Response to "Improving abundance estimation by combining capture-recapture and presence-absence data: example with a large carnivore"

Jack H. W. Thomas[1]

Simon J. Bonner[2]

Laura L. E. Cowen[1]

[1]Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 2Y2

[2]Department of Statistical and Actuarial Sciences, Western University, London, ON, N6A 3K7

SUMMARY.

1. A key area of research in ecological statistics involves combining data sources from multiple streams to improve population estimates. One such model attempts to integrate capture-recapture and presence-absence data to estimate the population size of Eurasian lynx in the Jura Mountains, eastern France.

2. This model has been observed to underestimate population sizes. We conducted an extensive simulation study to evaluate the model's performance. We describe our methods for generation of simultaneous capture-recapture and presence-absence data and demonstrate that the model is flawed.

3. Solution: Finally, we give an outline of our hypothesis on why the model underestimates population sizes.

Keywords: Bayesian inference, data integration, ecological model, ecosystem, endangered species, integrated population models, population dynamics, population size, wildlife management

1   Introduction

Accurately estimating the size of a population from ecological data is crucial for wildlife management. To address endangered species recovery, resource allocation decisions are often made based on population size and conservation decisions are triaged (Wiedenfeld et al., 2021). Traditional methods for monitoring wildlife populations generally fall into one of two categories. Presence-absence or occupancy studies (Bailey et al., 2014) attempt to find signs of presence in the field (hair, tracks, scat, etc.). These studies are inexpensive (MacKenzie and Nichols, 2004) and can

generate large datasets, but presence-absence data alone provide limited information about abundance (Efford, 2014). Alternatively, capture-recapture studies follow animals over time if individuals can be identified via tags or natural markings. This provides more information about the population including direct information about abundance, but mark-recapture methods are generally more costly, both in money and time (Lieury et al., 2017) and can be more invasive to the species being studied. The analysis of sparse data is particularly a concern as data are often limited in studies of the most threatened, rare, or elusive populations (Tourani et al., 2020). Methods to deal with data sparcity include for example: aggregating over sampling elements by pooling over species, locations, or sampling periods (Litt and Steidl, 2010); implementing time-stratification into a hierarchical Bayesian model (Oldemeyer et al., 2018); or using an informative prior (Fraysse et al., 2023) using for example traditional ecological knowledge (Girondot and Rizzo, 2015). More recently, efforts have been made to combine data from different sources to obtain more accurate estimates for population parameters from sparse data (e.g., Murphy et al., 2018; Morera-Pujol et al., 2023). We focus on one approach for combining presence-absence and mark-recapture data suggested by Blanc et al. (2014), and show that the method suffers from a conceptual flaw which leads to erroneous estimates of the size of a population except in extreme circumstances. We offer this as a warning to both the statistical and wildlife management communities and emphasize that care must be taken when combining data collected by different means.

The first methods we are aware of to combine data on the same population collected from different sources were the integrated population models (IPMs) of Besbeas et al. (2002). The specific model combined census data with ring-recovery data using the Kalman filter to construct a joint likelihood. The authors noted that the same approach could be applied to combine census data with capture-recapture data, and Besbeas et al. (2005) extended the model using multivariate normal approximations to combine the likelihoods from separate data sources. In a recent overview of IPM, Zipkin and Saunders (2018) argued that local-scale studies are often not broad enough to cover an entire population and make accurate estimates of abundance. They conclude that IPMs make the best use of all the available data by borrowing information from other sub-models to produce the most accurate population estimates.

More recently, variations of the IPM have been proposed to combine different types of data with different statistical methods. Renner et al. (2019) combined data from three sources to to estimate the abundance of Eurasian lynx in the Jura Mountains, eastern France: sightings of individuals, interference with livestock, and pictures of lynx taken by camera traps. The combined model improved population size estimation by incorporating an area-interaction into the likelihood (Renner et al., 2019) but required the assumption that occupancy is fixed (i.e., sites are always occupied or always unoccupied). The model of Renner et al. (2019) also incorporates spatial dependence to avoid the assumption that observations at different locations are independent. Strebel et al. (2022) estimated abundances for two species of bird in Switzerland, the common nightingale in the lowlands and rock bunting birds in the Southern Alps, by combining data from four sources without any marked individuals through a binomial N-mixture model: count data, presence-absence data, presence-only data, and absence-only data. Strebel et al. (2022) found that their approach

improved estimates when any single source of data is sparse, but cautioned against the use of some sources, particularly presence-only data, as bias might be introduced if some sites are preferentially sampled.

Specific to combining capture-recapture with presence absence-data, Conroy et al. (2008) considered a two-stage experiment in which an occupancy study is conducted first and then followed by mark-recapture restricted to regions in which the species of interest was detected. Their specific application assumes that individuals do not move between the regions, so that the mark-recapture samples are independent. Jiménez et al. (2022) combined presence-absence data with capture-recapture data to estimate the abundance of stone marten. In this case, the data all came from camera traps. However, only a portion of the population had sufficient markings to be individually identified meaning that some photographs can be connected to construct capture histories while others provide only presence-absence data. More generally, Chandler and Clark (2014) developed a model that explicitly combines mark-recapture and presence-absence data by modeling the movements of all individuals within a population. Latent variables are introduced to match individuals, whether marked or unmarked, with each detection in the presence-absence data and the model is fit in the Bayesian framework using Markov chain Monte Carlo to sample from the posterior distribution including these variables. The specific application is somewhat limited in that only one type of data is collected on each occasion (e.g., alternating between mark-recapture and occupancy studies). The authors also note that the method is "computationally demanding, especially when the number of individuals in the population is large" (Chandler and Clark, 2014, pg. 1359). All of these authors note that the inclusion of presence-absence data produced more accurate and precise estimates of the population size than the capture-recapture data alone. However, it is clear that further development in this area required to create widely applicable methods that overcome the computational complexities.

