**Supplementary Materials**

## Decision Nodes

### Survey Design

The choice of survey design determines how bends are selected each year for sampling. Two survey design choices were considered:

1. Random-yearly and protocols,and
2. Fixed-bends within segment are chosen randomly in the first year of the study and then are sampled annually.

The fixed survey design was included because it was required to estimate bend-level values using a mulit-state robust design model which in turn may affect segment-level estimates.

Both survey designs used river bend within segment as the basic sampling unit for the Pallid Sturgeon Population Assessment Program (PSPAP; Welker and others, 2017). In defining river segments and bends, we use a general hierarchical framework for river classification originally published by Frissell and others, (1986). ~~Segments were defined where major tributaries enter the mainstem or other independent geologic or engineering structures control the nature of the river’s flow regime or morphology. In the next level of the spatial hierarchy, bends are identified as geomorphic units with repeatable sequences of macro-habitats, typically cross-over – bend – cross-over, equivalent to riffle-pool-riffle sequences in smaller rivers.~~

### Number of Secondary Occasions

bend-level to capture and reacapture fish and s estimation of. Secondary occasions occur over a short period of time (e.g., days), short enough that closure of the population from demographic processes (e.g., recruitment, mortality, immigration, emigration) can be assumed. However repeatedly sampling a bendssampling Additionally, t and. Either

1. 1
2. 2,
3. 3, or
4. 4 secondary sampling occasions per sampled bend (i.e., a crew repeatedly sampling a bend over 2, 3 or 4 days) can be chosen as a sturgeon-season sampling strategy.

Increasing the number of sampling occasions beyond 4 makes it more likely that assumptions of closure would not be met and were not considered in this analysis. Specifically, the logistics of a 5 day work week will result in sa weekends without sampling increases the time since the first and last sampling occasion and increases the liklihood that the closure assumption is violated.

### Gears

The gear decision node specifies potential basin-dependent gear choices used to capture pallid sturgeon in each bend. The three gear choices were evaluated in this analysis:

1. Trotlines:-annual sampling is completed each year by deploying trotlines in both the upper and lower basin,
2. Gill Nets (LB)/Trammel Nets (UB): annual sampling is completed each year by deploying gill nets in the lower basin and trammel nets in the upper basin, and
3. Combination: annual sampling is completed each year by deploying trotlines during some secondary sampling occassions within each sampled bend and either gill nets (lower basin) or trammel nets (lower basin) during other secondary sampling occasions within the same bend.

Gear choices for sturgeon-season sampling were selected based on basin-specific gear efficiencies, expert advice, and maintaining compatibility with legacy PSPAP data.

### Estimator

The estimator decision node consists of 8 choices for a trend and abundance estimator. Each of the 8 choices combines 1 of 2 approaches for estimating population trend with 1 of 7 approaches for estimating segment-level abundance (table 2).

Trend estimators included 2 approaches:

1. Catch-effort data (CPUE), as well as those that directly
2. estimated trend from estimated abundance data (abundance based).

Abundance estimators evaluated included

1. single-occasion catch data approaches (minimum known alive),
2. as well as closed mark-recapture methods (M0,
3. Mt,
4. multi-state robust design).

All basin-level abundance estimates were computed as the sum of its segment-level abundances given an abundance estimate was available for each segment within the basin. If any segment-level abundance estimate was missing, the weighted mean aggregation method was used to estimate the overall basin abundance. Each trend estimator, abundance estimator, and aggregation method are described in further detail below.

### Number of Recruitment Detection Trawls

The number of recruitment detection trawls sets the protocol for how many otter trawl (OT04) deployments should be utilized within each selected bend for detect age-0 recruits. Nine choices for the number of recruitment detection trawls were considered: 0, 2, 3, 4, 5-9, 10-19, 20-29, 30-39, and 40-49. With many uncertainties (e.g., recruitment level, interception locations, detection probability) associated with pallid sturgeon recruitment, it is important that decision makers have a wide range of choices when it comes to the number of trawls used in recruitment detection. Low detection of age-1 recruits in past sampling efforts suggests that a large number of trawls (e.g., 40-49) may be required to increase the chances that the monitoring program will detect a recruit. However, recruitment-detection trawling may not always be worth the effort, and in these situations decision makers can choose not to trawl for recruits (0 trawls).

