

Estimation d'abondances relatives d'espèces à l'aide de données opportunistes

Camille Coron

Université Paris-Sud

26e Journées CaSciModOT

Tours

22 Juin 2017

Contexte

- Co-auteurs : Clément Calenge (ONCFS), Christophe Giraud (LMO, Orsay) et Romain Julliard (MNHN).
- But: Comparer, pour chaque espèce observée, son abondance à différents endroits/moments.

Approche

- Données de sciences participatives
- Différents niveaux de protocoles

⇒ Combiner plusieurs jeux de données

Observations d'oiseaux en Aquitaine:

- (A): données standardisées, 63 transects, 12 espèces
- (B): données opportunistes, 440496 observations dans 2086 communes, 34 espèces.
- (C): données standardisées, 48 sites, 34 espèces.

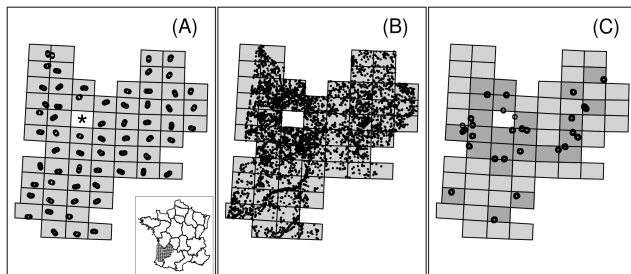


Figure: Positions des données ACT (A), LPO (B) et STOC (C).

Combinaison des données

- Espace divisé en régions indicées par $j \in \llbracket 1, J \rrbracket$.
- Plusieurs espèces indicées par $i \in \llbracket 1, I \rrbracket$: N_{ij} est l'abondance de l'espèce i dans la région j .

⇒ On cherche N_{ij}/N_{i1} pour tout i, j .

Combinaison des données

- Espace divisé en régions indicées par $j \in \llbracket 1, J \rrbracket$.
- Plusieurs espèces indicées par $i \in \llbracket 1, I \rrbracket$: N_{ij} est l'abondance de l'espèce i dans la région j .

⇒ On cherche N_{ij}/N_{i1} pour tout i, j .

- 2 jeux de données: standardisé ($k = 0$) et opportuniste ($k = 1$)

$$X_{ijk} \sim \mathcal{P}(N_{ij}O_{ijk}).$$

Combinaison des données

- Espace divisé en régions indicées par $j \in \llbracket 1, J \rrbracket$.
- Plusieurs espèces indicées par $i \in \llbracket 1, I \rrbracket$: N_{ij} est l'abondance de l'espèce i dans la région j .

⇒ On cherche N_{ij}/N_{i1} pour tout i, j .

- 2 jeux de données: standardisé ($k = 0$) et opportuniste ($k = 1$)

$$X_{ijk} \sim \mathcal{P}(N_{ij}O_{ijk}).$$

$$O_{ijk} = P_{ik}E_{jk}.$$

Identifiabilité

$$X_{ijk} \sim \mathcal{P}(N_{ij}P_{ik}E_{jk}).$$

Identifiabilité

$$X_{ijk} \sim \mathcal{P}(N_{ij}P_{ik}E_{jk}).$$

- Si E_{j0}/E_{10} est connu pour tout j .

Identifiabilité

$$X_{ijk} \sim \mathcal{P}(N_{ij}P_{ik}E_{jk}).$$

- Si E_{j0}/E_{10} est connu pour tout j .

$\Rightarrow N_{ij}E_{jk}P_{ik} = \tilde{N}_{ij}\tilde{E}_{jk}\tilde{P}_{ik}$, où \tilde{N}_{ij} , \tilde{E}_{jk} , \tilde{P}_{ik} sont identifiables, et

$$\tilde{N}_{ij}/\tilde{N}_{i1} = N_{ij}/N_{i1} \text{ pour tous } i, j.$$

Résultat théorique

$$\text{var}(\widehat{N}_{ij}) \xrightarrow{E_{j1} \rightarrow \infty} \text{var}(\widehat{N}_{ij}^0) \times \frac{P_{i0} N_{ij}}{\sum_l P_{l0} N_{lj}}. \quad (1)$$

Gain important quand:

- L'espèce i est rare
- L'espèce i est peu détectée dans les données opportunistes
- On considère un grand nombre d'espèces.

On peut calculer \widehat{N}_{ij} même si i n'est pas observée pour $k = 0$.

Données réelles

Données	Espèces dans ACT	Espèces hors ACT
ACT	0.27	—
$ACT + LPO$	0.55	0.35
ACT'	0.06	—
$ACT' + LPO$	0.54	0.28

Table: Médiane des coefficients de corrélation de Pearson entre les abondances relatives estimées et de référence.

