October 5, 2016

David Garshelis

Forest Wildlife Populations and Research Group

Minnesota Department of Natural Resources

1201 E Hwy 2

Grand Rapids, MN 55744

RE: Report on Genetic-based estimates of bear populations: maximizing bang for the buck

**Problem statement:** Genetic-based mark–recapture population estimates from hair-snaring are now routinely used for bears. The number of samples obtained typically far exceeds the budget allotted for genetic analysis, and moreover, many samples are likely to be redundant (same visit by same bear). No guidance presently exists for selecting which and how many samples to genotype. Sample selection is expected to significantly affect spatially-explicit capture-recapture (SECR) models, which rely on the spatial distribution of capture samples.

**Objectives of this project:**

1. Compare abundance and density estimates and precision obtained from the 2012 genetic CMR using different subsamples of data — derive the best estimates from the existing data.
2. Assess effects of subsampling and behavioral characteristics of the bears on the bias and precision of genetic CMR estimates based on capture histories similar to that observed in 2012.
3. Provide guidance for study design of future genetic CMR estimates of bears in Minnesota.

**Obejctive 1: Subsampling 2012 genetic CMR data**

During the summer of 2015, I fit several different SECR models using all available data (roughly 70% of hair clusters were analyzed). Density estimates ranged from a low of 13 bears/100mi2 to a high of 15 bears/100mi2 (Table 1). The best-fit model, based on AICc, produced an estimate of 14 bears/100mi2 (95% CI = 10,18).

Table 1 Spatial Mark-Recapture Estimates of Population Density

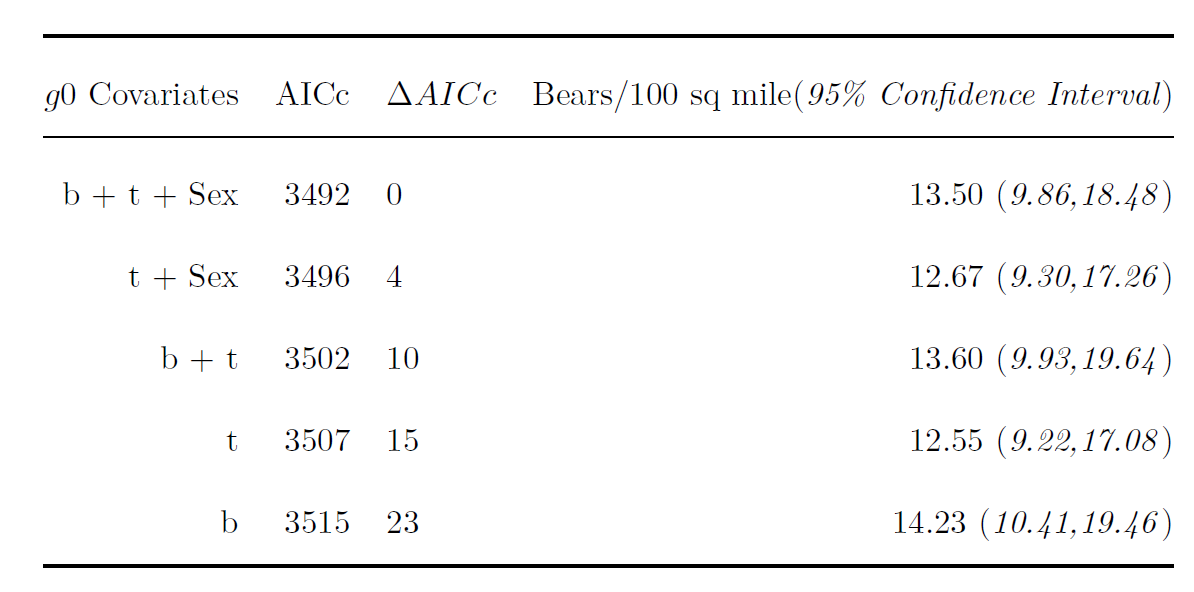
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Program | Density Estimate  (bears/100mi2) | 95% LCL | 95%UCL | AICc | deltaAIC |
| go~b+t+Sex, σ~Sex | secrAll2 | 14 | 10 | 18 | 3492 | 0 |
| go~t+Sex, σ~Sex | secrAll2 | 13 | 9 | 17 | 3496 | 4 |
| go~b+t, σ~Sex | secrAll1 | 14 | 10 | 19 | 3502 | 10 |
| go~t, σ~Sex | secrAll1 | 13 | 9 | 17 | 3507 | 15 |
| go~b+Sex, σ~Sex | secrAll2 | 14 | 10 | 19 | 3507 | 15 |
| go~b+T+Sex, σ~Sex | secrAll2 | 14 | 11 | 20 | 3507 | 15 |
| go~b, σ~Sex | secrAll1 | 14 | 10 | 19 | 3515 | 23 |
| go~b+T, σ~Sex | secrAll1 | 15 | 11 | 20 | 3516 | 24 |
| go~T+Sex, σ~Sex | secrAll2 | 13 | 9 | 17 | 3543 | 51 |
| go~T, σ~Sex | secrAll1 | 13 | 9 | 17 | 3554 | 62 |

