Modèles loglinéaires pour l'étude des associations entre captures par unité d'effort et variables hydroclimatiques

Jacques Badia [(1)] et Patrick Prouzet [(2)]

[(1)] *INRA, Unité de Biométrie et d'Intelligence artificielle, B.P. 27, 31326 Castanet Tolosan Cedex, France.*

[(2)] *IFREMER/INRA, Station d'Hydrobiologie de Saint-Pée-sur-Nivelle, B.P. 3, 64310 Ascain, France.*

Reçu le 14 septembre 1995; accepté le 7 mars 1996.

Badia J., P. Pouzet. *Aquat. Living Resour.*, 1996, **9**, 217-223.

Loglinear models for studying associations between catches per unit effort and hydroclimatic variables.

Abstract

In this paper we study associations between categorical variables. We present an approach based on loglinear modelling. A better knowledge of the influence of the intensity of tidal coefficient (cm) and of the rate of river flow (df) on catch per unit effort (cpue) of allis shad (*Alosa alosa*) was obtained. Analysis performed indicated that an increase of catch per unit effort is proportionally associated with the increase in the tidal coefficient. We also noted the influence of the rate of flow.

Keywords: Loglinear model, catch per unit effort, environmental variables, *Alosa alosa*, allis shad.

Résumé

Dans cet article, nous étudions les associations entre variables qualitatives. Nous présentons une approche basée sur l'utilisation du modèle loglinéaire. Appliqués aux associations entre captures par unité d'effort (cpue) d'aloses (*Alosa alosa*), coefficient de marée (cm) et débit fluvial (df), différents modèles mettent en évidence que les captures par unité d'effort augmentent proportionnellement à la valeur du coefficient de marée et qu'elles ont des évolutions conjointes moins structurées avec le débit fluvial.

Mots-clés : Modèle loglinéaire, capture par unité d'effort, variables environnementales, *Alosa alosa*, grande alose ou alose vraie.

INTRODUCTION

Le modèle linéaire généralisé est un outil efficace d'analyse des données qualitatives. En particulier, lorsque des variables catégorielles sont construites à partir de variables continues, les modèles loglinéaires permettent de définir des mesures d'associations. Dans ce qui suit, nous nous intéressons essentiellement à la méthodologie statistique exposée dans de nombreux ouvrages tels que Bishop *et al.* (1975), Fienberg (1980) et Agresti (1990), pour montrer la richesse de cette approche peu utilisée en hydrobiologie ou en halieutique. Nous présentons différentes modélisations décrivant les associations entre variables catégorielles, nous les illustrons par les données du tableau 1 qui se réfèrent aux résultats des pêches d'aloses (*Alosa alosa*) dans le fleuve Adour entre 1988 et 1991 (Prouzet *et al.*, 1994). Chaque cellule de ce tableau contient le nombre d'observations pour chaque combinaison des modalités des variables captures par unité d'effort (cpue), débit

Tableau 1. – Observations par combinaison des modalités des variables captures par unité d'effort, débit fluvial et coefficient de marée.
Observations for the combination of catch per unit effort, rate of flow and tidal coefficient variables.

Débit fluvial	Coefficient de marée	Captures par unité d'effort			
		Faible	Faible à Moyen	Moyen à Fort	Fort
Faible	Faible	11	5	2	1
	Faible à Moyen	4	11	4	2
	Moyen à Fort	7	5	9	4
	Fort	5	1	5	4
Faible à Moyen	Faible	4	4	7	5
	Faible à Moyen	8	6	4	6
	Moyen à Fort	4	3	6	8
	Fort	4	7	7	7
Moyen à Fort	Faible	5	4	5	9
	Faible à Moyen	3	8	4	6
	Moyen à Fort	2	3	5	6
	Fort	3	4	9	7
Fort	Faible	15	7	3	2
	Faible à Moyen	9	4	2	3
	Moyen à Fort	4	1	5	9
	Fort	7	6	2	10

fluvial (df) et coefficient de marée (cm). Dans une première partie, nous présentons brièvement le modèle loglinéaire standard pour des tableaux de contingences à deux et à trois dimensions. Dans une deuxième partie, nous analysons les tableaux marginaux cm et cpue (Tableau 2a) puis df et cpue (Tableau 2b) pour illustrer l'introduction dans le modèle standard de la relation d'ordre entre les modalités des variables. Nous étudions ensuite le tableau complet (Tableau 1) qui résume les informations sur les captures et sur les conditions hydroclimatiques des pêches.

