Modèle hiérarchique bayésien pour analyser les contenus stomacaux de prédateurs marins : application aux thons tropicaux

Marie Morfin, Marie-Pierre Etienne, Frédéric Ménard

▶ **To cite this version:**

Marie Morfin, Marie-Pierre Etienne, Frédéric Ménard. Modèle hiérarchique bayésien pour analyser les contenus stomacaux de prédateurs marins : application aux thons tropicaux. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386667

HAL Id: inria-00386667

https://inria.hal.science/inria-00386667

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèle hiérarchique bayésien pour analyser les contenus stomacaux de prédateurs marins: application aux thons tropicaux

Marie Morfin[1,2], Marie-Pierre Etienne[1], Frédéric Ménard[3]

1 UMR 518 MORSE, INRA AgroParisTech, 16 rue Claude Bernard, 75 231 Paris cedex 05, France

2 UMR 211 Agronomie, INRA AgroParisTech, 78 850 Thiverval-Grignon, France

3 UR 109 THETIS, IRD, Centre de Recherche Halieutique Méditerranéenne et Tropicale, Avenue Jean Monnet, 34203 Sète cedex, France

Résumé

L'analyse du contenu stomacal d'un animal permet d'étudier son comportement alimentaire. Dans le cas de prédateurs supérieurs, cette analyse est d'autant plus importante puisqu'elle présente le haut de la chaîne trophique. Les données stomacales peuvent consister en un dénombrement de proies dans l'estomac ou en une mesure de poids de proies. Ce type de données présente deux particularités : 1. elles sont agrégées, 2. de très nombreux estomacs sont vides (données zero-inflated).
Le travail présenté ici propose d'utiliser un modèle de Poisson composé pour modéliser des contenus stomacaux dans le cas de données de comptage. Le modèle présenté permet en outre de tenir compte de variables explicatives telles que la taille du prédateur, les différences entre espèces ainsi que la structuration en groupes des prédateurs étudiés. Le modèle ainsi construit est un modèle hiérarchique et l'inférence est menée dans un cadre bayésien. Une étude par simulation a permis d'établir l'importance de la structuration en bancs pour l'amélioration de la qualité d'estimation. Ce modèle a été particulièrement développé pour analyser les contenus stomacaux de deux espèces de thons tropicaux.

Abstract

The feeding behavior of an animal can be studied by observing its stomach content. The case of top predators is all the more important because they affect the whole food chain. Stomach data may be a specific enumeration of the preys in the stomach or a global weight of preys. Such data have two characteristics: 1. they are aggregated, 2. many stomachs are empty (zero-inflated data).
The work presented here proposes to use a compound Poisson model of stomach contents in the case of counting data. This model can also take into account variables such as the size, the differences between species and the structuring in schools of the predators

studied. The resulting model is hierarchical and the inference is conducted in a Bayesian framework. A simulation study allows to evaluate the importance of the structuring in schools in the quality of parameters estimation. This model has been developed to analyze the stomach contents of two tropical tuna species.

Mot-clés : Données "zero-inflated", modèles de Poisson-composés, inférence bayésienne, algorithmes MCMC, prédateurs marins.

Introduction

Les prélèvements de prédateurs supérieurs tels que le thon ont des répercussions sur l'ensemble du réseau trophique par de complexes interactions prédateurs-proies (Shin *et al.* (2003)). La connaissance du régime alimentaire de ces prédateurs supérieurs donne accès à un échantillon des niveaux intermédiaires de leur écosystème (céphalopodes, poissons, mollusques...). Cependant ces prédateurs sont difficiles à observer au quotidien, c'est pourquoi les écologues étudient leur alimentation à travers l'observation de leur contenu stomacal. L'analyse du contenu stomacal permet de déterminer le nombre de proies ingérées par le thon pendant une période t, correspondant au temps de digestion supposé commun à tous les thons. Cette observation est le résultat de deux phénomènes inobservés et aléatoires : la rencontre du prédateur avec un agrégat de proies et le nombre de proies qu'il a attrapées sur cet agrégat.

L'article de Magnússon et Aspelund (1997) propose d'utiliser un modèle de Poisson Composé pour rendre compte de ces deux sources de variabilité. Ce type de modèle permet également de traiter des données "zero-inflated", alors que les distributions standard s'ajustent mal aux données (Mayer *et al.* (2005)).

