A method for estimating population growth at a local scale.

An application using French breeding bird surveys data.

Adélie Erard [1,2], Raphaël Lachièze-Rey[1] and Romain Lorrillière[2]

[1]MAP5, Université Paris Cité

[2]CESCO, Muséum national d'Histoire naturelle

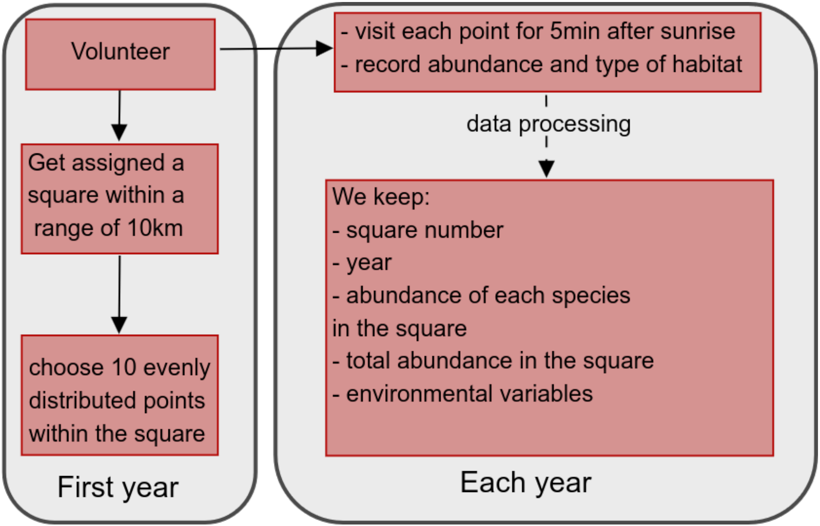
March 19, 2025 – Gothenburg statistics seminar

Breeding Bird Surveys

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: standardized protocol, geographical and temporal coverage.

French BBS program (STOC)



Goals

Square	Year	Abundance	Environmental variables	**Abundance next year**
10295	2003	5	🦜🦜 🪿 🪿 🌡️☀️🌲	2
11158	2006	1	🪿 🪿 🌧️🌡️🌽	4
20204	2015	6	🐦🐦🐦 ☀️ ❄️🌽	7
30363	2019	8	🐦🐦🪿🦜 🌧️❄️🌽	9
950294	2024	3	🪿 🪿 🐦 🌧️🌡️🌲	?

1. Give a method to estimate the growth of a population at a local scale (work in progress).
2. Find which environmental variables induce changes in abundance (work with Ottmar).

Base layers of the model

Birds at time t are represented as a point process:

$$\mathcal{P}_t := \sum_{x \in \mathcal{P}_t} \delta_x$$

Observers at time t are represented as a point process:

$$O_t := \sum_{y \in O_t} \delta_y$$

Environmental variables at time t are represented by a random field:

$$\Theta_t(\cdot) : \mathbb{R}^2 \to \mathbb{R}^q$$

Observed zone

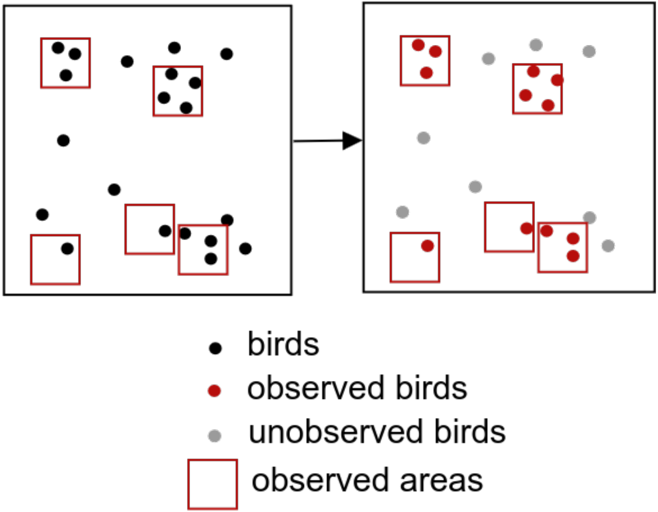
Let (O_t) be a spatial birth and death process with birth intensity $p_t(y)$ and death intensity γ

