Asymptotic properties of estimators for partially observed dependent spatial processes in a random environment.

Adélie Erard, Raphaël Lachièze-Rey and Romain Lorrillière

Statistiques au sommet de Rochebrune – March, 2026

Breeding Bird Surveys

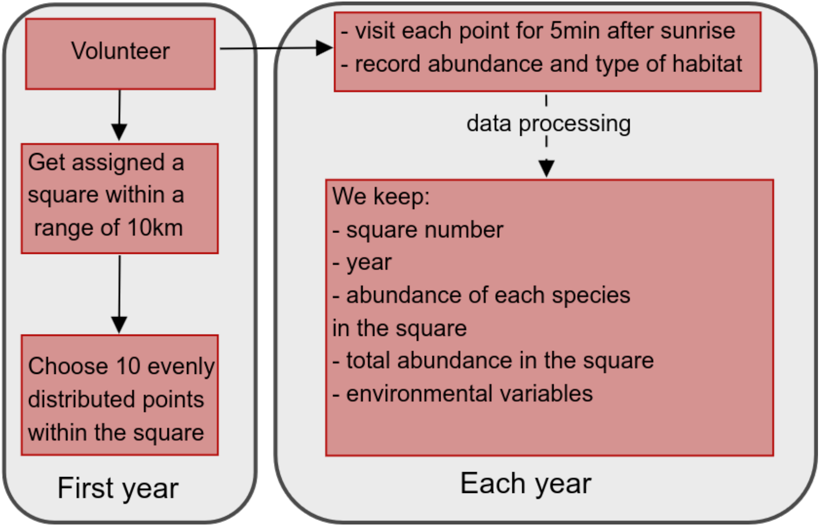
Meadow pipit (*Anthus pratensis*)

From Charles J. Sharp

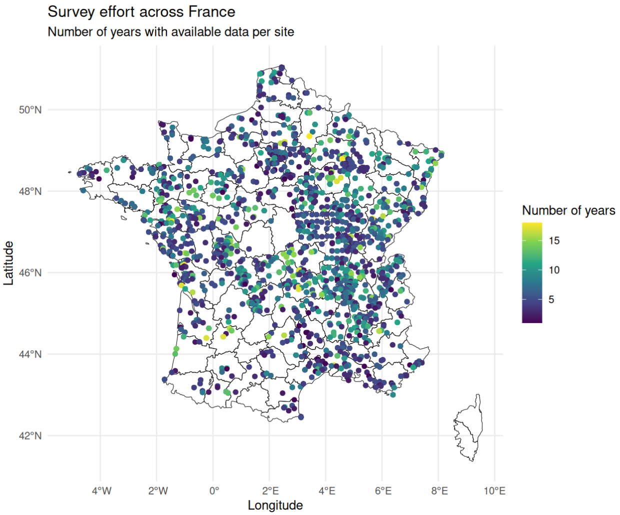
Breeding Bird Surveys (BBS) are long-term, large-scale, international avian monitoring programs designed to track the status and trends of bird populations.

Key features: standardized protocol, geographical and temporal coverage.

French BBS program (STOC)



What's in the data?



Environmental variables

For each observed point, we retrieve:

- Climate variables during spring (minimum and maximum temperature, total rain);
- Land uses in the square (% of agricultural, forest and urban land).

Goal

Give a method to **estimate future abundance** of birds at a **local scale**.

Birth and death model

Individuals at a time are represented as a point process $\mathcal{P}$:

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

Birth and death model

Individuals at a time are represented as a point process $\mathcal{P}$:

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

Field of covariates: $\theta := (\theta(x), x \in \mathbb{R}^2)$

Birth and death model

Individuals at a time are represented as a point process $\mathcal{P}$:

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

Field of covariates: $\theta := (\theta(x), x \in \mathbb{R}^2)$

The transition to a new state $\mathcal{P}'$ is governed, by:

- birth probability: $b(\tau_x \theta, \tau_x \mathcal{P})$
- death probability: $d(\tau_x \theta, \tau_x \mathcal{P})$

with τ_x a shift operator.

Effort rate

The observers form a random set in $\mathbb{R}^2$, $E = (E(x), x \in \mathbb{R}^2)$, that is independent of $\mathcal{P}$ and θ.

