

## Article

# An Enhanced Spatial Capture Model for Population Analysis Using Unidentified Counts through Camera Encounters

Mohamed Jaber <sup>†</sup>, Farag Hamad <sup>†,‡</sup>, Robert D. Breininger and Nezamoddin N. Kachouie <sup>\*</sup>

Department of Mathematics and Systems Engineering, Florida Institute of Technology, Melbourne, FL 32901, USA

<sup>\*</sup> Correspondence: nezamoddin@fit.edu<sup>†</sup> These authors contributed equally to this work.<sup>‡</sup> Current Affiliation: Department of Mathematics, Benghazi University, Benghazi, Libya.

**Abstract:** Spatial capture models are broadly used for population analysis in ecological statistics. Spatial capture models for unidentified individuals rely on data augmentation to create a zero-inflated population. The unknown true population size can be considered as the number of successes of a binomial distribution with an unknown number of independent trials and an unknown probability of success. Augmented population size is a realization of the unknown number of trials and is recommended to be much larger than the unknown population size. As a result, the probability of success of binomial distribution, i.e., the unknown probability that a hypothetical individual in the augmented population belongs to the true population, can be obtained by dividing the unknown true population size by the augmented population size. This is an inverse problem as neither the true population size nor the probability of success is known, and the accuracy of their estimates strongly relies on the augmented population size. Therefore, the estimated population size in spatial capture models is very sensitive to the size of a zero-inflated population and in turn to the estimated probability of success. This is an important issue in spatial capture models as a typical count model with censored data (unidentified and/or undetected). Hence, in this research, we investigated the sensitivity and accuracy of the spatial capture model to address this problem with the objective of improving the robustness of the model. We demonstrated that the estimated population size using the proposed enhanced capture model was more accurate in comparison with the previous spatial capture model.

**Keywords:** spatial Bayesian models; virtual trap; probability of success; distance sampling; count models; unidentified individuals; data augmentation; Markov chain Monte Carlo

**MSC:** 62P10; 62C10; 62N01; 60J05; 60J22; 60K40



**Citation:** Jaber, M.; Hamad, F.; Breininger, R.D.; Kachouie, N.N. An Enhanced Spatial Capture Model for Population Analysis Using Unidentified Counts through Camera Encounters. *Axioms* **2023**, *12*, 1094. <https://doi.org/10.3390/axioms12121094>

Academic Editors: Hans J. Haubold, Dragana Valjarević, Vladica Stojanović and Aleksandar Dj Valjarević

Received: 8 August 2023

Revised: 18 November 2023

Accepted: 23 November 2023

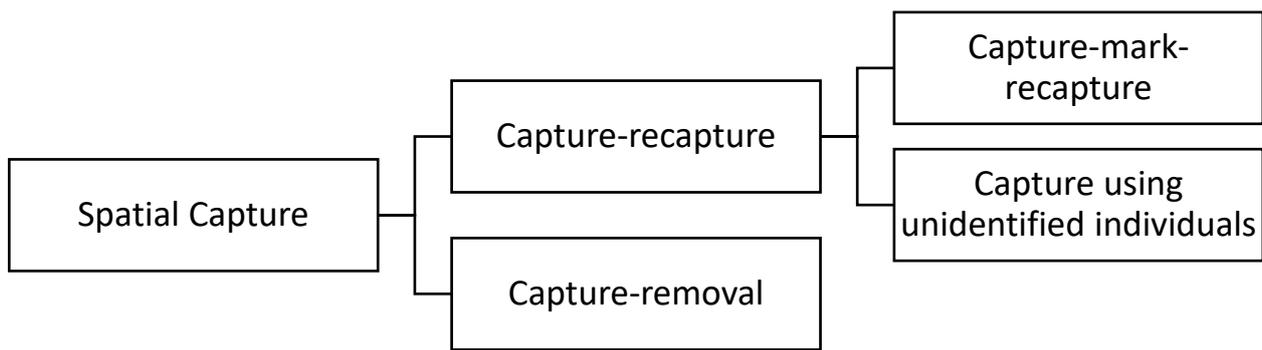
Published: 29 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Population analysis based on spatial sampling is an emerging field of research due to its broad range of applications. Some important applications of spatial population analysis are to preserve endangered species and to control invasive species. One of the first steps required to manage a population is estimating the population size. Several methods have been developed to estimate the abundance of animals [1–3]. A standard approach is counting the members of the population in a random sample to estimate the population size. Population density can be then estimated by dividing the population size by the area of the animals' habitat. The spatial capture methods can be split into two major groups, capture–removal methods and capture–recapture methods, as demonstrated in Figure 1. In capture–removal methods, the individuals that are captured will be counted, recorded, and then removed from the habitat.



**Figure 1.** Lineage of spatial capture methods.

In contrast, in capture–recapture methods, the individuals that are captured will be counted, recorded, and then released back into the habitat. Capture–recapture methods are commonly used in ecological statistics for estimating population size [1,3–7]. Capture–recapture methods can be split into two sub-groups, capture–mark–recapture and capture–unidentified individuals. The distinction between the two sub-groups is whether the captured individuals are marked, identified, and released, or whether individuals are virtually captured without being identified. The intuition behind the capture–mark–recapture method is that the proportion of the individuals that were captured, marked, and released in the first sample and were then recaptured in the second sample can be used to estimate the population size [1]. In the capture–unidentified–individuals method, animals are virtually captured without being identified, obtaining only the number of encountered individuals.

Capture–mark–recapture methods rely on the physical capture of individuals to collect the individual encounter history. Due to technological advances, the ability to capture individuals has been improved through more efficient methods, such as camera traps [8], acoustic recordings, and DNA samples [9,10]. Individuals can be virtually captured using their signs without being identified. Camera traps can be effectively used for the virtual capture of individuals. The inherent characteristics of sampling using camera traps include the following: (1) the same individual can be captured multiple times by the same camera, (2) the same individual can be captured by different cameras, (3) often, only a subset of individuals will be captured, and (4) captured individuals are not identified.

Conventionally, capture methods do not use spatial information about the captured individuals. Advanced spatial capture models have been implemented in the past decade [11–16]. A spatial capture model that has been broadly used was introduced by Chandler and Royle [17]. This model yields promising outcomes using unmarked or partially marked individuals to estimate the population size and density. However, due to the complexity of the spatial capture–unidentified approach, the error in the estimated population size and density could be prohibitively high. An open-ended problem is to address the shortcomings of this method and make the model robust regarding spatial sampling and spatiotemporal population analysis. The objective of this study is to improve the estimated population size by reducing the estimation error to make it robust. To this end, the proposed method performs the following tasks:

- (1) Employs a prior distribution for the essential parameter of the zero-inflated population;
- (2) Regularizes the Markov chain Monte Carlo (MCMC) by controlling the effective sample size;
- (3) Reduces the order of the chain by controlling the correlation of generated samples through Gibbs sampling [18–20].

## 2. Methods

The capture–recapture method is one of the most common methods to estimate the population size. The capture–recapture method has its strengths and limitations based on the species, habitat, and available resources.

### 2.1. Hierarchical Spatial Capture–Recapture Model

Hierarchical spatial capture–recapture (HSCR) is a statistical model widely used in wildlife ecology to estimate population size [11]. It combines information from multiple trapping sites and accounts for spatial dependences among the captured encounters to estimate the population size [21].

In this model, each individual is associated with an unknown spatial activity center and home range radius. Considering population size  $N$ , there are  $N$  unknown activity centers. It is assumed that individuals in the same habitat have the same unknown home range radius. Individual encounters in the study area are recorded using camera traps. The schematic of the capture–recapture model [22] is depicted in Figure 2 and is discussed below.

