bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Using visual encounter data to improve capture-recapture abundance estimates

Maxwell B. Joseph* Roland A. Knapp†

Abstract

Capture-recapture studies are widely used in ecology to estimate population sizes and demographic rates. In some capture-recapture studies, individuals may be visually encountered but not identified. For example, if individual identification is only possible upon capture and individuals escape capture, visual encounters can result in failed captures where individual identities are unknown. In such cases, the data consist of capture histories with known individual identities, and counts of failed captures for individuals with unknown identities. These failed captures are ignored in traditional capture-recapture analyses that require known individual identities. Here we show that if animals can be encountered at most once per sampling occasion, failed captures provide lower bounds on population size that can increase the precision of abundance estimates. Analytical results and simulations indicate that visual encounter data improve abundance estimates when capture probabilities are low, and when there are few repeat surveys. We present a hierarchical Bayesian approach for integrating failed captures and auxiliary encounter data in statistical capture-recapture models. This approach can be integrated with existing capture-recapture models, and may prove particularly useful for hard to capture species in data-limited settings.

*maxwell.b.joseph@colorado.edu, Earth Lab, University of Colorado, Boulder, CO 80303, USA

†Sierra Nevada Aquatic Research Laboratory, University of California, Mammoth Lakes, CA 93546, USA

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Capture-recapture studies are widely used for estimating abundance and demographic rates, using information about the identities of captured individuals (Jolly 1965). This paper examines the case where individual identification requires capture, and identities of animals that are visually encountered but not captured are unknown. In such cases, "capture" is synonymous with "identification". This often applies in capture-recapture studies of amphibians, where individual identification is only possible with the animal in hand. We also assume that if an individual escapes capture, it is not encountered again in the sampling occasion because it is hiding or otherwise inaccessible (Bailey and Nichols 2010, Joseph and Knapp 2018). Under these conditions, encounters leading to failed captures provide information about abundance, but this information is not readily used in traditional capture-recapture models.

When individuals can be encountered at most once on a survey, encounter and capture data both provide lower bounds on the total number of individuals in the population. Total abundance must be greater than or equal to the number of animals encountered in a survey. Similarly, total abundance must be greater than or equal to the number of unique individuals identified in the capture data. Capture data differ however, in that information accumulates over multiple surveys (Pollock 1982). For example, if two surveys occur on consecutive days in a closed population, then the total number of unique individuals captured across both surveys provides a lower bound on abundance.

Here, we show how visual encounter data can improve abundance estimates in capture-recapture studies for study designs where 1) individual identification requires capture, 2) a subset of encountered individuals are captured, and 3) individuals can be encountered at most once per sampling occasion. We develop a modified capture-recapture observation model, and investigate conditions under which encounter data improve abundance estimates. The methods presented here accommodate both failed captures and counts of animals

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

collected separately from capture-recapture surveys, and can be integrated with existing capture-recapture models.

Methods

Model description

We adopt a hierarchical Bayesian approach in which an observation model depends on a state model, and both depend on some parameters (Berliner 1996). The state model describes the presence or absence of individuals in a population, and the observation model describes the visual encounter and capture process. The parameter model represents prior distributions for all remaining unknowns.

State model

Abundance can be estimated from capture-recapture data in a Bayesian framework using parameter-expanded data augmentation (Royle and Dorazio 2012). Here, N^* unique individuals are observed, but $M > N^*$ individuals are modeled, augmenting the observed data with $M - N^*$ additional capture histories of animals that were never captured. The assumption is that the true abundance N is less than M, but greater than the number of observed individuals N^*.

Individuals $i = 1, ..., M$ are either "in the population" ($z_i = 1$) or not ($z_i = 0$), where the parameters $z_1, ..., z_M$ are state parameters to be estimated, and abundance is $N = \sum_i z_i$. These states can be modeled as conditionally independent Bernoulli random variables with probability parameter ω, where ω is the probability of an individual being in the population:

$$z_i \sim \text{Bernoulli}(\omega).$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Observation model

On surveys $k = 1, ..., K$ an observer searches for individuals, encountering each individual with probability η. We assume individuals can only be encountered once at most. If an animal is encountered, it is captured with probability κ. Because individuals must be captured to be identified, encounter data are observed for captured individuals, but not for failed captures. Thus, encounter histories are partly observed. We assume the number of failed captures on each survey is observed.

Let $y^*_{i,k}$ represent the categorical outcome for individual i on survey k. There are three possibilities (Figure 1):

1. The individual was not encountered ($y^*_{i,k} = 1$), with probability $z_i(1 - \eta) + 1 - z_i$.
2. The individual was encountered but not captured ($y^*_{i,k} = 2$), with probability $z_i\eta(1 - \kappa)$.
3. The individual was captured ($y^*_{i,k} = 3$), with probability $z_i\eta\kappa$.