Dans une première partie nous détaillons pas à pas la construction hiérarchique du modèle. Après une étude par simulation en deuxième partie pour vérifier la qualité des estimations proposées, nous proposons d'utiliser ce modèle pour répondre à deux questions biologiques concernant l'influence des Dispositifs de Concentration de Poissons (DCP)[1] sur le régime alimentaire du thon tropical d'une part ainsi que les différences de comportement alimentaire selon les espèces.

Construction du modèle

Le nombre de proies Y_k observées dans l'estomac du thon k est considéré comme la somme du nombre de proies X_{kl} mangées à chaque rencontre l avec un agrégat de proies (taille du repas), sur le nombre total N_l de rencontres (repas). N_l est modélisé par une loi de Poisson de paramètre μ, et X_{kl} suit une loi géométrique[2] de paramètre p. Ce premier modèle peut s'écrire :

[1]Un DCP est une épave flottante disposée intentionnellement ou non. Les thons comme de nombreux autres poissons ont tendance à s'aggréger sous ces épaves.

[2]La loi géométrique est définie par $P(X = l) = p(1-p)^l$, donc $E(X) = \frac{1-p}{p}$.

$$Y_k = \sum_{l=1}^{N_k} X_{kl} \quad \text{avec} \quad N_k \sim \mathcal{P}(\mu) \text{ et } (X_{kl})_{l=1,\dots,N_k} \overset{i.i.d}{\sim} \mathcal{G}(p)$$

Pour introduire plus de variabilité entre les repas on généralise la loi de X par une binomiale négative de paramètres r et p. r est alors un paramètre de nuisance, qui permet à la loi de X d'être surdispersée par rapport à une géométrique lorsqu'il est inférieur à 1.

Sachant que les proies concernées s'agrègent dans des bancs de très grande taille, on suppose que le prédateur se nourrit jusqu'à satiété à chacun de ses repas. Il est alors raisonnable de supposer que la taille du repas X est proportionnelle au volume V de l'estomac du thon. Comme $E(X) = r\,(1-p)/p$ et le paramètre de nuisance r n'a aucune raison de dépendre des caractéristiques du prédateur, on propose une relation de proportionnalité entre le volume V de l'estomac et la quantité $\frac{(1-p)}{p}$. De plus le volume V de l'estomac d'un prédateur est lié à sa longueur l par une relation allométrique de la forme $V = c\,l^3$.

En passant au logarithme, on introduit naturellement un lien logistique entre le paramètre p de la binomiale négative et la longueur l du prédateur :

$$\ln(\frac{1-p}{p}) = -logit(p) = \alpha + \beta\,\ln(l) \tag{1}$$

Dans le cadre de cette étude on s'intéresse particulièrement aux thons tropicaux qui s'agrègent en surface pour chasser. La senne, l'engin de pêche qui capture les thons, préserve cette organisation des thons en bancs. Ainsi il est fréquent d'observer plusieurs estomacs provenant d'un même coup de pêche (calée). Sachant que les thons restent généralement dans le même banc pendant une durée supérieure à la période de digestion t, il est important que le modèle construit intègre cette notion de groupes. On a donc imposé aux prédateurs issus d'un même groupe d'avoir rencontré les même agrégats de proies. Ainsi tous les thons k de la calée i ont rencontré N_i agrégats de proies.

Parmi les thons tropicaux on distingue plusieurs espèces de taille très variable, on introduit donc un effet de l'espèce du prédateur sur le couple de paramètres (α, β), traduisant une relation différente entre taille du prédateur et taille moyenne des repas. Le modèle final ainsi construit s'écrit :

$$\begin{aligned}(M)\quad & Y_{ijk}|N_i > 0 \sim \mathcal{BN}(r\,N_i, p_{ijk}) \quad \text{et} \quad Y_{ijk}|N_i = 0 \sim \delta_0,\\ & \text{où } (N_i)_{i=1,\dots,I} \overset{i.i.d}{\sim} \mathcal{P}(\mu), \quad logit(p_{ijk}) = \alpha_j + \beta_j\,\ln(l_{ijk}),\\ & \text{avec } i = 1,\dots,I \text{ la calée}, \quad j = 1,\dots,J \text{ l'espèce et } k = 1,\dots,K_{ij} \text{ le thon.}\end{aligned}$$

