière, perdent de l'importance, pendant que d'autres, dont la surface aquatique, gagnent de l'importance avec l'augmentation de la taille du quadrat. Selon elle, la forêt possède une forte hétérogénéité spatiale intra-habitat et donc une forte richesse spécifique aviaire à petite échelle. À une échelle plus grande, la forêt est un type commun d'habitat qui n'augmente pas la diversité du paysage et qui n'augmente que très peu la richesse spécifique. Par contre, le milieu aquatique possède une très faible hétérogénéité spatiale intra-habitat et abrite peu d'espèces à une petite échelle. À l'échelle du paysage, ce type d'habitat, qui n'est pas commun, augmente la diversité spatiale et donc la richesse spécifique (Fig. 11). Il apparaît donc que l'hétérogénéité spatiale à une certaine échelle prédit la richesse spécifique à cette même échelle et non à une échelle plus grande ou plus petite. Par conséquent, il semble évident que la taille du quadrat peut influencer fortement les résultats d'une étude et les conclusions établies à une certaine échelle ne peuvent être transposées à une autre échelle. Pour les programmes de protection de l'environnement, ce type de considérations est à prendre en compte lors de la délimitation des futures réserves naturelles.

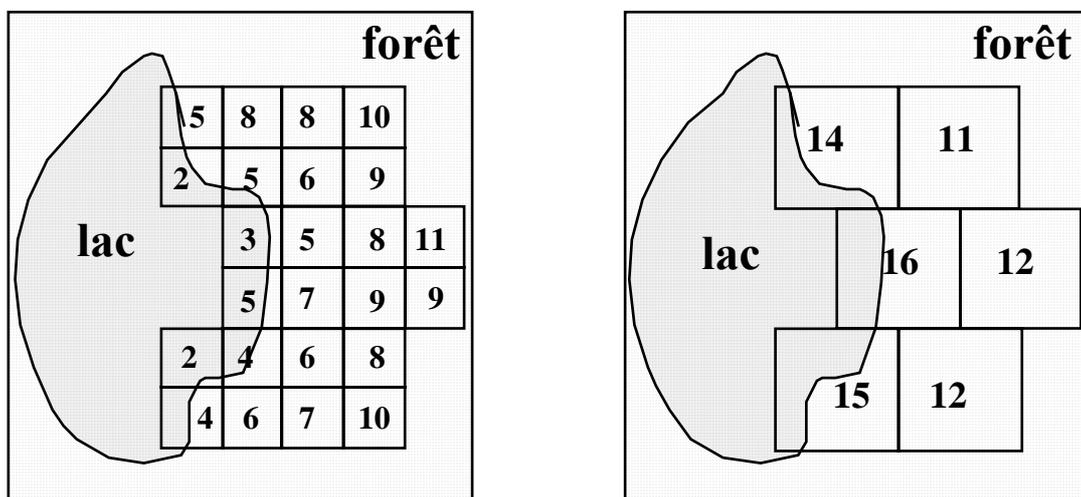


Figure 11 : Exemple fictif illustrant l'influence de la taille du quadrat sur la mesure de la richesse spécifique aviaire. À une petite échelle d'étude, la richesse spécifique est plus importante pour les quadrats exclusivement dans la forêt. Si la taille du quadrat augmente (ici x4), ce sont les quadrats les plus près du lac qui possèdent la plus forte richesse spécifique. Pour le milieu forestier, il existe une forte hétérogénéité spatiale intra-habitat qui induit une forte richesse spécifique à petite échelle. Si la taille du quadrat augmente, la présence du lac ajoute des espèces "aquatiques" aux quadrats abritant déjà des espèces "forestières".