Here, we focus on the method developed by Blanc et al. (2014) to estimate the size of a population from combined presence-absence and capture-recapture data. The method attempts to connect the two sources of data by equating the probability that a site is occupied and the probability that the population size is greater than zero. The researchers applied their method to estimate the population size of Eurasian lynx in the Jura Mountains, eastern France. Capture-recapture data were obtained using camera traps, and individuals were identifiable by natural marks. Observers also collected separate presence-absence data through extensive surveys of hair samples, tracks, scat, and/or livestock and wildlife killed. A total of nine unique individuals were identified during the study, and the final estimate of the population size was 10, after rounding, (95% CI: 9.0; 13.0). In comparison, their model of the capture-recapture data alone produced an estimate of 14.46 (95% CI: 9.0; 35.0) and previous spatially explicit capture-recapture analysis by a subset of the authors (Blanc et al., 2013) produced an estimate of 12.04 (95% CI: 9.0; 18) individuals. However, Blanc et al. (2014) did not provide results from either a mathematical analysis of the model or a simulation study to assess the performance of their combined data approach.

We believe that Blanc et al. (2014)'s method is flawed and systematically underestimates population sizes. In a previous study that included this method, Jahid et al. (2022) compared results from six methods to estimate the

population size of grizzly bears in the Rocky Mountains, Alberta, using camera trap (presence-absence) data, hair trap (DNA capture-recapture) data, or both. These six models were:

1. Closed capture-recapture model, which is the traditional approach of Otis et al. (1978).

2. Combined model of Blanc et al. (2014).

3. Spatially explicit capture-recapture (SECR) model from detection data of the hair traps of Efford (2004).

4. Combined detection data from hair traps and detection from camera traps model (SECR-O) from Tourani et al. (2020).

5. Model SECR with sex as a covariate.

6. Model SECR-O with sex as a covariate.

In their analysis, Jahid et al. (2022) found that the method of Blanc et al. (2014) produced results that were very different from the other methods with a much lower estimate of abundance and a very small posterior standard deviation. The capture-recapture data contain records of 37 individually identified bears over 8 occasions. Of these, 19 were captured only once, and 13 were captured twice. The estimate of the population size from the method of Blanc et al. (2014) was only 40 (SD = 1.56) while the capture-recapture-only model estimated 73.36 bears (SD = 31.84), and the SECR model estimated 380.28 bears (SD = 87.92). Although, Jahid et al. (2022) questioned the validity of the estimate from Blanc et al. (2014)'s model, they did not identify the source of the difference or determine whether this was due to a flaw in the model or a violation of assumptions specific to the data set. We explain the reason for the discrepancy and show mathematically and through simulation that Blanc et al. (2014)'s method contains a systematic error that leads to erroneous estimates of abundance.

2 Materials and methods

2.1 Simulation Study

The method developed by Blanc et al. (2014) starts by assuming that the population over the entire study area N is a realization of a Poisson distribution with rate parameter λ. The probability that a new population would have size n is

$$P(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}. \tag{1}$$

Next, Blanc et al. (2014) model presence-absence at each site with a Bernoulli random variable. They let $Z_i = 1$ and $Z_i = 0$ to indicate if site i is occupied or not occupied, respectively, and denote the probability that a site i is occupied

as

$$\psi_i = P(Z_i = 1). \tag{2}$$

113 To combine the two data types, Blanc et al. (2014) equate the probability that a site is occupied with the probability

114 that there is at least one animal in the overall population:

$$P(Z_i = 1) = P(N > 0). \tag{3}$$

115 Assuming that N follows (1), Blanc et al. (2014) conclude that:

$$P(N > 0) = 1 - P(N = 0) = 1 - e^{-\lambda}. \tag{4}$$

116 Using (2), they define:

$$\psi_i = 1 - e^{-\lambda}. \tag{5}$$

117 Solving for λ, they obtain:

$$\lambda = -\log(1 - \psi_i). \tag{6}$$

118 This is the key equation of Blanc et al. (2014), which connects the capture-recapture and presence-absence components

119 of the model together. The problem is that it operates across two different scales by linking site specific presence-

120 absence, ψ_i, with the overall population size, N. As discussed in Efford and Dawson (2012), the link between

121 presence-absence of individual sites and the total abundance depends heavily on the movement of individuals and the

122 relative sizes of the sampling sites and the individuals' ranges. If individuals have large ranges and move quickly then

123 they may leave signs in multiple sites in a single sampling period. In the limit as either the animal speed or the length

124 of the period increases, only one individual would be needed to occupy all sites on a single occasion. At the opposite

125 end of the spectrum, individuals will only occupy one site on each occasion if their home ranges are small relative to

126 the size of the sampling sites or if they move slowly.

127   We have assessed the model of Blanc et al. (2014) by conducting an extensive simulation study. On each replicate,

128 we simulated trap locations and a population of individuals and derived both the capture-recapture and presence-

129 absence data from these individuals. The first step to generating spatial capture-recapture and presence-absence data

130 is defining home ranges. As in Efford et al. (2000), we assumed animals in a population tend to stay within a certain

131 range, denoted as their home range. We assumed the study area to be a unit square for simplicity. Given a fixed pop-

132 ulation size (N), we used a Poisson process to generate home range centres (X_i, Y_i), where $X_i, Y_i \stackrel{iid}{\sim} \text{Uniform}(0, 1)$

133 for individuals $i = 1, 2, \ldots, N$.

