

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

Jack H. W. Thomas<sup>1</sup>

<sup>1</sup>Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 2Y2

**SUMMARY.** 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. 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. Finally, we give an outline of our hypothesis on why the model underestimates population sizes.

**Keywords:** capture-recapture, presence-absence, ecosystem, wildlife, abundance, data integration, integrated population models

## 1 Introduction

Accurately estimating the size of a population from ecological data is crucial for wildlife management. To help endangered species recover, we must be able to allocate resources to areas that need it the most. The most common methods for monitoring species are presence-absence and capture-recapture-like studies. Presence-absence studies attempt to find signs of presence in the field (hair, tracks, scat, etc.). These studies are inexpensive and can generate large datasets. Presence-absence data tell us whether at least one individual has been within a site during a certain time period. In capture-recapture studies animals are assumed to be identifiable either from unique tags or natural marks. Then the marked animals can be tracked over time. An approach gaining in popularity is to combine data from different sources to obtain more accurate estimates for population parameters. We emphasize that great care must be taken when implementing these new models.

Blanc et al. (2014) proposed a statistical model that combined capture-recapture and presence-absence data to estimate population sizes. The researchers estimated the population size of Eurasian lynx in the Jura Mountains, eastern France. Capture-recapture data of lynx were obtained using camera traps. Lynx were identifiable by natural marks. Observers collected presence-absence data by an extensive sign survey for evidence of presence (markers) in the field. Blanc et al. (2014) defined marker discovery by hair, tracks, scat, and/or livestock and wildlife killed.

We believe that Blanc et al. (2014)'s method underestimates population sizes. Jahid et al. (2022) discussed and applied two integrated population models that combine presence-absence data from camera traps and capture-recapture data from hair traps to estimate the population size of grizzly bears in the central Rocky Mountains of Alberta, Canada. Blanc et al. (2014)'s model was one of the models used to estimate the grizzly bear population. The approach of Blanc et al. (2014) produced a point estimate of forty bears, despite other methods estimating four hundred bears or more. The main problem with the method is that it fails to account for animal movements when using spatial capture-recapture data. Blanc et al. (2014) did not assess the performance of the method either analytically or via a simulation study. The authors analyzed a dataset but provided no ground truth. We have conducted a simulation study to explore the model and assess the performance of the resultant population estimates. We hypothesized that the model would produce erroneous estimates in all but two extreme cases: first, when animals do not move between sites during the study, and second when animals move throughout all the sites equally. This work is important because researchers may use the model of Blanc et al. (2014) and plan management strategies without realizing that the population size is underestimated. This could have grave wildlife management consequences.

Several different methods have been proposed to combine presence-only, absence-only, count, and full presence-absence data with capture-recapture data. Integrated population models (IPM) attempt to combine individual and population-level data into a single statistical model. Besbeas et al. (2002) demonstrated a state-space model that combined census data with ring-recovery data to obtain maximum likelihood estimates for population parameters of the northern lapwing and the grey heron. Census data were defined to be territory counts of animals. The authors used a Kalman filter to construct the likelihood of the census data. Ring-recovery data were obtained by ringing chicks at birth each year and recovering the rings from the animals at death. The likelihood for the ring-recovery data were constructed by a product of multinomial distributions as described by Catchpole et al. (1999). The authors stated that the model can be extended to integrate capture-recapture data. Besbeas et al. (2002) found that a decline in population for the northern lapwing could be due to a decrease in productivity. For the grey heron, the authors found that there was no decline in productivity and that the integrated population model improved the strength of logistic regressions of survival on winter severity. Besbeas et al. (2005) extended the model by using multivariate normal approximations to combine separate likelihoods.

More recently, Zipkin and Saunders (2018) gave an overview of integrated population models and how they can address incomplete data. Local-scale studies are often not broad enough to obtain population-level data to make

accurate estimates of population sizes (Zipkin and Saunders (2018)). The authors describe three components of IPMs: a population model linking abundance and demographic rates, individual likelihoods, and a joint likelihood (Zipkin and Saunders (2018)). The authors 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.

As an extension of the combined likelihood approach of IPMs, Renner et al. (2019) combined multiple presence-only data into a single closed-population likelihood to estimate the abundance of Eurasian lynx in the Jura Mountains, eastern France. The three sources of data consisted of sightings, 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 assumes that sites are always occupied or always unoccupied. Altitude, forest cover, and human population density were treated as environmental variables. Previous work by Renner et al. (2019) showed that there is more domestic interference in areas with higher human population densities. The usual assumption that multiple point locations are independently distributed is violated in real-life data. The model of Renner et al. (2019) accounts for this by using spatial dependence. Lasso-type (least absolute shrinkage and selection operator) penalties are also applied to the model to prevent potential over-fitting. A simulation study using generated data confirmed that the model performance was improved when spatial dependence was present in the data. The method takes advantage of prior knowledge in selecting presence-only datasets. Further extensions to the model include adding different types of spatial dependence to the model.

In another example Jiménez et al. (2022) combined presence-absence data with capture-recapture data to estimate the abundance of stone marten. Camera traps were used to identify some individuals, which constitutes the capture-recapture data. However, many individuals were unidentifiable, leading to the creation of presence-absence data using the camera traps as well. In essence, the camera traps generated both capture-recapture and presence-absence data. The paper found that the integrated model had more accurate and precise estimates by adding presence-absence data. One factor that limits the model is the sparsity of camera traps. Presence-absence data could only be collected by camera traps in the study. Thus, a full geographic site survey would not be incorporated into the model. To compensate, telemetry data was collected from tagged individuals.