Effort rate

The observers form a random set in $\mathbb{R}^2$, $E = (E(x), x \in \mathbb{R}^2)$, that is independent of $\mathcal{P}$ and θ.

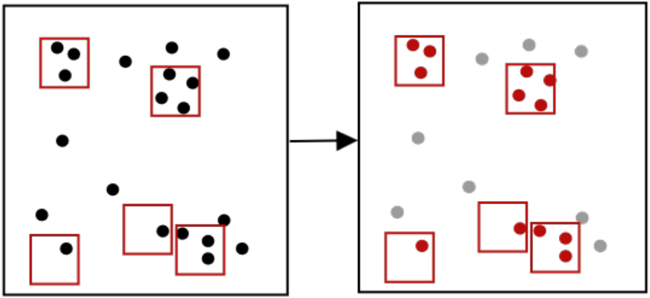
Example

$E_0 = \bigcup_e B(e, \rho)$ initial observed zone with e the location of the observers and ρ the (random) range of observation

To go to the next state of observed zone:

- each point is removed with constant rate;
- new points arrive according to a homogeneous Poisson process

Representation of the model



- birth and death process
- observed zones
- observed individuals
- non observed individuals

(Non) stationarity assumptions

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

What do we want to predict?

Given a population $\mathcal{P}_0$ and covariates $\theta = (\theta(x), x \in \mathbb{R}^2)$.

What will be the new state $\mathcal{P}_1$ of the population around a point x at the next observation time?

What do we want to predict?

Given a population $\mathcal{P}_0$ and covariates $\theta = (\theta(x), x \in \mathbb{R}^2)$.

What will be the new state $\mathcal{P}_1$ of the population around a point x at the next observation time?

Two possible observables:

- Next year abundance around x: $N_1(x) = \#\{\mathcal{P}_1 \cap B(x, \rho)\}$
- Variation of abundance between the two states at x: $\Delta(x) = N_1(x) - N_0(x)$

Construction of $\hat{N}$

Let $(\theta', \mathcal{C})$ be a deterministic configuration of interest - somewhere where we have data at the initial state and want the abundance at the next state.

How do we construct our estimator?

Construction of $\hat{N}$

Let $(\theta', \mathcal{C})$ be a deterministic configuration of interest - somewhere where we have data at the initial state and want the abundance at the next state.

How do we construct our estimator?

1. Construct a similarity function to compare two contexts.
2. Calculate similarity between $(\theta', \mathcal{C})$ and know contexts (from the data).
3. Do a weighted mean to have an estimation of next year abundance (or variation of abundance).

Estimator

Let $(\theta', \mathcal{C})$ be a deterministic configuration of interest.

$$\hat{N}_n^{(\theta',\mathcal{C})} := \frac{1}{\sum_{x_j \in E} k(\tau_{x_j}(\theta, \mathcal{P}_n), (\theta', \mathcal{C}))} \sum_{x_i \in E} N_1(x_i) k(\tau_{x_i}(\theta, \mathcal{P}_n), (\theta', \mathcal{C}))$$

where $\mathcal{P}_n := \mathcal{P} \cap [-n/2, n/2]^2$

What we want to show

Proposition

Under some conditions:

1. *on the process* $\mathcal{P}$ *(exponential mixing);*
2. *on the covariate field* θ *(exponential mixing);*
3. *on the similarity function* k *(stabilization).*

We have:

What we want to show

Proposition

Under some conditions:

1. *on the process $\mathcal{P}$ (exponential mixing);*
2. *on the covariate field θ (exponential mixing);*
3. *on the similarity function k (stabilization).*

We have:

$$\left(\mathrm{Var}\left(\hat{N}_n^{(\theta',\mathcal{C})}\right)\right)^{-1/2}\left(\hat{N}_n^{(\theta',\mathcal{C})} - \mathbb{E}[\hat{N}_n^{(\theta',\mathcal{C})}]\right) \xrightarrow{d} \mathcal{N}(0,1).$$

Examples of processes and similarity functions

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

Examples of processes and similarity functions

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

k verifies stabilization hypothesis: if we add a remove a point far from our point of interest then the similarity does not change.