2.2 Generating Capture-Recapture Data

Capture-recapture data were simulated to mimic camera traps. Individuals interact with the camera traps and are captured based on probabilities. Given a fixed number of cameras (m_1), we generated their locations in the study area, $(X_j^{(c)}, Y_j^{(c)})$, using a Poisson process with $X_j^{(c)}, Y_j^{(c)} \overset{iid}{\sim}$ Uniform$(0, 1)$ for camera traps $j = 1, 2, \ldots, m_1$. We assumed that the probability an individual is detected by a camera is related to the distance between a home range centre and a camera location as well as the amount of movement the species exhibits. Specifically, we implemented a half-normal detection function so that the probability individual i is detected by trap j on each occasion is

$$p_{ij} = p e^{(-d_{ij}^2/\tau^2)} \tag{7}$$

where $p \in (0, 1)$ is the baseline capture probability for a camera trap at distance 0 from the home range centre, $\tau > 0$ is the movement parameter for the species, and d_{ij} is the distance between the home-range centre i and the camera location j. When τ is large, animals move throughout the entire study area and in the limit as $\tau \to \infty$ $p_{ij} \to p$ for all i and j. On the other hand, when τ is close to zero, animals do not move significantly from their home-range centres and $p_{ij} \approx 0$, unless the location of camera j happens to coincide with the home-range centre. Once all the camera traps are placed, we simulate the interactions of individuals with the camera traps. We denote the number of capture-recapture sampling occasions by K_1 and generate a three-dimensional array in which cell i, j, k identifies whether individual i was captured by trap j at time k. This three-dimensional array contains the capture events $C_{ijk} \sim$ Bernoulli(p_{ij}).

2.3 Generating Presence-Absence Data

Presence-absence data generation starts with dividing the study area into regions. We do this simply by dividing the square study area into a grid of equally sized square cells. Following the terminology of Blanc et al. (2014), each grid cell is henceforth referred to as a site. We let m_2 denote the number of sites and K_2 the number of times each site is searched for signs of presence. To model the presence-absence observations, we assume that the number of signs left by an individual within a site follows a Poisson distribution dependent on the proportion of time the individual spends within the site, ψ_{ij}, and the average number of signs it deposits per sampling period, δ. Let Z_{ijk} be the number of signs that individual i deposits in cell j during period k and $D_{ijk} = 1(Z_{ijk} > 0)$. Assuming that animals move about their home ranges according to a circular bivariate normal distribution with variance τ^2 (movement parameter squared), the proportion of time that individual i spends within cell j is

$$\psi_{ij} = \int_{x_{0j}}^{x_{1j}} \int_{y_{0j}}^{y_{1j}} f(x, y | x_i, y_i, \tau) \, dx \, dy \tag{8}$$

6

where $x_{0j}, y_{0j}, x_{1j}, y_{1j}$ denote the bounds of the cell. The mean number of signs the individual deposits in cell j is then equal to $\delta\psi_{ij}$ so that

$$Z_{ijk} \sim \text{Poisson}(\delta\psi_{ij}) \tag{9}$$

and

$$\mathbb{P}(D_{ijk} = 1) = \mathbb{P}(Z_{ijk} > 0) = 1 - e^{-\delta\psi_{ij}}. \tag{10}$$

The final indicator of presence-absence in cell j during period k is then

$$D_{\cdot jk} = 1\left(\sum_{i=1}^{N} D_{ijk} > 0\right) \tag{11}$$

which follows a Bernoulli distribution with

$$P(D_{\cdot jk} = 1) = 1 - \prod_{i=1}^{n} e^{-\delta\psi_{ij}} = 1 - \exp\left(-\delta\sum_{i=1}^{n}\psi_{ij}\right). \tag{12}$$

Note that the probability of detecting presence within a cell increases as either the number of signs deposited (δ) or the proportion of time individuals spend within the cell (ψ_{ij}) increase.

2.4   Testing the Model

Blanc et al. (2014) provided supplementary code to implement their model in JAGS (Just Another Gibbs Sampler) using the R package rjags as an interface (Denwood, 2016). We used R version 4.1.2 (R Core Team, 2021) to implement our simulation. We compared the averaged population sizes with the true population sizes by computing the bias and mean square error of the point estimates (posterior mean), the bias of the posterior standard errors, and the coverage probability and width of the 95% credible intervals. We defined standard deviation bias to be the difference between Blanc et al. (2014)'s model output posterior standard deviations for the population size and our own calculated standard deviations of the population estimates based on all simulation replicates. We define the coverage probability to be the percentage of cases where the true population value was within the 95% credible interval reported by the model output. The generated capture-recapture and presence-absence data were used as input for the supplementary code from Blanc et al. (2014). Parameters for each simulation scenario were the true population size (N), movement parameter (τ), number of sites in the study area (m_1), number of capture-recapture sampling occasions (K_1), number of presence-absence sampling occasions (K_2), and the number of camera traps in the study (m_2).

We present two case studies, A and B. The aim of case study A was to determine if the performance of the estimates from the model of Blanc et al. (2014) depend on the true population size, N. To study this, we varied N, τ, m_1, and m_2, while holding p, δ, K_1, and K_2 as constants. We formed different scenarios from all possible

7

Table 1: Case study A parameter settings to assess the effect of N, τ, m_1, and m_2 on the population size estimates.