MATÉRIELS ET MÉTHODES

Les observations

On trouvera dans Prouzet *et al.*, (1994), une description précise du protocole de recueil des données et des caractéristiques de la pêcherie. Très brièvement, les données analysées sont les captures moyennes journalières entre 1988 et 1991 relevées dans les carnets de pêche mis en place par le Comité Interprofessionnel des Poissons migrateurs et des Estuaires (CIPE) et deux paramètres physiques qui caractérisent le milieu aquatique : le coefficient de marée qui est un coefficient astronomique exprimé en centième dont la valeur peut, sur l'année, être comprise entre 20 pour les marées de morte-eau les plus faibles et 120 pour les marées de vive-eau d'équinoxe; le débit fluvial journalier exprimé en mètre cube par seconde, est mesuré dans des stations de jaugeage. Les découpages en classes ont été effectués selon les quartiles pour les cpue. Pour prendre en compte les variations annuelles de l'abondance, la répartition en classe de cpue a été faite par année. Pour les variables physiques cm et df, le codage a été effectué d'après les histogrammes des valeurs, toutes années confondues.

Les modèles

Pour le Tableau 2, nous adoptons les notations suivantes : i est l'indice de ligne, j l'indice de colonne, $I\,(=4)$ le nombre de lignes et $J\,(=4)$ le nombre de colonnes. Les $n\,(=342)$ observations de l'échantillon sont réparties dans les $N = IJ\,(=16)$ cellules du tableau. L'effectif de chaque cellule ij est n_{ij}, il y a n_{i+} observations dans la ligne i et n_{+j} dans la colonne j.

Dans le cas d'un échantillon multinomial, les probabilités $\{\pi_{ij}\}$ des cellules $\{ij\}$ représentent la distribution conjointe des deux catégories de réponses. Si les observations sont indépendantes ($\pi_{ij} = \pi_{i+}\pi_{+j}$), les effectifs espérés ($m_{ij} = n\,\pi_{ij}$) sont

Tableau 2. – Observations par combinaison des modalités des variables captures par unité d'effort: a) coefficient de marée, b) débit fluvial.
Observations for the combination of catch per unit effort: a) tidal coefficient variables, b) rate of flow variables.

	Captures par unité d'effort			
	Faible	Faible à Moyen	Moyen à Fort	Fort
a) Coefficient de marée				
Faible	35	20	17	17
Faible à Moyen	24	29	14	17
Moyen à Fort	17	12	25	27
Fort	19	18	23	28
b) Débit fluvial				
Faible	27	22	20	11
Faible à Moyen	20	20	24	26
Moyen à Fort	13	19	23	28
Fort	35	18	12	24

égaux à $m_{ij} = n\,\pi_{i+}\,\pi_{+j}$ pour tout i et j. Sur l'échelle logarithmique, l'indépendance s'exprime sous la forme additive $\log m_{ij} = \log n + \log \pi_{i+} + \log \pi_{+j}$. En notant X et Y les variables ligne et colonne, l'équation précédente est équivalente à

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y \tag{1}$$

avec μ le paramètre général, λ_i^X le paramètre associé à la modalité i de la variable X et λ_j^Y le paramètre associé à la modalité j de la variable Y. Pour être identifiables, ces paramètres satisfont à des contraintes telles que $\sum_i \lambda_i^X = \sum_j \lambda_j^Y = 0$.

S'il existe une dépendance entre les variables X et Y, le modèle s'écrit

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \tag{2}$$

Les nouveaux paramètres λ_{ij}^{XY} mesurent les écarts à l'indépendance des variables X et Y, ce sont des paramètres d'association. Pour être identifiables, ces paramètres satisfont aux contraintes $\sum_i \lambda_{ij}^X = \sum_j \lambda_{ij}^Y = 0$.

