THE PALLID STURGEON POPULATION ASSESSMENT PROGRAM

POWER ANALYSIS

PREPARED FOR THE U.S. ARMY CORPS OF ENGINEERS

BY

ADAM W. SCHAPAUGH

ANDREW J. TYRE

SCHOOL OF NATURAL RESOURCES

UNIVERSITY OF Nebraska Lincoln

Pioneering new frontiers.

EXECUTIVE SUMMARY

The Pallid Sturgeon Population Assessment Program (PSPAP) was developed to provide the information needed to detect changes in population size and habitat preferences for pallid sturgeon (*Scaphirhynchus albus*) and other native species in the Missouri River Basin. The PSPAP thus requires a survey design capable of detecting trends in population size that actually occur. We carried out a three-part analysis of the PSPAP to address the following objectives:

(1) Evaluate trends in relative abundance (catch-per-unit-effort [CPUE]) of pallid sturgeon from 2005 through 2010, and identify potential determinants of CPUE.

(2) Compare gear types to identify unnecessary or redundant gears, based on effectiveness and ability to sample the entire size distribution of pallid sturgeon.

(3) Compare the performance of alternative survey designs in three monitoring areas of the Missouri River, and then identify the design in each that maximizes statistical power with respect to fiscal and logistical considerations.

To address Objective (1), we fit generalized linear mixed models to evaluate how spatial (segment), temporal (year, month), and environmental (meso-habitat, temperature) features explain variation in CPUE of pallid sturgeon. To address Objective (2), we generated plots of the cumulative distribution function (CDF) for lengths of fish captured using each active and passive gear type. Gear selectivity was then evaluated graphically by comparing the CDF of each gear type to the CDF of all gear types. To address Objective (3), we employed generalized linear mixed models, where the discrete counts were treated as a Quasi-Poisson process, and used a simulation-based approach to explore how different survey designs influence statistical power.

OBJECTIVE 1: TREND ANALYSIS

Results from the analysis addressing Objective (1) indicate a decreasing trend in relative abundance as you move from the Upper Monitoring Area downstream to the Middle and Lower Monitoring Areas. This result is most likely due to the Pallid Sturgeon Propagation and Population Augmentation Program, which utilizes six hatcheries located in the northern portion of the basin. In the Upper Monitoring Area, estimated coefficients for segment and year effects were dissimilar, indicating a strong, positive temporal trend, particularly in Segment 4. This positive trend was not observed elsewhere, as CPUE typically peaked in 2007 and 2008, followed by declines to the present.

- **CONCLUSION:** This analysis supports PSPAP hypothesis 1.1A – there are differences between years in pallid sturgeon abundance, as well as hypothesis 1.2A – there are differences between segments in abundance, with abundance decreasing downstream. There is insufficient evidence to reject the null hypotheses 1.3 and 1.4; between year variation is large and not in a consistent direction.

OBJECTIVE 2: GEAR COMPARISONS

Results from the analysis addressing Objective (2) support the use of multiple gear types to ensure all size classes are adequately sampled. We found that six gear/ deployment combinations (TN25S, GN14S, GN18S, GN41S, GN81S, and TLC2S) are ineffective at sampling individuals less than 250 mm in length. Conversely, one gear type (OT16S) is highly effective at sampling individuals of this length. Gear type TNS is the most effective at sampling the entire size distribution of pallid sturgeon, but will tend to miss the largest individuals.

- **RECOMMENDATION:** Maintain the use of multiple gear types to ensure all size classes are adequately sampled.

OBJECTIVE 3: POWER ANALYSIS

The power to detect increases in CPUE of age 1 and younger fish and adults was generally low (less than 0.80), unless the configuration (number of bends/segment/year) *and* level (number of gear deployments/macro-habitat/bend) of effort were increased relative to the existing design. The probability of trend detection did increase with increases in the number of bends (while maintaining the number of gear deployments) and, generally to a lesser degree, to increases in the number of gear deployments (while maintaining the number of bends).