The first two outcomes are not observed. We observe a binary record of whether individual i was captured on survey k: $y_{i,k}$, so that $y_{i,k} = I(y^*_{i,k} = 3)$, where I is an indicator function that is equal to one if the condition inside the parentheses is satisfied, otherwise it equals zero. In other words, $y^*_{i,k}$ is observed only if $y^*_{i,k} = 3$. Additionally, the observed number of failed captures f_k corresponds to the sum $f_k = \sum_i I(y^*_{i,k} = 2)$. The observation model consists of two parts: one for the capture data:

$$[y_{i,k} \mid y^*_{i,k}] = \text{Bernoulli}(I(y^*_{i,k} = 3)),$$

and another for the failed capture counts:

$$[f_k \mid y^*_{1,k}, ..., y^*_{M,k}] = I(f_k = \sum_i I(y^*_{i,k} = 2)),$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where square brackets denote a probability function. Alternatively, a "soft" constraint can be imposed as an approximation or to account for uncertainty in the number of failed captures:

$$[f_k \mid y^*_{1,k}, ..., y^*_{M,k}] = \text{Normal}(f_k \mid \sum_i I(y^*_{i,k} = 2), \sigma),$$

with σ set to some small fixed value.

Parameter model

The hierarchical model specification is completed by specifying prior distributions for remaining unknowns. Here, we use independent Uniform(0, 1) priors for all probabilities. This prior over the inclusion probability ω implies a discrete uniform prior for the true abundance from 0 to M.

Posterior distribution

The parameters consist of states $z_1, ..., z_M$, outcomes $y^*_{1,1}, ..., y^*_{M,K}$, and the probabilities of inclusion (ω), encounter (η), and capture (κ). The data consist of the capture histories $y_{1,1}, ..., y_{M,K}$, and counts of failed captures from each survey $f_1, ..., f_K$. The posterior distribution of parameters given data, $[z_1, ..., z_M, y^*_{1,1}, ..., y^*_{M,K}, \omega, \eta, \kappa \mid y_{1,1}, ..., y_{M,K}, f_1, ..., f_K]$, is proportional to:

$$\overbrace{\prod_i \prod_k [y_{i,k} \mid y^*_{i,k}]}^{\text{Captures}} \times \overbrace{\prod_k [f_k \mid y^*_{1,k}, ..., y^*_{M,k}]}^{\text{Failed captures}} \times$$
$$\overbrace{\prod_i \prod_k [y^*_{i,k} \mid z_i, \eta, \kappa]}^{\text{Encounter model}} \times \overbrace{\prod_i [z_i \mid \omega]}^{\text{State model}} \times \overbrace{[\omega][\eta][\kappa]}^{\text{Priors}}.$$

This model can be implemented in the popular BUGS language (Lunn et al. 2009), and we provide example specifications in Appendix S1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Auxiliary encounter data

In some cases, auxiliary encounter data are collected, such as during visual encounter surveys where individuals are counted on surveys, but no captures are attempted (Crump and Scott Jr 1994). Let a_k represent the number of unique individuals that were encountered on survey k. If encounters of individuals are conditionally independent, a binomial observation model provides a reasonable choice, where the number of trials is the population abundance $N = \sum_i z_i$ and the probability of success is the encounter probability η:

$$[a_k \mid \eta, z_1, ..., z_M] = \text{Binomial}(\eta, \sum_i z_i),$$

for surveys $k = 1, ..., K$. This is the observation model used in N-mixture models (Royle 2004). In addition to potentially providing a higher lower bound on abundance, auxiliary encounters also provide additional information about encounter probabilities, because the encounter data are conditionally independent from encounters leading to captures. When combined with the capture-recapture model outlined above, the joint model of captures, failed captures, and auxiliary encounters comprise an integrated population model (Besbeas et al. 2002, Abadi et al. 2010). The posterior distribution is proportional to:

$$\overbrace{\prod_i \prod_k [y_{i,k} \mid y^*_{i,k}]}^{\text{Captures}} \times \overbrace{\prod_k [f_k \mid y^*_{1,k}, ..., y^*_{M,k}]}^{\text{Failed captures}} \times \overbrace{\prod_k [a_k \mid \sum_i z_i, \eta]}^{\text{Auxiliary encounters}} \times$$
$$\overbrace{\prod_i \prod_k [y^*_{i,k} \mid z_i, \eta, \kappa]}^{\text{Encounter model}} \times \overbrace{\prod_i [z_i \mid \omega]}^{\text{State model}} \times \overbrace{[\omega][\eta][\kappa]}^{\text{Priors}}.$$

Abundance lower bounds from encounter and capture data

The total population size is bounded from below by the number animals encountered on any one survey, assuming each animal can be encountered once at most (i.e., individuals are not double-counted). If n_k is the number of unique animals encountered on survey k,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the lower bound on abundance from encounter data $n_{\min} = \max(n_1, ..., n_K)$ is the maximum of K independent binomial random variables with sample size N and probability η. The probability mass function of this lower bound is thus given by:

$$\Pr(n_{\min} = n) = F(n)^K - F(n-1)^K,$$

where $F(n)$ is the cumulative distribution function of a binomial random variable.

Population size must also be greater than or equal to the number of unique captured individuals. A probability mass function for the lower bound on abundance from capture data ($c_{\min}$: the number of unique captured individuals) can be derived with a binomial distribution. The binomial sample size is the true population size (N), and the probability of success is the probability of being captured one or more times: $1 - (1 - \eta\kappa)^K$. The probability mass function for $c_{\min}$, the abundance lower bound derived from the capture data, is:

$$\Pr(c_{\min} = n) = \text{Binomial}(n \mid 1 - (1 - \eta\kappa)^K, N).$$

When the expected lower bound from encounter data exceeds the expected lower bound from capture data ($\mathbb{E}(n_{\min}) > \mathbb{E}(c_{\min})$), encounter data are expected to increase the precision of abundance estimates.

