EFFECTS OF HAIR-SNARE SUBSAMPLING ON SPATIALLY EXPLICIT CAPTURE–RECAPTURE POPULATION ESTIMATES OF BLACK BEARS

Nick Gondek[1], David L. Garshelis, Karen V. Noyce, and John R. Fieberg[1]

SUMMARY OF FINDINGS

Genetic mark-recapture studies estimate animal abundance using non-invasive DNA identification methods to "capture" and subsequently "recapture" individuals that leave genetic material at trap sites. Due to the cost of genotypic analysis, researchers often choose to process only a subsample of this genetic material. Traditional (non-spatial) mark-recapture estimators of abundance have been shown to be biased in this case, especially when individuals display a behavioral trap response following initial capture. Less is known about the impact of subsampling genetic mark-recapture data when using spatially explicit capture-recapture (SECR) models to estimate abundance. We are exploring the effect of subsampling on SECR estimators using hair-snare data obtained from a 2012 genetic mark-recapture study of black bears (*Ursus americanus)* from the Chippewa National Forest, north-central Minnesota. Non-proportional subsampling may be preferable to simple random sampling, despite the inherent violations of SECR assumptions that may result.

INTRODUCTION

Mark–recapture studies are routinely used by wildlife managers to estimate animal abundance. Especially in the case of endangered species and game animals, abundance and its associated temporal trends are of critical importance for making informed management decisions. Hair snares offer a minimally invasive technique for obtaining capture and recapture samples. Specifically, hair left at the trap can be genotyped to identify individuals. However, the number of samples left at barbed wire hair traps typically far exceeds the budget allotted for genetic analysis, in part because a single animal often leaves hair on multiple barbs as it passes in and out of the corral of wire. Thus, it is common practice to genotype only a subsample of hair, knowing that much of it is redundant. Subsampling has been shown to negatively bias density estimates in the context of Huggins mark-recapture models because trap-shy animals are inadequately represented (Augustine et al. 2014). Less is known regarding the impact of subsampling on spatially-explicit capture–recapture models (SECR).

A fundamental difference between SECR and non-spatial capture–recapture models relates to how the estimators make use of multiple captures of the same individual during the same time period (trapping session) at different traps. Whereas non-spatial models collapse these into a single capture event, SECR models use multiple captures within a session to inform parameters that quantify individual movement characteristics (Borchers 2012, Royle et al. 2013). Thus,

---

[1] Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul.

samples that are redundant in a non-spatial model may be of critical importance to SECR models. Researchers may limit the number of genetically analyzed hair samples from a single trap within a single session because multiple visits by the same individual, hours or days apart, are impossible to distinguish from a single visit with genetic sampling alone; hence, multiple samples from the same individual at the same site-session are not informative and not worth the cost. However, if sites are visited by multiple animals, and the data from these sites are subsampled non-randomly or non-proportionally, then the spatial distribution of *processed samples* may result in a biased estimator of the spatial distribution of *visits*; hence, SECR-based density estimators may also be biased. A number of studies have examined the effects of subsampling hair samples on non-spatial capture–recapture estimates of black bears (Tredick et al. 2007, Dreher et al. 2009, Laufenberg et al. 2013), but we are not aware of any that have investigated effects on SECR-based estimates.

OBJECTIVES

1. Compare abundance and density estimates and precision obtained from the 2012 genetic capture–mark–recapture (CMR) using different subsamples of data (i.e., derive the best estimates from the existing data).
2. Provide guidance for study design of future genetic CMR estimates of bears in Minnesota.

METHODS

Data Collection

We used data from a 2012 genetic mark-recapture study of American black bears (*Ursus americanus*) mainly within the Chippewa National Forest, northern Minnesota (Noyce and Garshelis 2013). We obtained bear hair samples from 121 stationary traps, spaced within grid cells of 1 mi^2 and checked 6 times at 10-day intervals (trapping sessions) during May–July. Traps were constructed of 2 strands of barbed wire encircling a suspended bait of bacon and scent lures. We collected bear hair from 2,784 barbs that occurred in 1,642 separate clusters of 1–11 adjacent barbs (considering both upper and lower strands of wire). We considered each cluster (not the individual barbs) a sampling unit, meaning that at most, we sampled only 1 barb from each cluster (although a single barb with hair, not adjacent to any other barb with hair, was also considered a cluster). Of these 1,642 clusters, 1,113 were sent to a genetics laboratory for genotypic analysis, and 1,019 samples were successfully linked to specific individuals.

