

RESEARCH ARTICLE

WILEY

A vector of point processes for modeling interactions between and within species using capture-recapture data

Alex Diana¹ | Eleni Matechou¹ | Jim E. Griffin² | Yadvendradev Jhala³ | Qamar Qureshi³

¹Division of Computing, Engineering and Mathematical Sciences, University of Kent, Canterbury, UK

²Department of Statistical Science, University College London, London, UK

³Wildlife Institute of India, Dehradun, India

Correspondence

Alex Diana, Division of Computing, Engineering and Mathematical Sciences, University of Kent, Canterbury, UK.
Email: a.diana@kent.ac.uk

Abstract

Capture-recapture (CR) data and corresponding models have been used extensively to estimate the size of wildlife populations when detection probability is less than 1. When the locations of traps or cameras used to capture or detect individuals are known, spatially-explicit CR models are used to infer the spatial pattern of the individual locations and population density. Individual locations, referred to as activity centers (ACs), are defined as the locations around which the individuals move. These ACs are typically assumed to be independent, and their spatial pattern is modeled using homogeneous Poisson processes. However, this assumption is often unrealistic, since individuals can interact with each other, either within a species or between different species. In this article, we consider a vector of point processes from the general class of interaction point processes and develop a model for CR data that can account for interactions, in particular repulsions, between and within multiple species. Interaction point processes present a challenge from an inferential perspective because of the intractability of the normalizing constant of the likelihood function, and hence standard Markov chain Monte Carlo procedures to perform Bayesian inference cannot be applied. Therefore, we adopt an inference procedure based on the Monte Carlo Metropolis Hastings algorithm, which scales well when modeling more than one species. Finally, we adopt an inference method for jointly sampling the latent ACs and the population size based on birth and death processes. This approach also allows us to adaptively tune the proposal distribution of new points, which leads to better mixing especially in the case of non-uniformly distributed traps. We apply the model to a CR data-set on leopards and tigers collected at the Corbett Tiger Reserve in India. Our findings suggest that between species repulsion is stronger than within species, while tiger population density is higher than leopard population density at the park.

KEYWORDS

animal interactions, Bayesian analysis, interaction point processes, spatial capture-recapture

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Environmetrics* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Conservation scientists, researchers, and charities, as well as government organizations worldwide, are interested in monitoring sizes of populations, distributions of species in the landscape and spatial patterns of individuals in the wild. Capture-recapture (CR) is one of the most commonly employed surveying tools to achieve this goal. According to the CR protocol, an attempt is made to capture individuals on repeated sampling occasions at one or more traps, and all newly caught individuals are uniquely marked. Several models have been developed for estimating population sizes (Royle, 2004), arrival and departure patterns (Diana et al., 2020; Pledger et al., 2009) or survival patterns (Brownie, 1985) using CR data. When more than one trap or camera are used to sample individuals, and the locations of these traps are known, then spatially-explicit CR (SCR) models (Borchers & Efford, 2008; Efford, 2004; Royle & Young, 2008) can be used to infer the activity centers (ACs) of individuals, defined as the locations around which the individuals move when performing their daily activity (Royle et al., 2013). Even in cases when the spatial pattern is not of main interest, SCR models can provide more reliable estimates of population density, as they account for the fact that individuals are more likely to be caught in traps that are in close proximity with their AC (Reich & Gardner, 2014). For a nice review of the literature on SCR models see Borchers and Fewster (2016).

When using SCR models, the locations of the individual ACs are latent, which gives rise to challenges from an inferential perspective. SCR models can be developed in a classical (Borchers & Efford, 2008) or Bayesian (Royle et al., 2013) framework. The classical approach relies on maximizing the likelihood function after integrating out the ACs, while the Bayesian approach relies on sampling from the posterior distribution of the ACs as well as of the other parameters of interest. Estimation in the frequentist case has been performed in the case where the ACs are assumed to come from a tractable distribution, which makes the computations analytically possible. However, in cases where individuals are not independent in terms of the locations of their ACs, that is, when there is interaction between individuals, the resulting processes are quite complex, and the likelihood function is intractable. Hence, the classical approach to inference is not available, and instead a Bayesian approach needs to be considered.

A common assumption in SCR models is that the spatial pattern of ACs is assumed to be described by a homogeneous Poisson point process (PP), which is equivalent to assuming no interaction between individuals. In this context, when individuals interact then the probability density of a given point (i.e., the AC of one member of the species) depends on the location of the other points (the AC of the other members). This dependence can manifest either as *repulsion*, where the presence of a point at a particular location decreases the probability of another point at a nearby location, or as *attraction*, where the opposite is true and points tend to be clustered. To our knowledge, the only SCR models that do not assume a Poisson process are presented in Reich and Gardner (2014) and McLaughlin and Bar (2021). Reich and Gardner (2014) used a Strauss process (Strauss, 1975) to model the ACs of a single species while accounting for interaction within the species, while McLaughlin and Bar (2021) assume no interaction between individuals but instead assume interaction between the ACs of each individual over time. As shown by Reich and Gardner (2014), it is important to account for interaction since failing to account for it can lead to biases in the estimate of the population size. In this article we focus on the case of repulsion between as well as within species.

Spatial PPs are a major area of research in spatial statistics and a useful modeling tool when the objects of interest are locations of objects in space (Cressie, 1993; Ripley, 1977). The simplest example of a PP is the Poisson process, where the locations of the individual points are drawn independently of each other and do not exhibit interaction. One extension of the Poisson process is the class of Cox processes, which are processes for clustered or aggregated point patterns (Neyman & Scott, 1958). Another extension of point processes is the class of interaction PPs (Preston, 1976; Ruelle, 1969), which are especially suited to modeling interactions between points. Interaction PPs are more suited to modeling repulsion, which is the focus of this article, whereas Cox processes are more suited to modeling attraction.

The computational challenge that arises when working with interaction PPs is that the normalizing constant of the likelihood function is usually unavailable in closed form apart from trivial cases (Møller & Waagepetersen, 2003). Distributions of this kind are known in the literature as *doubly-intractable* distributions, as both the marginal likelihood of the data, with parameters integrated out, and the likelihood function are intractable. In a frequentist setting, this is usually dealt with by replacing the likelihood function with a pseudo-likelihood (Besag, 1978), for which asymptotic results are available (Jensen & Møller, 1991). However, in the SCR case where the ACs are latent, the use of the pseudo-likelihood does not overcome the challenge of integrating out the ACs from the likelihood function, since integrals with respect to interaction point processes cannot be computed in closed form (Møller & Waagepetersen, 2003). In a Bayesian setting, the problem usually involves computing a ratio of normalizing constants, as the ratio arises naturally when using a Metropolis–Hastings (MH) algorithm for sampling from the posterior distribution of the parameters of the PP. Popular

inference procedures include the auxiliary variable method of Møller et al. (2006), path sampling (Ogata & Tanemura, 1984), exchange sampling (Murray et al., 2012) and the Monte Carlo MH (MCMH) algorithm (Liang & Jin, 2013).

When working with several species, it is necessary to work with vectors of PPs instead of a single PP, since each PP corresponds to a single species. Although an extensive literature exists for univariate interaction PPs, vectors of interaction PPs have only recently gained attention in the literature (Rajala et al., 2018; Waagepetersen et al., 2016). For example, Waagepetersen et al. (2016) developed a model that takes into account correlation between more than one process using a Log-Gaussian Cox process, while Rajala et al. (2018) proposed a model to take into account interactions between many processes using a Gibbs model, which is a particular case of interaction PPs.

In this article, we extend the model of Reich and Gardner (2014) by developing a model for more than one species. To this end, we introduce a vector of PPs for modeling the locations of the ACs and build a multi-species SCR model based on this newly defined vector of PPs. Working with vectors of PPs brings additional challenges from an inferential perspective that cannot be overcome by the inferential methods used in Reich and Gardner (2014) as they do not scale for vectors of PPs. Hence, we propose to use the MCMH algorithm, as this approach is easily scalable as the number of parameters grows.