Let C_y be the square centered in y with 2km side for all $y \in X$

Then the observer process $(\mathcal{O}_t)$ is defined by the following closed random set:

$$\mathcal{O}_t = \bigcup_{y \in O_t \cap O_{t+1}} C_y$$

Representation of the model



Marked process

For all $x \in \mathcal{P}_t$ pose:

$$U_x = (\mathbb{1}_{x \in \mathcal{O}_t}, \Theta(\cdot - x))$$

Future abundance/variation of abundance around x:

$$\beta_x := \beta(U_x, B(x, R) \cap \mathcal{P})$$

Marked process:

$$\overline{\mathcal{P}}_t := \sum_{x \in \mathcal{P}_t} \delta_{(x, U_x, \beta_x)}$$

(Non) stationarity assumptions

- No temporal stationarity nor equilibrium
- Not in an high density limit
- We assume our process to be spatially ergodic

We consider the following point process of observed birds:

$$\overline{\mathcal{P}} = \bigcup_t \overline{\mathcal{P}}_t \cap \mathcal{O}_t$$

Estimator

Let $\mathcal{C}$ be a deterministic configuration of marked points (x, U_x), set:

$$\hat{N}_n^{\mathcal{C}}(\overline{\mathcal{P}}) := \frac{1}{\sum_{y \in \mathcal{P}_n} k(\overline{C_y}, \mathcal{C})} \sum_{x \in \mathcal{P}_n} k(\overline{C_x}, \mathcal{C}) \beta_x$$

where k is a similarity function, and $\mathcal{P}_n := \mathcal{P} \cap [-n/2, n/2]^2$ and $\overline{C_x}$ contains information on the process in the square centered at x

Similarity function

Let x, y be two lines of the database and p the number of variables. We set:

$$k(x,y) := \frac{1}{p}\sum_{i=1}^{p} s_i(x,y),$$

$$s_i(x,y) := 1 - \frac{|x_i - y_i|}{R_i} \text{ for quantitative variables,}$$

$$s_i(x,y) := \mathbb{1}_{x_i = y_i} \text{ for categorical variables.}$$

Convergence results

Theorem (5.2 of Błaszczyszyn, Yogeshwaran, and Yukich 2025+)

Let $\overline{\mathcal{P}}$ be a marked point process of $\overline{\mathcal{N}}$ having exponential mixing correlations. Let $\xi : \mathbb{R}^d \times \mathcal{M} \times \overline{\mathcal{N}} \to \mathbb{R}$ be a score function that is:

- *fast BL-localizing on finite windows of $\overline{\mathcal{P}}$;*
- *verifying the p moment condition on finite windows of $\overline{\mathcal{P}}$ for all $p \geq 1$.*

If $\mathrm{Var}\left(\mu_n^\xi\right) = \Omega(n^\nu)$ *for* $\nu > 0$. *Then, as* $n \to \infty$:

$$\left(\mathrm{Var}\left(\mu_n^\xi\right)\right)^{-1/2}\left(\mu_n^\xi - \mathbb{E}[\mu_n^\xi]\right) \overset{d}{\Longrightarrow} Z$$

with Z a standard normal random variable and $\mu_n^\xi = \sum_{x \in \mathcal{P}_n} \xi((x, U_x), \overline{\mathcal{P}})$

In our case

If $\overline{\mathcal{P}}$ is a log Gaussian Cox process then it has exponential mixing correlations.

Let $\xi((x, U_x, \beta_x), \overline{\mathcal{P}}) = \frac{1}{\sum_{y \in \mathcal{P}_n} k(\overline{C_y}, \mathcal{C})} k(\overline{C_x}, \mathcal{C})\beta_x$. It verifies the stabilization hypothesis and the moment hypothesis.

