Modeling breeding bird surveys

Adélie Erard (aerard@math.cnrs.fr) under the supervision of
Raphaël Lachièze-Rey and Romain Lorrillière

Université Paris Cité, MAP5 & Muséum national d'Histoire naturelle, CESCO

Biological context

Breeding Bird Surveys (BBS) are long-term, large-scale, international avian monitoring programs designed to track the status and trends of bird populations. Key features of the BBS include: standardized protocol, volunteer participation, geographical coverage, data collection and trend analysis. In our case we will look at the French protocol named STOC [3] for Temporal Monitoring of Common Birds. We aim to determine when the population will grow or not at a local scale.

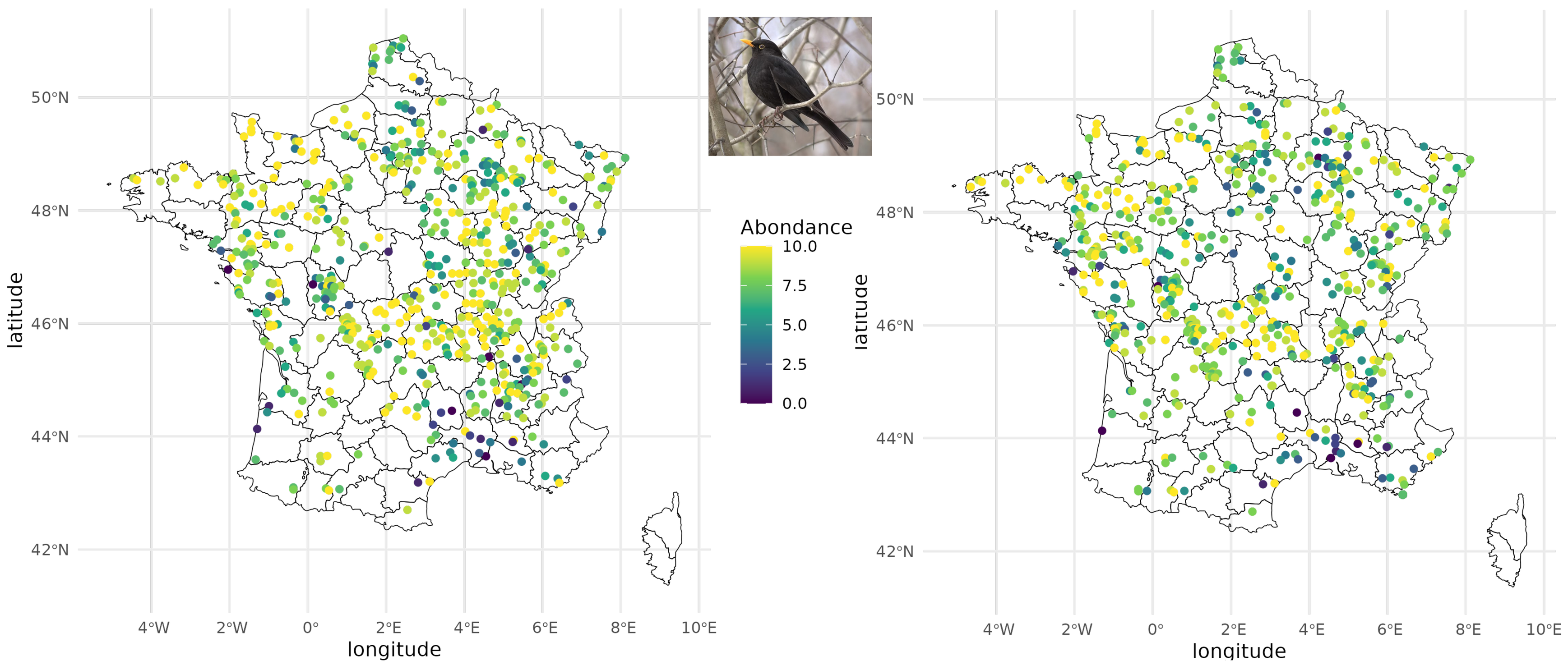
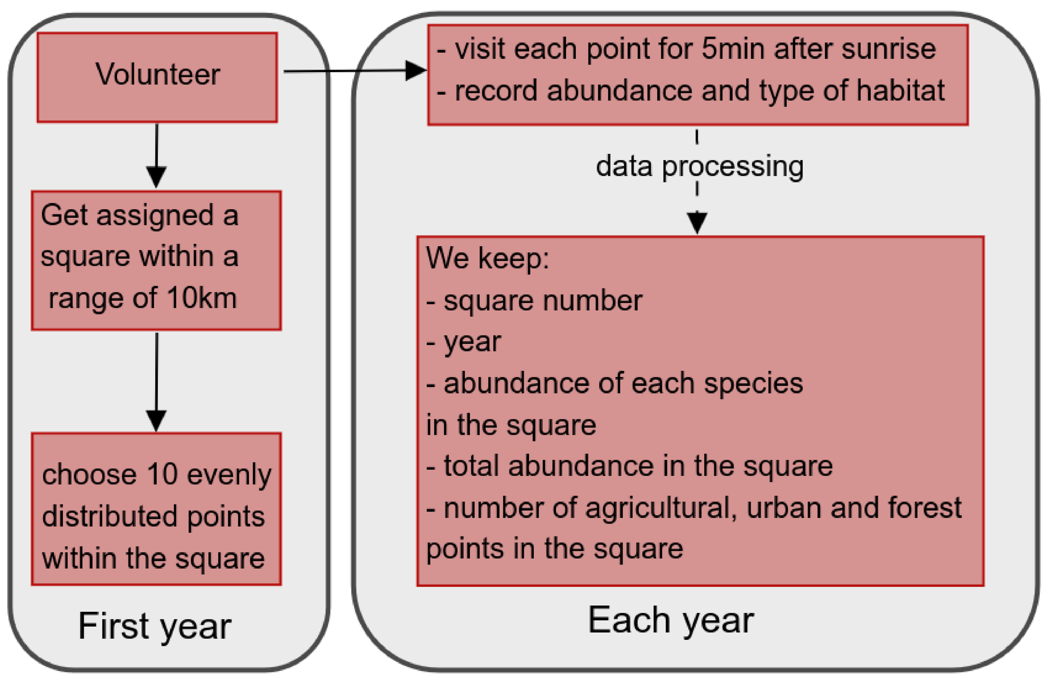