The type of PPs with which we work in this article assume a joint distribution between the number of points and their locations and are known in the literature as *unconditional* PPs (Møller & Waagepetersen, 2003), as opposed to *conditional* PPs, which assume a generic distribution for the number of points and model the locations of the points conditional on their number. Although inference in SCR models is usually performed using data augmentation (DA) (Converse & Royle, 2012; Royle, 2009), this approach is not an option when working with unconditional PPs and therefore we consider an alternative approach relying on a general algorithm for sampling spatial PPs introduced by Geyer and Møller (1994). This algorithm samples from the joint posterior distribution of the latent ACs by either moving existing points, adding new points or deleting existing points from the PP. We tune this approach in two ways. First, we choose the number of proposed and removed points adaptively in order to achieve the optimal mixing of the algorithm. Second, we adaptively learn the distribution of the locations of the newly proposed points.

We apply our new model to CR data-sets on tigers and leopards, collected in the Jim Corbett National Reserve in India. We are interested in estimating population density of the two species, which are known to be territorial and hence repulse each other within and between species, although this has never been quantified using CR data before.

The article is organized as follows. In Section 2, we briefly summarize the theory of PPs, as well as the extension to vectors of PPs and the main approaches available for performing inference. In Section 3, we define the spatial model for CR data. In Section 4, we present results of a simulation study and in Section 5 we apply the model to the tiger and leopard CR data-sets. Section 6 concludes the article and introduces some potential future directions.

2 | INTERACTION POINT PROCESSES

A spatial PP on a space S is a random variable whose realizations are sets of points in S . A PP can be defined conditional on a fixed number of points, in which case it is said to be a *conditional* PP, or the number of points can be sampled together with their locations, in which case it is said to be an *unconditional* PP.

The most basic example of a PP is the Poisson process, which is defined by an intensity that models the expected number of points in each region. A Poisson process is said to be homogeneous if the intensity is constant over the space, and inhomogeneous otherwise. The Poisson process is an unconditional PP, and its conditional equivalent is the binomial PP, where the number of points is fixed in advance. According to the Poisson or binomial process, points are independent of each other, in the sense that the location of a point does not affect the location of another point.

PPs can be defined by a general intensity function $f(x_1, \dots, x_n)$ where $x_1, \dots, x_n$ are the locations of the observed points. The normalizing constant of the intensity f is

$$\begin{cases} \int_S f(x_1, \dots, x_n) dx_1, \dots, dx_n & \text{conditional case} \\ \sum_{n=0}^{\infty} \frac{\exp(-|S|)}{n!} \int_S f(x_1, \dots, x_n) dx_1, \dots, dx_n & \text{unconditional case,} \end{cases}$$

which are both intractable (except in the case of a Poisson or binomial process) and hence f is usually expressed up to proportionality $f \propto h$.

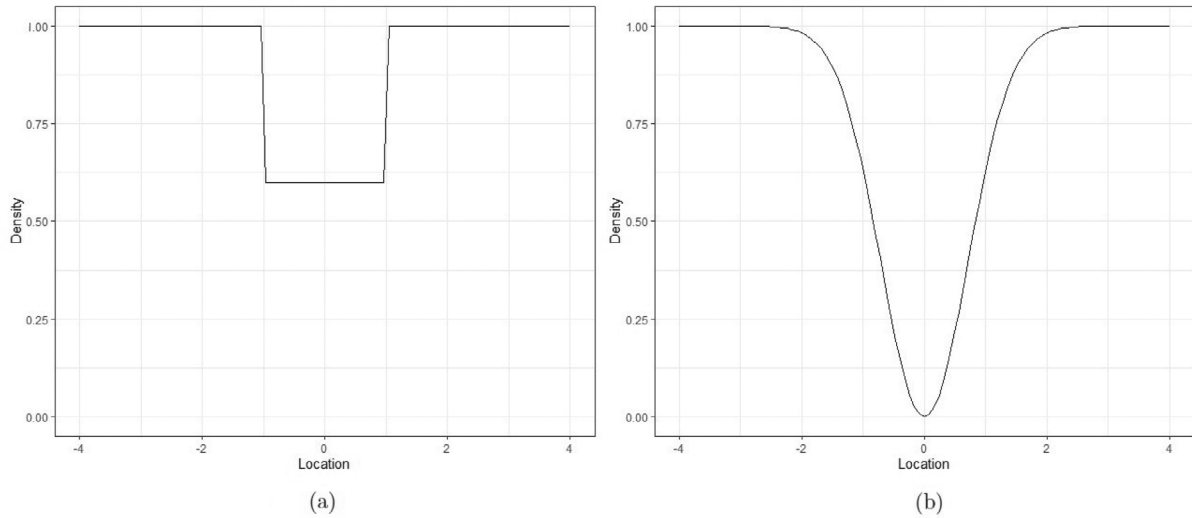


FIGURE 1 Interaction function $g(|x - y|) = \phi_2(x, y)$ for a Strauss process with parameters $\lambda = 0.6$, $R = 1$ (a) and soft-core process with parameter $\gamma = 2$ (b).

Additionally, in the unconditional case, the marginal distribution on the number of points is

$$\mathbb{P}(n(X) = n) = \frac{\exp(-|S|)}{n!} \int_{S^n} f(x_1, \dots, x_n) dx_1, \dots, dx_n. \quad (1)$$

Expression (1) is also intractable and hence it is necessary to resort to simulation to evaluate the distribution on the number of points. This is an advantage of the conditional case, as the distribution of the number of points can be chosen explicitly and independently on the locations of the points.

In this article, we choose f from the class of pairwise interaction functions, which have the form

$$f(\{x_1, \dots, x_n\}) \propto \prod_{i=1}^n \phi_1(x_i) \prod_{i < j} \phi_2(x_i, x_j | \theta), \quad (2)$$

where ϕ_1 is the intensity function and ϕ_2 is the interaction function. Whenever $\phi_2 < (>) 1$, there is repulsion (attraction) between points. In the following, we set ϕ_1 equal to a constant β .

Pairwise interaction PPs can thus be defined by simply specifying the function ϕ_2 that determines the interaction between two points. For example, the Strauss process (Strauss, 1975) is obtained when $\phi_2(x_1, x_2 | \lambda, R) \propto \lambda^{1[|x_1 - x_2| < R]}$, where $1[A] = 1$ if condition A is satisfied and 0 otherwise. The interpretation of this process is that points interact only if they are at a distance less than R . The Strauss process is well-defined only for $\lambda \leq 1$ (for $\lambda = 1$ it corresponds to the Poisson process), since if $\lambda > 1$ the number of points explodes. The limit of the Strauss process as $\lambda \rightarrow 0$ is termed the hard-core process and is a process where there are no two points at distance $< R$. Another example of interaction PP is the soft-core process (Ogata & Tanemura, 1984), defined by the interaction function

$$\phi_2(x_1, x_2 | \gamma) \propto 1 - \exp\left(-\frac{(x_1 - x_2)^2}{\gamma}\right), \quad (3)$$

which, as opposed to the Strauss process, allows points to interact at any distance. Examples of interaction functions for the two processes are shown in Figure 1.

2.1 | Vectors of point processes

Similarly to a single PP, a vector of D PPs can be defined as a set of D different point processes. Analogously to a single PP, we will define a vector of PPs $(X_1, \dots, X_D)$ with density $f = (f_1, \dots, f_D)$.

If f can be factorized as $f(X_1, \dots, X_D) = f_1(X_1) \dots f_D(X_D)$, the vector of PPs reduces to a product of D independent single interaction PPs. We consider, as before, pairwise interaction functions w.r.t. a vector of independent unit rate Poisson processes, that is, we choose f as

$$f(\mathbf{x}_1, \dots, \mathbf{x}_D) \propto \underbrace{\prod_{d=1}^D \prod_{i=1}^{N_d} \phi_1(x_i^d)}_{\text{single point intensities}} \underbrace{\prod_{d=1}^D \prod_{i=1}^{N_d} \prod_{j=i+1}^{N_d} \phi_2(x_i^d, x_j^d | \theta_d^1)}_{\text{within PP interactions}} \underbrace{\prod_{d_1=1}^D \prod_{d_2=2}^{d_1-1} \prod_{i=1}^{N_{d_1}} \prod_{j=1}^{N_{d_2}} \phi_2(x_i^{d_1}, x_j^{d_2} | \theta_{d_1, d_2}^2)}_{\text{between PP interactions}},$$

where $(\mathbf{x}_1, \dots, \mathbf{x}_D) = ((x_1^1, \dots, x_{N_1}^1), \dots, (x_1^D, \dots, x_{N_D}^D))$, θ_d^1 is the parameter of the within-points interaction of $\mathbf{x}_d$ and θ_{d_1, d_2}^2 is the parameter of the between-points interaction between $\mathbf{x}_{d_1}$ and $\mathbf{x}_{d_2}$. For example, ϕ_2 can be taken as the interaction function of the Strauss or soft-core process. Similarly to the Strauss process, a sufficient condition to guarantee that the PP does not explode is $\phi_2(x_1, x_2) < 1 \forall x_1, x_2$. A similar construction to the one provided above is given in Nightingale et al. (2015).

