

MODELISATION DES ARRONDIS LORS DES
RELEVÉS SCIENTIFIQUES EN MER:
UNE APPROCHE BAYÉSIENNE

MODELING HEAPING IN MARINE SURVEYS:
A BAYESIAN APPROACH

Éric Parent ¹, Matthieu Authier ²,
Sophie Donnet ¹, Isabelle Albert ¹ & Daouda Ba ¹

¹ UMR MIA-Paris, AgroParisTech, INRA,
Université Paris-Saclay, 75005, Paris, France
² UMS 3462, CNRS-La Rochelle Université, France

Résumé. Les écosystèmes marins sont soumis à de nombreuses pressions (e.g. changement climatique), notamment liées aux activités humaines comme le transport maritime ou l'extraction de ressources (e.g. pêche). Les campagnes écosystémiques, c'est-à-dire les campagnes scientifiques de mesures *in situ* des diverses composantes physico-chimiques et biologiques d'un écosystème, permettent d'étudier les variations de l'abondance des espèces marines afin d'estimer et d'anticiper l'état des stocks. Les politiques de gestion appropriées se fondent sur les études issues de campagnes écosystémiques, par exemple, les campagnes mises en oeuvre par l'Ifremer sur le navire océanographique *Thalassa* dans l'Atlantique du Nord-Est. Le programme MEGASCOPE, déployé depuis 2004 dans le cadre des campagnes écosystémiques de l'Ifremer, vise à estimer l'abondance de la mégafaune marine, et en particulier des mammifères marins, à partir d'observations visuelles depuis le navire. L'observation des mammifères marins est difficile et perturbée de bien des manières : l'état de la mer, l'angle d'observation, la distance des animaux au bateau, leur trajectoire imprévisible et leur comportement (e.g. plongée) *etc.* En outre, une caractéristique particulière des données collectées est liée aux observateurs eux-mêmes: ces derniers ont l'habitude d'arrondir le nombre d'animaux observés lorsque les animaux sont présents en grands nombres. Ces difficultés de dénombrement rendent l'inférence de l'abondance vraie délicate et, par conséquent, l'évaluation fragile si celles-ci sont ignorées lors de l'analyse. Dans cette communication, nous développons une approche bayésienne afin de faire la part des choses entre les incertitudes provenant du processus de comptage, et notamment la tendance à arrondir des chiffres; et celles résultant de la stochasticité environnementale.

Mots-clés. Environnement, Méthodes bayésiennes

Abstract. Marine ecosystems face a myriad of pressures, including climate change and anthropogenic activities such as increased shipping or fishing. Ecosystemic surveys,

that is scientific integrated survey designed to monitor holistically an ecosystem, provide the best means to study variations in species abundance at the ecosystem scale, and thereby to anticipate stock changes and design adequate management policies. Integrated surveys are routinely carried out in the North-East Atlantic Ocean by Ifremer onboard R/V *Thalassa*. Since 2004, the MEGASCOPE program is part of such integrated surveys and aims specifically at documenting from visual observations the abundance and distribution of marine megafauna, and marine mammals in particular. Visual observation however is complicated by various factors such as sea state, observation angle, distance of animals to the boat, their unpredictable movements, their behaviour (e.g. diving), *etc.* In addition, an intrinsic feature of the reported count data is the tendency of observers to round the number of observed animals when these are very numerous. All these measurement difficulties result into making inference on true abundance rather tricky and, if ignored, into fragile assessment. In this paper, we develop a Bayesian approach to disentangle the uncertainty due to the counting process, and in particular due to rounding, and the uncertainty stemming from environmental stochasticity.

Keywords. Environment, Bayesian methods

1 Les relevés scientifiques de présence de dauphins

Les campagnes écosystémiques permettent d'étudier l'ensemble des compartiments d'un écosystème et représentent à ce titre un outil de suivi particulièrement appropriés pour une gestion holistique de l'environnement marin (Doray et al., 2018). La mégafaune marine inclut l'ensemble des animaux marins visibles à l'oeil nu depuis un navire. Elle comprend notamment les oiseaux marins, les tortues, les grands poissons (e.g. requins) et les mammifères marins (e.g. dauphins). Ces derniers sont des espèces charismatiques, protégées et souvent menacées. Les données collectées lors des campagnes écosystémiques permettent de construire une estimation de l'abondance de l'espèce cible, par exemple le dauphin commun *Delphinus delphis*. Le programme MEGASCOPE, par exemple, est mis en oeuvre depuis 2004 sur les campagnes écosystémiques de l'Ifremer à bord du navire océanographique *Thalassa* dans l'Atlantique Nord-Est. Calculé annuellement, l'indicateur statistique de l'abondance rend compte des variations de stocks et fournit une base scientifique pour l'élaboration de stratégies de conservation. La figure 1 présente la localisation des relevés scientifiques de présence de dauphins communs et dauphins bleu et blanc (*Stenella coeruleoalba*) lors des campagnes 2004 à 2016 dans le golfe de Gascogne et en Manche Ouest. Mais l'observation de la présence de dauphins, classiquement réalisée par deux observateurs à l'avant du pont d'un bateau n'est pas aisée. D'une part, le champ d'observation est limité lorsque la mesure est réalisée depuis un bateau (voir figure 2). D'autre part, la trajectoire des mammifères est imprévisible, la distance des groupes d'individus au vaisseau est variable, la profondeur a un effet déformant, on constate les

