

## Analyse des Correspondances de Données Planifiées: Etude de la Chémotaxie de la Larve Infestante d'un Parasite

Laurent Fraile,<sup>1</sup> Yves Escoufier,<sup>2</sup> et André Raibaut<sup>1</sup>

<sup>1</sup>Laboratoire d'Ichthyologie et de Parasitologie Générale,  
Université Montpellier II, place E. Bataillon  
34095 Montpellier Cedex 5, France

<sup>2</sup>Unité de Biométrie, Ensa. M-Inra-Um 2, 9 place Pierre Viala  
34060 Montpellier Cedex 01, France

### RÉSUMÉ

Une étude de l'influence du mucus sur le déplacement du copépodite infestant, *Caligus minimus*, parasite buccal du loup de mer *Dicentrarchus labrax*, conduit à un tableau de données constitué de 42 histogrammes donnant la position finale de ces larves en fonction de la durée de l'expérience et du mucus (nature, quantité).

Mariant la technique de l'analyse des correspondances adaptée à la comparaison des histogrammes à des procédures de projection et de tests inspirées de l'analyse de variance, on met en évidence un effet significatif de glissement de tendance centrale dû au mucus du loup et un effet significatif d'étalement des histogrammes dû à la durée.

### 1. Introduction

Les présentations de la statistique partitionnent fréquemment les démarches et les méthodes proposées en une classe dite exploratoire et une classe dite confirmatoire (Tukey et Mallows, 1982).

Les travaux relevant de la première classe ont pour objectif de faire apparaître des associations d'individus ou des liaisons entre variables tandis que ceux qui relèvent de la seconde classe sont plus dédiés à la confrontation entre des hypothèses et les données disponibles.

L'analyse des correspondances a été dès l'origine présentée comme une méthode exploratoire (Benzécri, 1973). Elle a été et est utilisée largement dans cet esprit en particulier lorsque les ensembles de données soumis à l'étude sont volumineux et issus d'enquêtes sociologiques n'ayant pas suivi des plans de sondages stricts. Aux réponses aux questions d'une enquête s'ajoutent souvent des informations sur les caractéristiques des sujets enquêtés; ces caractéristiques sont rarement définies à l'avance et sont simplement observées au même titre que les réponses aux questions. La pratique usuelle consiste alors à considérer ces caractéristiques comme des variables illustratives et à étudier leur lien avec les réponses aux questions par la procédure des points supplémentaires (Lebart, Morineau, et Fenelon, 1979).

Pénétrant les disciplines plus expérimentales des sciences de la nature ou de la vie, l'analyse des correspondances est mise en oeuvre par des chercheurs habitués à mener des expériences contrôlées, c'est-à-dire à définir avant toute expérience les caractéristiques des individus sur lesquels les observations seront faites ainsi que les conditions dans lesquelles ces observations seront faites. Il devient alors illogique de n'envisager d'utiliser ces caractéristiques et ces conditions que comme des variables illustratives après avoir fait tous les calculs alors que leur choix est significatif des interrogations qui justifient l'expérience: des individus de caractéristiques différentes fournissent-ils des réponses différentes? Les conditions expérimentales influencent-elles les résultats?

De nombreuses publications récentes ont montré l'intérêt de cette approche. S'inspirant des résultats de l'analyse en composantes principales par rapport à des variables instrumentales (Rao, 1964; Robert et Escoufier, 1976) Sabatier (1987) a fourni les bases mathématiques de ce développement. De nombreuses applications écologiques ont été publiées: Cazes, Chessel, et Doledec (1988), Lebreton et Yoccoz (1987), Lebreton et al. (1988), Sabatier, Lebreton, et Chessel (1989), Yoccoz et Chessel (1988), Ter Braak (1986, 1987). Ces travaux rejoignent des articles récents de Nishisato et Lawrence (1988),

---

*Mots clés:* Analyse des correspondances; Analyse des données planifiées; Chémotaxie; Copépodes parasites; Décomposition de l'inertie; Poissons; Tests de permutation.

Pontier, Jolicoeur, et Pernin (1987) et prolongent des propositions plus anciennes de Lebart et Fenelon (1975) et Rouanet et Lepine (1976).

L'originalité du travail présenté dans ce texte réside dans la prise en compte stricte d'un plan d'expérience à deux facteurs, dans le cadre d'une analyse des correspondances, l'utilisation systématique de tests de permutation pour apprécier la signification des effets, la mise en évidence d'un effet de glissement de la tendance centrale pour l'un des facteurs et d'un effet de modification de la dispersion avec maintien de la tendance centrale pour l'autre.

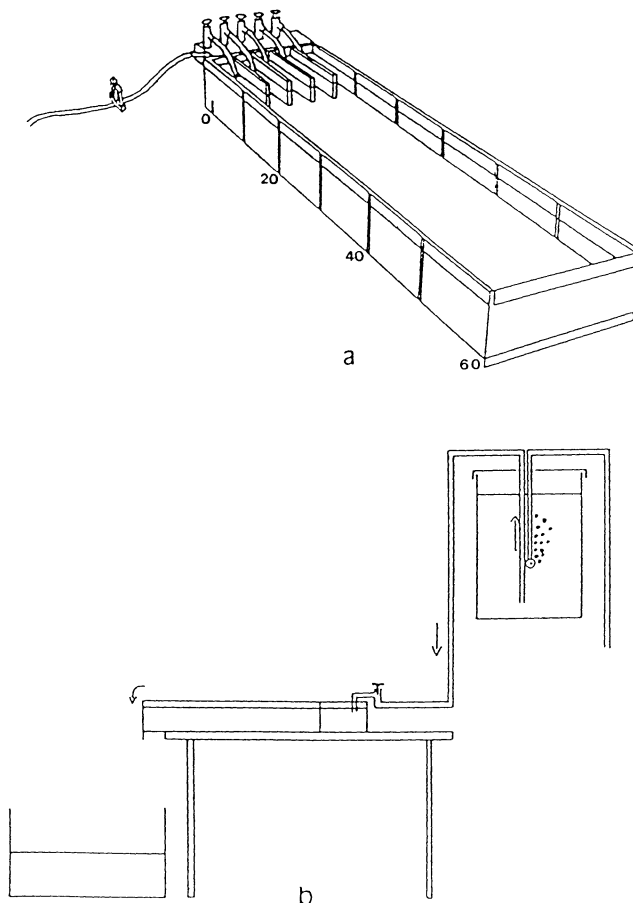
## 2. Présentation des Données

Les copépodes parasites (crustacés) ont dans la majorité des cas une larve nageante, active, appelée copépodite qui doit impérativement trouver un animal-hôte (poisson par exemple) pour se fixer et se développer en un adulte. Il existe des facteurs susceptibles de favoriser la rencontre, la chémotaxie et la rhéotaxie semblant être parmi les plus importants (Gotto, 1959, 1962; Carton, 1968a, 1968b; Briggs, 1974; Ache, 1982; Raibaut, 1985).