A different approach by Strebel et al. (2022) combined count, detection/non-detection, presence-only, and absence-only data into a binomial N-mixture model to estimate the abundance of bird species in Switzerland. The four types of data were described as follows: Count data were simply observed counts of animals in a certain site in a certain time range. This is different from capture-recapture data because individuals are not identified or recaptured. Next, detection/non-detection data are described as detecting certain species using a checklist (presence-absence data). Presence-only data is a set of sites that animals are known to inhabit, such as breeding sites. The data is a list of species that were present during the breeding season at a certain site (Strebel et al. (2022)). Finally, absence-only data is a set of sites that animals are known to not inhabit. This may include environments that are not suitable for the ani-

mal, such as steep elevation gradients. Combining presence-absence data and presence-only data increased the spatial coverage compared to using presence-absence data alone. Absence-only data added more information to the model, which allowed more accurate parameter estimates. The authors caution against the use of presence-only data as some sites are more likely to be visited than others and this introduces bias (Strebel et al. (2022)). The authors assert that their approach is valuable when a single source of data is sparse and adding more sources may improve parameter estimates.

Finally, Jahid et al. (2022) compared integrated multi-sampling models for camera trap and hair trap data. The hair trap data constituted capture-recapture data, while the camera traps constituted the presence-absence data. The paper compared six different models that estimated the population size of grizzly bears in the Rocky Mountains, Alberta:

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.

Jahid et al. (2022) found that the precision of the population estimates did not improve when presence-absence data was added to the model. One of the assumptions for these integrated models is that the data sources are independent. In the discussion, the paper noted that hair traps and camera traps were not independent. Hair traps and camera traps were often in the same locations. Bears that were captured through hair trapping were likely to be captured by camera traps as well. Jahid et al. (2022) found that the method of Blanc et al. (2014) had an unusually small posterior standard deviation and that the abundance estimates were almost equal to the number of animals captured. The number of stations with bear detection was 38, which is close to the estimated population of 40 by the Blanc et al. (2014) model. The capture-recapture-only model estimated 73.36 bears, and the SECR model estimated 380.28 bears. The authors suggest a simulation study be performed to test the model of Blanc et al. (2014).

The objective of our study is to test Blanc et al. (2014)'s model through a simulation study to determine the bias of the estimates under various scenarios and to study the behaviour of the model.

## 2 Methods

The model 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  $\lambda$ . Then the probability that a new population would have size  $n$  is calculated using the probability mass function of the Poisson random variable:

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

Next, Blanc et al. (2014) model presence-absence data 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. Then, the probability that a site  $i$  is occupied is:

$$\psi_i = P(Z_i = 1). \quad (2)$$

To combine the two data types, Blanc et al. (2014) equate the probability that a site is occupied with the probability that there is at least one animal in the overall population:

$$P(Z_i = 1) = P(N > 0). \quad (3)$$

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

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

Using (2), they define:

$$\psi_i = 1 - e^{-\lambda}. \quad (5)$$

Solving for  $\lambda$ , they obtain:

$$\lambda = -\log(1 - \psi_i). \quad (6)$$

This is the key equation of Blanc et al. (2014), which links capture-recapture and presence-absence together. In principle, the likelihood functions for both variables inform abundance. The problem with the model is that equation (6) links site specific presence-absence with the overall population size. There are two different scales being merged into one equation. The model of Blanc et al. (2014) does not sample animals instantaneously, it tracks animals over sampling occasions. Site population estimates do not have a logical interpretation because summing up the population estimates for all the sites would not give the overall population size since there would be overlap between the sites. The relationship between overall population size and site specific occupancy is defined by relating occupancy and

movement with time. Every site will become occupied by a single individual given enough time.

We assess the performance of the Blanc et al. (2014) model through a simulation study. Two sets of data were created for each replication: capture-recapture data and presence-absence data. We simulated a set of individuals and derived the capture-recapture and presence-absence data from these individuals. This meant that there could be different numbers of trapping occasions and detection occasions. The data sets are independent conditional on the population, but are not marginally independent.

## 2.1 Simulating the Population

The first step to generating spatial capture-recapture and presence-absence data is defining home ranges. Animals in a population tend to stay within a certain range, denoted as their home range (Efford et al., 2000). We assumed the study area to be a unit square for simplicity. We used a Poisson process to generate  $n$  home range centres  $(x_i, y_i)$ , where  $X_i, Y_i \stackrel{iid}{\sim} \text{Uniform}(0, 1)$  for individuals  $i = 1, 2, \dots, n$ . This is our simulated population of size  $n$ .

## 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. We placed the  $m$  camera traps  $(x_j, y_j)$  into the study area using a Poisson process with  $X_j, Y_j \stackrel{iid}{\sim} \text{Uniform}(0, 1)$  for camera traps  $j = 1, 2, \dots, m$ . We assumed that capture probability is related to the distance between a home range centre and a camera location as well as the amount of movement the species exhibits. A half-normal detection function modeled the capture probability of individual  $i$  by trap  $j$ :

$$p_{ij} = pe^{(-d_{ij}^2/\tau^2)} \quad (7)$$

where  $p$  is the capture probability for a camera trap at distance 0 from the home range centre,  $\tau$  is the movement parameter for the species and  $d_{ij}$  is the distance between the home-range centre  $i$  and the camera location  $j$ . The value  $\tau$  is defined to be any positive real number. When  $\tau$  is large, animals move throughout the entire study area. On the other hand, when  $\tau$  is close to zero, animals do not move significantly from their home-range centres. As  $\tau$  increases the capture probability also increases with a limit of  $p$  when  $\tau = \infty$ . As  $d_{ij}$  increases, the camera trap capture probability  $p_{ij}$  decreases. Once all the camera traps are placed, we simulate the interactions of individuals with the camera traps. We define the number of sampling occasions to be the number of times we check the camera trap data to record the number of captures. Let the number of capture-recapture sampling occasions be defined as  $k = 1, 2, \dots, K$ . Then we obtain a three-dimensional array in which cell  $i, j, k$  determines whether individual  $i$  was captured by trap  $j$  at time  $k$ . This three-dimensional array contains the capture events  $C_{ijk} \sim \text{Binomial}(p_{ij})$ .

