

L'analyse des données en relation avec l'échantillonnage¹

PIERRE LEGENDRE

Département de Sciences biologiques
Université de Montréal
C.P. 6128, Succursale A
Montréal, Québec H3C 3J7

Résumé

Le premier objectif de cet article est de montrer pourquoi l'échantillonnage et le traitement des données doivent être conçus conjointement, lors de la préparation d'un projet de recherche. Pour ce faire, les différentes catégories d'analyse de données sont inventoriées afin de dégager dans chaque cas quelles sont les contraintes que s'imposent réciproquement l'échantillonnage et l'analyse. La question connexe de la précision des données écologiques est aussi abordée par le biais de la question suivante: vaut-il la peine de chercher à mesurer les phénomènes écologiques avec beaucoup de précision?

MOTS CLÉS: Analyse des données, échantillonnage, écologie, entomologie, modèle, précision des données, structure écologique.

Introduction

L'objectif premier de cet article est de montrer pourquoi l'échantillonnage et le traitement des données doivent être envisagés conjointement, lors de la conception d'un projet de recherche écologique où l'on tente d'appréhender des relations complexes entre composantes des écosystèmes. Quoique les remarques ci-dessous puissent être transposées aux études de dynamique des populations qui considèrent l'influence de plusieurs facteurs, ce texte est surtout conçu en fonction de l'étude des communautés (synécologie), là où les phénomènes à comprendre (structure du peuplement ou de la communauté) de même que les explications possibles sont essentiellement de nature multidimensionnelle.

1. Conférence prononcée à l'occasion du symposium «Les problèmes d'échantillonnage en entomologie: bilan et perspectives», organisé à l'occasion de la 110^e réunion annuelle de la Société d'entomologie du Québec, le 3 novembre 1983, à l'Université du Québec à Trois-Rivières.

Pour atteindre cet objectif, les différentes catégories d'analyse de données seront considérées à tour de rôle, afin de dégager dans chaque cas quelles sont les contraintes que s'imposent réciproquement les étapes d'échantillonnage et d'analyse des données. Nous considérerons différents types d'analyse exploratoire servant à la recherche, puis à l'interprétation des structures écologiques, avant de dire quelques mots de l'analyse confirmatoire d'un modèle écologique.

La question connexe de la précision de l'information écologique à acquérir sera aussi abordée, de façon à lever le mythe selon lequel seule l'information quantitative à haute précision numérique peut faire l'objet d'une analyse.

Types d'analyse de données

Dans les paragraphes qui suivent, les méthodes d'analyse de données seront regroupées en fonction des objectifs des différentes étapes de l'analyse écologique, afin de mettre en évidence leurs relations avec l'échantillonnage. Cet exposé ne constitue pas un cours complet d'analyse des données; les lecteurs intéressés pourront se référer à Pielou (1977), Orłóci (1978), Gauch (1982) ou Legendre et Legendre (1983, 1984a) pour des exposés détaillés sur les méthodes d'analyse des données écologiques, ou encore à Green (1979) ou à Frontier (1982) pour les relations entre les stratégies d'échantillonnage et l'analyse des données.

1. Méthodes de recherche d'une structure

On entend par **structure** l'organisation des échantillons ou des variables écologiques en groupes, ou encore en gradients. Quoique l'observation de certaines variables (plutôt que d'autres) se fasse toujours selon un pré-modèle, implicite ou explicite (Frontier 1982: chapitre 1), l'analyse d'un écosystème commence souvent par une phase **exploratoire** qui a pour but de mettre en évidence l'existence de certaines interrelations (sans présomption de causalité; voir la section 3 ci-dessous à cet effet); elle permet surtout d'éliminer certaines autres relations possibles, que ne supportent pas les données.

Les données récoltées par les écologistes des communautés suivent trois axes: les différentes variables (ou descripteurs) observées, l'espace (stations) et le temps (moments). Si on représente toutes ces observations au sein d'un cube de données (fig. 1), il s'en dégage six modes principaux d'analyse (Cattell 1966):

- **Mode O** = comparaison entre les moments, pour l'ensemble des descripteurs (une seule station);
- **Mode P** = comparaison entre descripteurs, pour l'ensemble des moments (une seule station);
- **Mode Q** = comparaison entre stations, pour l'ensemble des descripteurs (un seul moment);
- **Mode R** = comparaison entre descripteurs, pour l'ensemble des stations (un seul moment);
- **Mode S** = comparaison entre stations, pour l'ensemble des moments (un seul descripteur);
- **Mode T** = comparaison entre moments, pour l'ensemble des stations (un seul descripteur).

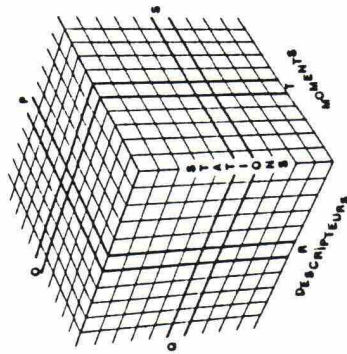


Figure 1. Le cube des données de l'écologiste. Adapté de Cattell (1966).