Simulations

We empirically verified our theoretical results about the expected lower bounds on abundance provided by encounter and capture data using Monte Carlo simulation. We generated five replicate encounter-capture-recapture data sets for each parameter combination of $N = 10, 50, 100$, $K = 3, 6, 9$, and η and κ ranging from 0.01 to 0.99 in increments of 0.01, resulting in 441,045 unique data sets. For each parameter combination, we computed the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

empirical mean lower bounds from encounter and capture data, averaging over the five replicate iterations, and compared the results to the theoretical expectations generated from the probability mass functions for $n_{\min}$ and $c_{\min}$.

To understand the implications of bounding abundance for other parameters, we used a simulation study with known parameters across a range of repeat surveys ($K = 3$, $K = 6$, and $K = 9$). We visualized the joint posterior distribution of abundance and the probability of being encountered and captured, and compared these results to a simpler model: M_0 - a capture-recapture model of a closed population with identical detection probabilities $p_1 = ... = p_K = p$, which ignores encounters that do not lead to captures (Royle and Dorazio 2008). This model can only estimate the marginal probability of capture $p = \eta\kappa$. The observation model is $y_{i,k} \sim \text{Bernoulli}(z_i p)$ for individuals $i = 1, ..., M$ on survey $k = 1, ..., K$. The state model for z is unchanged, and uniform priors over (0, 1) were assigned to ω and p.

Draws from the posterior distributions of all models were simulated using JAGS, with six parallel Markov chain Monte Carlo chains, and 400,000 iterations per chain with an adaptation period of 200,000, a burn-in period of 40,000, and posterior thinning by 400 to reduce memory usage (Plummer and others 2003). Convergence was assessed using visual inspection of traceplots, and the potential scale reduction factor ($\hat{R}$) statistic (Gelman et al. 1992). All code to replicate the analyses is available in a research compendium at https://github.com/mbjoseph/secmr.

Results

Across a range of abundances, encounter data are expected to provide a higher lower bound on abundance when capture probabilities are low, and when there are few repeat surveys (Figure 2). The boundary in the bivariate encounter-capture parameter space delineating the region where $\mathbb{E}(c_{\min} - n_{\min}) < 0$ shifts toward lower capture probabilities as the number

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of repeat surveys increases, and as abundance increases. Empirical average bounds from simulated capture-recapture data were in agreement with the theoretical expectations derived from the probability mass functions of $n_{\min}$ and $c_{\min}$ (Figure 3).

When the lower bound on abundance is greater for encounter data than capture data, the joint model of encounters and captures produces a more precise estimate of abundance because there is zero probability mass below the lower bound on abundance. Furthermore, if there is posterior correlation between abundance and another parameter, encounter data can also increase the precision of the correlated parameter. For example, the marginal probability of capture and abundance are correlated in the posterior, so that increased posterior precision for abundance implies increased posterior precision of marginal capture probability (Figure 4).

Discussion

Failed captures and auxiliary encounters are likely to be most useful in capture-recapture studies when animals are hard to capture and the number of surveys is small. This expectation holds across a range of population abundance and encounter probabilities. In such cases, encounter data increase the precision of population abundance estimates by increasing the lower bound on abundance. Such data are essentially "free" in encounter-capture-recapture study designs, and can be included by modifying the likelihood function (and not the underlying state model) of capture-recapture models.

In addition to increasing the precision of abundance estimates, encounter data can increase the precision of parameter estimates for parameters that are correlated with abundance in the posterior distribution. For the simple model presented here, this includes the detection and inclusion probabilities. For more complex models that allow state evolution through time, this might include survival and recruitment probabilities. Thus, we expect that encounter

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

data in general might provide information about abundance, and parameters relating to abundance and the measurement process.

These results are consistent with related findings for mark-resight studies where marked individuals are subject to incomplete identification. In such studies, accounting for failed identifications of marked individuals – analogous to failed captures – is most advantageous when identification probabilities are low – analogous to capture probabilities being low (McClintock et al. 2014b). However, in the encounter-capture-recapture scenario considered here, whether an animal is marked or not is unknown until it is captured, as would be the case for subdermal passive integrated transponder tags in an amphibian (Gibbons and Andrews 2004).

Here we assumed that individuals could be encountered at most once per sampling occasion. In other words, there is no double counting: each failed capture corresponds to one unique individual. The observation model developed here is not robust to violations of this assumption, because the total number of encounters sets a lower bound on the true population size. As a consequence, if individuals escape capture multiple times in the sampling sampling occasion, it is possible that the posterior for abundance might be misleadingly precise (i.e., the lower bound on population size would be too high). Therefore, we do not recommend this approach if individuals might be encountered multiple times on the same sampling occasion. This assumption is likely to hold for example in capture-recapture studies of amphibians in high elevation lakes, where individual animals that escape capture hide afterwards, e.g., underwater where they cannot be seen or captured (Joseph and Knapp 2018). If individuals are captured multiple times in the same sampling occasion, then it will be clear that the assumption of at most one encounter has been violated. In such cases, alternative encounter models may be necessary, e.g., a Poisson model that allows repeat encounters on a sampling occasion (Royle et al. 2009).

