

Optimally designing drone-based surveys for wildlife abundance estimation with N-mixture models

Ismael V. Brack¹  | Andreas Kindel¹ | Luiz Flamarion B. de Oliveira² | José J. Lahoz-Monfort^{3,4}

¹Graduate Program in Ecology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

²Department of Vertebrates, National Museum, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

³Quantitative and Applied Ecology Group, School of Biosciences, University of Melbourne, Melbourne, Victoria, Australia

⁴Pyrenean Institute of Ecology (CSIC), Jaca, Spain

Correspondence

Ismael V. Brack

Email: ismaelbrack@hotmail.com

Funding information

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: Finance Code 001

Handling Editor: Chris Sutherland

Abstract

1. Hierarchical N-mixture models have been suggested for abundance estimation from spatiotemporally replicated drone-based count surveys, since they allow modeling abundance of unmarked individuals while accounting for detection errors. However, it is still necessary to understand how these models perform in the wide variety of contexts and species in which drone surveys are being used. This knowledge is fundamental to plan study designs with optimal allocation of scarce resources in ecology and conservation.
2. We conduct a simulation study to address N-mixture model (binomial and multinomial) performance and optimal survey effort allocation in different scenarios of local abundance and detectability of individuals, focusing on their application for drone-based surveys. We also investigate the benefits of using a double-observer protocol (either human or algorithm) in image review to decompose the detection process in availability and perception. Finally, we illustrate our simulation-based survey design considerations by applying them to abundance estimation of marsh deer in the Pantanal wetland (Brazil).
3. Accuracy of abundance estimation with N-mixture models increases with local abundance in sites and especially with the availability of individuals. The optimal design requires more visits at fewer sites when the availability probability is lower, and the optimal design is more flexible as local abundance increases. Two observers checking images can increase the estimator performance even at very high perception probabilities. We quantified how much the use of a double-observer protocol in image review can reduce fieldwork effort while achieving the same accuracy.
4. N-mixture models can deliver accurate abundance estimates from spatiotemporally replicated drone surveys in a wide variety of contexts while accounting for imperfect detection. The improvements achieved by a consciously planned design, rearranging survey efforts among sites and visits, as well as using a second observer in image review, can be crucial to detect trends when monitoring a population or to categorize a species as threatened or not.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

KEYWORDS

abundance modelling, aerial surveys, count data, double observer, effort allocation, hierarchical models, imperfect detection, sampling design

1 | INTRODUCTION

The use of drones (a.k.a. Unmanned Aerial Vehicles or Remotely Piloted Aircrafts) to survey wildlife populations has spread to many different species in many different contexts. Drones are replacing conventional aircraft to survey for instance large herbivores (e.g. Barasona et al., 2014; Rey et al., 2017; Vermeulen et al., 2013), waterbirds (e.g. Hodgson et al., 2018; Hong et al., 2019), marine mammals (e.g. Goebel et al., 2015; Hodgson et al., 2013, 2017) and crocodiles (e.g. Ezat et al., 2018). Furthermore, given the flexibility of this tool, drone surveys are being explored for situations where conventional aircraft were unsuitable before, expanding the ability of aerials surveys to sample smaller species (e.g. canids, Bushaw & Ringelman, 2019; sharks and rays, Kiszka et al., 2016) and in forested areas (e.g. primates, de Melo, 2021; koalas, Hamilton et al., 2020). Thus, drone surveys are improving the way we collect abundance data for ecological studies (e.g. population dynamics) or conservation/management issues (e.g. threatened or invasive species monitoring). The recent developments in machine learning algorithms for image processing (e.g. convolutional neural networks; LeCun et al., 2015; Weinstein, 2017) have greatly improved the efficiency of drone-based surveys to generate abundance estimates.

To estimate the abundance of wildlife populations using drones, counts need to be carried out by searching for a species of interest on the images or footage obtained along with the flights. Counts, either if conducted manually by human observers or using an automated algorithm, are subject to errors that might bias estimation of abundance if not properly addressed. Some individuals may be hidden (e.g. below vegetation, inside a burrow, or underwater) and therefore are not visible on images and undetectable by any reviewing method, or may be similar to the background making them difficult to detect either by a human or an algorithm. An individual can thus be missed, yielding false-negative errors in counts (i.e. imperfect detection), by two different processes: (i) it is unavailable for detection at the time of the flight or (ii) it appears on an image but can be missed by a human observer or an algorithm during reviewing (Brack et al., 2018). As in other wildlife survey methods, processes driving these detection errors in drone-based counts can vary in space and time and may depend on species characteristics (e.g. conspicuity and behavior), habitat features (e.g. tree coverage or water turbidity) and conditions when surveying (e.g. weather; Guillera-Aroita, 2017). Furthermore, some characteristics specific to drone surveys could also affect detectability, such as sensor type (e.g. thermal or visible), pixel resolution and flight height. Manual or automated reviewing procedures may also result in different detection probabilities. While false negatives are expected to be the major source of error in drone aerial counts (Brack et al., 2018), false-positive can also affect counts in drone surveys. For example, other similar species or

a background feature could be misidentified as the target species (misidentification error) or the same individual could be counted more than once (double counts) because of either appearing in overlapped pictures or moving between lines during the flight.

Hierarchical N-mixture models have been suggested to model abundance from spatiotemporally replicated aerial counts since they are a valuable framework for studying unmarked populations while accounting for the sources of imperfect detection (Brack et al., 2018; Christensen et al., 2021; Martin et al., 2015; Williams et al., 2017). In such approach, count data obtained for each visit (i.e. repeated flight) in each site are modelled as a result of (at least) two hierarchically connected processes: local abundance at sites and observation (detection) process of individuals in each visit for each site (susceptible to imperfect detection). Moreover, a double-observer protocol can be applied to image review to permit decomposing the detection process in two components: *availability* of individuals to detection in the images and *perception errors* (not detecting those available individuals in the images). By fitting these double-observer count data on spatiotemporally replicated surveys, it is possible to address the two common sources of false-negative errors in aerial surveys—availability and perception – without resorting to auxiliary data such as biotelemetry marked individuals (Brack et al., 2018). Such double-observer protocol can be composed by two human observers, a human observer and an algorithm or even two different algorithms. Approaches to accommodate false positives in the modelling structure are still scarce and they are typically addressed by avoiding them in sampling design and reviewing process (see Brack et al., 2018 for further discussion).

The accuracy (bias and precision) of hierarchical models for estimating abundance depends on several factors, including population density, sample size and detectability (e.g. availability and perceptibility), so their efficient use depends on an adequate sampling design (Field et al., 2005). Furthermore, the reliability of drone surveys to assess abundance in different contexts depends on assessing how these abundance estimators perform in a wide variety of scenarios of population densities and detection probabilities of individuals. Since the total effort in fieldwork is commonly limited by a fixed budget B , the sampling design of hierarchical models must distribute this budget in a specific combination of number of sites S and visits J so that $B = S \times J$. A survey design could be planned to spend the budget by visiting more sites but fewer times; this gives more sampling units to better estimate abundance, but fewer records per site and hence less information to estimate detection probabilities. On the other hand, one could choose to characterize the detection process more thoroughly by spending the budget with more repeated visits to fewer sites, but possibly foregoing the ability to capture variations in abundance. Thus, for each scenario of population density and detectability, there is an optimal combination of survey design elements in this trade-off (i.e.

prioritize more sites and less visits, or more visits in fewer sites) that produces the most accurate estimation of abundance.

As different scenarios demand different optimal effort allocations, it is important to optimize the available budget to an efficient spending of the usually scarce resources for ecological studies (i.e. conscious sampling design; e.g. Conn et al., 2016; Knights et al., 2021). Simulation experiments are a powerful approach to address the performance of estimators in different contexts and assess optimal survey allocation to assist sampling design decisions (Kéry & Royle, 2016; Zurell et al., 2010). In short, this kind of experiment encompasses a computer-based stochastic simulation of the data collection process over a simulated population for which the true underlying parameters are known at both levels (biological and observational). Then, simulated data are fitted to the respective (or an alternative) model structure, and the capacity of the estimator to rescue true parameter values is evaluated (performance).

Performance of N-mixture models, particularly the binomial N-mixture model for single-observer counts, has been evaluated before but in limited scenarios of abundance, detection probability and number of sites and visits, or evaluating model performance in the presence of assumption violations (Duarte et al., 2018; Kéry & Royle, 2016; Veech et al., 2016; Yamaura, 2013). Here, we provide a wide and systematic scan of N-mixture models' performance and survey effort allocation (as for occupancy-detection models in Bailey et al., 2007; Mackenzie & Royle, 2005) in different scenarios of population abundance and detectability of individuals, focusing on their application for drone-based surveys. We also investigate the benefits of using a double-observer protocol to decompose the detection process in availability and perception. We conducted three experiments creating simulated count data from spatiotemporally replicated surveys and fitting N-mixture models to the data aiming to:

1. Assess the performance of N-mixture models using double- and single-observer counts under different scenarios of local abundance and detection probability and address the optimal survey effort allocation in terms of spatial versus temporal prioritization for each scenario.
2. Investigate how the use of a double-observer protocol increases model performance and affects optimal survey effort allocation.
3. Evaluate if the use of a double-observer protocol can reduce the fieldwork effort required to match the performance of the single-observer approach in the same circumstances.