Considérons maintenant les observations du Tableau 1 relatives aux variables $X\,(df)$ et $Y\,(cm)$ croisées avec la variable $Z\,(cpue)$. Avec des notations analogues aux précédentes :

(i) si les trois variables sont mutuellement indépendantes ($\pi_{ijk} = \pi_{i++}\,\pi_{+j+}\,\pi_{++k}$ pour tout i, j et k), le modèle d'indépendance mutuelle s'écrit $\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$.

(ii) si la variable Z est indépendante de X et de Y ($\pi_{ijk} = \pi_{ij+}\,\pi_{++k}$ pour tout i, j et k), le modèle d'indépendance jointe s'écrit $\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$.

(iii) si X et Y sont indépendantes dans le tableau partiel associé à la modalité k de Z, X et Y sont dites conditionnellement indépendantes au niveau k de Z. Comme $\pi_{ij|k} = \pi_{i+|k}\,\pi_{+j|k}$ pour tout i et j avec, $\pi_{ij|k} = \pi_{ijk}/\pi_{++k}$ la distribution conjointe de X et Y au niveau k de Z, le modèle d'indépendance conditionnelle de X et de Y s'écrit :

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \tag{3}$$

Pour les tableaux à trois dimensions, $\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$ est le modèle loglinéaire le plus général qui peut être analysé. Pour l'interpréter, il faut décrire les associations λ_{ij}^{XY}, λ_{ik}^{XZ} et λ_{jk}^{YZ} qui sont des fonctions des rapports des chances (*odds ratios*) conditionnels (Agresti, 1990).

Dans le tableau marginal $(X \times Y)$, obtenu en sommant sur toutes les modalités k de la variable Z, il y a $(I-1)\ (J-1)$ *odds ratios* marginaux

$$\theta_{ij}^{XY} = \frac{\pi_{ij+}\,\pi_{i+1,j+1,+}}{\pi_{i+1,j,+}\,\pi_{i,j+1,+}} \quad (1 \le i \le I-1,\ 1 \le j \le J-1).$$

qui décrivent l'association marginale de X et de Y.

Dans le tableau partiel associé à la modalité k de la variable Z il y a $(I-1)(J-1)$ *odds ratios* conditionnels

$$\theta_{ij(k)} = \frac{\pi_{ijk}\,\pi_{i+1,j+1,k}}{\pi_{i,j+1,k}\,\pi_{i+1,j,k}} \quad (1 \le i \le I-1,\ 1 \le j \le J-1)$$

qui décrivent l'association conditionnelle de X et de Y. Ces *odds ratios* sont les mêmes si Z et X ou Z et Y sont conditionnellement indépendantes (Bishop, 1971). De même, les tableaux marginaux $(X \times Z)$ et $(Y \times Z)$ donnent les mêmes *odds ratios* que les tableaux partiels correspondants si X et Y sont conditionnellement indépendantes. A partir de ces résultats, on peut avoir une démarche pas à pas pour mesurer et interpréter les associations entre une variable réponse Z (*cpue*) et des variables facteurs $X\,(df)$ et $Y\,(cm)$. Elle consiste à étudier les tableaux marginaux $(X \times Z)$ et $(Y \times Z)$ puis à vérifier sur le tableau $(X \times Y \times Z)$ que X et Y sont conditionnellement indépendantes.

Pour vérifier le bon ajustement de ces modèles aux observations, différentes statistiques de test sont possibles, nous utilisons la statistique de Wald et la statistique G^2 du rapport de vraisemblance qui permet de tester l'hypothèse que les effectifs espérés de la population satisfont à un modèle donné. Pour les grands échantillons, ces statistiques sont distribuées (on dit asymptotiquement distribuées) comme des χ^2, la dimension de l'espace des paramètres correspondant à l'hypothèse alternative et à celle de l'hypothèse nulle. Si on ne fait pas d'hypothèse d'ordre sur les modalités des variables, les estimations des paramètres et les statistiques de test sont invariantes quel que soit l'ordre des catégories. Si on exploite le caractère ordonné des modalités des facteurs, on peut mettre en évidence des associations qui quantifient des tendances monotones uniformes par exemple (les *cpue* varient proportionnellement aux *cm*) ou des ordonnancements stochastiques (les *df* ont des effets sur les *cpue* mais ces effets ne sont pas les mêmes sur les modalités de *cpue*). Ces derniers modèles sont plus faciles à interpréter que les modèles nominaux, ils peuvent être non saturés alors que leurs équivalents sur l'échelle nominale le sont, leurs tests sont plus puissants pour détecter certains types d'associations.

