Modeling Fine-Scale Abundance Dynamics:

A Dual Frequentist and Bayesian Approach Applied to Common Birds

Adélie Erard, Ottmar Cronie, Raphaël Lachièze-Rey and Romain Lorrillière

Aussois — June 17, 2025

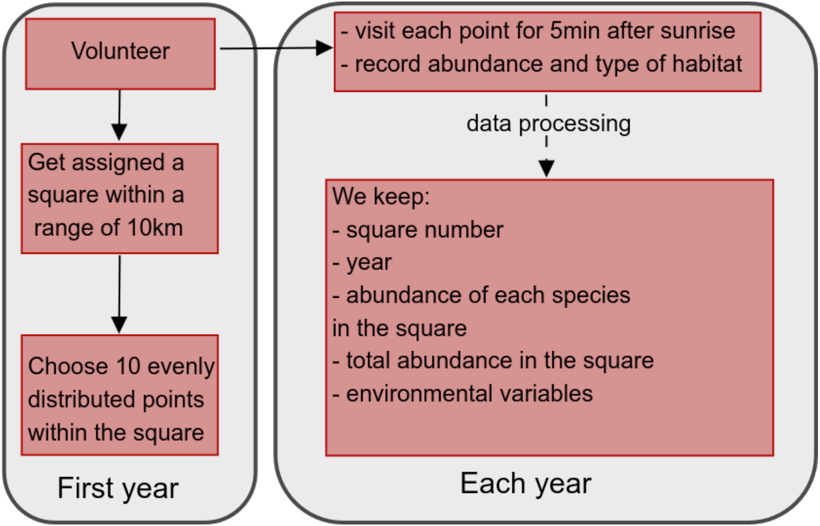
LABORATOIRE MAP5
CESCO Centre d'Écologie et des Sciences de la Conservation
Modélisation Mathématique et Biodiversité

Breeding Bird Surveys

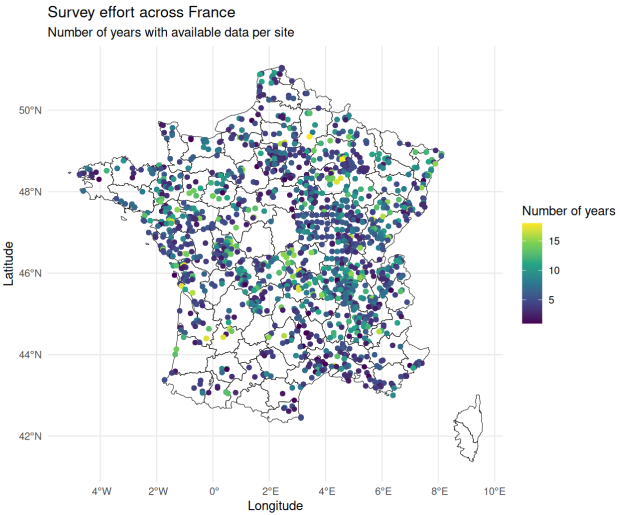
Breeding Bird Surveys (BBS) are long-term, large-scale, international avian monitoring programs designed to track the status and trends of bird populations.

Key features: standardized protocol, geographical and temporal coverage.

French BBS program (STOC)



What's in the data?



Goals

1. Give a method to estimate future abundance of birds at a local scale.
2. Find which environmental variables induce changes in abundance.