The model developed here relates to other approaches for handling imperfect individual

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

identification in capture-recapture studies including misidentification (Link et al. 2010, McClintock et al. 2014a, Schofield and Bonner 2015) and partial identification (Augustine et al. 2018), and also to approaches that account for latent encounter histories with observed summary counts (Chandler et al. 2013). In particular, this approach might be viewed as an aspatial analog of the spatial model presented in Chandler et al. (2013) in which a subset of individual identities are available, with a Bernoulli (instead of a Poisson) encounter model, and a stochastic second stage "identification upon encounter" component. This approach can also be seen as a degenerate case of a partial identification model, in which there are no spatial (Augustine et al. 2018) or genetic data (Wright et al. 2009) available to inform the identities of individuals that have escaped capture.

Looking ahead, there are opportunities to build upon this approach. First, in terms of implementation, marginalization over the discrete latent variables might allow more efficient sampling from the posterior distribution. Second, because this model includes separate parameters for encounter and capture probabilities, covariates can be included separately for each of these components. This could be useful for example to account for predator avoidance behavior that might influence capture probabilities, and weather conditions that might influence encounter probabilities. Observer effects provide an additional use case: some observers might be better than others at finding or capturing individual animals.

In this paper, we presented a motivation for including encounter data in capture-recapture studies based on abundance lower bounds from encounter and capture data. Given that encounter data are included via a modified likelihood and not a modified state model, this approach can be readily integrated with a variety of capture-recapture models, and may be useful for hard to capture species in data-limited settings.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements

This work was motivated by years experience in the field capturing (and failing to capture) amphibians in the Sierra Nevada and the Klamath mountains, and funded by a grant from the Yosemite Conservancy. We thank Ben Augustine for helpful comments on an earlier version of the manuscript.

References

Abadi, F., O. Gimenez, R. Arlettaz, and M. Schaub. 2010. An assessment of integrated population models: Bias, accuracy, and violation of the assumption of independence. Ecology 91:7–14.

Augustine, B. C., J. A. Royle, M. J. Kelly, C. B. Satter, R. S. Alonso, E. E. Boydston, K. R. Crooks, and others. 2018. Spatial capture–recapture with partial identity: An application to camera traps. The Annals of Applied Statistics 12:67–95.

Bailey, L. L., and J. D. Nichols. 2010. Capture–mark–recapture, removal sampling, and occupancy models. Amphibian Ecology and Conservation: A Handbook of Techniques. CK Dodd (ed.). Oxford University Press, New York:447–463.

Berliner, L. M. 1996. Hierarchical bayesian time series models. Pages 15–22 *in* Maximum entropy and bayesian methods. Springer.

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

Chandler, R. B., J. A. Royle, and others. 2013. Spatially explicit models for inference about density in unmarked or partially marked populations. The Annals of Applied Statistics

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

7:936–954.

Crump, M., and N. Scott Jr. 1994. Visual encounter surveys. In "measuring and monitoring biological diversity: Standard methods for amphibians". Smithsonian Institution: Washington, DC.

Gelman, A., D. B. Rubin, and others. 1992. Inference from iterative simulation using multiple sequences. Statistical science 7:457–472.

Gibbons, W. J., and K. M. Andrews. 2004. PIT tagging: Simple technology at its best. Bioscience 54:447–454.

Jolly, G. M. 1965. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. Biometrika 52:225–247.

Joseph, M. B., and R. A. Knapp. 2018. Disease and climate effects on individuals drive post-reintroduction population dynamics of an endangered amphibian. Ecosphere 9.

Link, W. A., J. Yoshizaki, L. L. Bailey, and K. H. Pollock. 2010. Uncovering a latent multinomial: Analysis of mark–recapture data with misidentification. Biometrics 66:178–185.

Lunn, D., D. Spiegelhalter, A. Thomas, and N. Best. 2009. The bugs project: Evolution, critique and future directions. Statistics in medicine 28:3049–3067.

McClintock, B. T., L. L. Bailey, B. P. Dreher, W. A. Link, and others. 2014a. Probit models for capture–recapture data subject to imperfect detection, individual heterogeneity and misidentification. The Annals of Applied Statistics 8:2461–2484.

McClintock, B. T., J. M. Hill, L. Fritz, K. Chumbley, K. Luxa, and D. R. Diefenbach. 2014b. Mark-resight abundance estimation under incomplete identification of marked individuals. Methods in Ecology and Evolution 5:1294–1304.

Plummer, M., and others. 2003. JAGS: A program for analysis of bayesian graphical models using gibbs sampling. Page 10 *in* Proceedings of the 3rd international workshop on distributed

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

statistical computing. Vienna, Austria.

Pollock, K. H. 1982. A capture-recapture design robust to unequal probability of capture. The Journal of Wildlife Management 46:752–757.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities. Elsevier.

Royle, J. A., and R. M. Dorazio. 2012. Parameter-expanded data augmentation for bayesian analysis of capture–recapture models. Journal of Ornithology 152:521–537.

Royle, J. A., K. U. Karanth, A. M. Gopalaswamy, and N. S. Kumar. 2009. Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. Ecology 90:3233–3244.

Schofield, M. R., and S. J. Bonner. 2015. Connecting the latent multinomial. Biometrics 71:1070–1080.