Après avoir adopté une taille de quadrat, il faut choisir le modèle destiné à expliquer la richesse spécifique à l'aide des variables environnementales. Cela peut être un modèle de régression linéaire simple (Pearson & Ghorpade 1989, Linder 1991) ou multiple (Owen 1989) ou multiple pas à pas (Linder 1991, Böhning-Gaese 1997). Mais ce type de modèles pose des hypothèses, sur la distribution des erreurs notamment, qui ne peuvent être vérifiées et autorise des valeurs prédites (nombre d'espèces) négatives. Une solution appropriée apparaît dans l'utilisation d'un modèle linéaire généralisé. Ainsi Mourelle & Ezcurra (1996) et Heikkinen (1998) emploient un modèle logarithmique avec l'hypothèse d'une distribution poissonnienne des erreurs car le nombre d'espèces est une variable discrète. Certains auteurs ont également essayé de considérer les particularités des listes d'occurrences dans leurs modèles (cf. III. pour Mourelle & Ezcurra 1996). Par exemple,

Connor & Simberloff (1978) qui étudient la flore des îles Galápagos, ont inventorié, grâce à la littérature et à des communications personnelles, le nombre de visites de botanistes que chaque île a reçu. Dans leur modèle, il apparaît que le nombre de voyages ou son logarithme est la variable qui explique la plus grande part de variance du nombre d'espèces ou du logarithme du nombre d'espèces (respectivement 97% et 92%). Afin de déterminer si la corrélation entre le nombre d'espèces et le nombre de voyages, et non la corrélation entre l'aire de l'île et le nombre de voyages, était responsable de ces résultats, les auteurs ont utilisé une régression multiple forcée. Ils concluent que dans le modèle log-log le nombre de voyages et l'aire de l'île sont interchangeable même si le logarithme du nombre de voyages a une corrélation plus forte avec le nombre d'espèces (0,959 vs. 0,891). Pour le modèle non transformé, le nombre de voyages reste un meilleur prédicteur que l'aire de l'île. S'ils ne tiennent pas compte du nombre de voyages, l'altitude est le meilleur prédicteur dans le modèle non transformé. Mais, quand ils introduisent la variable du nombre de voyages, l'altitude n'entre plus dans le modèle de régression. Par conséquent, les résultats de certaines études, concluant à une forte corrélation entre l'altitude et le nombre d'espèces à l'aide d'un modèle non transformé, peuvent être faux. Pour analyser les données d'un atlas ornithologique, Böhning-Gaese (1997) introduit la qualité de l'observateur dans les analyses statistiques. Pour se faire, les organisateurs de l'atlas ont estimé la capacité de chaque observateur à reconnaître toutes les espèces d'oiseaux présentes en utilisant trois classes de qualité. Dans la régression pas à pas, cette variable n'entre pas dans le modèle et elle n'explique qu'une part très faible de la variance dans la régression multiple.

L'étude de la richesse spécifique à l'aide de facteurs environnementaux n'utilise pas toute l'information disponible. En effet, dans ce type d'analyse on s'intéresse seulement au nombre d'espèces enregistrées dans un quadrat et non aux espèces elles-mêmes. Une autre approche consiste donc à examiner les relations entre facteurs environnementaux et composition spécifique.

IV.3. Composition spécifique et facteurs environnementaux

Pour analyser la composition spécifique à l'aide de variables de milieu, l'approche est assez similaire à celle présenté ci-dessus. La zone d'étude est découpée en quadrats auxquels on attribue les occurrences enregistrées ainsi que les valeurs moyennes des variables environnementales. Il en résulte un tableau **F** formé par les **q** quadrats en lignes et les **s** espèces en colonne et un tableau **M** formé par les **q** quadrats en lignes et les **m** variables mésologiques en colonne (Fig. 12) :

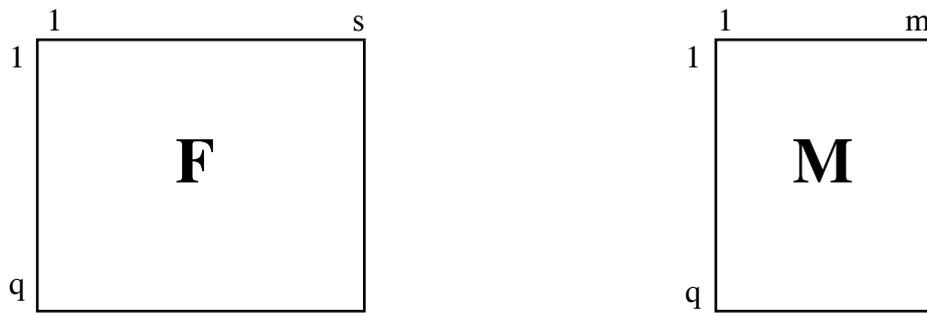


Figure 12 : Présentation des données utilisées. Le tableau **F** est formé par les **q** quadrats en lignes et les **s** espèces en colonne et le tableau **M** est formé par les **q** quadrats en lignes et les **m** variables mésologiques en colonne.