## Nature Nodes

### Basin

The monitoring program is designed to assess pallid sturgeon populations in two functionally independent population zones of the Missouri River, typically referred to as the upper basin and the lower basin. Designation of “basin” is somewhat misleading as the assessment focuses on the mainstems and does not include all tributaries. In particular, we use the terms “upper basin” to describe the Upper Missouri River from Fort Peck Dam to the headwaters of Lake Sakakawea, including the Lower Yellowstone River up to the Intake Diversion Dam, and “lower basin” to describe the Lower Missouri River from Gavins Point Dam to the confluence with the Mississippi River (fig. 1).

The upper and lower basins differ in many aspects, including length and number of river bends. These differences affect the pallid sturgeon populations (e.g., differing drift distances available for free embryos) and the monitoring program (e.g., difference in gears that can be effectively used to capture pallid sturgeon). Therefore, monitoring program results may differ by basin. To adequately address basin-level differences, the catchability nodes (which rely on choice of gear) have been broken down by basin, and the “Basin” nature node (table 3) has been included as a direct affect on the outcomes of recruitment-detection sampling. As success of the monitoring program in both basins is crucial, the basin nature node has been parameterized as 50-50.

### Recruitment Detection

#### Recruitment Level

We considered basin recruitment levels, given recruitment occurred, that ranged from 1 recruit to 1000 recruits. The full range of recruitment was broken down into different levels: 1-30, 30-60, 60-100, 100-500, and 500-1000 recruits (table 4), with the range of each bin increasing as the number of recruits increased. As recruitment level is uknown we weighted each bin evenly. This more heavily weights there being a lower number of recruits. For example, less than 100 age-1 pallid sturgeon are recruited to the population 3/5 of the time.

#### Interception Location

The success of recruitment-detection sampling depends on where along the river drifting free embryos are intercepted and remain to develop. Four interception locations based on differing drift distance scenarios were considered: anywhere in the basin, lower two-thirds of the basin, lower one-third of the basin, and outside of the basin (table 5). The second and third scenarios are based on basin length in river kilometers, and the fourth scenario represents the drift of free embryos into Lake Sakakawea (upper basin) or the Mississippi River (lower basin). None of the nodes are assumed to affect interception location. As interception location is uncertain, we gave each scenario an equal likelihood of occurring. This weighting is approximately equivalent to assuming interception occurs in the upper third of the basin 1/12 of the time, the middle third of the basin with a probability of 5/24, the lower thrid of the basin with a probability of 11/24, and outside of the basin 1/4 of the time. As recent experiments with beads support longer drift distances of free embryos (citation), this initial parameterization allows for a higher likelihood that interception occurs in the lower third or outside of the basin, while still supporting the potential for interception in the upper two-thirds of the basin.

#### Detection Probability

The detection probability (table 6) of a single deployment, given the presence of a recruit, is an important variable in determining whether or not a recruit will be detected within the basin given a certain sampling strategy. All recruitment-detection sampling uses otter trawls (OT04) because of their effectiveness in capturing young pallid sturgeon. Gear choice, therefore, does not affect detection probability, which is assumed to be homogeneous among trawls within a given basin. Dectection of recruits is extremely rare, so there may be a high likelihood that detection probability is low. However, detection probabilities could be higher (up to 0.1) with the lack of recruitment detection due to a lack of spawning or survival to age-1. Considering these uncertainties and lack of information, we parameterized the detection probability nature node, which ranges from 0 to 0.1 and is descritized into 5 equal sized bins, with a uniform distribution.

#### Probability of Detecting a Recruit

The nature node “Probability of Detecting a Recruit” represents the probability that at least one recruit is captured during an entire year’s worth of sampling. This probability is dependent on basin, recruitment level, interception location, and detection probability. The node, which ranges from 0 to 1, was descritized into 10 equal levels (table 7). We parameterized the node using simulations that consisted of several steps:

1. Simulate the occupancy status of each bend within basin given basin, recruitment level, and interception location.
2. Simulate the recruitment detection status of all sampled bends within the basin given the detection probability of each trawl, the number of trawls, and the occupancy status of the bend.
3. Calculate the recruitment status of the basin at the end of the sampling year.
4. Repeat this process 500 times for each combination of basin, recruitment level, interception location, and detection probability, using the probability of success as the probability of detecting a recruit.