- **RECOMMENDATION:** To increase the probability of trend detection, allocate an increased number of gear deployments to more bends, with emphasis on the number of bends.

We found that trend detection is sensitive to the relative magnitude of the year effects. Even small values of variance components (due to river bend and subsample) can overwhelm the regression coefficients in a model. If their magnitude is large, power will increase primarily with time; adding bends or gear deployments will likely make little improvement. The sensitivity of trend detection to variance components implies multiple avenues for reducing their effect.

- **RECOMMENDATION:** Hypothesize and evaluate additional causal covariates (e.g., discharge, acres shallow-water habitat, etc) linking pallid sturgeon abundance with environmental features.
- **RECOMMENDATION:** Identify sections of the river with internally similar variance structures to improve trend detection at larger scales.

- **Recommendation:** Collect reproductive and mortality data, which would enhance trend detection with the added advantage of giving ecological information useful in devising a recovery response.

Our power analysis also yielded insights regarding the trade-offs between allocating an increased number of gear deployments to fewer bends, as opposed to a fewer number of gear deployments to more bends. A relevant general remark is that the number of bends is usually the most restrictive element in the design. In some cases, increasing the number of subsamples can be effective because little bend-to-bend variation occurs. When among-bend variation is greater, adding subsamples within a bend produces little data of relevance for inference across the set of bends.

- **Recommendation:** Maintain, if not increase, the configuration (number of bends sampled/segment/year) of the existing design.

Demonstrating trends in survey data requires a long-term commitment to monitoring and consistent data collection. It may be too obvious to state that concentrating survey effort within a year is not adequate to generate the data necessary to evaluate trends across years, regardless of whether the interest is within segments, monitoring areas, or basin-wide. In the case of long-lived, rare species, trends of interest are likely to be of a magnitude of 1–3% per year or less. The objective of the PSPAP is timely detection of biologically significant trends; yet pallid sturgeon would most likely grow (or decline) at a lesser rate. On the other hand, population indices for fish can be so variable that a statistically significant "trend" of ±10% per year may occur over a period of only a few years due to factors other than a change in abundance. In that case, the threshold for statistical significance bears little relationship to biological significance. Thus, the greater challenge is not statistical power per se, but a better understanding of what is meaningful from a population perspective.

- **Recommendation:** Clarify the meaning of "long term trends" in PSPAP hypotheses 1.3 and 1.4.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

INTRODUCTION

Monitoring abundance over time is a key aspect of many conservation initiatives. Demonstrating trends in survey data is, however, difficult (Field et al. 2005). A ubiquitous concern is misinterpretation: given a null hypothesis of no trend, two errors are possible – inferring a trend when none exists (Type I error) or inferring no trend when in fact one does (Type II error). Many biological monitoring programs presuppose survey designs that are capable of detecting trends that actually occur. That is, having clearly defined the magnitude, or effect size, of interest, subsequent data analysis must carry adequate statistical power to permit sound decision-making (Field et al. 2004).

Cohen (1962; 1969) introduced power analysis as a technique suitable for estimating trend detection probabilities. Given that most landscape-scale monitoring programs are subject to tight financial and logistical imperatives, *efficient* survey design is vital. Efficient survey design maximizes statistical power within constraints imposed by available effort (Field et al. 2005). Power analysis in this context requires the ability to estimate statistical power *and* cost of alternative designs. These, in turn, must be integrated in an analysis weighing the trade-offs between performance and expense.

The Pallid Sturgeon Population Assessment Program (PSPAP) was developed to provide the information needed to detect changes in population size and habitat preferences for pallid sturgeon (*Scaphirhynchus albus*) and other native species in the Missouri River Basin. The PSPAP thus requires a survey design capable of detecting trends in population size that actually occur. We carried out a three-part analysis of the PSPAP to address the following objectives:

(1) Evaluate trends in relative abundance (catch-per-unit-effort [CPUE]) of pallid sturgeon from 2005 through 2010, and identify potential determinants of CPUE.