RÉSULTATS

Pour étudier les associations entre les variables hydroclimatiques (cm et df) et la variable de production ($cpue$), nous avons analysé les Tableaux marginaux $2a$ et b puis le Tableau 1. Les calculs ont été réalisés avec la procédure PROC CATMOD du logiciel SAS/STAT. Les procédures de tendance uniforme, d'ordonnancement stochastique et nominale sont données en annexe.

Analyse des tableaux marginaux

Les tableaux marginaux sont obtenus en sommant les comptages des cellules du Tableau 1 sur les modalités de df (Tableau $2a$) et sur les modalités de cm (Tableau $2b$). Les modèles loglinéaires nominaux les plus généraux appliqués aux données du Tableau $2a$ ($\log m_{jk} = \mu + \lambda_j^Y + \lambda_k^Z$) et à celles du Tableau $2b$ ($\log m_{ik} = \mu + \lambda_i^X + \lambda_k^Z$) sont des modèles d'indépendance. Les statistiques de test $G^2_{(\text{tabl. } 2a)} = 23,88$ (Prob.$= 0,005$) et $G^2_{(\text{tabl. } 2b)} = 25,17$ (Prob.$= 0,003$) pour 9 degrés de liberté montrent que, dans chacun de ces cas, les variables ne peuvent être considérées comme indépendantes, les effectifs estimés sont très différents des effectifs observés, certains résidus ajustés (Haberman, 1973) sont élevés (Tableau $3a$ et b). Pour le modèle d'indépendance associé au Tableau $2a$ les résidus ajustés sont définis par :

$$r_{jk} = \frac{n_{jk} - \hat{m}_{jk}}{[\hat{m}_{jk}\,(1 - p_{j+})\,(1 - p_{+k})]^{1/2}}$$

avec $\hat{m}_{jk} = np_{j+}\,p_{+k}$, p_{j+} et p_{+k} étant les proportions marginales lignes et colonnes du Tableau $2a$ (on a une formule analogue pour le modèle d'indépendance associé au Tableau $2b$). Dans le Tableau $3a$, les plus forts résidus positifs sont sur la première diagonale, les plus forts résidus négatifs sont sur la seconde diagonale. Cette structure particulière suggère que les variables cm et $cpue$ évoluent dans le même sens, il importe donc de modéliser cette tendance. Pour le Tableau $3b$, on ne trouve pas de structure comparable des résidus, les évolutions conjointes de df et $cpue$ sont plus ératiques.

Analyse de la liaison captures par unité d'effort-coefficient de marée

On introduit dans le modèle un paramètre d'association entre cm et $cpue$ en assignant, pour respecter l'ordre des modalités des variables, des scores croissants $\{u_j\}$ et $\{v_k\}$ aux lignes et aux colonnes du Tableau $2a$. Le modèle peut alors s'écrire :

$$\log m_{jk} = \mu + \lambda_j^Y + \lambda_k^Z + \beta\, u_j\, v_k \qquad (4)$$

Tableau 3. – Valeurs estimées des observations et résidus ajustés (entre parenthèses). Captures par unité d'effort : a) coefficient de marée; b) débit fluvial.
Estimated values of observations and adjusted residuals (in brackets). Catch per unit effort: a) tidal coefficient variables; b) rate of flow variables.

	Captures par unité d'effort			
	Faible	Faible à Moyen	Moyen à Fort	Fort
a) Coefficient de marée				
Faible	24,72 (2,82)	20,56 (-0,16)	20,56 (-1,04)	23,16 (-1,73)
Faible à Moyen	23,33 (0,18)	19,40 (2,86)	19,40 (-1,61)	21,86 (-1,39)
Moyen à Fort	22,50 (-1,57)	18,71 (-2,03)	18,71 (1,90)	21,08 (1,72)
Fort	24,44 (-1,51)	20,33 (-0,68)	20,33 (0,78)	22,90 (1,44)
b) Débit fluvial				
Faible	22,22 (1,36)	18,48 (1,07)	18,48 (0,46)	20,82 (-2,86)
Faible à Moyen	25,00 (-1,37)	20,79 (-0,23)	20,79 (0,94)	23,42 (0,72)
Moyen à Fort	23,06 (-2,83)	19,17 (-0,05)	19,17 (1,15)	21,60 (1,84)
Fort	24,72 (2,83)	20,56 (-0,75)	20,56 (-2,50)	23,16 (0,24)