Proposition

$$\left(\mathrm{Var}\left(\hat{N}_n^{\mathcal{C}}(\overline{\mathcal{P}})\right)\right)^{-1/2}\left(\hat{N}_n^{\mathcal{C}}(\overline{\mathcal{P}}) - \mathbb{E}[\hat{N}_n^{\mathcal{C}}(\overline{\mathcal{P}})]\right) \xrightarrow{d} Z$$

Method for data analysis

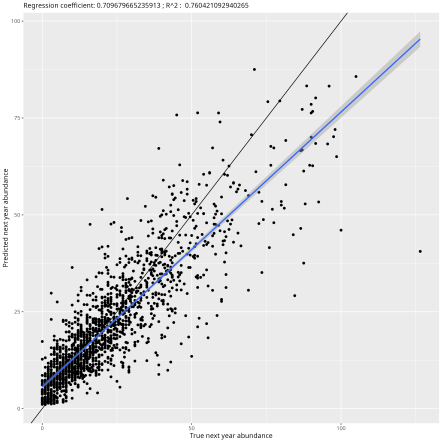
1. Calculate the similarity between lines based on all variables except the target;
2. Split the database into train and test;
3. For each line in test set, calculate $\hat{N}_n^{\mathcal{C}}(\overline{\mathcal{P}})$ with $\overline{\mathcal{P}}$ as the train set.

What's in the database?

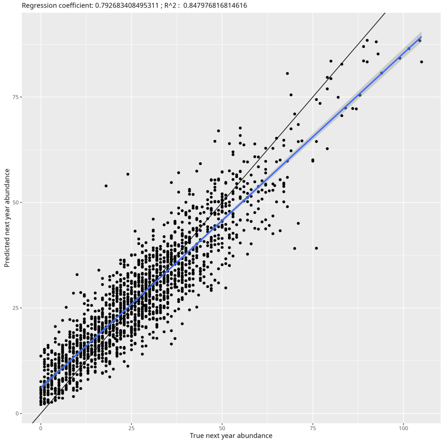
Square	Year	Abundance	Environmental variables	**Abundance next year**
10295	2003	5	🦜🦜 🪿 🪿 🌡️☀️ 🌲	2
11158	2006	1	🪿 🪿 🌧️🌡️ 🌽	4
20204	2015	6	🐦 🐦 🐦 ☀️ ❄️ 🌽	7
30363	2019	8	🐦 🐦 🪿🦜 🌧️ ❄️ 🌽	9
950294	2024	3	🪿 🪿 🐦 🌧️🌡️🌲	?

The abundance of the square 950294 in 2025 is estimated taking the mean over the most similar rows, e.g first and second row, and thus should be 3.

Results with real data



Agricole species



Forest species

Perspectives

- Find other classes of process $\mathcal{P}$ and function ξ verifying the convergence theorem (e.g Gibbs processes, functional giving other information than future abundance...).

- Extend the quantity of environmental variables we use. For example we are trying to add a pressure variable that we construct by a krigging procedure.

- Find what induce changes in abundance.

Model

- $(\Theta_t)_t$ represent the environment.
- $(\mathcal{P}_t)_t$ represent birds with $\mathcal{P}_{t+1} \sim \text{Cox}\left(f(\mathcal{P}_t, \Theta_t)\right)$.
- $(O_t)_t$ is a spatial birth and death process for the observers. At year t there are N_t points.

Each year t we observe the marked process $Y_t = \{O_i^t, M_i^t\}_{i=1}^{N_t}$, where M_i^t is the realization of $\mathcal{P}_t$ around O_i^t.

Spatiotemporal prediction for log-Gaussian Cox processes, Brix and Diggle 2001

- $\mathcal{P}_t$ is a log-Gaussian Cox process with intensity $\Lambda_t = \exp G_t$, G_t a Gaussian field.
- Observation of $\mathcal{P}_t$ in a grid where **all cells are observed**.
- G_t is modelled with a spatial Ornstein Uhlenbeck process that is **stationary in time**.
- Parameter estimation and intensity prediction.

Thank you for your attention!

- Błaszczyszyn, B., D. Yogeshwaran, and J. E. Yukich (2025+). "Limit theory for statistics of Lipschitz-localized stochastic processes in spatial random models". In: *in preparation*. +.
- Brix, Anders and Peter J. Diggle (2001). "Spatiotemporal prediction for log-Gaussian Cox processes". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.4, pp. 823–841. ISSN: 1467-9868. DOI: 10.1111/1467-9868.00315.