Les données traitées concernent l'influence du mucus sur le déplacement des copépodites infestants de *Caligus minimus* (Otto) parasite buccal du loup de mer ou bar *Dicentrarchus labrax* (Lime).

Le dispositif expérimental est composé d'une cuve en polychlorure de vinyl transparent de 62 cm de long, 13,5 cm de large et 5,5 cm de profondeur (Figure 1). Sous le fond est installée une feuille de papier millimétré, graduée de cm en cm. Les douze premiers centimètres de la cuve sont divisés en cinq compartiments égaux par quatre cloisons. Ces dernières ont pour but de régulariser le flux d'eau et de le rendre le plus laminaire possible.

L'alimentation en eau de mer se fait simplement par gravité grâce à un réservoir surélevé dans lequel l'eau est oxygénée en permanence par un bulleur à air comprimé. L'arrivée de l'eau dans la cuve est réglée grâce à cinq robinets disposés chacun dans un des compartiments. A l'autre extrémité, un simple trop plein évacue l'eau. Les cinq arrivées qui plongent de 2 cm dans l'eau ainsi que le trop



**Figure 1.** Schéma du dispositif expérimental; la figure a représente la cuve; la figure b l'ensemble du montage assurant l'alimentation en eau de mer.

plein sont munis d'un filet à plancton de 100  $\mu\text{m}$  qui évite les pertes d'individus au cours des expériences.

Les expériences sont réalisées à température constante (18°) et à un courant correspondant à un débit du trop plein compris entre 400 et 800 ml/heure. En effet, Fraile (1989) a montré que ces conditions de courant n'ont aucune influence sur le déplacement des copépodites. En outre, étant donné la phototaxie positive de ces larves, les expériences se sont déroulées dans l'obscurité. Les copépodites, âgées de 12 heures environ, sont placés dans la cuve les uns après les autres. La nature et la quantité du *mucus* placé dans la cuve pour attirer les copépodites ainsi que la *durée* de l'expérience c'est-à-dire le temps laissé aux larves pour se déplacer sont contrôlées et déterminent les conditions des expériences.

Pratiquement le copépodite est placé au cm 35 à l'instant initial. Le facteur chimique est placé simultanément dans le compartiment central au cm 5. Au bout de la durée *T* choisie la lumière est allumée. Le copépodite est recherché et visé sur le papier millimétré. (La lumière pouvant créer des effets parasites sur les déplacements, toute visée demandant plus de 20 secondes est éliminée).

La mesure faite, le copépodite est éliminé et un autre le remplace au cm 35.

**Tableau 1**

*Données observées: Chaque ligne donne la répartition des 30 copépodites dans chacune des 12 classes d'observations. La treizième colonne identifie l'expérience: les durées sont codées de 1 à 6; les mucus de A à G*

0	0	0	0	2	4	7	6	6	3	2	0	1A
0	0	0	0	3	6	7	7	5	2	0	0	1B
0	0	0	3	1	4	8	7	7	0	0	0	1C
0	0	0	2	2	4	9	8	5	0	0	0	1D
0	0	0	0	2	7	9	7	4	1	0	0	1E
0	0	0	1	2	5	9	6	4	3	0	0	1F
0	0	0	2	3	5	8	7	3	2	0	0	1G
0	0	0	0	2	4	5	4	7	5	3	0	2A
0	0	0	2	1	3	4	7	5	3	4	1	2B
0	0	0	3	3	3	6	4	9	2	0	0	2C
0	0	0	4	3	4	4	5	8	2	0	0	2D
0	0	1	1	2	6	5	6	5	4	0	0	2E
0	0	1	1	3	6	8	4	5	2	0	0	2F
0	1	1	2	5	5	7	5	3	1	0	0	2G
0	0	0	0	1	3	6	4	7	6	2	1	3A
0	0	0	1	3	5	4	6	4	2	3	2	3B
1	1	1	4	4	5	4	5	2	2	1	0	3C
0	3	2	3	4	4	4	4	4	2	0	0	3D
0	0	1	1	1	5	8	6	4	3	1	0	3E
0	0	1	1	5	8	6	6	2	1	0	0	3F
2	1	2	4	5	6	5	3	2	0	0	0	3G
0	0	0	0	1	1	3	5	7	6	4	3	4A
0	0	0	0	3	3	4	7	5	4	2	2	4B
3	0	2	1	3	7	8	4	0	0	2	0	4C
2	0	2	0	5	1	2	5	8	3	1	1	4D
0	1	2	1	1	3	6	6	4	4	2	0	4E
0	1	2	1	5	7	6	4	2	1	1	0	4F
0	2	3	5	5	3	5	3	3	1	0	0	4G
0	0	0	0	2	3	3	4	6	4	3	5	5A
0	0	0	0	1	3	4	6	4	5	3	4	5B
0	1	2	6	2	5	4	4	3	1	2	0	5C
3	3	3	2	2	5	4	4	3	0	0	1	5D
1	2	0	2	1	5	3	5	4	2	4	1	5E
0	1	1	2	4	5	7	6	3	1	0	0	5F
0	2	2	3	4	6	6	4	1	1	1	0	5G
0	0	0	1	2	3	2	6	6	4	3	3	6A
0	0	1	1	2	2	3	5	7	4	2	3	6B
0	0	0	2	3	6	5	6	1	2	3	2	6C
3	1	1	6	2	4	3	5	2	0	1	2	6D
3	4	2	2	4	1	2	3	2	0	4	3	6E
0	2	1	1	5	6	6	4	3	1	1	0	6F
0	1	2	2	3	5	5	4	3	2	2	1	6G

Six durées ont été utilisées régulièrement échelonnées de 10 mn à 60 mn. Elles sont notées de 1 à 6.

Six mucus différents sont expérimentés: les écailles de muge, les écailles de loup, le mucus de loup en pastille, et le mucus de loup lyophilisé pastille à 30, 40 ou 50 mg. Une expérience est réalisée en l'absence de tout mucus. Elle est identifiée par la lettre A. Les six autres sont repérées par les lettres B à G.

Pour chacune des associations durée  $\times$  mucus, trente copépodites sont lâchés successivement dans la cuve. Pour chacune de ces associations, les résultats sont conservés sous la forme d'un histogramme en 12 classes d'amplitude 5 cm décrivant la répartition des positions atteintes par les copépodites.