(2) Compare gear types to identify unnecessary or redundant gears, based on effectiveness and ability to sample the entire size distribution of pallid sturgeon.

(3) Compare the performance of alternative survey designs in three monitoring areas of the Missouri River, and then identify the design in each that maximizes statistical power with respect to fiscal and logistical considerations.

Often, in studies involving count data, a generalized linear model with a Poisson error distribution is used. However, the PSPAP survey design has a nested structure such that random effects arise due to samples and subsamples. Correlated observations may also lead to over-dispersion, which is problematic given only one free parameter and the inability to adjust the variance independently of the mean. To address Objective (1), we thus fit generalized linear mixed models to evaluate how spatial (segment), temporal (year, month), and environmental (meso-habitat, temperature) features explain variation in CPUE of pallid sturgeon. To address Objective (2), we generated plots of the cumulative distribution function (CDF) for lengths of fish captured using each active and passive gear type. Gear selectivity was then evaluated graphically by comparing the CDF of each gear type to the CDF of all gear types. To address Objective (3), we employed generalized linear mixed models, where the discrete counts were treated as a Quasi-Poisson process, and used a simulation-based approach to explore how different survey designs influence statistical power.

METHODS

Access to PSPAP data for 2005 - 2010 was provided by the United States Army Corps of Engineers. Our database extract contained 41,790 individual net deployments; each included an identification number, a record of set date, sample period, year, gear type, and trawl effort, and the following location records: segment, bend, river mile, and macro- and meso-habitat at the net site. A count and the Missouri River Standard Operating Procedures for Fish Sampling and Data Collection code (Welker & Drobish 2010) of the species sampled were also recorded.

We briefly describe the PSPAP survey design and sampling protocols relevant to the analysis. Exhaustive detail is provided by Welker and Drobish (2010). The PSPAP survey design treats river bends as the experimental units, which is the largest scale at which measurements are replicated. River bends include three continuous macro-habitats (outside bend, inside bend, channel crossover) and up to 12 discrete macro-habitats (large and small tributary mouths, confluences, large and small secondary connected channels, non-connected secondary channels, deranged, braided, and dendritic). Within each year and segment combination, a random sample of river bends is selected and, within each river bend, gear deployments are distributed across the macro-habitats present. Gear types are active (otter trawl [OT] and trammel net [TN]) and passive (gill net [GN], mini-fyke net [MF], and trotline [TL]). The unit of measure for active gears is catch-per-unit-effort (CPUE) based on distance trawled and minimum deployment area. The unit of measure for passive gears is CPUE based on soak time and minimum deployment area or number of hooks.

ABUNDANCE ANALYSIS

We fit generalized linear mixed models to evaluate how spatial (segment), temporal (year, month), and environmental (meso-habitat, temperature) features explain variation in the relative abundance (catch-per-unit-effort [CPUE]) of pallid sturgeon. We first treated the discrete counts as a Poisson process:

$$y_{ijklm} \sim Poisson(\lambda_{ijklm}\, d_{ijklm});$$

where y_{ijklm} is the CPUE for gear i, year j, segment k, river bend l, and subsample m. The term d_{ijklm} is an offset that effectively resulted in a rate (number of fish collected/effort) by accounting for variation among individual samples that arose primarily through trawling distance (in the case of active gears) or soak time (in the case of passive gears). Due to correlated observations, we treated the discrete counts as an over-dispersed or Quasi-Poisson process (i.e., to inflate the variance, the dispersion parameter was not fixed at one). In addition, we assumed that the log of λ_{ijklm} followed the linear equation:

$$log\ (\lambda_{ijklm}\, d_{ijklm})\ = \mu + \ \alpha_i + \beta_j + \gamma_k + \beta\gamma_{jk}\ + A_{kl} + \ S_{ijklm} + log\ (d_{ijklm});$$