### Abundance and Trend Estimates

#### Movement

Movement between sampling years was modeled on two levels: “None” and “Little”. In the case of no movement all pallid sturgeon return to or remain in the same bend of the Missouri River, so that during any sampling occasion each living fish is available for capture in the same bend each year. In the case of little between year movment, all pallid sturgeon remain within the same bend during any given sampling year; however, there is a small probability that a surviving fish will be available for capture in a new bend the following sampling year. We quantified “Little” between year movment as having an average probability less than 0.25 of moving to a different bend within basin the following sampling year. Moreover, for “Little” between year movement no fish had a probability of moving that was greater than 0.5. Movement was restricted to within basin movment due to a high likelihood that the existing physical barriers (i.e., damns) prevent inter-basin migration. We parametrized the between year movement node with equal probabilities, as movement of pallid sturgeon in the Missouri River is uncertain. No decision or other nature nodes considered are expected to affect pallid sturgeon movement. However, between year movement may have an affect on an estimator’s ability to approximate population abundance and trend from capture history data, and therefore, movement is included as an input to each of the estimate performance metrics.

#### Recruitment

As population trend is a function of survival and recruitment, the performance of population trend estimates may be affected by varying recruitment of pallid sturgeon over several years. Annual recruitment of age-1 pallid sturgeon to the population was included in the network at two levels: “None” and “Low”. We defined “None” to mean that with 100% certainty recruitment did not occur during any year of the 10-year program. If recruitment is “Low” then both the frequency and level of recruitment are stochastic. Consequently, when recruitment is low, there is still a chance that recruitment does not occur during any given year of the study, but the probability that recruitment never occurred during the 10-year program is very small. The frequency of recruitment and the recruitment level are both basin-specific. Therefore, when recruitment does occur, it could occur in the Upper, the Lower, or both Missouri River basins. In the past 14 years of the PSPAP, only a handful of age-1 pallid strugeon have been captured. The absence of age-1 pallid sturgeon captures could be a result of lack of spawning, lack of survival to age-1, or a lack of power to detect low levels of recruitment. Given that spawning, survival, and age-1 detection probabilty are uncertain, we parameterized the recruitment nature node with equal probabilities. No nodes of the BDN affect recruitment; however, recruitment is linked to the population estimates (e.g., population trend depends on survival and recruitment).

#### Catchability

Catchability relates catch, effort, and population abundance as

*C*=*q*⋅*f*⋅*N*, (8)

where

*C* is the number of fish caught,

*f* is the fishing effort,

*N* is the number of fish available to be caught, and

*q* is the catchability value.

This relationship implies that catchability also relates effort to capture probability, *p*, by the equation *p*=*q*⋅*f*, giving rise to the definition of catchability as the probability of capturing any single fish with one unit of effort, given the fish is available for capture. The probability that a fish is caught at any point in time depends on several variables, making catchability a complex value that incorporates uncertainties from various biological and technilogical sources. Most importantly for addressing the problem at hand, catchability is dependent on gear. Hence, all nature nodes describing catchability are a function of the gear choice decision node.

While the monitoring program uses standardized protocols for each gear type to minimize catchability variation due to gear size, mesh, hook type, bait type, and duration of use, factors such as water temperature, water velocity and the experience of the crew make variation in catchability values unavoidable. To account for the inherent variation of catchability in field studies, we modeled catchability at the lowest level of the monitoring program, the deployment level. During each sampling occasion, 8 gear deployments are used within the bend to capture fish. To ensure the deployment-level catchability value is a probability, catchability is assumed to be logit-normally distributed:

 (9)

where

is the catchability of gear *g*,

is the standard deviation of the log-odds catchability for gear *g*, and

is the log-odds of the median catchability value for gear *g*, or

 (10)

where  is the median catchability value.

The baseline catchability nodes for the upper and lower basins represent values for the median deployment-level catchability, . Since different basins may use different gears (e.g., trammel nets in the upper basin and gill nets in the lower) each basin has its own separate nature node for baseline catchability. Reasonable ranges for baseline catchability values by gear were approximated based on resulting capture probabilities. Each gear considered was associated with a basin-specific effort distribution, where effort was placed on the common scale of time per deployment in minutes. Active gears considered (trammel nets) are deployed for significantly shorter periods of time than the passive gears considered (gill nets and trotlines) (table 1). When effort is measured in time of deployment, catchability values must be significantly higher for the active gears in order to acheive the same level of capture probabilities. We allowed the maximum baseline catchability node to extend to values that would give capture probabilities of as high as 0.4 for the typical trammel net deployment. The full range for baseline catchability was set from 0 to 0.006 (table 10). Since there is no evidence that passive gears have significantly higher capture probabilities than passive gears, the range for passive gears was set to the smaller range 0 to 0.00005. These different ranges in catchability reflect the different ranges in effort and allow for both active and passive gears to obtain similar capture probabilities (figs. 2-5). For trotlines and gill nets, the passive catchability range was parameterized with a uniform distribution and descritized into 4 evenly spaced levels. All catchabilities above 0.0005 were parameterized with 0 weight for the passive gears. The active range was also parameterized with a uniform distribution. The descritization for this range was developed to included 4 evenly spaced levels, where the lower level consists of 5 sublevels to account for the smaller mesh of the passive gears.