En définitive, les données se présentent sous la forme de quarante deux histogrammes correspondant chacun à la répartition de 30 individus en 12 classes. Chaque histogramme est associé à une information de durée et une information de mucus précisant les conditions de son obtention.

Le Tableau 1 fournit les données. La treizième colonne identifie l'expérience en indiquant la durée et le type de mucus utilisé; par exemple 1A correspond à une durée de 10 minutes en l'absence de tout mucus.

### 3. L'Analyse Classique

La nature du recueil des données conduit naturellement à la mise en oeuvre d'une analyse des correspondances, méthode particulièrement adaptée à la comparaison des histogrammes. Les variables durée et mucus sont prises dans une telle démarche comme des variables qualitatives supplémentaires et pourront être représentées par exemple par le point moyen de chacune de leurs modalités.

Les résultats sont présentés dans le Tableau 2 et les Figures 2 et 3. La lecture des axes principaux, coordonnées des points représentant les colonnes, laisse prévoir un effet Guttman: les valeurs du premier axe suivent une fonction linéaire décroissante du numéro des colonnes tandis que les valeurs du second axe évoluent comme un polynôme d'ordre deux. Les figures confirment cet effet qui traduit que certains histogrammes chargent plus les classes terminales de la cuve tandis que d'autres chargent plus les classes proches du compartiment contenant le mucus.

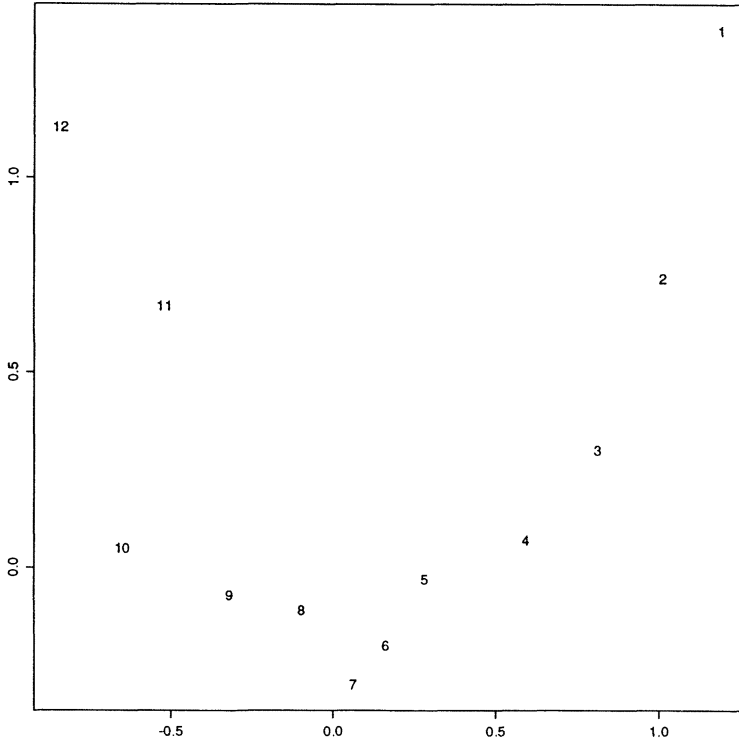
La projection des points moyens des classes (Figures 4 et 5) permet de mieux voir le rôle joué par la durée et le facteur chimique dans cette répartition. On constate que les modalités de ces deux variables ne s'organisent pas de façon comparable.

Les durées s'organisent assez simplement le long du second axe signifiant que l'allongement de l'expérience ne se traduit pas par un déplacement massif des copépodites vers l'une ou l'autre des extrémités de la cuve mais au contraire par un étalement progressif de l'histogramme. Ainsi à mesure que s'allonge la durée d'exposition, les copépodites s'éloignent de leur position initiale de façon équilibrée soit vers le compartiment contenant le mucus, soit vers le trop plein (Figure 6).

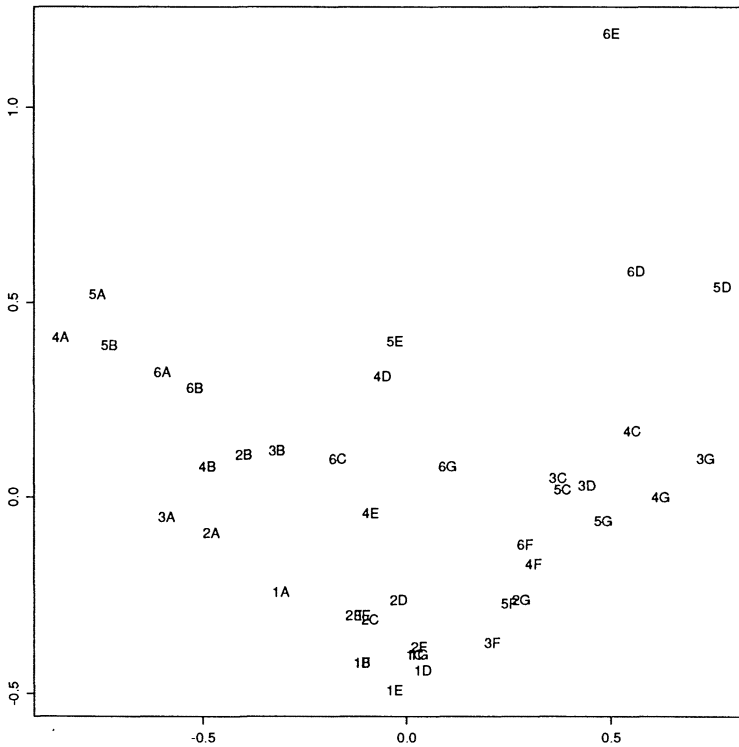
Les classes de mucus se rangent le long de l'axe 1 et montrent une opposition nette des modalités "pas de mucus" et "Ecaille de Muge" aux autres. On peut donc interpréter cet axe en terme d'attrait du mucus pour les copépodites. Une modalité attrayante se traduira par une surcharge des classes proches du compartiment contenant le mucus donc par des valeurs positives sur l'axe 1. Au contraire, en présence d'une modalité peu attrayante, les copépodites ne chercheront pas à remonter le courant. Ceci se traduira par une surcharge des classes proches du trop plein et donc par une valeur négative sur l'axe 1 (Figure 7).