Description	Parameter	Parameter settings
Population size	N	$10, 50, 100, 200, 500$
Movement parameter	τ	$0.15, 0.25, 0.35, 0.45, 0.55$
Capture probability	p	0.75
Mean sign deposition	δ	4
Number of camera traps	m_1	$5, 15, 30$
Number of sites	m_2	$64, 81, 100$
Number of capture-recapture sampling occasions	K_1	7
Number of presence-absence sampling occasions	K_2	5

Table 2: Case study B parameter settings to assess the effect of p and δ on the population size estimates.

Description	Parameter	Parameter settings
Population size	N	$10, 500$
Movement parameter	τ	$0.15, 0.25, 0.35, 0.45, 0.55$
Capture probability	p	$0.1, 0.25, 0.5, 0.75$
Mean sign deposition	δ	$1, 4, 8, 16$
Number of camera traps	m_1	$5, 15, 30$
Number of sites	m_2	64
Number of capture-recapture sampling occasions	K_1	7
Number of presence-absence sampling occasions	K_2	5

combinations of the parameter values in Table 1. For case study B, our goal was to determine if the behaviour of the model changed from case study A by varying p, and δ. Case study B only had two different values of N, while varying τ, p, and δ, and holding K_1, m_1, m_2 and K_2 as constants. We formed different scenarios from all possible combinations of the parameter values in Table 2. In both case studies, we generated 50 replicate data sets per scenario. Note that some replicates produced no captures of individuals. This happens particularly when the capture probability is low, there are few traps, and the movement parameter is small so that individuals are unlikely to encounter traps and unlikely to be captured if they do. In these cases, we only analysed the data sets that produced some captures since it is not possible to estimate abundance from presence-absence data alone. Summaries of the unique number of individuals captured and the number of presence-absence signs detected in the different scenarios are provided in Tables A1-A4 in Web Appendix B in the online Supporting Information.

3 Results

The first key result of our study is that the estimate of the population size depended only on the number of individuals captured at least one time, not the true population size, how often individuals were caprured, or the value of any of the parameters (see Figure 1). In fact, the estimate of the population size was almost exactly linearly related to the number of animals captured with correlation $\rho = 0.999$. This held even across different scenarios with very different numbers of individuals captured but the same true population size. The estimates of the population size did approach
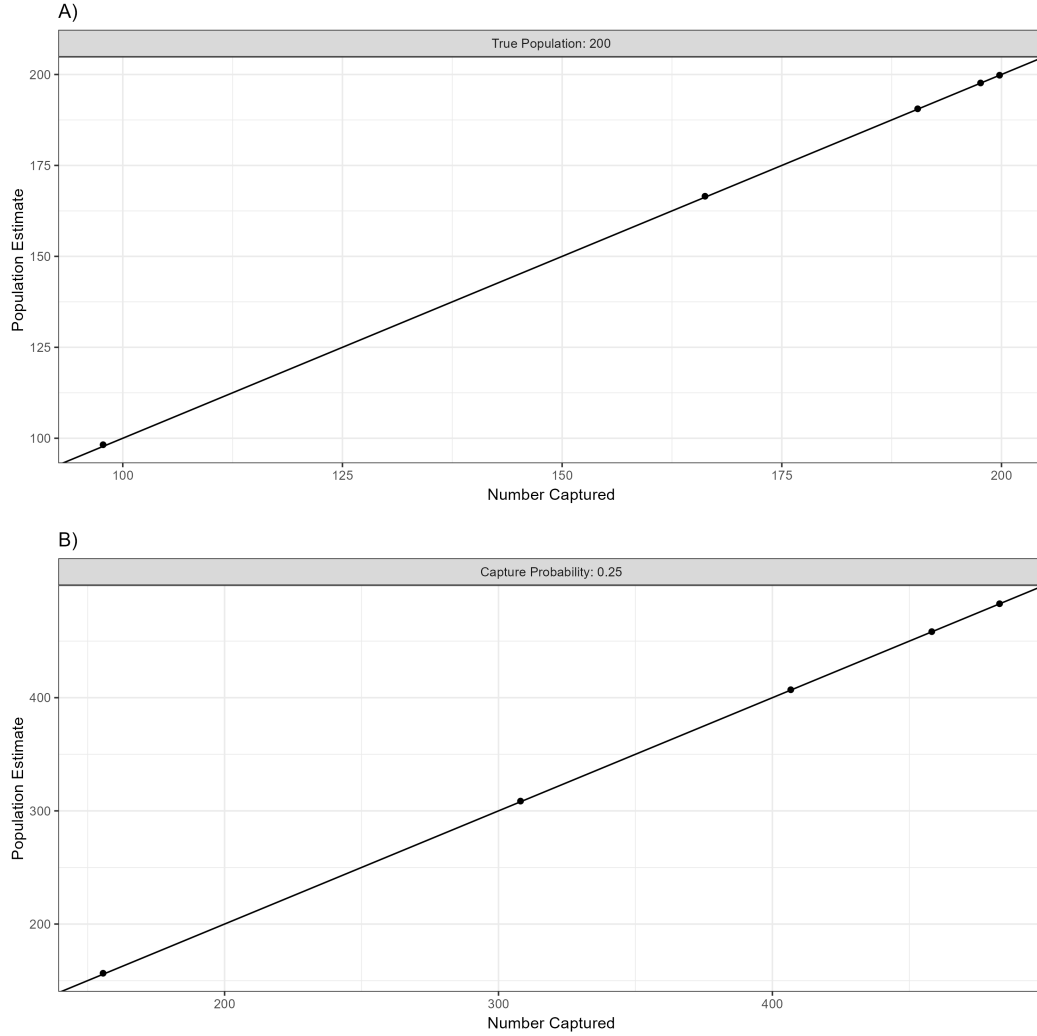
Figure 1: Correlation of the estimated population size versus number of observed individuals (Number Captured) during the study period for selected parameter sets from case studies A and B. A) had: N=200, m_1=5, m_2=100,. B) had: N=500, m_2=5, m_2=64. Both A) and B) had: movement parameter $\tau \in (0.15, 0.25, 0.35, 0.45, 0.55)$, number of capture-recapture sampling occasions $K_1 = 7$, and number of presence-absence sampling occasions $K_2 = 5$. Each point on A) and B) represent a value of τ.