## 2.3 Generating Presence-Absence Data

Presence-absence data generation starts with dividing the study area into regions. We define a grid such that each grid cell is equal in size. Following the terminology of Blanc et al. (2014), each grid cell is henceforth referred to as a site. We construct a definition of occupancy (presence-absence) probability. Each individual has a home range that expands or shrinks based on the movement parameter  $\tau$ . Let the proportion of time spent by individual  $i$  in each grid cell  $j$  equal the probability of occupancy  $\psi_{ij}$ . A site is occupied if there is at least one individual within. To calculate  $\psi_{ij}$ , we perform a double integration of the bivariate normal distribution function with the variance of  $X$  and  $Y$  equal to  $\tau^2$  and covariance between  $X$  and  $Y$  equal to 0. We have:

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

where  $f(x, y | x_i, y_i, \tau)$  is the bivariate normal distribution function given home-range  $x_i, y_i$ , and covariance  $\tau$ . The bounds of the grid cell  $j$  are  $x_{0j}, y_{0j}, x_{1j}, y_{1j}$ . Next, we define detection probability. This is the probability that we actually detect an animal in a site through marker discovery. We define the marker deposition rate (e.g. tracks, hair) for a species as the number of markings each individual leaves behind in each sampling period. A higher marker deposition rate results in a higher probability that we will detect an individual in an area. Let  $\sigma$  denote the marker deposition rate for the species. Now, we need to take into account the amount of time between sampling occasions. If there is more time between sampling occasions, there are more markers in each site. Let  $\delta$  denote the time between sampling occasions. We assume that the number of signs animal  $i$  leaves in cell  $j$  is a Poisson random variable  $Z_{ij} \sim \text{Poisson}(\sigma\delta\psi_{ij})$ . Then the detection probability is defined to be:

$$\mathbb{P}(\text{presence detection}) = \mathbb{P}(Z_{ij} > 0) = 1 - e^{-\sigma\delta\psi_{ij}} \quad (9)$$

Each individual  $i$  has a two-dimensional matrix representing the study area grid where each entry contains the detection probability of that site. This probability matrix yields our presence-absence matrix which indicates whether at least one animal was detected at each site for each period. The two-dimensional presence-absence array is the detection event  $D_{ij} \sim \text{Binomial}(\mathbb{P}(Z_{ij} > 0))$ .

## 2.4 Testing the Model

Supplementary code is provided by Blanc et al. (2014) that implements 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

Table 1: Simulation case study *A*. We varied population size, movement parameter, the number of sites, and the number of camera traps. The aim of case study *A* was to see if the model produced erroneous population estimates across all  $N$ .

Population Size	$N$	10, 25, 50, 75, 100, 125, 150, 175, 200, 500
Movement Parameter	$\tau$	0.15, 0.20, 0.25, 0.30, 0.35, 0.4, 0.45, 0.5, 0.55
Capture Probability	$p$	0.75
Marker Deposition Rate	$\sigma$	2
Presence-Absence Sampling Interval Time	$\delta$	2
Number of Sites	$n_{sites}$	64, 81, 100
Number of Capture-Recapture Sampling Occasions	$n_{CR}$	7
Number of Presence-Absence Sampling Occasions	$n_{PA}$	5
Number of Camera Traps	$n_{traps}$	3, 4, 5, 6

errors, and the coverage probability and width of the 95% credible intervals. We defined standard deviation bias to be the difference between the model output posterior standard deviations for the population size and our own calculated standard deviations of the population estimate. We define the coverage probability to be the percentage of cases where the actual 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). The R package “rjags” generated samples from the JAGS model. Parameters for each simulation were the actual population size, movement parameter, number of sites in the study area, number of capture-recapture sampling occasions, number of presence-absence sampling occasions, and the number of camera traps in the study. Each simulation was repeated 50 times and was defined by a set of parameters. We studied all possible combinations of the parameter values in Table 1 and Table 2.

We present two case studies, *A* and *B*. In case study *A* we varied  $N$ ,  $\tau$ ,  $n_{sites}$ , and  $n_{traps}$ , while holding  $p$ ,  $\sigma$ ,  $\delta$ ,  $n_{CR}$ , and  $n_{PA}$  as constants. 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,  $N$ . Case study *B* only had two different values of  $N$ , while varying  $\tau$ ,  $p$ ,  $\sigma$ , and  $\delta$ , and holding  $n_{CR}$ ,  $n_{sites}$ ,  $n_{traps}$  and  $n_{PA}$  as constants. The aim of case study *B* was to see if the behavior of the model changed from case study *A* by varying  $p$ ,  $\sigma$ , and  $\delta$ .



Table 2: Simulation case study *B*. Varying movement parameter, capture probability, marker deposition rate, presence-absence sampling interval time, the number of sites, and the number of camera traps. The aim of case study *B* was to see if varying  $p$ ,  $\sigma$ , and  $\delta$  would affect the behavior from case study *A*; erroneous population estimates.

Population Size	$N$	10, 500
Movement Parameter	$\tau$	0.15, 0.20, 0.25, 0.30, 0.35, 0.4, 0.45, 0.5, 0.55
Capture Probability	$p$	0.1, 0.25, 0.5, 0.75
Marker Deposition Rate	$\sigma$	1, 2, 3, 4
Presence-Absence Sampling Interval Time	$\delta$	1, 2, 3, 4
Number of Sites	$n_{sites}$	64
Number of Capture-Recapture Sampling Occasions	$n_{CR}$	7
Number of Presence-Absence Sampling Occasions	$n_{PA}$	5
Number of Camera Traps	$n_{traps}$	5