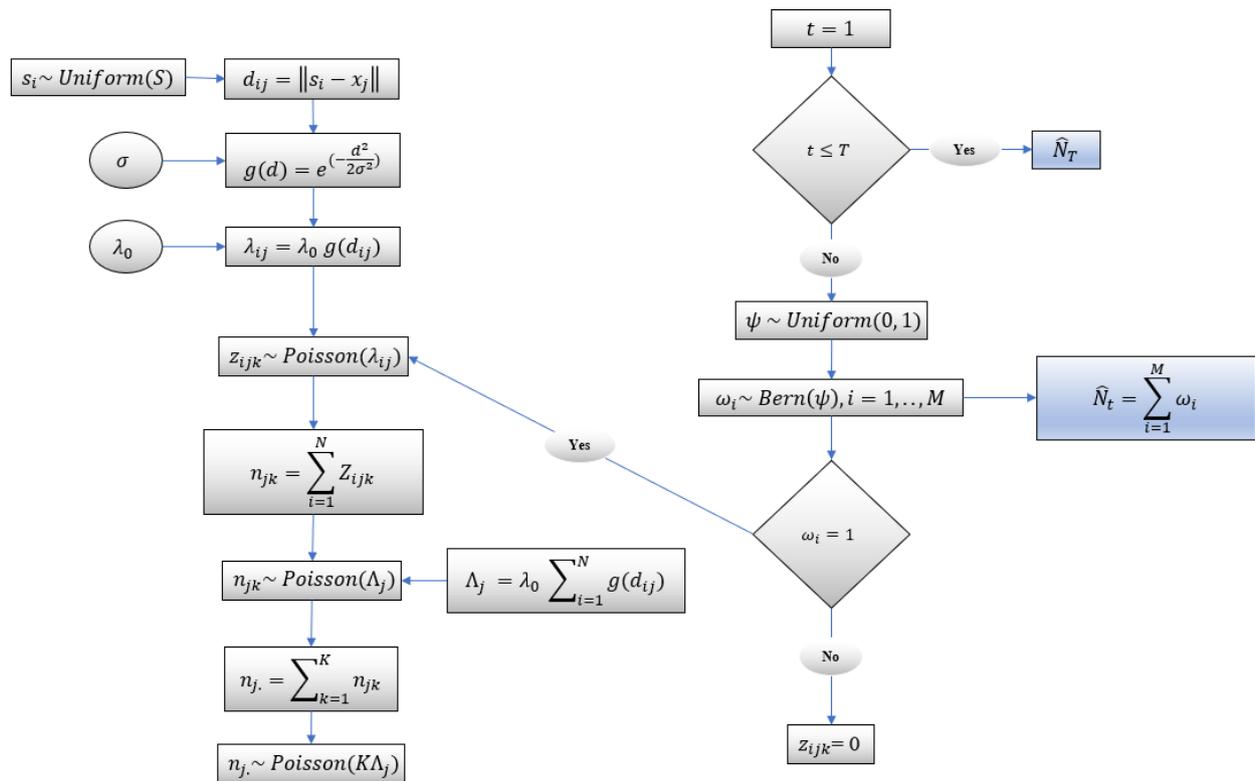


Figure 2. Schematic of capture–recapture model.

Using distance sampling, the distance between the camera locations and the unknown center of activities is reflected in the model. It is assumed that an individual  $i$  has a fixed center of activity defined with the coordinates  $s_i = (s_x, s_y)$  where  $i = 1, 2, \dots, N$ , and  $N$  centers of activities are randomly distributed over the area of study  $S$ . A bivariate uniform prior is used to model the unknown activity center  $s_i$ :

$$s_i \sim \text{Uniform}(S), \tag{1}$$

There are  $J$  camera locations; each is defined by the coordinate  $x_j, j = 1, 2, \dots, J$ . Notice that an individual can be detected at multiple cameras and/or at multiple times by the same camera during a sampling occasion. A Poisson distribution is used to model a camera encounter history  $z_{ijk}$  for individual  $i$ , at camera  $j$ , on occasion  $k$ :

$$z_{ijk} \sim \text{Poisson}(\lambda_{ij}), \tag{2}$$

where  $\lambda_{ij}$  is the encounter rate for individual  $i$  at camera  $j$ . The expected number of the captures or detections of an individual  $i$  at camera  $j$ , which is a function of the Euclidean distance between activity center  $s_i$  and the camera location  $d_{ij} = \|s_i - x_j\|$ , is

$$\lambda_{ij} = \lambda_0 g_{ij}, \tag{3}$$

where  $\lambda_0$  is the baseline encounter rate, and  $g_{ij}$  is a function of the distance which monotonically decreases and is modeled using a half Gaussian function:

$$g_{ij} = \exp\left(\frac{-d_{ij}^2}{\sigma^2}\right), \tag{4}$$

where  $\sigma$  is a scale parameter and will be estimated using the collected data. If an individual can be captured once during a sampling occasion, the encounter history takes binary values; that is,  $z_{ijk}$  takes a value of one if the individual  $i$  is captured, or zero otherwise. However, an individual can be captured more than once during a sampling occasion. In this case,  $z_{ijk}$  will be the number of times that the individual  $i$  has been encountered at camera  $j$  on occasion  $k$ . Therefore, a  $(J \times K)$  encounter history matrix is considered for each individual. Obviously, the capture histories  $z_{ijk}$  cannot be directly observed for unmarked individuals. A data augmentation method has been implemented to estimate the unknown population size. The number of camera encounters at camera  $j$  on occasion  $k$  is modeled by

$$n_{jk} = \sum_{i=1}^N z_{ijk}. \tag{5}$$

The full conditional latent encounter data are defined by a multinomial distribution:

$$\{z_{1jk}, z_{2jk}, \dots, z_{Njk}\} \sim \text{Multinomial}\left(n_{jk} \{\pi_{1j}, \pi_{2j}, \dots, \pi_{Nj}\}\right), \tag{6}$$

where  $\pi_{ij} = \frac{\lambda_{ij}}{\sum_{i=1}^N \lambda_{ij}}$ . The camera encounter counts are modeled using a Poisson distribution:

$$n_{jk} \sim \text{Poisson}(\Lambda_j), \tag{7}$$

where

$$\Lambda_j = \lambda_0 \sum_{i=1}^N g_{ij}. \tag{8}$$

The number of camera encounters at camera  $j$  can be obtained by

$$n_j = \sum_{k=1}^K n_{jk}. \tag{9}$$

Because  $\Lambda_j$  and  $K$  are independent,

$$n_j \sim \text{Poisson}(K\Lambda_j). \tag{10}$$

In the data augmentation method used in Chandler and Royle [22], Royle et al. [17], and Royle and Dorazio [23], the camera encounter histories were augmented with a set of all-zero camera encounter histories to create a hypothetical augmented population of size  $M$  in the study area. The augmented parameter  $M$  is an integer and is recommended to be much greater than unknown  $N$ , i.e.,  $M \gg N$ , to avoid the truncation of the posterior distribution of  $N$ . Notice that, a very large value of  $M$  will increase the computational time. Uninformative prior distributions are assumed for the unknown parameters. Prior distributions of  $\lambda_0$ ,  $\sigma$ , and  $\psi$  are considered Uniform  $(0, 1)$ , where  $\psi$ , probability of success, is the probability that an individual in the occupancy model of size  $M$  is a member of

the true population of size  $N$ . A binomial prior distribution,  $N \sim \text{Binomial}(M, \psi)$ , is assumed for  $N$  where  $\psi \sim \text{Uniform}(0, 1)$ . Assuming a discrete uniform distribution for the detection of individuals in the hypothetical population of size  $M$ ,  $L = M - n$  individuals are associated with all-zero encounter histories. In turn, indicator variables  $\omega_1, \omega_2, \dots, \omega_M$  are introduced such that

$$\omega_i = \begin{cases} 0, & \text{if the individual } i \text{ is not a member of the population} \\ 1, & \text{if the individual } i \text{ is a member of the population} \end{cases}, \tag{11}$$

where  $\omega_i \sim \text{Bernoulli}(\psi)$ ,  $i = 1, 2, \dots, M$ , with expected value  $E(\omega_i) = \psi$  and variance  $\text{Var}(\omega_i) = \psi(1 - \psi)$ . Hence, the encounter data for each individual in the augmented population can be modeled by

$$z_{ijk} \mid \omega_i = 1 \sim \text{Poisson}(\lambda_{ij}\omega_i), z_{ijk} \mid \omega_i = 0 \sim I(z_{ijk} = 0), \tag{12}$$

and in turn, the population size can be obtained by

$$\hat{N} = \sum_{i=1}^M \omega_i. \tag{13}$$

Assuming mutual independence of the prior distributions, the joint prior distribution is

$$P(\psi, \lambda_0, \sigma) \propto P(\psi)P(\lambda_0)P(\sigma) \tag{14}$$

and in turn, the joint posterior distribution of the parameters is

$$[z, \omega, s, \psi, \lambda_0, \sigma \mid n, X] \propto \left\{ \prod_{i=1}^M \left\{ \prod_{j=1}^J \prod_{k=1}^K [n_{jk} \mid z_{ijk}] [z_{ijk} \mid \omega_i, s_i, \sigma, \lambda_0] \right\} [\omega_i \mid \psi][s_i] \right\} [\psi][\lambda_0][\sigma]. \tag{15}$$

where  $X$  is a coordinate matrix of camera locations. Notice that in the original model, the assumed prior distributions for  $\lambda_0$  and  $\sigma$  are uninformative. A spatial Metropolis–Gibbs Markov chain Monte Carlo (MCMC) algorithm for estimating the model parameters is used in Chandler and Royle [22].