During the 2015-2016 Academic year, I worked with an honor’s student to explore the effect of subsampling on SECR estimates of population abundance. We began by comparing SECR-based estimates of population size when using all available data to estimates obtained by subsampling *n* clusters of hair. We compared different subsampling strategies, but focused primarily on 2 subsampling strategies:

1. *Simple Random Sampling*: *n* samples were chosen at random from the entire set of hair clusters (pooled across the different sites and sampling periods). This type of sampling should, on average, result in data that are representative of the population of hair clusters.
2. A strategy that we refer to as *Spread.one*: to understand this strategy, define “site-sessions” as the combination of sampling site and sampling period (e.g., site 1 and sampling period 2).
   1. If *n* < number of unique site-sessions with at least 1 capture, we randomly selected *n* site-sessions containing hair data, and then randomly selected 1 hair cluster from each of these randomly selected site-sessions.
   2. If *n* > number of unique site-sessions with at least 1 capture, we randomly selected 1 hair cluster from each unique site-session containing hair. We then selected hair clusters randomly from the rest of the pooled data until a total of *n* samples are selected.

We then fit 5 different SECR models (Table 2) to the subsampled data across a range of sample sizes, *n*.

Table 2. Models fit to data subsampled from the 2012 CMR study.



The methods, results, and conclusions of this work are described in more detail in Nick Gondek’s honors thesis (attached). The main take away messages were:

* SECR density estimators, like their non-spatial counterparts, are biased low when applied to subsampled data. These effects were minimal, however, when using the S*pread.one* sampling strategy, provided that a behavioral effect was included in the model. This behavioral effect allows recapture rates to differ from initial capture rates. By contrast, S*imple Random Sampling* often resulted in density estimates that were considerably lower than the full data estimates.
* Non-random sampling increased the number of uniquely identified individuals, which outweighed any costs of violating SECR’s assumptions (as a result of non-random sampling of the hair data).

Based on this analysis, and the fact we were able to process close to 70% of the hair clusters collected during the 2012 CMR study, I would suggest using the density estimate from the best fitting model in Table 1 and Table 2 for management purposes: = 13.50 bears/mi2 (95% CI = 9.86, 18.48).

**Objective 2: Assess effects of subsampling and behavioral characteristics of the bears on the bias and precision of genetic CMR estimates based on capture histories similar to that observed in 2012.**

This summer, Nick and I have been using simulations where the population size and density of bears is known to determine if the patterns observed when subsampling the real data hold true more generally. A short description of simulation scenarios and preliminary results are given below.

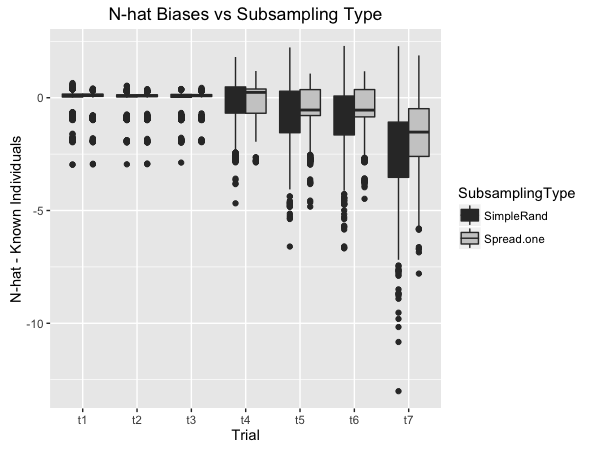
Using a simulation framework, we created seven scenarios with various behavioral effects and population demography:

|  |  |  |
| --- | --- | --- |
| **Trial** | **Active Effects** | **Description** |
| t1 | none | No behavioral effects. Likelihood of capture for each bear at any trap determined solely by distance from that bear’s activity center to that trap. |
| t2 | trap-shyness | Bears are less likely to be subsequently captured after initial capture. |
| t3 | individual heterogeneity | Certain bears are more likely to be captured in general, and some are less likely. These effects are constant throughout each simulation. |
| t4 | redundancy | After a bear is captured, that bear may leave more than one sample at a trap (ie, the bear may leave redundant samples) |
| t5 | individual heterogeneity + redundancy | Certain bears are more likely to be captured in general, and those bears are also more likely to leave redundant data. The converse is true for bears unlikely to be captured in general. |
| t6 | trap-shyness + individual heterogeneity + redundancy | Certain bears are more likely to be captured in general, and those bears are also more likely to leave redundant data. The converse is true for bears unlikely to be captured in general. All bears are less likely to be captured after initial capture. |
| t7 | trap-shyness + individual heterogeneity + redundancy + uneven density | ~75% of bears occupy 50% of trapping grid. Certain bears are more likely to be captured in general, and those bears are also more likely to leave redundant data. The converse is true for bears unlikely to be captured in general. All bears are less likely to be captured after initial capture. |

All but one scenario has been completed. The remaining scenario is identical to t1, with uneven bear density and no behavioral effects.

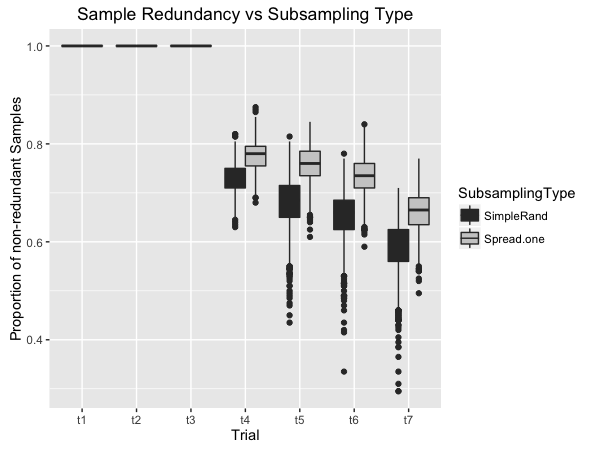
Each scenario contains 30 bears in the area of the trapping grid. The trapping grid consists of 36 traps, arranged systematically in a square, separated by 800 meters. At this point, only one secr model has been fit to the ~1200 simulated datasets of each scenario, where intercept capture probability is influenced by whether a bear has been captured in a previous session. We plan to fit two more models – one with a behavioral and time trend influencing intercept capture probability, and one with a time trend only.

We compared the derived number of individuals, as well as the proportion of non-redundant data analyzed, from each fitted model with respect to the subsampling type. Our results indicate that the spread.one subsampling strategy produces density estimates that were closer, on average, to the true number of individuals in the population (figure 1).



*Figure 1: Mean biases of n-hat vs subsampling method on secr models fitted to ~8400 simulated capture histories.*

We hypothesized, based on spread.one’s superior performance analyzing empirical data, that these differences in n-hat were caused by the proportion of non-redundant data in any given subsample; specifically, that spread.one resulted in less redundant data being included for analysis. A comparison of sample redundancy with respect to subsampling type in our simulations supports this notion (figure 2). On average, spread.one selected more non-redundant data in each trial that included sample redundancy.



*Figure 2: Mean proportion of non-redundant data analyzed with respect to subsampling type for each of the 7 scenarios. t1, t2 and t3 cannot produce redundant data by design.*

**Objective 3: Provide guidance for study design of future genetic CMR estimates of bears in Minnesota.**

Based on the two different simulation studies we have conducted (subsampling hair data from the 2012 CMR study and simulating capture histories), we conclude that the *spread.one* strategy provides a cost-effective approach to estimating abundance when the number of samples obtained far exceeds the budget allotted for genetic analysis.

We are committed to working with David Garshelis and Karen Noyce over the coming year to summarize these results in a peer-reviewed manuscript.