Wright, J. A., R. J. Barker, M. R. Schofield, A. C. Frantz, A. E. Byrom, and D. M. Gleeson. 2009. Incorporating genotype uncertainty into mark–recapture-type models for estimating abundance using dna samples. Biometrics 65:833–840.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure legends

Figure 1

Conceptual diagram to represent the data model for individuals $i = 1, ..., M$ and sampling occasions or surveys $k = 1, ..., K$. Each individual is either in the population ($z_i = 1$) or not ($z_i = 0$). Those that are not are never encountered. Those that are may be encountered on occasion k or not, and encountered individuals may or may not be captured and identified (we assume that identification requires capture, so that capture and identification are synonymous). Each path leads to a value of the partly observed quantity $y^*_{i,k}$, where $y^*_{i,k} = 1$ when animals are not encountered, $y^*_{i,k} = 2$ when animals are encountered but not identified, and $y^*_{i,k} = 3$ when animals are encountered and identified.

Figure 2

Expectations for the difference in abundance lower bounds provided by capture and encounter data as a function of the number of surveys K, abundance N, the encounter probability η, and the probability of capture conditional on an encounter κ. When the surface is red, encounter data are expected to increase the precision of abundance estimates by increasing the lower bound on true abundance. The heavy black line marks the null isocline where the expected difference is zero. Lighter lines represent contours spaced by 2 individuals.

Figure 3

Empirical verification of theoretical expectations for lower bounds on abundance provided by encounter and capture data. Each point represents the empirical average of five replicate simulations across a range of parameter values, with panels separated by population size (N) and whether the lower bounds are derived from capture data ($c_{\min}$) or encounter data ($n_{\min}$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

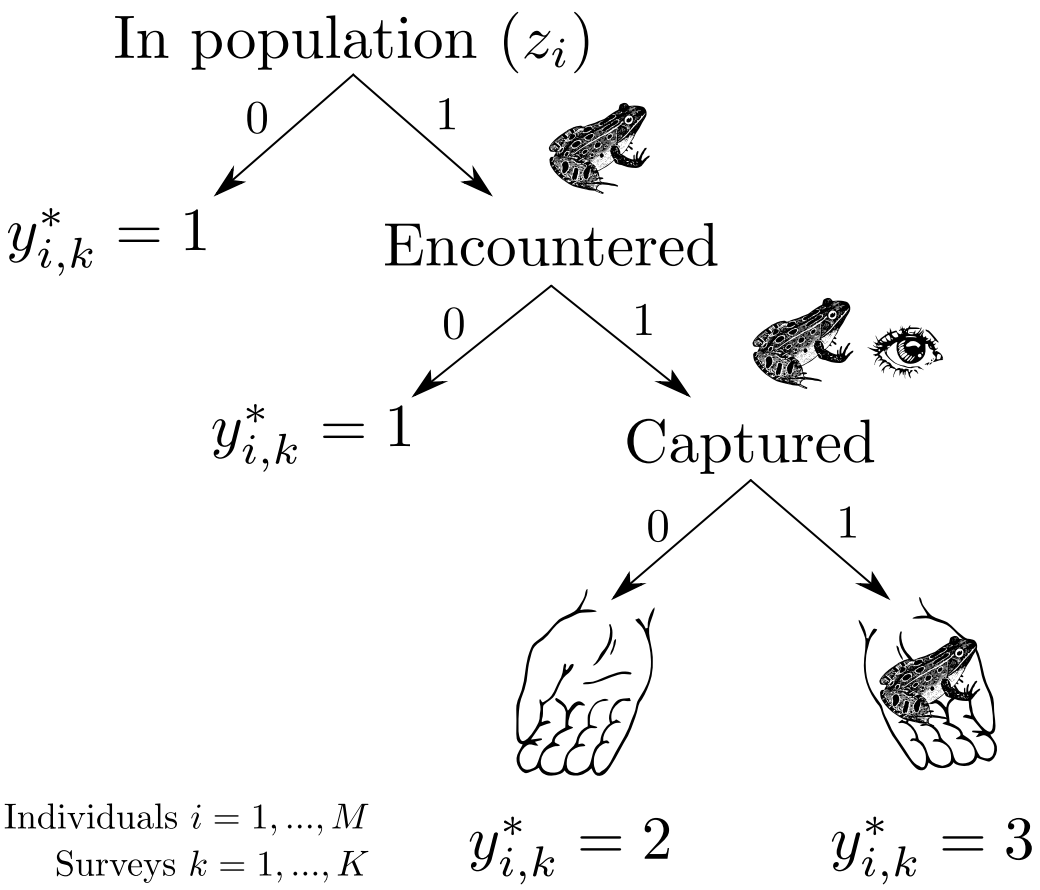
Figure 4

Samples from the posterior distribution of abundance (N, x-axis) and marginal capture probability $p = \eta\kappa$ for an individual in the population (y-axis). Each point is a sample from the posterior. Black points correspond to the baseline capture-recapture model M_0, which does not include encounter data. Blue points correspond to an encounter-capture-recapture model that uses encounter data to bound abundance. Vertical dashed lines are shown for the lower bounds on abundance derived from encounter (n_{min}) and capture (c_{min}) data.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