In CR data, individual ACs are latent. Additionally, CR data can be sparse, with only a small number of individuals being observed more than once. Hence, it is important to consider parsimonious models and limit the number of parameters that need to be estimated. For this reason, we choose to work with the soft-core instead of the Strauss process. For comparison, we note that when $D = 2$, using a Strauss process requires eight parameters (each pair (λ, R) is repeated for the two between and the within species interactions plus the two intensities β), as opposed to the five required for the soft-core process.

If $D = 2$ and the interaction function is taken to be the one of the soft-core process, we term the resulting PP a *bivariate soft-core process* and we use the notation $\text{BivSC}(\gamma_1, \gamma_2, \gamma_{12}, \beta_1, \beta_2)$ for the distribution of this process. In this case, f takes the form

$$f(\{x_1^1, \dots, x_n^1\}, \{x_1^2, \dots, x_m^2\} | \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_{12}) \propto \beta_1^n \beta_2^m \prod_{i < j} \left\{ 1 - \exp\left(-\frac{(x_i^1 - x_j^1)^2}{\gamma_1}\right) \right\} \prod_{i < j} \left\{ 1 - \exp\left(-\frac{(x_i^2 - x_j^2)^2}{\gamma_2}\right) \right\} \prod_{ij} \left\{ 1 - \exp\left(-\frac{(x_i^1 - x_j^2)^2}{\gamma_{12}}\right) \right\}, \quad (4)$$

where γ_1 and γ_2 model the interaction within each of the two species and γ_{12} models the interaction between the species. Similarly to what was described in the previous section for univariate PPs, the process defines marginally a joint distribution on the number of points n_1 and n_2 of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively.

In Figure 2, we show several sets of simulations from the process with $D = 2$. Increasing the interaction parameters leads to a reduction in the number of points that can be located within a given area. Hence, to show how the pattern of points changes as the interaction parameters vary, we simulate from a *conditional* point process. The first column shows samples with minimal within-species interaction, while in the second column the interaction is at least as strong as the between-species interaction. As the parameter γ_{12} increases (first row to second row), the points of different colors become more separated. The bottom right plot is equivalent to a single realization from a single soft-core PP, as the interaction is the same for each pair of points, regardless of color.

2.2 | Bayesian inference for interaction point processes

As mentioned in the introduction, posterior inference for the parameters $\theta = (\{\theta_1^d\}_{d=1, \dots, D}, \{\theta_2^{d_1, d_2}\}_{d_1, d_2=1, \dots, D})$ of an interaction PP X is complicated as the normalizing constant is not available in closed form. Given a sample from a PP $y \sim X | \theta$ and a prior π for θ , posterior inference can be performed using a MH sampler to sample from the posterior distribution $p(\theta | y) = \pi(\theta) \frac{h(y | \theta)}{Z_\theta}$, where we have written the likelihood of the PP in an unnormalized form $f(y | \theta) = \frac{h(y | \theta)}{Z_\theta}$. The MH ratio for the parameter θ takes the form

$$H(\theta' | \theta) = \frac{\pi(\theta') h(y | \theta') q(\theta | \theta')}{\pi(\theta) h(y | \theta) q(\theta' | \theta)} \frac{Z_\theta}{Z_{\theta'}}, \quad (5)$$

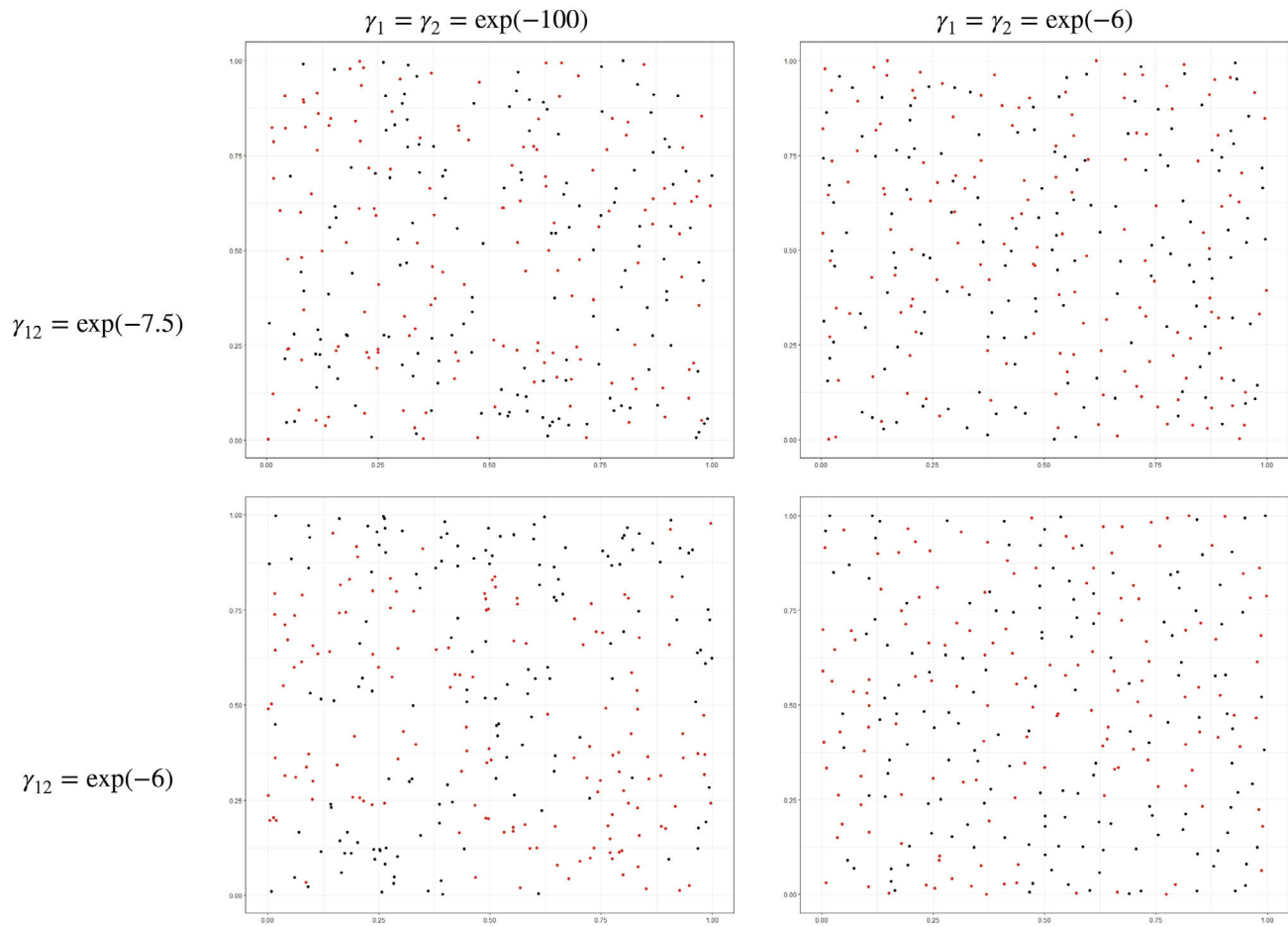


FIGURE 2 Simulations from a conditional BivSC process with $N_1 = N_2 = 100$, where γ_1 and γ_2 are the within-species interaction parameters for the two processes and γ_{12} is the between-species interaction parameter. The top row and bottom represent cases of low and high between-species interaction, respectively. The first and second column represent cases of low and high within-species interaction, respectively.

where q is the proposal distribution. The unavailability of the ratio $\frac{Z_\theta}{Z_{\theta'}}$ makes the use of standard Markov chain Monte Carlo (MCMC) techniques infeasible and thus a different sampling procedure has to be used. For ease of notation, we express $f(y|\theta)$ and $h(y|\theta)$ as $f_\theta(y)$ and $h_\theta(y)$, respectively.