### 3 Results

All model population estimates were almost one-to-one with the number of animals captured in the simulation (Figure 1) across all capture probabilities  $p$ . There was an almost perfect linear correlation ( $\rho = 0.999$ ) between the population estimate and the number of animals captured, even when the true population size remained fixed. Varying the species capture probability  $p$  did not affect the correlation value. The model estimates the population to be almost exactly the number of animals captured. As  $p$  increased to 0.75, there was more clustering of the number of animals captured towards the actual population size of 500, but the result was the same. Figure 2 shows that as  $\tau$  increased from 0.15 to 0.55, more animals were captured in the simulation until eventually all animals in the population were captured. Varying  $\sigma$  and  $\delta$  did not change the near one-to-one correspondence of the number of captured individuals and the population estimate from the model (Figure 3). Except for extreme situations, these represent gross underestimates of the actual population size. As the number of animals captured converged to the actual population, the number of presence-absence detections decreased (Figure 4). As  $\tau$  decreases, fewer animals are captured by camera traps but animals spend more of their time in fewer sites. When animals stay near a small number of sites, we consistently get more presence-absence detections. As  $\tau$  increased from 0.15 to 0.55, the site occupancy probability decreased from 0.9916 to 0.9892 (Figure 5). In other words, as animals moved more they spent less time in the overall study area and there was a small decrease in site presence-absence probability. There was convergence of the mean  $\hat{\psi}$  as  $\tau$  increased (Figure 5). The mean  $\hat{\psi}_0$  parameter was close to one in every simulation. Figure 6 shows how the coverage of the 95% credible intervals changed with  $\tau$  and  $p$ . When  $p$  was high enough to capture enough animals to be close to the true population ( $p = 0.5$ ), the coverage probability increased systematically from 0 when  $\tau$  was small (.10 or .20)

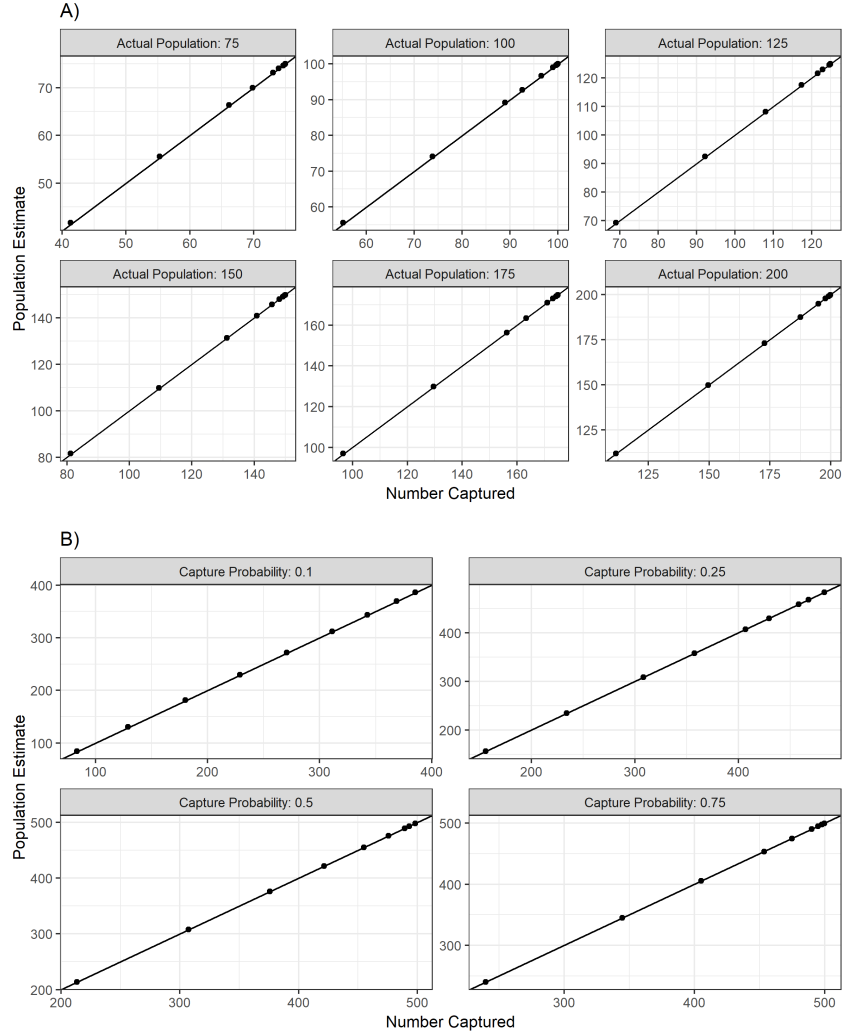


Figure 1: Correlation of the estimated population size versus number of observed individuals (Number Captured) during the study period varied by A) true population size (75, 100, 125, 150, 175, 200) with capture probability set to  $p = 0.75$ , 100 sites, and 6 camera traps and B) capture probability (0.1, 0.25, 0.5, 0.75) with true population size  $N = 500$ , 64 sites, and 5 camera traps. Both A) and B) had: movement parameter  $\tau$  (0.15, 0.20, 0.25, 0.30, 0.35, 0.4, 0.45, 0.5, 0.55), number of CR sampling occasions (7), and number of PA sampling occasions (5). Each point on A) and B) represent a value of  $\tau$ .

