**Too Much Hair? Subsampling for Cost Efficiency with Spatial Capture-Recapture Models**

Nick Gondek1

1Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota; email:gonde012@umn.edu

**ABSTRACT**

**INTRODUCTION**

Mark-recapture studies have long since established themselves as the most prolific field method for the estimation of animal abundance (bunch of papers about MR? maybe a review?). Especially in the case of endangered species and game animals, abundance and its associated temporal trends are of critical importance for informed management decisions (same thing?). One long-elusive aspect of the process of mark-recapture methodology has been the selection of an effective sample area, which is necessary for the purpose of large-scale extrapolation by way of density (Royle et al 2013). In the case of endangered species, a site-specific abundance may be sufficient or even appropriate, but in the case of game species, extrapolation to a regional scale is almost always desired, if not necessary (Royle et al 2013, Borchers et al 2002). As such, abundance estimates without associated reliable density estimates may be of limited use to managers.

The relatively new development of spatially-explicit capture recapture methods (SECR) represent a potential solution, by linking estimations of abundance to their associated study area in a more statistically rigourous way (Borchers 2010, Royle et al 2013). Where the estimation of density was previously derived in a post-hoc and somewhat arbitrary way, SECR estimates of abundance scale directly with a given sample area, potentially allowing for more standardization of estimates across space or time. A secondary and certainly benefical outcome of this unambiguous linkage is that individual’s probability of capture in a given trap is allowed to vary, fundamentally, by the indvidual’s distance to that trap (Borchers 2010, Royle et al 2013, Meredith ??). Given that individual heterogeneity of capture has been established as a potential issue in many capture-recpture estimates (Ebert et al 2010), including the animal’s physical proximity to the trap grid as a predictor for capture probability may reduce a significant biological and physical source of that heterogeneity (Royle et al 2013, probably need more because this is a bold claim).

Presented with these benefits, a manager or applied ecologist may wonder whether it is still appropriate to opt for a non-spatial capture-recapture model. The answer may lie in a constraint that is not biologically or statistical – the project’s budget. As previously alluded to, a crucial aspect of a SECR analysis involves treating capture probability as a function of an animal’s activity center, where this activity center is derived by averaging the locations of the animal’s capture history (Borchers 2010, Royle et al 2013, Meredith ??). Depending on the methodology of the field study, capturing (or detecting) an animal may be costly; here, we take the case of an emergent practice in wildlife and conservation biology, genetic mark-recapture, where individuals are identified by sequencing the genome of residual genetic material left at a trap site (Boulanger 2004, Buckworth and Territory 2012, Petit and Valiere 2006, Gervasi et al 2008).

The project analyzed in this study aimed to enumerate northern Minnesota black bears using hair samples obtained by hair-snaring (Garshelis 2013 (?) ). Though recent advancements in the field of genetics have drastically reduced the cost of identifying individuals using hair samples, this step is still likely to represent a large portion of the operating budget in a genetic mark-recapture study. This project in particular is unique in that 1019 samples from a single season were successfully genotyped, far surpassing comparable studies (Coster et al 2011, Boulanger 2004, Gervasi et al 2008, Boerson 2003). **In this study, our objective was to compare abundance and density estimates using (1) various subsampling strategies utilized by wildlife managers and researchers and (2) various subsample sizes that may be forced due to budget constraints. Using these results, we provide guidance for genetic mark-recapture estimates when budget constraints limit effective sample size.**

