**Introduction**

One of the fundamental objectives of the Pallid Sturgeon Population Assessment Project (PSPAP) is to quantify pallid sturgeon *Scaphirhynchus albus* trend and abundance. Achieving this objective warrants the identification of abundance and trend estimators that will provide optimal estimates given budget constraints. Hence, it is pertinent to compare metrics of estimator success and associated sampling costs across multiple estimators and sampling strategies while accounting for uncertainties. The following sections outline an approach for such a comparison: first discussing methods for evaluating metrics of estimator success, followed by a discussion of sampling cost, and concluding with methods for combining metrics into a comprehensive assessment of estimator-sampling design utility in the face of uncertainty.

**Evaluating Estimator Metrics**

Bias, precision, and performance metrics are used to evaluate estimator success. The following approach allows the calculation of all three metrics for various abundance and trend estimators. Additionally, it provides opportunities for evaluation of other estimates (e.g. length structure) of interest to stakeholders. In short, this approach has four steps: (1) Simulate a known reference PS population, (2) Simulate sampling catch data, (3) Calculate estimates from the catch data, (4) Compare the estimates to the reference population.

The pallid sturgeon reference population is initialized using data from the PSPAP database and the pallid sturgeon literature. Each river bend is populated with pallids based on expected segment-level densities, while each individual fish is assigned an initial length and von Bertalanffy growth parameters and from segment-level and basin-level distributions, respectively. Individual fish are then tracked for 10 years, recording individual survival status, bend location, and length on a yearly basis. Survival is binomially distributed with fixed parameter , and growth is projected by individualized von Bertalanffy growth curves. Within basin movement is based on a pallid’s current bend location with the probability of being in a particular bend the following year increasing as distance to that bend decreases. Recruitment occurs randomly with a fixed expected frequency (e.g. once every 3 years), and the number of recruits, given there is recruitment, is drawn from a basin dependent Poisson distribution. Each new recruit is tracked after being randomly assigned an age-0 location within basin, an age-0 length of 200mm, and von Bertalanffy growth parameters.

For each of the simulated reference populations, various sampling designs can be implemented to obtain simulated catch data. All sampling designs include segments 2-4, 7-10, 13, and 14, and at minimum, the number of bends sampled within a segment matches those listed in Table A1 of Welker et al. (2017). The way bends are chosen, however, may vary. In a random sampling design bends within segment are chosen randomly each year, while in a fixed sampling design they are chosen randomly once and then fixed to be sampled each of the following years. A combination of fixed and random designs can also be implemented. Sampling strategies can also differ in the number of sampling occasions (times within a year that each bend is sampled), as well as the number of gear deployments within a particular occasion. Additionally, each sampling design can be implemented with different types of gears: gill nets, trammel nets, otter trawls, trotlines, or a combination of these.

Once a sampling design is chosen, 10 years of occasion-level catch data is simulated for each of the selected bends. Each fish within a selected bend has a probability, , of being captured. This occasion-level capture probability varies from occasion to occasion, as it is calculated from the individualized deployment catchability and effort values. For each deployment, effort values are drawn from a gear and basin specific gamma distribution, which was fit to PSPAP effort data. Deployment catchability , or the probability of catching a single fish with one unit of effort, is drawn from a gear specific distribution. Deployment specific capture probabilities, , are calculated as and then aggregated to the occasion level to obtain . When a fish is successfully caught, fish id and length, location (bend) and timing (occasion within year) of catch, and gear used are recorded, simulating a complete capture history at the bend level for the given sampling design.

Several catch data simulations are made per reference population. All estimators are applied to each of the catch data simulations except where sampling design limits the use of an estimator, forcing the application of a smaller subset. Abundance estimates are first computed on the bend-level using various estimators: closed population and , Cormack-Jolly-Seber, robust design, and for single occasion estimates, a count of minimum known alive. Bend-level abundance estimates are then aggregated to the segment-level by converting bend abundance to bend density, estimating segment density as the average bend density within a segment, and converting segment density to abundance. Segment-level standard errors are calculated from estimator reported bend-level standard errors and the delta method. Trend estimates are computed as the slope of the linear model , where is either segment-level abundance or segment-level catch per unit effort. Additionally, estimates of other population attributes of interest are calculated, e.g. mean fish length.