L'analyse de ces deux tableaux permet de s'intéresser aux relations entre biocénose et biotope et d'essayer de déterminer quels facteurs environnementaux influencent la présence d'une espèce. Il est alors possible de définir l'étendue de la niche d'une espèce et de séparer les niches entre les espèces, et ainsi d'essayer de prévoir la présence d'une espèce en fonction des valeurs des paramètres mésologiques.

L'analyse du couple de tableaux (F, M) peut se faire soit par une approche symétrique soit par une approche dissymétrique. Pour les méthodes symétriques, les deux jeux de données sont traités simultanément avec la même considération. Dans le cas des méthodes dissymétriques, on considère les variables d'un tableau comme prédicteurs des variations observées dans le second (Chessel & Mercier 1993). Dans notre cas, le choix d'une stratégie dissymétrique est préférable car c'est essentiellement le milieu qui influence la présence des espèces (la réciproque est vraie mais moins importante). Parmi les stratégies dissymétriques, deux approches sont possibles : les méthodes d'ordination directes ou indirectes.

Les méthodes d'ordination indirecte consistent en un positionnement des quadrats selon leur composition spécifique (on ne travaille que sur le tableau F). Les quadrats sont alors définis par leurs coordonnées factorielles. Parmi les méthodes couramment utilisées pour ordonner ce type de tableaux, on peut citer l'analyse des correspondances (AFC) et sa version détendancée (DCA), ainsi que l'analyse en composante principale (ACP). Lors d'une deuxième étape, on interprète les coordonnées factorielles d'un quadrat par ses paramètres mésologiques. Pour ce faire, on peut projeter les scores des quadrats sur une carte d'une variable de milieu. Par exemple, Gimaret-Carpentier et al. (1998) réalisent une analyse non symétrique des correspondances (ANSC) sur 3947 occurrences d'espèces d'arbres regroupées en 140 quadrats. Puis, ils projettent les scores factoriels sur des cartes où l'altitude et la durée de la saison sèche sont indiquées afin d'étudier les relations entre ces variables et la composition spécifique (Fig.13). Les quadrats dont le score factoriel sur le premier axe est positif se trouvent dans les zones de basse altitude, alors que les scores négatifs se trouvent à moyenne et haute altitude. Pour le deuxième axe, les scores positifs sont associés aux zones où la saison sèche est longue alors que les scores négatifs sont associés à une courte durée de saison sèche. Il apparaît donc que ces deux facteurs influencent fortement la composition spécifique.