Les différents types d'analyses seront situés par rapport à ces six modes. Rappelons enfin que les descripteurs se présentent sous trois formes mathématiques: descripteurs quantitatifs, semi-quantitatifs et qualitatifs.

1.1 - Analyses générales

Les méthodes appartenant à cette catégorie sont utilisées par toutes les sciences d'observation pour l'analyse des données multidimensionnelles. Elles ne contiennent aucune prémisses particulière à l'une de ces sciences et ne prennent pas en compte l'existence de relations *a priori* (spatiales ou temporelles) entre les échantillons écologiques, comme ce sera le cas à la section 1.2. Ces méthodes visent essentiellement à former des groupes ou à identifier des gradients. Groupes et gradients sont deux structures que l'écologiste peut chercher à reconnaître dans ses données.

Cette recherche procède habituellement selon le schéma de la figure 2. À partir d'un tableau de données d'observations — correspondant à l'un des plans du cube de données —, on calcule une matrice de ressemblance, à l'aide de coefficients de similarité, de distance ou de dépendance (similarités entre objets, en haut et à gauche de la fig. 2). Cette matrice, carrée et habituellement symétrique, est à la base des groupements (hiérarchique, dans la figure) ou des ordinations en espace réduit (la méthode illustrée dans la figure est l'analyse en coordonnées principales). Les réseaux de groupement peuvent être superposés aux ordinations pour faciliter la compréhension de la structure (dans la figure, on a superposé à l'ordination le réseau de groupement formé par la chaîne primaire de liens de similarité, plutôt qu'une partition des objets extraite de la hiérarchie). Enfin, la figure montre qu'à ces structures dérivées — groupements et ordinations — correspondent des matrices de ressemblance qui peuvent être comparées à leur tour à la matrice de ressemblance initiale, pour juger de la perte d'information consécutive à la synthèse obtenue. La chaîne d'analyses

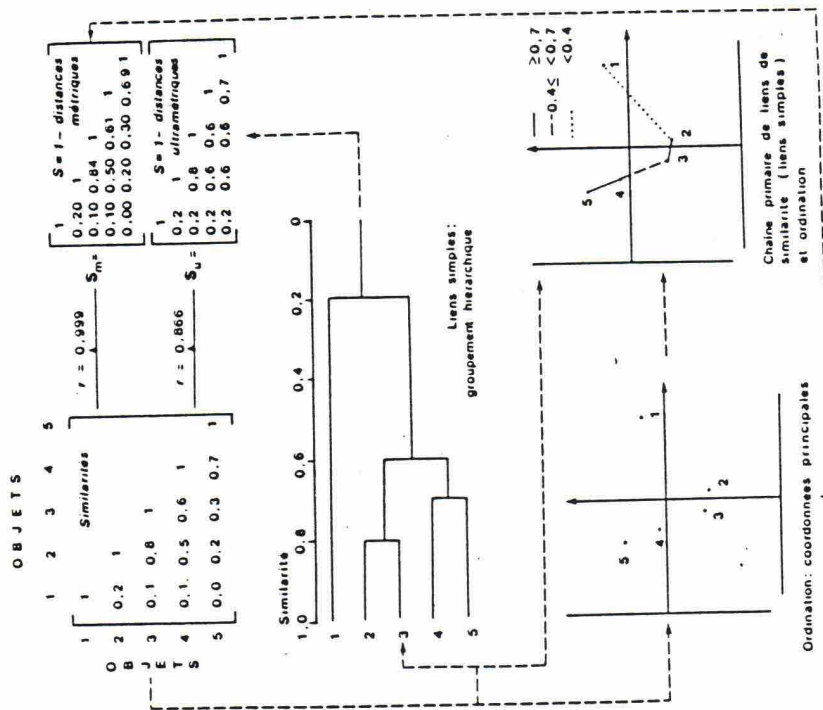


Figure 2. Exemple de recherche de la structure d'une matrice fictive de similarité, par groupement et ordination en espace réduit. Tiré de Legendre et Legendre (1982).

Tableau 1. Coefficients d'association en mode Q, classés selon le type de données auxquelles ils s'appliquent. Modifié de Legendre et Legendre (1982).

	Abondances d'espèces: coefficients asymétriques	Autres descripteurs: coefficients symétriques
Descripteurs quantitatifs	Dénombréments bruts: Similarités de Steinhaus et de Kulczynski, indice d'association de Whittaker, métrique du χ^2 , distance du χ^2 , mesure de corde, métrique de Canberra et ses variantes.	Entre échantillons: Similarité de Gower (symétrique), distance euclidienne, distance moyenne («taxonomique»), distance de Manhattan, différence moyenne des descripteurs.
Descripteurs semi-quantitatifs	Données normalisées (distribution symétrique): Comme dans le cas semi-quantitatif, ci-dessous.	Entre groupes d'échantillons: Distance généralisée de Mahalanobis, coefficient de ressemblance raciale.
Descripteurs qualitatifs	Similarités de Steinhaus et Kulczynski, similarité de Gower (asymétrique), coefficient de Legendre & Chodorowski, coefficient probabiliste de Goodall.	Coefficient général de similarité d'Estabrook & Rogers.
	Coefficient de communauté (Jaccard), indice de Sørensen et leurs variantes.	Similarité de Gower (symétrique), coefficient d'Estabrook & Rogers, coefficient de simple concordance et ses variantes.