**Tableau 2**  
Résultats de l'analyse des correspondances des données initiales

Premières valeurs propres												
	Valeur propre		Pourcentage		Pourcentage cumulé							
	0,17168		36,43		36,43							
	0,12291		26,08		62,51							
	0,04160		8,83		71,33							
	0,03599		7,64		78,97							
	0,03164		6,71		85,68							
Deux premiers axes principaux												
Colonnes	1	2	3	4	5	6	7	8	9	10	11	12
Axe 1	1,19	1,01	0,81	0,59	0,28	0,16	0,06	-0,10	-0,32	-0,65	-0,52	-0,84
Axe 2	1,38	0,74	0,30	0,07	-0,03	-0,20	-0,30	-0,11	-0,07	0,05	0,67	1,13



**Figure 2.** Représentation des colonnes dans l'analyse des correspondances.



**Figure 3.** Représentation des lignes dans l'analyse des correspondances.

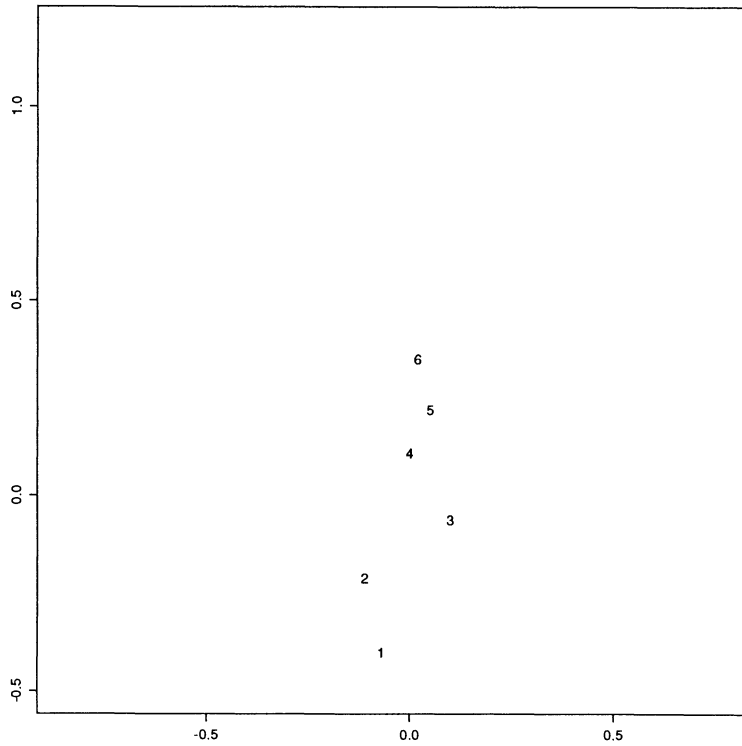


Figure 4. Représentation des moyennes des durées.

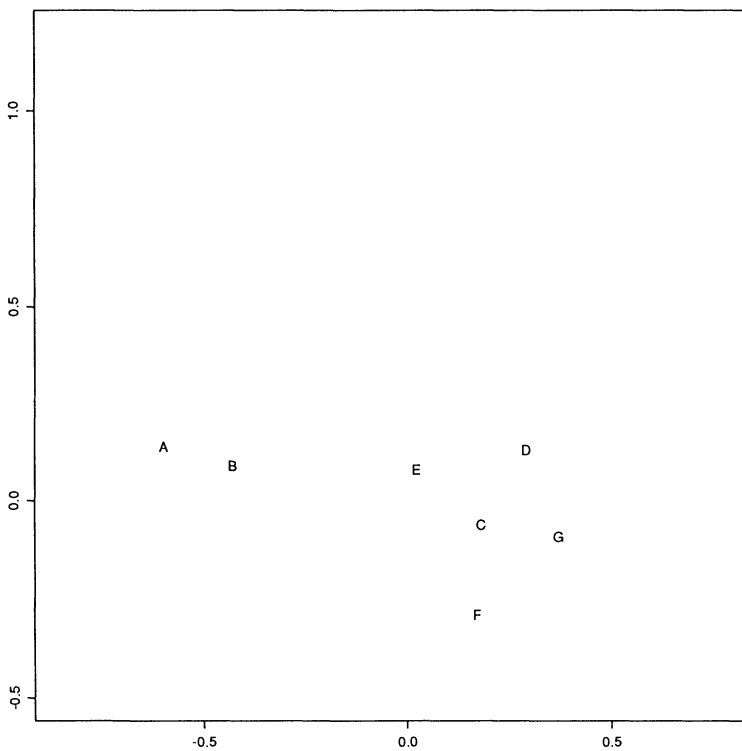
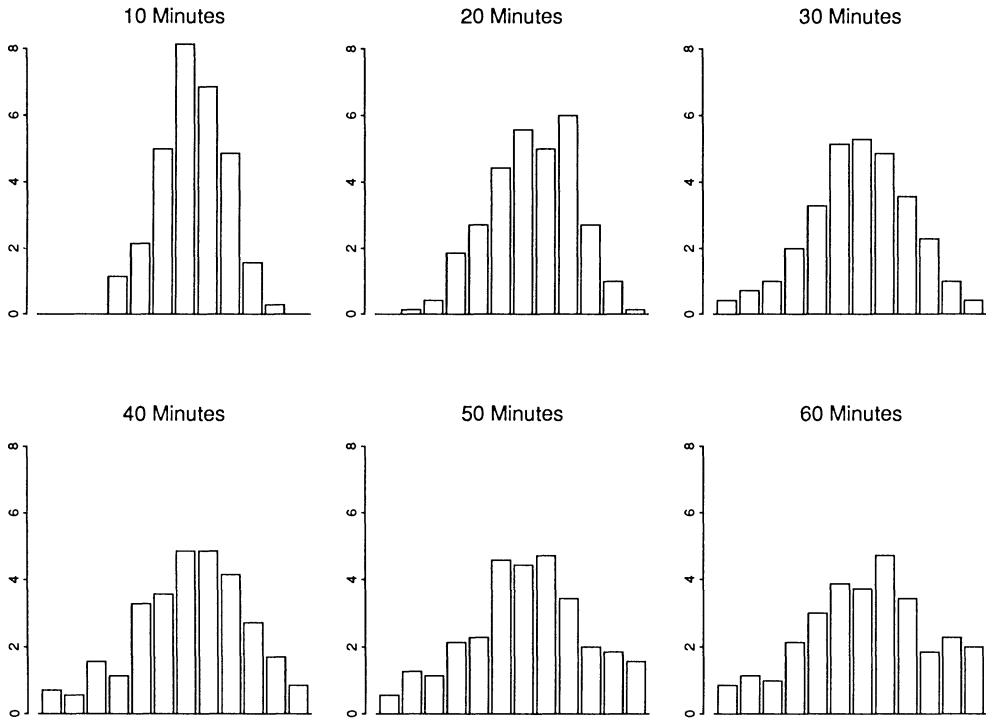
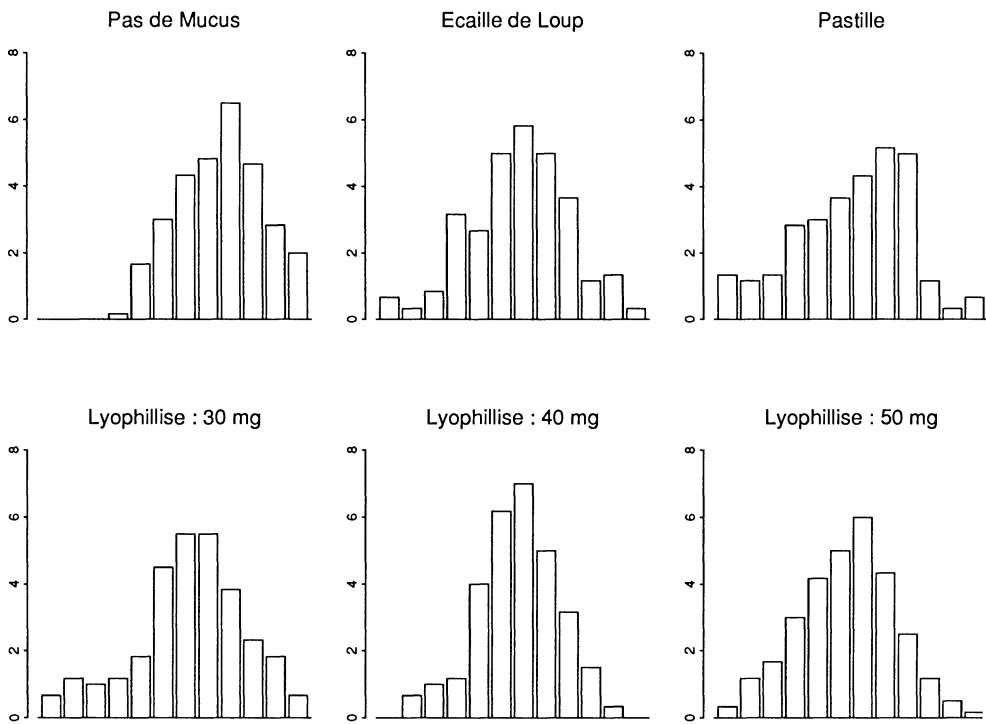