An interesting method to avoid calculating the ratio is the approach of Møller et al. (2006), which relies on the introduction of an auxiliary variable to simplify the ratio of normalizing constants (RNC). However, the method requires knowledge of the density of the auxiliary variable, which is known only if the auxiliary variable takes standard forms, and hence it is difficult to apply in practice. A more applicable generalization is the exchange sampling algorithm of Murray et al. (2012), which only requires simulating from the process of interest.

Another interesting inference procedure is path sampling (Ogata & Tanemura, 1984), which is also used in the CR model of Reich and Gardner (2014). Path sampling is based on the identity $\log \left(\frac{Z_{\theta_i}}{Z_{\theta'_i}} \right) = \int_0^1 G_i(\theta(s)) s(t) dt$ where $G_i(\theta(s))$ is a function of θ and $s(t)$ is a path connecting θ and θ' . By computing the function $G_i(\cdot)$ on a grid of values of θ in advance, the integral can be approximated numerically and used in (5). However, this quickly becomes infeasible if the dimension of the parameter θ is greater than 2, as for example in the case of vectors of PPs, as the grid becomes prohibitively large.

A more scalable approach is based on the MCMH algorithm of Liang and Jin (2013) (LJ13), which is based on replacing the RNC with an importance sampling estimate. LJ13 propose three versions of the MCMH algorithm and we adopt here the algorithm termed MCMH-I, summarized in Algorithm 1.

The MCMH algorithm can be seen as a generalization of the exchange sampling algorithm, as the latter is obtained when $M = 1$ in the MCMH-III algorithm of LJ13. However, the advantage of the MCMH over the exchange sampling algorithm is that by drawing the samples $X_1, \dots, X_M$ in parallel, the MCMH can run at comparable speed to the exchange

Algorithm 1. Monte Carlo MH algorithm-I

1. Sample $\theta' \sim q(\cdot|\theta)$.
2. Accept θ' with probability $\min(1, H(\theta'|\theta))$ where

$$H(\theta'|\theta) = \frac{\pi(\theta')h(y|\theta')q(\theta|\theta')}{\pi(\theta)h(y|\theta)q(\theta'|\theta)} \frac{1}{R_M(\theta, \theta', X)},$$

where $R_M(\theta, \theta', y) = \frac{1}{M} \sum_{m=1}^M \frac{h(X_m|\theta')}{h(X_m|\theta)}$ and M is the number of samples used to compute the Monte Carlo approximation.

3. If θ' is accepted, sample $(X_1, \dots, X_m)$ from $f(x|\theta')$.

sampling algorithm, while providing a substantially better estimate of the RNC. A further advantage of the MCMH is that it does not depend on the dimension of θ , which makes it useful for application to vectors of PPs.

The samples $(X_1, \dots, X_m)$ in step 3 can be obtained through either MH (Geyer & Møller, 1994) or through perfect simulation (Kendall & Møller, 2000). A review of the existing algorithms for simulating interaction PPs is presented in Møller and Waagepetersen (2003). We choose to use an MH algorithm as, when a new θ' is accepted, the current samples $(X_1, \dots, X_m)$ from $f(x|\theta)$ can be used as a warm start for the MH algorithm to sample from $f(x|\theta')$, significantly reducing the burn-in period. In fact, if θ and θ' are sufficiently close, we also expect samples from PPs with densities $f_{\theta'}$ and f_{θ} to be close. On the other hand, perfect simulation is computationally very intensive and hence not feasible to perform at each iteration. We describe the MH algorithm to obtain samples from an interaction PP in the Appendix.

3 | SPATIAL CAPTURE-RECAPTURE MODEL

We now define the spatial model for CR data collected on two or more species. The prior for the ACs of individuals is defined using the soft-core model introduced in Section 2. For simplicity, we present here a model for two species but we note that the case of more than two species is constructed similarly.

We assume that there are K traps, placed in the study area at locations $(x_1, \dots, x_K)$, and that sampling is performed for T sampling occasions in each trap. For two species, the data can be summarized in two matrices, H^1 and H^2 , where H_{ik}^l is the number of times individual i of species l is captured at trap k .

Each individual is assumed to have an AC s_i^l for the duration of the study. Given the AC s_i^l , the probability of capturing individual i of species l at trap k is a decreasing function of the distance between the location of the AC of the individual, s_i^l , and the trap location x_k . We choose here an exponential function, which leads to probability of capture $p(s, x) = p_0 \exp(-\frac{1}{2\sigma^2}||s - x||^2)$ between AC s and trap location x , where p_0 is the baseline probability of detection if the distance is 0, and σ is a scale parameter. We assume that the two populations are closed during the study, that is, there is no mortality or recruitment and there is no immigration or emigration, and we define by N_1 and N_2 the number of individuals in the first and second population, respectively. The two vectors of ACs $(\mathbf{s}^1, \mathbf{s}^2) = ((s_1^1, \dots, s_{N_1}^1), (s_1^2, \dots, s_{N_2}^2))$ are assumed to have a BivSC($\beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_{12}$) prior. Since this process is an unconditional PP, this also defines a prior for N_1 and N_2 , given by Equation (1) in the case of single PP. As stated in Section 2, we resort to simulations to evaluate the prior. We assume gamma prior distributions for $\gamma_1, \gamma_2, \gamma_{12}$, and σ_l and beta prior distributions for p_0^l . The model can be summarized hierarchically as

$$\begin{cases} H_{ik}^l = \text{Binomial}(T, p_{ik}^l) & i = 1, \dots, N_l, \quad k = 1, \dots, K \quad l = 1, 2 \\ p_{ik}^l = p_0^l \exp\left(-\frac{1}{2\sigma_l^2}||s_i^l - x_k||^2\right) & i = 1, \dots, N_l, \quad k = 1, \dots, K \quad l = 1, 2 \\ (\mathbf{s}^1, \mathbf{s}^2) \sim \text{BivSC}(\gamma_1, \gamma_2, \gamma_{12}, \beta_1, \beta_2) \\ \beta_j \sim \text{Gamma}(a_\beta, b_\beta) & j = \{1\}, \{2\} \\ \gamma_j \sim \text{Gamma}(a_\gamma, b_\gamma) & j = \{1\}, \{2\}, \{12\} \\ p_0^l \sim \text{Beta}(a_p, b_p) \\ \sigma_l \sim \text{Gamma}(a_\sigma, b_\sigma). \end{cases}$$

We emphasize that comparing within-species interactions across different species, that is, comparing the absolute values of γ_1 and γ_2 is not meaningful. Interpretation of the strength of a within-species interaction does not only depend on the value of the corresponding γ parameter, but also on the population density for a given area size. For example, in a high-density population, even a small γ value can lead to substantial repulsion between individuals, as space is limited. On the other hand, in a sparse population, the same value of γ will have a much smaller effect, as individuals are further away from each other. Nonetheless, the within species interaction parameters, γ_1 and γ_2 , can still be compared to the between species interaction parameter, γ_{12} , to determine whether the between species interaction is stronger than the within species interaction. For this reason, in what follows we focus on the log ratios $\log\left(\frac{\gamma_{12}}{\gamma_1}\right)$ and $\log\left(\frac{\gamma_{12}}{\gamma_2}\right)$.

3.1 | Inference

We sample from the posterior distribution of the parameters $(\mathbf{s}^1, \mathbf{s}^2, p_0^l, \sigma_l, \beta_j, \gamma_j)$ by using an MCMC approach. The parameters p_0^l, σ_l can be sampled straightforwardly from their full conditional, while the PP parameters $(\beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_{12})$ can be sampled using the MCMH algorithm described in Section 2.2.