Structuration spatiale

- Différents types d'habitats
- Différentes préférences à l'habitat
- Habitat inconnu

⇒ Habitat: variable cachée qui induit un biais

Habitat et individus observés

- Occupation du sol : habitats indicés par $h \in \llbracket 1, H \rrbracket$.
- V_{hj} : aire du domaine j couverte par h .
- Densité de la position de l'espèce i dans le domaine j :

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

Habitat et observateurs

- 2 jeux de données indicés par $k \in \{1, 2\}$.
- Cellules indicées par $c \in \llbracket 1, C \rrbracket$.
- Densité de la position d'un observateur de la cellule c :

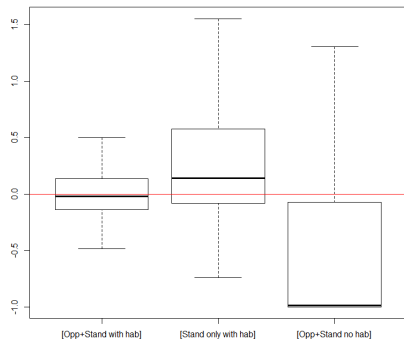
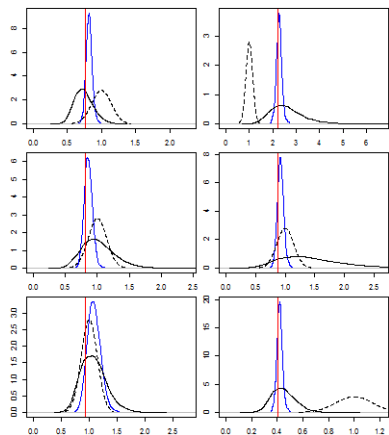
$$\frac{q_{h(x)k}}{\sum_{h'} q_{h'k} V_{h'c}}.$$

Loi des observations

Sur la cellule c et pour le jeu de données k , on observe

$$\begin{aligned}
 X_{ick} &\sim \mathcal{P} \left(\int_{\mathcal{A}_c} N_{ij} \frac{S_{ih(x)}}{\sum_{h'} S_{ih'} V_{h'j}} \times E_{ck} \frac{q_{h(x)k}}{\sum_{h'} q_{h'k} V_{h'c}} \times P_{ik} dx \right) \\
 &= \mathcal{P} \left(N_{ij} E_{ck} P_{ik} \sum_h \frac{q_{hk}}{\sum_{h'} q_{h'k} V_{h'c}} \frac{S_{ih}}{\sum_{h'} S_{ih'} V_{h'j}} V_{hc} \right).
 \end{aligned}$$

Données simulées: distributions postérieures



Distributions postérieures des abondances relatives (gauche) et distances aux paramètres à estimer (droite).

Données réelles: performance des estimations

Modèle et données	Espèces dans ACT	Espèces hors ACT
[Opp+Stand with hab]	0.49 (0.30–0.54)	0.39 (0.12–0.54)
[Stand only with hab]	0.29 (0.03–0.46)	–
[Opp+Stand no hab]	0.44 (0.32–0.68)	0.31 (0.19–0.42)

Table: Médiane des coefficients de corrélation de Pearson entre les estimations des abondances relatives et celles de référence.

Données réelles: préférences

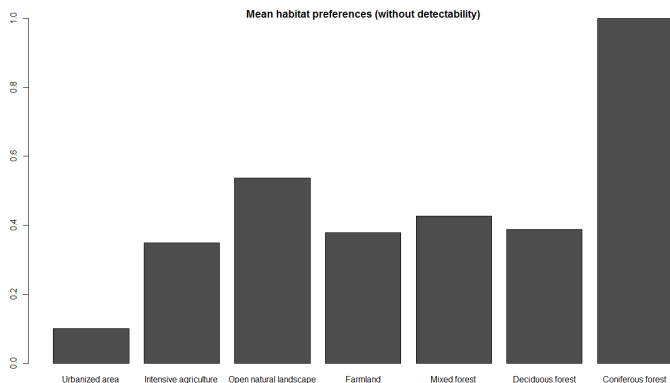
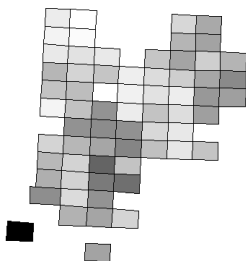


Figure: Préférences moyennes pour chaque habitat

Données réelles: Carte

Eurasian nuthatch (without habitat)



Eurasian nuthatch (with habitat)

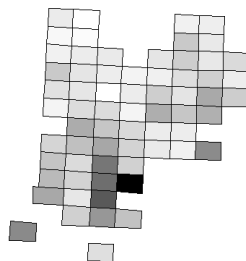


Figure: Carte d'abondance de l'épervier avec et sans prise en compte de l'habitat

Références



Giraud, C., Calenge, C., Coron, C. et Julliard, R.: Capitalizing on opportunistic data for monitoring species relative abundances. *Biometrics* 2015.



Calenge, C., Coron, C., Giraud, C., et Julliard, R.: Opportunistic data and estimation of species abundances on a habitat structured space.