Appendix A: Optimization of a stratified random design

The SamplingStrata R package (Barcaroli (2014)) was used to optimize the placement of strata boundaries and allocation of effort across strata subject to user-inputted pre-specified precision targets for each taxon. A genetic algorithm was used to conduct a partial search for the optimal placement of strata created using two stratum variables: latitude and distance to shore. These stratum variables were specified for each unit in the sampling domain and generally characterize the diverse set of distributions exhibited across the species set. A candidate set of solutions (i.e., stratifications) are initialized and each solution is given a score, the optimal total effort allocated across strata which is calculated by employing the Bethel algorithm (Bethel (1989)). The Bethel algorithm calculates the optimal effort allocated across strata subject to user-inputted precision constraints for each taxon (see below for how these constraints were initialized and tuned). The lower the score for a candidate solution, the more “fit” the solution is. A new set of candidate solutions are produced from the current set of solutions, employing aspects of natural selection like elitism, crossover, and mutations to create a new set of solutions that reflect the fitter solutions in the candidate set. Over many iterations, the average fitness of the solution set increases (i.e., stratifications are more efficient) and asymptotes, at which point the partial search is terminated. The fittest (lowest sample size) solution is chosen from that terminal solution set is the optimized solution.

The task of setting the user-inputted precision constraints are set across taxa is a nontrivial one. We developed a three-step process to tune the precision constraints to a particular level of total survey effort ranging from roughly 50-175 stations.

1. Initial stratification solution: the optimization was first run with initial CV constraints set to the SRS CVs across the species set assuming 200 stations. The algorithm settings included: 100 candidate solutions, a 10% elitism rate, and a mutation rate inversely proportional to the total number of strata specified. The algorithm was run for 300 iterations and the best solution (i.e., lowest sample size) solution was chosen as the optimal solution.
2. Single-species solutions: Given this initial solution, the Bethel algorithm was conducted to calculate the lowest CV for the range of total stations of interest for each species separately. This is done by iteratively conducting the Bethel algorithm, modifying the CV constraint until the sample sizes of interest are calculated. This serves two purposes. First is that by evaluating how stations are allocated for each species individually, we can visualize the tradeoffs in survey design across the species set. Second, these CVs can be used as lower limits in the multispecies optimization because we assume that optimizing the bethel algorithm with only one species would provide the lowest CV given a stratification. Thus, taken together the single-species optimized CV and the SRS CVs for a given sample size is the range of CVs that will constrain the multispecies optimization.
3. Multispecies solutions: Given the initial stratification, the Bethel algorithm was iterated to calculate the lowest CV for the range of total stations of interests across the species set. This was done by modifying the CV constraint for each species by a given rate (e.g., 0.001) scaled by the difference between the current CV constraint and the single species CV (if aiming to increase the current total sample size) or the SRS CV (if aiming to decrease the total sample size). The advantage of this approach is that the multispecies optimal CV will be below the SRS CV which we expect due to the nature of stratification (Cochran (1977)) but above the single species optimal CV due to the inherent interspecific tradeoffs in optimizing survey design.

# References

Barcaroli, Giulio. 2014. “SamplingStrata: An R Package for the Optimization of Stratified Sampling.” *Journal of Statistical Software* 61 (4): 1–24. <https://doi.org/10.18637/jss.v061.i04>.

Bethel, JAMES. 1989. “Sample Allocation in Multivariate Surveys.” *Survey Methodology* 15 (1): 47–57.

Cochran, William G. 1977. *Sampling Techniques*. John Wiley & Sons.