formée d'un calcul de ressemblance, suivi de méthodes générales de groupement et d'ordination, est employée principalement en écologie dans les modes P, Q, R et S de la figure 1. La spécificité de l'analyse provient en général de la mesure de ressemblance, dont le choix est crucial puisque celle-ci doit s'adapter au type mathématique des données (quantitatives, semi-quantitatives ou qualitatives) ainsi qu'au mode d'analyse. Il existe cependant des coefficients de ressemblance adaptés à tous les types mathématiques de données (voir par exemple le tableau 1, pour le mode Q d'analyse). Certains coefficients permettent d'ailleurs de combiner plusieurs types de descripteurs en une mesure unique de la ressemblance. De même, la panoplie des méthodes de groupement et d'ordination permet le traitement de tous les types de descripteurs (tableau 2).

Tableau 2. Méthodes de recherche de la structures d'un tableau de données écologiques. Modifié de Legendre et Legendre (1982).

Groupement	Descripteurs quantitatifs seulement	Descripteurs de tout niveau de précision
	Tous les types de groupement, à partir d'un coefficient d'association approprié (tableau 1).	
Ordination	Analyse en composantes principales	Analyse des correspondances - Analyse en coordonnées principales - Cadrage métrique ou non-métrique - etc.

Depierreux *et al.* (1983) ont étudié la colonisation de substrats artificiels par les invertébrés benthiques, en rivière, à l'aide de différentes méthodes de groupement hiérarchique ainsi que de l'analyse en composantes principales. Ils ont ainsi mis en évidence des groupes de stations écologiques correspondant à un gradient de qualité des eaux. D'autres auteurs ont réalisé des divisions de ruisseaux ou de rivières en zones homogènes quant à leur peuplement de macroinvertébrés benthiques, par analyse des correspondances associée ou non aux résultats d'un groupement hiérarchique (Lapchin et Roux 1977; Nelva 1979; Culp et Davies 1980).

On peut aussi employer ces méthodes générales d'analyse pour l'étude de problèmes particuliers à l'écologie, par exemple la recherche d'associations d'espèces, dans l'espace (mode R d'analyse) ou selon l'axe temporel (mode P). Ce problème écologique se traduit dans la méthodologie par le choix de mesures de ressemblance spéciales, ainsi que par la recherche d'une partition (division en groupes) plutôt que d'un groupement hiérarchique (voir Legendre et Legendre 1984a). Le chapitre 13 de Southwood (1978) présente un exposé méthodologique de cette question, axé sur l'étude des peuplements d'insectes. Ghetti *et al.* (1979) ont employé le groupement agglomératif hiérarchique pour la recherche d'associations d'espèces dans la communauté de macroinvertébrés benthiques d'une rivière d'Italie, alors que Nelva (1979) a utilisé un groupement non-hiérarchique aux mêmes fins, pour la faune benthique de ruisseaux du Jura.

Le plan d'échantillonnage n'impose que rarement des contraintes quant au choix d'une méthode d'analyse de données appartenant à la catégorie des analyses générales. Mentionnons cependant un exemple: un échantillonnage aléatoire simple ou un échantillonnage systématique conviennent au groupement selon l'association moyenne («UPGMA» dans Sneath et Sokal 1973) ainsi qu'au groupement centride («UPGMC»), alors qu'un échantillonnage stratifié ou à probabilités inégales requerra plutôt un groupement à poids proportionnels («WPGMA») ou médian («WPGMC»). Dans la majorité des cas, cependant, c'est au niveau de l'interprétation des résultats des analyses qu'il convient de faire intervenir le plan d'échantillonnage: la structure observée reproduit-elle simplement le plan d'échantillonnage, ou apporte-t-elle en plus de l'information quant à la nature des processus écologiques ayant cours dans l'écosystème? (Voir aussi Frontier 1974).

1.2. Analyses prenant en compte les relations a priori entre échantillons

De nouvelles méthodes ont été développées récemment, qui incluent certains éléments du modèle écologique dans le cours de l'analyse (de la même façon qu'en taxonomie numérique, les méthodes cladistiques quantitatives introduisent dans le procédé d'analyse certains éléments de la théorie évolutive). Les écologistes ont ainsi emprunté depuis quelques années certains modèles d'analyse des séries de données, correspondant bien aux besoins d'analyse des processus écologiques temporels. Puisqu'il s'agit d'un champ de recherche en développement actif, nous nous contenterons de l'illustrer par quelques exemples.