The catchability variation nodes represent ranges for the logit-transformed standard deviation, . "Low" variation is defined as , "Medium" as , and "High" as  (table 11). Similar, to baseline catchability, each basin has it own nature node for catchability variation in order to account for the use of different gears in different basins. As catchability variation is largely unknown for all gears, both catchability variation nodes were parameterized with equal probabilities for each of the three levels.

#### Mean

We used the probability of a fish not being captured during the sampling year () to summarize the capture probabilities of all sampling occasions within the year with a single value. Moreover, since the capture probabilities of all sampling occasions within the sampling year are direclty related to the capture history frequencies that affect estimator outcomes,  has the potential to serve as a useful simplified approach to linking sampling with estimator performance metrics (e.g., bias, precision). Indeed, for the simulations considered, we found a replicate’s average probability of not being captured during a sampling year (mean ) to be a good predictor of each of the five replicate-level estimator performance metric utilities described below (figs. 6-10). Therefore, the Mean  nature node is a key node in connecting sampling decisions (gear choice and number of secondary sampling occasions) and uncertainties in catchability to the estimator performance metric utilities that assess PSPAP v. 2.0 fundamental objective 2: quantify pallid strugeon population status and trend.

The probability of not being captured during the sampling year () is a function of the occasion-level capture probability and the number of secondary sampling occasions for a particular gear. Assuming independence, the probability of not being captured during the sampling year, is the product of the probabilities of not being captured during each sampling occasion that year, or

 (11)

where

is the probability of not being captured during the sampling year,

is the occasion-level capture probability during the *i*th sampling occasion,

is the number of sampling occasions during the sampling year, and

*i* indexes occasions within the sampling year.

As we simulated gear-specific occasion level capture probabilities as a function of deployment-level effort and catchability, the Mean  nature node is linked with gear (which directly determined effort) and the 4 nodes that determine the distribution of catchabilities. In particular,  is connected to effort and catchability by the combination of equation (11) and letting:

 (12)

where

 is the catchability value for the *k*th deployment of gear *g*  and

 is the effort (in minutes) for the *k*th deployment of gear *g* in basin *b*.

The 8 gear- and basin-specific effort values are randomly generated from a gamma distribution whose parameters were derived as the best fit to the gear- and basin-specific PSPAP effort data:

 (13)

where the gear- and basin-specific shape and rate values used in the parameterization of the Mean nature node are given in table 1. The 8 catchability values were drawn from the gear-specific logit-normal distribution described by equation (9).

We descritized the Mean  nature node into 10 equal sized bins ranging from 0 to 1 and parameterized it with simulations (table 12). For a particular combination of gear choice, lower basin baseline catchability bin, lower basin catchability variation bin, upper basin baseline catchability bin, upper basin catchability variation bin, and number of secondary occasions, each basin was first considered seperately. For the lower basin, one value was calculated for each of the 82 sampled bends during each of the 10 years using equations (9)-(13), where  and  were fixed but chosen uniformly at random from the given lower basin baseline catchability and catchability variation bins, respectively. Then the lower basin mean  value was calculated as the mean of all lower basin  values. Similarly, values were calculated for the upper basin, where only 45 bends were sampled each year, from the same equations with the exception that the fixed values of  and  were chosen uniformly at random from the given upper basin baseline catchability and catchability variation bins, respectively. We averaged the mean of the upper basin  values with the mean of the lower basin  values to obtain the overall mean value. This process was repeated 10,000 times for each combination of gear choice, associated catchability bins, and number of secondary occasions. We constructed the conditional probability table entries for the particular combination of inputs from the resulting probability density of the 10,000 mean  values.

#### Estimate Performance Metrics

The monitoring program objectives require population abundance and trend estimates to be computed on an annual basis. Estimates with a small bias and high precision are preferred to those with a large bias and low precision. To quantify the performance of the population abundance and trend estimates produced by a particular choice of monitoring program, the bias and precision of each estimate was considered. Additionally, a highly accurate and precise estimator is not as useful to the monitoring program if it almost always fails to produce an estimate. To account for this the “reliablity” of the estimator was also taken into account. In total, 5 estimator performance metric nodes were included in the BDN model:

1. Bias: Population Abundance,

2. Precision: Population Abundance,

3. Reliability: Population Abundance,

4. Bias: Population Trend, and

5. Precision: Population Trend.

Five other nodes in the BDN model are assumed to directly affect each of the estimator performance metric nature nodes listed above: Recruitment, Between Year Movement, Mean , Survey Design, and Estimator. Additionally, the nodes representing gear choice, baseline catchbility, catchability variation, and the number of secondary sampling occasions, all indirectly affect the estimator performance metrics through their relationship with the Mean  nature node. All but the Estimator input node affect data collection and may result in significant differences in capture histories. The interaction between the structure of the capture history data and the choice of estimator used for data analysis determines if an estimate can be made and affects the estimate’s precision and accuracy.

