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

Adélie Erard,Romain Lorrillière and Raphaël Lachièze-Rey

Séminaire Proba/Stat Poitiers – January 15, 2026

LABORATOIRE MAP5

CESCO Centre d'Écologie et des Sciences de la Conservation

Modélisation Mathématique et Biodiversité

# Breeding Bird Surveys

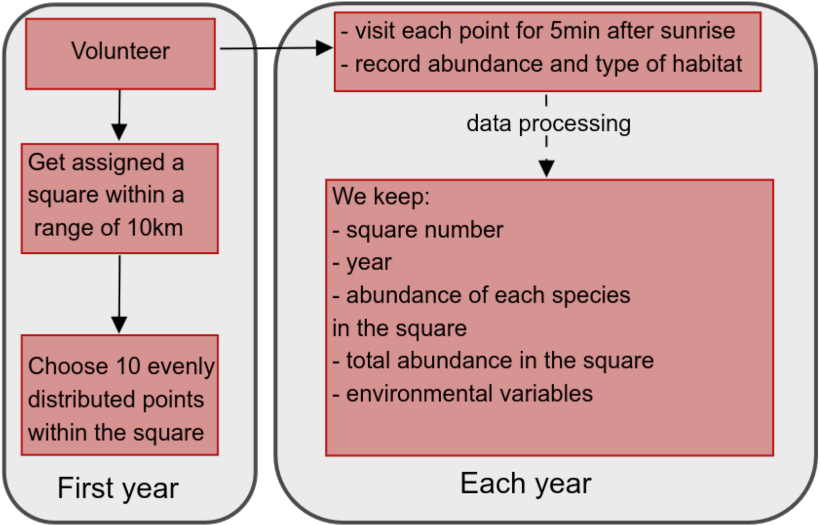
Meadow pipit (*Anthus pratensis*)

From Charles J. Sharp

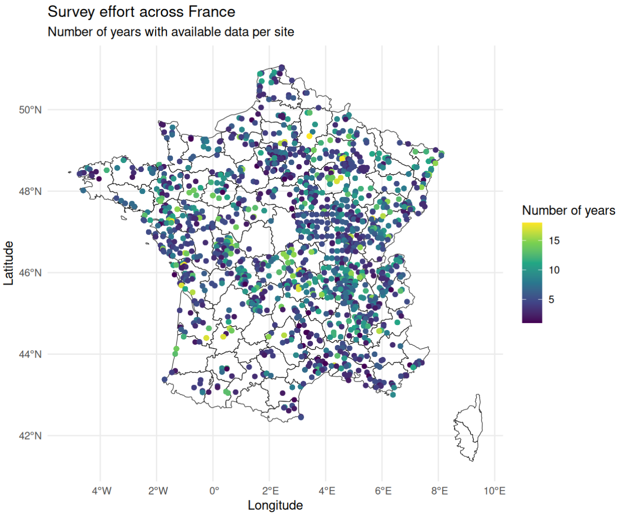
**Breeding Bird Surveys** (BBS) are long-term, large-scale, international avian monitoring programs designed to track the status and trends of bird populations.

**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);
- Indices on how the agricultural land is fragmented 10km around the observation.

# Goals

1. Give a method to estimate the variation of abundance (or future abundance) at a local scale.

2. Find which environmental variables induce changes in abundance.

# 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 $\tau_x$ 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 $\theta$.

# 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 $\theta$.

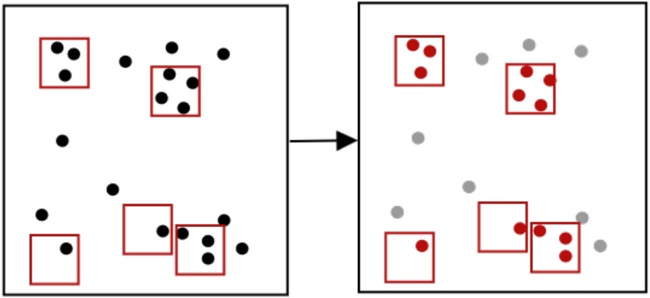
**Example**

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

# Construction of $\hat{N}$

Let $k$ be a similarity function. We define two statistics:

# Construction of $\hat{N}$

Let $k$ be a similarity function. We define two statistics:

$$F_{(\theta',\mathcal{C})}(\theta,\mathcal{P}_0) = \sum_{x_i \in E} N_1(x_i) k(\tau_{x_i}(\theta,\mathcal{P}_0),(\theta',\mathcal{C}))$$