S'agissant d'une première étude sur le thon, nous ne disposions d'aucune connaissance a priori sur le nombre moyen de repas pour un thon (paramètre μ). L'intuition biologique

donne une valeur attendue pour β proche de 3 mais ce travail a justement pour but de vérifier cette intuition, aussi nous n'avons pas utilisé cette information pour construire un prior. Les paramètres des lois a priori ont donc été choisis pour les rendre plates. Le choix des formes de ces lois a été guidé par des raisons techniques et de façon à favoriser des propriétes de conjugaisons partielles. Finalement les priors choisis sont

$$\mu \sim \Gamma(0.001, 0.001) \quad \alpha_j \sim \mathcal{N}(0, 1000) \text{ et } \beta_j \sim \mathcal{N}(0, 1000), \forall\ j = 1, \dots J, \quad r \sim \mathcal{U}(0, 20)$$

L'inférence de ce modèle hiérarchique a été menée grâce à un algorithme Metropolis Within Gibbs (Gilks *et al.* (1996)). Dans les chaînes de Markov construites la mise à jour sur le paramètre μ se fait par Gibbs en profitant de la conjugaison Poisson-Gamma (Parent et Bernier (2007)) et par un pas de Metropolis pour α_j, β_j et r. Lorsque l'effet de la taille du thon n'est pas prise en compte, on pose un prior beta(1,1) sur le paramètre p et la mise à jour peut se faire par un pas de Gibbs. L'algorithme d'estimation a été implémenté[3] sous R (R Develpoment Core Team, 2008). La convergence a été systématiquement vérifiée en utilisant le critère de Gelman-Rubin (Gelman et Rubin (1992), Cowles et Carlin (1996)). Elle a lieu en général au bout de 10000 itérations.

Etude par simulation

Pour étudier la précision des distributions a posteriori en fonction du nombre d'observations et de la structuration en groupes de ces observations, nous proposons l'étude par simulation suivante. Pour un jeu de paramètres fixé ($\mu = 2, \alpha = -8.3, \beta = 2.54, r = 2$) on simule selon le modèle en faisant varier le nombre de groupes $i = 5, 10, 20, 50, 100$ et le nombre prédateurs par groupes $K_i = 1, 5, 10, 20$. Pour chaque répartition on simule 200 jeux de données et on estime les intervalles de crédibilité à 90%.

Cette étude a mis en évidence le rôle crucial de la structuration en groupes. En effet, observer plusieurs individus issus d'un même groupe facilite la désagrégation des données et donc l'estimation des paramètres. Ainsi les intervalles de crédibilité moyens des paramètres μ, α et β estimés à partir de 100 thons indépendants ont des largeurs respectives de $4.25, 0.94$ et 2.51 tandis que les intervalles de crédibilité lorsqu'on dispose de 10 groupes de 10 thons chacun valent en moyenne $3.26, 0.73$ et 1.48, soit une réduction de $23\%, 22\%$ et 41%. L'intervalle du paramètre μ, correspondant à l'espérance du nombre de repas d'un thon, évolue dans le sens inverse car tous les thons d'un même groupe sont supposés avoir rencontré les mêmes agrégats de proies.

Applications sur l'alimentation des thons tropicaux

Beaucoup d'espèces de poissons pélagiques s'agrègent naturellement autour de DCP (Dispositifs de Concentration de Poisson), désignant n'importe quel objet flottant à la surface de l'océan. Des DCPs artificiels sont placés par les thoniers pour augmenter leurs rendements, et sont souvent présentés comme un piège écologique ((Marsac *et al.* 2000).

[3]Les codes R sont disponibles sur simple demande à l'un des auteurs.

Les pêcheurs recherchent aussi les bancs de thons libres qui bien souvent se concentrent pour se nourrir dans des zones riches en agrégations de proies.

Nous disposons d'un jeu de données récoltées au large des Seychelles par des observateurs scientifiques dans une zone riche en bancs de *Natosquilla investigatoris*, un petit crustacé que l'on retrouve en forte abondance dans les estomacs de thons.

A chaque prise, on dispose de la localisation spatiale et temporelle, ainsi que du type de banc (sous DCP ou libre). Puis un échantillon de thons a été prélevé, pour identifier l'espèce auquel il appartient, sa longueur, et le nombre de proies par espèce contenues dans son estomac.