79 70 60 56
0987654321098765432109876

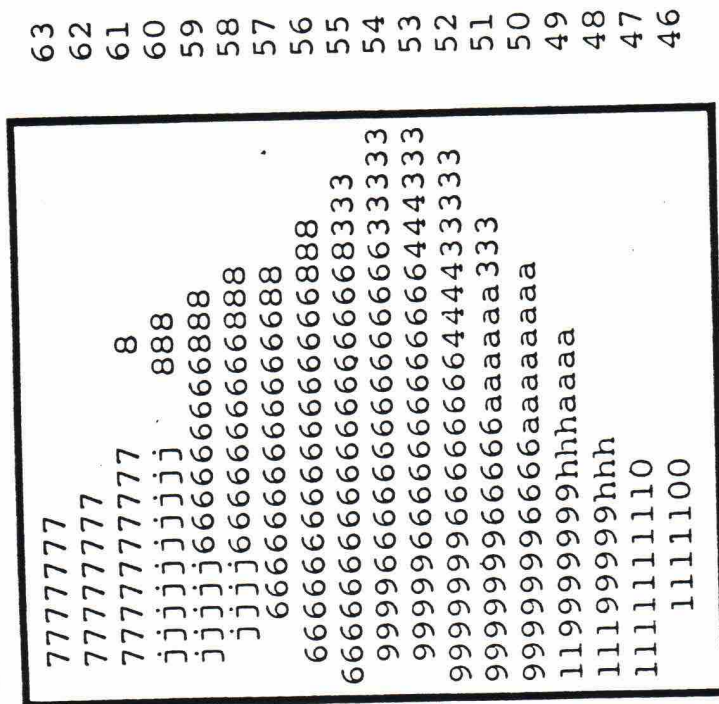


Figure 3. Résultat d'un groupement biogéographique avec contrainte de contiguïté spatiale (carte inédite de l'étude de Legendre 1984b, sur les poissons du Québec). Cette carte du Québec montre l'agglomération réalisée au niveau de similarité de 0,70. Tous les pixels (surfaces d'un degré de côté) réunis dans une même région sont représentés par une même lettre ou un même chiffre. Les degrés de longitude ouest et de latitude nord sont indiqués par le programme en marge de la carte.

1. Les programmes de groupement biogéographique, écrits en langage PASCAL, sont disponibles gratuitement pour les chercheurs gouvernementaux et universitaires.

Dans les deux exemples ci-dessus, il n'est pas nécessaire que l'échantillonnage soit systématique, quoiqu'un tel échantillonnage puisse faciliter l'interprétation des résultats. Il faut cependant connaître la relation (spatiale ou temporelle) existant entre les échantillons, puisque l'ordre des échantillons intervient dans l'analyse. Si l'échantillonnage n'est pas systématique, l'écologiste devra en tenir compte lors de l'interprétation des résultats, puisque sa maille d'observation variera d'une région à l'autre, dans l'espace ou dans le temps.

La recherche de phénomènes périodiques dans le comportement temporel des écosystèmes intéresse de nombreux écologistes des populations et des communautés, puisque la reconnaissance de telles périodes constitue un élément important des modèles de prédiction des variations d'abondance des espèces. Ces fluctuations périodiques peuvent être contrôlées par la périodicité des phénomènes géophysiques (rythme circadien, marées, système de circulation atmosphérique, rythmes annuels...) ou encore être engendrées par l'écosystème lui-même. Les méthodes permettant la recherche de telles périodes sont présentées au tableau 3. On y voit que plusieurs d'entre elles — celles qui permettent l'analyse de descripteurs semi-quantitatifs ou qualitatifs — sont d'un développement récent, axé souvent sur les problèmes écologiques. Comme exemple d'application à la biologie des populations d'insectes, Campbell et Shipp (1974) ont employé l'analyse spectrale pour la recherche des périodes caractéristiques des rythmes de l'activité locomotrice chez un criquet australien, *Teleogryllus commodus*.

Tableau 3. Méthodes d'analyse des périodes caractéristiques de séries de données écologiques, selon leur niveau de précision. Modifié de Legendre et Legendre (1984).

	Descripteurs quantitatifs seulement	Descripteurs de tout niveau de précision
Un seul descripteur	<ul style="list-style-type: none"> - Corrélogramme - Périodogramme - Spectres (de variance, par entropie maximum) 	<ul style="list-style-type: none"> - Périodogramme de contingence (Legendre <i>et al.</i> 1981) - Spectre (descripteurs binaires) (Kedem 1980)
Deux descripteurs	<ul style="list-style-type: none"> - Corrélation paramétrique avec retard - Spectres de cohérence et/ou de phase 	<ul style="list-style-type: none"> - Corrélation non-paramétrique avec retard (Fortier et Legendre 1979) - Contingence avec retard (Legendre et Legendre 1982)
Multidimensionnel	<ul style="list-style-type: none"> - Auto D² (Ibanez 1976) - Spectres multivariés (Brillinger 1975) 	<ul style="list-style-type: none"> - Groupement suivi d'un périodogramme de contingence (Legendre <i>et al.</i> 1981)

Ces méthodes ont en commun la nécessité d'un échantillonnage systématique à intervalles fixes (on peut cependant concevoir que l'échelle sous-jacente puisse être autre que le temps, par exemple le nombre de degrés-jours ou une autre variable appropriée, chez les poikilothermes). Par ailleurs, la recherche de phénomènes périodiques génère l'obligation d'un échantillonnage suffisamment serré pour résoudre les périodes les plus fines, dans la gamme des longueurs d'ondes qui intéressent l'écologiste, et aussi suffisamment prolongé pour couvrir au moins deux fois la période la plus longue.