### 2.2. Proposed Method

For a random sample  $Y$  from a given population with unknown probability distribution  $f(Y, \theta)$ , the unknown parameter  $\theta$  can be estimated using a point estimator to construct a confidence interval for the unknown parameter. In Bayesian statistics, the samples are often generated from an uninformative prior. However, an informative prior to sample parameter  $\theta$  can be inferred empirically or could be available from previous studies [21–23]. In the next section, an informative prior to sample  $\psi$  is introduced.

### 2.3. Sensitivity of the Model to $\psi$

The probability of success  $\psi$  that an individual is a member of the population is

$$\psi = \frac{N}{M}, \tag{16}$$

where  $N$  is the unknown true population size, and  $M$  is the size of the zero-inflated population. In the data augmentation model, an arbitrary large zero-inflated population is generated. As a result, the upper bound of augmented population size  $M$  can increase indefinitely. It has been demonstrated using simulations that by increasing the zero-inflated population size, the true estimation error will be increased. In turn, the spatial models can substantially overestimate the unknown population size. To address the inflated estimation error due to the inflated size of the population, we suggest bounding the size of the augmented data. A practical lower bound of the augmented data is twice as high as the

unknown population size. By increasing the size of the inflated population above twice the true population size, the error of the estimated population size will increase, and in turn, the width of the credible interval will be increased.

To investigate the sensitivity of the model to  $\psi$ , simulations were designed, and the impact of  $\psi$  on the estimated population size was studied. In the first set of simulations, an uninformative prior with no constraints was considered to sample the probability of success  $\psi$ . It means all values in the range of  $[0, 1]$  will be accepted for  $\psi$  by the Gibbs sampling. In the next set of simulations, different constraints for sampling  $\psi$  were enforced. An estimate of  $\psi$  can be obtained by  $\hat{\psi} = \frac{\hat{N}}{M}$ , where  $\hat{\psi}$  is sampled from a Beta distribution with parameters:

$$\alpha = 1 + \sum_{i=1}^M \omega_i, \tag{17}$$

and

$$\beta = 1 + M - \sum_{i=1}^M \omega_i, \tag{18}$$

The estimated number of individuals, in Equation (13), is given by  $\hat{N} = \sum_{i=1}^M \hat{\omega}_i$ . The expected population size at camera  $j$  is given by

$$E(\hat{N}) = E\left[\sum_{i=1}^M \hat{\omega}_i\right] = \sum_{i=1}^M E[\hat{\omega}_i] = \sum_{i=1}^M I_i \hat{\psi}, \tag{19}$$

where  $I_i$  is the indicator function for individual  $i$ , and  $\hat{\psi}$  is the estimated probability of success, that is, the probability that an individual belongs to the true population. The variance of the estimated  $\hat{N}$  can be derived from

$$Var(\hat{N}) = Var\left[\sum_{i=1}^M \hat{\omega}_i\right] = \sum_{i=1}^M Var[\hat{\omega}_i] = \sum_{i=1}^M I_i \hat{\psi}(1 - \hat{\psi}), \tag{20}$$

and the standard deviation of the estimated  $\hat{N}$  is

$$Sd_{\hat{N}} = \sqrt{\sum_{i=1}^M I_i \hat{\psi}(1 - \hat{\psi})} = \sqrt{M \hat{\psi}(1 - \hat{\psi})}, \tag{21}$$

$$Sd_{\hat{\psi}} = \sqrt{\frac{\hat{\psi}(1 - \hat{\psi})}{M}}. \tag{22}$$

#### 2.4. Autocorrelation Plot

The autocorrelation plot is an effective tool to assess the correlation of the samples produced by a Markov chain and inspect whether the samples are well mixed [24]. The correlation coefficient range is  $[-1, 1]$ , with  $-1$  indicating perfect negative correlation, zero representing no correlation, and one indicating perfect positive correlation. It will also quantify the correlation between the current value of the chain and its past values (lags). Generated samples by MCMC from one iteration to the next will be somewhat correlated. In a well-mixed chain, the correlation is small, and the autocorrelation drops relatively quickly. However, if the chain does not mix well, samples will be highly correlated, and the correlation will decay slowly. In turn, a large number of iterations is required to reach the stationary distribution of the Markov chain [25,26]. Lag  $k$  autocorrelation represents the autocorrelation between the current sample and  $k$ th preceding sample [19].

#### 2.5. Effective Sample Size

Effective sample size (ESS) is another way to study the convergence of the chain. ESS provides an estimated number of independent observations equivalent to the samples

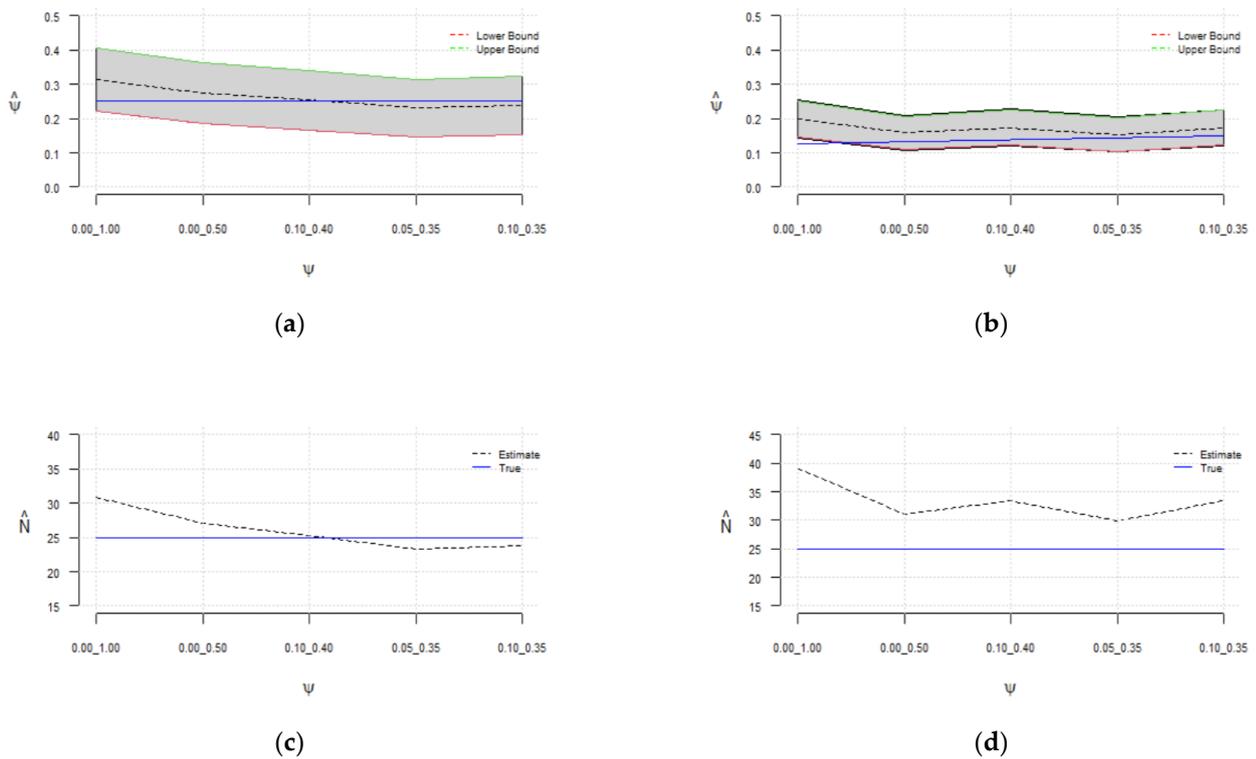
generated by a Markov chain [27] that iterates  $T$  times (Figure 2). In other words, it represents the number of independent samples in the simulation and gives an estimate of how well the simulation represents the target distribution. As we have mentioned, the samples generated by MCMC are somewhat correlated. It means, less information is provided by highly correlated or poor mixing chains. A high ESS indicates a well-mixed chain, while a low ESS indicates poor convergence. A chain with a low ESS must run for a larger number of iterations to improve the convergence of the chain toward its stationary distribution [26].