Finally, we showcase our simulation-based survey design considerations by applying them to estimate the abundance of marsh deer *Blastocerus dichotomus* in the Pantanal wetland (Brazil).

2 | N-MIXTURE MODELS FOR SPATIOTEMPORALLY REPLICATED DRONE SURVEYS

To estimate abundance with N-mixture models, drone flights are conducted with repeated visits in multiple sites (i.e. spatiotemporally

replicated) and counts are carried out in the collected imagery by either one or two observers (human and/or an algorithm; Figure 1). Assuming abundance M_i at sites $i \in \{1, \dots, S\}$ is constant along visits $j \in \{1, \dots, J\}$, we can model this local abundance at sites under a Poisson distribution (or other distribution for count data) with mean (and variance) λ :

$$M_i \sim \text{Poisson}(\lambda).$$

With a single-observer reviewing imagery, counts Y_{ij} for each visit j at each site i are determined by a binomial distribution in which each individual of the local population M_i has a probability p^* of being detected:

$$Y_{ij} \sim \text{binomial}(M_i, p^*), \text{ where } p^* = \varphi \cdot p.$$

In this case, it is not possible to decompose the detection probability p^* , assumed to be the product of two processes: availability probability φ and perception probability p . This two-level model for single-observer counts is called binomial N-mixture model (Kéry et al., 2005; Royle, 2004).

When using a double-observer protocol for image review, the observation process can be segregated into two levels. Each individual of M_i has a probability φ of being available for detection in (the image correspondent to) each visit j , so de facto, of the M_i individuals, only N_{ij} will be available for detection in the imagery:

$$N_{ij} \sim \text{binomial}(M_i, \varphi).$$

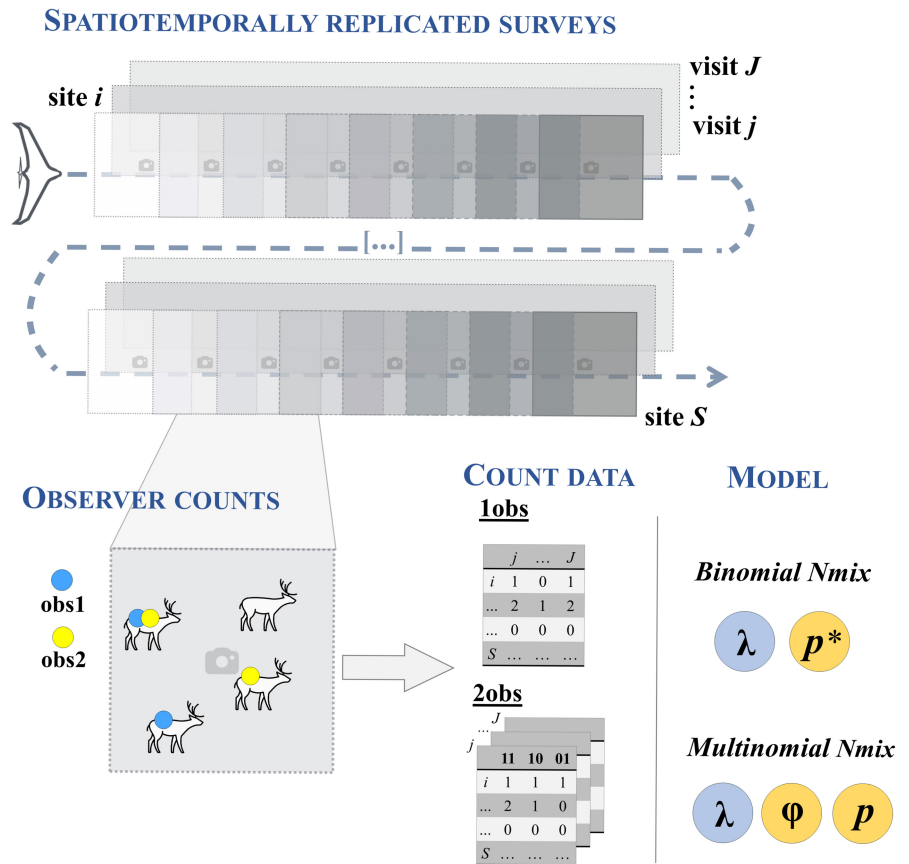
When sampled sites are truly closed throughout visits (no entries nor departures of individuals), the availability process described by φ corresponds to the probability of an individual present at the site i to not being hidden. However, not rarely when sampling wildlife species, individuals could move in and out of the surveyed sites. Then, availability would correspond to two processes: (i) the probability of the individual (that has its home range overlapping the sampled site) being present at site i in visit j and (ii) not being hidden (Brack et al., 2018; Chandler et al., 2011; Kéry & Royle, 2016).

Finally, each individual from the available pool N_{ij} , has a probability p of being detected by each observer (independently), resulting in four possible *encounter histories* (as in a capture-recapture procedure): $k_1 = '11'$ detected by both observers; $k_2 = '10'$ only detected by the first observer; $k_3 = '01'$ only detected by the second observer and $k_4 = '00'$ not detected by any observer. Thus, the resulting count data Y_{ijk} for each visit j , at each site i , and under each observable encounter histories $k \in \{1, 2, 3\}$ is modeled as a function of multinomial conditional cell probabilities π_k :

$$Y_{ijk} \sim \text{multinomial}(N_{ij}, \pi(p)_k).$$

in which, the probability of each observable encounter history is determined as $k_{11} = p^2$; $k_{10} = p(1-p)$ and $k_{01} = (1-p)p$. This three-level model for independent double-observer counts is known as multinomial N-mixture model with a temporary emigration component,

FIGURE 1 Sampling design, count data and N-mixture models for spatiotemporally replicated drone-based surveys with a single or multiple observers reviewing images. Single-observer counts are fitted with binomial N-mixture models and multiple observer counts under a Multinomial N-mixture model with a temporary emigration component. Parameters: λ = expected local abundance; p^* = overall detection probability; φ = availability probability and p = perception probability. Note that observers can be either a human or an algorithm. Figure adapted from Brack et al. (2021).



hereafter “multinomial” (Chandler et al., 2011). The double-observer protocol can be applied in only a proportion (i.e. subset) of the imagery (flights), still being possible to segregate the two observation levels by modeling the data from the mixed double- and single-observer protocols.

For each of the presented levels, heterogeneity in the basic parameters (λ , φ , p , p^*) can be modeled as a function of explanatory variables through linear regression and an appropriate link function (log for λ and logit for probabilities). Count data, with either single- or double-observer (or mixed) protocols, can be obtained manually by human observers or by an algorithm trained to detect the target species in the imagery.

If the spatial temporary emigration process is expected to be significant, the local abundance parameter should be interpreted with caution. A more adequate interpretation for λ would be the expected number of individuals of the population that uses a given site (i.e. intensity of use). To reduce the effects of spatial temporary emigration, one could plan sampling design to (i) shorten the time interval between repeated visits (but avoiding temporal autocorrelation) and (ii) define site size as relatively large in comparison with the expected area used by the individuals during sampling. Christensen et al. (2021) proposed a post hoc sensitivity analysis to define the best length to split flight lines into sites and found that the ideal size is similar to the home range size reported to the target species. Auxiliary data from telemetry or some marked individuals can provide information to segregate the two components of temporary emigration (spatial and temporal; Brack et al., 2018).

In these formulations of N-mixture models, count data are assumed to not have false-positive errors. This might be particularly important when using counts from an algorithm, that can be then post-checked by an observer to avoid false positives. Although this is an area of research that needs further development, there are available approaches for some cases. Misidentification between similar species can be addressed in N-mixture models by modeling the uncertain detections (Chambert et al., 2016). If a fully automated approach is used, the multiple-observer approach, with a human observer reviewing only a subset of the imagery, could be used to estimate both perception and misidentification errors of the algorithm (Conn et al., 2013, 2014). Double counts commonly can be avoided with decisions in sampling design or review process (e.g. separate flight strips, count in orthomosaic; Brack et al., 2018). Accommodating false-positive errors resulting from the movement of individuals between lines could be modeled, for example, from telemetry data (Terletzky & Koons, 2016).