Cette dernière remarque s'applique aussi à la recherche de structures régulières dans l'espace. La méthode de l'autocorrélation spatiale (Cliff et Ord 1981; Legendre et Legendre 1984), ainsi que le variogramme (qui fait

partie de la méthode géostatistique, revue par David 1977, ainsi que par Journel et Huijbregts 1978), permettent d'identifier la présence de phénomènes périodiques dans l'espace (structures agrégées ou en taches) ou encore des gradients spatiaux. Pour l'analyse spatiale des phénomènes écologiques, il est préférable de disposer d'un échantillonnage systématique de la surface, quoique les techniques puissent s'accommoder d'autres types d'échantillonnage.

2. Méthodes d'interprétation d'une structure

Rappelons brièvement le principe suivi par nombre d'analyses écologiques. Les écologistes cherchent souvent à établir des relations entre le monde biologique et le monde abiotique, ou encore entre deux compartiments biologiques, de façon à confirmer des hypothèses quant aux mécanismes contrôlant le monde biologique, ou encore pour prévoir son évolution. Pour ce faire, au moins deux ensembles de données doivent être récoltés: par exemple, un tableau de données concernant un compartiment biologique, un autre tableau de données concernant les aspects physiques, chimiques, géologiques, géomorphologiques ou autres des mêmes stations d'échantillonnage. Les analyses exploratoires visant à décrire la structure (section 1, ci-dessus) sont habituellement menées sur l'un de ces tableaux, par exemple les données biologiques, la structure résultante étant par la suite comparée aux données du deuxième tableau pour fins d'interprétation. Ceci constitue un schéma asymétrique d'analyse. Plus rarement, une structure est extraite de façon symétrique de chacun des tableaux, puis ces structures sont comparées entre elles. Legendre et Legendre (1984, ch. 10) présentent des tableaux comparatifs des différentes méthodes disponibles pour l'analyse d'une structure (groupement ou gradient). Plusieurs de ces méthodes, parmi celles adaptées aux descripteurs semi-quantitatifs ou qualitatifs, ont été développées récemment et sont dues aux psychométriciens.

Il suffit, pour notre propos, de noter que ces méthodes n'imposent aucune contrainte particulière à l'échantillonnage, sauf bien sûr la nécessité de récolter aussi lors de l'échantillonnage les données qui seront nécessaires à l'interprétation de la structure.

3. Analyse confirmatoire d'un modèle

La modélisation écologique est un exercice de représentation de relations causales multiples en vue de faire des prévisions ou des prédictions (voir Legendre et Legendre 1984, ch. 10 pour la distinction de ces deux objectifs). Ces relations auront elles-mêmes été découvertes par réflexion sur la théorie écologique, ainsi que par les phases d'analyse des données mentionnées ci-dessus. Un modèle peut à son tour être confirmé (ou infirmé) par l'analyse de données pertinentes. Nous en donnerons trois exemples.

Les modèles de causalité constituent une famille de modèles intéressants pour l'écologiste (fig. 4). Dans ces modèles, on essaie de prédire le comportement d'un ou de plusieurs descripteurs (Y) à l'aide de variables prédictives (X) mesurées dans le milieu. Le modèle doit prévoir des variables résiduelles (U) représentant l'ensemble des facteurs non tenus en compte par le modèle, qui influencent le comportement des variables Y. De tels modèles peuvent être confirmés à l'aide de l'analyse des coefficients de direction («path analysis», en anglais) qui est elle-même dérivée de l'analyse de régression.

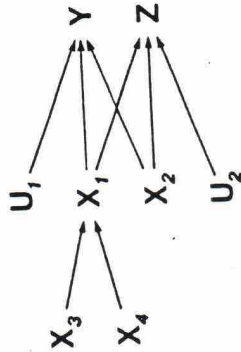


Figure 4. Exemple de modèle de causalité (adapté de Sokal et Rohlf 1981). Y et Z sont deux variables à prédire à l'aide des prédicteurs X_1 à X_4 . U_1 et U_2 sont des variables non mesurées, représentant la variance résiduelle de Y et Z respectivement (variance non expliquée par le modèle).

Un second exemple touche les relations temporelles de cause à effet que peut contenir un modèle. Il est en effet possible de confirmer l'existence de telles relations, même si un retard existe entre l'effet et sa cause présumée. Les méthodes appropriées portent le nom de corrélation avec retard, contingence avec retard, spectre de cohérence ou de phase (tableau 3). Leur mise en oeuvre requiert bien entendu un échantillonnage suffisamment fin pour permettre la mise en évidence du délai d'action présumé.

Un dernier exemple touche les hypothèses que l'écologiste peut formuler quant aux processus générateurs des structures de répartition spatiale des organismes. De telles hypothèses peuvent être testées, soit par la méthode de l'autocorrélation spatiale mentionnée plus haut (dans le cas de structures régulières), soit à l'aide du test de Mantel (1967; Sokal 1979) qui est l'équivalent d'une corrélation adaptée aux prélèvements liés par une structure d'échantillonnage non-aléatoire.

Ces analyses confirmatoires de modèles n'apportent par elles-mêmes aucune contrainte nouvelle à l'échantillonnage. Cependant, la phase de validation d'un modèle, déjà confirmée à l'aide d'un premier jeu de données, nécessite le recours à un nouvel ensemble de données afin de vérifier la valeur prédictive du modèle (Swartzman 1980).