Figure 1: Observed blackbirds (*Turdus merula*) in 2007 (left) and 2008 (right), in the STOC program.

STOC protocol – bird data



Environmental data

- Climate data from Wordclim: minimum and maximum temperature in the year and maximum rainfall amount of wettest month
- Pesticide uses: pesticide sales in each postcode each year (from BNVD database) and cluster of use from [2]
- CORINE landcover: surface of each type of soil per postcode

Which variable will induce changes in abundance?

Method

Our goal is to predict the trend in abundance variation for the following year. We thus add a column for next year trend: ratio of next year's abundance to the current year's. Each row is finally classified into three categories $(-1, 0, 1)$ respectively classified by:

- Abundance decreases by at least 10% the next year (-1).
- Abundance increases by at least 10% the next year (1).
- Abundance change is less than 10% (0).

We then run random forest algorithms (RF) from scikit learn package and keep the feature importance to see wich variables induce changes in abundance. Note that we did not run the algorithm on all the dataset but only on some species (most occurent species) or group of species (considering type of habitat or diet).

Results

Since our objective is to find the variables that most impact changes in abundance from one year to the next, we do not need to consider time separately and combine all years without distinction.

For all species groups, abundance is the most important feature, likely indicating some spatial structure governing changes in abundance. The other important variables are environmental, which aligns with ecological literature.

For all the groups of species, the abundance is the most important feature, it probably implies that there is some spatial structure governing changes in abundance. The other impotant variables are environmental variables which is comforting regarding the ecological litterature.

Group	Number of observations	Sucess of RF
Blackbird	15500	75%
Eurasian blackcap	15540	76%
Agricultural species	123338	67%
Insectivorous species	492333	67%

Table 1: Size of some groups we tried random forest on and results of the algorithm

A stochastic model considering local effects

In this part our goal is to construct a model respresenting birds demography and to propose a local estimator of the population's growth: births − deaths in a zone between two times.

For $t \in \mathbb{N}$ a given year and $x \in \mathbb{R}^2$. Let $\Theta_t(x) \in \mathbb{R}^q$ be the field of environmental parameters at time t (climate, pesticide...) around x. Let $\mathscr{N}$ be the space of configurations of points of $\mathbb{R}^2$.