3 | SIMULATIONS GENERAL DESIGN

To assess the performance and optimal design of N-mixture models for drone-based surveys, we simulated count data under single- and double-observer review protocols and analyzed them with the correspondent N-mixture model structure (binomial or multinomial). In software R (R Core Team, 2020), we simulated the local abundance as Poisson distributed, a binomial outcome for available individuals

(in the multinomial model) and the single-observer counts under overall detection (in the binomial model), and a multinomial outcome derived from the perception probability by the observers for the double-observer counts. For each simulation, we defined the number of sites, visits and observers and the true local abundance, availability and perception probabilities. We considered all model parameters constant (i.e. no covariates), a common practice in survey design studies that aim to find general design principles (e.g. Guillera-Arroita et al., 2010; Mackenzie & Royle, 2005; Yamaura, 2013). We fitted simulated data using maximum likelihood estimation with the R package UNMARKED (Fiske & Chandler, 2011), using the *pcount* function for the single-observer data and *gmultmix* for the mixed- and double-observer protocols.

All simulation results and code to reproduce the simulations and results of this paper are available at <https://github.com/ismaelvbrack/designNmix4droneSurveys>. We provide R code to simulate and analyze data under both modeling structures (single binomial and double multinomial N-mixture models, including the mixed-protocol model) using maximum likelihood estimation, as well as BUGS code to conduct Bayesian analysis for the same models using JAGS software (Plummer, 2003) from R (JAGSUI package; Kellner, 2015). As supplementary material, we provide an interactive “ready-to-consult” website (<https://ismaelvbrack.github.io/designNmix4droneSurveys>) to assist future studies with planning drone-based surveys according to specific scenarios of expected local abundance and detectability. There, it is possible to explore all results of this study on the performance of N-mixture models with single and double observers and the optimal survey effort allocation in a user-friendly interface, besides simulating examples of these models.

For each simulation study, we defined different scenarios considering local abundance, availability and perception probability. For each of these, we simulated data collection and analysis for different combinations of number of sites, visits and observers, and calculated the estimator accuracy for the expected local abundance parameter based on 2000 iterations of the simulation using the root mean squared error relative to the true parameter value ($\text{rel.RMSE} = \frac{1}{\lambda} \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\lambda}_i - \lambda)^2}$; in which n is the number of iterations, $\hat{\lambda}$ is the estimated expected local abundance and λ is the true known expected local abundance). The lowest rel.RMSE value for each scenario of $\{\lambda, \varphi\}$ was used to define the optimal survey design $J_{\text{opt}} = B/S_{\text{opt}}$. From the 2000 iterations of each scenario, we excluded those with no convergence or with infinite abundance estimates (upper 95% confidence intervals >200 individuals, associated with detection probabilities near zero). We are using here the rel.RMSE as a measure of accuracy in the sense that it represents bias and precision (Hone, 2008). For the purpose of using N-mixture models for drone-based counts, we fixed the perception probability p at 0.8, considering what we have seen as a moderately low perception threshold in studies with drone surveys with both human observers (e.g. Brack et al., 2021; Patterson et al., 2016; Preston et al., 2021; Vermeulen et al., 2013) and algorithms (e.g. Dujon et al., 2021; Eikelboom et al., 2019; Kellenberger et al., 2018). However, for other applications of N-mixture models with availability and perception

observation processes (e.g. auditory bird point counts, terrestrial strip transects of burrowing species, or boat-based counts of marine mammals), scenarios with lower perception probabilities should be explored (with just slight modifications in the code provided). We also note that, if the study focus is not on the local abundance level but on other parameters (e.g. availability or detection), defining optimality with respect to these parameters is likely to lead to different optimal survey designs (Guillera-Arroita et al., 2010).

4 | SIMULATION STUDY 1: OPTIMAL DESIGN OF COUNT SURVEYS FOR N-MIXTURE ABUNDANCE ESTIMATION

We assessed the performance and optimal effort allocation of N-mixture models for drone surveys in 73 scenarios defined from the combination of 11 mean local abundances in sites λ {0.1; 0.2; 0.3; 0.5; 1; 2; 4; 8; 12; 20; 40} and eight availability probabilities φ {0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8}. For the three highest values of local abundance, we only simulated scenarios with availabilities of {0.1; 0.2; 0.5} to check general trends for higher local abundances while mitigating the time consumed in simulation tasks. Perception probability p was fixed at 0.8. We defined a fixed total effort (budget) of $B = 2000$ flights (i.e. 2000 site visits); we purposely chose a high number here to represent a case where data are not a limiting factor for analysis (i.e. large sample dataset) and hence ensure that large-sample approximations from maximum likelihood estimation hold in scenarios of low local abundance and/or detectability. For each scenario, we tested different combinations of this total effort B on the distribution of sites S and visits J so that $B = S \cdot J$ (e.g., 1000 sites with 2 visits, 250 sites with 8 visits, 100 sites with 20 visits and so on). We ranged the number of visits from $J = 2$ until either $J = 10$ visits or until being sure that the rel.RMSE was rising up (i.e. the estimator accuracy was decreasing).

We then assessed the accuracy of the estimator of local abundance (based on the rel.RMSE from the 2000 iterations) for each number of visits in each scenario of $\{\lambda, \varphi\}$. We found the optimal survey design for each scenario based on the lowest rel.RMSE value. Because of the difference in rel.RMSE between numbers of visits was very low for many scenarios, we also obtained the range of designs (S^*J) for which the rel.RMSE was lower than 0.5% of the rel.RMSE under the optimal design and considered the accuracy inside this range as equivalent. This means that a range of designs results in very similar performance and thus other criteria (e.g. travel times between sites) could be used to choose among those possibilities when designing a study.

We repeated this procedure on the 73 scenarios considering the single-observer counts (binomial N-mixture model) and the (full) double-observer protocol (multinomial N-mixture model). Finally, we checked whether the pattern we found for optimal designs across scenarios was consistent for a different total survey effort using $B = 4000$ flights, that is, if the pattern for J_{opt} is independent of the total effort, as it is observed for the related occupancy-detection models (Mackenzie & Royle, 2005). For this last analysis, we only

considered six scenarios using expected local abundances {0.2; 1} and availability probabilities {0.2; 0.4; 0.6}.

4.1 | Results

Both model structures—single-observer binomial and double-observer multinomial N-mixture models—presented unbiased estimations of local abundance for all evaluated scenarios (Figure S6). The rel.RMSE of N-mixture models under the optimal number of visits in each scenario of expected local abundance λ and availability φ ranged from 0.024 to 0.426 for the single-observer model and from 0.017 to 0.378 for the double-observer model. That is, under optimal survey effort allocation, the mean error of the estimator relative to the true expected abundance ranged between ~2% and ~40%, depending on population density (local abundance in sites) and availability of individuals. Estimators reached much larger rel. RMSE values (near one) for designs with a small number of visits in scenarios of low expected local abundance ≤ 0.5 (Figures S1 and S2). Only in one scenario ($\lambda = 0.1$; $\varphi = 0.1$), the rel.RMSE was higher than 30% (for both modeling approaches).

The loss of performance from the single- to double-observer approaches (i.e. difference in rel.RMSE) was 0.05 lower for the $\{\lambda = 0.1$; $\varphi = 0.1\}$ scenario, and this difference reduced with the increase in local abundance and/or availability (Figures S1 and S2). The accuracy of both model structures increased with local abundance λ and especially with the availability of individuals φ .

The optimal number of visits varied from $J = 28$ to $J = 2$ for the single-observer model and from $J = 22$ to $J = 2$ for the double-observer model depending on the scenario of local abundance and availability (Table 1). Generally, J_{opt} decreased with increasing availability φ . J_{opt} was lower for the double-observer model than for the single-observer one. The range of possible optimal number of visits J_{opt} (i.e. considered equivalent with a difference in rel. RMSE $< 0.5\%$) increased with expected local abundance λ (Table 1).

In the six scenarios for which we evaluated a different total effort $B = 4000$ ($\lambda = \{0.2; 1\}$; $\varphi = \{0.2; 0.4; 0.6\}$), we found a very similar J_{opt} (except for a few stochastic differences because of the number of iterations). We interpret this as supporting the idea that the optimal number of visits is independent of the total effort in large sample sizes for both binomial and multinomial N-mixture models (Figures S3-S5).

5 | SIMULATION STUDY 2: EXPLORING THE BENEFIT OF THE DOUBLE-OBSERVER PROTOCOL

To assess the improvement in accuracy with the use of a double-observer protocol, we defined six scenarios by combining expected local abundances λ {0.2; 1} and availability probabilities φ {0.2; 0.4; 0.6}, with fixed $p = 0.8$. For each scenario, we tested different combinations of number of sites S and visits J using a fixed total effort of

$B = 2000$, for different proportions of the flights checked by two observers (double-observer protocol): {0%; 20%; 40%; 60%; 80%; 100%}. The 0% corresponds to the single-observer binomial model, the 100% corresponds to the full double-observer multinomial model (both considered in simulation experiment 1), and the rest are mixed protocols with only a percentage of flights reviewed by a second observer.