the truth as the capture probability increased to 0.75, but this occurred only because the probability that the entire population was captured at least once became very close to 1 (effectively, the population was censused). Figure 2 shows that the same result held as τ increased from 0.15 to 0.55. Increasing animal movement (τ), increases the probability that all individuals are captured at least once, and eventually the estimate coincides with the truth. Varying the mean number of signs deposited per period (δ) did not change the near one-to-one correspondence of the number of captured individuals and the population estimate from the model (Figure 3). Our conclusion is that the method of Blanc et al. (2014) always underestimates the true population size except for extreme situations in which all individuals are captured.

Our second key conclusion is that the estimate of the population size is completely independent of the parameters that control how signs are deposited. Figure 4 shows how the mean estimate population size varies with δ, τ, and p. The results show that the mean estimate is completely determined by τ and p and does not vary at all with δ. This implies that the estimate of N does not depend on the parameters of the presence-absence surveys once the parameters controlling the mark-recapture portion of the study are set, and, hence, that the estimate of abundance is unaffected by the presence-absence data.

Credible intervals and estimates of uncertainty based on the method of Blanc et al. (2014) also performed very poorly. The coverage of the 95% credible intervals increased systematically with τ when p was large enough for the entire population to be captured ($p > 0.5$) ranging from a value of only 0.1 when τ was small to almost 1.0 when τ was large (see Figure 5). Similarly, posterior standard deviations can vastly underestimate the true uncertainty in the estimate of the population size (see Figures A1 and A2 in Web Appendix 1 in the online Supporting Information). The posterior standard deviation underestimated the standard deviation of the estimates of population size from the simulated data sets in all scenarios, and the the magnitude of the mean relative bias was as high in magnitude as 0.95. The bias was generally reduced as p and τ both increased, but the effect was small and the mean relative bias remained below -0.5 in all cases.
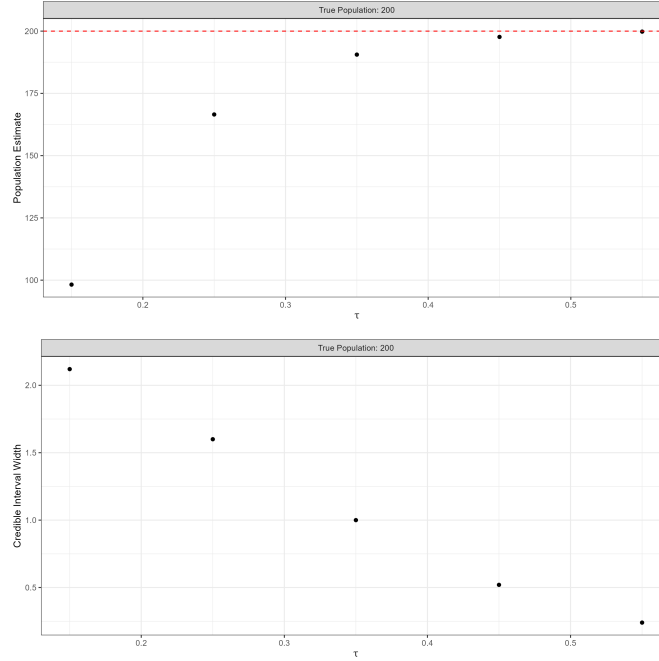
Figure 2: Population estimate versus τ (top) and credible interval width versus τ (bottom). Parameters selected from case study A. True population size was set to 200 with capture probability set to $p = 0.75$, number of camera traps $m_1 = 5$, and number of sites $m_2 = 100$. The dotted red line in the top panel represents the true population.

4 Discussion

The results of our simulation study show that the estimates of abundance produced by Blanc et al. (2014)'s model do not accurately reflect the true population size. The estimate of abundance was always equal to, or just slightly greater than, the number of individuals captured during the study. This quantity was determined by both the true population size and the overall probability that an individual was captured during the study, which was determined by the movement parameter, τ, the baseline capture probability p, and the number of capture occasions K_1. The number of individuals captured during a study must always be less than the true population size, barring violations of the basic mark-recapture assumptions like tag-loss, and so the method of Blanc et al. (2014) was severely negatively biased in most of the scenarios. Further to this, the uncertainty was also vastly underestimated. The posterior standard deviations showed severe negative bias in all scenarios, and the 95% credible intervals computed from Blanc et al. (2014)'s model failed to capture the true population size adequately. For the scenarios in which $p < 0.5$ and $\tau \leq 0.55$ the coverage was as low as 0% and increased only as high as 75%, never reaching the nominal coverage rate (Figure 5).