Spatially-Explicit Capture–Recapture Estimators

Detection probabilities in spatial mark-recapture models are assumed to decrease as a function of distance between each trap and an individual's activity center, a latent variable in the model. There are a number of detection functions that can be used to model detection probabilities. We fit a half-normal detection function (the default), which in its most simple form has 2 parameters: g0, which determines the detection probability at the activity center, and σ, which controls how quickly detection probabilities decrease with distance from the activity center. Intuitively, σ will depend on how much individuals move and will thus be related to home range size.

Either of these parameters can, in turn, be modeled as a function of covariates (e.g., sex) or time (sampling session). There are 2 options for how parameters, and thus detection probabilities, may depend on time–parameters can vary linearly (on a transformed scale), denoted by T, or they may vary in an unstructured way, allowing each sampling session to vary independently, denoted by t. In addition, models can allow for a "behavioral effect," whereby parameters for recapture probabilities differ from those for initial capture probabilities. Models that allow parameters to change following an initial capture are denoted using the following syntax: b indicates a behavioral effect that applies to future capture probabilities at any site, and b_k indicates the behavioral effect only applies to future capture probabilities at the site (*k*) where the animal was previously captured.

We fit 4 SECR models to the full data set, each with a different combination of explanatory variables: g0 ~ 1 (i.e., capture probabilities only depend on the distance between a bear's activity center and the trap location), g0 ~ t, g0 ~ b_k, and g0 ~ b_k + t. For all models, we also assumed that σ varied by sex. Models were fit using the Program R (R Core Team 2015), package 'secr' for fitting SECR models and packages 'foreach' and 'doParallel' for optimization of model fitting using parallel processing (Revolution Analytics and Weston 2015a,b; Efford 2017).

Subsampling Methods

We are exploring 2 subsampling strategies: simple random sampling (SRS) and a subsampling method that gives preference to unique site-sessions, which we refer to as site-session preferred (SSP). With SRS, *n* samples were chosen at random from the set of hair clusters pooled across the different sites and trapping sessions. Alternatively, with SSP, we tried to maximize the number of unique site-sessions represented in the subsample. Let *m* represent the number of unique site-sessions with hair in the full dataset. If $m \le n$, we randomly choose *n* unique site-sessions, with 1 sample randomly selected from each of these site-sessions (in the survey sampling literature, this is referred to as a 2-stage cluster sample). When $n > m$, we chose 1 sample at random from each unique site-session and then took a second simple random subsample of size $n - m$ from the remaining clusters (from the pooled data) to give a total of *n* samples. This approach attempts to serve as a compromise between minimizing the amount of redundant data (accomplished by sampling an equal, or near equal, number of observations from each unique site-session) and maximizing the representativeness of the sample (as accomplished by SRS).

We considered 3 subsample sizes, *n* = 250, 550, and 850. For each subsample, we determined the number of unique combinations of (individual x site x session). We subtracted this number from *n* to determine the amount of redundant data in the subsample. We are in the process of developing R code that will allow us to fit the same 4 SECR models to each subsampled data set. This will allow us to compare density estimates from the full dataset to those obtained with various levels of subsampling.

RESULTS AND DISCUSSION

SECR Models Fit to Full Dataset

The best-fitting SECR model for the full hair-snare data set included a trap-specific behavior response and an unstructured time covariate, g0 ~ b_k + t. Estimates of recapture probabilities were substantially greater than original capture probabilities (Figure 1). Capture and recapture probabilities also varied considerably among the different sampling sessions and decreased more quickly with distance for females than for males (Figure 1). Using this model we estimated there were 12.4 bears per 100 mi^2 (95% CI = 9.06–16.87, Table 1).