The simulation procedure was as in simulation study 1 (rel.RMSE from 2000 iterations under each number of visits; total survey budget $B = 2000$ flights) for each proportion of the double-observer protocol. For each double-observer proportion, we found the optimal design J_{opt} for each scenario of $\{\lambda, \varphi\}$ with the lowest rel. RMSE. Because of the stochasticity from the 2000 iterations, we estimated a linear trend of the rel.RMSE in relation to the proportion of double-observer review and calculated the linear decrease in rel. RMSE relative to the single-observer model (0%).

5.1 | Results

The accuracy of the abundance estimator increased (decreasing rel. RMSE) with the proportion of images checked by two observers (Figure 2). The difference in accuracy reached almost 15% for the full double-observer model in the $\{\lambda = 1$; $\varphi = 0.6\}$ scenario (compared with the baseline performance of a single observer). The optimal allocation of effort showed a tendency to decrease the number of repeat visits (lower J), thus allowing visiting more sites (higher S), as the proportion of double-observer protocol increases (Figure S4).

6 | SIMULATION STUDY 3: REDUCING FIELDWORK EFFORT BY EMPLOYING A DOUBLE-OBSERVER PROTOCOL

We have seen how the double-observer protocol leads to increased accuracy in the estimation of local abundance. Alternatively, double observers could be used to reduce the amount of fieldwork effort needed to match the same accuracy that is obtained from an optimal design of the single-observer model. We explored this strategy using the same six scenarios of simulation study 2 ($\lambda = \{0.2; 1\}$; $\varphi = \{0.2; 0.4; 0.6\}$; $p = 0.8$) by calculating the proportional reduction of total fieldwork effort ($B = 2000$ flights) that can be achieved with a double-observer protocol while still matching the accuracy (lowest rel.RMSE) of the single-observer model under optimal design J_{opt} .

To do this, we started with the full budget $B = 2000$ for the double-observer model and reduced B (under its optimal number of visits, thus reducing the number of sites) until we reached an accuracy (rel.RMSE) similar to the target single-observer rel.RMSE. Because of the long computational time, we conducted this search in only six incremental steps. In the first step, we reduced the total effort to 50% ($B = 1000$ flights). Then for each subsequent search step, the difference in B was reduced by 50% (“half-way” distance) between the last step and either the full reference effort ($B = 2000$) or the effort in the previous step, depending on if the rel.RMSE was

TABLE 1 Optimal number of visits (J_{opt}) in spatiotemporally replicated drone-based surveys for abundance modeling with N-mixture models, under different scenarios of expected local abundance (λ) and availability probability (ϕ). (a) Binomial N-mixture model for single-observer counts. (b) Multinomial N-mixture model for double-observer counts. The optimal J is obtained from the lowest relative root mean square error (rel.RMSE) calculated from 2000 iterations for each combination of the number of sites S and visits J . In brackets, the range of visits J for which the performance can be considered equivalent (i.e. rel.RMSE < 0.5%). Shading of the table cells (from light gray to black) indicates an increasing J_{opt}

λ											
0.1		0.2	0.3	0.5	1	2	4	8	12	20	40
(a) Single-observer (Binomial) N-mixture model											
ϕ	0.1	20 (20-20)	23 (19-26)	21 (21-25)	27 (21-28)	19 (19-19)	20 (11-24)	15 (15-15)	11 (11-17)	17 (12-19)	21 (11-23)
	0.2	13 (8-13)	14 (8-14)	12 (10-13)	12 (9-18)	14 (9-19)	10 (9-19)	9 (7-12)	10 (7-15)	13 (7-15)	11 (7-13)
	0.3	7 (5-8)	7 (6-8)	5 (5-9)	7 (5-10)	13 (6-15)	10 (7-16)	8 (7-12)	-	-	-
	0.4	4 (4-6)	5 (4-6)	4 (4-8)	5 (4-8)	6 (5-10)	9 (5-10)	9 (4-10)	-	-	-
	0.5	4 (3-4)	4 (3-4)	4 (3-5)	4 (3-7)	6 (4-8)	6 (4-10)	8 (4-10)	5 (3-15)	11 (4-15)	10 (4-15)
	0.6	3 (3-3)	3 (3-3)	4 (3-5)	3 (3-5)	4 (3-7)	4 (3-10)	6 (3-10)	-	-	-
	0.7	2 (2-3)	2 (2-3)	3 (2-4)	3 (2-4)	3 (2-6)	4 (3-9)	5 (3-10)	-	-	-
	0.8	2 (2-2)	2 (2-2)	2 (2-3)	2 (2-3)	3 (2-5)	3 (2-7)	4 (2-9)	-	-	-
(b) Double-observer (Multinomial) N-mixture model											
ϕ	0.1	22 (15-22)	19 (16-19)	18 (18-23)	18 (16-27)	18 (16-26)	15 (9-22)	16 (14-17)	15 (11-18)	11 (11-11)	13 (11-19)
	0.2	7 (7-10)	9 (7-11)	8 (7-11)	10 (7-14)	8 (7-17)	12 (7-18)	8 (7-13)	11 (6-13)	12 (7-13)	11 (5-15)
	0.3	5 (4-6)	6 (4-7)	6 (5-6)	6 (5-9)	10 (5-12)	9 (6-16)	8 (5-20)	-	-	-
	0.4	4 (3-5)	4 (3-5)	4 (3-5)	4 (3-7)	6 (4-10)	6 (4-10)	7 (5-10)	-	-	-
	0.5	3 (3-4)	3 (3-4)	3 (3-4)	3 (3-5)	4 (3-7)	4 (3-10)	7 (3-10)	9 (3-15)	9 (3-15)	14 (3-15)
	0.6	2 (2-3)	2 (2-3)	3 (2-3)	3 (2-4)	3 (2-4)	3 (2-7)	5 (3-10)	-	-	-
	0.7	2 (2-2)	2 (2-2)	2 (2-3)	2 (2-3)	3 (2-4)	3 (2-6)	4 (2-9)	-	-	-
	0.8	2 (2-2)	2 (2-2)	2 (2-2)	2 (2-3)	2 (2-3)	2 (2-4)	3 (2-7)	-	-	-

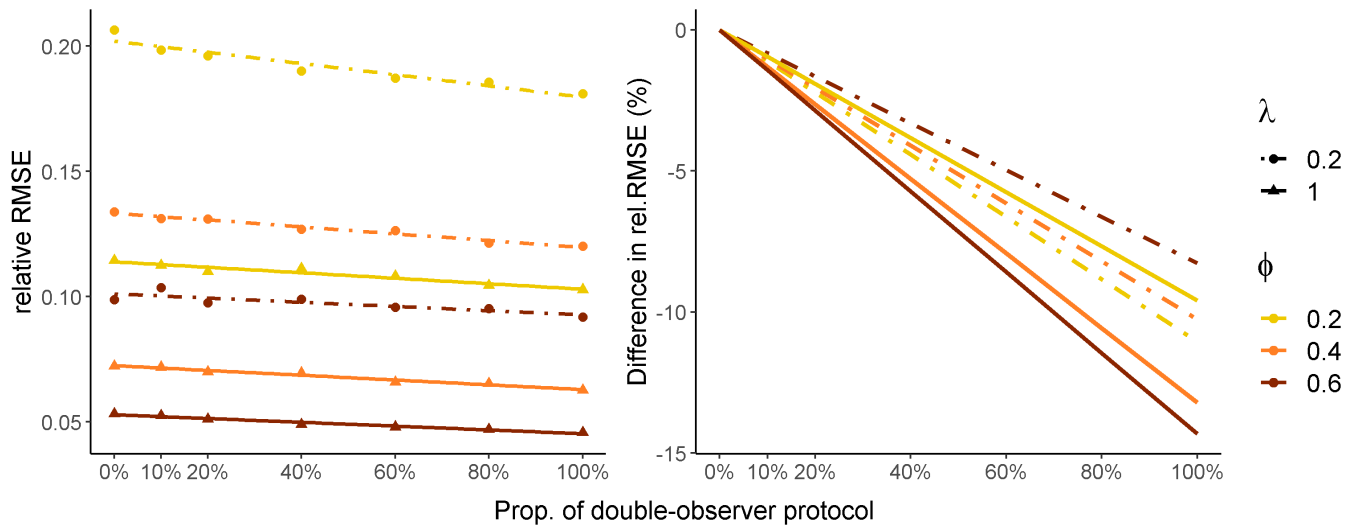


FIGURE 2 Accuracy (relative RMSE) and increase in accuracy (difference in rel.RMSE) in abundance estimation with N-mixture models in relation to the proportion of double-observer protocol in image reviewing, and under different scenarios of expected local abundance in sites (λ) and individual availability probability (ϕ). 0% = single-observer binomial N-mixture model; 100% = full double-observer multinomial N-mixture model; the others are mixed protocols of multinomial N-mixture models.