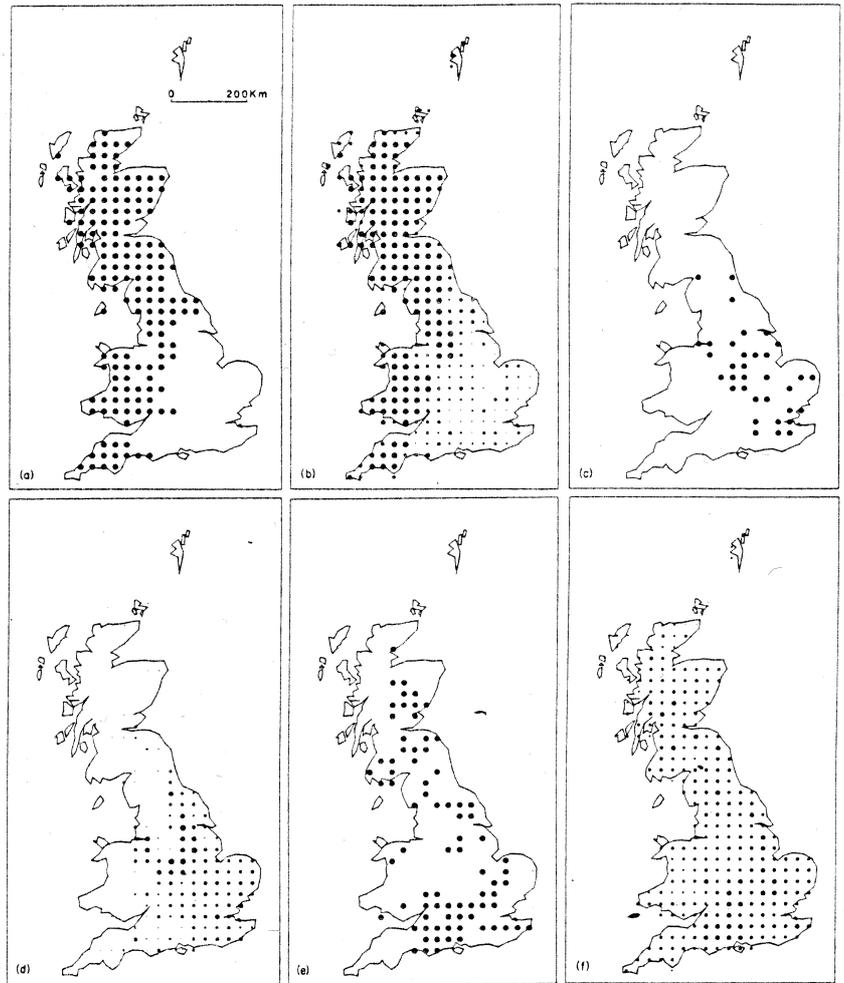


Figure 13 : Projection des scores factoriels des quadrats sur des cartes de la région d'étude (d'après Gimaret-Carpentier et al. (1998)). (A) projection des scores sur le premier axe factoriel sur une carte topographique. La légende du bas représente les ordres de grandeurs des scores, celle du haut représente les altitudes moyennes des quadrats en mètres. (B) : projection des scores sur le deuxième axe factoriel sur une carte de longueur de la saison sèche. La légende du bas représente les ordres de grandeurs des scores, celle du haut représente les durées moyennes de la saison sèche par quadrat en mois.

Mais ces approches, par leur caractère indirect, ne garantissent en aucun cas une corrélation entre les axes factoriels et les variables de milieu prises en compte. Il n'y a donc aucune raison que les scores factoriels soient facilement interprétables écologiquement.

Pour les méthodes d'ordination directe, l'ordination des quadrats ne se fait plus selon leur composition spécifique mais en fonction d'un ou plusieurs gradients de milieu. On ordonne alors les quadrats par les variables mésologiques et on étudie le mode de répartition des espèces sur les axes d'ordination ainsi définis. Une autre approche est d'étudier simultanément les tableaux M et F. Ceci peut être fait par l'analyse canonique des correspondances (ACC) de Ter Braak (1986). Cette méthode a répondu à une grosse demande des écologistes et elle est couramment utilisée pour analyser les relations entre les espèces et leur milieu à partir de relevés ou de listes d'occurrences regroupées en quadrats. De plus, elle permet d'essayer de prédire la présence d'une espèce en fonction de paramètres mésologiques. Par exemple, Hill (1991) réalise une régression logit pour déterminer la probabilité de présence d'une espèce dans un quadrat en utilisant les scores sur les quatre premiers axes d'une CCA comme variables explicatives (Fig. 14).

Figure 14 : Distributions actuelles (a, c, e) et prédites (b, d, f) en utilisant une régression logit (d'après Hill (1991)). Les espèces étudiées sont le cincle (*Cinclus cinclus*) (a, b), le pluvier doré (*Charadrius dubius*) (c, d) et le liseron (*Helianthemum nummularium*) (e, f). Les points de taille croissante représentent les catégories de probabilité de présence d'une espèce, dans un quadrat, suivantes : 1-4%, 5-10%, 11-30%, 31-50%, 51-75%, et 76-100%.



Cependant, l'étape de découpage de la zone en quadrats nécessite l'affectation d'une valeur moyenne des variables mésologiques pour chacun des quadrats. Par conséquent, il n'est pas possible d'identifier l'hétérogénéité en milieu que peut contenir un quadrat. Les conséquences sur l'interprétation écologique peuvent être assez dramatiques surtout lorsque l'on travaille à grande échelle dans des milieux très hétérogènes. Il serait donc judicieux de travailler directement sur la liste d'occurrences et d'estimer les paramètres mésologiques pour chacune des occurrences.