Effects of Subsampling on Redundancy

We collected data from m = 377 unique site-sessions with hair. As the SSP subsampling strategy initially collects 1 sample from each unique site-session, subsamples with $n \le 377$ will not include any redundant data. The same is not true for SRS, which may include multiple samples from the same individual at the same trap, particularly for site-sessions with many clusters of hair. We found that, at low sample sizes, SSP selected far fewer redundant samples than SRS, and that this advantage diminished as sample size increased (Figure 2).

Future Work

Our simulations suggest that using a subsampling strategy that maximizes the number of unique site-sessions can reduce the likelihood of analyzing redundant samples, but more work is required to determine whether this benefit outweighs the inherent loss of movement information incurred using this method. In either case, the optimal strategy likely depends on the characteristics of the observed study population with respect to the spacing of the traps. Further simulation is needed before we can make general conclusions and recommendations regarding the effect of subsampling methodology on SECR estimates. In the future, we plan to fit SECR models to the subsampled datasets. In addition, we plan to simulate genetic mark-recapture data with varying degrees of behavioral responses to evaluate subsampling methods across a range of scenarios where the true population size is known.

ACKNOWLEDGMENTS

This project was funded by the Wildlife Restoration (Pittman-Robertson) Program grant W-68-D-15.

LITERATURE CITED

Augustine, B.C., C. A. Tredick, and S. J. Bonner. 2014. Accounting for behavioural response to capture when estimating population size from hair snare studies with missing data. Methods in Ecology and Evolution 5:1154–1161.

Borchers, D. 2012. A non-technical overview of spatially explicit capture–recapture models. Journal of Ornithology 152 (Supplement): S435–S444.

Dreher, B. P., G. J. M. Rosa, P. M. Lukacs, K. T. Scribner, and S. R. Winterstein. 2009. Subsampling hair samples affects accuracy and precision of DNA-based population estimates. Journal of Wildlife Management 73:1184–1188.

Efford, M. 2017. Secr: Spatially explicit capture-recapture models. R package version 3.0.1. https://CRAN.R-project.org/package-secr. Accessed June 12, 2017.

Laufenberg, J.S., F. T. van Manen, and J. D. Clark. 2013. Effects of sampling conditions on DNA-based estimates of American black bear abundance. Journal of Wildlife Management 77:1010–1020.

Noyce, K.N., and D.L. Garshelis. 2013. Capture heterogeneity in hair-trapping of bears. Pages 71–85 in L. Cornicelli, M. Carstensen, M. Grund, M.A. Larson, and J.S. Lawrence, editors Summaries of wildlife research findings 2013. Minnesota Department of Natural Resources, St. Paul, MN.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Revolution Analytics and S. Weston. 2015a. foreach: Provides Foreach Looping Construct for R. R package version 1.4.3. https://CRAN.R-project.org/package=foreach. Accessed June 12, 2017.

Revolution Analytics and S. Weston. 2015b. doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.10. https://CRAN.R-project.org/package=doParallel. Accessed June 12, 2017.

Royle, J., R. Chandler, R. Sollmann, and B. Gardner. 2013. Spatial capture-recapture. Academic Press.

Tredick, C.A., M.R. Vaughan, D.F. Stauffer, S.L. Simek, and T. Eason. 2007. Sub-sampling genetic data to estimate black bear population size: a case study. Ursus 18:179–188.

Table 1. Density estimates and Aikaike Information Criterion (AICc) scores associated with spatially-explicit capture–recapture models fit to data from 1,019 genetically-identified hair samples from black bears that visited hair-snares in Minnesota, May–July, 2012. ΔAICc represents the difference from the lowest scoring model and the compared model. All models were fit using Program R and package secr. For capture probabilities, notation 'b_k' represents trap-specific behavior,'*t*' represents a non-linear time effect, and 1 indicates an intercept-only model (i.e., detection depends only on distance from the animal's activity center to the trap). All 4 models assumed the scale parameter varied by sex, σ ~ sex.