TABLE 2 Proportion of the total effort (budget) of the double-observer model (multinomial N-mixture) needed to obtain similar performance to the single-observer model (binomial N-mixture) for different scenarios of expected local abundance λ and availability of individuals ϕ

		λ	
		0.2	1
ϕ	0.2	81%–88%	75%–83%
	0.4	75%–81%	75%–86%
	0.6	88%–91%	63%–72%

over or under the target rel.RMSE, respectively. Because of small differences in rel.RMSE values and stochasticity from 2000 iterations, we provide an interval of survey effort B that is close to the target single-observer rel.RMSE.

6.1 | Results

If a full double-observer protocol is used in image reviewing, the total fieldwork budget may be reduced from 9% to 37% (depending on the scenario of $\{\lambda, \phi\}$), while still matching the best accuracy obtained with a single-observer protocol (Table 2, Figure S5).

7 | CASE EXAMPLE: MARSH DEER DRONE-BASED SURVEYS IN PANTANAL WETLAND

The marsh deer *Blastocerus dichotomus*, the largest cervid in South America (up to 150kg), is a habitat-specific species associated with

wetlands (Piovezan et al., 2010) and threatened with extinction (IUCN; Duarte et al., 2016). Because of the inaccessibility of its habitat, marsh deer population estimates are usually obtained from aerial surveys (e.g. Andriolo et al., 2005; Mourão et al., 2000; Ríos-Uzeda & Mourão, 2012). In 2017, Brack et al. (2021) conducted spatiotemporally replicated drone-based count surveys to estimate the abundance of marsh deer in the Pantanal wetland (Sesc Pantanal Private Natural Reserve; 108,000ha) and explored the use of this method to monitor that species. Six flight paths (32–42 km) were flown from two to six times each using a fixed-wing drone equipped with an RGB camera. The six flight paths were split into 203 1-km sites, and two observers carried out deer counts in a manual review of the collected imagery. The first observer reviewed the entire image set (~25,000 images) and the second observer reviewed only 20% of the flights. Count data were fitted using the three-level multinomial N-mixture model for the mixed single- and double-observer protocol. The estimated marsh deer mean local abundance λ was 0.33 (95% CI = 0.23–0.48), availability probability ϕ was 0.14 (0.10–0.19) and perception probability by observers p was 0.93 (0.82–0.97).

Aiming at improving accuracy in marsh deer abundance estimation and optimally plan survey design for upcoming assessments, we conducted a simulation study based on the 2017 population survey. We considered the point estimates of the 2017's population assessment ($\lambda = 0.33$; $\phi = 0.14$ and $p = 0.93$), a total effort of $B = 813$ (from $S = 203$ and $J = 4$), and different proportions of the images checked by two observers (double-observer protocol): {0%; 20%; 50%; 100%}. We explored the number of visits between $J = 4$ and $J = 26$ and, for each number of visits and each proportion of double observers, we ran 2000 iterations from which we calculated the relative RMSE.

7.1 | Results

The optimal number of visits J_{opt} for the defined budget ($B = 813$) varied from $J = 14$ to $J = 24$ depending on the proportion of flights reviewed by double observers. In comparison with the survey design used in the 2017's marsh deer survey ($J = 4$), the rel.RMSE reduced around two-thirds (62%–64%) under the J_{opt} . Under the optimal design, there was a reduction in rel.RMSE of about 11% between the single-observer model and the full double observer, even with the high perception probability used ($p = 0.93$). For the proportion of 20% of double-observer protocol, as used in the 2017's assessment, the increase in accuracy of local abundance estimation (in comparison with the single-observer model) is generally very small.

8 | DISCUSSION

We presented here a comprehensive assessment of N-mixture models' performance and optimal effort allocation to assist sampling designs for estimating abundance with drone-based surveys. We found that spatiotemporally replicated drone-based counts analyzed with N-mixture models can be applied in a wide variety of contexts of population densities and availability of individuals. Interestingly, these results suggest that drone-based wildlife surveys can be suitable to estimate abundance even when the availability of individuals is low (e.g. forested areas, burrowing species and marine animals that rarely remain at the water surface) and for low-density species (which are often threatened). Nonetheless, as already shown in previous simulation studies with binomial N-mixture models (Joseph et al., 2009; Yamaura, 2013), extra care should be taken in scenarios of very low local abundance and availability (here $\lambda = 0.1$; $\varphi = 0.1$, where abundance estimation performance was unstable). The results of this study are directly applicable for large sample sizes; for small sample sizes, the performance of N-mixture models is expected to be poorer (including biased abundance estimation), particularly in scenarios of low detection probability (Kéry, 2018; Yamaura, 2013).

We found that consciously planning drone-based surveys by optimally allocating survey efforts can have a great impact on the performance of N-mixture models for abundance estimation. Our results indicate that survey effort should prioritize repeating more visits in the same sites when availability probability is expected to be low, while more sites with fewer visits should be selected to optimally allocate effort in higher availabilities. For example, in the marsh deer study case shown, we found that an optimal sampling design can increase the accuracy in abundance estimation by up to two-thirds compared with a non-optimal design. To optimize sampling designs, pilot studies or previous knowledge on the species are fundamental to provide guesses about the parameters for the simulations. From the comparison between the two N-mixture model structures (single- and double-observer counts), the single-observer model generally demands allocating effort in more temporal replicates than the double-observer model, for the same scenario. Lastly, as abundance estimation accuracy notably increases with the

availability of individuals, logistic adjustments in surveys to sample in moments that individuals are more available (e.g. time of the day, season) have a great potential to improve abundance estimation performance.

The optimal design becomes less sensitive to the choice of number of visits and sites when local abundance is higher (rel.RMSE curves become “flatter”). Therefore, there is a wider range of efficient designs with similar accuracy in the estimation of local abundance (i.e. close performance to the optimal design). This flexibility allows for other design considerations when planning surveys, for example, choosing to sample more sites if there is interest in the spatial variation of abundance (relationships with covariates) or more visits if travel cost to new sites is expensive. It is important to note that, in “real-case” studies such as the marsh deer example, when modeling the spatial variation in local abundance using covariates (e.g. Brack et al., 2021), the optimal number of visits would be pushed towards lower numbers of visits, to cover more sites.

There is a fundamental difference between fieldwork sampling design (number of visits and sites), which must be planned before data collection, and the reviewing procedure protocol (to use one or two observers to check images). The decision to use a double-observer protocol can be made after fieldwork and might improve the estimation accuracy. Notably, the accuracy in abundance estimation can be considerably improved even at very high perception probabilities by the observers, as illustrated in the marsh deer case study (11% increase at $p = 0.93$). Moreover, the improvement in the estimator performance can be achieved with the second observer only reviewing a proportion of the images. For the perception probability considered in the simulations ($p = 0.8$, representative of human observer counts from drone imagery), this benefit is more evident when the second observer reviews more than 20% of the image set. This result is particularly interesting when an automated process is used in image review: if the computer algorithm counts are considered the first observer and a human the second observer, reviewing only a subset of the drone imagery already provides gains in accuracy. In this case, the perception probability of both the algorithm and the human observer should be estimated separately.

We showed here that it is possible to reduce considerably the effort in fieldwork by doubling the effort in image reviewing in lab (i.e. double instead of single-observer counts) while achieving a similar accuracy in abundance estimation. This consideration is particularly interesting when budget or conditions (e.g. weather, visibility or loss of equipment) are limited for fieldwork. However, depending on the time needed to manually review images, a double-observer protocol can be prohibitive. Here, we have limited the budget for optimal allocation only considering fieldwork effort, under the assumption that lab effort is more flexible and commonly less costly. If image review is also costly and can be budgeted in comparison with fieldwork expenses, the simulations of this study could be expanded to consider this extra aspect of survey design (e.g. proportion of images with a double-observer protocol). This would require including costs for the different units of effort (fieldwork and image review) explicitly in the evaluations of optimal effort allocation. Nonetheless, with

the increasing development of machine learning algorithms (Christin et al., 2019; Dujon & Schofield, 2019), trained specifically for each context and potentially replacing manual reviewing, the costs of image reviewing are expected to drastically reduce in the future.