### 3. Results

The simulations were implemented using R v4.2.2 (a statistical analysis programming language) within the RStudio and a range of specialized packages for Markov chain Monte Carlo (MCMC), the Metropolis–Hasting algorithm, and the Gibbs sampler (see [19,28–31]). Statistical simulations were implemented to study the sensitivity of the spatial capture model to the data augmentation parameter  $L$  (added number of zeros) by estimating unknown population size  $N$ , home range  $\sigma$ , and  $\lambda_0$ . In this model, a hypothetical population size ( $M$ ) for the upper bound of the true population size ( $N$ ) is selected such that  $M = N + L$ . In turn, the estimated  $N$  can assume values between zero and  $M$ . Several simulations were performed to test the sensitivity of the model to the selected value of  $M$  for the true value of  $N = 25$ , and the results are shown in Table 1 and Figure 3. The estimated values of  $\sigma$ ,  $\lambda_0$ ,  $\psi$ , and  $N$  were calculated along with the ESS and  $Lag10$  autocorrelation. The estimated population size  $N$  is obtained by running simulations with and without constraint on the sampling range of  $\psi$  (Table 1). For  $M = 100$ , the true probability of success ( $\psi = \frac{N}{M}$ ) is 0.25, and it is 0.125 for  $M = 200$ .

**Table 1.** Summary of the estimated mean of  $\sigma$ ,  $\lambda_0$ ,  $\psi$ , and population size  $N$  for different ranges of  $\psi$  and  $M \in \{100, 200\}$ .

$M$	$\psi$	$\hat{\sigma}$	$\hat{\lambda}_0$	$\hat{\psi}$	$Sd_{\hat{\psi}}$	$ \%e_{\hat{\psi}} $	$\hat{N}$	$Sd_{\hat{N}}$	$ESS_N$	$ESS_{\psi}$	$Lag10_N$	$Lag10_{\psi}$
100	0.00–1.00	0.437	0.589	0.313	0.046	25.200	30.887	2.577	1113.2	1297.6	0.668	0.625
	0.00–0.50	0.451	0.606	0.274	0.045	9.600	27.107	2.322	1850.8	2180.9	0.482	0.412
	0.10–0.40	0.465	0.585	0.253	0.043	1.200	25.255	2.185	2679.4	3232.7	0.403	0.324
	0.05–0.35	0.479	0.584	0.231	0.042	7.600	23.326	2.036	2505.5	3145.0	0.395	0.309
	0.10–0.35	0.471	0.589	0.237	0.043	5.200	23.862	2.077	2738.4	3681.9	0.352	0.260
200	0.00–1.00	0.416	0.596	0.199	0.028	59.200	39.231	2.501	250.6	263.3	0.909	0.894
	0.00–0.50	0.443	0.579	0.158	0.026	26.400	30.979	2.030	1066.8	1140.4	0.674	0.623
	0.10–0.40	0.415	0.579	0.174	0.027	39.200	33.433	2.192	2340.8	2727.2	0.496	0.431
	0.05–0.35	0.443	0.579	0.153	0.025	22.400	29.922	1.969	2082.8	2203.2	0.547	0.486
	0.10–0.35	0.411	0.592	0.173	0.027	38.400	33.340	2.184	3017.5	3701.7	0.446	0.375



**Figure 3.** Sensitivity of the estimated probability of success ( $\hat{\psi}$ ) to the selected range of  $\psi$  for  $M = 100$  (a) and  $M = 200$  (b); sensitivity of the estimated population size ( $\hat{N}$ ) to the selected range of  $\psi$  for  $M = 100$  (c) and  $M = 200$  (d).

It can be observed (in Table 1 and Figure 3) that the estimated  $N$  has assumed a broad range of values. The sensitivity of the model to the added number of zeros  $L$  is more noticeable when sampling the probability of success  $\psi$  from  $[0, 1]$ , i.e., without constraint. In all cases, the estimated values of  $N$  are more accurate with the constraints on  $\psi$ . It was observed that by decreasing the sampling range of the  $\psi$ , the effective sample size will increase, while the rejection rate and in turn computational cost increase.

It is recommended to choose large values for the data augmentation parameter  $M$ . However, it must be pointed out that  $M$  may not be increased indefinitely, as it may produce very small  $\psi = \frac{N}{M}$  close to zero and in turn significant overestimation of  $N$ . Hence, two different constraints,  $[0.0, 0.5]$  and  $[0.1, 0.4]$ , were set for the range of parameter  $\psi$ , and the simulation results were compared with the uninformed prior of  $\psi$  and sampling it from  $[0, 1]$ . We ran 50 simulations for each scenario and computed the average to be able to generalize the results.

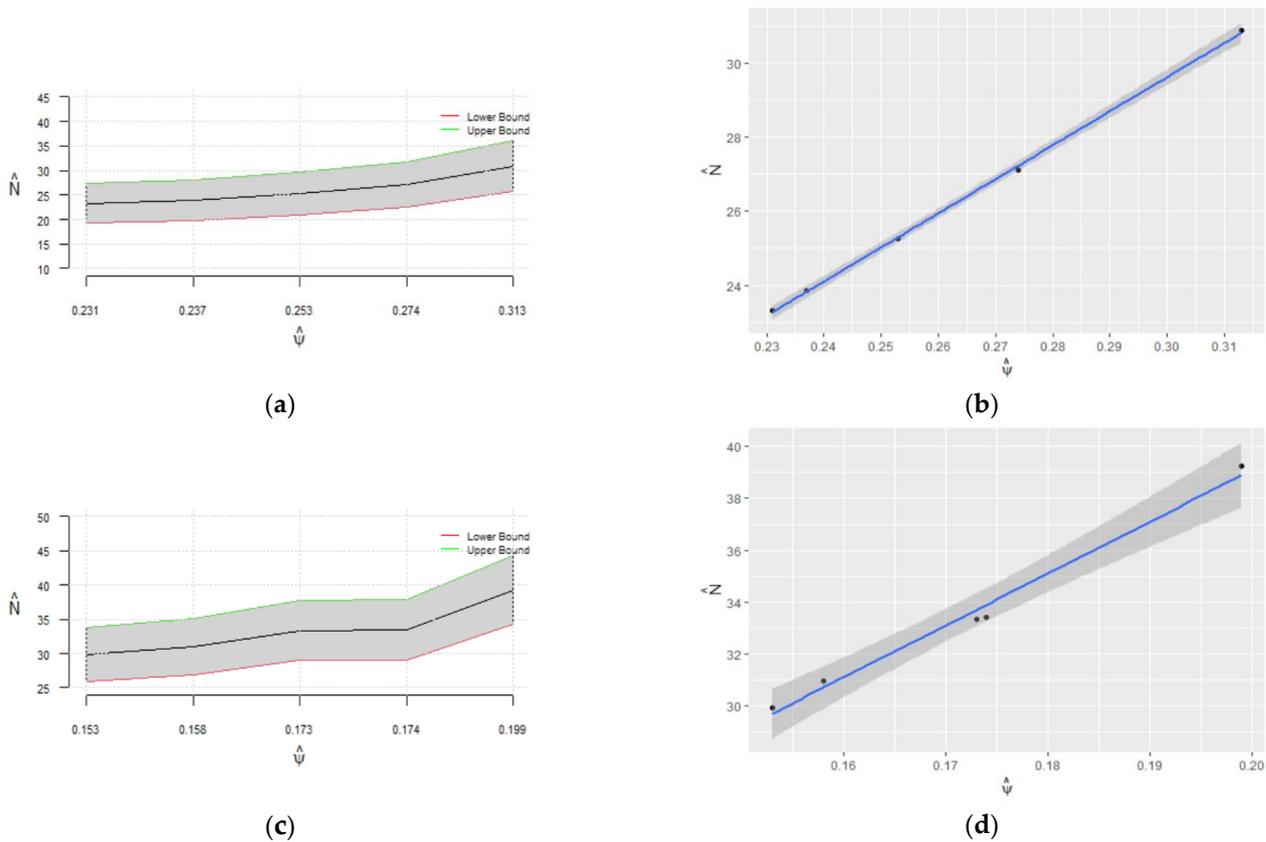
As depicted in Figure 3, regardless of the number of added zeros  $L$ , the sensitivity decreases by constraining the probability of success, and true  $\psi$  is contained in the estimated confidence interval highlighted in gray. Furthermore, the estimated  $N$  is consistent for  $\psi = 0.25$ , as depicted in Figure 2. A fair estimate of  $N$  is obtained when sampling  $\psi$  from  $[0.0, 0.5]$  regardless of  $L$ . The results show that a reasonable estimate of  $N$  between 23 and 33 can be obtained by constraining  $\psi$ , and with a higher probability of success of 0.25, the estimated  $N$  is more accurate (between 23 and 27).