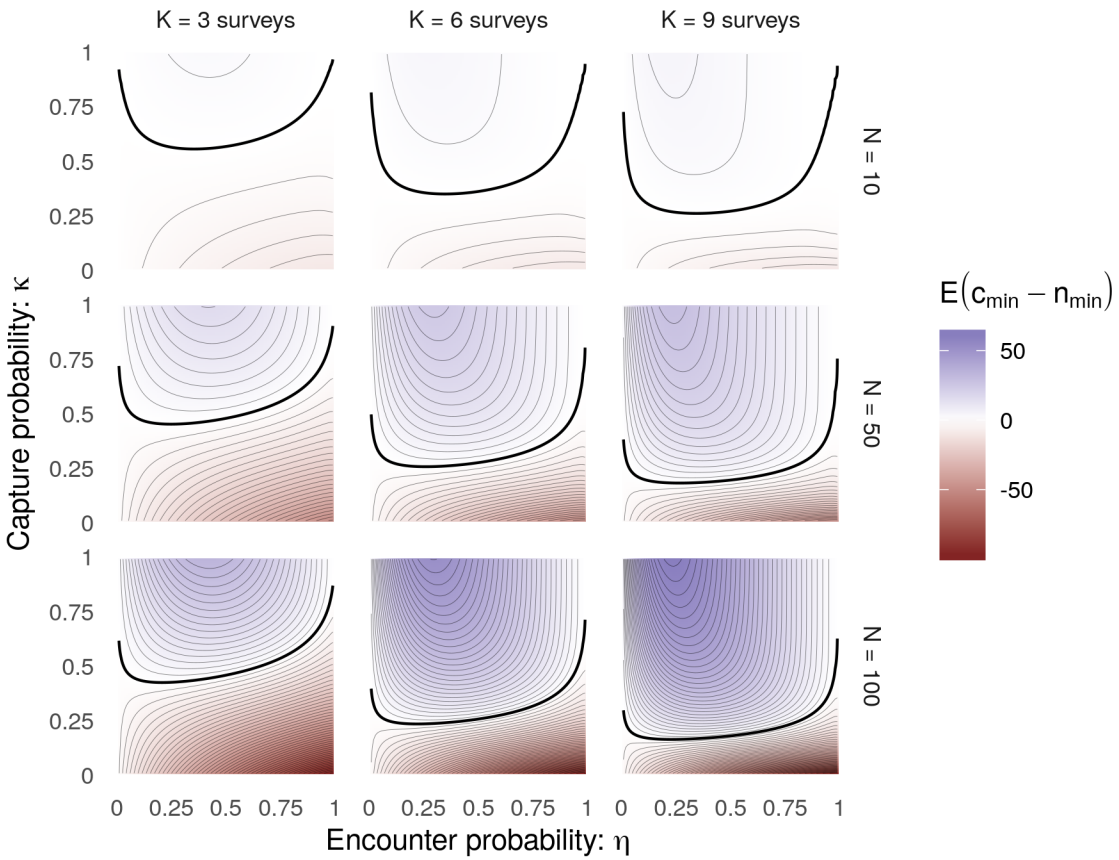
Figures

Figure 1



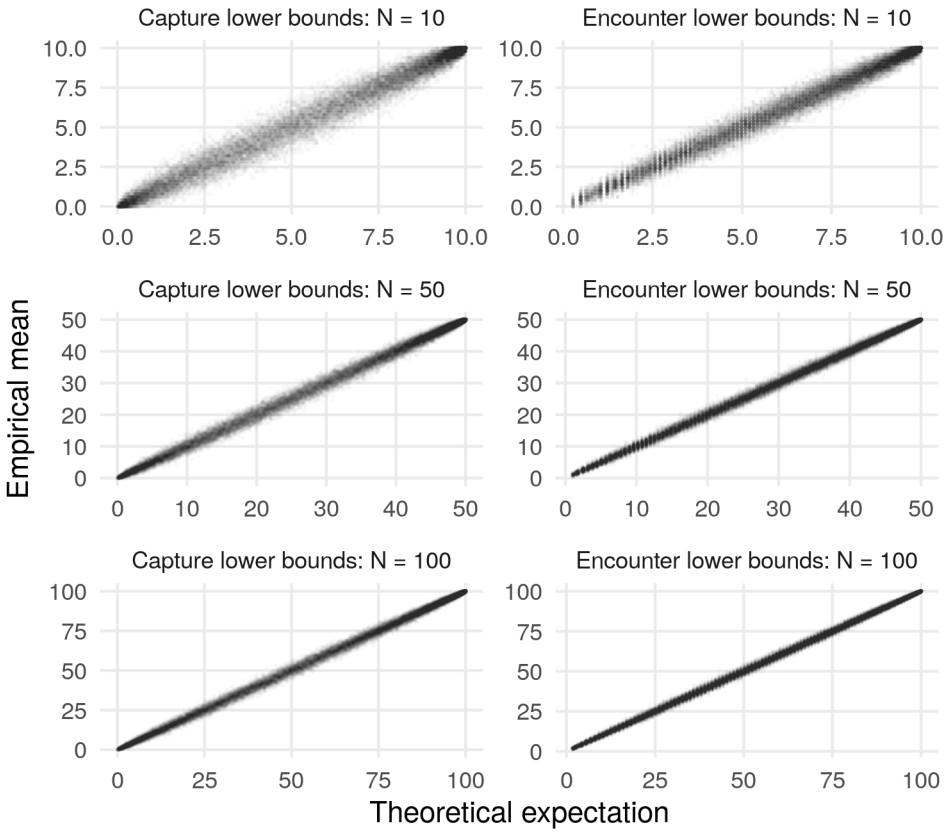
bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 2