Figure 5. Représentation des moyennes des mucus.



**Figure 6.** Histogramme moyen par durée.



**Figure 7.** Histogramme moyen par mucus (pour des raisons de place la modalité Ecaille de muge est omise).

4. Quantification des Effets

4.1. Appelons  $N$  le tableau  $(p = 42) \times (q = 12)$  des données observées.  $N_i^j$  est l'élément de la ligne  $i$  et de la colonne  $j$  de ce tableau. Les choix faits pour mener l'expérimentation font que pour tout  $i$ :  $\sum_{j=1}^q N_i^j = 30$  et  $\sum_{i=1}^p \sum_{j=1}^q N_i^j = 42 \times 30 = 1260$ .

Appelons alors  $F$  le tableau déduit de  $N$  en divisant chacun de ses éléments par 1260. Notons  $D_I$  la matrice diagonale  $p \times p$  dont les éléments diagonaux sont les  $f_i$  totaux des lignes des  $F$  et  $D_J$  la matrice diagonale  $q \times q$  dont les éléments diagonaux sont les  $f_j$  totaux des colonnes de  $F$ .

Si  $H$  est la matrice identité de  $R^p$ , on sait que pour ce qui concerne la représentation des lignes de  $F$ , l'analyse des correspondances est équivalente (Van Der Heijden, De Falguerolles, et De Leeuw, 1989) à l'analyse en composantes principales du triplet  $(\tilde{F} = (H - \mathbf{1}_p \mathbf{1}_p' D_I)F, D_J^{-1}, D_I)$ .

L'inertie totale de cette analyse est égale à

$$\text{Tr}(\tilde{F} D_J^{-1} \tilde{F}' D_I) = \text{Tr}(\tilde{F}' D_I \tilde{F} D_J^{-1}) = \sum_{j=1}^q \frac{\tilde{F}'^j D_I \tilde{F}^j}{f_j} \tag{1}$$

Le terme  $\tilde{F}'^j D_I \tilde{F}^j / f_j$  mesure la part prise dans l'inertie totale par la colonne  $j$  du tableau initial. La première colonne du Tableau 3 montre que ce sont les classes extrêmes qui ont les plus grandes participations. On sait que ceci provient de la propriété connue de l'analyse des correspondances de valoriser les différences entre lignes dans les colonnes aux effectifs marginaux faibles (présence des  $f_j$  en dénominateur). Dans cet exemple, cette particularité n'apparaît pas dramatique au sens où on ne voit pas une ou deux colonnes expliquer à elles seules la plus grande partie de l'inertie.

4.2. Appelons  $U_T$  le tableau disjonctif complet  $(p = 42) \times (t = 6)$  associé aux classes de durée et  $P_T = U_T(U_T' D_I U_T)^{-1} U_T' D_I$  le projecteur  $D_I$ -orthogonal sur le sous espace de  $R^p$  engendré par les  $t$  colonnes de ce tableau.  $P_T \tilde{F}$  est un tableau  $p \times q$  dans lequel une ligne est égale à la moyenne des lignes  $\tilde{F}_i$  associées à une même modalité de  $T$ . Introduisons  $\tilde{F} - P_T \tilde{F} = (H - P_T) \tilde{F}$  le tableau des écarts à ces moyennes. Un calcul simple établit que

$$\text{Tr}(\tilde{F}' D_I \tilde{F} D_J^{-1}) = \text{Tr}((P_T \tilde{F})' D_I (P_T \tilde{F}) D_J^{-1}) + \text{Tr}(((H - P_T) \tilde{F})' D_I ((H - P_T) \tilde{F}) D_J^{-1}) \tag{2}$$

et encore que

$$\frac{\tilde{F}'^j D_I \tilde{F}^j}{f_j} = \frac{(P_T \tilde{F})'^j D_I (P_T \tilde{F})^j}{f_j} + \frac{((H - P_T) \tilde{F})'^j D_I ((H - P_T) \tilde{F})^j}{f_j} \tag{3}$$