Parameterization of each of the performance metrics nodes was done through data simulation. A set of 400 reference populations were simulated. Half of these had no movement and no recruitment, while the other half had little movment and low recruitment. For each combination of survey design, gear choice, and number of secondary sampling occasions considered, 4 different catch-data sets were simulated for each population, resulting in a total of 25,600 sets. For all catch-data sets, baseline catchabilty and catchability variation were chosen at random from the gear-dependent ranges described above, resulting in various mean  values. Not all estimators could analyze each of these catch-data sets. For example, all of the capture-mark-recapture estimators require there to be at least 2 secondary capture occasions. Hence, these estimators can only be ran on three-fourths (19,200 sets) of the total number of simulated catch-data sets. With the excpetion of the CRDMS estimator, all other estimators have been applied to each of the catch-data sets they are compatible with. The CRDMS estimator takes a significant amount of computation time and has currently only been ran on 4,800 of the 19,200 catch-data sets available for analysis. Nonetheless, the replicate sample size used to parameterize the estimator performance metric nodes for each combination of input nodes were reasonably large for the combinations simulated. No reference populations were simulated that combined no movement with low recruitment or little between year movement with no recruitment, resulting in a gap in the parameterization of the reliability node for cases that include these combinations.

For each catch-data replicate described above, a single perfromance metric value was calculated. The details of this calculation, the construction of the conditional probability table, and the meaning of each performance metric node are described in the following sections.

##### Reliability: Population Abundance

The relibability of an estimator to produce annual segment-level population estimates is important when calculating if a particular estimator will provide relavent input data for the collaborative population model (PSPAP v. 2.0 fundamental objective 4). The reliability of a population abundance estimator is defined as the probability the estimator will successfully compute a segment-level abundance estimate under the given monitoring program and conditions. For each replicate catch-data set, the reliablity of the estimator was first calculated for each basin separately and then the basin-level data was averaged, reflecting the equal importance of assessing the pallid sturgeon population in each basin. Basin-level reliability was calculated as the proportion of segment-level estimates that were successfully computed across the 10-year study (i.e., the total number of segment-level abundance estimates successfully computed across the 10-year study divided by the product of 10 and the total number of segments within the basin). For the purposes of this analysis, an estimator successfully computed an abundance estimate if it did not fail and it did not give any warnings or errors. All segment-level estimates derived from bend-level estimates were successful if there was at least one bend-level estimate successfully computed. The probability density of replicate reliabilty values by input combination were then used to build the conditional probability table for the "Reliability: Population Abundance" node, which ranges from 0 to 1 and is discretized into 10 equally sized components (table 13).

##### Bias: Population Abundance

The “Bias: Population Abundance” nature node represents the mean utility of the basin-level abundance estimate bias. Each catch-data replicate analyzed by a particular estimator is associated with one bias utility value that is calculated from the following steps:

1. Calculate the relative absolute bias, or bias metric, of each basin-level abundance estimate for each of the 10 years.

2. Calculate the utility of each relative absolute bias value.

3. Calculate the mean of the utility values.

For each estimate, we calculated the bias metric as the ratio of the absolute value of the bias to the actual population abundance, where the bias was calculated as the population abundance estimated from the catch data minus the actual population abundance of the reference population from which the catch data was simulated. We then calculted the utility of the bias metric by scaling it to a value between 0 and 1, with the linear equation:

 (14)

where

is the utility of the abundance bias performance metric, or scaled bias performance metric,

*x* is the value of the bias metric, or absolute relative bias of the estimate, and

*x*=NA, if and only if the estimator failed to produce a basin-level abundance estimate for

the given basin during the given year.

For all cases where an estimator failed to produce a basin-level abundance estimate during a given year, the utility was set to 0 for that basin and year. For cases where an estimate was produced, when the estimate bias is near 0, as prefered, the utility is high (near 1). As the absolute relative bias increases towards 1, the utility of the bias performance metric decreases towards 0, and for absolute relative bias values greater than or equal to 1, the utility is 0. We chose 1 as an upper bound for the utility function because an absolute relative abundance bias that is greater than 1 would mean that the basin-level population estimate was more than double the actual population size.