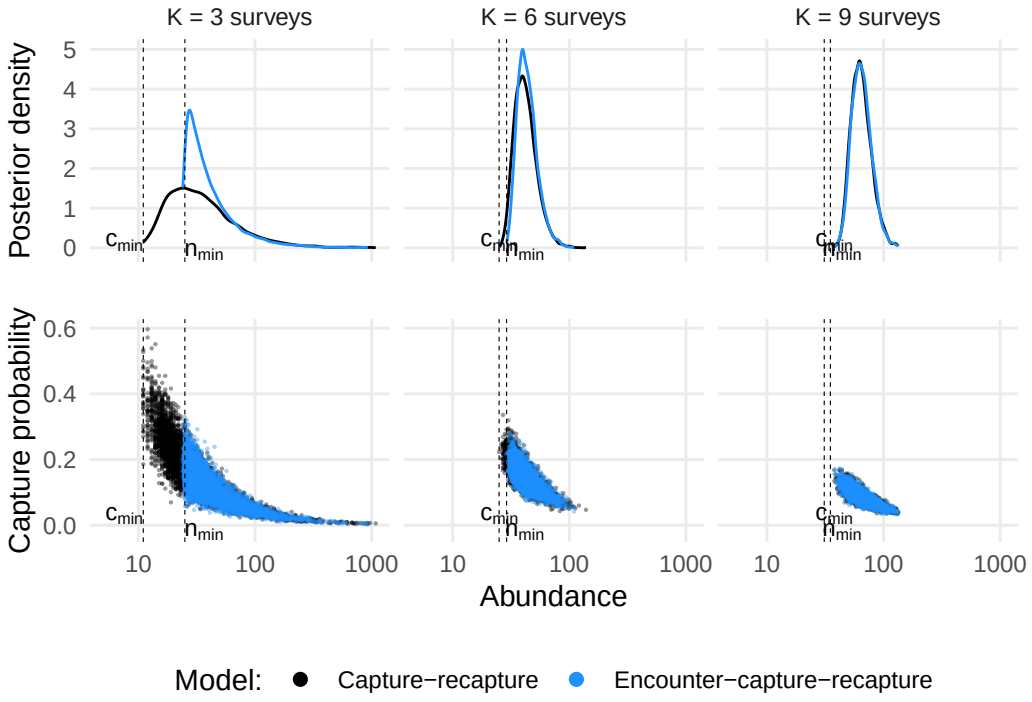
bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Appendix S1

This appendix includes model specifications and BUGS/JAGS code for failed capture and auxiliary encounter models. These models differ in the observation process, but the model of presence/absence states is identical. Each individual $i = 1, ..., M$ is either in the population ($z_i = 1$) or not ($z_i = 0$), where $z_i \sim \text{Bernoulli}(\omega)$, and ω is an inclusion probability parameter. Abundance is the sum of these values $N = \sum_i z_i$.

Baseline model

The baseline model is M_0 and assumes equal detection probabilities of individuals $i = 1, ..., M$ on surveys $k = 1, ..., K$, where $y_{i,k} = 1$ represents a capture of individual i on survey k, and $y_{i,k} = 0$ indicates no capture. If p is the capture probability, then $y_{i,k} \sim \text{Bernoulli}(z_i p)$. The posterior distribution is:

$$\left[z_1, ..., z_M, \omega, p \mid y_{1,1}, ..., y_{M,K}\right] \propto$$
$$\overbrace{\prod_i \prod_k [y_{i,k} \mid z_i, p]}^{\text{Captures}} \times \times \overbrace{\prod_i [z_i \mid \omega]}^{\text{State model}} \times \overbrace{[\omega][p]}^{\text{Priors}}.$$

JAGS code for this model with uniform priors over ω and p is:

```
model {
omega ~ dbeta(1, 1)
p ~ dbeta(1, 1)

for (i in 1:M) {
  z[i] ~ dbern(omega)
  for (k in 1:K) {
    y[i, k] ~ dbern(z[i] * p)
```

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

```
  }
}
}
```

Baseline model with failed captures

We add failed captures to the baseline model by introducing a categorical parameter $y^*_{i,k}$, which represents "not encountered" ($y^*_{i,k} = 1$), "failed capture" ($y^*_{i,k} = 2$), or "capture" ($y^*_{i,k} = 3$). If η is the encounter probability, κ is the probability of capture conditional on encounter, and f_k is the number of failed captures, then as described in the main text the posterior is:

$$\left[z_1, ..., z_M, y^*_{1,1}, ..., y^*_{M,K}, \omega, \eta, \kappa \mid y_{1,1}, ..., y_{M,K}, f_1, ..., f_K\right] \propto \overbrace{\prod_i \prod_k [y_{i,k} \mid y^*_{i,k}]}^{\text{Captures}} \times \overbrace{\prod_k [f_k \mid y^*_{1,k}, ..., y^*_{M,k}]}^{\text{Failed captures}} \times \overbrace{\prod_i \prod_k [y^*_{i,k} \mid z_i, \eta, \kappa]}^{\text{Encounter model}} \times \overbrace{\prod_i [z_i \mid \omega]}^{\text{State model}} \times \overbrace{[\omega][\eta][\kappa]}^{\text{Priors}}.$$

