Biodiversity monitoring using citizen science

Camille Coron

Université Paris-Saclay

EcoDep conference
Paris
June 22, 2022

Context

- Clément Calenge (ONCFS), Christophe Giraud (U. Paris-Saclay), Romain Julliard (MNHN).
- Aim: Compare a given species abundances at different places/times.

Approach

- 2 datasets with different levels of protocole
- Opportunistic data

$\Rightarrow$ Combine several datasets

Data

Birds observations in Aquitaine:

- (A): Standardized data, 63 transects, 12 species
- (B): Opportunistic data, 440496 observations in 2086 cities, 34 species.
- (C): Standardized data, 48 sites, 34 species.

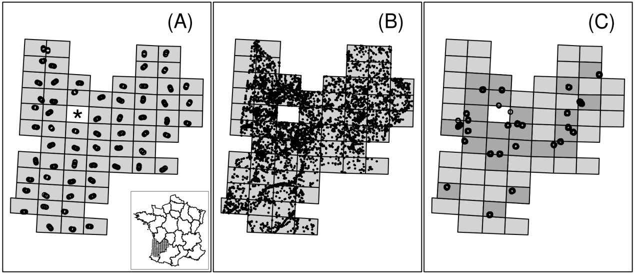


Figure: Data positions.

Aim

- Species relative abundances maps
- Interest of opportunistic data

Model

- Space divided in regions indexed by $j \in [\![1, J]\!]$.
- Species indexed by $i \in [\![1, I]\!]$.
- N_{ij} is the abundance of species i in region j.

$\Rightarrow$ We seek N_{ij}/N_{i1} for all i, j.

Model

- Space divided in regions indexed by $j \in [\![1, J]\!]$.
- Species indexed by $i \in [\![1, I]\!]$.
- N_{ij} is the abundance of species i in region j.

$\Rightarrow$ We seek N_{ij}/N_{i1} for all i, j.

- 2 datasets: standardized $(k = 0)$ and opportunistic $(k = 1)$

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

Model

- Space divided in regions indexed by $j \in [\![1, J]\!]$.
- Species indexed by $i \in [\![1, I]\!]$.
- N_{ij} is the abundance of species i in region j.

$\Rightarrow$ We seek N_{ij}/N_{i1} for all i, j.

- 2 datasets: standardized $(k = 0)$ and opportunistic $(k = 1)$

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

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

Identifiability

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

Identifiability

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

- E_{j0}/E_{10} is known for all j (standardized data).

Identifiability

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

- E_{j0}/E_{10} is known for all j (standardized data).

$\Rightarrow N_{ij}E_{jk}P_{ik} = \tilde{N}_{ij}\tilde{E}_{jk}\tilde{P}_{ik},$ where $\tilde{N}_{ij}$, $\tilde{E}_{jk}$, $\tilde{P}_{ik}$ are identifiable, and

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

Estimators

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

Generalized linear model

$\widehat{N}_{ij}$ = maximum likelihood estimator of N_{ij}/N_{i1}

$\widehat{N}_{ij}^0$ = maximum likelihood estimator of N_{ij}/N_{i1} when using only standardized data.

Theoretical result

$$\mathbb{E}(\widehat{N}_{ij}) \overset{E_{j1}\to\infty}{\longrightarrow} \tilde{N}_{ij}$$

$$\mathrm{var}(\widehat{N}_{ij}) \overset{E_{j1}\to\infty}{\longrightarrow} \mathrm{var}(\widehat{N}^0_{ij}) \times \frac{P_{i0}N_{ij}}{\sum_l P_{l0}N_{lj}} \,.$$

Interesting when

- Species i is rare
- Species i is hardly detected in standardized data
- Considering a large number of species

$\hat{N}_{ij}$ is defined even if species i is not observed in standardized data

Real data

Data	Species in ACT	Species not in ACT
ACT	0.27	—
$ACT + LPO$	0.55	0.35
ACT'	0.06	—
$ACT' + LPO$	0.54	0.28

Table: Median of Pearson correlation coefficients between estimated and reference relative abundances.