For the remaining set of parameters, $(\mathbf{s}^1, \mathbf{s}^2)$, we note that since we use an unconditional PP prior, the DA scheme, used for example in Reich and Gardner (2014) and McLaughlin and Bar (2021), cannot be used for inference, as the DA scheme assumes a Binomial prior on the population size independently on the point locations. Hence, we sample the latent ACs $(\mathbf{s}^1, \mathbf{s}^2)$ from their full conditional by generalizing the algorithm for simulating PPs of Geyer and Møller (1994), according to which, in each step, one of the following moves is proposed: changing the location of the existing points, proposing a random number n_b of new points and deleting a random number n_d of existing points. The full algorithm is described in the Appendix.

We propose two additional strategies to tune the algorithm. First, the distribution of the number of new points proposed can be learnt adaptively in order to target an optimal acceptance rate. Second, we adaptively learn the distribution of the locations of newly proposed points, q_b , in order to propose points where they are more likely to be accepted.

To achieve the former, we assume that the number of proposed points to add/remove follows a $\text{Pois}(\lambda)$ distribution, and we learn λ adaptively using a Robbins–Monro update (Robbins & Monro, 1951). More specifically, every n_c iterations, we update λ according to $\lambda^{(t+1)} = \lambda^{(t)} + \frac{1}{t}w(\bar{h} - h^*)$, where $\bar{h}$ is the average value of the MH ratio over the previous n_c iterations, h^* is the targeted acceptance rate and w is a fixed step size. In our case, we set $n_c = 50$, $h^* = 0.4$ and $w = 10$. This procedure is repeated separately for each PP. We note that we divide the step size by the number of updates t in order to obtain decreasing to 0 step sizes and satisfy the diminishing adaptation condition necessary to maintain ergodicity (Andrieu & Thoms, 2008).

To adaptively learn q_b , we model $q_b = \pi U(S) + (1 - \pi) \sum_{k=1}^K w_k N(\mu_k, \Sigma_k)$ where $U(S)$ is the uniform distribution over the sample space S and π is the weight of the mixture, thereby using an inhomogeneous Poisson process to sample the new locations. The weight of the mixture has been chosen equal to 0.95. The mixture $\sum_{k=1}^K w_k N(\mu_k, \Sigma_k)$ is fitted on the locations on the accepted points during the burn-in phase of the MCMC using a k-means algorithm, with the number of clusters chosen using the Akaike information criterion (Akaike, 1974). Since adaptation can perturb the ergodicity properties of the MCMC, we perform this step only at the end of the burn-in phase. Further details are presented in the Appendix. A similar approach for adaptive estimation of mixture of normals is presented in Giordani and Kohn (2010).

Adapting the proposal for the population size is necessary for large population sizes, since only proposing a single point at a time could lead to a very slow mixing rate, whereas learning the proposal distribution is particularly efficient when a large number of traps is used, since as the number of traps becomes larger, when using a uniform proposal it is more likely that a point is proposed close to a trap, where it is less likely to be accepted since only uncaptured individuals are proposed (as captured individuals are already known to be in the population).

4 | SIMULATION STUDY

In order to assess the performance of the model and the sensitivity to changes in study design and model parameters, we performed several simulations by varying the following parameters. First, we have simulated 100 or 200 individuals, whose locations are sampled from a conditional BivSC process (Equation 4) with $\gamma_1 = \gamma_2 = \exp(-100)$ and γ_{12} set to $\exp(-6)$ (high) or $\exp(-7.5)$ (low). Next, we set the number of traps to 10^2 or 15^2 , placed uniformly in the unit square and simulated capture histories using a baseline capture probability p_0 of 0.03 or 0.3 and $T = 10$ sampling occasions. For

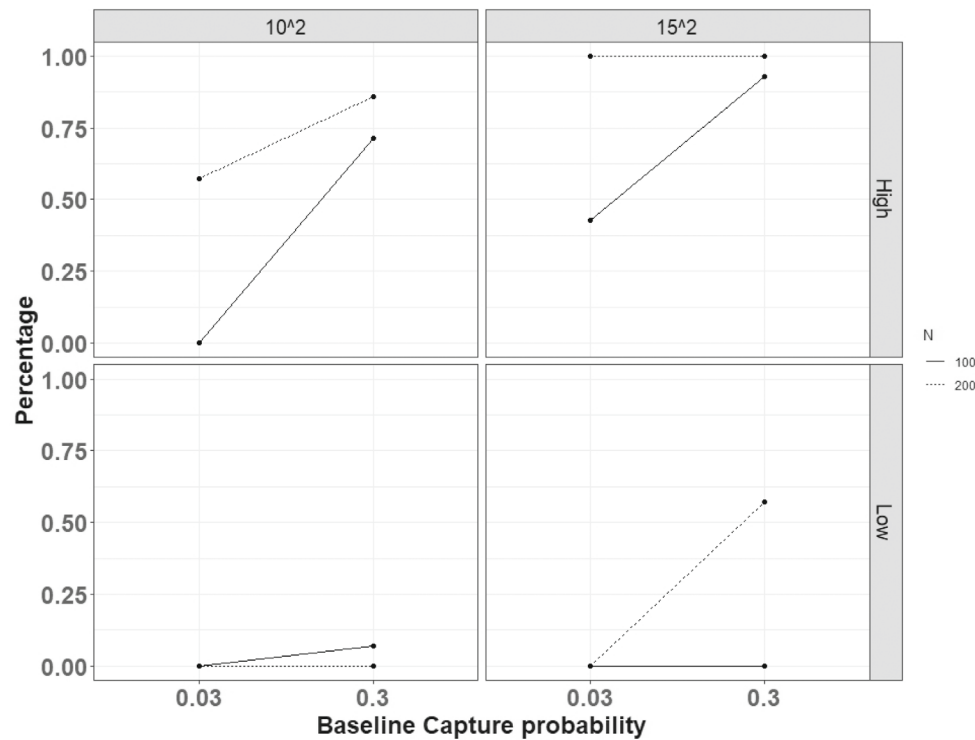


FIGURE 3 Simulation study. We vary the interaction strength in the rows and the number of traps in the columns. The dashed and solid lines represent the simulations with 200 and 100 individuals, respectively. The baseline capture probability is reported on the x-axis. In each plot, we report the percentage of times the true effect, $\log\left(\frac{\gamma_1}{\gamma_{12}}\right) < 0$ and $\log\left(\frac{\gamma_2}{\gamma_{12}}\right) < 0$, was identified over the 10 runs and over the 2 effects.

TABLE 1 Simulation study. Effective sample size per second for the MCMC algorithm with adaptation and without adaptation

Run	Without adaptation	With adaptation
1	0.032	0.308
2	0.013	0.250
3	0.009	0.219
4	0.015	0.288

each of the previous 16 scenarios, we have ran the model 10 times and reported the proportion of times the 95% posterior credible interval (PCI) of $\log\left(\frac{\gamma_1}{\gamma_{12}}\right)$ (or $\log\left(\frac{\gamma_2}{\gamma_{12}}\right)$) is below 0.

Results are presented in Figure 3. As expected, the power increases as the capture probability increases. In the low interaction scenario, the model is rarely able to detect the effect, except if number of traps, capture probability and number of individuals are at their highest setting. As expected, in the high interaction case, the power is always greater. Therefore, detection of interaction effects is possible only if the effect is strong and the capture probability is not very low.

We perform an additional simulation study to check the improvements to the mixing brought by the adaptation techniques described in Section 2.2. For simplicity, we work with only one species and without interaction. We have simulated the locations of 200 traps at random over the unit square and simulated the location of individuals from a homogeneous Poisson process with intensity $\beta = 300$ and capture histories using $p_0 = 0.02$ and $\sigma = 0.05$ and compared the MCMC algorithm without using any adaptation techniques to the one we have devised by comparing the effective sample size of the Markov chain obtained for N divided by the total time to run the algorithm. Results are presented in Table 1, where it can be seen that the effective sample size per second is substantially higher using adaptation.