A partir de ces données nous avons sélectionné plusieurs échantillons d'individus pour répondre aux questions posées : effet de la taille ou de l'espèce de thon sur son alimentation, effet du type d'association (sous DCP ou en bancs libres).

La mise en œuvre d'une inférence sur tous les modèles possibles et la sélection du meilleur modèle au sens du critère DIC (Spiegelhalter *et al.* (2002)) a mis en évidence l'importance de l'effet taille pour le thon albacore. De plus l'intervalle de crédibilité à 90% pour β_{alb} est $[3.04; 4.10]$ ce qui est proche de la valeur de 3 attendue.

Grâce aux données disponibles sur les thons albacore et listao, nous avons montré que le nombre moyen de repas pour les thons sous DCP, μ_{DCP}, était bien inférieur au nombre moyen de repas pour les thons en banc libre μ_{libre}. La taille des repas est également plus petite.

Les thons listao sont en moyenne bien plus petits que les thons albacore. Ainsi ce sont les juvéniles albacores qui ont des tailles similaires aux thons listao. On peut donc se demander si à longueur comparable un listao et un albacore ont la même taille de repas. Or le modèle le mieux ajusté aux données est celui sans la prise en compte de l'espèce, ce qui indique que la relation liant longueur et taille de repas est la même quelle que soit l'espèce.

Conclusions et perspectives

Le modèle proposé pour étudier les contenus stomacaux prend en compte la structuration en bancs des proies et la structuration en groupe des prédateurs. Grâce à sa structure hiérarchique, le modèle prend en compte des paramètres ayant une signification biologique, ce qui les rend facilement interprétables. Les conclusions obtenues sur les thons tropicaux confirment les intuitions des biologistes en ce qui concerne l'impact des DCP sur l'alimentation du thon mais pas l'influence de l'espèce.

Le modèle néglige l'effet de satiété, ce qui est validé a posteriori par le fait que le nombre moyen de repas pour une durée de digestion t est de 1. Néanmoins il serait intéressant pour certaines autres espèces de prendre en compte cet effet de satiété en supprimant l'hypothèse d'indépendance entre les repas. De plus, nous avons sélectionné des périodes de l'année et des régions pour lesquelles le thon a une alimentation monospécifique car le modèle ne permet pas de prendre en compte différents types de proies. La modélisation du poids du contenu stomacal plutôt que du nombre de proies ingérées permettra d'étudier

des espèces s'alimentant sur plusieurs espèces de proies et de tenir compte de l'effet de satiété.

Enfin ce modèle a été développé pour l'analyse des contenus stomacaux de thons mais il est tout à fait utilisable pour d'autres types de prédateurs se nourrissant sur des proies structurées en bancs. Ainsi un travail similaire est en cours sur l'analyse des contenus stomacaux des manchots royaux.

Bibliographie

[1] Cowles, K. et Carlin, B. (1996) Markov chain monte carlo convergence diagnostics : a comparative review. *Journal of the American Statistical Association*, 91, 883–904.

[2] Gelman, A. et Rubin, D. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–472.

[3] Gilks, W., Richardson, S. et Spiegelhalter, D. (1996) *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC.

[4] Magnússon, K.G. et Aspelund, T. (1997) A model for estimating meal frequency and meal size from stomach data with an application to atlantic cod (*gadus morhua*) feeding on capelin (*mallotus villosus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 54, 876–889.

[5] Marsac, F., Fonteneau, A., et Ménard, F. (2000) Drifting FADs used in tuna fisheries: an ecological trap ? *Pêche thonière et dispositifs de concentration de poissons.* 28. Ed. ifremer, actes colloq. edition.

[6] Mayer, D., Roy, D., Robins, J., Halliday, I. et Sellin, M. (2005) Modelling zero-inflated fish counts in estuaries - a comparison of alternate statistical distributions, *MODSIM 2005 International Congress on Modelling and Simulation*, Zerger, A. and Argent, R., 2581–2587.

[7] Parent, E. et Bernier, J. (2007) Le raisonnement bayésien *Statistiques et Probabilités appliquées*, Springer.

[8] R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

[9] Spiegelhalter, D., Best, J., Nicola, G., Carlin, B.P. et Van der Linde, A.(2002) Bayesian measures of model complexity and fit *Journal of Royal the Society*, Serie B, 64, 583–639.