Table 2 shows the estimated values of  $\sigma, \lambda_0, \psi$ , and  $N$ , along with ESS and  $Lag10$  autocorrelation for  $\psi$  and  $N$ . For  $M = 100$ , the estimated value of the population size  $N$  in the case of no constraint on the parameter  $\psi$  is about 32, with an ESS equal to 1862. After enforcing the constraint  $[0, 0.5]$  on the range of  $\psi$  to reject all samples with  $\psi > 0.50$ , the estimated value of  $N$  is about 27 with an ESS of 4368. Moreover, by limiting the range of  $\psi$  to  $[0.10, 0.40]$ , the estimated population size is 26 with the ESS of 7678. Pointwise nonparametric confidence intervals for population size given the range of  $\psi$  are compared

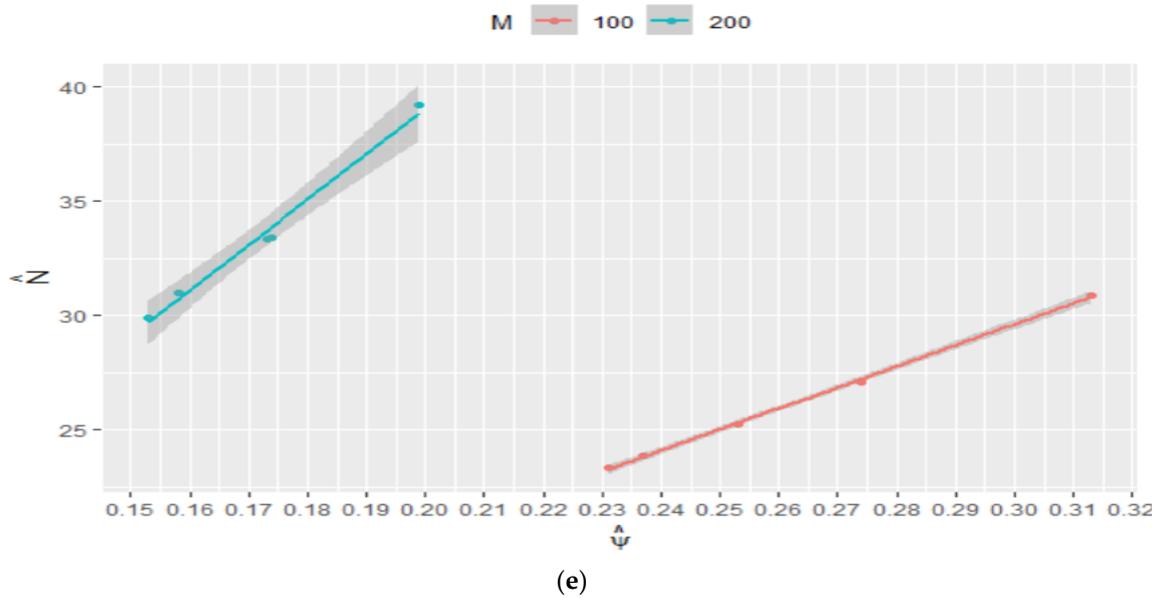
for different values of  $M$  (Figure 4). As we can see in this figure, more robust estimates of the population size and  $\psi$  are obtained for  $M = 100$ , while the estimation error and bias are increased for  $M = 200$ . Figure 5 shows the estimated  $\psi$  and its confidence interval regarding the ESS of  $\hat{N}$  and ESS of  $\hat{\psi}$  for different values of  $M$ . It is clear that the estimated  $\psi$  is more robust and converges to its true value within an ESS of 3000 for  $M = 100$ , while it is overestimated and does not converge to the true value of  $\psi$  for  $M = 200$ .

**Table 2.** Summary of 50 runs of the estimated mean of  $\sigma$ ,  $\lambda_0$ ,  $\psi$ , and population size  $N$  for different ranges of  $\psi$  and  $M \in \{100, 200\}$ .

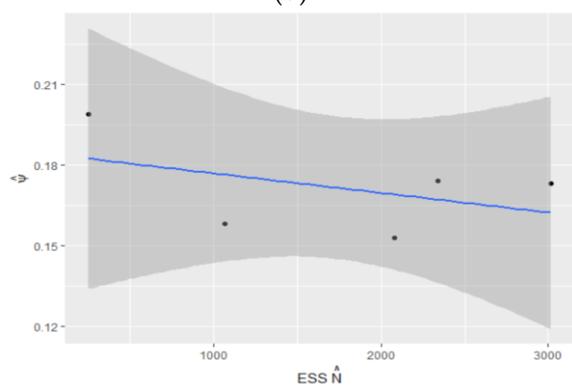
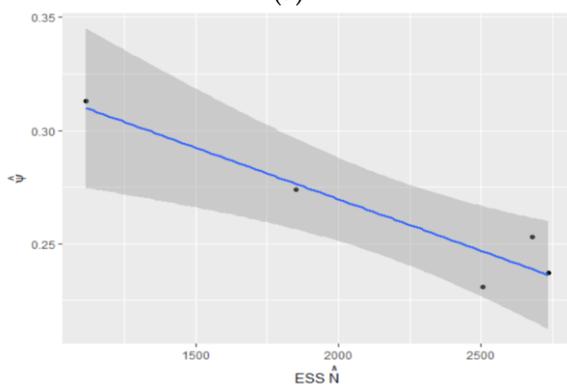
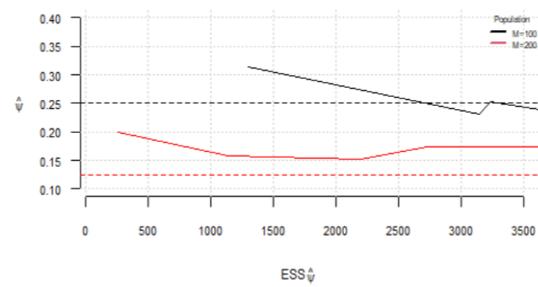
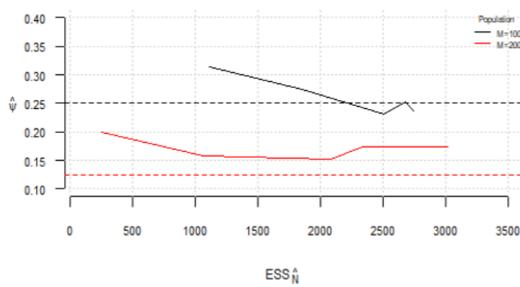
$M$	$\psi$	$\hat{\sigma}$	$\hat{\lambda}_0$	$\hat{\psi}$	$ \%e_{\hat{\psi}} $	$\hat{N}$	$ESS_N$	$ESS_{\psi}$	$Lag10_N$	$Lag10_{\psi}$
100	0.00–1.00	0.560	0.497	0.323	29.200	31.943	1674.3	1862.4	0.635	0.593
	0.00–0.50	0.563	0.546	0.273	9.200	27.173	3661.7	4368.0	0.393	0.334
	0.10–0.40	0.516	0.557	0.261	4.400	26.251	5987.0	7677.7	0.270	0.205
200	0.00–1.00	0.546	0.527	0.170	36.000	33.321	1490.3	1778.8	0.670	0.620
	0.00–0.50	0.527	0.564	0.158	26.400	31.026	2647.5	3152.3	0.528	0.473
	0.10–0.40	0.496	0.545	0.165	32.00	31.168	3880.0	4606.6	0.412	0.355