where μ is the mean over all gears, years, segments, bends and subsamples; α_i is the fixed effect due to gear (i = 1,…, I); β_j is the fixed effect due to year (j = 1,…, J); γ_k is the fixed effect due to segment (k = 1,…, K); A_{kl} is the random effect due to river bend (l = 1,.., L); and S_{ijklm} is the random effect due to subsample (m = 1,…, M). This multi-gear model allowed us to assess CPUE patterns in space and time without disregarding the effects of gear efficiency and variable detection probabilities (for a discussion of this approach, see Arab et al. 2008).

Variable selection is a central component in this and other regression techniques, and a number of fundamentally different approaches are possible: use of all collated candidate predictors, assuming that the model can isolate the relevant variables by assigning coefficients reliably (a poor strategy in the case of rare species – see Elith & Leathwick 2009); use of a selection algorithm, such as stepwise selection (again, problematic – see Whittingham et al. 2006); or creating multiple models with different subsets of predictors. Model selection has become popular in ecology and in applied settings (see, for example, Burnham & Anderson 2002). Part of the appeal is the ability to address the tradeoff between the maximization of model fit and the minimization of prediction error, in particular avoiding the problem of over-fitting. Over-fitting produces a model that is adapted to the patterns in the training data to such an extent that it no longer generalizes well. The primary objective in fitting a predictive model is thus to control the process such that major trends in the data are captured while sample specific noise is ignored. This can be achieved in several ways, but the most common is to use an information criterion to select the best subset of models from many, based on different combinations of predictors (Elith & Leathwick 2009).

We thus constructed an identical set of *a priori* models (Table 1) for each of three monitoring areas: Upper Monitoring Area (Segments 2, 3, 4); Middle Monitoring Area (Segments 5, 6); and Lower Monitoring Area (Segments 7, 8, 9, 10, 13, 14) (Figure 1), and used information-theoretic methods (Akaike Information Criterion; AIC) to rank models based on complexity and fit (Burnham & Anderson 2002). The variables year and segment were included in every model (but the null model) because we expected CPUE to vary spatially and temporally. Month was an additional covariate included to account for within-year variability. Meso-habitat and temperature were included because we anticipated that CPUE would vary with these small-scale environmental variables. Due to the large number of terms in some models, it was convenient in Table 1 to adopt a 'coefficient free' form of model specification (sensu Szabo *et al.* 2009). For example, $x + y$ specifies a linear predictor of the form $\beta_0 + \beta_1 x + \beta_2 y$, while $x \times z + y \times z$ is interpreted as $\beta_0 + \beta_1 x + \beta_2 y + \beta_3 z + \beta_4 xz + \beta_5 yz$. In other words, multiplication signs (×) indicate interactions between the terms attached to the operator. Following Bolker et al. (2008), we used Laplace approximations to calculate the likelihoods, which estimate the standard deviations of the random effects by assuming that the fixed-effect estimates are precisely correct. We used package lme4 (Bates & Maechler 2009) and package nlme (Pinheiro et al. 2009) in Program R 2.10.1 (R Development Core Team 2009) for this analysis.

Gear Selectivity

Plots of the cumulative distribution function (CDF) were generated for lengths of fish captured using each active and passive gear (described below). A CDF provides information on the percent of captured individuals that lie below or are equal to a given indicator value (in this case, fork length). Gear selectivity was then evaluated graphically by comparing the CDF of each gear type to the CDF of all gear types. Based on results from an initial analysis, we also produced separate plots for the same gears having different orientation to the current. We used package stats (R Development Core Team 2009) in Program R 2.10.1 (R Development Core Team 2009) for this analysis.