JAGS code for this model with uniform priors over ω, η, and κ is:

```
model {
  omega ~ dbeta(1, 1)
  p_encounter ~ dbeta(1, 1)
  p_capture ~ dbeta(1, 1)
  p_encounter_and_capture = p_encounter * p_capture
  p_failed_capture = p_encounter * (1 - p_capture)
  p_not_encountered = 1 - p_encounter
```

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

```
  for (i in 1:M) {
    z[i] ~ dbern(omega)
    pi[i, 1] = z[i] * p_not_encountered + 1 - z[i]
    pi[i, 2] = z[i] * p_failed_capture
    pi[i, 3] = z[i] * p_encounter_and_capture
    for (k in 1:K) {
      y_star[i, k] ~ dcat(pi[i, ])
      y[i, k] ~ dbern(y_star[i, k] == 3)
    }
  }

  # number of failed captures (soft constraint)
  for (k in 1:K) {
    failures[k] ~ dnorm(sum(y_star[, k] == 2), 4)
  }
}
```

Note that the normal distribution for failed captures imposes a "soft" constraint on the sum $\sum_i y^*_{i,k}$. By adjusting the normal precision parameter, this constraint can be relaxed or made more strict. In the strictest case a hard constraint can be imposed via the "ones trick", where a vector of ones is passed as data, and the likelihood for failed captures is represented as `ones[k] ~ dbern(sum(y_star[, k]) == failures[k])`.

Baseline model with auxiliary encounters

To add auxiliary encounter data to the baseline model, we use the following likelihood for the number of auxiliary encounters a_k in survey k conditional on the true abundance N and the encounter probability η:

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$a_k \sim \text{Binomial}(\eta, N),$$

where $N = \sum_i z_i$. Note that there need not be an equal number (K) capture-recapture surveys and auxiliary encounter surveys, but here we assume this is the case to simplify notation. This model permits decomposing the parameter p, which represents the marginal probability of capture, into the product of the encounter probability η and the probability of capture conditional on encounter κ.

Then, the posterior is:

$$\left[z_1, ..., z_M, \omega, \eta, \kappa \mid y_{1,1}, ..., y_{M,K}, a_1, ..., a_K\right] \propto$$
$$\overbrace{\prod_i \prod_k [y_{i,k} \mid z_i, \eta, \kappa]}^{\text{Captures}} \times \overbrace{\prod_k [a_k \mid \sum_i z_i, \eta]}^{\text{Auxiliary encounters}} \times \times \overbrace{\prod_i [z_i \mid \omega]}^{\text{State model}} \times \overbrace{[\omega][\eta][\kappa]}^{\text{Priors}}.$$

JAGS code for this model with uniform priors over ω, η, and κ is:

```
model {
omega ~ dbeta(1, 1)
p_encounter ~ dbeta(1, 1)
p_capture ~ dbeta(1, 1)

for (i in 1:M) {
  z[i] ~ dbern(omega)
  for (k in 1:K) {
    y[i, k] ~ dbern(z[i] * p_encounter * p_capture)
  }
}
```

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

```
N <- sum(z[])
for (k in 1:K) {
  a[k] ~ dbinom(p_encounter, N)
}
}
```

Baseline model with failed captures and auxiliary encounters

Combining failed captures and auxiliary encounters includes the N-mixture model likelihood for auxiliary encounters in the failed capture model. As described in the main text, the posterior is:

$$\left[z_1, ..., z_M, y^*_{1,1}, ..., y^*_{M,K}, \omega, \eta, \kappa \mid y_{1,1}, ..., y_{M,K}, f_1, ..., f_K, a_1, ..., a_K\right] \propto$$
$$\overbrace{\prod_i \prod_k [y_{i,k} \mid y^*_{i,k}]}^{\text{Captures}} \times \overbrace{\prod_k [f_k \mid y^*_{1,k}, ..., y^*_{M,k}]}^{\text{Failed captures}} \times \overbrace{\prod_k [a_k \mid \sum_i z_i, \eta]}^{\text{Auxiliary encounters}} \times$$
$$\overbrace{\prod_i \prod_k [y^*_{i,k} \mid z_i, \eta, \kappa]}^{\text{Encounter model}} \times \overbrace{\prod_i [z_i \mid \omega]}^{\text{State model}} \times \overbrace{[\omega][\eta][\kappa]}^{\text{Priors}}.$$

JAGS code for this model with uniform priors over ω, η, and κ is:

```
model {
  omega ~ dbeta(1, 1)
  p_encounter ~ dbeta(1, 1)
  p_capture ~ dbeta(1, 1)
  p_encounter_and_capture = p_encounter * p_capture
  p_failed_capture = p_encounter * (1 - p_capture)
  p_not_encountered = 1 - p_encounter
```

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.22.915314; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

```
  for (i in 1:M) {
    z[i] ~ dbern(omega)
    pi[i, 1] = z[i] * p_not_encountered + 1 - z[i]
    pi[i, 2] = z[i] * p_failed_capture
    pi[i, 3] = z[i] * p_encounter_and_capture
    for (k in 1:K) {
      y_star[i, k] ~ dcat(pi[i, ])
      y[i, k] ~ dbern(y_star[i, k] == 3)
    }
  }

  # number of failed captures (soft constraint)
  for (k in 1:K) {
    failures[k] ~ dnorm(sum(y_star[, k] == 2), 4)
  }

  N <- sum(z[])
  for (k in 1:K) {
    a[k] ~ dbinom(p_encounter, N)
  }
}
```