# Construction of $\hat{N}$

Let $k$ be a similarity function. We define two statistics:

$$F_{(\theta',\mathcal{C})}(\theta,\mathcal{P}_0)=\sum_{x_i\in E}N_1(x_i)k(\tau_{x_i}(\theta,\mathcal{P}_0),(\theta',\mathcal{C}))$$

$$Z(\theta,\mathcal{P}_0)=\sum_{x_j\in E}k(\tau_{x_j}(\theta,\mathcal{P}_0),(\theta',\mathcal{C})).$$

# Estimator

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

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

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 $\mathcal{P}$;*
2. *on the covariate field $\theta$ ;*
3. *on the similarity function $k$.*

*We have:*

# What we want to show

## Proposition

*Under some conditions:*

1. *on the process* $\mathcal{P}$*;*
2. *on the covariate field* $\theta$ *;*
3. *on the similarity function* $k$*.*

*We have:*

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

# Hypothesis on the process

## Definition

The p-th order correlation function of $\mathcal{P}$ are defined as follows:

$$\mathbb{E}[\prod_{i=1}^{p} \mathcal{P}(B_i)] = \int_{B_1 \times \dots \times B_p} \rho^{(p)}(x_1, \dots, x_p) dx_1 \dots dx_p$$

# Hypothesis on the process

## Definition (BYY26+)

$\mathcal{P}$ has exponentially fast mixing correlation if for all $p, q \in \mathbb{N}$ and all $(x_1, \dots, x_{p+q}) \in (\mathbb{R}^2)^{p+q}$,

$$|\rho^{(p+q)}(x_1, \dots, x_{p+q}) - \rho^{(p)}(x_1, \dots, x_p)\rho^{(q)}(x_{p+1}, \dots, x_{p+q})| \leq C_{p+q}\phi(s)$$

where $\phi$ is a fast correlation decay function verifying:

$$-\infty \leq \liminf_{r \to \infty} \frac{\log \phi(r)}{r^b} < 0, \quad 0 < b$$

# A class of admissible process

## Proposition

*Let $\Psi$ be a Cox process with intensity $\Lambda$. We suppose that $\Lambda$ has a (random) density with respect to the Lebesgue measure:*
$\Lambda(B) = \int_B \lambda(x)dx$.
*The p-th correlation function of $\Psi$ is given by:*

$$\rho^{(p)}(x_1, \ldots, x_p) = \mathbb{E}[\prod_{i=1}^{p} \lambda(x_i)]$$

*If $\{\lambda(x)\}_{x \in \mathbb{R}^d}$ has fast decay of correlation then $\Psi$ has fast decay of correlation.*

# Log Gaussian Cox processes

**Definition**

Let $\Lambda$ be the measure with density $\lambda(u) = \exp(G(u))$, where $G$ is a Gaussian random field. Then the Cox process generated by $\Lambda$ is a log-Gaussian Cox process.

# Log Gaussian Cox processes

**Definition**

Let $\Lambda$ be the measure with density $\lambda(u) = \exp(G(u))$, where $G$ is a Gaussian random field. Then the Cox process generated by $\Lambda$ is a log-Gaussian Cox process.

**Proposition**

*If $G$ is a stationary Gaussian random field and has fast decay of correlation, then the LGCP defined above verifies mixing property.*

# Hypothesis on the covariate field

Let $\theta_x$ be the value of the covariate field at a point $x \in \mathcal{P}$.

## Definition (BYY26+)

$(\theta_x)_{x\in\mathcal{P}}$ has exponentially fast mixing correlation if for all $p, q \in \mathbb{N}$ and all $(\theta_{x,1}, \dots, \theta_{x,p+q}) \in (\mathbb{R}^2)^{p+q}$,

$$\begin{aligned}|\mathbb{E}[f(\theta_{x,1}, \dots, \theta_{x,p})g(\theta_{x,p+1}, \dots, \theta_{x,p+q})]\\ - \mathbb{E}[f(\theta_{x,1}, \dots, \theta_{x,p})]\mathbb{E}[g(\theta_{x,p+1}, \dots, \theta_{x,p+q})]| \leq C_{p+q}\phi(s)\end{aligned}$$

where $\phi$ is a fast correlation decay function verifying:

$$-\infty \leq \liminf_{r\to\infty} \frac{\log \phi(r)}{r^b} < 0, \quad 0 < b$$

# Stabilization hypothesis

Consider the marked process $\overline{\mathcal{P}} = \sum_{x \in \mathcal{P}} \delta_{(x,\theta_x)}$ and the score function $\xi : (x, \theta_x, \overline{\mathcal{P}}) \mapsto \mathbb{R}$.