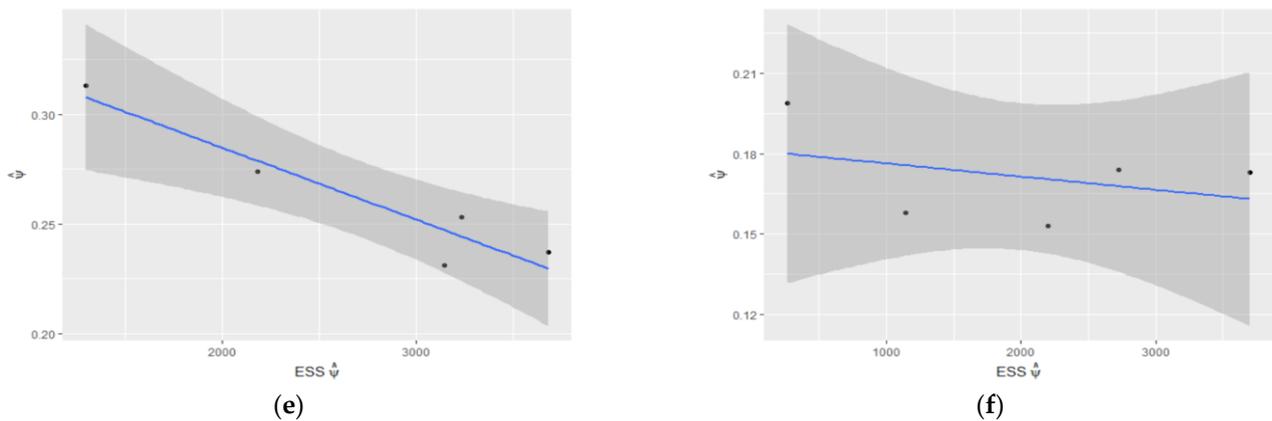
**Figure 4.** Cont.



**Figure 4.** Pointwise 95% confidence intervals for  $N$  regarding  $\hat{\psi}$  for  $M = 100$  (a) and  $M = 200$  (c); nonparametric 95% confidence intervals for  $N$  regarding  $\hat{\psi}$  for  $M = 100$  (b) and  $M = 200$  (d); (e) combined plot of the confidence intervals in panels (b,d) for comparison.

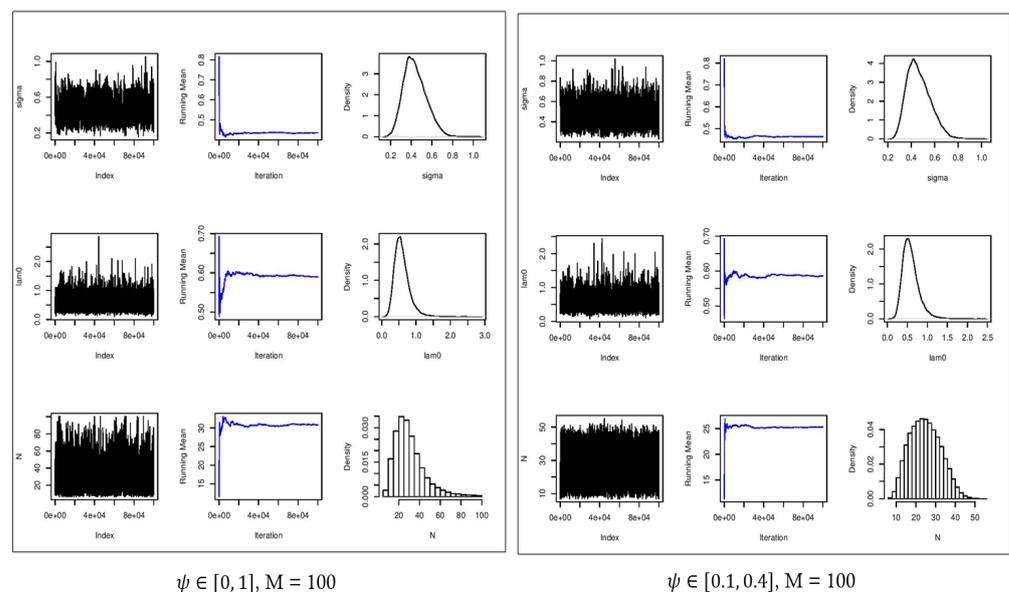


**Figure 5.** Cont.

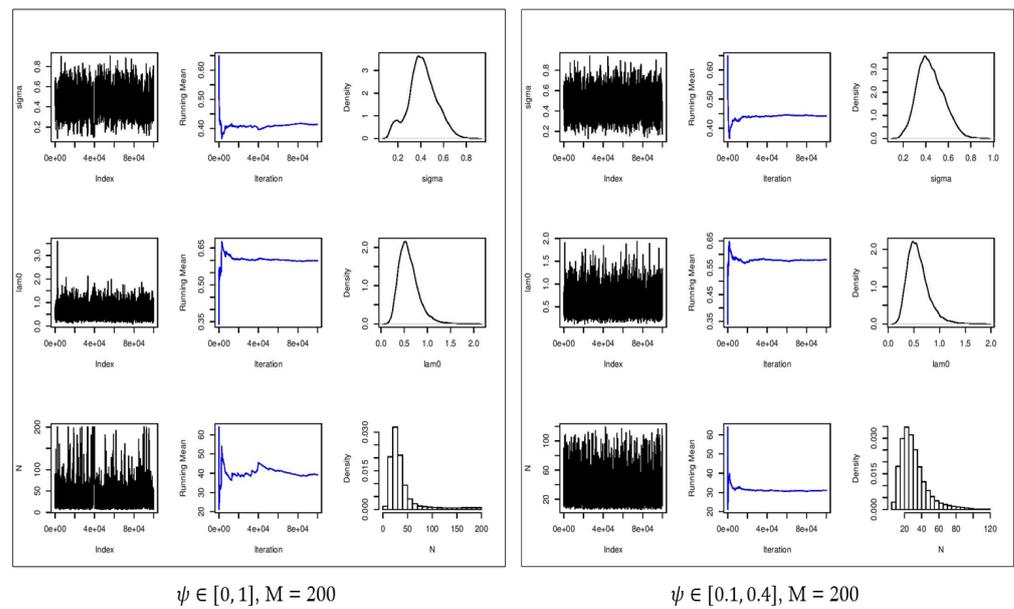


**Figure 5.** Estimated  $\psi$ : (a) vs. ESS of  $\hat{N}$  and (b) vs. ESS of  $\hat{\psi}$ ; estimated 95% confidence interval for  $\psi$  regarding ESS of  $\hat{N}$  for  $M = 100$  (c) and  $M = 200$  (d); estimated 95% confidence interval for  $\psi$  regarding ESS of  $\hat{\psi}$  for  $M = 100$  (e) and for  $M = 200$  (f).

Figures 6 and 7 show the estimated parameters  $\sigma$  and  $\lambda_0$  in comparison with their true values. In all cases, the densities of the estimated parameters converged to the stationary distribution. With the constraint on parameter  $\psi$ , the mixing of samples in the chains were improved, providing more accurate estimates of  $N$ . With no constraint on the parameter  $\psi$  and  $M = 200$ , the Monte Carlo average of the estimated population size  $N$  was about 33 with an ESS of 1779. By enforcing the constraints on the parameter  $\psi$ , the estimated population size  $N$  was 31 which demonstrated a smaller absolute error as well as improved accuracy with a narrower confidence interval in comparison with the scenario with an uninformative prior distribution (Table 2). We should point out that with  $M = 200$ , the augmented population size is eight times larger than the true population size of  $N = 25$ , as a result of which,  $N$  is substantially overestimated, and  $N = 25$  is well outside of the confidence interval of  $[LB = 34.0, UB = 44.0]$  (Table 3).



**Figure 6.** Convergence plots (first column), running mean (second column), and estimated posterior densities (third column) for  $\sigma$  (first row),  $\lambda_0$  (second row), and  $N$  (third row) for  $M = 100$ . Left: no constraint on  $\psi$ . Right:  $\psi$  constrained on  $[0.1, 0.4]$ .



**Figure 7.** Convergence plots (first column), running mean (second column), and estimated posterior densities (third column) for  $\sigma$  (first row),  $\lambda_0$  (second row), and  $N$  (third row) for  $M = 200$ . Left: no constraint on  $\psi$ . Right:  $\psi$  constrained on  $[0.1, 0.4]$ .