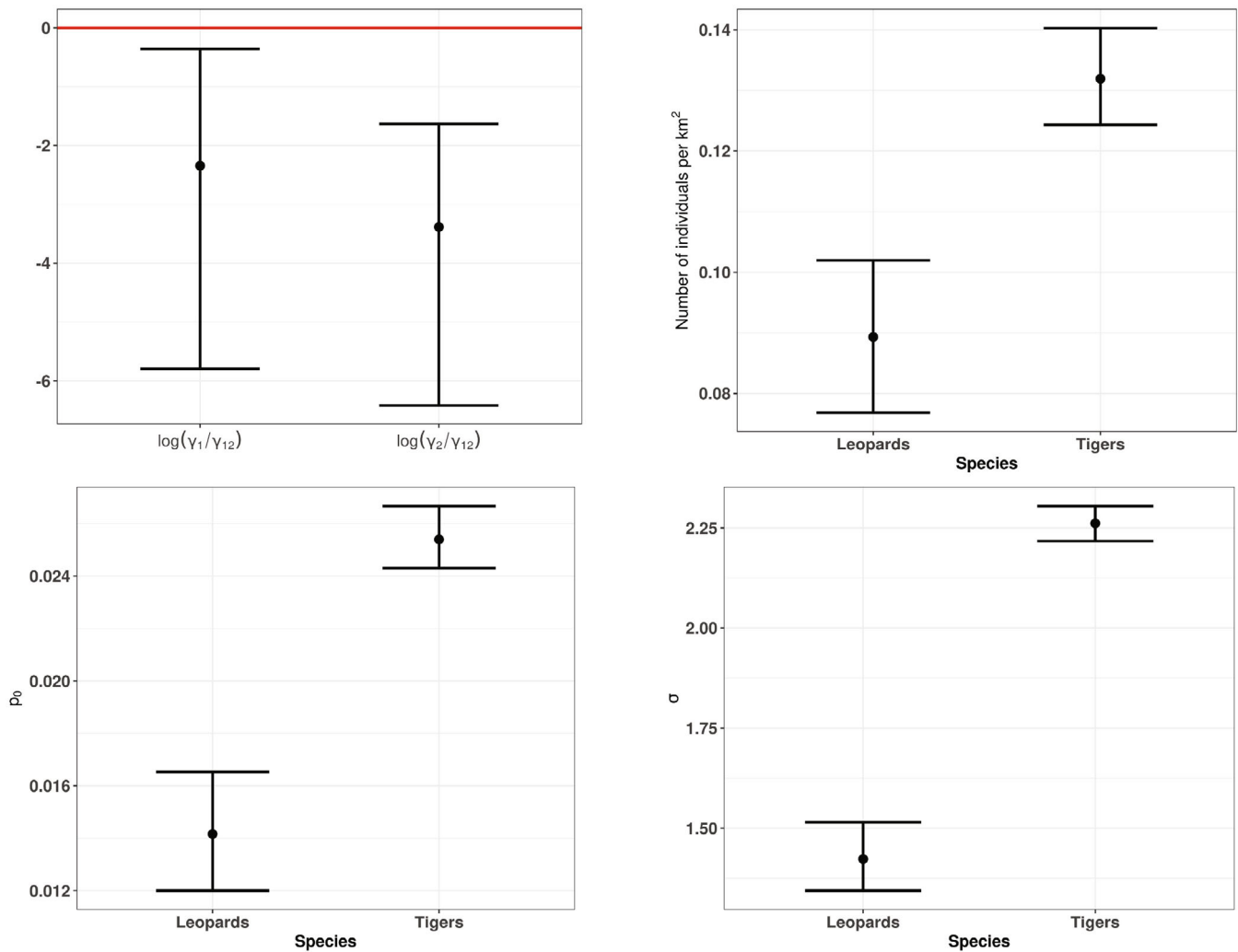


FIGURE 4 Case study. 95% PCI of the log ratios $\log\left(\frac{\gamma_1}{\gamma_{12}}\right)$ and $\log\left(\frac{\gamma_2}{\gamma_{12}}\right)$ (top left), where the index 1 refers to leopards and 2 refers to tigers. 95% PCI of the population densities (top right). 95% PCI of p_0 (bottom left). 95% PCI of σ (bottom right).

5 | CASE STUDY: CR DATA ON LEOPARDS AND TIGERS

We apply the model to two CR data of leopards (*Panthera pardus fusca*) and tigers (*Panthera tigris tigris*), collected at the Corbett Tiger Reserve in northern India's Uttarakhand State spanning an area of approximately 1100 km². The study area has been obtained by considering the region of the park of at most 8 km distant from the closest trap. Camera traps were placed in 529 locations. We define each sampling occasion as a 24 h interval in which the trap was active. The number of sampling occasions varies by trap but 90% of the traps were active for more than 40 days. All the sampling occasions belong to a single season. A total of 123 leopards and 231 tigers have been detected at least once during the study.

We set the following prior distributions $\log(\gamma) \sim N(-30, 20^2)$ and $\beta_{1,2} \sim \text{Gamma}(250, 100^2)$. The marginal prior distributions on N_1 and N_2 are obtained through simulations, performed via first sampling β and γ from their priors and next via generating a sample from a BivSC with the sampled parameters β and θ . The resulting distribution has 95% central interval equal to (94,450) and median equal to 232.

The posterior distributions of the log ratios between interaction parameters are shown in Figure 4. The distributions suggest that leopards and tigers tend to repulse each other more than they repulse individuals of the same species. This is expected since top carnivores often cause declines in the second-order carnivores by direct killing and out-competing them for resources (Kumar et al., 2019; Odden et al., 2010; Seidensticker, 1976). However, it is important to note that the simulation study suggests that power to detect this effect is expected to be small with low detection probabilities.

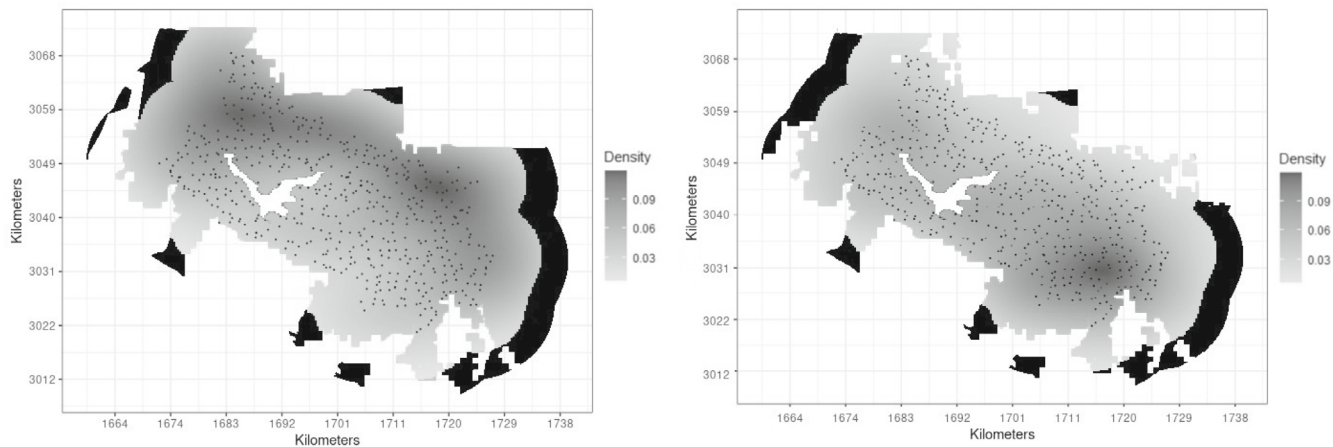


FIGURE 5 Case study. Maps of posterior densities of the locations of ACs of leopards (left) and tigers (right). Darker regions represent higher density. The black region is the part of the habitat not considered in the study area as is more than 8 km distant from the closest trap. The white area in the middle corresponds to a lake.

In Figure 4, we also present the estimates of the baseline capture probability, p_0 , and of σ . Tigers were observed to have a larger σ than leopards which is in agreement with larger energy demands of tigers that as a consequence have larger territory sizes compared to leopards (Kumar et al., 2019). Also, the camera trap deployment was designed to maximize captures of tigers and this is reflected in a higher capture probability of tigers compared to leopards.

In Figure 5, we show maps of posterior densities of the locations of ACs for both species. The prime habitats consisting of flat broad valleys, and undulating hills, that constitute the central, southern, and eastern parts of the Corbett Tiger Reserve have higher tiger densities, while leopards can achieve higher densities only in marginal tiger habitat that is, the more rugged Himalayan foot-hills on the northern boundary of the reserve.