Examples of processes and similarity functions

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

k verifies stabilization hypothesis: if we add a remove a point far from our point of interest then the similarity does not change.

Proposition

$$\left(\mathrm{Var}\left(\hat{N}_n^{(\theta',\mathcal{C})}\right)\right)^{-1/2}\left(\hat{N}_n^{(\theta',\mathcal{C})} - \mathbb{E}[\hat{N}_n^{(\theta',\mathcal{C})}]\right) \xrightarrow{d} \mathcal{N}(0,1)$$

Proof with a theorem from BYY 2025

How does it work with a toy data set ?

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	🪿 🪿 🐦🌧️🌡️🌲	?

In our example we say that the similarity only depend on the number of geese and the habitat:

$$k(L_1, L_2) = \frac{1}{2}\left(\mathbb{1}_{\text{number of geese of } L_1 = \text{number of geese of } L_2} + \mathbb{1}_{\text{same habitat}}\right)$$

Similarity matrix

Square	Year	Abundance	Environmental variables	Abundance next year	Similarity with square of interest
10295	2003	5	🦜🦜 🪿 🪿 🌡️☀️🌲	2	1
11158	2006	1	🪿 🪿 🌧️🌡️ 🌽	4	0.5
20204	2015	6	🐦🐦🐦☀️❄️ 🌽	7	0
30363	2019	8	🐦🐦🪿🦜 🌧️ ❄️ 🌲	9	0.5
950294	2024	3	🪿 🪿 🐦🌧️🌡️🌲	?	1

Similarity matrix

Square	Year	Abundance	Environmental variables	Abundance next year	Similarity with square of interest
10295	2003	5	🦜🦜 🪿 🪿 🌡️☀️🌲	2	1
11158	2006	1	🪿 🪿 🌧️🌡️ 🌽	4	0.5
20204	2015	6	🐦🐦🐦☀️❄️🌽	7	0
30363	2019	8	🐦🐦🪿🦜 🌧️ ❄️ 🌲	9	0.5
950294	2024	3	🪿 🪿 🐦🌧️🌡️🌲	?	1

$$\hat{N} = \frac{2 \times 1 + 4 \times 0.5 + 7 \times 0 + 9 \times 0.5}{1 + 0.5 + 0 + 0.5 + 1}$$
$$= 2.125$$

What with real data?

1. Divide into train and test datasets.
2. Similarity calculations:

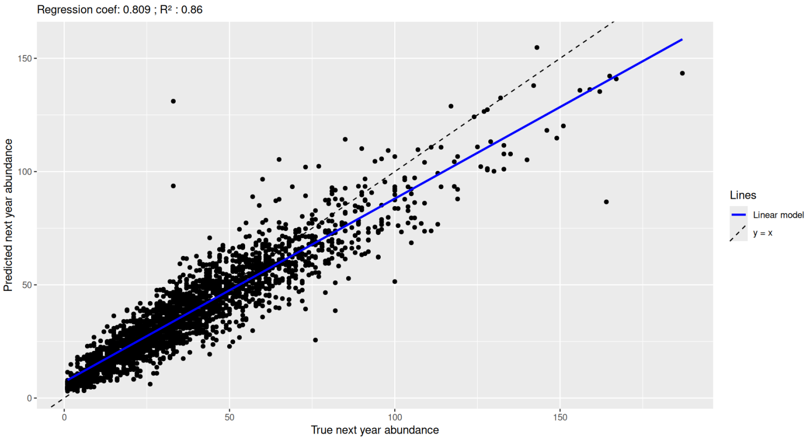
$$K(L_i, L_j) = \mathbb{1}_{k(L_1,L_2)\geq\alpha}$$
$$k(L_i, L_j) = \frac{\sum_n \min(L_i^n, L_j^n)}{\sum_i \max(L_i^n, L_j^n)}$$

with L_i in the train set and L_j in the test set and α a threshold.

3. Apply estimator for each line in the test set:

$$\hat{N}^{L_j} = \frac{1}{\sum_{L_i} k(L_i, L_j)} \sum_{L_i} N(L_i) k(L_i, L_j)$$

Results with real data



Forest species, $\alpha = 0.8$

Meadow pipit (*Anthus pratensis*)

From Charles J. Sharp.

Thank you !