Power Analysis

Given the nested structure of the PSPAP survey design, calculating power for alternative configurations and levels of survey effort requires knowledge of variance components associated with random effects due to river bend and subsample. In studies involving grouped or hierarchical count data, estimating the variance of a statistic requires resolving issues not faced in a typical linear setting: the mean is presumed to vary multiplicatively, sampling variance is an increasing function of the mean, and subsample and sample variances may be estimated on different scales (Gray & Burlew 2007). To this we add that grouped data, count or otherwise, will generally exhibit group effects on the mean, and that the variation among these groups may substantially influence the variance of trend estimates and, hence, power to detect trends. Values of variance components were thus inferred (and subsequently considered known) based on estimates obtained from a generalized linear mixed model, as described above, consisting of three fixed effects (gear, year, and segment).

The primary goal of this analysis was to explore how different survey designs influence the ability to detect an increase in CPUE over a ten-year period. To do so, we simulated data from the model described above, assuming that the sample from each trawl was Negative Binomial distributed with the mean given by the model, and a dispersion parameter $k = 0.5$. This generated data similar overdispersion characteristics similar to the original dataset. We simulated data for 10 years, 2011 – 2020, assuming that the year effects were compounding rates of increase. We transformed all simulated data by adding 1 to the observed count and dividing by the set effort. This CPUE was subjected to a Box-Cox transformation with λ = -0.45, which was the best fit value for the same transformation on the original dataset. Although this transformed data was more difficult to interpret, we were able to use linear mixed models instead of generalized linear mixed models, which dramatically sped up the simulation process. Each simulated dataset was analyzed with a linear mixed effects model with the same random effects structure as the original dataset, and fixed effects as described above.

We simulated 100 datasets for each of the scenarios described below, and recorded the number in which a trend was detected as a measure of statistical power (i.e., the probability of detecting an increase in CPUE). The appropriate null hypothesis is: $\mathrm{H}_0\colon \bar{\mu}_{2011} = \bar{\mu}_{2020}$. The appropriate alternative hypothesis is $\mathrm{H}_\mathrm{A}\colon \bar{\mu}_{2011} < \bar{\mu}_{2020}$. We tested the power to detect different rates of increase in relative abundance using a *t*-test on the estimated yearly coefficients of 2011 and 2020. The compounding rates of increase were 1, 5, and 10% per annum. Power is given by the probability of rejecting the null hypothesis given that the alternative is true. This is the same as the probability that the test statistic is greater than or equal to the critical value of the test given significance level, α, under the *t*-distribution with the appropriate degrees of freedom.

We evaluated nine alternative design scenarios–the configuration (number of bends/segment/year) and level (number of gear deployments/macro-habitat/bend) of these are shown in Table 2. In brief, Scenarios 1-3 proposed maintaining the existing number of bends/segment/year, Scenarios 4-6 proposed decreasing the number of bends/year by 10% in each segment, and Scenarios 7-9 proposed increasing the number of bends/year by 10% in each segment. Within these three alternative configurations, we considered three levels of survey effort: 1, 2, and 3 gear deployments per macro-habitat per bend (5, 10, and 15 gear deployments/bend, respectively [Middle Monitoring Area]; 9, 18, 27 gear deployments/bend [Upper and Lower Monitoring Areas]). To clarify, Scenario 1 maintained the existing number of bends with 1 gear deployment per macro-habitat, Scenario 2, which represents the existing survey design, increased this level from 1 to 2 gear deployments per macro-habitat, and Scenario 3 evaluated 3 gear deployments per macro-habitat. This pattern was repeated for Scenarios 4-6 and again for Scenarios 7-9. These scenarios were evaluated separately for each monitoring area, repeated for three compounding rates of increase as described above, and each of three size classes: age 1 and younger (fork length < 300 mm); juveniles (300 mm < fork length < 750 mm); and adults (fork length > 750 mm).

We used package lme4 (Bates & Maechler 2009), package nlme (Pinheiro et al. 2009), package doBy (Hojsgaard et al. 2011) and package snow (Tierney et al. 2011) in Program R 2.10.1 (R Development Core Team 2009) for all simulations.