# Stabilization hypothesis

Consider the marked process $\overline{\mathcal{P}} = \sum_{x \in \mathcal{P}} \delta_{(x,\theta_x)}$ and the score function $\xi : (x, \theta_x, \overline{\mathcal{P}}) \mapsto \mathbb{R}$.

## Definition (BYY19)

The radius of stabilization, $R_x^{\mathcal{P}}$, of $\xi$ at $x$ and on the marked process $\overline{\mathcal{P}}$ is defined as the smallest $r$ such that:

$$\xi(x, \theta_x, \overline{\mathcal{P}} \cap B(x,r)) = \xi(x, \theta_x, (\overline{\mathcal{P}} \cap B(x,r)) \cup (\mathcal{A} \cap B^c(x,r))),$$

with $\mathcal{A} \subset \mathbb{R}^2$.

# Stabilization hypothesis

Consider the marked process $\overline{\mathcal{P}} = \sum_{x \in \mathcal{P}} \delta_{(x, \theta_x)}$ and the score function $\xi : (x, \theta_x, \overline{\mathcal{P}}) \mapsto \mathbb{R}$.

## Definition (BYY19)

The radius of stabilization, $R_x^{\mathcal{P}}$, of $\xi$ at $x$ and on the marked process $\overline{\mathcal{P}}$ is defined as the smallest $r$ such that:

$$\xi(x, \theta_x, \overline{\mathcal{P}} \cap B(x,r)) = \xi(x, \theta_x, (\overline{\mathcal{P}} \cap B(x,r)) \cup (\mathcal{A} \cap B^c(x,r))),$$

with $\mathcal{A} \subset \mathbb{R}^2$. $\xi$ is said fast to be stabilizing on $\overline{\mathcal{P}}$ if for all $l \in \mathbb{N}$ there exists a fast decaying function $\phi$ such that:

$$\sup_{1 \leq n \leq \infty} \sup_{x_1, \ldots x_l \in [-n/2, n/2]^2} \mathbb{P}_{x_1, \ldots x_l}(R_x^{\mathcal{P}} > t) \leq C\phi(t)$$

# Admissible similarity function

Recall that we want to calculate similarity between a given context $(\theta', \mathcal{C})$ and a context around an observed point $x$ with: $\xi(x, \theta_x, \overline{\mathcal{P}}) = k(\tau_x(\theta, \mathcal{P}), \theta', \mathcal{C})$

# Admissible similarity function

Recall that we want to calculate similarity between a given context $(\theta', \mathcal{C})$ and a context around an observed point $x$ with: $\xi(x, \theta_x, \overline{\mathcal{P}}) = k(\tau_x(\theta, \mathcal{P}), \theta', \mathcal{C})$

Easy case: we compare $(\theta', \mathcal{C})$ with $\overline{\mathcal{P}} \cap B(x, \rho)$ with $\rho$ a finite range of observation.

# Admissible similarity function

Recall that we want to calculate similarity between a given context $(\theta', \mathcal{C})$ and a context around an observed point $x$ with: $\xi(x, \theta_x, \overline{\mathcal{P}}) = k(\tau_x(\theta, \mathcal{P}), \theta', \mathcal{C})$

Easy case: we compare $(\theta', \mathcal{C})$ with $\overline{\mathcal{P}} \cap B(x, \rho)$ with $\rho$ a finite range of observation.

A more challenging case: we compare $(\theta', \mathcal{C})$ with the n closest neighbor of $x$.

# A useful CLT

**Theorem (5.2 of BYY26+)**

*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 $n \to \infty$:*

$$\left(\mathrm{Var}\left(\mu_n^\xi\right)\right)^{-1/2}\left(\mu_n^\xi - \mathbb{E}[\mu_n^\xi]\right) \xrightarrow{d} \mathcal{N}(0,1)$$

*with $\mu_n^\xi = \sum_{x \in \mathcal{P}_n} \xi((x, U_x), \overline{\mathcal{P}})$*

# In our case

If $\mathcal{P}$ has exponential mixing correlations and $k$ verifies the stabilization hypothesis we have:

## Lemma

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

$$\left(\mathrm{Var}(Z(\theta,\mathcal{P}_n))\right)^{-1/2}\left(Z(\theta,\mathcal{P}_n)-\mathbb{E}[Z(\theta,\mathcal{P}_n)]\right)\stackrel{d}{\Longrightarrow}\mathcal{N}(0,1)$$

# What about the convergence of the ratio?

In [BYY26+] they also have a LLN under the same kind of conditions.