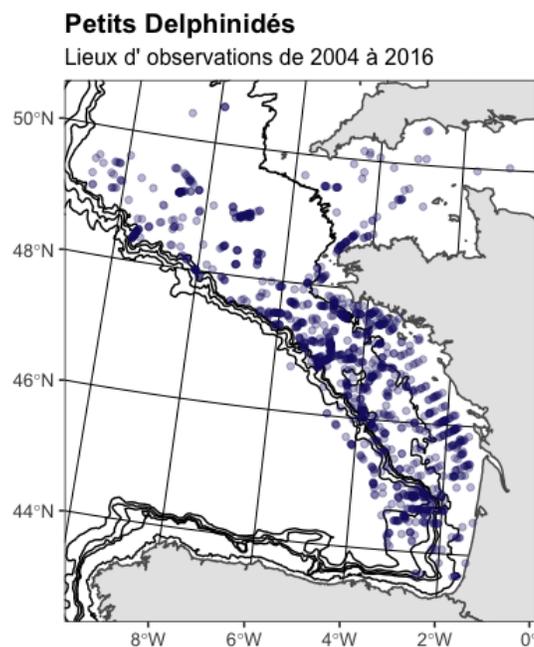


Figure 1: Localisation des relevés scientifiques de présence de Delphinidés lors des campagnes 2004 à 2016 dans le golfe de Gascogne et en Manche Ouest.

effets de parallaxe dus à l'angle d'observation, etc. Pour toute ces raisons et à cause de la courte durée du phénomène observé, les observateurs arrondissent les comptages qu'ils enregistrent, comme le montre la figure 3. La nécessité de prendre en compte les effets des arrondis est bien connue en médecine et a été étudiée par Wang and Heitjan (2008) , Crawford et al. (2015) ainsi que par Allen et al. (2017).

Mais qu'en est-il de l'impact des arrondis en halieuthique? Comment biaisent-ils l'estimation de l'abondance? Ces arrondis entraînent bien sûr une augmentation de l'incertitude qui porte sur cette grandeur. Vaut-il la peine de la prendre en compte? Nous développons ci-après un modèle statistique hiérarchique pour tenter de mettre en lumière toutes les sources d'incertitude qui relèvent de cette question.

2 Objectifs du travail

Un modèle hiérarchique statistique Une première étude des données suggère un modèle poissonien (voir panel de gauche de la figure 3). Le phénomène d'arrondi est mis en évidence par l'échelle log du panel de droite de la figure 3 montrant la fréquence des relevés : on constate sur cette figure, une sur-représentation des observations multiples de 10 ou 100.

Appelons :

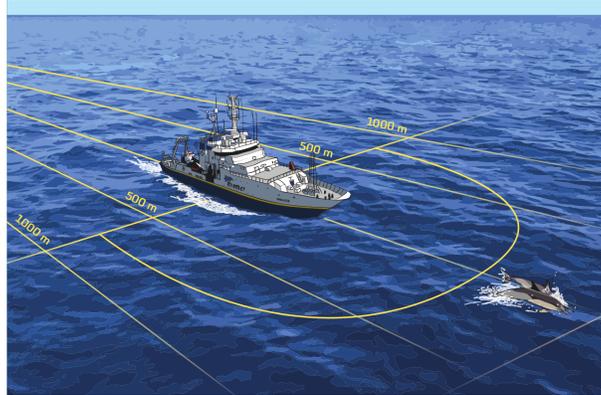


Figure 2: Les données sont recueillies par deux observateurs à l’avant du bateau.

- Y_i le relevé avec arrondis par les observateurs du nombre de dauphins au site i ,
- Z_i le nombre exact (sans arrondi) de dauphins au site i ,
- X_i les covariables explicatives de la présence de dauphins au site i telles que la bathymétrie, la distance à la côte, etc.

Nous proposons le modèle structuré hiérarchiquement suivant:

$$\log(\lambda(X_i)) = X_i\beta + \alpha_{t(i)} + \varepsilon_i \quad (1)$$

$$Z_i \sim \text{Poisson}(\lambda(X_i)) \quad (2)$$

$$Y_i \sim \mathcal{H}(Z_i, \delta) \quad (3)$$

L’équation 2 traduit le caractère poissonien des données de comptages. On choisit classiquement une fonction de lien (équation 1), par exemple *log* pour relier le nombre moyen d’individus λ aux variables explicatives (coefficient β). On peut introduire un effet aléatoire sur l’intercept α selon l’année (choisi souvent normal de variance Σ_α) et un effet aléatoire ε de proximité spatiale (sous hypothèse de normalité, Σ_ε représente ici la matrice de variance–covariance spatiale).