Precision de l'information

Considérant les différents niveaux de précision de l'information, les écologistes initiés à la statistique paramétrique peuvent avoir l'impression que les descripteurs quantitatifs ont une plus grande valeur intrinsèque, pour fins d'analyse de données, que les descripteurs semi-quantitatifs ou

qualitatifs. Cette croyance peut déterminer différentes composantes de l'échantillonnage: les descripteurs à observer, les appareils de mesure ou d'enregistrement nécessaires, ainsi que le nombre d'échantillons qu'il est possible de traiter, toutes considérations qui déterminent ou qui sont déterminées par le budget de la campagne d'échantillonnage. Il convient donc de se poser la question suivante: est-il nécessaire — ou vaut-il la peine — de chercher à mesurer l'information écologique avec grande précision?

Trois éléments complémentaires de réponse peuvent être apportés à cette question:

1. Les phénomènes écologiques mesurables en nature contiennent en général une composante aléatoire tellement importante que la notation semi-quantitative ou qualitative de l'information est souvent suffisante, les phénomènes plus fins n'étant possiblement mesurables qu'en laboratoire, dans des conditions expérimentales contrôlées.
2. La notation de l'information sous forme semi-quantitative (par exemple, par utilisation d'une cotation d'abondance au lieu du dénombrement exhaustif des espèces) ou qualitative permet dans la plupart des cas de sauver du temps, ce qui permet de réduire le budget, ou encore d'augmenter la représentativité de l'échantillonnage par l'analyse d'un plus grand nombre d'échantillons.
3. Par ailleurs, on a vu (tableaux 1 à 3) que la panoplie des méthodes d'analyse de données disponibles n'impose plus de contrainte à l'écologiste quant au choix d'un niveau de précision irréaliste.

Pour toutes ces raisons, l'écologiste qui juge suffisant d'observer la variabilité de ses échantillons avec un niveau de précision intermédiaire ou bas devrait le faire, puisque ceci lui permettra d'optimiser le rendement de son investissement (en temps ou en argent). Cette tactique peut par ailleurs lui permettre d'augmenter le nombre d'échantillons observés, et par là, sa connaissance de la variabilité spatiale ou temporelle du phénomène, ce qui est hautement souhaitable.

Plan d'une recherche écologique

Les paragraphes précédents ont tenté de mettre en évidence les interrelations existant entre les objectifs d'une étude, l'échantillonnage et le traitement des données. La figure 5 place ces relations dans la perspective d'un schéma global de la recherche écologique.

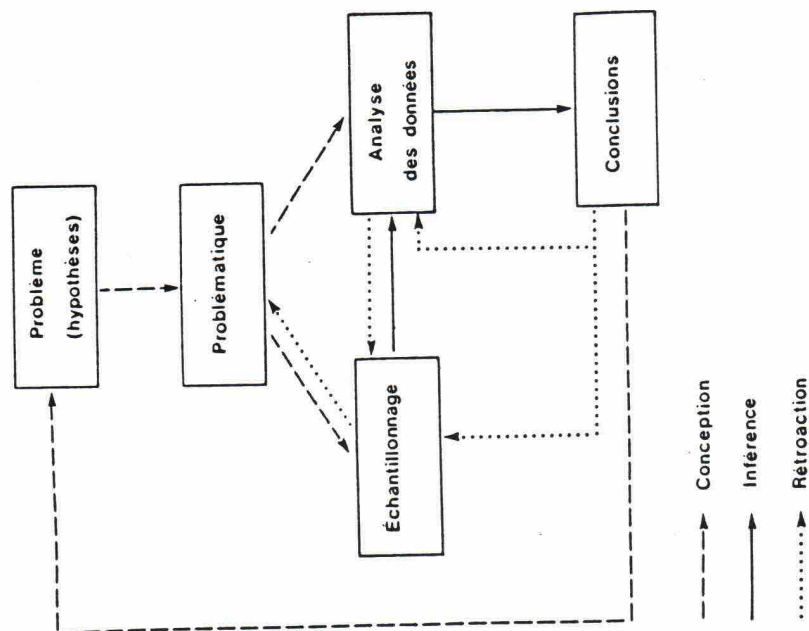


Figure 5. Relations entre les étapes de la recherche écologique. Tiré de Legendre et Legendre (1982).

La relation entre la problématique (objectifs, hypothèses tirées d'un prémodèle) et l'échantillonnage est bien connue. Rappelons à titre d'exemple qu'une population ne doit pas être échantillonnée de la même façon qu'un peuplement. Dans le premier cas, il suffit de prélever, dans l'habitat de l'espèce-cible, un échantillon suffisamment grand d'organismes appartenant à cette espèce pour que la composition de cet échantillon se rapproche de celle de la population visée par l'étude. Dans le second cas, il est nécessaire de délimiter la zone étudiée et d'y répartir l'effort d'échantillonnage de façon aléatoire, afin d'estimer correctement la composition spécifique du peuplement, ainsi que sa répartition spatiale entre les différentes composantes de la mosaïque des habitats qu'on y trouve, s'il y a lieu.

