

SCIENCES PARTICIPATIVES ET SUIVI DE LA BIODIVERSITÉ

Camille Coron ¹ & Clément Calenge ² & Christophe Giraud ³ & Romain Julliard ⁴

¹ *Université Paris Sud, Bâtiment 425, 15 Rue Georges Clemenceau, 91400 Orsay,
camille.coron@math.u-psud.fr*

² *Office National de la Chasse et de la Faune Sauvage, 5 rue de Saint-Thibaud, 78610
Auffargis, clement.calenge@oncfs.gouv.fr*

³ *Université Paris Sud, Bâtiment 425, 15 Rue Georges Clemenceau, 91400 Orsay,
christophe.giraud@math.u-psud.fr*

⁴ *Museum National d'Histoire Naturelle, 61 rue Buffon, 75 005 Paris,
julliard@mnhn.fr*

Résumé. Les programmes de sciences participatives se sont beaucoup développés au cours des vingt dernières années, notamment dans le domaine de l'écologie et du naturalisme. Nous disposons dans ce travail de deux jeux de données issus de tels programmes: un jeu de données dites standardisées et un jeu de données dites opportunistes. Le premier présente un nombre plus faible de données, mais qui sont associées à un protocole d'observation et de récolte très précis, tandis que le deuxième fournit un très grand nombre de données rapportées selon les souhaits et habitudes des observateurs. Grâce à ces données, nous cherchons à estimer les abondances relatives de plusieurs espèces sur différentes régions de l'espace et/ou à différents moments. Nous prouvons dans un premier temps que la combinaison de ces deux types de jeux de données permet une estimation plus précise de ces abondances, que l'utilisation des seules données standardisées. La prise en compte dans un deuxième temps des préférences respectives des observateurs et des observés pour les différents types d'habitats présents sur l'espace considéré permet des estimations plus réalistes et applicables.

Mots-clés. Sciences participatives, données opportunistes, estimation d'abondances relatives d'espèces.

Abstract. Citizen science programs have been largely developed for the last twenty years, notably in the field of ecology and naturalism. In this work, we use two datasets resulting from such programs: a standardized dataset and an opportunistic dataset. The first one presents a smaller amount of data, that are associated with a precise observation and reporting protocole, while the second one gathers a very large number of data, collected according to unknown observers wills and habits. Thanks to these data, we aim at estimating relative abundances of several species on different regions of space and/or at different moments. We first prove that combining these two types of datasets leads to a more precise estimation of these abundances, than provided by the sole standardized

dataset. Second, taking into account habitat structuration of space and respective preferences of observers and observed individuals for different habitat types allows a more realistic and more proper application of our estimations.

Keywords. Citizen sciences, opportunistic data, species relative abundances estimation.

1 Motivation et données

Nous cherchons, à l'aide de données issus de programmes de sciences participatives, à estimer et suivre les abondances (nombres d'individus) de plusieurs espèces sur une région donnée. N'importe quel citoyen volontaire peut participer à ces programmes, qui présentent un protocole plus ou moins précis. L'idée de ce travail consiste à combiner deux jeux de données: un jeu de données comportant relativement peu de données, issues d'un protocole précis (données standardisées), et un jeu de données comportant un très grand nombre de données, récoltées sans protocole (données opportunistes). Pour nos résultats numériques nous utiliserons plus spécifiquement deux jeux de données portant sur l'observation d'oiseaux communs en Aquitaine. Le jeu de données standardisées est fourni par le programme de Suivi Temporel des Oiseaux Communs (STOC) de Vigie Nature. Lorsqu'un citoyen veut y participer, on lui assigne un carré de 2 km de côté, choisi uniformément au hasard dans un rayon de 10 km autour son habitation. L'observateur doit alors disposer 10 points dans ce carré, dont la répartition doit être représentative des habitats présents dans le carré. Deux fois par an, à une période donnée de l'année, l'observateur doit se rendre dans son carré, passer 5 minutes à chaque point, et rapporter toutes les observations d'oiseaux qu'il aura effectuées (Jiguet et al (2012)). Les carrés ainsi visités entre les années 2008 et 2011 en Aquitaine, sont représentés en rouge sur la Figure 1. Le jeu de données opportunistes est fourni par l'organisme Faune Aquitaine: n'importe quel citoyen peut y faire part de n'importe quelle observation d'oiseaux en Aquitaine, sans contrainte. Les communes visitées entre les années 2008 et 2011 en Aquitaine sont représentées en noir sur la Figure 1.

2 Modèle

Nous introduisons deux modèles: un premier modèle basique dont le but est d'appréhender et de quantifier l'intérêt de la combinaison des deux jeux de données (Giraud et al (2015)), et un modèle plus sophistiqué qui prend en compte la structuration de l'espace en différents habitats et dont l'utilisation est plus intéressante et pertinente pour les écologues (Calenge et al (2015)).

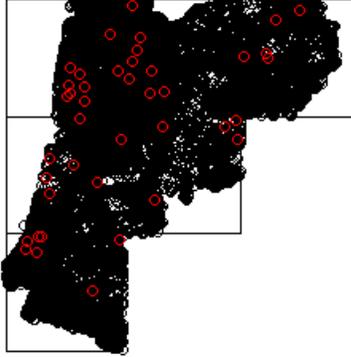


Figure 1: Position des données STOC (en rouge) et Faune Aquitaine (en noir) récoltées entre les années 2008 et 2011.