L’équation 3 permet de passer du *vrai* comptage au comptage reporté (et possiblement arrondi). Nous supposons qu’il y a un mécanisme aléatoire d’arrondi \mathcal{H} qui dépend de la valeur Z et paramétré par δ , donnant finalement la valeur Y , celle qui va être reportée. Wang and Heitjan (2008) propose un modèle d’arrondi adapté aux données de tabagisme. Ce modèle, trop spécifique, ne peut être utilisé dans le cas des données halieuthiques.

Le premier objectif de ce travail est donc de proposer un ou plusieurs modèles d’arrondis \mathcal{H} adapté(s) aux données.

Inférence bayésienne L’inconnue θ de la procédure d’inférence est alors formée de la collection de paramètres $(\alpha, \beta, \delta, \Sigma_\alpha, \Sigma_\varepsilon)$. Nous envisageons une approche bayésienne. En

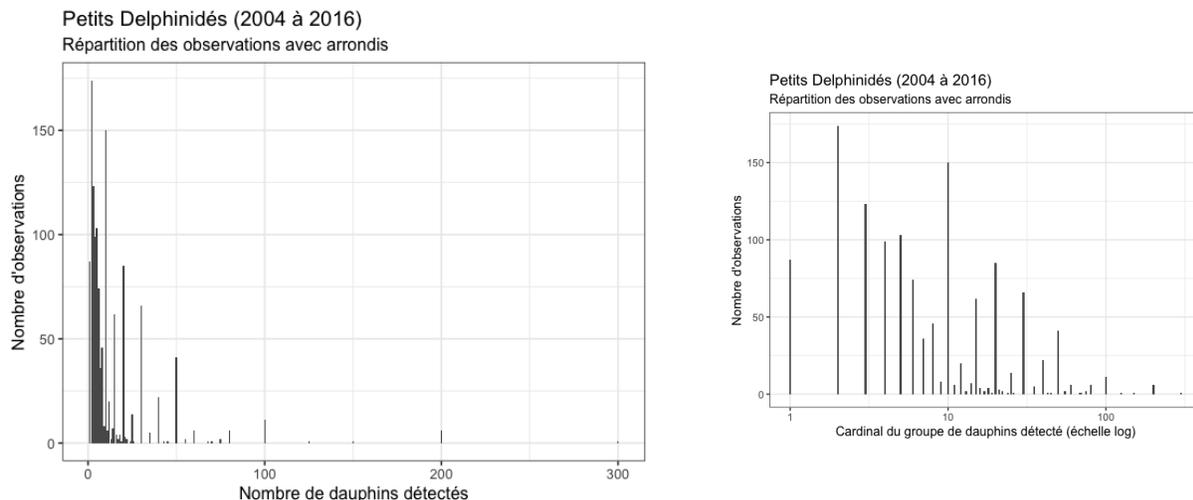


Figure 3: Répartition du nombre de dauphins observés lors des campagnes 2004 à 2016 dans le golfe de Gascogne et en Manche Ouest et manifestation des arrondis grâce à l'échelle Log.

effet, on comprend tout l'intérêt de réaliser l'inférence selon le paradigme bayésien puisque les données à elles seules n'ont guère de pouvoir explicatif du mécanisme d'arrondi. Il nous faut comprendre comment fonctionne l'observateur, décrire le caractère subjectif qu'il introduit dans les résultats avec ses a priori pour l'encoder dans \mathcal{H} . Un deuxième objectif de ce travail est donc de mettre en place une stratégie d'éllicitation de loi a priori avec dires d'experts (ici les observateurs).

Nous serons alors en mesure d'analyser les effets et les conséquences sur l'aide à la gestion des stocks qui s'appuie sur ce type d'observations avec arrondis.

3 Remerciements

Le stage de master de Daouda BA, portant sur la représentation des arrondis lors des relevés scientifiques en mer, a été financé par une bourse de l'INRA.

References

- Allen, C. M., Griffith, S. D., Shiffman, S., and Heitjan, D. F. (2017). Proximity and gravity: modeling heaped self-reports. *Statistics in medicine*, 36(20):3200–3215.
- Crawford, F. W., Weiss, R. E., and Suchard, M. A. (2015). Sex, lies and self-reported counts: Bayesian mixture models for heaping in longitudinal count data via birth-death processes. *The annals of applied statistics*, 9(2):572.

- Doray, M., Petitgas, P., Romagnan, J., Huret, M., Duhamel, E., Dupuy, C., Spitz, J., Authier, M., Sanchez, F., Berger, L., Dorémus, G., Bourriau, P., Grellier, P., and Massé, J. (2018). The PELGAS survey: Ship-based integrated monitoring of the Bay of Biscay pelagic ecosystem. *Progress in Oceanography*, 166:15–29.
- Wang, H. and Heitjan, D. F. (2008). Modeling heaping in self-reported cigarette counts. *Statistics in medicine*, 27(19):3789–3804.