As a cautionary note on the use of N-mixture models, there have been concerns on the sensitivity of these models to assumption violations and the claim that repeated counts have little information about the detection process for parameter identifiability (Barker et al., 2018). Evaluations of assumption violations in binomial N-mixture models have shown that unmodeled heterogeneity in detection probability (especially directional/non-random) can bias the estimation of abundance (Duarte et al., 2018; Knape et al., 2018; Link et al., 2018). However, despite these raised concerns, studies comparing N-mixture models with more traditional and well-established approaches (i.e. capture-recapture models or distance sampling) have generally shown similar results (Christensen et al., 2021; Ficetola et al., 2018; Keever et al., 2017; Kéry, 2018). In the context of drone-based surveys, the impact of these findings would be expected to be more concerning for the availability process (non-modelled heterogeneity in unavailability). Heterogeneity in perception can be addressed by including individual covariates for each record and model this observation process using a data augmentation approach (Royle et al., 2007). To improve robustness against unaccounted heterogeneity, additional information of some marked individuals may be collected (see Dunstan et al., 2020 for an example of drone surveys of marked individuals). In the context of drone-based surveys, Corcoran et al. (2020) found that N-mixture models can overestimate abundance in comparison with an adapted Horvitz-Thompson estimator. However, we argue their comparison between methods is inadequate because it ignored basic recommendations for study design and analysis of N-mixture models: (i) very few sites sampled; (ii) large time interval between visits, which can prevent interpretations about temporary emigration processes and thus local abundance (Chandler et al., 2011; Williams et al., 2017); (iii) use of a negative binomial distribution for local abundance, prone to provide unrealistic high abundance estimates (Joseph et al., 2009; Kéry, 2018; Kéry et al., 2005; Knape et al., 2018); (iv) did not provide any uncertainty for the N-mixture model estimates, precluding a proper comparison and (v) did not account for false-positive errors, which were known to occur in their system.

The sampling design considerations provided by our study are directly applicable to other contexts in which N-mixture models can be used with a double-observer protocol to segregate the availability process from the ability to detect individuals by the observers, such as birds' auditory point transect surveys (e.g. Amundson et al., 2014) or terrestrial strip transects of elusive species (e.g. burrowing animals, Zylstra et al., 2009). Differences in performance between the two modeling approaches used in this study are expected to be more pronounced the lower the perception probability, while differences tend to zero as perception gets close to one. Furthermore, the basic structure of the presented N-mixture models (closed population, single-season and single-species) could be expanded to accommodate other sources of variation (e.g. overdispersion in detection probability, Knape et al., 2018; group detection, Martin et al., 2011;

spatial autocorrelation, Guélat et al., 2018), include multiple species (multi-species random effects, Dorazio et al., 2015; Sollmann et al., 2016), or incorporate dynamics in abundance (trend models, Kéry et al., 2009; models with explicit dynamics, Bellier et al., 2016; Dail & Madsen, 2011; Zipkin et al., 2014).

The use of drones to survey wildlife is quickly spreading to many species in different habitats. Nevertheless, few studies go further than simple tests of detecting species in drone-based images and apply drone surveys to model the abundance of wildlife populations (Brack et al., 2018; Linchant et al., 2015). It is imperative to plan sampling designs considering sources of detection errors and, for this, spatiotemporally replicated flights and N-mixture models have proven to be a useful straightforward approach. Survey design efficiency to make the best use of available resources is a key consideration in wildlife monitoring and conservation, particularly given that these resources are usually limited. The improvements achieved by an optimal design, rearranging survey efforts among sites and visits to obtain the most accurate abundance estimation can be crucial, for example, to detect trends when monitoring a population or to categorize a species as threatened or not.

AUTHOR CONTRIBUTIONS

Ismael V. Brack and José J. Lahoz-Monfort conceived the ideas and designed the experiments. Ismael V. Brack conducted the simulation experiments and analyzed the data. All authors discussed results and applications. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

IB was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes; Finance Code 001).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14054>.

DATA AVAILABILITY STATEMENT

All simulation results, data and code to reproduce the simulations and results of this paper are available at <https://doi.org/10.5281/zenodo.7348035> (Brack, 2022). We also provide a website (<https://ismaelvbrack.github.io/designNmix4droneSurveys>) to simulate scenarios and see the results of this paper.

ORCID

Ismael V. Brack  <https://orcid.org/0000-0003-2988-9811>

REFERENCES

Amundson, C. L., Royle, J. A., & Handel, C. M. (2014). A hierarchical model combining distance sampling and time removal to estimate

- detection probability during avian point counts. *The Auk*, 131(4), 476–494. <https://doi.org/10.1642/AUK-14-11.1>
- Andriolo, A., Piovezan, U., da Costa, M. J. R. P., Laake, J., & Duarte, J. M. B. (2005). Aerial line transect survey to estimate abundance of marsh deer (*Blastocerus dichotomus*) (Illiger, 1815). *Brazilian Archives of Biology and Technology*, 48(5), 807–814. <https://doi.org/10.1590/S1516-89132005000600017>
- Bailey, L. L., Hines, J. E., Nichols, J. D., & MacKenzie, D. I. (2007). Sampling design trade-offs in occupancy studies with imperfect detection: Examples and software. *Ecological Applications*, 17(1), 281–290. [https://doi.org/10.1890/1051-0761\(2007\)017\[0281:SDTIO S\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2007)017[0281:SDTIO S]2.0.CO;2)
- Barasona, J. Á., Mulero-Pázmány, M., Acevedo, P., Negro, J. J., Torres, M. J., Gortázar, C., & Vicente, J. (2014). Unmanned Aircraft Systems for studying spatial abundance of ungulates: Relevance to spatial epidemiology. *PLoS ONE*, 9(12), e115608. <https://doi.org/10.1371/journal.pone.0115608>
- Barker, R. J., Schofield, M. R., Link, W. A., & Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics*, 74(1), 369–377. <https://doi.org/10.1111/biom.12734>
- Bellier, E., Kéry, M., & Schaub, M. (2016). Simulation-based assessment of dynamic N-mixture models with density-dependence and environmental stochasticity in vital rates. *Methods in Ecology and Evolution*, 7(9), 1029–1040. <https://doi.org/10.1111/2041-210X.12572>
- Brack, I. V. (2022). Designing drone surveys for N-mixture models (v1.0) [source code]. <https://doi.org/10.5281/zenodo.7348035>
- Brack, I. V., Kindel, A., Berto, D. O., Cordeiro, J. L. P., Coelho, I. P., Lahoz-Monfort, J., & Oliveira, L. F. B. (2021). Drone surveys revealed bottom-up, and not top-down, effects on the marsh deer local abundance. <https://doi.org/10.21203/RS.3.RS-884511/V1>
- Brack, I. V., Kindel, A., & Oliveira, L. F. B. (2018). Detection errors in wildlife abundance estimates from Unmanned Aerial Systems (UAS) surveys: Synthesis, solutions, and challenges. *Methods in Ecology and Evolution*, 9(8), 1864–1873. <https://doi.org/10.1111/2041-210X.13026>
- Bushaw, J. D., & Ringelman, K. M. (2019). Applications of unmanned aerial vehicles to survey mesocarnivores. (pp. 1–9). <https://doi.org/10.3390/drones3010028>
- Chambert, T., Hossack, B. R., Fishback, L., & Davenport, J. M. (2016). Estimating abundance in the presence of species uncertainty. *Methods in Ecology and Evolution*, 7(9), 1041–1049. <https://doi.org/10.1111/2041-210X.12570>
- Chandler, R. B., Royle, J. A., & King, D. I. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology*, 92(7), 1429–1435. <https://doi.org/10.1890/10-2433.1>
- Christensen, S. A., Farr, M. T., & Williams, D. M. (2021). Assessment and novel application of N-mixture models for aerial surveys of wildlife. *Ecosphere*, 12(8), e03725. <https://doi.org/10.1002/ecs2.3725>
- Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- Conn, P. B., McClintock, B. T., Cameron, M. F., Johnson, D. S., Moreland, E. E., & Boveng, P. L. (2013). Accommodating species identification errors in transect surveys. *Ecology*, 94(11), 2607–2618. <https://doi.org/10.1890/12-2124.1>
- Conn, P. B., Moreland, E. E., Regehr, E. V., Richmond, E. L., Cameron, M. F., & Boveng, P. L. (2016). Using simulation to evaluate wildlife survey designs: Polar bears and seals in the Chukchi Sea. *Royal Society Open Science*, 3(1), 150561. <https://doi.org/10.1098/rsos.150561>
- Conn, P. B., Ver Hoef, J. M., McClintock, B. T., Moreland, E. E., London, J. M., Cameron, M. F., ... Boveng, P. L. (2014). Estimating multispecies abundance using automated detection systems: Ice-associated seals in the Bering Sea. *Methods in Ecology and Evolution*, 5(12), 1280–1293. <https://doi.org/10.1111/2041-210X.12127>
- Corcoran, E., Denman, S., & Hamilton, G. (2019). New technologies in the mix: Assessing N-mixture models for abundance estimation using automated detection data from drone surveys. *Ecology and Evolution*, 2020, 1–10. <https://doi.org/10.1002/ece3.6522>
- Dail, D., & Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, 67(2), 577–587. <https://doi.org/10.1111/j.1541-0420.2010.01465.x>
- de Melo, F. R. (2021). Drones for conservation: New techniques to monitor muriquis. *Oryx*, 55(2), 171. <https://doi.org/10.1017/S0030605321000028>
- Dorazio, R. M., Connor, E. F., & Askins, R. A. (2015). Estimating the Effects of Habitat and Biological Interactions in an Avian Community. *PLoS ONE*, 10(8), e0135987. <https://doi.org/10.1371/journal.pone.0135987>
- Duarte, A., Adams, M. J., & Peterson, J. T. (2018). Fitting N-mixture models to count data with unmodeled heterogeneity: Bias, diagnostics, and alternative approaches. *Ecological Modelling*, 374(January), 51–59. <https://doi.org/10.1016/j.ecolmodel.2018.02.007>
- Duarte, J. M. B., Varela, D., Piovezan, U., Beccaceci, M. D., & Garcia, J. E. (2016). *Blastocerus dichotomus*. The IUCN Red List of threatened species 2016: e.T2828A22160916. <https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T2828A22160916.en>
- Dujon, A. M., Ierodiaconou, D., Geeson, J. J., Arnould, J. P. Y., Allan, B. M., Katselidis, K. A., & Schofield, G. (2021). Machine learning to detect marine animals in UAV imagery: Effect of morphology, spacing, behaviour and habitat. *Remote Sensing in Ecology and Conservation*, 7(3), 341–354. <https://doi.org/10.1002/rse2.205>
- Dujon, A. M., & Schofield, G. (2019). Importance of machine learning for enhancing ecological studies using information-rich imagery. *Endangered Species Research*, 39, 91–104. <https://doi.org/10.3354/esr00958>
- Dunstan, A., Robertson, K., Fitzpatrick, R., Pickford, J., & Meager, J. (2020). Use of unmanned aerial vehicles (UAVs) for mark-resight nesting population estimation of adult female green sea turtles at Raine Island. *PLoS ONE*, 15(6), 1–18. <https://doi.org/10.1371/journal.pone.0228524>
- Eikelboom, J. A. J., Wind, J., van de Ven, E., Kenana, L. M., Schroder, B., de Knegt, H. J., van Langevelde, F., & Prins, H. H. T. (2019). Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, 10(11), 1875–1887. <https://doi.org/10.1111/2041-210X.13277>
- Ezat, M. A., Fritsch, C. J., & Downs, C. T. (2018). Use of an unmanned aerial vehicle (drone) to survey Nile crocodile populations: A case study at Lake Nyamithi, Ndumo game reserve, South Africa. *Biological Conservation*, 223(January), 76–81. <https://doi.org/10.1016/j.biocon.2018.04.032>
- Ficetola, G. F., Barzaghi, B., Melotto, A., Muraro, M., Lunghi, E., Canedoli, C., Lo Parrino, E., Nanni, V., Silva-Rocha, I., Urso, A., Carretero, M. A., Salvi, D., Scali, S., Scari, G., Pennati, R., Andreone, F., & Manenti, R. (2018). N-mixture models reliably estimate the abundance of small vertebrates. *Scientific Reports*, 8(1), 1–8. <https://doi.org/10.1038/s41598-018-28432-8>
- Field, S. A., Tyre, A. J., & Possingham, H. P. (2005). Optimizing allocation of monitoring effort under economic and observational constraints. *The Journal of Wildlife Management*, 69(2), 473–482. [https://doi.org/10.2193/0022-541X\(2005\)069\[0473:OAOME U\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2005)069[0473:OAOME U]2.0.CO;2)
- Fiske, I. J., & Chandler, R. B. (2011). unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10), 1–23. <https://doi.org/10.1002/wics.10>
- Goebel, M. E., Perryman, W. L., Hinke, J. T., Krause, D. J., Hann, N. A., Gardner, S., & LeRo, D. J. (2015). A small unmanned aerial system for estimating abundance and size of Antarctic predators. *Polar Biology*, 38(5), 619–630. <https://doi.org/10.1007/s00300-014-1625-4>