De l'équation (2), il vient que l'inertie totale de l'étude peut être décomposée en une part  $I_1$  expliquée par le facteur durée et une part résiduelle  $I_2$ :

$$I_1 = \text{Tr}((P_T \tilde{F})' D_I (P_T \tilde{F}) D_J^{-1}),$$

$$I_2 = \text{Tr}(((H - P_T) \tilde{F})' D_I ((H - P_T) \tilde{F}) D_J^{-1}).$$

**Tableau 3**  
Participations à l'inertie

Colonnes	Total		Durée		Mucus		Résidus/ Durée + Mucus
1	63,81	=	7,62	+	14,92	+	41,27
2	46,27	=	10,40	+	11,64	+	24,23
3	30,95	=	9,05	+	10,67	+	11,23
4	43,61	=	3,22	+	22,45	+	17,94
5	19,25	=	2,18	+	9,21	+	7,86
6	18,43	=	2,16	+	6,66	+	9,61
7	22,41	=	10,89	+	5,22	+	6,30
8	9,38	=	3,37	+	1,94	+	4,07
9	30,95	=	6,20	+	11,06	+	13,69
10	37,33	=	2,48	+	21,68	+	13,17
11	41,92	=	9,74	+	19,55	+	12,63
12	59,86	=	19,45	+	21,43	+	18,98
Total	424,16	=	86,76	+	156,42	+	180,98
Test			Significatif		Significatif		

Tous les nombres sont multipliés par  $30 \times 30 = 900$ .



De l'équation (3), il vient que cette décomposition peut se faire colonne par colonne. La seconde colonne du Tableau 3 rassemble ces participations. On voit que le temps explique une variabilité de 86,76/900 soit 20,45% de la variabilité totale égale à 424,16/900.

Les colonnes 2, 3, 7, 8, 9, 11, 12 du tableau des données correspondent à des rapports entre la colonne 2 et la colonne 1 du Tableau 3 égaux ou supérieurs à 20%. On pourra dire que ces colonnes sont fortement sensibles au facteur durée.

Afin de donner une signification à la part de la variabilité expliquée par la durée de l'expérience, un test de permutation peut être réalisé qui consiste à affecter aléatoirement chacune des lignes du tableau des données observées aux modalités du facteur durée (en respectant les effectifs de ces modalités) et à calculer la trace correspondante. En 100 permutations successives la valeur trouvée 86,76 n'a été dépassée que 4 fois, ce qui permet de dire qu'au seuil de 5%, le facteur durée a un effet significatif sur l'expérience.

La Figure 6 qui montre les histogrammes moyens pour chacune des durées suggère une évolution régulière des histogrammes en fonction du temps ayant pour effet de vider les classes centrales au bénéfice des classes extrêmes. Ceci suggère de substituer au tableau disjonctif  $U_T$  la variable quantitative  $T$  ayant pour valeur la durée de l'expérience. Soit  $\tilde{T} = (H - \mathbf{1}_p \mathbf{1}_p' D_T)T$  son expression centrée et  $\Pi_T = \tilde{T}(\tilde{T}' D_T \tilde{T})^{-1} \tilde{T} D_T$  le projecteur  $D_T$ -orthogonal sur le sous-espace de  $R^p$  engendré par  $\tilde{T}$ . On peut faire pour  $\Pi_{\tilde{T}}$  ce qui a été fait pour  $P_T$ .

La part de variation globale absorbée par  $\Pi_{\tilde{T}} \tilde{F}$  est égale à 64,75. Remarquons qu'il est normal qu'elle soit inférieure à celle absorbée par  $P_T \tilde{F}$  puisque dans la seconde chaque histogramme est remplacé par la moyenne des histogrammes de la classe de temps auquel il appartient alors que dans la première, c'est une approximation linéaire qui est utilisée.

Le test de permutation peut être effectué sur  $T$  comme il l'a été sur  $U_T$ . La valeur 64,75 n'est jamais dépassée en 100 permutations ce qui permet de conserver l'hypothèse d'une évolution linéaire des histogrammes.

**4.3.** Appelons de façon analogue  $U_M$  le tableau disjonctif complet ( $p = 42$ )  $\times$  ( $m = 7$ ) associé aux classes de mucus et  $P_M$  le projecteur associé.

La troisième colonne du Tableau 3 donne les quantités permettant d'apprécier l'effet du mucus sur les expériences. On ne peut pas comparer directement la part de variabilité expliquée par le mucus (156,42) qui a 7 modalités à celle expliquée par la durée qui n'en a que 6. 100 permutations successives des classes de mucus ne donnent jamais de variabilité expliquée supérieure à 156,42. On peut donc conclure à un effet très significatif du mucus.

La Figure 5 suggère une dichotomie des classes du mucus  $\{A, B\}$  et  $\{C, D, E, F, G\}$ . Appelons  $U_d$ , le tableau disjonctif complet ( $p = 42$ )  $\times$  ( $r = 2$ ) associé à cette dichotomie et  $P_d$  le projecteur associé. On peut faire pour  $P_d$  ce qui a été fait pour  $P_M$ . On trouve une variabilité expliquée de 99,76 valeur qui n'est jamais atteinte en 100 permutations. On peut donc conclure à une différence significative entre  $\{A, B\}$  et  $\{C, D, E, F, G\}$ . "A" correspondant à l'absence de mucus, on voit que la modalité "B", écaille de muge est peu attractive sur les copépodites.

**4.4.** Appelons  $v_t$  le nombre d'expériences réalisées dans la modalité  $t$  du facteur durée et  $v_m$  le nombre d'expériences réalisées dans la modalité  $m$  du facteur mucus. Soit  $v_{tm}$  le nombre d'expériences réalisées pour la conjonction des modalités  $t$  et  $m$ . Notons par  $v$  le nombre total d'expériences réalisées. Ici  $v = 42$ ,  $v_t = 7$ ,  $v_m = 6$ , et  $v_{tm} = 1$  si bien qu'on a l'égalité:

$$v_{tm} = \frac{v_t v_m}{v} \tag{4}$$

Lorsqu'une telle condition est réalisée pour tous les couples  $(t, m)$ , on dit que les facteurs sont orthogonaux, ce qui assure une décomposition additive des parts de variabilité.

Autrement dit si on note  $U_{(T+M)}$  la juxtaposition  $(U_T; U_M)$  des tableaux  $U_T$  et  $U_M$  et  $P_{T+M}$  le projecteur sur  $U_{(T+M)}$  alors  $P_{T+M} = P_T + P_M$  si bien que  $P_{T+M} \tilde{F} = P_T \tilde{F} + P_M \tilde{F}$ .

Lorsque la condition (4) n'est pas remplie  $P_{T+M} \neq P_T + P_M$ ; deux décompositions symétriques de  $P_{T+M}$  sont envisageables:

$$P_{T+M} = P_T + P_{M \cap T^\perp} \tag{5}$$

et

$$P_{T+M} = P_M + P_{T \cap M^\perp} \tag{6}$$

Pratiquement la décomposition (5) revient à utiliser  $P_T \tilde{F}$  comme nous l'avons fait puis  $P_M((H - P_T) \tilde{F})$  c'est-à-dire à chercher l'effet du second facteur  $M$  sur le résidu par rapport au premier.