In the last step, the mean of the 20 abundance bias utility values (one for each basin for each of the 10 years) is calculated for each replicate, and the density of the replicate means is used to parameterize the “Bias: Population Abundance” nature node, which ranges from 0 to 1 and is discretized into 10 equally sized bins (table 14).

##### Precision: Population Abundance

The “Precision: Population Abundance” nature node represents the mean utility of the precision performance metric, i.e., the basin-level abundance estimate’s coefficient of variation. For each basin-level abundance estimate, the coefficient of variation was calculated as the ratio of the standard error of the estimate to the estimate value. The utility of each precision performance metric was calculated by scaling the coefficient of variation as

 (15)

where

is the utility of the abundance precision performance metric (scaled precision performance metric),

*x* is the value of the precision performance metric (coefficient of variation), and

*x*=NA, if and only if the estimator failed to produce a basin-level abundance estimate for

the given basin during the given year.

The utility function scales each coefficient of variation between 0 and 1, assigning smaller coefficients of variation to higher utility values (near 1). Additionally the function assigns a utility of 0 to all cases of estimator failure or any coefficient of variation larger than 0.3. The mean of the 20 abundance precision utility values is calculated for each replicate, and the density of the replicate means is used to parameterize the “Precision: Population Abundance” nature node, which was discretized into 10 equally sized bins (table 15).

##### Bias: Population Trend

The “Bias: Population Trend” nature node represents the utility of the trend bias performance metric. We chose the trend bias performance metric to be the absolute value of the bias, where the bias of the trend estimate was calculated as the difference between the trend estimate and the actual trend. Similar to the trend estimate, the actual trend was calculated as

, (16)

where

λ represents the actual population trend, and

β is the slope of the linear model fit to the log-transformed annual segment-level abun-

dances as known from the reference population.

Each trend bias performance metric was scaled as

 (17)

where

 is the utility of the trend bias performance metric,

*x* is the value of the bias metric, or absolute bias of the estimate, and

*x*=NA, if and only if the estimator failed to produce a population trend estimate for the

catch-data replicate.

The scaled performance metric, or utility value, is a value between 0 and 1, with small absolute biases resulting in high (near 1) utility values. As the absolute bias of the trend estimate increases towards 0.7, the trend bias utility decreases towards 0. For any absolute bias values above 0.7, the trend estimate predicted a population growth or decline that was more than 70% off of the actual trend, and the bias utility is set to 0. Additionally, if the estimator was not able to produce a trend estimate for the 10-year data set, then the trend bias utility was also assigned a value of 0.

The “Bias: Population Trend” nature node extends across the full range of possible utility values (0 to 1) and is discretized into 10 equally sized bins (table 16). The node was parameterized by building the conditional probability table from the density of trend bias utility values among the replicates with the given combination of inputs.

##### Precision: Population Trend

The “Precision: Population Trend” nature node represents the utility of the trend precision performance metric. The precision of the trend estimate was calculated as the estimate’s coefficient of variation, or the ratio of the standard error of the estimate to the absoluate value of the estimate. To be placed on a similar scale to all other performance metrics, the trend precision performance metric was then scaled using the linear equation:

 (18)

where

 is the utility of the trend precision performance metric,

*x* is the value of the precision metric (coefficient of variation), and

*x*=NA, if and only if the estimator failed to produce a population trend estimate for the

catch-data replicate.

Similar to the abundance precision utility, the trend precision utility is near 1 when the coefficient of variation is near 0. As the coefficient of variation increases towards 0.375, the trend precision utility decreases towards 0 and remains at 0 for any estimates with a coefficient of variation greater than 0.375. Additionally, if the estimator was not able to produce a trend estimate for the 10-year data set, then the estimator was assigned a trend precision utility of 0.

We descritized the range of the “Precision: Population Trend” nature node into 10 equal sized ranges and parameterized the node by building the conditional probability table from the density of trend bias utility values among the replicates with the given combination of inputs (table 17).

### Cost

#### Recruitment Occasions

The number of recruitment occasions is the number of days each of the sampled bends will be visited in order to achieve the prescribed amount of recruitment-detection sampling. This value is a deterministic function of the number of trawls used to sample each bend. If recruitment-detection sampling is not included as part of the monitoring program (i.e., 0 trawls per bend), then there will be 0 recruitment occasions. If recruitment-detection sampling is included in the monitoring program, then anywhere between 1 and 49 trawls may be used. For less than 50 trawls per day at an average of 4 minutes per trawl, all detection recruitment trawls within a bend are assumed to be completed in a single day. Therefore, for all non-zero numbers of trawls per bend considered, the number of recruitment occasions is 1 (table 18). Hence, the recruitment occasions nature node is parameterized as

 (19)

where

*f* is the number of recuitment occasions, and

*x* is the number of recruitment-detection trawls per bend.