Green (1979) nous rappelle par ailleurs qu'il faut limiter la récolte des données à celles qui sont clairement identifiées comme pertinentes aux objectifs de l'étude, puisque la plupart des données qui contribuent à la solution de certains problèmes n'apportent aucune information en relation avec d'autres problèmes (tableau 1.1 de Green). De plus, les objectifs d'une étude déterminent la maille de l'échantillonnage, spatialement et

temporellement. La relation opposée montrant l'influence de l'échantillonnage sur les objectifs rappelle que les contraintes techniques et logistiques obligent parfois à modifier les objectifs d'une étude, surtout lorsque de l'équipement lourd ou dispendieux est nécessaire, ou lorsque la réalisation d'un objectif implique la collaboration de personnel extérieur à l'équipe immédiate.

La case «échantillonnage» de la figure 5 est en fait constituée de deux composantes: le QUOI (choix des variables) et le COMMENT (plan d'échantillonnage). L'influence de l'échantillonnage sur l'analyse des données consiste d'abord en ce que le choix des variables, ainsi que leur niveau de précision, déterminent la méthode d'analyse à employer pour atteindre chaque objectif particulier. Il importe donc que l'écologiste connaisse un ensemble de méthodes d'analyse, plutôt que de se trouver limité dans son choix de descripteurs par une palette de méthodes trop limitée. D'autre part, le plan d'échantillonnage détermine la population statistique de référence dont on veut connaître les caractéristiques et à laquelle on pourra, en retour, appliquer les conclusions de l'analyse. La flèche opposée, pointant de l'analyse des données vers l'échantillonnage, rappelle que dans certains cas particuliers, la méthode d'analyse des données impose un plan d'échantillonnage particulier. Par exemple, l'échantillonnage doit être systématique pour certaines analyses spatiales ou temporelles, ou encore aléatoire simple ou systématique pour certaines méthodes de groupement.

Enfin, l'influence de la problématique sur le plan d'analyse des données réside dans la nécessité de choisir une méthode d'analyse permettant de répondre le mieux possible à la question écologique posée. La recherche méthodologique se poursuit, pour développer — ou emprunter de domaines connexes comme la sociométrie, la psychométrie, l'économétrie ou la géographie — des modèles d'analyse de données contenant le plus possible d'éléments ou de contraintes qui correspondent à la formulation du problème écologique.

Au vu de cet ensemble de relations de dépendance, il devient évident que problématique, plan d'échantillonnage et plan d'analyse des données doivent être conçus simultanément, au cours de la phase de préparation du projet de recherche, quitte à être améliorés ou complétés en cours d'étude.

Références

- BRILLINGER, D.R. 1975. Time series: data analysis and theory. Holt, Rinehart & Winston, New York. xii + 500 pp.
- CAMPBELL, D.J. and E. SHIPP. 1974. Spectral analysis of cyclic behaviour with examples from the field cricket *Teleogryllus commodus* (Walk.). *Anim. Behav.* 22: 862-875.
- CATTELL, R.B. 1966. The data box: its ordering of total resources in terms of possible relational systems. pp. 67-128 in R.B. Cattell (Ed.) *Handbook of multivariate experimental psychology*. Rand McNally & Co., Chicago. xxxi + 959 pp.
- CUFF, A.D. and J.K. ORR. 1981. Spatial processes: models and applications. Pion Limited, London. 266 pp.
- CUPP, J.M. and R.W. DAVIES. 1980. Reciprocal averaging and polar ordination as techniques for analyzing lotic macroinvertebrate communities. *Can. J. Fish. Aquat. Sci.* 37: 1358-1364.
- DAVID, M. 1977. Geostatistical ore reserve estimation. Developments in geostatistics, 2. Elsevier Scient. Publ. Co., Amsterdam. xx + 364 pp.
- DEPIERUX, E., E. FERTMANS et J.-C. MICHA. 1983. Utilisation critique de l'analyse en composantes principales et du cluster analysis pour la description d'échantillons d'invertébrés benthiques en eau douce. *Oikos* 40: 81-84.