6 | DISCUSSION

We have developed a SCR model for CR data collected on two or more species accounting for interactions between and within the species using a vector of PPs. Inference for the parameters of the PP is performed using the MCMH algorithm of Liang and Jin (2013), which we think provides the best balance in terms of speed, accuracy and scalability in the context of vectors of PPs. Additionally, we present an efficient inferential approach that adaptively learns the population size and the locations of the individuals never caught.

Although the model is developed for CR data, it is in theory possible to extend it to other sampling protocols where individual information is not present. The most popular sampling method for monitoring species with individuals that are not uniquely identifiable is the collection of count data. Chandler & Royle (2013) defined a Bayesian spatial model for count data, which works by inferring the latent capture histories of individuals as well as the latent individual ACs. However, this is highly inefficient in practice as reconstructing all these latent variables is computationally expensive and can lead to slow mixing. A possible alternative is to use a likelihood-free approach such as approximate Bayesian computation (Sisson et al., 2018).

We used the soft-core process for modeling the interaction between individuals, as this process is summarized using a single parameter and is able to model repulsion, but not attraction between points. It is generally difficult to model attraction using interaction PPs because if the interaction function ϕ_2 in Equation (2) is greater than 1 there is the potential for the number of points to explode. Being able to model attraction and repulsion at the same time is useful in the case of particular wildlife populations, as some species present attraction in small groups, such as for example within a family of individuals, and repulsion in bigger groups, such as between different families. This suggests for example a modeling approach where a pairwise and a triple-wise interaction function are considered together in the interaction function, and hence the process is bounded from exploding as the two interaction functions balance each other. An alternative modeling approach is to work with a *conditional* PP, since this issue does not arise in a conditional PP. However, assuming that the number of points is independent of the spatial structure of the process is often unrealistic.

We have assumed population closure and fixed ACs for the duration of the study. This short duration study does not motivate a model that relaxes these assumptions. However, in studies with longer duration or species that change AC frequently, a spatio-temporal approach where the locations of the points are allowed to change over time is worth considering. Although spatio-temporal models for log-Gaussian Cox process are available (Brix & Diggle, 2001; Brix & Møller, 2001), analogous models for Markov point process are not available.

ACKNOWLEDGMENT

We thank Ujjwal Kumar for assisting in the interpretation of the data and providing habitat masks used in the case study.

REFERENCES

- Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974;19(6):716-723.
- Andrieu C, Thoms J. A tutorial on adaptive MCMC. *Statistics and Computing*. 2008;18(4):343-373.
- Besag J. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*. 1978;47(2):77-92.
- Borchers D, Efford M. Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*. 2008;64(2):377-385.
- Borchers D, Fewster R. Spatial capture-recapture models. *Statistical Science*. 2016;31:219-232.
- Brix A, Diggle PJ. Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001;63(4):823-841.
- Brix A, Møller J. Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scandinavian Journal of Statistics*. 2001;28(3):471-488.
- Brownie C. *Statistical inference from band recovery data: A handbook*. United States Department of the Interior, Fish and Wildlife Service; 1985.
- Chandler RB, Royle JA. Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics*. 2013;7(2):936-954.
- Converse SJ, Royle JA. Dealing with incomplete and variable detectability in multi-year, multi-site monitoring of ecological populations. *Design and Analysis of Long-term Ecological Monitoring Studies*. 2012;426-442.
- Cressie NA. *Statistics for spatial data*. John Wiley and Sons, Inc.; 1993.
- Diana A, Matechou E, Griffin J, Johnston A. A hierarchical dependent Dirichlet process prior for modelling bird migration patterns in the UK. *The Annals of Applied Statistics*. 2020;14(1):473-493.
- Efford M. Density estimation in live-trapping studies. *Oikos*. 2004;106(3):598-610.
- Geyer CJ, Møller J. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*. 1994;1:359-373.
- Giordani P, Kohn R. Adaptive independent metropolis-Hastings by fast estimation of mixtures of normals. *Journal of Computational and Graphical Statistics*. 2010;19(2):243-259.
- Jensen JL, Møller J. Pseudolikelihood for exponential family models of spatial point processes. *The Annals of Applied Probability*. 1991;1(3):445-461.
- Kendall WS, Møller J. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*. 2000;32(3):844-865.
- Kumar U, Awasthi N, Qureshi Q, Jhala Y. Do conservation strategies that increase tiger populations have consequences for other wild carnivores like leopards? *Scientific Reports*. 2019;9(1):1-8.
- Liang F, Jin I-H. A Monte Carlo metropolis-Hastings algorithm for sampling from distributions with intractable normalizing constants. *Neural computation*. 2013;25(8):2199-2234.
- McLaughlin P, Bar H. A spatial capture-recapture model with attractions between individuals. *Environmetrics*. 2021;32(1):e2653.
- Møller J, Pettitt AN, Reeves R, Berthelsen KK. An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*. 2006;93(2):451-458.
- Møller J, Waagepetersen RP. *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC Press; 2003.
- Murray I, Ghahramani Z, & MacKay D. (2012). *MCMC for doubly-intractable distributions*. Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, UAI 2006.
- Neyman J, Scott EL. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1958;20(1):1-29.
- Nightingale GF, Illian JB, King R. Pairwise interaction point processes for modelling bivariate spatial point patterns in the presence of interaction uncertainty. *Journal of Environmental Statistics*. 2015;7.
- Odden M, Wegge P, Fredriksen T. Do tigers displace leopards? If so, why? *Ecological Research*. 2010;25(4):875-881.
- Ogata Y, Tanemura M. Likelihood analysis of spatial point patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1984;46(3):496-518.
- Pledger S, Efford M, Pollock K, Collazo J, Lyons J. *Stopover duration analysis with departure probability dependent on unknown time since arrival. Modeling demographic processes in marked populations*. Springer; 2009:349-363.
- Preston C. *Random fields*. Vol 534. Springer-Verlag; 1976.
- Rajala T, Murrell D, Olhede S. Detecting multivariate interactions in spatial point patterns with Gibbs models and variable selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2018;67(5):1237-1273.

- Reich BJ, Gardner B. A spatial capture-recapture model for territorial species. *Environmetrics*. 2014;25(8):630-637.
- Ripley BD. Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1977;39(2):172-192.
- Robbins H, Monro S. A stochastic approximation method. *The Annals of Mathematical Statistics*. 1951;22:400-407.
- Royle JA. N-mixture models for estimating population size from spatially replicated counts. *Biometrics*. 2004;60(1):108-115.
- Royle JA. Analysis of capture-recapture models with individual covariates using data augmentation. *Biometrics*. 2009;65(1):267-274.
- Royle JA, Chandler RB, Sollmann R, Gardner B. *Spatial capture-recapture*. Academic Press; 2013.
- Royle JA, Young KV. A hierarchical model for spatial capture-recapture data. *Ecology*. 2008;89(8):2281-2289.
- Ruelle D. *Statistical mechanics: Rigorous results*. World Scientific; 1969.
- Seidensticker J. On the ecological separation between tigers and leopards. *Biotropica*. 1976;1:225-234.
- Sisson SA, Fan Y, Beaumont M. *Handbook of approximate Bayesian computation*. CRC Press; 2018.
- Strauss DJ. A model for clustering. *Biometrika*. 1975;62(2):467-475.
- Waagepetersen R, Guan Y, Jalilian A, Mateu J. Analysis of multispecies point patterns by using multivariate log-Gaussian Cox processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2016;65(1):77-96.

How to cite this article: Diana, A., Matechou, E., Griffin, J. E., Jhala, Y., & Qureshi, Q. (2022). A vector of point processes for modeling interactions between and within species using capture-recapture data. *Environmetrics*, 33(8), e2781. <https://doi.org/10.1002/env.2781>

APPENDIX A

A.1 Simulation algorithm for point processes

We can easily extend the algorithm for sampling interaction point processes of Geyer and Møller (1994) by adding or deleting more than one points at a time. Let X be a point process with density f w.r.t. the homogeneous Poisson process with intensity 1. We can sample X with the following procedure.