il est intermédiaire entre le modèle d'indépendance ($\beta = 0$) et le modèle saturé ($\beta\, u_j\, v_k$ est une forme particulière de λ_{jk}^{YZ}), il ne fait perdre qu'un seul degré de liberté à la résiduelle. En utilisant les scores $\{u_j = j\}$ et $\{v_k = k\}$, espacés d'une unité, on remarque que pour deux lignes adjacentes (j et $j+1$) et deux colonnes adjacentes (k et $k+1$), $\log[(m_{jk}\, m_{j+1,\,k+1})/(m_{j,\,k+1}\, m_{j+1,\,k})] = \beta\,(u_{j+1} - u_j)\;(v_{k+1} - v_k) = \beta$. Ainsi, $\theta_{jk} = \exp(\beta)$ est un *odds ratio* commun à toutes les paires de lignes et de colonnes adjacentes, il se réfère à des associations uniformes. En utilisant ces scores, on obtient : $G^2_{(\text{tabl. } 2a)} = 10,90$ (Prob. $= 0,208$) pour 8 degrés de liberté, $\hat{\beta} = 0,152$ avec $\hat{\sigma}_{\hat{\beta}} = 0,04$ pour écart-type asymptotique. La valeur positive de $\hat{\beta}$ indique que les $cpue$ tendent à être d'autant plus grandes que les valeurs des modalités de cm sont plus grandes. L'estimation du *odds ratio* uniforme ($\exp(0,152) = 1,16$) indique que par rapport à la catégorie k, la $cpue$ de la catégorie $k+1$ a augmenté d'un facteur multiplicatif 1,16 pour chaque changement croissant de catégorie pour cm. En considérant les cellules des coins extrêmes du Tableau $2a$, le *odds ratio* estimé pour une forte $cpue$ est $\exp[0,152\,(4-1)\,(4-1)] = 3,93$ fois plus important pour la catégorie "Fort" que pour la catégorie "Faible" de cm. Un intervalle de confiance à 95 % pour $\exp(\beta)$ est $\exp(0,152 \pm 1,96 \times 0,04)$, soit $(1,08;\, 1,26)$.

Le modèle (4) a structurellement la forme du logarithme de la densité d'une distribution binormale. Ce modèle tend à donner un bon ajustement quand la distribution continue sous-jacente est approximativement binormale. Pour des scores standardisés, β est comparable à $\rho/(1-\rho^2)$, où ρ est le coefficient de corrélation (Agresti, 1990). Si les associations sont faibles $\beta \simeq \rho$. L'analyse avec les scores standardisés donne : $G^2_{(\text{tabl. 2}a)} = 11,77$ (Prob.= $0,162$) pour 8 degrés de liberté, $\hat{\beta} = 0,196$ avec $\hat{\sigma}_{\hat{\beta}} = 0,06$ pour écart-type asymptotique. S'il y a une distribution sous-jacente binormale, la corrélation estimée est $\hat{\rho} = 0,189$. Ainsi, le modèle d'association uniforme met en évidence une association positive relativement faible.

Analyse de la liaison captures par unité d'effort-débit fluvial

Nous avons vu, dans l'analyse des résidus, que les évolutions conjointes des *df* et des *cpue* ne mettaient pas en évidence l'existence éventuelle d'une tendance (Tableau 2*b*). Le modèle précédent n'est donc pas adapté, on ne peut plus considérer qu'il y a une relation d'ordre sur *df* d'une part et sur *cpue* d'autre part. Puisque nous cherchons à expliquer l'importance de la variable réponse *cpue* par la variable facteur *df*, nous ne prendrons en compte que la relation d'ordre sur les *cpue*. Dans ce cas, le modèle s'écrit :

$$\log m_{ik} = \mu + \lambda_i^X + \lambda_k^Z + \mu_i v_k \qquad (5)$$

les $\{v_k\}$ sont des scores et les $\{\mu_i\}$ sont des effets lignes. Le modèle (5) a $(I-1)$ paramètres de plus que le modèle d'indépendance dans lequel tous les μ_i seraient égaux.