Conclusions et Perspectives

Les listes d'occurrences spécifiques constituent un type de données foncièrement original. Une occurrence résulte de la rencontre particulière entre un spécimen animal ou végétal et un observateur (ou collecteur) en un lieu et à une date donnée. La compilation d'occurrences à partir de spécimens préservés dans les muséums ou dans les herbiers, ou bien à partir d'atlas constitue une banque de données importante pour étudier la répartition spatiale d'animaux ou de végétaux. Le développement des ordinateurs et des méthodes statistiques a permis aux biogéographes et aux écologistes d'exploiter ces données initialement réservées à des travaux de systématique. Ainsi de nombreuses études ont été réalisées à partir de listes d'occurrences pour analyser les patterns de diversité et leurs causes, l'aire de distribution d'une espèce, etc... Dans la plupart des cas, les occurrences ne sont pas traitées indépendamment les unes des autres, elles sont regroupées en quadrats grâce à un découpage *a posteriori* de la zone d'étude. Les surfaces élémentaires ainsi définies, de nombreux auteurs traitent ces données comme si elles équivalaient à des relevés expérimentaux classiques. D'ailleurs, certains n'hésitent pas à compléter des listes occurrences avec des données issues de relevés.

Mais, la nature particulière des listes d'occurrences nécessite de prendre quelques précautions lors de leur utilisation. N'étant pas prédestiné à la biogéographie ou à l'écologie, ce type de données peut induire certains biais lorsque l'on s'intéresse à la répartition spatiale des espèces :

- L'intensité d'observation peut ne pas être uniforme sur l'ensemble de la zone d'étude. Mais, aucune information ne nous permet de le confirmer ou de l'infirmier.

- Une liste d'occurrences ne donne aucun renseignement sur l'abondance : la probabilité d'existence d'une occurrence dépend à la fois de l'abondance d'une espèce mais également des motivations personnelles de l'observateur.

- Un quadrat n'est pas équivalent à une station. L'assemblage d'espèces observé en une station reflète l'existence d'une communauté, ce n'est pas obligatoirement le cas pour une compilation d'occurrences.

Par conséquent, il est impératif d'analyser ces données comme un ensemble de rencontres réalisées et notifiées. Aucune supposition ne peut et ne doit être faite sur la "non-existence" d'une occurrence : peut-être que l'espèce n'était pas présente, peut-être que l'espèce était présente et l'observateur n'a pas pu ou voulu l'indiquer. On ne le sait pas et on ne le saura jamais.

Les listes d'occurrences peuvent se prêter assez bien à l'analyse des relations espèces-milieu. En effet, on sait qu'une espèce a été trouvée en un point géographique bien précis et, si l'on considère

qu'elle était en cet endroit car le milieu satisfaisait ses exigences de niche, il est alors possible d'essayer de déterminer quels sont les facteurs environnementaux les plus corrélés à cette présence. Les méthodes de couplage de tableaux permettent de répondre de façon efficace à ce type de questions. Cependant, leur utilisation en écologie se fait exclusivement par l'intermédiaire de quadrats auxquels sont affectées des valeurs moyennes des variables environnementales, sans tenir compte de l'hétérogénéité mésologique intra-quadrat. L'analyse des listes d'occurrences, dans leur format original, pourrait permettre de tenir compte de cette hétérogénéité. Ainsi, l'étude des relations espèces-milieu gagnerait en précision au niveau géographique, mais surtout écologique. Ce type d'approche a commencé à être exploré par Gimaret-Carpentier (1999) grâce à une nouvelle vision de l'analyse canonique des correspondances.

La suite de mon travail sera donc axée, dans un premier temps, sur une exploration des aspects techniques de cette nouvelle perspective en l'appliquant à l'inventaire des carabidae réalisé par le réseau entomologique rhône-alpin. En outre, il est assez saisissant de voir les similarités entre les compilations d'occurrences réalisées à une échelle régionale et les cartographies d'arbres sur une parcelle, pratique courante de la foresterie (dont la bibliographie n'a pas été abordée lors de ce rapport). Lors de ce type d'inventaire, chaque arbre est caractérisé par son espèce, ses coordonnées spatiales, son milieu local afin d'étudier les structures horizontales du peuplement. Nous essaierons donc, dans un deuxième temps, de comparer l'usage de l'analyse canonique des correspondances sur ces deux types de données, listes d'occurrences spatialisées et forêts cartographiées, en réalisant les quelques ajustements qui pourraient s'avérer nécessaires.