#### Expected Cost

The expected cost of the monitoring program is a function of the number of days each sampled bend needs to be visited in order to acheive the decided upon sampling protocols. The total number of sampling occasions per sampled bend is the number of sampling occasions chosen for the sturgeon-season sampling plus the number of occasions for the recruitment-detection sampling. The expected cost per sampling occasion per bend was roughly estimated using past annual field crew budgets and the associated sampling data by field crew. As it was not feasible to tease apart non-sampling and sampling costs, the expected cost indirectly includes overhead costs and data-processing costs without requiring separate nodes for inputs. The expected cost per sampling occasion was calculated as approximately $0.3 million, and the nature node (table 19) was parameterized as

, (20)

where

*r* is the number of recruitment-detection sampling occasions from the “Recruitment

Occasions” nature node,

*s* is the number of sturgeon-season sampling occasions from the “Number of Secondary

Occasions” decision node, and

*C* is the expected cost of the monitoring design.

## Utility Nodes

In order to combine monitoring design fundamental objective outcomes into an overall monitoring design utility, the outcomes need to be on a similar scale. Therefore, each fundamental objective outcome was ultimately measured in terms of a utility value that took on values between 0 and 1, where 0 was undesirable and 1 was desirable. Fundamental objective performance metric outcomes that took on values outside of the 0 to 1 range, were scaled to vary between 0 and 1.

### Scaled Fundamental Objective 1

The performance metric for quantifying fundamental objective 1—quantify pallid sturgeon recruitment to age-1—was the power to detect recruitment if recruitment occurs. The power to detect, or the probability of detecting a recruit within a sampling year given basin-level recruitment, was already a value between 0 and 1, where 0 was the least desirable outcome and 1 was the most desireable outcome. Therefore, no scaling out the “Probability of Detecting a Recruit” nature node was necessary in evaluating the “Scaled FO1 – Recruitment” node. The utility value for fundamental objective 1 are the same as the probability of detecting a recruit, allowing the “Scaled FO1 – Recruitment” node to be parameterized by the deterministic function,

 (21)

where

 is the utility value of fundamental objective 1, and

 is the probability of detecting a recruit at the basin-level during a sampling year given recruitment occurred.

### Scaled Fundamental Objective 2

There were 4 performance metrics used to quantify fundamental objective 2—quantify pallid sturgeon population status and trends. The 4 performance metrics quantified the precision and accuracy (bias) of the basin-level population abundance estimates and the population trend estimate. In raw form, these metrics all differed in range and magnitude; therefore, they were each scaled to vary from 0 to 1 with the most desireable outcome receiving a utility of 1. Scaling the performance metrics to utilities not only allows the outcomes of the 4 metrics to be compared, but it also allows for them to be combined in a way that reflects the values of the decision-makers. The weighted mean of the 4 perfomance metric utilities was used to parameterize the “Scaled FO2 – Trend & Abundance” node, so that

 (22)

where

 is the utility value of fundamental objective 2,

 is one of the 4 performance metric utilities,

 is the weight for performance metric utility  and ,

 indicates whether the performance utility is for the bias or precision (prec) metric, and

 indicates whether the perfomance utility is for the abundance (abund) or trend estimates.

Calculations for each  (i.e.,    and ) are given by equations (14)-(17) and described in detail above. The weighted mean of the perfomarnce metric utility values results in a value between 0 and 1, with 1 being the most desireable outcome and 0 being the least, as needed. Future work includes collecting decision-maker swing-weighting data for each performance metric in order to use weights that reflect how decision makers value each of the four fundamental objective 2 performance metrics. As decision-maker input has not yet been collected, all weights were assumed equal for the baseline analysis. To understand the how the model behaves under different weighting schemes other weighting scenarios were investigated and a sensitivity analysis was performed.

### Scaled Fundamental Objective 3

Fundamental objectve 3—maintian compatiblity with PSPAP v. 1.0 data to the extent possible—was influenced by only one node, the “Survey Design” nature node. The performance metric for fundamental objective 3 was the proportion of randomly selected bends within a segment. Since all monitoring designs perserved the number of bends to be sampled within each segment, the only difference is whether bends were selected randomly or not. Therefore the “Scaled F03 –Legacy Compatibility” node is parameterized by the deterministic function:

 (23)

where

 is the utility value of fundamental objective 3, and

 is the survey design selected by the decision-makers.

