**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**

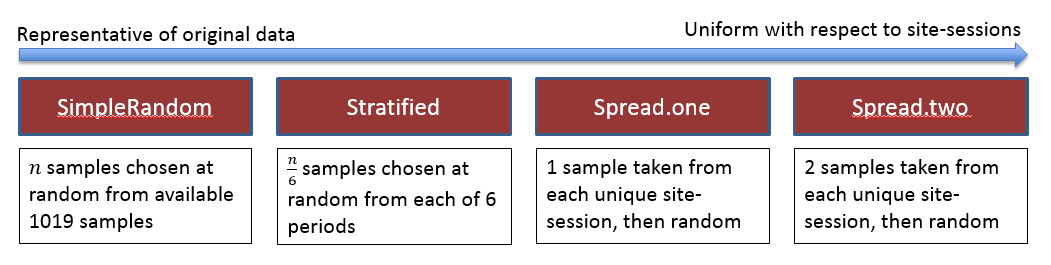
**INTRODUCTION**

**METHODS**

**Model Structure**

**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)*** |

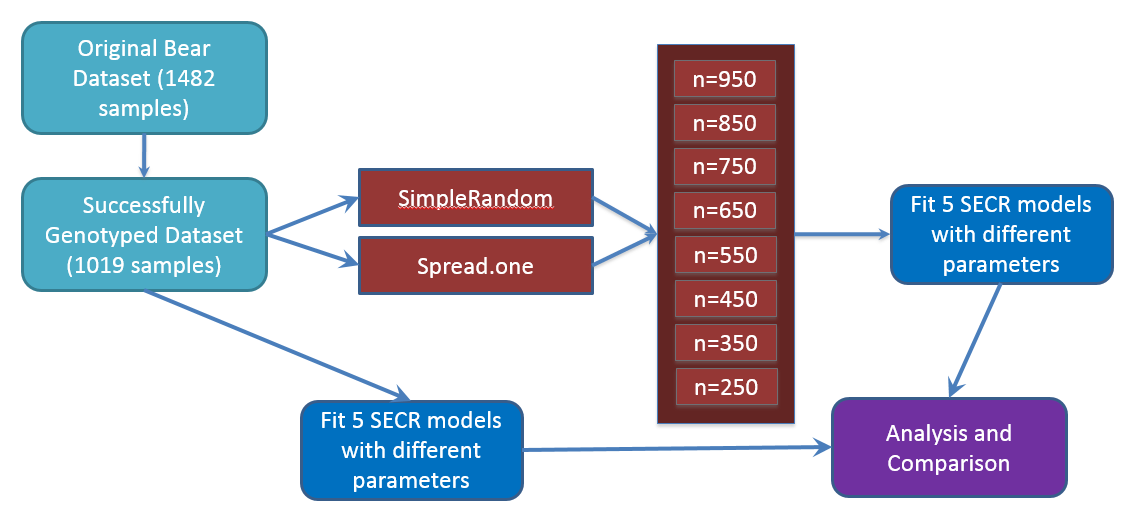
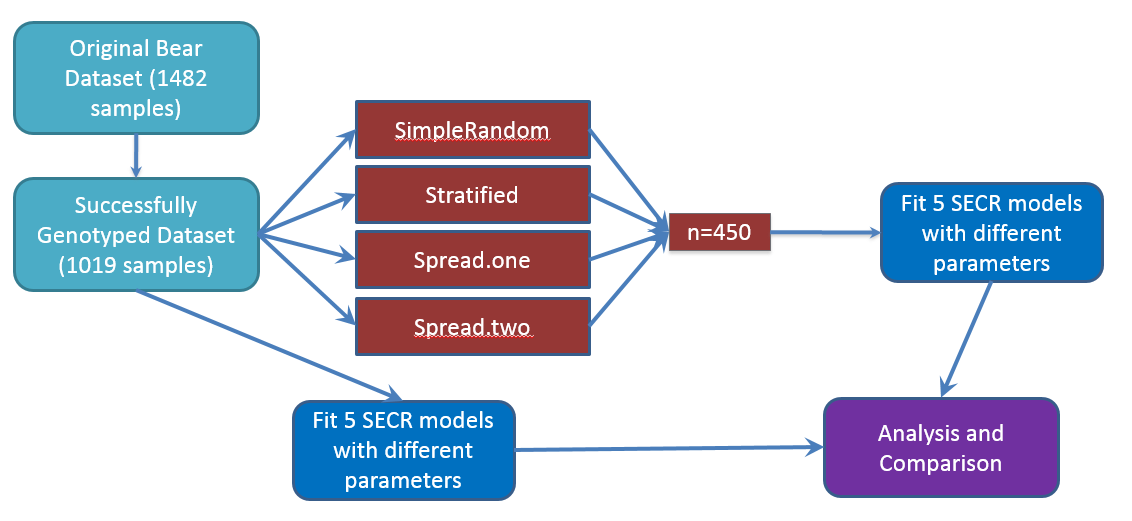
***Table 1.* Models chosen with accompanying AICc scores, ∆AICc,**

**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.**

**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**