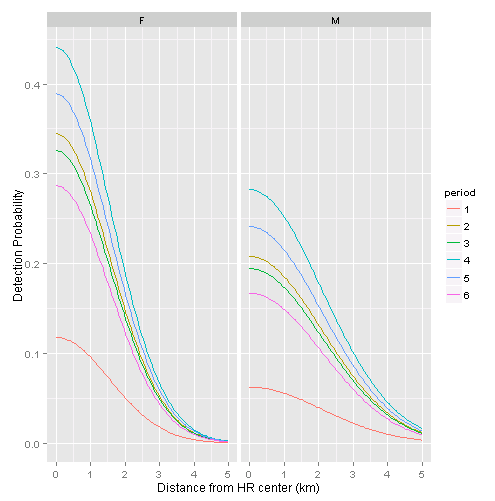
**METHODS**

**Model Structure**

**A SECR model is unique from other mark-recapture models in that an animal’s capture probability is derived using the animal’s homerange center (AC), but the parameters that characterize this relationship are derived in a similar way to other population models** (Royle et al 2013, Borchers et al 2002)**. Though many curves are used to characterize this relationship, a common and readily understood choice is a half-normal curve, using two parameters: g0 and σ. g0 represents the probability of detecting an animal whose homerange is centered exactly at the trap location, and is the intercept of the half-normal curve. It is important to note that g0 is not constrained at exactly one, which is the case in non-spatial distance sampling models, where detection is assumed to be perfect at distance zero (**Royle et al 2013).  **σ represents the rate at which this probability decreases as an animal’s homerange center moves further away from the trap, and is the distance along the x-axis of the half-normal curve where detection probability flattens and becomes less steep (Meredith ???).**

**Similar to other population models, these parameters are allowed to vary by sex (Sex), time (t), and behavioral trap response (b). In figure 1 below, g0 varied by time and sex (g0 ~ t + Sex) and σ varied by sex only (σ ~ Sex); as such, each time and sex combination has its own intercept, and both sexes have their own rate of capture probability depletion. Five different combinations of parameters were investigated in this study; these are covered in more detail in subsection “Model Fitting”.**

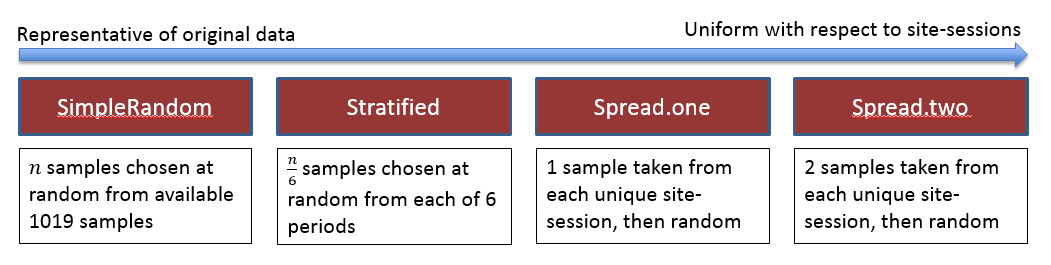
**Using these detection probability curves and a given sample area, the SECR model derives a population that scales appropriately with the sample area provided. In this study, both eventual density estimates as well as covariate values for parameters g0 and σ were saved and compared in order to deduce the influence of sample size and subsampling methodology on eventual population estimates.**



***Figure 1.*  Example half-normal detection curves for a SECR model, using all 1019 samples, where g0 varied by time and sex (g0 ~ t + Sex) and σ varied by sex only (σ ~ Sex).**

**Subsampling**

**The dataset analyzed in this experiment represents the 1019 successfully genotyped hair clusters of the 1642 clusters collected in Noyce and Garshelis 2013. In the effort to identify individuals using these samples, 83.24% of samples were submitted for analysis (1234 of 1482), and 82.58% of these submitted samples resulted in the successful identification of an individual (1019 of 1234). Only successful samples were included in subsampling in order to avoid confounding subsampling type with number of samples analyzed by the SECR model.**



***Figure 2.* Visual representation of subsampling types.**

**Figure 2 illustrates the four subsampling types chosen in this experiment. These were chosen in accordance with existing subsampling strategies used by wildlife managers (citation needed). The subsampling types are:**