BIBLIOGRAPHIE

- BLACKBURN, T. M. & GASTON, K. J.** (1996a) Spatial patterns in the species richness of birds in the New World. *Ecography* : 19, 369-376.
- BLACKBURN, T. M. & GASTON, K. J.** (1996b) Spatial patterns in the geographic range sizes of bird species in the New World. *Philosophical Transactions of the Royal Society, B* : 351, 897-912.
- BÖHNING-GAESE, K.** (1997) Determinants of avian species richness at different spatial scales. *Journal of Biogeography* : 24, 49-60.
- BROWN, J. W. & OPLER, P. A.** (1990) Patterns of butterfly species density in peninsular Florida. *Journal of Biogeography* : 17, 615-622.
- 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. 15-43.
- CONNOR, E. F. & SIMBERLOFF, D.** (1978) Species number and compositional similarity of the Galapagos flora and avifauna. *Ecological Monographs* : 48, 219-248.
- CRONK, Q. C. B.** (1989) The past and the present vegetation of Saint Helena. *Journal of Biogeography* : 16, 47-64.
- DONALD, P. F. & FULLER, R. J.** (1998) Ornithological atlas data: a review of uses and limitations. *Bird Study* : 45, 129-145.
- ELPHICK, C. S.** (1997) Correcting avian richness estimates for unequal sample effort in atlas studies. *IBIS* : 139, 189-190.
- FERRARI, C., BONAFEDE, F. & ALESSANDRINI, A.** (1993) Rare plants of the Emilia-Romagna region (Northern Italy): a data bank and computer-mapped atlas for conservation purposes. *Biological Conservation* : 64, 11-18.
- FLAHAULT, CH.** (1900) Premier essai de nomenclature phytogéographique. In : *Compte rendu du congrès international de botanique*. Paris. 427-450.
- GASTON, K.J. & BLACKBURN, T.M.** (1996) Range size-body size relationships: evidence of scale dependence. *Oikos* : 75, 479-485.
- GASTON, K. J., QUINN, R. M., WOOD, S. & ARNOLD, H. R.** (1996) Measures of geographic range size: effects of sample size. *Ecography* : 19, 259-268.
- GASTON, K. J. & LAWTON, J. H.** (1990) Effects of scale and habitat on the relationship between regional distribution and local abundance. *Oikos* : 58, 329-335.
- GASTON, K. J.** (1991) How large is a species' geographic range ?. *Oikos* : 61, 434-438.
- GASTON, K.J.** (1994) Measuring geographic range sizes. *Ecography* : 17, 198-205.
- GASTON, K. J.** (1996) The multiple forms of the interspecific abundance-distribution relationship. *Oikos* : 76, 211-220.
- GIMARET-CARPENTIER, C., CHESSEL, D. & PASCAL, J.P.** (1998) Non-symmetric correspondence analysis: an alternative for species occurrences data. *Plant Ecology* : 138, 97-112.
- GIMARET-CARPENTIER, C.** (1999) Analyse de la biodiversité à partir d'une liste d'occurrences d'espèces : nouvelles méthodes d'ordination appliquées à l'étude de l'endémisme dans les Ghats occidentaux. *Thèse Doctorat d'état, Université Lyon I* : 235
- GLEICK, J.** (1989) *La Théorie du chaos*. Champs Flammarion. 431.