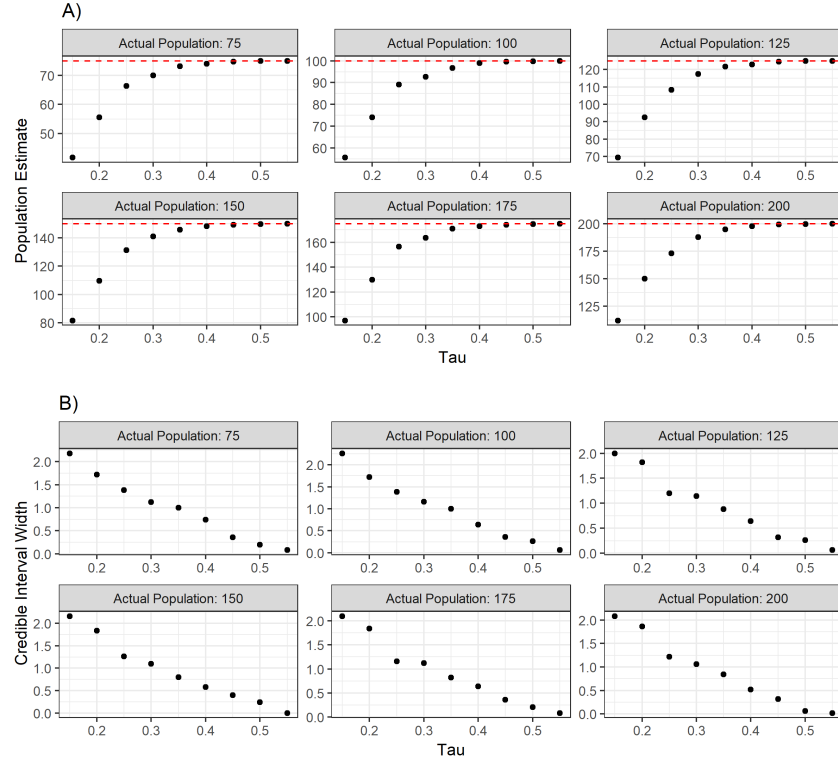


Figure 2: Population estimate versus tau A) and credible interval width versus tau B). True population size was varied in ( 75, 100, 125, 150, 175, 200) with capture probability set to  $p = 0.75$ , number of sites  $n_{sites} = 100$ , and number of camera traps  $n_{traps} = 6$ . The dotted red line in A) represents the actual population.

to almost 1.0 when  $\tau$  was large (0.55). Figure S1 shows that as  $\tau$  increases, the standard deviation bias goes to zero across all population sizes because the population is being censused. Figure S1 also shows the effect of  $p$  and  $\tau$  on the standard deviation bias behaviour. When  $p$  is small and we capture few animals, the standard deviation bias magnitude increases to over -0.95 as  $\tau$  increases to 0.55. When  $p$  is large ( $p = 0.75$ ) and we capture many animals then as  $\tau$  increases to 0.55 we see a small decrease in magnitude of the standard deviation bias to -0.87. The number of animals captured directly affects the standard deviation bias, which is always negatively biased. Increasing the proportion of animals captured decreases the magnitude of the standard deviation bias. When  $p$  is small ( $p = 0.10$ ) then the standard deviation bias increased in magnitude as the population proportion captured increased to 1 (when  $\tau$  increases to 0.55) (Figure S1).

## 4 Discussion

The results of the simulation study in Section 3 show that the population estimates of Blanc et al. (2014)'s model are completely dependent on the number of animals captured. Blanc et al. (2014)'s model estimates the population size to be slightly greater than the number of individuals captured (Figure 1). The positive correlation between population

estimate and the number of animals captured is almost exactly one-to-one ( $\rho = 0.999$ ). Since the population size is fixed for each simulation, the one-to-one correlation between the population estimate and the number of animals captured is a problem because the model is severely underestimating the population size in almost every case (except when we capture all the animals in the population). The parameter that affects the number of individuals captured is  $\tau$ , the movement parameter, and  $p$ , the species capture probability. In Figure 2 we see the convergence of the population size estimate to the actual population value as  $\tau$  increases to 0.55. As  $\tau$  increases, the bias of the standard deviation goes to zero (Figure S2). The magnitude of the bias of the population estimate is always significantly negative unless the probability of capturing an individual is close to one or the animals are moving throughout the entire study area (Figure 2). For the first case, we are censusing the population. For the second case, there is no real spatial effect.

The credible interval values produced by Blanc et al. (2014)'s model are highly suspect. In all cases, the lower 2.5% quantile of the posterior distribution is the number of individuals captured. We would expect that the 0% quantile would be the number of individuals captured and that the 2.5% quantile would be larger. However, this was not the case. Rather, for most values of  $\tau$  and  $p$ , the 95% credible intervals do not capture the true population size. For  $p < 0.5$  and  $\tau \leq 0.55$  the coverage is as low as 0% and as high as 75%, never reaching 95% coverage (Figure 6). As  $\tau$  increases to 0.55 and the species capture probability increases to 0.75, the number of animals captured also increases and we eventually capture the entire population. Once  $\tau$  and  $p$  have increased to sufficient values, we essentially perform a population census and the 95% credible interval contains the true population size, but only because the number of individuals captured is almost equal to the true population size. This behaviour can be seen in Figure 6. As the population proportion captured increases to 1, the 95% credible interval eventually contains the true population value. The amount of animal movement is intrinsically linked to the model's performance. As  $\tau$  increases, the credible interval width decreases, converging to the actual population size (Figure 2).

Another important result of the relationship between  $\tau$  and  $p$  is that the number of presence detections did not prove useful in influencing the population estimates of Blanc et al. (2014)'s model. When we increase  $\tau$  and  $p$  more animals are captured by the camera traps in the simulation. The number of presence detections have a negative correlation with both population size estimates from Blanc et al. (2014)'s model and the number of captured individuals in the simulation. The number of presence detections decrease as the movement parameter  $\tau$  increases because animals do not linger in single sites when  $\tau$  is high. When  $\tau$  decreases, we see a higher number of detections because we have a higher probability of seeing an animal near its home-range. Since the number of observed individuals has a nearly one-to-one correlation with population size estimate, we see that Figure 4 A) and B) look almost identical.

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.