Using Slutsky's lemma and the continuous mapping theorem we should have the convergence of the ratio and thus the following proposition:

# What about the convergence of the ratio?

In [BYY26+] they also have a LLN under the same kind of conditions.

Using Slutsky's lemma and the continuous mapping theorem we should have the convergence of the ratio and thus the following proposition:

## Proposition

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

# 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:

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

# 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 |

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

# Gower's distance function

Let $L_1$, $L_2$ be two lines in the data set. Suppose we have $p$ variables of interest. The Gower distance is defined as:

# Gower's distance function

Let $L_1$, $L_2$ be two lines in the data set. Suppose we have $p$ variables of interest. The Gower distance is defined as:

$$g(L_1, L_2) := \frac{1}{p} \sum_{i=1}^{p} d_i(L_1, L_2),$$

$$d_i(L_1, L_2) := \frac{|(L_1)_i - (L_2)_i|}{R_i} \text{ for quantitative variables,}$$

$$d_i(L_1, L_2) := \mathbb{1}_{(L_1)_i \neq (L_2)_i} \text{ for categorical variables,}$$

where $R_i$ is the span of the i-th covariate.

$$k(L_1, L_2) = \mathbb{1}_{g(L_1, L_2) \leq \alpha},$$

$$0 \leq \alpha \leq 1$$

# What happen with the Gower similarity function?

| 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 | 🦢 🦢 🐦🌧️🌡️🌲 | ? |

From the table above:

$$
\begin{aligned}
g(L_1, L_5) &= \frac{1}{6}\left(\frac{|5-3|}{7} + \frac{|2-0|}{2} + \frac{|2-2|}{2} + \frac{|0-1|}{3} + 0 + 0 + 1\right) \\
&= \frac{55}{126} \approx 0.44
\end{aligned}
$$

# What with real data?

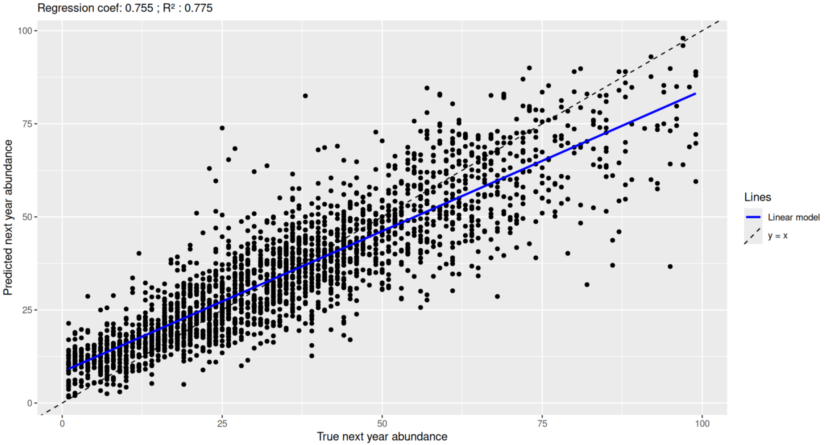
1. Divide into train and test datasets
2. Similarity calculations:

$$k(L_i, L_j) = \mathbb{1}_{g(L_1,L_2)\leq 0.05}$$

   with $L_i$ in the train set and $L_j$ i the test set
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

# What's next?

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

Meadow pipit (*Anthus pratensis*)

From Charles J. Sharp.

Thank you !

Błaszczyszyn, B., D. Yogeshwaran, and J. E. Yukich (2019). "Limit Theory for Geometric Statistics of Point Processes Having Fast Decay of Correlations". In: *The Annals of Probability* 47.2. Publisher: Institute of Mathematical Statistics, pp. 835–895. ISSN: 0091-1798. URL: http://www.jstor.org/stable/26613430 (visited on 10/31/2023).

— (2026). "Limit theory for statistics of Lipschitz-localized stochastic processes in spatial random models". In: *in preparation*. +.