* ***SimpleRandom* – n samples are chosen at random from the entire data set, without respect to period or site. This type of sampling is, on average, most representative of the original data set; for example, site-sessions that have a large number of samples would have a the largest number of samples in the subsample, and site-sessions with only one sample are unlikely to be chosen.**
* ***Stratified* – n/6 samples are chosen at random from each of 6 sessions, without respect to site. It is important to note that this differs from a true stratified random sample, where the number of samples chosen from each session would be weighted by the number of samples in each session relative to the number of samples overall.**
* ***Spread.one* ­– one sample is chosen from unique site-session. After this, samples are chosen randomly until n samples are aggregated.**
* ***Spread.two* ­– two samples are chosen from each unique site-session. After this, samples are chosen randomly until n samples are aggregated. This method is least representative of the original data, and most uniform with respect to site-session; for example, each unique site-session with only one or two samples have both of those samples chosen, and site-sessions with large amounts of samples are under-represented relative to the original dataset.**

**Model Fitting**

**After subsampling, a SECR model was fit with 5 varying covariate combinations. These parameters were first chosen by AIC selection with respect to the original 1019 sample dataset, then by increasing parsimony. Each model fitting assumed δ varied only by sex (insert explanation/source – because bears are known to vary their movement by sex, and sigma depends heavily on the movement patterns/distances). The five models chosen can be found in Table 1.**

|  |  |  |  |
| --- | --- | --- | --- |
| ****Covariates**** | ****AICc**** | ****∆AICc**** | ****Density Estimate *(95% CI)***** |
|  | **3492** | **0** | **13.50 *(9.86, 18.48)*** |
|  | **3496** | **4** | **12.67 *(9.30, 17.26)*** |
|  | **3502** | **10** | **13.60 *(9.93, 19.64)*** |
|  | **3507** | **15** | **12.55 *(9.22, 17.08)*** |
|  | **3515** | **23** | **14.23 *(10.41, 19.46)*** |

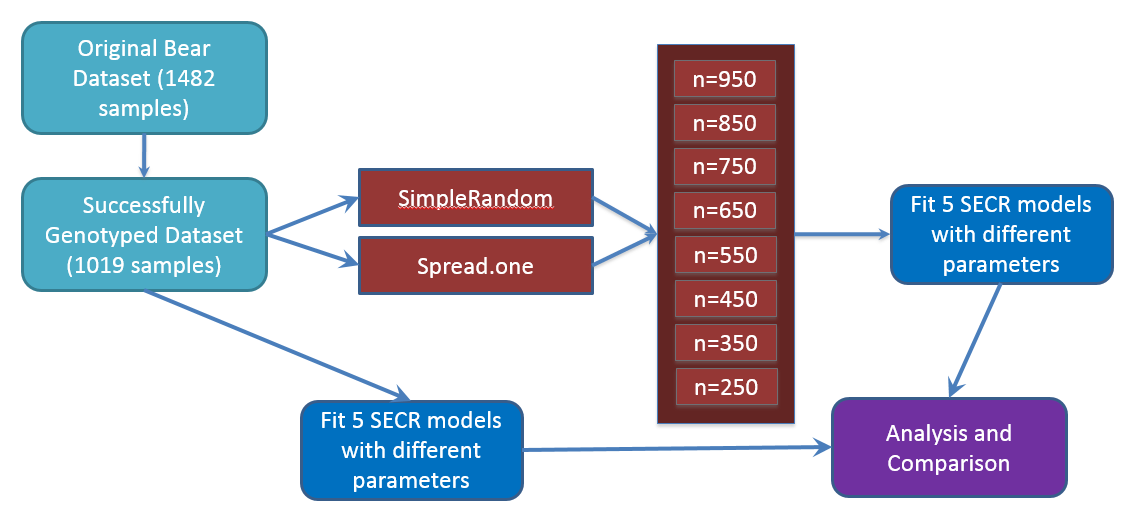
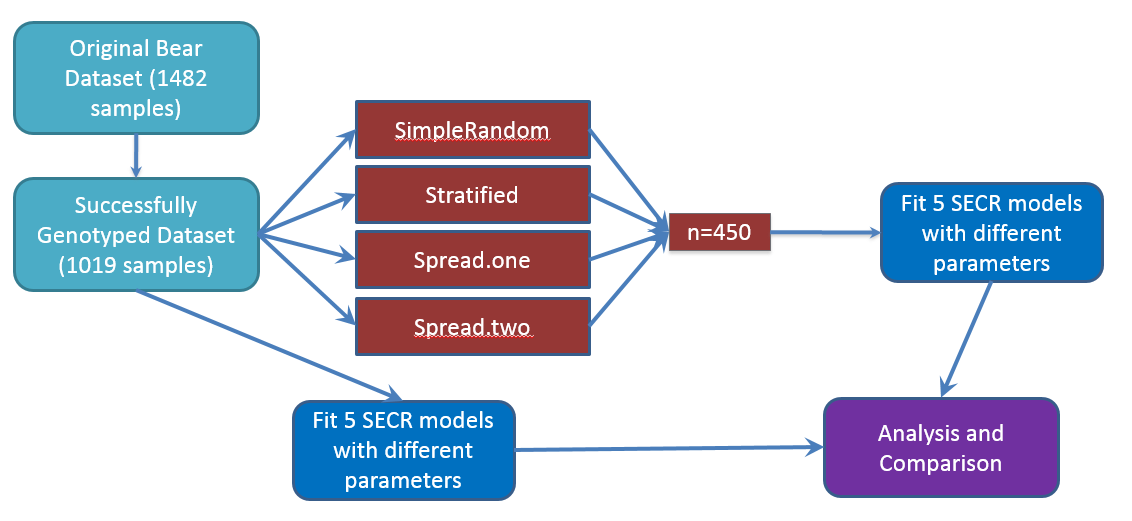
***Table 1.* Models chosen with accompanying AICc scores, ∆AICc,**