We summarize the results of our simulation study by stating that Blanc et al. (2014)'s model is invalid for all but one

extreme case: when there is effectively only one site because individuals move through the entire study area. Jahid et al. (2022) speculated that the model would be valid if animals ranged over the entire study area (when  $\tau$  is sufficiently high). We have shown this to be true; when the animal movement is high enough to capture all individuals in the population, the model could be considered valid. When there is low movement, the model produces population size estimates that match the number of individuals captured (Figure 1). Why does the model produce these results? The problem is that two different scales are being equated in Equation 6. The mean abundance  $\lambda$  for the entire population is related with the probability that a single site is occupied  $\psi$ . The equation does not take into account animal movements across sites or the size of the sites. Blanc et al. (2014)’s model does not sample animals instantaneously and does not account for the fact that all sites will become occupied by every individual given enough time. We have seen that the study-area sites become saturated in the presence-absence data generation, thus the presence probability tends to one. The probability that a site is never occupied is low because the number of individuals tends to overlap the number of sites. In Blanc et al. (2014),  $\lambda$  is an increasing function of  $\psi$  which makes sense because higher abundance means higher presence-absence probability and vice versa. But the data on presence-absence pushes  $\psi$  down because knowing some sites are unoccupied means that  $\psi < 1$ . However, the data on abundance pushes  $\lambda$  and  $\psi$  up. Thus  $\lambda$  increases just enough to make the observed number of individuals possible but then stops to avoid dragging  $\psi$  up too high. The resulting population estimate is more about the interplay between  $\lambda$  and  $\psi$  being on different spatial scales and essentially ignores the presence-absence data. More care is needed when implementing new models; we need to ask when it makes sense to combine data types especially when they concern different spatial scales.

## ACKNOWLEDGEMENTS

This research was enabled in part by support provided by a Canadian Statistical Sciences Institute Collaborative Research Team grant to SB and LC and the Digital Research Alliance of Canada ([alliance.can.ca](http://alliance.can.ca)).

## Territory Acknowledgement

We acknowledge and respect the łə́kʷəjən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day.

## References

- Besbeas, P., Freeman, S., and Morgan, B. (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., 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.
- Catchpole, E. A., Morgan, B. J. T., Freeman, S. N., and Peach, W. J. (1999). Modelling the survival of british lapwings *vanellus vanellus* using ring-recovery data and weather covariates. *Bird Study*, 46(sup1):S5–S13.
- 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.
- Jahid, M., Steeves, H. N., Fisher, J. T., Bonner, S. J., Muthukumarana, S., and Cowen, L. L. (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.
- 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.
- Strebel, N., Kéry, M., Guélat, J., and Sattler, T. (2022). Spatiotemporal modelling of abundance from multiple data sources in an integrated spatial distribution model. *Journal of Biogeography*, 49(3):563–575.
- Tourani, M., Dupont, P., Nawaz, M. A., and Bischof, R. (2020). Multiple observation processes in spatial capture–recapture models: How much do we gain? *Ecology*, 101(7):e03030.
- Zipkin, E. F. and Saunders, S. P. (2018). Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation*, 217:240–250.

## 5 Appendix

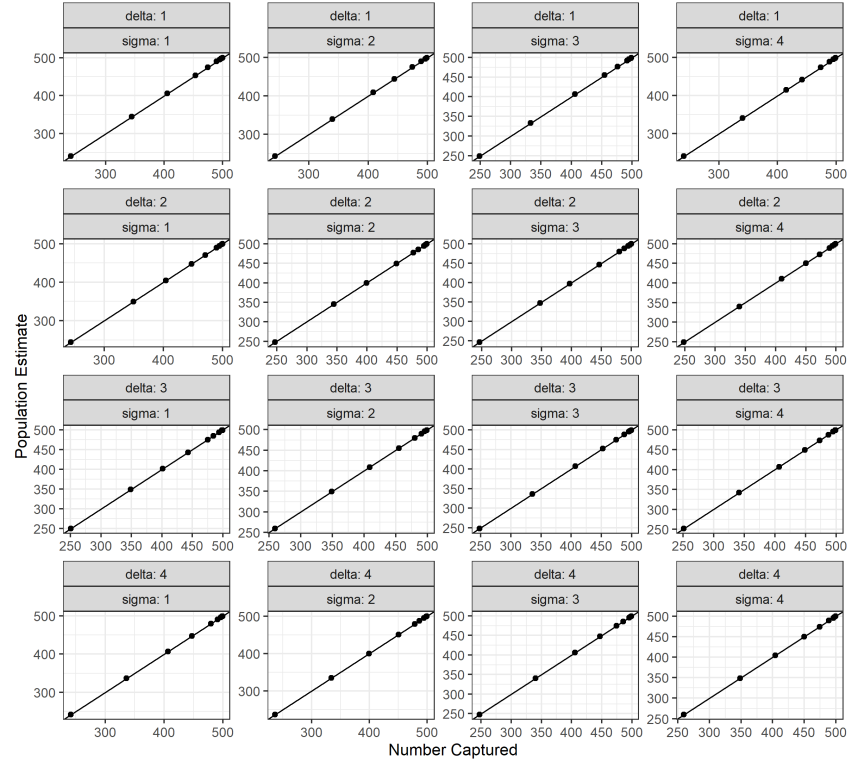


Figure 3: Population estimate versus number of observed individuals (Number Captured) during the study period varied by marker deposition rate ( $\delta$ ) and presence-absence sampling interval time ( $\sigma$ ). Actual population was  $N = 500$ , number of sites  $n_{sites} = 64$ , number of camera traps  $n_{traps} = 5$ , and capture probability  $p = 0.75$ . A slope 1 line is plotted.