We represent birds at time t by a point process $\mathscr{P}$ of unknown intensity depending of the field Θ, with possibly spatial dependence, such as for a Cox process or a Gibbs process, it takes values in $\mathscr{N}$. This process is written as:

$$\mathscr{P} := \sum_{x \in \mathscr{P}} \delta_x$$

As we do not observe all birds, we need to thin the process $\mathscr{P}$. Let Y_t be an homogeneous Poisson birth and death process with no memory representing observers. The observed area is a booleen process $\mathscr{O}_t$ defined as:

$$\mathscr{O}_t = \bigcup_{y \in Y_t \cap Y_{t+1}} C_y$$

where C_y is a square centered in y. Note that we only keep observations in $Y_t \cap Y_{t+1}$ to observe changes in the population.

For all $x \in \mathscr{P}$ pose:

$$U_x = (\mathbb{1}_{x \in \mathscr{O}_t}, s_x, \Theta(\cdot - x)) \in \mathscr{M} := \{0,1\} \times \mathbb{Z} \times E$$

where E is a suitable space of functions from $\mathbb{R}^2$ to $\mathbb{R}^q$. The roles of U_x components are respectively to:

- $(\mathbb{1}_{x \in \mathscr{O}_t})$ Indicates wether x is in an observed square or not.
- (s_x) Assign the population variation in the square of x: $s_x = 0$ if x is not observed, otherwise s_y where y is the observer observing x with s_y the number of births minus the number of deaths in the square centered in y. We assume that s_x is ruled by a "birth and death" rate, i.e. $\mathbb{E}[s_x] = r(\mathscr{P}(\cdot - x), \Theta(\cdot - x))$, where we wish to estimate the function r.
- $(\Theta(\cdot - x))$ gives informations about the environment around x.

Let $\overline{\mathscr{P}} := (x, U_x)_{x \in \mathscr{P}}$ be the marked process.

Second order and asymptotic normality of estimators based on score functions

We consider observables under the general form:

$$H_n^{\xi}(\overline{\mathscr{P}}) := \sum_{(x,U_x) \in \overline{\mathscr{P}} \cap W_n} \xi(x, U_x, \mathscr{P}_n)$$

where $\mathscr{P}_n := \mathscr{P} \cap W_n$ with W_n a window of volume n and $\xi : (\mathbb{R}^2 \times \mathscr{M}) \times \mathscr{N} \to \mathbb{R}$ is a score function.

Now suppose we want to predict the population evolution in a given context $(\mathscr{X}, \theta) \in E \times \mathscr{N}$, i.e. we wish to estimate $r(\mathscr{X}, \theta)$ the trend of variation in abundance. We define the estimator:

$$\hat{s}_n((\mathscr{X}, \theta)) := \frac{1}{\sum_{(x,U_x) \in \overline{\mathscr{P}} \cap W_n} s_x} \sum_{(x,U_x) \in \overline{\mathscr{P}} \cap W_n} k((C_x \cap \mathscr{P}_n, \Theta), (\mathscr{X}, \theta)) s_x$$

based on some similarity distance $k : (\mathscr{N} \times E)^2 \to \mathbb{R}$ that measure the similarity in abundance and parameters between observations $(C_x \cap \overline{\mathscr{P}}, \Theta)$ (C_x is the square centered in x) and $(\mathscr{X}, \theta)$. We wish to use the theory of geometric stabilisation to establish a Law of large numbers, variance asymptotics and Central Limit Theorem under hypotheses of weak dependency on the point process and the marks spatial dependency, following the method of [1].

[1] B. Błaszczyszyn, D. Yogeshwaran, and J. E. Yukich. Limit Theory for Geometric Statistics of Point Processes Having Fast Decay of Correlations. *The Annals of Probability*, 47(2):835–895, 2019. Publisher: Institute of Mathematical Statistics.

[2] Milena Cairo, Anne-Christine Monnet, Stéphane Robin, Emmanuelle Porcher, and Colin Fontaine. Identifying pesticide mixtures at country-wide scale. March 2023.

[3] Benoît Fontaine, Caroline Moussy, Jules Chiffard Carricaburu, Jérémy Dupuis, Emmanuelle Corolleur, Lucie Schmaltz, Romain Lorrillière, Grégoire Loïs, and Clémence Gaudard. *Suivi des oiseaux communs en France 1989-2019 : 30 ans de suivis participatifs*. 2021.