2.1 Modèle de base

On considère I espèces, indicées par $i \in \llbracket 1, I \rrbracket$, l'espace est divisé en J sites, indicés par $j \in \llbracket 1, J \rrbracket$, et nous disposons de 2 jeux de données, indicés par $k \in \{0, 1\}$ ($k = 0$ correspond au jeu de données standardisées, et $k = 1$ au jeu de données opportunistes). Notons N_{ij} le nombre d'individus de l'espèce i dans le site j . Nous cherchons à estimer N_{ij}/N_{i1} pour tous i et j . Notons X_{ijk} la somme des observations de l'espèce i sur le site j pour le jeu de données k . Nous supposons que

$$X_{ijk} \sim \text{Poisson}(N_{ij}P_{ik}E_{jk}), \quad (1)$$

où P_{ik} représente la probabilité de détection et de report d'un individu de l'espèce i pour le jeu de données k et E_{jk} représente l'intensité d'observation (typiquement le temps passé par les observateurs) sur le site j , pour le jeu de données k . Pour le jeu de données standardisées ($k = 0$), on supposera naturellement que E_{j0}/E_{10} est connu pour tout j .

2.2 Prise en compte de l'habitat

Les écologues observent que chaque espèce a des préférences différentes pour chaque type d'habitats (Bellamy et al (1998), Fithian et al (2015)), ce qui est aussi le cas des observateurs. Il est dès lors indispensable de prendre en compte et donc de modéliser ces préférences pour éviter des biais dans nos estimations. Concernant les individus observés tout d'abord, nous supposons donc que la loi de la position d'un individu de l'espèce i et du site j a pour densité:

$$\frac{S_{ih(x)}}{\sum_{h'} S_{ih'} V_{h'j}},$$

où V_{hj} est l'aire occupée par l'habitat h dans le site j .

On sait dans quelle commune (resp. dans quel carré) a été faite chaque donnée opportuniste (resp. standardisée). Pour chaque commune (resp. carré) c , on suppose qu'un observateur du jeu de données k se rend dans l'habitat h avec probabilité

$$\frac{q_{hk}V_{hc}}{\sum_{h'} q_{h'k}V_{h'c}}.$$

En particulier on supposera que $q_{h0} = 1$ pour tout h .

Enfin, la détectabilité des espèces n'est pas la même dans tous les habitats. On note $P_{ik}\alpha_h$ la probabilité de détection et de report de l'espèce i dans l'habitat h et pour le jeu de données k . Alors la somme des observations de l'espèce i dans la commune (ou le carré) c suit la loi

$$X_{ick} \sim \text{Poisson} \left(N_{ij} P_{ik} E_{ck} \sum_h \frac{q_{hk} V_{hc}}{\sum_{h'} q_{h'k} V_{h'c}} \alpha_h \frac{S_{ih} V_{hc}}{\sum_{h'} S_{ih'} V_{h'j}} \right). \quad (2)$$

Notons que l'habitat associé à chaque observation n'est pas supposé connu et n'est pas non plus réduit à la composition moyenne des habitats de la commune.

3 Résultats

Tout d'abord, on prouve pour les deux modèles que les abondances relatives N_{ij}/N_{i1} sont bien identifiables à l'aide des données $(X_{ijk})_{i,j,k}$ ou $(X_{ick})_{i,c,k}$. Le premier modèle peut-être reformulé sous la forme d'un modèle linéaire généralisé dont les paramètres sont estimés sous R.

Le deuxième modèle est implémenté en utilisant JAGS, avec des données simulées dans un premier temps, puis avec les données réelles que nous avons à notre disposition. L'utilisation des données simulées confirme notamment que la prise en compte de l'habitat permet d'éviter des biais dans nos estimations dues aux préférences distinctes des observateurs et des observés aux différents types d'habitat.

Par ailleurs, nous prouvons que l'utilisation des données opportunistes permet de diminuer la variance des estimateurs des abondances relatives, ce qui est retrouvé dans l'application numérique.

Bibliographie

- [1] Giraud, C., Calenge, C., Coron, C. et Julliard, R. (2015), Capitalising on opportunistic data for monitoring biodiversity. Prépublication.
- [2] Calenge, C., Coron, C., Giraud, C. et Julliard, R. (2015), Citizen Science and estimation of species abundances in a habitat-structured space. Prépublication.

- [3] Fithian, W., Elith, J., Hastie, T. et Keith, D.A. (2015), Bias Correction in Species Distribution Models: Pooling Survey and Collection Data for Multiple Species. *Methods for Ecology and Evolution*.
- [4] Bellamy, P. E., Brown, N. J., Enoksson, B., Firbank, L. G., Fuller, R. J., Hinsley, S. A., et Schotman, A. G. M. (1998), The influences of habitat, landscape structure and climate on local distribution patterns of the nuthatch (*Sitta europaea* L.). *Oecologia*.
- [5] Jiguet, F., Devictor, V., Julliard, R. et Couvet, D. (2012), French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica*.