Given the value at time m , $X_m = \{x_1, \dots, x_n\}$, generate $X_{m+1}|X_m$ by applying with equal probability one of the following steps:

1. Move: Draw $i \sim \text{Uniform}(\{1, \dots, n\})$, propose $\xi \sim q_i(X_m, \cdot)$, where q_i is the proposal distribution for the i th point and set

$$X_{m+1} = \begin{cases} \{x_1, \dots, x_{i-1}, \xi, x_{i+1}, \dots, x_n\} & \text{if } U([0, 1]) \leq r_i(X_m, \xi) \\ X_m & \text{otherwise,} \end{cases}$$

where $r_i(X, \xi)$ is the Hastings ratio

$$r_i(X, \xi) = \frac{f((X \setminus x_i) \cup \xi) q_i(\{x_1, \dots, x_{i-1}, \xi, x_{i+1}, \dots, x_n\}, x_i)}{f(X) q_i(X, \xi)}.$$

2. Add: Propose $N \sim \text{Pois}(\lambda_1)$ new points $(\xi_1, \dots, \xi_N)$, with $\xi_i \sim q_b(\cdot)$ independently and set

$$X_{m+1} = \begin{cases} X_m \cup (\xi_1, \dots, \xi_N) & \text{if } U([0, 1]) \leq r_b(X_m, \xi) \\ X_m & \text{otherwise,} \end{cases}$$

where $r_b(X, \xi_1, \dots, \xi_N)$ is the Hastings ratio

$$r_b(X, (\xi_1, \dots, \xi_N)) = \frac{f(X \cup (\xi_1, \dots, \xi_N)) \text{Pois}(N, \lambda_2) \prod_{i=1}^N q_d(\xi_i)}{f(X) \text{Pois}(N, \lambda_1) \prod_{i=1}^N q_b(\xi_i)}.$$

3. Delete: Sample $N \sim \text{Pois}(\lambda_2)$ existing points x_i with probability proportional to $q_d(\cdot)$ and set

$$X_{m+1} = \begin{cases} X_m \setminus (x_1, \dots, x_N) & \text{if } U([0, 1]) \leq r_d(X_m, (x_1, \dots, x_N)) \\ X_m & \text{otherwise,} \end{cases}$$

where $r_d(X, \xi)$ is the Hastings ratio

$$r_d(X, \xi) = \frac{1}{r_b(X \setminus (x_1, \dots, x_N), (x_1, \dots, x_N))}.$$

The generalization to a pair of point processes is straightforward.

A.2 MCMC algorithm

A.2.1 Update s_{il}

Similarly to the previous section, we update the pair of point processes $(\mathbf{s}^1, \mathbf{s}^2) = (s_1^1, \dots, s_{N_1}^1), (s_1^2, \dots, s_{N_2}^2)$ by using with equal probability a move/add/delete step for each of the two processes.

The posterior distribution of $(\mathbf{s}^1, \mathbf{s}^2)$ can be written as

$$p(\mathbf{s}^1, \mathbf{s}^2 | \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_{12}, H^1, H^2, p_0^1, \sigma_1, p_0^2, \sigma_2) \propto p(\mathbf{s}^1, \mathbf{s}^2 | \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_{12}) p(H^1 | \mathbf{s}^1, p_0^1, \alpha^1) / p(H^2 | \mathbf{s}^2, p_0^2, \alpha^2),$$

where

$$p(\mathbf{s}^1, \mathbf{s}^2 | \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_{12}) \propto \beta_1^{N_1} \beta_2^{N_2} \prod_{i < j} 1 - \exp\left(-\frac{(s_i^1 - s_j^1)^2}{\gamma_1}\right) \prod_{i < j} 1 - \exp\left(-\frac{(s_i^2 - s_j^2)^2}{\gamma_2}\right) \prod_{i,j} 1 - \exp\left(-\frac{(s_i^1 - s_j^2)^2}{\gamma_{12}}\right), \quad (\text{A1})$$

and $p(H^l | s_i^l)$ is the probability of the captures locations, which is equal to

$$\left(\prod_{i=1}^D \prod_{k=1}^K p(H_{ik}^l | s_i^l, x_k, p_0^l, \alpha^l) \right) \binom{N}{N-D} \prod_{i=1}^{N_D} \prod_{k=1}^K p(H_0 | s_i^l, x_k, p_0^l, \alpha^l),$$

where H_0 is a capture history of an individual never captured. The first term is the contribution of the captured individual and the second is the contribution of the uncaptured individuals.

We can update $(\mathbf{s}^1, \mathbf{s}^2)$ using the algorithm explained in the previous section by taking $f(\mathbf{s}^1) \propto p(\mathbf{s}^1 | \mathbf{s}^2, \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_{12}) p(H^1 | \mathbf{s}^1, p_0^1, \alpha^1)$.

We choose $\lambda_1 = \lambda_2$ and we select the points to delete with equal probability, which implies $q_d(\xi_1, \dots, \xi_n) = \frac{1}{n}$.

At each step, we update the point process $\mathbf{s}^1$ by choosing one of the next three steps with equal probability ($\mathbf{s}^2$ is updated analogously):

- Move

For each point s_i , propose $\xi \sim N(s_i, \text{diag}(\sigma_1^2, \sigma_2^2))$ and accept with probability

$$\frac{p((\mathbf{s}^1 \setminus s_i^1) \cup \xi, \mathbf{s}^2 | \gamma) p(H^1 | (\mathbf{s}^1 \setminus s_i^1) \cup \xi, p)}{p(\mathbf{s}^1, \mathbf{s}^2 | \gamma) p(H^1 | \mathbf{s}^1, \mathbf{s}^2)} = \frac{f((\mathbf{s}^1 \setminus s_i^1) \cup \xi, \mathbf{s}^1)}{f(\mathbf{s}^1, \mathbf{s}^2)} \cdot \frac{p(H_i^1 | \xi, p)}{p(H_i^1 | s_i^1, p)}.$$

- Add

Propose $m \sim \text{Pois}(\lambda_1)$ new points $(\xi_1, \dots, \xi_m)$, with $\xi_i \sim q_b(\cdot)$ independently and accept the points with probability

$$\frac{p(\mathbf{s}^1 \cup (\xi_1, \dots, \xi_m) | \mathbf{s}^2, \beta_1, \gamma_1, \gamma_2, \gamma_{12})}{p(\mathbf{s}^1 | \mathbf{s}^2, \beta_1, \gamma_1, \gamma_2, \gamma_{12})} \frac{\left(\prod_{l=1}^m \frac{(N_1+l)}{(N_1+l-D_1)} p(H_0 | \xi_l, p_0^1, \sigma) \right)}{\prod_{l=1}^m q_b(\xi_l) \prod_{l=1}^m (N+l)},$$

where H_0 is the capture history of an individual never captured.

- Delete

Sample $m \sim \text{Pois}(\lambda_2)$ existing points x_i and set

$$\frac{p(\mathbf{s}^1 \setminus (\xi_1, \dots, \xi_m) | \mathbf{s}^2, \beta_1, \gamma_1, \gamma_2, \gamma_{12})}{p(\mathbf{s}^1 | \mathbf{s}^2, \beta_1, \gamma_1, \gamma_2, \gamma_{12})} \frac{\prod_{l=1}^m q_b(\xi_l)}{\left(\prod_{l=1}^m \frac{(N_1 - m + l)}{(N_1 - m + l - D_1)} p(H_0 | \xi_l, p_0^l, \sigma) \right) \prod_{l=1}^m (N - m + l)}.$$

A.2.2 Update γ_j and β_j

These variables are updated using the MCMH algorithm described in Section 2.2.

A.2.3 Update p_0^l and α^l

We update these variables using a simple Metropolis–Hastings update.

A.2.4 Adaptation of the proposal parameters λ and q_b

- λ .

Every k iterations, we compute the acceptance rate $\hat{p}$ and we update λ using the Robbins–Monro update as

$$\lambda^* = \lambda + (\hat{p} - p^*) \frac{w}{k_c},$$

where k_c is the number of updates performed so far and p^* is the targeted acceptance rate.

The values used were $w = 10$, $p^* = 0.3$ and $k = 50$.

- q_b

We term as $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ the set of all the accepted points during the burn-in. At the end of the burn-in phase, we fit several mixture models $\sum_{k=1}^K w_k N(\mu_k, \Sigma_k)$ on the set of all the accepted points $\mathbf{x}$ for various values of K and we choose the model with the best AIC. We set $q_b = \sum_{k=1}^{\hat{K}} w_k N(\mu_k, \Sigma_k)$ where $\hat{K}$ is the optimal value of K .