- Guélat, J., Kéry, M., Gu, M. E., Id, O., & Isaac, N. (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, 9(6), 1614–1625. <https://doi.org/10.1111/2041-210X.12983>
- Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, 40(2), 281–295. <https://doi.org/10.1111/ecog.02445>
- Guillera-Arroita, G., Ridout, M. S., & Morgan, B. J. T. (2010). Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, 2002, 131–139. <https://doi.org/10.1111/j.2041-210X.2010.00017.x>
- Hamilton, G., Corcoran, E., Denman, S., Hennekam, M. E., & Koh, L. P. (2020). When you can't see the koalas for the trees: Using drones and machine learning in complex environments. *Biological Conservation*, 247 (September 2019), 108598. <https://doi.org/10.1016/j.biocon.2020.108598>
- Hodgson, A., Kelly, N., & Peel, D. (2013). Unmanned Aerial Vehicles (UAVs) for surveying marine fauna: A dugong case study. *PLoS ONE*, 8(11), e79556. <https://doi.org/10.1371/journal.pone.0079556>
- Hodgson, A., Peel, D., & Kelly, N. (2017). Unmanned aerial vehicles for surveying marine fauna: Assessing detection probability. *Ecological Applications*, 27(4), 1253–1267. <https://doi.org/10.1002/eap.1519>
- Hodgson, J. C., Mott, R., Baylis, S. M., Pham, T. T., Wotherspoon, S., Kilpatrick, A. D., Raja Segaran, R., Reid, I., Terauds, A., & Koh, L. P. (2018). Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, 9(5), 1160–1167. <https://doi.org/10.1111/2041-210X.12974>
- Hone, J. (2008). On bias, precision and accuracy in wildlife aerial surveys. *Wildlife Research*, 35(4), 253–257. <https://doi.org/10.1071/WR07144>
- Hong, S.-J., Han, Y., Kim, S.-Y., Lee, A.-Y., & Kim, G. (2019). Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Sensors*, 19(7), 1651. <https://doi.org/10.3390/s19071651>
- Joseph, L. N., Elkin, C., Martin, T. G., & Possingham, H. P. (2009). Modeling abundance using N-mixture models: The importance of considering ecological mechanisms. *Ecological Applications*, 19(3), 631–642. <https://doi.org/10.1890/07-2107.1>
- Keever, A. C., McGowan, C. P., Ditchkoff, S. S., Acker, P. K., Grand, J. B., & Newbolt, C. H. (2017). Efficacy of N-mixture models for surveying and monitoring white-tailed deer populations. *Mammal Research*, 62(4), 413–422. <https://doi.org/10.1007/s13364-017-0319-z>
- Kellenberger, B., Marcos, D., & Tuia, D. (2018). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, 216, 139–153. <https://doi.org/10.1016/j.rse.2018.06.028>
- Kellner, K. F. (2015). jagsUI: A wrapper around rjags to streamline JAGS analyses. R package version 1.3.1.
- Kéry, M. (2018). Identifiability in N-mixture models: A large-scale screening test with bird data. *Ecology*, 99(2), 281–288. <https://doi.org/10.1002/ecy.2093>
- Kéry, M., Dorazio, R. M., Soldaat, L., Van Strien, A., Zuiderwijk, A., & Royle, J. A. (2009). Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, 46, 1163–1172. <https://doi.org/10.1111/j.1365-2664.2009.01724.x>
- Kéry, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology* (Vol. 1) (Prelude and Static model). Academic Press.
- Kéry, M., Royle, J. A., & Schmid, H. (2005). Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications*, 15(4), 1450–1461. <https://doi.org/10.1890/04-1120>
- Kiszka, J., Mourier, J., Gastrich, K., & Heithaus, M. (2016). Using unmanned aerial vehicles (UAVs) to investigate shark and ray densities in a shallow coral lagoon. *Marine Ecology Progress Series*, 560(November), 237–242. <https://doi.org/10.3354/meps11945>
- Knape, J., Arlt, D., Barraquand, F., Berg, Å., Chevalier, M., Pärt, T., Ruete, A., & Zmihorski, M. (2018). Sensitivity of binomial N-mixture models to overdispersion: The importance of assessing model fit. *Methods in Ecology and Evolution*, 2018(July), 1–13. <https://doi.org/10.1111/2041-210X.13062>
- Knights, K., McCarthy, M. A., Camac, J., & Guillera-Arroita, G. (2021). Efficient effort allocation in line-transect distance sampling of high-density species: When to walk further, measure less-often and gain precision. *Methods in Ecology and Evolution*, 2021(February), 1–9. <https://doi.org/10.1111/2041-210X.13589>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Linchant, J., Lisein, J., Semeki, J., Lejeune, P., & Vermeulen, C. (2015). Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges. *Mammal Review*, 45(4), 239–252. <https://doi.org/10.1111/mam.12046>
- Link, W. A., Schofield, M. R., Barker, R. J., & Sauer, J. R. (2018). On the robustness of N-mixture models. *Ecology*, 99(7), 1547–1551. <https://doi.org/10.1002/ecy.2362>
- Mackenzie, D. I., & Royle, J. A. (2005). Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology*, 42(6), 1105–1114. <https://doi.org/10.1111/j.1365-2664.2005.01098.x>
- Martin, J., Edwards, H. H., Fonnesebeck, C. J., Koslovsky, S. M., Harmak, C. W., & Dane, T. M. (2015). Combining information for monitoring at large spatial scales: First statewide abundance estimate of the Florida manatee. *Biological Conservation*, 186, 44–51. <https://doi.org/10.1016/j.biocon.2015.02.029>
- Martin, J., Royle, J. A., Mackenzie, D. I., Edwards, H. H., Kéry, M., & Gardner, B. (2011). Accounting for non-independent detection when estimating abundance of organisms with a Bayesian approach. *Methods in Ecology and Evolution*, 2(6), 595–601. <https://doi.org/10.1111/j.2041-210X.2011.00113.x>
- Mourão, G., Coutinho, M., Mauro, R., Campos, Z., Tomás, W., & Magnusson, W. (2000). Aerial surveys of caiman, marsh deer and pampas deer in the Pantanal Wetland of Brazil. *Biological Conservation*, 92(2), 175–183. [https://doi.org/10.1016/S0006-3207\(99\)00051-8](https://doi.org/10.1016/S0006-3207(99)00051-8)
- Patterson, C., Koski, W., Pace, P., McLuckie, B., & Bird, D. M. (2016). Evaluation of an unmanned aircraft system for detecting surrogate caribou targets in Labrador. *Journal of Unmanned Vehicle Systems*, 4(1), 53–69. <https://doi.org/10.1139/juvs-2015-0014>
- Piovezan, U., Tiepolo, L. M., Tomas, W. M., Duarte, J. M. B., Varela, D., & Marinho-Filho, J. S. (2010). Marsh deer *Blastocercus dichotomus* (Illiger 1815). In J. M. B. Duarte & S. Gonzalez (Eds.), *Neotropical cervidology: Biology and medicine of Latin American deer* (pp. 66–76). Funep/IUCN.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.
- Preston, T. M., Survey, U. S. G., Rocky, N., & Science, M. (2021). Enumerating white-tailed deer using unmanned aerial vehicles. 1–12. <https://doi.org/10.1002/wsb.1149>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing <https://www.R-project.org/>
- Rey, N., Volpi, M., Joost, S., & Tuia, D. (2017). Detecting animals in African Savanna with UAVs and the crowds. *Remote Sensing of Environment*, 200, 341–351. <https://doi.org/10.1016/j.rse.2017.08.026>
- Ríos-Uzeda, B., & Mourão, G. (2012). Densities of the Vulnerable marsh deer *Blastocercus dichotomus* in Bolivia's northern savannahs. *Oryx*, 46(2), 260–265. <https://doi.org/10.1017/S0030605311000238>
- Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1), 108–115. <https://doi.org/10.1111/j.0006-341X.2004.00142.x>
- Royle, J. A., Dorazio, R. M., & Link, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *Journal*