The only scenarios in which Blanc et al. (2014)'s model did provide accurate estimates of the population size were those in which the movement parameter was large, $\tau = 0.55$ meaning, effectively, that the study comprises only one site because individuals move through the entire area. However, the probability that individuals are captured is almost

11

1 in these scenarios, meaning that almost all individuals are captured and the population is effectively censused. Moreover, the model still did not provide valid estimates of the uncertainty in these scenarios providing posterior standard deviations that still underestimated the true standard deviation of the estimator and credible intervals that failed to meet the nominal coverage probability.

Our other finding is that, in contrast to the aim of Blanc et al. (2014) which is to combine information from mark-recapture and presence-absence data, the presence-absence data does not contribute to the estimation of the population size. The number of presence detections does show a positive relationship with the estimate of population size, but this is driven simply by the relationship between the different types of data. Increasing the movement parameter, τ, increases both the probability that individuals are captured and the number of sites with signs of presence since individuals that move further will leave signs in more sites. Beyond this, the estimates of abundance show no relationship with the parameters that determine the number of signs an individual deposits for each presence-absence survey, δ. If the presence-absence data were contributing to the estimate, then we would expect the distribution of the estimator to change, and hopefully improve, as δ increases and more presence signs are found. Since this does not occur, we conclude that the relationship between the number of presence-absence detections and the estimate of abundance is being driven purely by the relationship of both to the number of individuals captured, and that the presence-absence data does not contribute at all.

Furthermore, varying the number of sites, number of capture-recapture sampling occasions, number of presence-absence sampling occasions, and the number of camera traps as in case study A and case study B did not affect the outcome of the model. The population size estimates were always just above the number of animals captured.

Why does the model produce these results? As mentioned in Section 2.1, the problem is that two different scales are being equated in Equation 6. The mean abundance for the population over the entire study area, λ, is related with the probability that a single site is occupied, ψ. The equation does not take into account animal movements across sites or the size of the sites. Blanc et al. (2014)'s model does account for the fact that sampling of the presence-absence data happens over a period of time, not instantaneously, so that individuals may leave signs in multiple sites during a single period and sites will become occupied by every individual given enough time (Efford and Dawson, 2012). To understand why the estimator of N essentially equals the number of captured animals consider the following. Since there is only one realization of N, the overall abundance, we would expect $\lambda \approx 10$. If, for example, $N = 10$, a rather small population size, then we would expect that $\lambda \approx 10$. Equation (6) would then imply that $\psi_i = 1 - e^{-10} = 1 - .00005$. This is so close to 1 that we would expect every site to be occupied. Since ψ_i is a monotone increasing function of λ, the situation only becomes worse as N increases. Failing to observe signs in any site is so unlikely in this case that even one site with no presence-absence signs found in one period forces ψ_i and λ to be as small as possible. Since λ is bounded below by the number of unique individuals captured, we essentially end up with $\lambda = n$ unless signs are detected at every site on every occasion.

Although Blanc et al. (2014) do discuss the use of SECR models, it is in reference to a violation of the assumption about the total abundance, N, being a Poisson random variable. In particular, they suggest that their model may underestimate the variance of N if abundance does not follow a "homogeneous Poisson random variable" (Blanc et al., 2014, pg. 1738). This statement seems to confuse two concepts: that of a Poisson random variable and that of a Poisson process. Homogeneity refers to whether the intensity of a Poisson *process* is constant over space or time. In fact, the overall abundance generated by a non-homogeneous Poisson process is still a Poisson random variable, so modelling the distribution of individuals across space as suggested would not affect the distribution of N. This means that incorporating SECR to model the distribution of animals will not solve any problems relating to the Poisson assumption. Moreover, as our simulation shows, the approach does not work even when the distribution of animals is homogeneous. Unmentioned by Blanc et al. (2014) is the issue of spacial scales, with the occupancy likelihood defined over site yet the capture-recapture model collapsed over all sites. It seems they did not recognize the spatial-scales issue even though they report an estimated abundance of 9.96 for their combined model and 14.46 [95% CI: 9·0; 35·0] for the CR model alone. We note that the lower bound of the interval estimate is 9.0, the number of observed individuals. Thus clearly Blanc et al. (2014) were encountering the issue of their integrated model returning the number of observed individuals, but they did not explore the underlying cause as we have done here.

Blanc et al. (2014) are correct that the problem of combining a SECR model likelihood with an occupancy model likelihood is still open (Chandler and Clark, 2014). Our simulation study demonstrates that parameter estimates can be erroneous if proper care is not taken to ensure that spatial scales are handled appropriately. In cases such as the Eurasian lynx, spatial information should be incorporated and a SECR-type model should be employed. More care is needed when implementing new models; we need to ask whether it makes sense to combine data types especially when they concern different spatial scales. If models such as Blanc et al. (2014)'s are used without such care, there could be repercussions on conservation decisions or future study designs.

Our results show that Blanc et al. (2014)'s approach systematically underestimates population sizes which could misdirect resources and neglect areas that need support, potentially jeopardizing the survival of endangered species. Additionally, our simulation methodology can serve as a tool for researchers developing and evaluating novel population estimation models that integrate spatial capture-recapture or presence-absence data. Researchers must carefully evaluate any novel population estimation model that integrates capture-recapture or presence-absence data before their application in actual conservation scenarios. Integrating multiple data sources to improve population estimates is a promising avenue for future research. However, a cautious approach that prioritizes methodological rigor and considers both model assumptions and the potential pitfalls of combining data sources, is paramount for effective conservation efforts.