En utilisant les scores $\{v_k = k\}$, on obtient : $G^2_{(\text{tabl. 2}b)} = 9,55$ (Prob.= $0,145$) pour 6 degrés de liberté. Pour les lignes h et i et les colonnes adjacentes k et $k+1$, le logarithme des *odds ratios* ($\log[(m_{hk}\, m_{i,k+1})/(m_{h,k+1}\, m_{ik})] = \mu_i - \mu_h$) est identique pour toutes les $K-1$ paires de colonnes adjacentes. Les $\{\mu_i - \mu_h\}$ décrivent les différences entre lignes. Quand $\mu_i = \mu_h$, les lignes i et h ont des distributions identiques. Si $\mu_i > \mu_h$, les *cpue* sont stochastiquement plus élevées dans la ligne i que dans la ligne h. Avec les contraintes d'identification des paramètres définies précédemment, on trouve $\mu_1 = -0,221$, $\mu_2 = 0,116$, $\mu_3 = 0,252$ et $\mu_4 = -0,147$. On en déduit que les *cpue* sont stochastiquement plus importantes pour la catégorie "Moyen à Fort" que pour la catégorie "Faible à Moyen". Des conclusions comparables peuvent être tirées en considérant les autres couples μ_i et μ_h.

La première analyse montre que le niveau des captures par unité d'effort augmente proportionnellement à la valeur du coefficient de marée. La deuxième analyse met en évidence un mécanisme d'action différent, les débits fluviaux "Faible" ou "Fort" sont peu propices à des *cpue* importantes alors que les débits fluviaux "Faible à Moyen" et "Moyen à Fort" sont plus favorables, ces relations n'étant que stochastiquement vérifiées. Ces conclusions étant déduites de deux analyses séparées, rien ne peut être dit sur les interactions éventuelles entre *cm* et *df*, elles doivent donc être complétées par une analyse globale du Tableau 1.

Analyse globale

Considérons maintenant le modèle

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \quad (6)$$

Les résultats de cette analyse sont donnés dans le Tableau 4, ils montrent la bonne adéquation des données au modèle, $G^2_{\text{résidu}} = 26,16$ (Prob.= $0,510$) pour 27 degrés de liberté précise que les résidus sont peu importants. Conditionnellement aux *cpue*, *cm* et *df* sont indépendantes, la statistique de Wald est égale à 8,51 (Prob.= $0,484$) pour 9 degrés de liberté, seules les interactions *cm* × *cpue* et *df* × *cpue* sont significatives, les statistiques de Wald étant respectivement égales à 24,66 (Prob.= $0,003$) et 23,51 (Prob.= $0,005$) pour 9 degrés de liberté. D'après les résultats du paragraphe sur les modèles, on est en droit d'accepter que les variables *cm* et *cpue* d'une part, *df* et *cpue* d'autre part sont liées par les relations décrites dans l'étude des tableaux marginaux.

Pour décrire ces données par un modèle plus parcimonieux, on structure l'interaction *cm*×*cpue* en assignant des scores croissants $\{u_j\}$ à *cm* et $\{v_k\}$ à *cpue*. Ce nouveau modèle s'écrit : $\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \beta u_j v_k$ (le terme λ_{jk}^{YZ} a été remplacé par $\beta u_j v_k$). Les *odds ratios* conditionnels satisfont à $\log[(m_{ijk}\, m_{i,j+1,k+1})/(m_{i,j,k+1}\, m_{i,j+1,k})] = \beta(u_{i+1} - u_i)\,(v_{j+1} - v_j) = \beta$ $(1 \le j \le J-1$, $1 \le k \le K-1)$ pour tout i, $G^2_{(\text{tabl. 1})} = 38,46$ (Prob.= $0,316$) pour 35 degrés de liberté, $\hat{\beta} = 0,155$ avec $\hat{\sigma}_{\hat{\beta}} = 0,04$ pour écart-type asymptotique. Les estimations de $\hat{\beta}$ et de $\hat{\sigma}_{\hat{\beta}}$ sont identiques à celles

Tableau 4. – Tableau d'analyse de variance : captures par unité d'effort (cpue), coefficient de marée (cm) et débit fluvial (df), degrés de liberté (d.d.l.).
Analysis of variance table: catch per unit effort (cpue), tidal coefficient (cm), rate of flow (df) variables, degrees of freedom (d.d.l.)..