Revue d'entomologie du Québec

Vol. 29 no 2

Novembre 1984

- FORTIER, L. et L. LEGENDRE. 1979. Le contrôle de la variabilité à court terme du phytoplancton estuarien: stabilité verticale et prolongement critique. *J. Fish. Res. Board Can.* 36: 1325-1335.
- FRONTIER, S. 1974. L'analyse factorielle est-elle heuristique en écologie du plancton? *Can. O.R.S.T.O.M., Sér. Océanogr.* 12: 77-81.
- FRONTIER, S. (Ed) 1982. Stratégies d'échantillonnage en écologie. Collection d'écologie, 17. Masson, Paris et les Presses de l'Université Laval. xviii + 494 pp.
- GAUCH, H.G., Jr. 1982. Multivariate analysis in community ecology. Cambridge Univ. Press, Cambridge. x + 298 pp.
- GHETTI, P.F., G. BONAZZI et M. ALFARONE. 1979. Il torrente Parma: criteri statistici per una tipizzazione biologica. *Lavori della Società Italiana di Biogeografia*, nuova serie 6 (1976): 461-478.
- GREEN, R.H. 1979. Sampling design and statistical methods for environmental biologists. Wiley, New York. xi + 257 pp.
- IBÁÑEZ, F. 1976. Contribution à l'analyse mathématique des événements en écologie planctonique. Optimisations méthodologiques: étude expérimentale en continu à petite échelle de l'hétérogénéité du plancton côtier. Thèse de doctorat d'état en sciences naturelles, Université Paris VI. *Bull. Inst. Océanogr.*, Monaco 72 (1431): 1-96.
- JOURNEL, A.G. and C.J. HUIBREGTS. 1978. Mining geostatistics. Academic Press, London. x + 600 pp.
- KEDEM, B. 1980. Binary time series. Lecture notes in pure and applied mathematics, vol. 39. Marcel Dekker, New York. ix + 140 pp.
- LAPCHIN, L. et C. Roux. 1977. Utilisation de l'analyse des correspondances pour l'étude de la répartition des invertébrés benthiques dans deux ruisseaux à salmonides de Bretagne. *Ann. Hydrobiol.* 8: 333-354.
- LEGENDRE, L., M. FRECHETTE and P. LEGENDRE. 1981. The contingency peridogram: a method of identifying rhythms in series of nonmetric ecological data. *J. Ecol.* 69: 965-979.
- LEGENDRE, L. et P. LEGENDRE. 1983. Numerical ecology. Developments in environmental modelling, 3. Elsevier Scient. Publ. Co., Amsterdam. xvi + 419 pp.
- LEGENDRE, L. et P. LEGENDRE. 1984a. Ecologie numérique (2ème édition revue et augmentée). Tome 1: Le traitement multiple des données écologiques. Tome 2: La structure des données écologiques. Collection d'écologie, 12 et 13. Masson Paris et les Presses de l'Université du Québec. 260 p. et 335 p.
- LEGENDRE, P., S. DALLOT and L. LEGENDRE. 1985. Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. *American Nat.* 125: (sous presse).
- LEGENDRE, P. & L. LEGENDRE. 1982. Échantillonnage et traitement des données. pp. 163-216 in S. Frontier (Ed.). *Stratégies d'échantillonnage en écologie*. Collection d'écologie, 17. Masson, Paris et les Presses de l'Université Laval. xviii + 494 pp.
- LEGENDRE, P. and V. LEGENDRE. 1984b. Post-glacial dispersal of freshwater fishes in the Québec peninsula. *Can. J. Fish. Aquat. Sci.* 41: 1781-1802.
- MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209-220.
- NELVA, A. 1978. Structure écologique de trois rivières du Jura méridional d'après l'analyse multivariée de la macrofaune benthique. *Arch. Hydrobiol.* 87: 327-346.
- ORLOCI, L. 1978. Multivariate analysis in vegetation research (2nd edition). W. Junk Publ. Co., The Hague. ix + 451 pp.
- PIELOU, E.C. 1977. Mathematical ecology (2nd edition). J. Wiley & Sons, New York. x + 385 pp.
- SNATH, P.H.A. and R.R. SOXAL. 1973. Numerical taxonomy — The principles and practice of numerical classification. W.H. Freeman, San Francisco. xv + 573 pp.
- SOXAL, R.R. 1979. Testing statistical significance of geographic variation patterns. *Syst. zool.* 28: 227-232.
- SOXAL, R.R. and F.J. ROHLF. 1981. Biometry — The principles and practice of statistics in biological research (2nd edition). W.H. Freeman, San Francisco. xviii + 859 pp.
- SOUTHWOOD, T.R.E. 1978. Ecological methods with particular reference to the study of insect populations. Second edition. Chapman and Hall, London. xxiv + 524 pp.
- SPRULS, W.G. 1980. Nonmetric multidimensional scaling analysis of temporal variation in the structure of limnetic zooplankton communities. *Hydrobiologia* 69: 139-146.
- SWARTZMAN, G. 1980. Evaluation of ecological simulation models. pp. 230-267 in W.M. Getz (Ed.). *Mathematical modelling in biology and ecology*. Lecture Notes in Biomathematics, 33. Springer-Verlag, Berlin. viii + 355 pp.

SOMMAIRE

A species of Acanthomidae (Prolura) new to Canada.....	55
<i>Hybomitra aequalincta</i> et <i>H. zonalis</i> (Diptera: Tabanidae): clé d'identification, distribution et abondance saisonnière au Québec-Labrador.....	62
Population fluctuations of northern corn rootworm (Coleoptera: Chrysomelidae) in a Quebec cornfield.....	74
<i>Alloperla scadiana</i> n.sp. (Plecoptera: Chloroperlidae) du Nouveau-Brunswick.....	83
Colloque: «Les problèmes d'échantillonnage en entomologie: bilan et perspectives».....	
L'analyse des données en relation avec l'échantillonnage.....	86
Stratégies d'échantillonnage établies pour la détection et la surveillance des insectes en milieu forestier.....	101
Objectifs et stratégies d'échantillonnage des insectes en milieu agricole.....	114
Index des auteurs.....	127
Index des sujets.....	128

(Reçu le 26 janvier 1984, accepté le 28 mai 1984)