Base layers of the model

Birds at time t are represented as a point process:

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

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:

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

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

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 $y \in X$

Observed zone

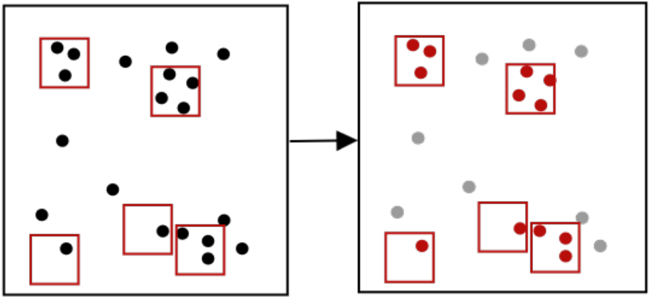
Let (O_t) be a spatial birth and death process with birth intensity $p_t(y)$ and death intensity γ

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



- birds $\mathcal{P}_t$, for example a Cox process
- observed zones $\mathcal{O}_t$, centers of squares follow a spatial birth and death model
- observed birds
- non observed birds

Marked process

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

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

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:

$$N_x := N(U_x, C_y(x) \cap \mathcal{P})$$

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:

$$N_x := N(U_x, C_y(x) \cap \mathcal{P})$$

Marked process:

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

(Non) stationarity assumptions

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

(Non) stationarity assumptions

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

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:

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

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

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* $n \to \infty$*:*

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

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 $\xi\left((x, U_x, N_x), \overline{\mathcal{P}}\right) = \frac{1}{\sum\limits_{y \in \mathcal{P}_n} k\left(\overline{C_y}, \mathcal{C}\right)} k\left(\overline{C_x}, \mathcal{C}\right) N_x$. It verifies the stabilization and the moment hypothesis.

In our case

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

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

Proposition

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

What's in the database?

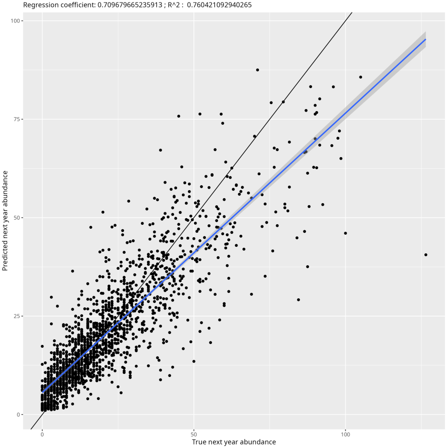
Square	Year	Abundance	Environmental variables	**Abundance next year**
10295	2003	5	🦜🦜 🦢 🦢 🌡️☀️🌲	2
11158	2006	1	🦢 🦢 🌧️🌡️ 🌽	4
20204	2015	6	🐦🐦🐦 ☀️ ❄️🌽	7
30363	2019	8	🐦🐦🦢🦜 🌧️❄️🌽	9
950294	2024	3	🦢 🦢 🐦 🌧️🌡️🌲	?

What's in the database?

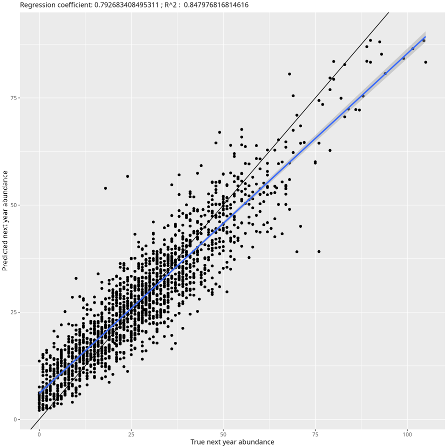
Square	Year	Abundance	Environmental variables	**Abundance next year**
10295	2003	5	🦜🦜 🦢 🦢 🌡️☀️ 🌲	2
11158	2006	1	🦢 🦢 🌧️🌡️ 🌽	4
20204	2015	6	🐦 🐦 🐦 ☀️ ❄️ 🌽	7
30363	2019	8	🐦 🐦 🦢 🦜 🌧️ ❄️ 🌽	9
950294	2024	3	🦢 🦢 🐦 🌧️🌡️ 🌲	?

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



Agricultural species



Forest species

What next on this project?

- Show the mixing property for other processes (Gibbs, Hawkes,. . .)
- Try other similarity functions
- Add new environmental variables

Some ideas on what can change bird's abundance

- Land use,
- climate,
- agricultural practices,
- biodiversity,
- . . .