Source	d.d.l.	Statistique de Wald	Probabilité
cpue	3	0,97	0,810
cm	3	0,69	0,875
df	3	1,35	0,717
cm × cpue	9	24,66	0,003
df × cpue	9	23,51	0,005
cm × df	9	8,51	0,484
Résidu	27	26,16	0,510

trouvées précédemment ($\hat{\beta} = 0,152$, $\hat{\sigma}_{\beta} = 0,04$). Ce dernier modèle est tout à fait satisfaisant, on ne peut plus le simplifier car, comme nous l'avons vu, les relations entre *df* et *cpue* ne sont que stochastiquement vérifiées.

DISCUSSION

Sur les données analysées, on aurait pu raisonner conditionnellement à *cm* et *df* et utiliser le modèle logit généralisé si on ne voulait pas tenir compte de l'ordre sur la variable réponse (Prouzet *et al.*, 1994) ou des modèles logit adjacent ou logit cumulé pour prendre en compte l'ordre sur les *cpue*. Appliqués aux aloses, tous ces modèles donnent des résultats comparables, ils suggèrent l'impact des variations conjuguées du débit fluvial et du coefficient de marée sur l'abondance des captures journalières, le modèle (6) est d'ailleurs identique au modèle logit généralisé (McCullagh and Nelder, 1989). Une autre approche possible, essentiellement exploratoire, serait l'analyse des correspondances du tableau de contingence croisant ($df \times cm$) avec *cpue* pour visualiser les profils lignes (Prouzet *et al.*, 1994) et tester approximativement l'influence de *df* et *cm* sur la dispersion de ces profils (Tenenhaus *et al.*, 1993). A toutes ces analyses, nous préférons celle basée sur la modélisation des interactions dans le modèle loglinéaire car elle permet de quantifier les liens entre deux variables. Ces liaisons n'étant pas identifiées par des expressions analytiques exactes, le choix *a priori* de scores régulièrement espacés semble raisonnable et de plus ils sont facilement interprétables. Si des scores régulièrement espacés ne s'imposaient pas de façon naturelle, il serait nécessaire de compléter ces analyses par une étude de sensibilité en éprouvant plusieurs ensembles de scores et en vérifiant si les conclusions diffèrent fondamentalement. D'un point de vue général les tableaux de contingences sont des représentations simplifiées d'observations continues, les catégories sont choisies par l'expérimentateur, elles dépendent de la distribution des variables continues sous-jacentes et des objectifs assignés à l'étude. Les scores doivent être le reflet de la construction du tableau de contingence et de son utilisation. Le problème traité en est un exemple, la modélisation des interactions à partir de scores également espacés est ici un choix admissible eu égard à la construction des catégories.

La mise en œuvre de ces modèles est simple (*cf.* annexe), elle peut se résumer en disant qu'elle débute par l'utilisation du modèle loglinéaire nominal le plus général qui s'ajuste bien aux données et qu'elle est poursuivie en recherchant le modèle le plus parcimonieux, les paramètres d'interaction étant remplacés par des termes structurés tenant compte de l'ordre sur les modalités des variables.

L'utilisation de ces modèles en hydrobiologie peut être généralisée à l'étude des interactions entre plusieurs variables dont l'une est à expliquer. Les exemples sont nombreux : variation des densités en fonction des caractéristiques du milieu; fréquence d'infestation d'un parasite en fonction des saisons, du milieu et des caractéristiques de l'hôte, etc. Plus généralement, ils permettent de compléter les résultats obtenus par des analyses descriptives multivariées (Tenenhaus *et al.*, 1993) par la modélisation de la loi de probabilité d'une variable y en fonction des variables explicatives et de décrire l'influence de celle-ci sur la loi de probabilité de y. Ainsi que le notent ces auteurs, le modèle linéaire généralisé permet de valider statistiquement des résultats que l'analyse des correspondances ne fait que suggérer.