- GREEN, R.H.** (1971) A multivariate statistical approach to the Hutchinsonian niche: bivalve Molluscs of Central Canada. *Ecology* : 52, 543-556.
- HAWKSWORTH, D. L.** (1995) The Resource Base for Biodiversity Assessments. *In : Global Biodiversity Assessment*. **HEYWOOD, V. H. (ED.)** Cambridge University Press. 545-605.
- HEIKKINEN, R. K.** (1998) Can richness patterns of rarities be predicted from mesoscale atlas data? A case study of vascular plants in the Kevo reserve. *Biological Conservation* : 83, 133-143.
- HILL, M. O.** (1991) Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *Journal of biogeography* : 18, 247-255.
- HURLBERT, S. H.** (1971) The non-concept of species diversity: a critique and alternative parameters. *Ecology* : 52, 577-586.
- HUTCHINSON, G. E.** (1957) Concluding remarks. *In : Population studies: animal ecology and demography*. Cold Spring Harbor Symposia on Quantitative Biology. 22, 415-427.
- KADMON, R. & DANIN, A.** (1997) Floristic variation in Israel: a GIS analysis. *Flora* : 192, 341-345.
- KEATING, K. A. & QUINN, J. F.** (1998) Estimating species richness: the Michaelis-Menten model revisited. *Oikos* : 81, 411-416.
- KERR, J. T.** (1996) Species richness, endemism, and choice of areas for conservation. *Conservation Biology* : 11, 1094-1100.
- LEGAKIS, A. & KYPRIOTAKIS, Z.** (1994) A biogeographical analysis of the island of Crete, Greece. *Journal of Biogeography* : 21, 441-445.
- LETCHER, A. J. & HARVEY, P. H.** (1994) Variation in geographical range size among mammals of the palearctic. *The American Naturalist* : 144,30-42.
- LINDER, H. P.** (1991) Environmental correlates of patterns of species richness in the southwestern Cape Province of South Africa. *Journal of Biogeography* : 18, 509-518.
- MACPHERSON, E. & DUARTE, C. M.** (1994) Patterns in species richness, size, and latitudinal range of East Atlantic fishes. *Ecography* : 17, 242-248.
- MAGURRAN, A. E.** (1988) *Ecological Diversity and Its Measurement*. Croom Helm Ltd., London. 179.
- MOURELLE, C. & EZCURRA, E.** (1996) Species richness of Argentine cacti: a test of biogeographic hypotheses. *Journal of Vegetation Science* : 7, 667-680.
- OWEN, J. G.** (1989) Patterns of herpetofaunal species richness: relation to temperature, precipitation and variance in elevation. *Journal of Biogeography* : 16, 141-150.
- PAGEL, M. D., MAY, R. M. & COLLIE, A. R.** (1991) Ecological aspects of the geographical distribution and diversity of mammalian species. *The American Naturalist* : 137,791-815.
- PATIL, G. P. & TAILLIE, C.** (1979) An overview of diversity. *In : Ecological Diversity in Theory and Practice*. **GRASSLE, J. F., PATIL, G. P., SMITH, W. & TAILLIE, C. (EDS.)** International Cooperative Publishing House, Fairland, MD. 3-27.
- PEARSON, D. L. & GHORPADE, K.** (1989) Geographical distribution and ecological history of tiger beetles (Coleoptera: Cicindelidae) of the Indian subcontinent. *Journal of Biogeography* : 16, 333-344.
- PRENDERGAST, J. R., WOOD, S. N., LAWTON, J. H. & EVERS HAM, B. C.** (1993) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity Letters* : 1, 39-53.

- RANJIT DANIELS, R. J.** (1992) Geographical distribution patterns of amphibians in the Western Ghats, India. *Journal of Biogeography* : 19, 521-529.
- REAKA, M.L.** (1980) Geographic range, life history patterns, and body size in a guild of coral-dwelling mantis shrimps. *Evolution* : 34, 1019-1030.
- RICH, T. C. G. & WOODRUFF, E. R.** (1992) Recording bias in botanical surveys. *Watsonia* : 19, 73-95.
- SCHOENER, T. W.** (1987) The geographical distribution of rarity. *Oecologia* : 74, 161-173.
- SOBERON, J. M. & LLORENTE, J. B.** (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology* : 7, 480-488.
- STEVENS, G.C.** (1986) Dissection of the species-area relationship among wood-boring insects and their host plants. *American Naturalist* : 128, 35-46.
- SWIHART, R. K., SLADE, N. A. & BERGSTROM, B. J.** (1988) Relating body size to the rate of home range use in mammals. *Ecology* : 69, 393-399.
- TERBRAAK, C. J. F.** (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* : 67, 1167-1179.
- TÓTHMÉRÉSZ, B.** (1995) Comparison of different methods for diversity ordering. *Journal of Vegetation Science* : 6, 283-290.
- VIERA DA SILVA, J.** (1979) *Introduction à la théorie écologique*. Masson. 112.
- WHITTAKER, R. H.** (1972) Evolution and measurement of species diversity. *Taxon* : 21, 213-251.
- WILLIAMS, P. H.** (1996) Mapping variations in the strength and breadth of biogeographic transition zones using species turnover. *Proceedings of the Royal Society of London, B* : 263, 579-588.