4.5. Notons  $U_{T\oplus M}$  le tableau disjonctif complet associé à la variable qualitative obtenue en croisant chacune des modalités de durée avec chacune des modalités de mucus.  $U_{T\oplus M}$  a  $(t \times m)$  colonnes, soit dans notre exemple 42 colonnes, ce qui traduit le cas particulier de cette expérimentation où une seule expérience (observation de 30 copépodites) a été faite pour une durée et un mucus donnés.

Il en découle que  $P_{T\oplus M} = H$  et que la projection de  $(H - P_T + P_M)\tilde{F}$  sur le sous espace engendré par  $U_{T\oplus M}$  correspond donc à  $(H - P_T + P_M)\tilde{F}$ .

Si les données ne présentaient pas cette particularité, l'étude de  $P_{T\oplus M}(H - P_T + P_M)\tilde{F}$  aurait permis d'apprécier les effets d'interaction entre les modalités de durée et de mucus réduisant d'autant la part non expliquée de variabilité.

Ici, la seule étude possible est l'analyse en composantes principales du triplet  $((H - P_T + P_M)\tilde{F}, D\bar{J}^1, D_I)$ . Le Tableau 4 en donne les valeurs propres et les axes principaux. La Figure 8 donne la représentation dans des individus dans le plan 1-2.

On voit que les individus 6E, 4C, 5D se détachent nettement vers la gauche. Les coefficients de

**Tableau 4**  
Résultats de l'analyse en composantes principales des résidus par rapport à la durée et au mucus

Premières valeurs propres		
Valeur propre	Pourcentage	Pourcentage cumulé
0,06447	32,06	32,06
0,03462	17,22	49,28
0,02751	13,68	62,96
0,01979	9,84	72,80
0,01685	8,38	81,18

Deux premiers axes principaux												
Colonnes	1	2	3	4	5	6	7	8	9	10	11	12
Axe 1	-0,67	-0,44	-0,30	-0,15	-0,29	-0,15	0,17	0,46	0,74	0,70	-0,13	0,06
Axe 2	0,31	-0,42	-0,25	-0,51	-0,13	0,09	0,08	0,21	-0,14	0,12	0,16	0,48

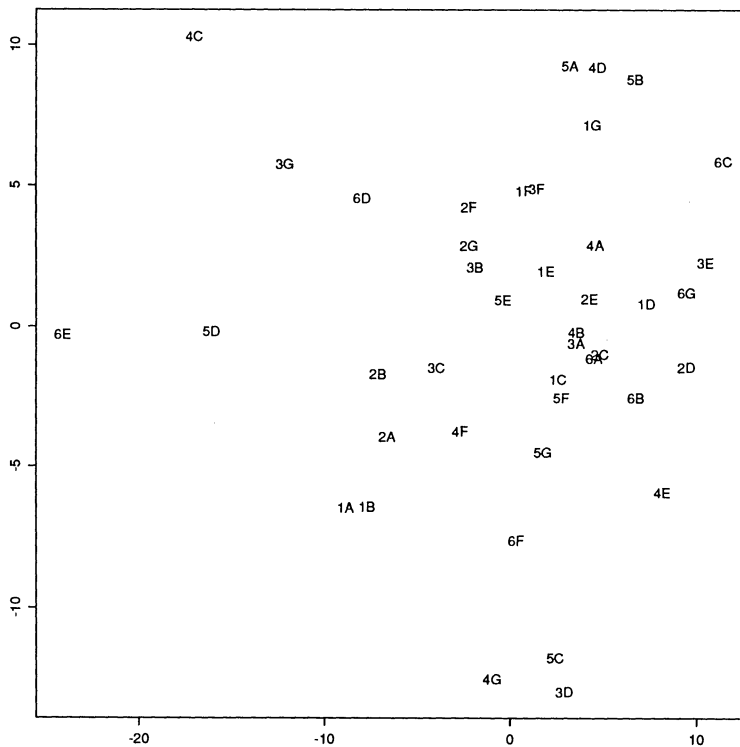


Figure 8. Représentation des lignes dans l'analyse en composantes principales des résidus par rapport à la durée et au mucus.

l'axe principal 1 permettent d'en conclure que ces individus ont des scores plus élevés sur les cases 1 et 2 et moins élevés sur les cases 09 et 10 que l'addition des effets mucus et durée ne permet de l'attendre. Ce sont donc des associations durée + mucus pour lesquelles les copépodites se sont plus déplacés que la simple addition des effets mucus et durée ne pouvait le faire espérer.

## 5. Conclusion

La démarche suivie est aisément généralisable à des plans expérimentaux plus complexes. Sa mise en oeuvre réside sur la possibilité de créer les tableaux projetés et résiduels tels que  $P_T\tilde{F}$  et  $(H - P_T)\tilde{F}$  et de les utiliser dans une analyse en composantes principales généralisée. Nous avons utilisé ici le logiciel LEAS (Delcamp, 1987) et son extension récente adaptée à la création du tableau projeté et du tableau résiduel. Ces fonctions sont également disponibles dans le logiciel Bioméco (Lebreton et al., 1986).

La disposition de telles fonctions ouvrent la possibilité de créer des ponts entre les démarches exploratoires et confirmatoires considérées souvent comme antinomiques.

Les résultats obtenus ici montrent que cette opposition peut être dépassée: ce sont les inerties de l'analyse des correspondances qui servent de support à des tests de permutation; et les effets jugés significatifs concernent non seulement un déplacement de tendance centrale mais aussi un étalement de distribution.

## SUMMARY

A study on the effect of mucus on the movement of infective copepodites of *Caligus minimus*, buccal parasite of sea bass, *Dicentrarchus labrax*, leads to a data array formed by 42 bar charts describing the distributions of these larvae depending on length of the experiment and the type and quantity of mucus.

A standard application of correspondence analysis to this data array, followed by the usual technique of supplementary points for the means of the experiments with the same length or the same mucus suggests an effect of length and mucus on the behavior of the copepodites: when the length of the experiment increases, the spread of the bar charts increases too; the central tendency of the bar charts moves with the attracting strength of the mucus.

In order to proceed from the qualitative result given by correspondence analysis to a quantification of the effects, we develop the following hypothesis testing approach. First, the total dispersion measure, called inertia  $I$ , of the study is divided into two parts,  $I_1$  explained for example by the mucus classes, and  $I_2 = I - I_1$ , the residual part. To do that, consider  $F$ , the table of available frequencies.  $D_i$  and  $D_j$  are respectively the diagonal matrices constructed with the row and column margins of  $F$ .  $\tilde{F}$  is the centered table obtained from  $F$  using the weights given by the diagonal of  $D_i$ .