Spatial structure

- Different types of habitats
- Different habitat preferences
- Unknown habitat

$\Rightarrow$ Habitat: hidden variable that induces bias

Habitat and species repartition

- Ground occupation : habitats indexed by $h \in [\![1, H]\!]$.
- V_{hj}: area of domain j covered by h.
- Density of species i in domain j:

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

Habitat and observers repartition

- 2 datasets indexed by $k \in \{1, 2\}$.
- Cells indexed by $c \in [\![1, C]\!]$.
- Density of observers in cell c:

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

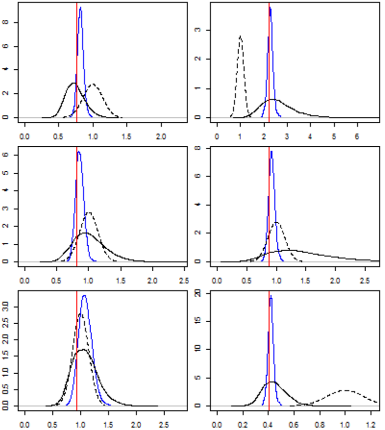
Law of observations

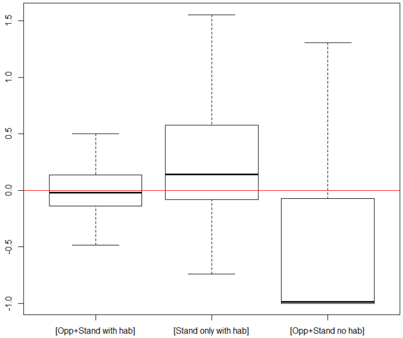
In cell c and for the dataset k,

$$X_{ick} \sim \mathcal{P}\left(\int_{\mathcal{A}_c} N_{ij} \frac{S_{ih(x)}}{\sum_{h'} S_{ih'} V_{h'j}} \times E_{ck} \frac{q_{h(x)k}}{\sum_{h'} q_{h'k} V_{h'c}} \times P_{ik}\, dx\right)$$

$$= \mathcal{P}\left(N_{ij} E_{ck} P_{ik} \sum_{h} \frac{q_{hk}}{\sum_{h'} q_{h'k} V_{h'c}} \frac{S_{ih}}{\sum_{h'} S_{ih'} V_{h'j}} V_{hc}\right).$$

Simulated data: posterior distributions





Posterior distributions of relative abundances (left) and distances to real values of parameters (right).

Real data: estimation performance

Model and data	Species in ACT	Species not in 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: Median of Pearson correlation coefficients between estimated and reference relative abundances estimations.

Real data: preferences

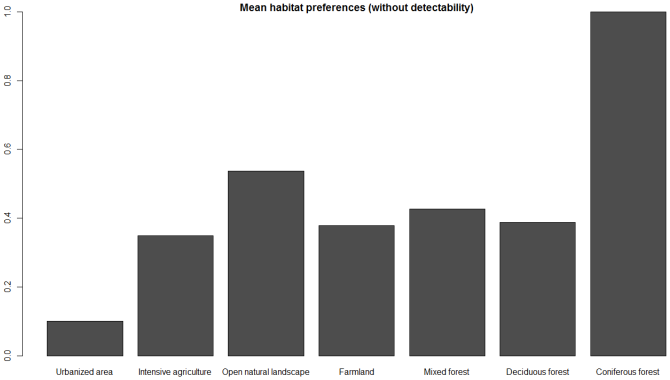


Figure: Mean preferences for each habitat

Real data: abundance maps

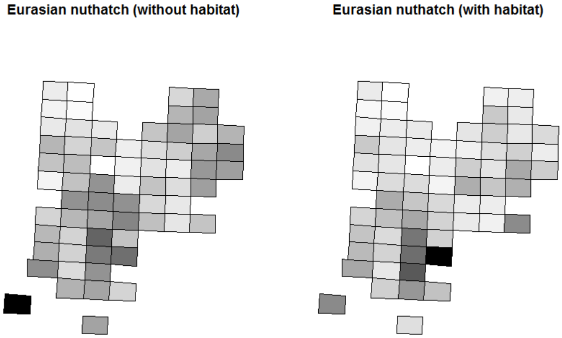


Figure: Relative abundance maps of Eurasian nuthatch with and without habitat

Air quality monitoring

Work in progress with Jean-Michel Poggi and Benjamin Auder.

Context: Air quality monitoring

Aim: Establish concentration maps of some particles : NO2, PM10.

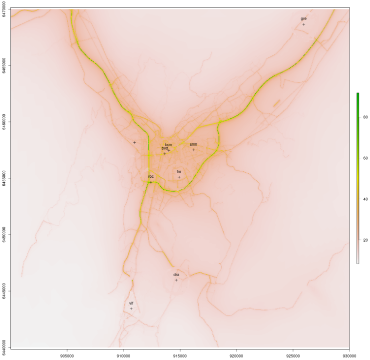


Figure: Example of concentration map (NO2 (SIRANE), Grenoble)

Data and questions

Data:

- Deterministic physico-chemical model
- Measures : 2 types of devices
  - Fixed stations (precise and expensive)
  - Micro-sensors (unprecise and cheap)

Approach and questions

Approach:

- Start with physico-chemical model
- Model its bias
- Estimate its bias using measures

Questions:

- Can micro-sensors help improve concentration maps ?
- How many micro-sensors ?
- What spatial distribution for fixed stations and micro-sensors ?

References

- 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.