**Models were fit using the R programming language (R Core Team, 2014), package ‘secr’ for the fitting of the SECR models, and packages ‘foreach’ and ‘doParallel’ for optimization of model fitting using parallel processing (Site murray and whoever made foreach). Using function *secr.fit* in package ‘secr’, subsamples in the transformed from of a capture history were accompanied with a trap grid (referred to as a ‘mask’, and used to define an effective state-space) to arrive at a derived density estimate and a covariate matrix, both of which were used for analysis (**Royle et al 2013).

**Simulation**

**In this study, a single ‘simulation’ can be broadly defined as the taking of a subsample according to a single subsampling strategy (Fig. 2), the fitting of five SECR models with varying covariates (Table 1), and the saving of model output for later comparison. In the initial stage of this experiment, subsampling strategy was manipulated while holding sample size constant at 450, in order to infer the influence of subsampling strategy on the derived density estimate of each SECR model (Fig. 3); these simulations are referred to as ‘subsampling trials’ in this document as well as in the accompanying code documentation. In these trials, a simulation occurred stepwise using each subsampling strategy perpetually for six weeks, using a (SONY VAIO SPECS).**

**Using the resultant density estimates from the subsampling trials, two subsampling types (SimpleRandom and Spread.one) were chosen for ‘size trials’, where each simulation varied by both subsampling strategy *and* sample size (Fig. 3). Due to computational constraints, all four subsampling strategies could not be analyzed in these size trials. In these trials, a simulation occurred stepwise using each subsampling strategy and each sample size (250-950, by 100) perpetually for three weeks.**



***Figure 3.* Methodology flowchart for ‘subsampling trials’ (above) and ‘size trials’ (below). Comparisons of density estimates from subsampling trials led to the subsampling strategy selection in the size trials.**

**RESULTS**

**DISCUSSION**

**ACKNOWLEDGEMENTS**

**REFERENCES**