of *Computational and Graphical Statistics*, 16(1), 67–85. <https://doi.org/10.1198/106186007X181425>

- Sollmann, R., Gardner, B., Williams, K. A., Gilbert, A. T., & Veit, R. R. (2016). A hierarchical distance sampling model to estimate abundance and covariate associations of species and communities. *Methods in Ecology and Evolution*, 7(5), 529–537. <https://doi.org/10.1111/2041-210X.12518>
- Terletzky, P. A., & Koons, D. N. (2016). Estimating ungulate abundance while accounting for multiple sources of observation error. *Wildlife Society Bulletin*, 40(3), 525–536. <https://doi.org/10.1002/wsb.672>
- Veech, J. A., Ott, J. R., & Troy, J. R. (2016). Intrinsic heterogeneity in detection probability and its effect on N-mixture models. *Methods in Ecology and Evolution*, 7(9), 1019–1028. <https://doi.org/10.1111/2041-210X.12566>
- Vermeulen, C., Lejeune, P., Lisein, J., Sawadogo, P., & Bouché, P. (2013). Unmanned Aerial Survey of Elephants. *PLoS ONE*, 8(2), e54700. <https://doi.org/10.1371/journal.pone.0054700>
- Weinstein, B. G. (2017). A computer vision for animal ecology. *Journal of Animal Ecology*, 87(3), 1–13. <https://doi.org/10.1111/1365-2656.12780>
- Williams, P. J., Hooten, M. B., Womble, J. N., & Bower, M. R. (2017). Estimating occupancy and abundance using aerial images with imperfect detection. *Methods in Ecology and Evolution*, 8(12), 1679–1689. <https://doi.org/10.1111/2041-210X.12815>
- Yamaura, Y. (2013). Confronting imperfect detection: Behavior of binomial mixture models under varying circumstances of visits, sampling sites, detectability, and abundance, in small-sample situations. *Ornithological Science*, 12(2), 73–88. <https://doi.org/10.2326/osj.12.73>
- Zipkin, E. F., Thorson, J. T., See, K., Lynch, H. J., Grant, E. H. C., Kanno, Y., Chandler, R. B., Letcher, B. H., & Royle, J. A. (2014). Modeling structured population dynamics using data from unmarked individuals. *Ecology*, 95(1), 22–29. <https://doi.org/10.1890/13-1131.1>
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, 119(4), 622–635. <https://doi.org/10.1111/j.1600-0706.2009.18284.x>
- Zylstra, E. R., Steidl, R. J., & Swann, D. E. (2009). Evaluating survey methods for monitoring a rare vertebrate, the Sonoran Desert Tortoise. *The Journal of Wildlife Management*, 74(6), 1311–1318. <https://doi.org/10.2193/2009-331>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. N-mixture model performance (relative RMSE curves) and optimal number of visits for single and double observer counts under different scenarios of local abundance and availability probability. Lower panels show the optimal number of visits (i.e. lowest rel. RMSE) and the bars correspond to the range of visits for which the performance can be considered equivalent (i.e. rel. RMSE < 0.5%).

Figure S2. N-mixture model performance (relative RMSE curves) and optimal number of visits for single and double observer counts under different scenarios of local abundance and availability probability. Lower panels show the optimal number of visits (i.e. lowest rel. RMSE) and the bars correspond to the range of visits for which the performance can be considered equivalent (i.e. rel. RMSE < 0.5%).

Figure S3. N-mixture model performance (relative RMSE curves) and optimal number of visits for two different total budgets (2k = 2000 flights, 4k = 4000 flights) in six scenarios of local abundance {0.2; 1} and availability probabilities {0.2; 0.4; 0.6}. Lower panels show the optimal number of visits (i.e. lowest rel. RMSE) and the bars correspond to the range of visits for which the performance can be considered equivalent (i.e. rel. RMSE < 0.5%).

Figure S4. N-mixture model performance (relative RMSE curves) and optimal number of visits using different proportions of double-observer protocol for six scenarios, from the combination of local abundances {0.2; 1} and availability probabilities {0.2; 0.4; 0.6}. Lower panels show the optimal number of visits (i.e. lowest rel. RMSE) and the bars correspond to the range of visits for which the performance can be considered equivalent (i.e. rel. RMSE < 0.5%).

Figure S5. Performance (relative RMSE) of the double observer (multinomial) N-mixture model for different total efforts (sites x visits) in comparison with the performance of a baseline single observer (binomial) N-mixture model with total effort = 2000.

Figure S6. Relative error in estimated local abundance for N-mixture models (difference between the mean point estimate and the real value divided by the real value) under different scenarios of local abundance and availability probability. Boxplots are composed of 2000 iterations. Boxes contain the 50% central values, the bold line indicates the median and the whiskers correspond to an extension of the box (upper and below) using 1.5 times the interquartile range. Points are outliers.

How to cite this article: Brack, I. V., Kindel, A., de Oliveira, L. F. B., & Lahoz-Monfort, J. J. (2023). Optimally designing drone-based surveys for wildlife abundance estimation with N-mixture models. *Methods in Ecology and Evolution*, 14, 898–910. <https://doi.org/10.1111/2041-210X.14054>