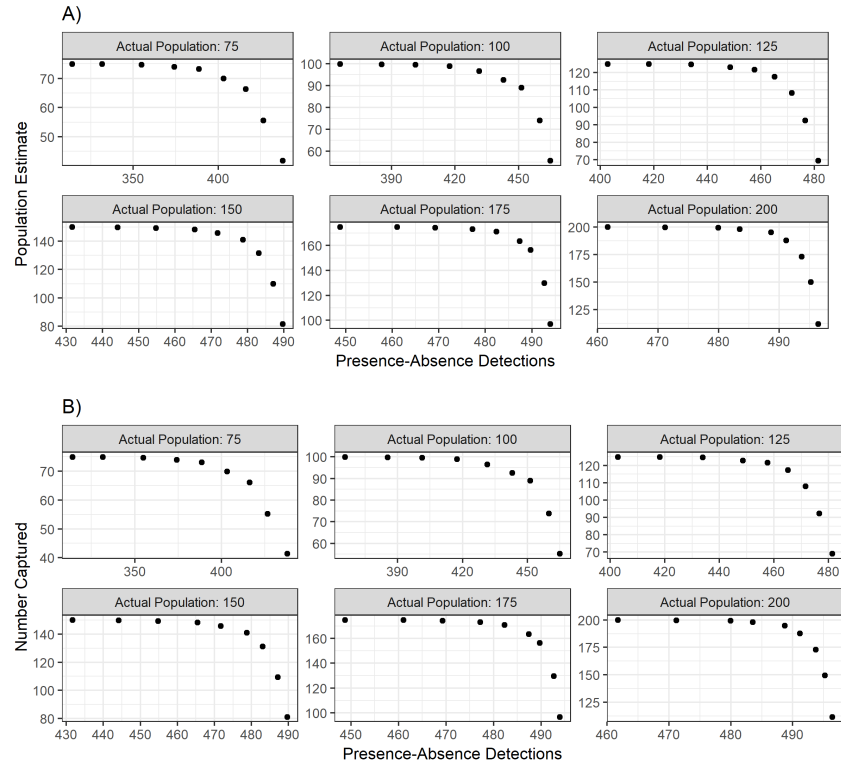


Figure 4: Population estimate versus presence-absence detections A) and number of observed individuals during the study period versus presence-absence detections B) varied by actual population (75, 100, 125, 150, 175, 200). Capture probability set to  $p = 0.75$ , number of sites  $n_{sites} = 100$ , and number of camera traps  $n_{traps} = 6$ .



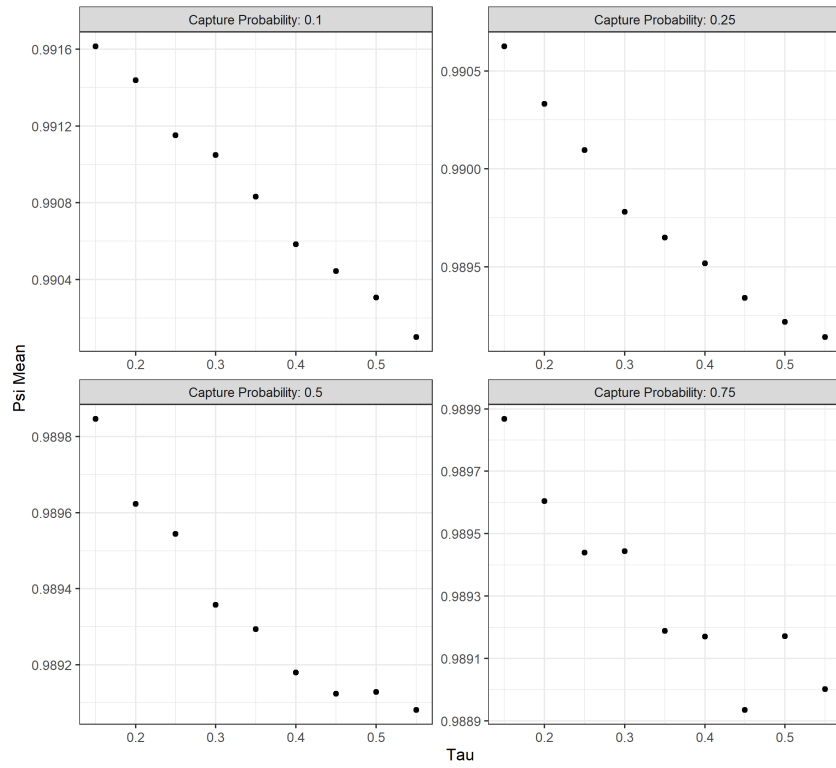


Figure 5: Psi mean versus movement parameter ( $\tau$ ) varied by capture probability (0.10, 0.25, 0.50, 0.75). Actual population  $N = 500$ , number of sites  $n_{sites} = 64$ , number of camera traps  $n_{traps} = 5$ , presence-absence sampling interval time  $\delta = 1$ , and marker deposition rate  $\sigma = 1$ .