Model	AICc	ΔAICc	Bears/100 mi^2 (95% CI)
$g0 \sim b_k + t$	3082	0	12.36 (9.06–16.87)
$g0 \sim b_k$	3129	47	12.31 (9.02–16.80)
$g0 \sim t$	3507	425	12.54 (9.22–17.08)
$g0 \sim 1$	3570	488	12.54 (9.21–17.07)

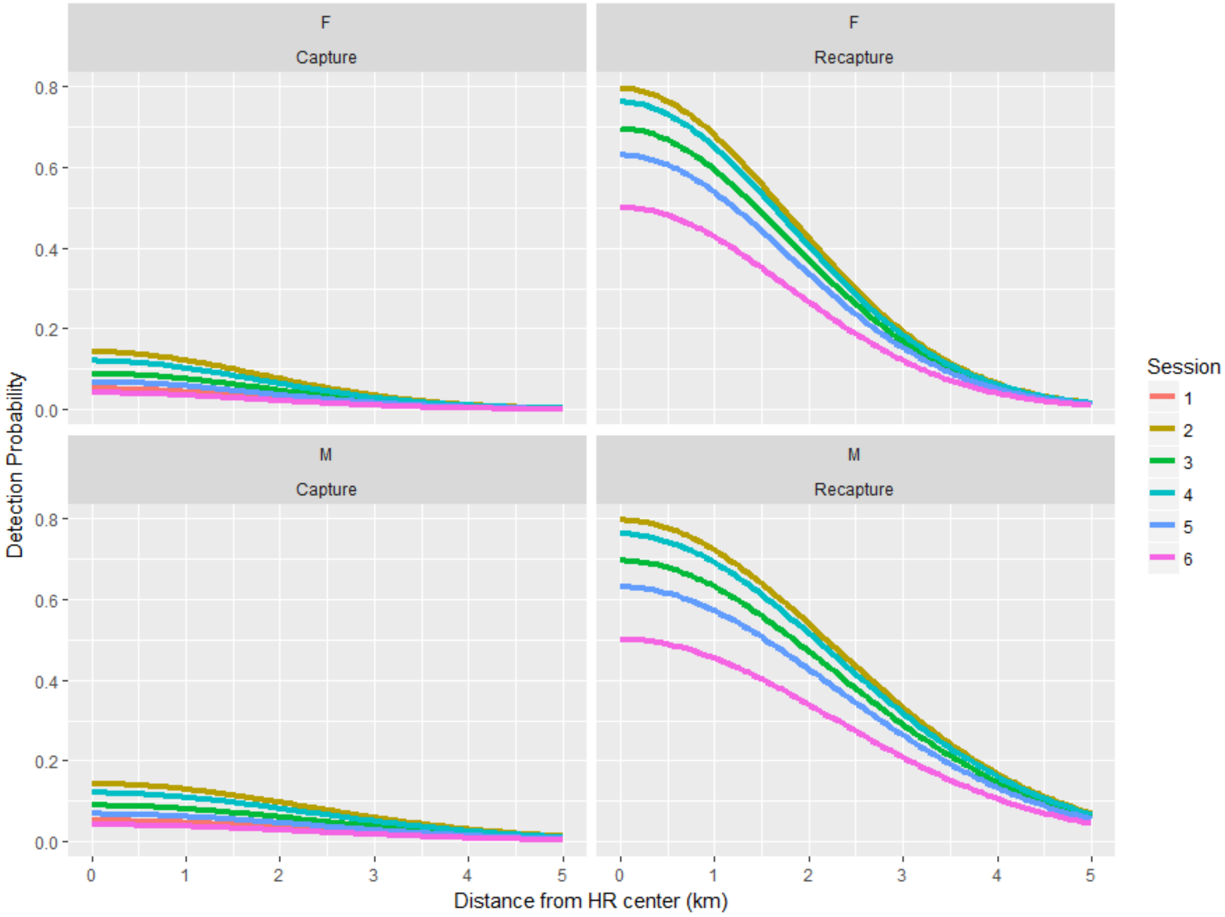

Figure 1. Estimated capture and recapture probabilities for male (M) and female (F) bears modeled as a function of the distance between a bear's estimated home range (HR) center (based on spatial distribution of recapture data) and a given trap. Estimates were obtained using model g0 ~ b_k + t, σ ~ sex. In all graphs, the 6 lines represent, from top to bottom, periods 2, 4, 3, 5, 1 (no recaptures for 1st period), and 6. Data are from a bear hair-snaring study in the Chippewa National Forest, Minnesota, in 2012.