ANNEXE

Les modèles utilisés sont ajustés en spécifiant la matrice du modèle d'un modèle logit généralisé.

Liaison coefficient de marée-captures par unité d'effort

Pour le modèle d'association uniforme, la matrice du modèle est de dimension 12×4 puisqu'il y a trois logits ($\log(m_{jk}/m_{j4})$) dans chaque ligne du Tableau 2*a* et quatre paramètres à estimer. Les trois premiers paramètres d'une ligne se réfèrent aux interceptions des trois logits. En utilisant les scores $u_j = j$ et $v_k = k$, le coefficient de β dans le modèle logit est $u_j(v_k - v_4)$.

```
OPTIONS PS=1000 LS=80;

  DATA tab2a;
  INFILE 'tab2a';
  INPUT rep cm cpue;
PROC CATMOD;
WEIGHT rep;
POPULATION cm;
       MODEL cpue=

     (100   -3,  010  -2,  001  -1,
      100   -6,  010  -4,  001  -2,
      100   -9,  010  -6,  001  -3,
      100  -12,  010  -8,  001  -4)

       /ML PRED=FREQ;
RUN;
```

Liaison débit fluvial et captures par unité d'effort

Le modèle d'association stochastique utilise la contrainte $\mu_4 = -(\mu_1 + \mu_2 + \mu_3)$, les quatrième, cinquième et sixième éléments de chaque ligne sont les coefficients $(v_k - v_4)$ de μ_1, μ_2 et μ_3 du modèle logit généralisé ($\log(m_{ik}/m_{i4})$).

```
OPTIONS PS=1000 LS=80;
  DATA tab2b;
  INFILE 'tab2b';
  INPUT rep df cpue;
PROC CATMOD;
WEIGHT rep;
POPULATION df;
       MODEL cpue=

(100 -3  0  0, 010 -2  0  0, 001 -1  0  0,
 100  0 -3  0, 010  0 -2  0, 001  0 -1  0,
 100  0  0 -3, 010  0  0 -2, 001  0  0 -1,
 100  3  3  3, 010  2  2  2, 001  1  1  1)

       /ML PRED=FREQ;
RUN;
```

Liaison coefficient de marée-débit fluvial-captures par unité d'effort

Pour le modèle loglinéaire nominal le plus général on utilise LOGLIN.

```
OPTIONS PS=1000 LS=80;
  DATA tab1;
  INFILE 'tab1';
  INPUT rep cm df cpue;
PROC CATMOD;
WEIGHT rep;
FACTORS cm $ 4, df $ 4, cpue $ 4;
  MODEL      cm*df*cpue=_RESPONSE_/ML
PRED=FREQ;
  LOGLIN cpue cm df cpue*cm cpue*df
cm*df;
RUN;
```

RÉFÉRENCES

Agresti A. 1990. Categorical Data Analysis. John Wiley & Sons. New York, 558p.

Bishop Y. V. V. 1971. Effects of collapsing multidimensional contingency tables. *Biometrics* **27**, 545-562.

Bishop Y. V. V., S. E. Fiengerg, P. Holland 1975. Discrete Multivariate Analysis: theory and practice, MIT Press, Cambridge, Mass.

Fienberg S. E., 1980. The analysis of cross classified categorical data, 2nd ed. MIT Press, Cambridge, Mass.

Haberman S. J. 1973. The analysis of residuals in cross-classification tables. *Biometrics* **29**, 205-220.

McCullagh P., J. A. Nelder 1989. Generalized Linear Models. Monographs on statistics and applied probability. Chapman and Hall, London.

Prouzet P., J. P. Martinet, J. Badia 1994. Caractérisation biologique et variations des captures de la grande alose (*Alosa alosa*) par unité d'effort sur le fleuve Adour, (Pyrénées Atlantiques, France). *Aquat. Living Resour.* **7**, 1-10.

SAS Institute 1990. SAS/STAT User's guide. Version 6, Fourth Edition.

Tenenhaus M., Y. Le Roux, C. Guimard, P. L. Gonzalez 1993. Modèle linéaire généralisé et analyse des correspondances. *Rev. Stat. Appl.* **41**, 59-86.