The total inertia of the study is:  $I = \text{Tr}(\tilde{F}'D_i\tilde{F}D_j^{-1})$ . Let  $U_T$  be the table of the indicator variables of the mucus classes and  $P_T = U_T(U_T'D_iU_T)^{-1}U_T'D_i$  the  $D_i$ -orthogonal projector associated with it.  $I_1 = \text{Tr}((P_T\tilde{F})'D_i(P_T\tilde{F})D_j^{-1})$  and  $I_2$  is the corresponding quantity for  $\tilde{F} - P_T\tilde{F}$ .

Under the null hypothesis of no effect of mucus on the behavior of copepodites,  $I_1$  can be compared with similar quantities obtained after random partitions of the copepodites with the same number of classes and the same number of elements in the classes. The generation of such random partitions gives the distribution curve of  $I_1$  under the null hypothesis. An extremal position for the observed  $I_1$  is interpreted as grounds for rejection of the null hypothesis. For the data array studied, the test rejects the null hypothesis of no effect for both mucus and length of the experiments.

The approach is used in this paper in a two-factor context. The paper extrapolates to how this approach can be extended to more complicated designs.

## BIBLIOGRAPHIE

- Ache, B. W. (1982). Chemoreception and thermoreception. In *The Biology of Crustacea*, Vol. 3, D. E. Bliss (ed.), 369–398. New York: Academic Press.
- Benzécri, J. P. (1973). *L'Analyse des Données—Tome 2: L'Analyse des Correspondances*. Paris: Dunod.
- Briggs, R. P. (1974). Aspects of the biology and structure of the associated copepod *Paranthesius anemoniae* Claus. Ph.D. thesis, Queen's University of Belfast, Vol. I, 1–183, Vol. II, 1–114.
- Carton, Y. (1968a). Spécificité parasitaire de *Sabelliphilus sarsi*, parasite de *Spirographis spallanzani*. III: Mise en évidence d'une attraction biochimique du Copépode par l'Annélide. *Archives de Zoologie Expérimentale et Générale* **109**, 123–144.
- Carton, Y. (1968b) Spécificité parasitaire de *Sabelliphilus sarsi*, parasite de *Spirographis spallanzani*. IV: Etude biochimique des hôtes et de leurs sécrétions. *Archives de Zoologie Expérimentale et Générale* **109**, 269–286.
- Cazes, P., Chessel, D., et Doledec, S. (1988). L'analyse des correspondances internes d'un tableau partitionné: son usage en hydrobiologie. *Revue de Statistique Appliquée*. **36**, 39–54.

- Delcamp, A. (1987). *L.E.A.S. Manuel d'utilisation*. Unité de Biométrie, 9 Place Viala, 34060 Montpellier Cedex (France).
- Fraile, L. (1989). Recherche sur les taxies des copépodes parasites de Poissons. Thèse de Doctorat, Université de Montpellier II.
- Gotto, R. V. (1959). An ascidian "hotel." *Irish Naturalists' Journal* **13**, 99.
- Gotto, R. V. (1962). Egg-number and ecology in commensal and parasitic Copepods. *Annals and Magazine of Natural History* **13**, 97–107.
- Lebart, L. et Fenelon, J. P. (1975). *Statistique et Informatique Appliquées*, 3e édition. Paris: Dunod.
- Lebart, L., Morineau, A., et Fenelon, J. P. (1979). *Traitement des Données Statistiques: Méthodes et Programmes*. Paris: Dunod.
- Lebreton, J. D., Chessel, D., Prodon, R., et Yoccoz, N. (1988). L'analyse des relations espèces milieu par l'analyse canonique des correspondances—1: Variables de milieu quantitatives. *Acta Oecologica—Oecologica Generalis* **9**, 53–67.
- Lebreton, J. D., Roux, M., Bacou, A. M., et Banco, G. (1986). *Biomeco, notice d'utilisation*. Documentation polycopiée CEPE/CNRS, BP 5051, 34033 Montpellier.
- Lebreton, J. D. and Yoccoz, N. (1987). Multivariate analysis of bird count data. *Acta Oecologica—Oecologica Generalis* **8**, 125–144.
- Nishisato, S. and Lawrence, D. R. (1988). Dual scaling of multiway matrices: Several variants. In *Multiway Data Analysis*, R. Coppi and S. Bolasco (eds), 317–326. Amsterdam: North Holland.
- Pontier, J., Jolicoeur, P., et Pernin, M. O. (1987). Analyse canonique complète. *Stat. et Anal. des Données* **12**, 124–148.
- Raubaut, A. (1985). Les cycles évolutifs des copépodes parasites et les modalités de l'infestation. *Ann. Biol.* **XXIV**, Fax 3, p. 16.
- Rao, C. R. (1964). The use and the interpretation of principal component analysis in applied research. *Sankhyā, Série A* **26**, 329–359.
- Robert, P. and Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The  $R_c$ -coefficient. *Applied Statistics* **15**, 257–265.
- Rouanet, H. et Lepine, D. (1976). Structures linéaires et analyse des comparaisons. *Math. Sci. Hum.* **56**, 5–46.
- Sabatier, R. (1987). Méthodes factorielles en analyse des données: approximation et prise en compte de variables concomitantes. Thèse d'Etat, Université de Montpellier II.
- Sabatier, R., Lebreton, J. D. and Chessel, D. (1989). Principal component analysis with instrumental variables as a tool for modelling composition data. In *Multiway Data Analysis*, R. Coppi and S. Bolasco (eds), 341–352. Amsterdam: North Holland.
- Ter Braak, C. J. F. (1986). The analysis of vegetation–environment relationships by canonical correspondence analysis. *Vegetatio* **69**, 69–77.
- Ter Braak, C. J. F. (1987). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**, 1167–1179.
- Tukey, J. W. and Mallows, C. (1982). An overview of techniques of data analysis emphasizing its exploratory aspects. In *Some Recent Advances in Statistics*, B. Epstein (ed.). London: Academic Press.
- Van Der Heijden, P. G. M., De Falguerolles, A., and De Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. *Applied Statistics* **38**, 249–292.
- Yoccoz, N. et Chessel, D. (1988). Ordination sous contraintes de relevés d'avifaune: éliminations d'effet dans un plan d'observations à deux facteurs. *CR Acad. Sci., Paris, Série III* **307**, 189–194.

Received November 1989; revised April 1991 and April 1992; accepted April 1992.