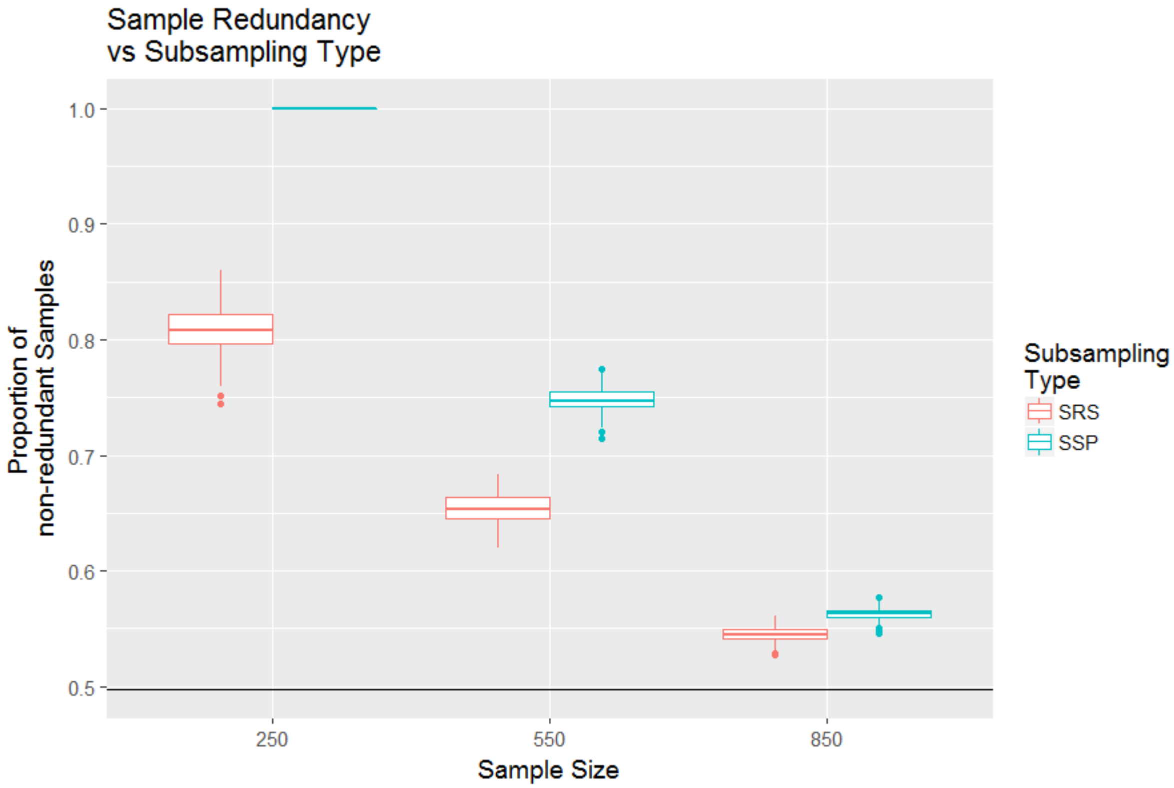

Figure 2. Proportion of non-redundant samples (i.e., different site-sessions) versus sample size when using simple random sampling (SRS) or site-session preferred sampling (SSP). Horizontal line near 0.5 represents the proportion of non-redundant samples in the full data set of genetically analyzed samples (which itself was already a subsample of 1,642 clusters of barbs with hair). When sample size is smaller than number of traps on the trapping grid, all samples chosen using SPP will, by definition, be non-redundant. As *n* increases, the difference in sample redundancy between SRS and SPP diminishes, converging to 0 when the full data set is utilized (i.e., when *n* = 1019). Data are from a bear hair-snaring study in the Chippewa National Forest, Minnesota, in 2012.