**Table 3.** Estimated 95% confidence interval for population size  $N$  and probability of success  $\psi$ .

	95% CI				95% CI	
	$\hat{N}$	$\hat{N} - 2 * Sd_{\hat{N}}$	$\hat{N} + 2 * Sd_{\hat{N}}$	$\hat{\psi}$	$\hat{\psi} - 2 * Sd_{\hat{\psi}}$	$\hat{\psi} + 2 * Sd_{\hat{\psi}}$
$M = 100$	30.887	25.733	36.041	0.313	0.220	0.406
	27.107	22.463	31.751	0.274	0.185	0.363
	25.255	20.886	29.624	0.253	0.166	0.340
	23.326	19.255	27.397	0.231	0.147	0.315
	23.862	19.707	28.017	0.237	0.152	0.322
$M = 200$	39.231	34.230	44.232	0.199	0.143	0.255
	30.979	26.919	35.039	0.158	0.106	0.210
	33.433	29.049	37.817	0.174	0.120	0.228
	29.922	25.984	33.860	0.153	0.102	0.204
	33.340	28.972	37.708	0.173	0.120	0.226

The estimated  $N$  for the constrained  $\psi$  in  $[0.10, 0.40]$  and  $M = 100$  has the minimum absolute error ( $|N - \hat{N}| = |25 - 26| = 1$ ) and provides the highest number of independent samples with  $ESS = 7678$  generated through 90,000 MCMC iterations. With an uninformative prior for  $\psi$ , the range of estimated  $\sigma$  is  $[0.30, 2.03]$  with an average of 0.56. By regularizing  $\psi$  and constraining its range to  $[0, 0.50]$ , the range of the estimated  $\sigma$  is  $[0.40, 1.70]$ . Moreover, by constraining  $\psi$  to  $[0.10, 0.40]$ , the range of the estimated  $\sigma$  is  $[0.39, 0.8]$  with an average of 0.52 and the narrowest range for the estimated  $\psi$  containing true  $\psi$  and with the lowest absolute error of the estimate. The estimated  $\lambda_0$  for the aforementioned ranges of  $\psi$  is 0.497, 0.546, and 0.557, respectively. The estimated population size ranges from 8 to 61 with an average of 32 for no restriction on  $\psi$ , from 6 to 37 with an average of 27 for  $\psi$  belongs to  $[0, 0.50]$ , and from 16 to 37 with an average of 26 for  $\psi$  belongs to  $[0.1, 0.40]$ . The estimated value of the population size is more accurate, and the range of the estimated value is narrower after regularizing  $\psi$  (Tables 1–3). Moreover, the

standard error for the estimated population size ranges from 5.8 to 24.7, with an average of 14.37 with no restriction on  $\psi$ . The standard error is between 5.8 and 13 with an average of 8.56 for  $\psi \leq 0.50$ . With  $\psi$  constrained between 0.10 and 0.40, the standard error ranges from 5 to 9 with an average of 6.8.

The estimated values for  $\sigma$  with  $M = 200$  are 0.546, 0.527, and 0.496 (Table 2) for all three scenarios regarding  $\psi$ , i.e., no constraint,  $[0, 0.5]$ , and  $[0.1, 0.4]$ . Also, the estimated values of  $\lambda_0$  are 0.527, 0.564, and 0.545 (Table 2) for the aforementioned constraints on  $\psi$ . As it can be observed, the range of the estimated parameter is smaller with a constraint on  $\psi$  (Tables 1 and 2). The estimated population size ranges from 6 to 107 with an average of 33 with no constraint on  $\psi$ , from 11 to 61 with an average of 31 by constraining  $\psi$  to  $[0, 0.5]$ , and from 19 to 47 with an average of 31 by constraining  $\psi$  to  $[0.10, 0.40]$ . As depicted in Figure 7, we can see that by constraining  $\psi$  to  $[0.10, 0.40]$ , the distribution of  $\hat{N}$  has a shorter tail and converges to the stationary posterior distribution. Moreover, there is better mixing of the chains providing more accurate estimates of the parameters.

#### 4. Discussion

By constraining the parameter  $\psi$  to the range  $[0, 0.50]$ ,  $M$  can assume any value greater than  $2N$  ( $M \geq 2N$ ). In contrast, by restricting  $\psi$  to the range  $[0.10, 0.40]$ ,  $M$  is double-bounded within  $2.5N \leq M \leq 10N$ . Enforcing an upper bound of  $10N$  is a sufficient assumption that satisfies  $M \gg N$  recommended for spatial capture models, while we prevent indefinitely large  $M$  that can result in prohibitively small estimates of  $\psi$  (close to zero) with a highly skewed distribution. In practical applications, determining an accurate prediction for the value of  $M$  can be challenging. In cases where no prior information about  $M$  is known, the range  $[0, 0.5]$  can be considered for  $\psi$ . However, if prior information about the population size is known, and the camera grid provides adequate coverage of the habitat of interest, the range  $[0.1, 0.4]$  for  $\psi$  is preferred to avoid underestimation of  $\psi$  toward zero.

It was observed that the estimated population size using the spatial capture models with camera encounters is subject to overestimation and bias. The estimation bias tends to increase by increasing  $M$ . ESS and lag 10 can be used to control the convergence of MCMC and in turn controlling the estimation bias. Specifically, lower ESS values are associated with higher bias, whereas higher ESS values are associated with lower bias. Additionally, larger values of lag 10 tend to correspond to higher levels of bias, and vice versa.

In terms of  $M$ , by setting it within five times of  $N$ , the estimated  $N$  is relatively accurate with low bias, while by increasing  $M$  toward 10 times of  $N$ , the estimated  $N$  converges to higher values (overestimates) with increased bias. It should be noticed that the values of  $\psi$  and  $\lambda_0$  depend on each other, and in turn, their estimated values are not mutually exclusive. Specifically, there is a trade-off between the two estimated parameters. If the estimate of  $\lambda_0$  increases, the estimates of  $\psi$  will decrease, and vice versa. Potentially, the estimated population size could be further improved by considering a prior distribution for  $\lambda_0$ , which is the subject of our future work. Nonetheless, regularizing the value of  $\psi$  by a prior distribution in conjunction with controlling the ESS and log 10 improved the accuracy of the estimated population size.

#### 5. Conclusions

Population management is important to preserve the populations of endangered species and to control the populations of invasive species. Estimating the population size is an essential task in managing the population. Collecting a random sample of individuals from the population is a feasible practice to study the population when it is not possible to count every individual. Capture–recapture methods based on spatial sampling and count models have become the standard methods in the analytical framework for ecological statistics and are widely used for population analysis to estimate population size and density.

The unknown size of a population is considered as the number of successes in a binomial distribution. The parameters of this binomial distribution are an unknown number of independent trials and an unknown probability of success. This is an inverse problem as

none of the parameters including the population size, the probability of success, and the number of independent trials is known. An initial realization of the unknown number of trials is required to approach a solution while the accuracy of the estimated parameters strongly relies on the initial value for the zero-inflated population size. Hence, the estimated population size in spatial capture models is quite sensitive to the size of a zero-inflated population and in turn to the estimated probability of success. This is a typical count model with censored data as captured individuals are not identified, and some individuals are not detected. To address this problem and improve the estimation accuracy, in this research, we investigated the sensitivity and accuracy of the spatial capture-identified models (using virtual encounters) with the objective of improving their robustness. Statistical simulations were implemented to study the sensitivity of the spatial capture models to the zero-inflated population parameter  $\psi$ .

In capture-identified models, augmented population size ( $M$ ) is allowed to be increased indefinitely, and so  $\psi$  will be allowed to belong to  $[0, 1]$ . As a result, the population size  $N$  will be highly overestimated, and the accuracy of the estimate declines for large values of  $M$  while the estimated  $\psi$  moves toward zero. Consequently, the credible interval becomes wider, and the distribution of  $\hat{N}$  displays a heavy and long right tail. Moreover, the true estimation error and estimation bias increase, while ESS tends to remain low. To address the aforementioned issues, a lower and an upper bound for  $\psi$  were recommended to prevent overestimation of population size that is caused by excessive underestimation of  $\psi$  when it is not regularized. In this way, the accuracy and bias were retained providing fairly narrow credible intervals, while ESS was improved.