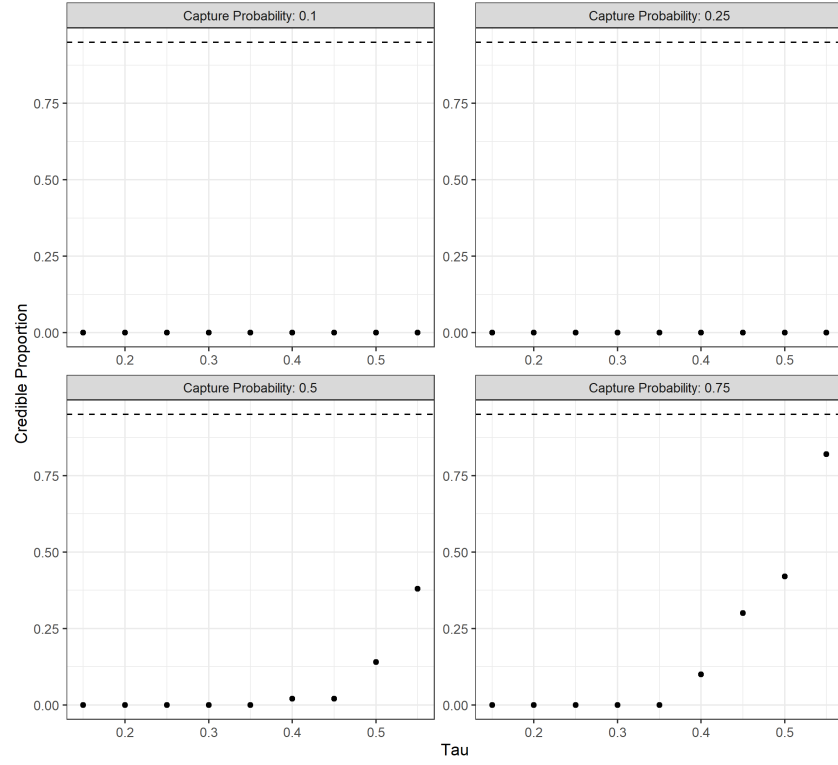


Figure 6: Credible proportion versus movement parameter ( $\tau$ ) varied by capture probability (0.1, 0.25, 0.5, 0.75). Actual population  $N = 500$ , number of sites  $n_{sites} = 64$ , number of camera traps  $n_{traps} = 5$ , presence-absence sampling interval time  $\delta = 1$ , and marker deposition rate  $\sigma = 1$ . Credible proportion is the percentage of times that the 95% credible interval contained the actual population value ( $N=500$ ). The 95% threshold is displayed as a dotted line.

## 6 Supplementary Material

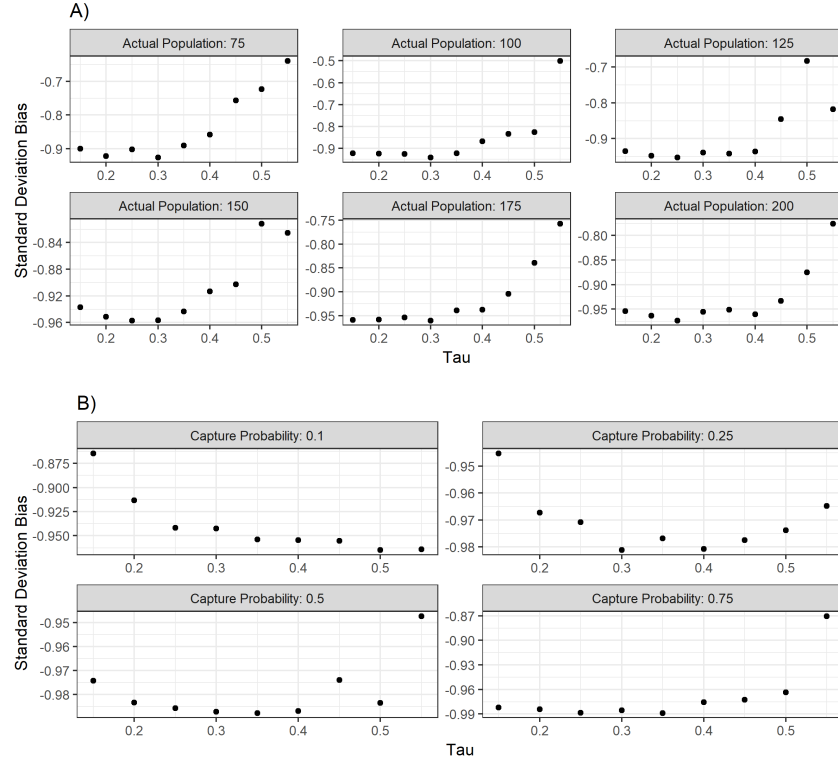


Figure S1: Standard deviation bias versus movement parameter  $\tau$  varied by A) actual population  $N$  (75, 100, 125, 150, 175, 200), number of sites  $n_{sites} = 100$ , number of camera traps  $n_{traps} = 6$  and B) capture probability  $p$  (0.1, 0.25, 0.5, 0.75), actual population  $N = 500$ , number of sites  $n_{sites} = 64$ , number of camera traps  $n_{traps} = 5$ .

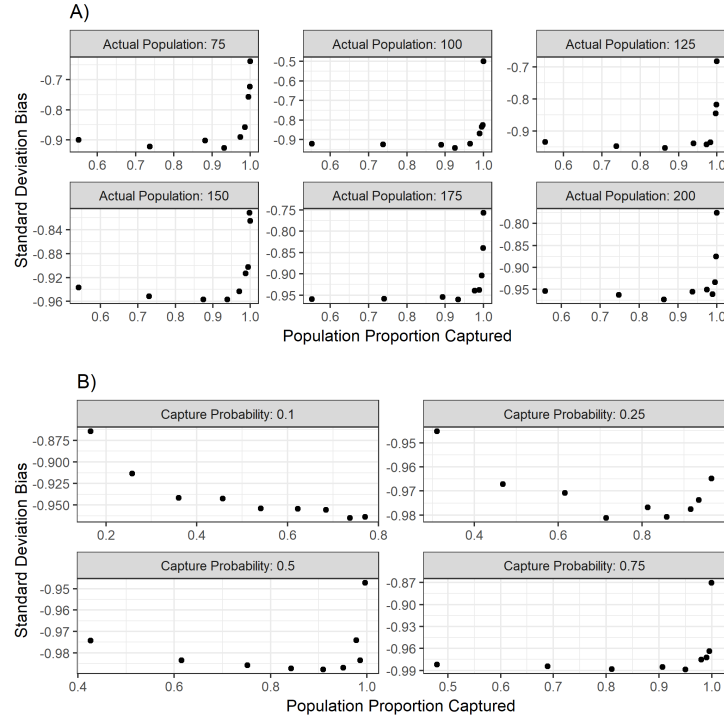


Figure S2: Standard deviation bias versus population proportion captured varied by A) actual population  $N$  (75, 100, 125, 150, 175, 200), number of sites  $n_{sites} = 100$ , number of camera traps  $n_{traps} = 6$  and B) capture probability  $p$  (0.1, 0.25, 0.5, 0.75), actual population  $N = 500$ , number of sites  $n_{sites} = 64$ , number of camera traps  $n_{traps} = 5$ . Standard deviation bias defined as the difference between model output posterior standard deviation and calculated standard deviation of the population estimates.