Author Contributions

Laura Cowen and Simon Bonner conceived the ideas and edited the manuscript. Jack Thomas designed the methodology. Jack Thomas wrote the code and performed the simulations for this study. Jack Thomas, Laura Cowen, and Simon Bonner analyzed and interpreted the results. Jack Thomas led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Conflict of Interest Statement

We, the authors, declare no conflict of interest.
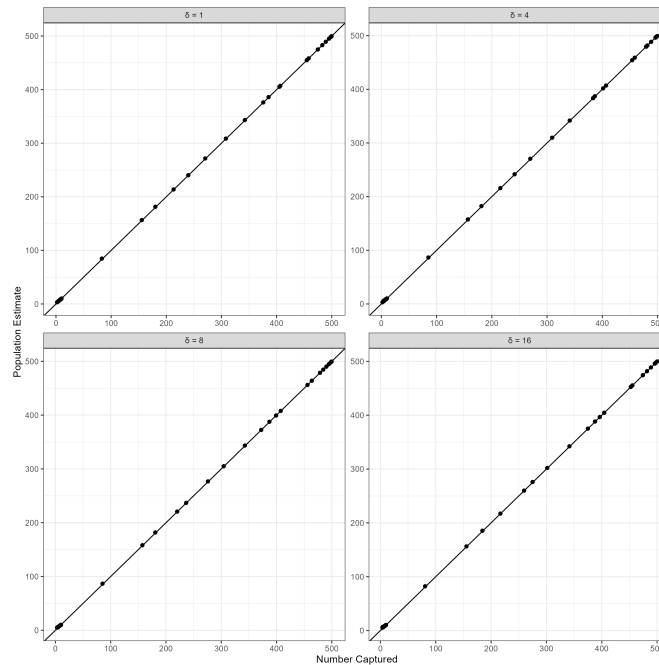
Acknowledgements

Figure 3: Population estimate versus number of observed individuals (Number Captured). Parameters were selected from case study B and varied by the mean number of signs deposited per survey period (δ). A slope 1 line is plotted.
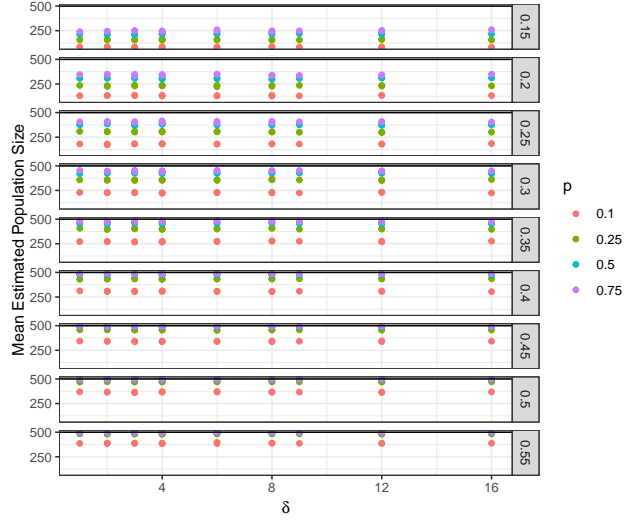
Figure 4: Mean population estimate versus the average number of signs deposited per survey period (δ). The true population size, $N = 500$, is indicated by the solid horizontal line in each panel. Panels are separated by the values of τ and the colours indicate the value of p. The number of camera traps and the number of sites were fixed at $m_1 = 5$ and $m_2 = 64$.
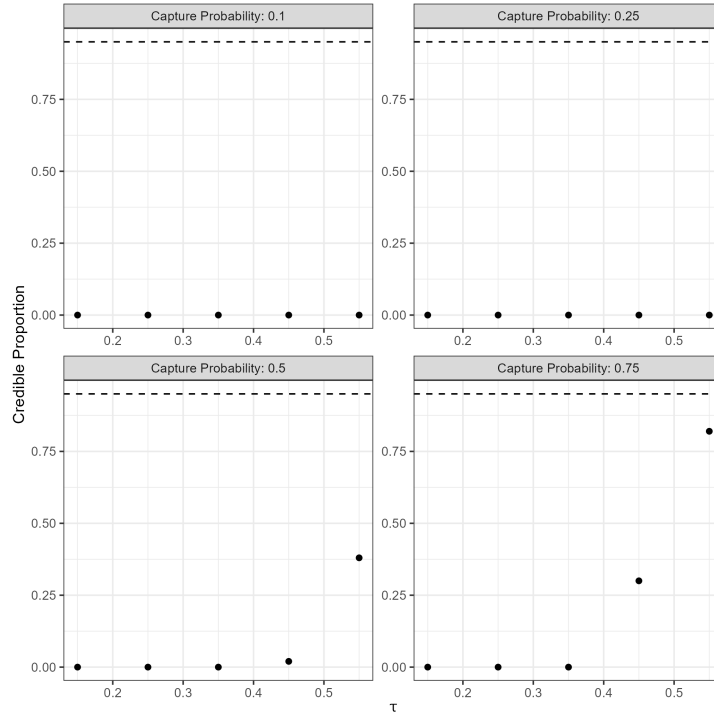

Figure 5: Credible proportion versus movement parameter (τ) varied by capture probability (0.1, 0.25, 0.5, 0.75). True population $N = 500$, number of camera traps $m_1 = 5$, number of sites $m_2 = 64$, and mean number of signs deposited per survey period $\delta = 1$. Credible proportion is the percentage of times that the 95% credible interval contained the true population value (N=500). The 95% threshold is displayed as a dotted line.

References

Bailey, L. L., MacKenzie, D. I., and Nichols, J. D. (2014). Advances and applications of occupancy models. *Methods in Ecology and Evolution*, 5(12):1269–1279.