Model

- We model bird counts using a **negative binomial regression model** to account for overdispersion.

Model

- We model bird counts using a **negative binomial regression model** to account for overdispersion.
- The expected count $\lambda(s,t)$ is modeled as:

$$\log \lambda(s,t) = \beta_0 + \sum_k \beta_k X_k(s,t) + \sum_{i,j} \beta_{i,j} X_i(s,t) X_j(s,t) + w_{s,t}$$

where:

  - $X_k(s,t)$: environmental covariates (climate, land cover, etc.)
  - $w_{s,t}$: spatio-temporal randomness

Bayesian inference with INLA Rue et al. 2017

- Deterministic approximations of the posterior law (Integrated Nested Laplace Approximation):
  - One for the latent field (β_i)
  - One for hyperparameters (spatial and time effects)
- Very fast but only for Latent Gaussian Model

Variables selection procedure

1. Test all possible models with interactions

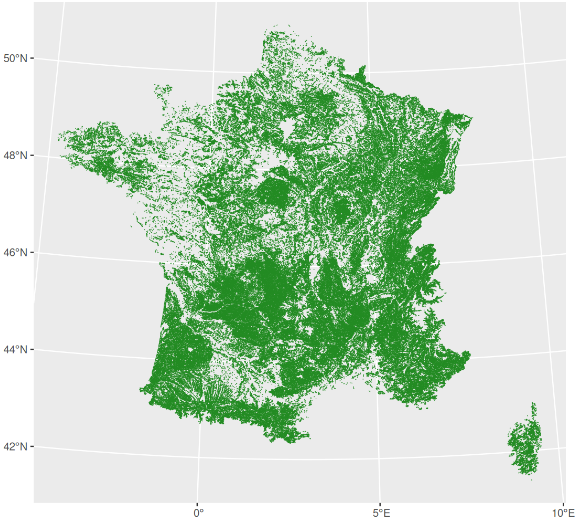
2. Keep the best WAIC

   - If all variables are significant we stop here
   - Otherwise we try to remove interactions one by one and keep the best WAIC

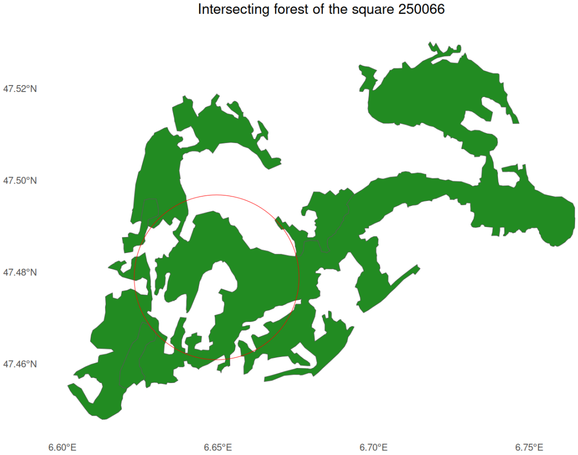
Results — Forest species — Bourgogne

Fixed effect	**Mean**	**2.5% Q**	**97.5% Q**
Intercept	1.686	1.508	1.859
Other birds in square	**0.067**	0.031	0.104
Artificial area	**-0.220**	-0.350	-0.089
Last year minimum temperature	0.011	-0.023	0.045
Other birds × artificial area	0.013	-0.021	0.046
Other birds × last year Tmin	**0.037**	0.017	0.056
Artificial area × last year Tmin	-0.021	-0.045	0.003

Adding land sharing/sparing covariates



Thank you for your attention!


Intersecting forest of the square 250066
47.52°N
47.50°N
47.48°N
47.46°N
6.60°E
6.65°E
6.70°E
6.75°E

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*. +.

Rue, Håvard et al. (Mar. 2017). "Bayesian Computing with INLA: A Review". fr. In: *Annual Review of Statistics and Its Application* 4.Volume 4, 2017. Publisher: Annual Reviews, pp. 395–421. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-060116-054045.