**Author Contributions:** Conceptualization, M.J., F.H., R.D.B. and N.N.K.; methodology, M.J., F.H., R.D.B. and N.N.K.; software, M.J., F.H., R.D.B. and N.N.K.; validation, M.J., F.H., R.D.B. and N.N.K.; formal analysis, M.J. and F.H.; investigation, M.J., F.H., R.D.B. and N.N.K.; resources, N.N.K.; data curation, M.J., F.H. and R.D.B.; writing—original draft preparation, M.J., F.H., R.D.B. and N.N.K.; writing—review and editing, M.J., F.H., R.D.B. and N.N.K.; visualization, M.J., F.H., R.D.B. and N.N.K.; supervision, N.N.K.; project administration, N.N.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** No data was collected. Rather, data was randomly sampled in the simulation study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

MCMC	Markov chain Monte Carlo
HSCR	Hierarchical spatial capture–recapture
$N$	Population size
$\hat{N}$	Estimated population size
$K$	Sampling occasion
$z_{ijk}$	Camera encounter history for individual $i$ , at camera $j$ , on occasion $k$
$\lambda_{ij}$	The encounter rate for individual $i$ at camera $j$
$\lambda_0$	The baseline encounter rate
$\sigma$	Home range radius.
$d_{ij}$	The Euclidean distance between activity center $s_i$ and the camera location $x_j$
$n_{jk}$	The number of camera encounters at camera $j$ on occasion $k$
$M$	The augmented parameter (the total number of hypothetical individuals)
$\psi$	Probability of success, i.e., the probability that an individual in the occupancy model of size $M$ is a member of the original model of size $N$
$\alpha$ and $\beta$	Parameters of Beta probability distribution
ESS	The effective sample size
$L$	The number of zeros added to the model $L = M - N$ (data augmentation size)
Lag $k$	The autocorrelation between the current sample and $k$ th preceding sample

## References

1. Pollock, K.H. Capture-Recapture Models: A Review of Current Methods, Assumptions and Experimental Design. *Stud. Avian Biol.* **1981**, *6*, 426–435.
2. Schwarz, C.J.; Seber, G.A. Estimating Animal Abundance: Review III. *Stat. Sci.* **1999**, *14*, 427–456. [[CrossRef](#)]
3. Nichols, J.D. Capture-Recapture Models. *BioScience* **1992**, *42*, 94–102. [[CrossRef](#)]
4. Pollock, K.H. A Capture-Recapture Design Robust to Unequal Probability of Capture. *J. Wildl. Manag.* **1982**, *46*, 752–757. [[CrossRef](#)]
5. Pollock, K.H.; Nichols, J.D.; Brownie, C.; Hines, J.E. Statistical Inference for Capture-Recapture Experiments. *Wildl. Monogr.* **1990**, *48*, 3–97.
6. Karanth, K.U. Estimating Tiger Panthera Tigris Populations from Camera-Trap Data Using Capture—Recapture Models. *Biol. Conserv.* **1995**, *71*, 333–338. [[CrossRef](#)]
7. O’Connell, A.F.; Nichols, J.D.; Karanth, K.U. *Camera Traps in Animal Ecology: Methods and Analyses*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 271.
8. Jaber, M.; Woesik, R.V.; Kachouie, N.N. *Probabilistic Detection Model for Population Estimation*; Old Dominion University: Norfolk, VA, USA, 2018; p. 43.
9. Distiller, G.; Borchers, D.L. A spatially explicit capture–recapture estimator for single-catch traps. *Ecol. Evol.* **2015**, *5*, 5075–5087. [[CrossRef](#)]
10. Engeman, R.; Massei, G.; Sage, M.; Gentle, M.N. Monitoring Wild Pig Populations: A Review of Methods. *Environ. Sci. Pollut. Res.* **2013**, *20*, 8077–8091. [[CrossRef](#)]
11. Royle, J.A.; Young, K.V. A Hierarchical Model for Spatial Capture–Recapture Data. *Ecology* **2008**, *89*, 2281–2289. [[CrossRef](#)]
12. Royle, J.A.; Karanth, K.U.; Gopalaswamy, A.M.; Kumar, N.S. Bayesian Inference in Camera Trapping Studies for a Class of Spatial Capture–Recapture Models. *Ecology* **2009**, *90*, 3233–3244. [[CrossRef](#)]
13. Kery, M.; Gardner, B.; Stoeckle, T.; Weber, D.; Royle, J.A. Use of Spatial Capture-recapture Modeling and DNA Data to Estimate Densities of Elusive Animals. *Conserv. Biol.* **2011**, *25*, 356–364. [[CrossRef](#)] [[PubMed](#)]
14. Borchers, D. A Non-Technical Overview of Spatially Explicit Capture–Recapture Models. *J. Ornithol.* **2012**, *152*, 435–444. [[CrossRef](#)]
15. Royle, J.A.; Chandler, R.B.; Sollmann, R.; Gardner, B. *Spatial Capture-Recapture*; Academic Press: Cambridge, MA, USA, 2013; ISBN 0-12-407152-X.
16. Gardner, B.; Reppucci, J.; Lucherini, M.; Royle, J.A. Spatially Explicit Inference for Open Populations: Estimating Demographic Parameters from Camera-trap Studies. *Ecology* **2010**, *91*, 3376–3383. [[CrossRef](#)] [[PubMed](#)]
17. Royle, J.A.; Dorazio, R.M.; Link, W.A. Analysis of Multinomial Models with Unknown Index Using Data Augmentation. *J. Comput. Graph. Stat.* **2007**, *16*, 67–85. [[CrossRef](#)]
18. Jiao, G.; Liang, J.; Wang, F.; Chen, X.; Chen, S.; Li, H.; Jin, J.; Cai, J.; Zhang, F. Longitudinal Data Analysis Based on Bayesian Semiparametric Method. *Axioms* **2023**, *12*, 431. [[CrossRef](#)]
19. Carlo, C.M. Markov Chain Monte Carlo and Gibbs Sampling. *Lect. Notes EEB* **2004**, *581*, 3.
20. Alotaibi, R.; Nassar, M.; Elshahhat, A. Statistical Analysis of Inverse Lindley Data Using Adaptive Type-II Progressively Hybrid Censoring with Applications. *Axioms* **2023**, *12*, 427. [[CrossRef](#)]
21. Sollmann, R.; Tôrres, N.M.; Furtado, M.M.; de Almeida Jácomo, A.T.; Palomares, F.; Roques, S.; Silveira, L. Combining Camera-Trapping and Noninvasive Genetic Data in a Spatial Capture–Recapture Framework Improves Density Estimates for the Jaguar. *Biol. Conserv.* **2013**, *167*, 242–247. [[CrossRef](#)]
22. Chandler, R.B.; Royle, J.A. Spatially explicit models for inference about density in unmarked or partially marked populations. *Ann. Appl. Stat.* **2013**, *7*, 936–954. [[CrossRef](#)]
23. Royle, J.A.; Dorazio, R.M. Parameter-Expanded Data Augmentation for Bayesian Analysis of Capture–Recapture Models. *J. Ornithol.* **2012**, *152*, 521–537. [[CrossRef](#)]
24. Christensen, O.F.; Roberts, G.O.; Sköld, M. Robust Markov Chain Monte Carlo Methods for Spatial Generalized Linear Mixed Models. *J. Comput. Graph. Stat.* **2006**, *15*, 1–17. [[CrossRef](#)]
25. Kuzmanovska, I. Markov Chain Monte Carlo Methods in Biological Mechanistic Models. 2012. Available online: <https://www.research-collection.ethz.ch/handle/20.500.11850/153590> (accessed on 8 August 2023).
26. Lewis, A. Efficient Sampling of Fast and Slow Cosmological Parameters. *Phys. Rev. D* **2013**, *87*, 103529. [[CrossRef](#)]
27. Martino, L.; Elvira, V.; Louzada, F. Effective Sample Size for Importance Sampling Based on Discrepancy Measures. *Signal Process.* **2017**, *131*, 386–401. [[CrossRef](#)]
28. Fabreti, L.G.; Höhna, S. Convergence Assessment for Bayesian Phylogenetic Analysis Using MCMC Simulation. *Methods Ecol. Evol.* **2022**, *13*, 77–90. [[CrossRef](#)]
29. Zhang, J.; Cui, S. Investigating the Number of Monte Carlo Simulations for Statistically Stationary Model Outputs. *Axioms* **2023**, *12*, 481. [[CrossRef](#)]

30. Jaber, M. A Spatiotemporal Bayesian Model for Population Analysis. 2022. Available online: <https://repository.fit.edu/etd/880/> (accessed on 8 August 2023).
31. Martin, A.D.; Quinn, K.M.; Park, J.H. MCMCpack: Markov Chain Monte Carlo in R. *J. Stat. Softw.* **2011**, *42*, 1–21. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.