At this point bias, precision, and performance can be computed for each estimator-sampling design combination. The bias of an estimator used with a particular sampling design is computed as the expected value of the difference between the estimated value and the actual value, where the expectation is taken over all estimates made by the given estimator on all catch data simulated under the given sampling design. Note, actual values are known since they are reported in or can be derived from the reference population data. Precision is calculated as the expected ratio of the standard error to the absolute value of the estimate (more precise estimators have lower values of precision). Since some catch data (e.g. sparse data) will lead to errors in certain abundance estimator calculations (e.g. non-convergence), a measure of estimator performance is calculated. The performance metric is a measure of how often the collected data leads to estimator errors relative to the amount of data collected.

**Sampling Costs**

The cost for each sampling design is evaluated taking several factors into account. While overhead costs are similar for all sampling designs, costs will vary among sampling designs with the number of sampling occasions within a year, the number of deployments per sampling occasion, sampling effort of each deployment, which gears are used, how much training is required for a particular gear, etc. Costs may also vary with estimators or what population characteristics are estimated, especially if there is a need to hire someone with a particular set of advanced skills. For example, if an estimate requires samples to be sent out for analysis there is an additional cost to the collection of the samples.

**Overall Utility and Assessing Uncertainty**

Each estimator-sampling design combination is associated with a measure of bias, precision, and performance. In order to understand the overall usefulness of an estimator-sampling design pair, it is essential to place each metric on a comparable scale and rank their importance. To this end, we compute metric utilities, which scale metric values to a common range: 0-1, with values scoring the closest to 1 giving the highest utility. This process leads to three utility values (a bias, precision, and performance utility) for each estimator-sampling design pair. The overall utility of the pair is evaluated as the weighted mean of its bias, precision, and performance utilities, where the weights are determined by the importance of each metric as established by stakeholders. Similarly, the overall utility of a suite of estimators (one abundance estimator, one trend estimator, etc.) used under the same sampling design is the weighted mean of the utilities for each estimator. For example, if stakeholders decide that abundance and trend estimates are equally important, then the utility, , of a particular sampling-estimation design would be calculated as , where and are the utilities of the chosen abundance estimator and the chosen trend estimator under the particular sampling design, respectively.

If the cost of any sampling-estimation design is greater than that allotted for in the budget, then it is removed from further analysis (essentially its utility becomes zero since it is not monetarily feasible to implement such a design). The remaining designs can be compared using relative costs, which can be scaled from 0 to 1. Utility and cost can now be weighted and combined into a single measure of a sampling-estimation design.

Bayesian nets are then used to compare different sampling design-estimator suites in the face of uncertainty. In the above stochastic simulations, many parameter distributions were derived from the PSPAP database or the pallid sturgeon literature. However, mean values and standard deviations for movement probabilities, recruitment frequency, number of recruits, and catchability are highly uncertain and not reliably available. To account for this, reference populations and catch data are simulated for a wide range of mean values for these parameters. For example, although all movement probabilities are a function of distance, we vary movement parameters so that some reference populations have high site fidelity while others have low site fidelity.

Using Netica and inputs from estimator results (ran in R), the expected utility of multiple sampling design-estimator suites can easily be calculated and compared under specific conditions. In particular, a comparison of various ranges of uncertain parameters can be made, leading to valuable information. For example, it is possible that for all ranges of catchability the robust design has the most utility, and hence management decisions can be made more confidently without knowing more precise values of catchability. However, if Strategy A is better given high catchability and Strategy B is better given low catchability, then we’ve discovered that it is important to learn more about catchability to make an informed decision. Furthermore, weighting uncertain situations by their expected occurrence, allows for the choice of an optimal strategy selection under the given assumptions and constraints, based on our current knowledge.

**References**

Welker, T.L., M.R. Drobish, and G.A. Williams (editors). 2017. Pallid Sturgeon Population Assessment

Project, Guiding Document, Volume 1.8. U.S. Army Corps of Engineers, Omaha District,

Yankton, SD.