Since the performance metric for fundamental objective 3 is a proportion with 0 being the least desireable and 1 being the most desirable, no further scaling was necessary.

### Scaled Fundamental Objective 4

The performance metric for fundamental objective 4—provide relevant demographic data for pallid sturgeon population model inputs—was the possible proportion of the population model inputs that are estimated as a result of monitoring design sampling. This node is dependent on the estimator package chosen, as well as the reliability of the abundance estimator. Sampling protocols for all monitoring designs are assumed to be sufficient in collecting weight and length data, allowing a weight and growth models to be constructed and the parameters used as inputs for the population model. Under current sampling protocols, none of the monitoring designs are expected to be able to estimate the probability of spawning, the probability of being sexually mature, fecundity model parameters, the sex ratio, the maximum age, or the stocking level. Some these inputs, however, may be able to be estimated from data collected by other sampling programs (e.g., broodstock collection). Abundance inputs and survival inputs are also needed for the population model. Because of lack of data for young sturgeon, abundance and trend estimators are not expected to be able to provide estimates for pallid sturgeon less than age-3. However, all abundance estimators are expected to produce segment-level abundance estimates for the adult population of pallid sturgeon (both natural and hatchery origin) whenever they successfully produce an estimate. Moreover, some measure of age-3+ survival can estimated for all estimator packages. Lastly, the population model requires transition probabilities for between year movement on the bend-level. While the mulit-state robust design abundance estimator is capable of producing transition probabilities at the segment-level, it failed to run at this level. Therefore, no estimators can provide the movement probabilities inputs required for the population model.

### Scaled Fundamental Objective 5

Estimated cost was the performance metric for fundamental objective 5—minimize costs. Estimated costs ranged anywhere from 0.3 to 1.5 million dollars for the monitoring designs considered, and must be scaled to be compared with the other fundamental objective outcomes. The higher the cost of a monitoring program, the less desireable. Therefore, estimated cost was scaled using an equation that scales the minimum cost of 0.3 million dollars to a utility of 1 and the maximum estimated cost of 1.5 million dollars to a utility of 0:

 (24)

where

 is the utility value, or scaled performance metric value, of fundamental objective 5, and

 is the expected cost of the monitoring program.

This scaling equation was used to parameterize the “Scaled FO5 – Minimize Cost” node.

### Overall Utility

The overall utility of a monitoring program (i.e., the specific combination of decisions) is calculated as the weighted mean of the 5 fundamental objective utilities:

 (25)

where

 is the overall utility of the monitoring program,

 is the utility value for fundamental objective ,

 is the weight given to fundamantal objective  where , and

=1, 2, 3, 4, or 5 indexes the fundamental objective.

The weighted mean both takes into account how the decision-makers value different objectives (weights), as well as their satisfaction with the outcomes and attitudes towards risk (utility functions). As decision-maker input has not yet been collected, all weights were assumed equal for the baseline analysis. To understand the how the model behaves under different weighting schemes other weighting scenarios were investigated and a sensitivity analysis was performed.

## Thoughts on Catchability Formulation

If deployment *i* catches

fish, then the number of fish caught during the occasion is

but the sum of the capture probabilities can be (and is for some of the occasions in our simulations) more than 1, so it is possible that *C*>*N*. While we have forced a cap on the sum of the deployment level capture probabilities so that *C* will not be larger than *N*, this does not line up with the mathematical formulation, which suggests that we either should not be summing the deployment level capture probabilities in order to obtain the occasion-level capture probability, that we should be using a different formulation for deployment-level catch. While we cannot change or replace the model simulations we currently have, we can change the model description and network so that catchability is considered at the occasion level instead of the deployment level. This is a good solutions as the simulations we have generated use only occasion-level catch. Moreover, we can use a logit-normal distribution for occasion level catchability to ensure it is a probability. This would require simulating a new CPT for mean p0, which could be done in about a day (mostly due to simulation run time). While we didn’t run the current estimate simulations with occasion level catchability, if we really think catchability only affects the performance metrics through its relationship with mean p0 as we have modeled it, then this should not be a problem. Moreover, I might be able to use the median and sd of the realized occasion level catchabilities of each replicate to compare the predicted outcomes of the BDN model to the simulated outcomes and see if they reasonably match. It won’t be a perfect comparison, but it should give us some idea on whether or not changing to the logit-normal occasion-level approach is valid for our simulations. Thoughts?