Besbeas, P., Freeman, S. N., and Morgan, B. J. T. (2005). The potential of integrated population modelling. *Australian & New Zealand Journal of Statistics*, 47(1):35–48.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58(3):540–547.

Blanc, L., Marboutin, E., Gatti, S., and Gimenez, O. (2013). Abundance of rare and elusive species: empirical investigation of closed versus spatially explicit capture–recapture models with lynx as a case study. *The Journal of Wildlife Management*, 77(2):372–378.

Blanc, L., Marboutin, E., Gatti, S., Zimmermann, F., and Gimenez, O. (2014). Improving abundance estimation by combining capture-recapture and occupancy data: Example with a large carnivore. *Journal of Applied Ecology*, 51(6):1733–1739.

Chandler, R. B. and Clark, J. D. (2014). Spatially explicit integrated population models. *Methods in Ecology and Evolution*, 5:1351–1360.

Conroy, M. J., Runge, J. P., Barker, R. J., Schofield, M. R., and Fonnesbeck, C. J. (2008). Efficient estimation of abundance for patchily distributed populations via two-phase, adaptive sampling. *Ecology*, 89(12):3362–3370.

Denwood, M. J. (2016). runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, 71(9):1–25.

Efford, M. (2004). Density estimation in live-trapping studies. *Oikos*, 106(3):598–610.

Efford, M., Warburton, B., and Spencer, N. (2000). Home-range changes by brushtail possums in response to control. *Wildlife Research*, 27:117–127.

Efford, M. G. (2014). Bias from heterogeneous usage of space in spatially explicit capture–recapture analyses. *Methods in Ecology and Evolution*, 5(7):599–602.

Efford, M. G. and Dawson, D. K. (2012). Occupancy in continuous habitat. *Ecosphere*, 3(4):1–15.

Fraysse, R., Choquet, R., Costantini, C., and Pradel, R. (2023). Population size estimation with capture-recapture in presence of individual misidentification and low recapture. *arXiv preprint arXiv:2304.00885*.

Girondot, M. and Rizzo, A. (2015). Bayesian framework to integrate traditional ecological knowledge into ecological modeling: a case study. *Journal of Ethnobiology*, 35(2):337–353.

Jahid, M., Steeves, H. N., Fisher, J. T., Bonner, S. J., Muthukumarana, S., and Cowen, L. L. E. (2022). Shooting for abundance: Comparing integrated multi-sampling models for camera trap and hair trap data. *Environmetrics*, e2761:1–15.

Jiménez, J., Díaz-Ruiz, F., Monterroso, P., Tobajas, J., and Ferreras, P. (2022). Occupancy data improves parameter precision in spatial capture–recapture models. *Ecology and Evolution*, 12(8):e9250.

Lieury, N., Devillard, S., Besnard, A., Gimenez, O., Hameau, O., Ponchon, C., and Millon, A. (2017). Designing cost-effective capture-recapture surveys for improving the monitoring of survival in bird populations. *Biological Conservation*, 214:233–241.

Litt, A. R. and Steidl, R. J. (2010). Improving estimates of abundance by aggregating sparse capture-recapture data. *Journal of agricultural, biological, and environmental statistics*, 15:228–247.

MacKenzie, D. I. and Nichols, J. D. (2004). Occupancy as a surrogate for abundance estimation. *Animal biodiversity and conservation*, 27(1):461–467.

Morera-Pujol, V., Mostert, P. S., Murphy, K. J., Burkitt, T., Coad, B., McMahon, B. J., Nieuwenhuis, M., Morelle, K., Ward, A. I., and Ciuti, S. (2023). Bayesian species distribution models integrate presence-only and presence–absence data to predict deer distribution and relative abundance. *Ecography*, 2023(2):e06451.

Murphy, S. M., Augustine, B. C., Adams, J. R., Waits, L. P., and Cox, J. J. (2018). Integrating multiple genetic detection methods to estimate population density of social and territorial carnivores. *Ecosphere*, 9(10):e02479.

Oldemeyer, B. N., Copeland, T., and Kennedy, B. P. (2018). A multiyear hierarchical bayesian mark-recapture model incorporating data on recurring salmonid behavior to account for sparse or missing data. *Fishery Bulletin*, 116(3-4):254–281.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, pages 3–135.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Renner, I. W., Louvrier, J., and Gimenez, O. (2019). Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization. *Methods in Ecology and Evolution*, 10(12):2118–2128.

369 Strebel, N., Kéry, M., Guélat, J., and Sattler, T. (2022). Spatiotemporal modelling of abundance from multiple data
370   sources in an integrated spatial distribution model. *Journal of Biogeography*, 49(3):563–575.

371 Tourani, M., Dupont, P., Nawaz, M. A., and Bischof, R. (2020). Multiple observation processes in spatial cap-
372   ture–recapture models: How much do we gain? *Ecology*, 101(7):e03030.

373 Wiedenfeld, D. A., Alberts, A. C., Angulo, A., Bennett, E. L., Byers, O., Contreras-MacBeath, T., Drummond, G.,
374   da Fonseca, G. A., Gascon, C., Harrison, I., et al. (2021). Conservation resource allocation, small population
375   resiliency, and the fallacy of conservation triage. *Conservation Biology*, 35(5):1388–1395.

376 Zipkin, E. F. and Saunders, S. P. (2018). Synthesizing multiple data types for biological conservation using integrated
377   population models. *Biological Conservation*, 217:240–250.