RESULTS

Abundance Analysis

For clarity and ease of presentation, we have organized results separately for each monitoring area. In each subsection, we first draw your attention to model selection and then highlight notable findings from model inference. All best-fit parameter estimates can be found in Tables 3-5.

Upper Monitoring Area

Four models were ranked with ΔAIC values less than or equal to 4.00 (Table 1). Together, they accounted for more than 99% of the Akaike weight, and each included the predictor variables: gear, segment × year, and month. The best model also included the predictor variable: meso-habitat. The next best model was more complex and had the best model nested inside it, therefore we only present parameter estimates from the model with ΔAIC = 0 (Table 3). Estimated coefficients for segment and year effects were dissimilar, indicating a strong, positive temporal trend, particularly in Segment 4 (Figure 2). Coefficients for gear and meso-habitat effects, though variable, were not significantly different from zero.

Middle Monitoring Area

A single model was ranked with ΔAIC value less than or equal to 4.00 (Table 1). It accounted for 75% of the Akaike weight and included the predictor variables: gear, segment × year, month, and temperature. Estimated coefficients for the segment × year interaction were dissimilar. In Segment 5, CPUE reached a peak in 2007, followed by a drop-off that persisted through 2010. In Segment 6, CPUE also reached a peak in 2007, but maintained a similar level of relative abundance through 2008 before again declining (Figure 3). Significant gear effects were present, and the best-fit parameter indicated a positive relationship between CPUE and water temperature (Table 4).

Lower Monitoring Area

Two models were ranked with ΔAIC values less than or equal to 4.00 (Table 1). Together, they accounted for 92% of the Akaike weight, and included the predictor variables: gear, segment × year, and temperature. The best model also included the meso-habitat × temperature interaction. Aside from Segment 13, estimated coefficients for the segment × year interaction were similar (Table 5). Significant gear effects were present and, though variable, meso-habitat and temperature effects were not significantly different from each other. Visual inspection of Figure 4 reveals a decreasing trend in relative abundance as you move downstream.

Gear Selectivity

Visual inspection of Figure 6 was used to compare the effectiveness of gear types for different size classes. A gear type is relatively more effective if the cumulative size distribution for that gear is above the cumulative size distribution for the entire sample, and relatively less effective if it is below the cumulative size distribution of the entire sample. Six gear types (TN25S, GN14S, GN18S, GN41S, GN81S, and TLC2S) are ineffective at sampling individuals less than 250 mm in length. Conversely, one gear type (OT16S) is highly effective at sampling individuals of this length; however, more than 75% of the individuals captured with this gear are less than 500 mm in length. Gear type TNS is probably the most effective at sampling the entire size distribution of pallid sturgeon, but will tend to miss the largest individuals. Gear type MFS is completely ineffective at sampling any size of pallid sturgeon.

POWER ANALYSIS

For clarity and ease of presentation, we have again organized this section by monitoring area. In some instances, insufficient catches prevented us from calculating power using our simulation-based approach (for young-of-the-year fish and adults). Therefore, within each subsection, we devote the most attention to results for the juvenile size class. Recall that we simulated data for 10 years, 2011 – 2020, assuming that the year effects were compounding rates of increase. Fixing trends of 1, 5, and 10% per year became problematic when the baseline rate was approximately less than 10 individuals per year, which is not terribly meaningful when discussing individuals. As we were dealing with CPUE, rather than actual abundance, this was technically not what occurred in the simulation, but it does illustrate the difficulty. We have identified in the text below where this occurred. We have also included the number of individuals in that size class that were captured and the total number of gear deployments over the period 2006 – 2010.

Upper Monitoring Area

The power to detect increases in CPUE of age 1 and younger fish was generally less than 10%, unless the configuration (number of bends/segment/year) *and* level (number of gear deployments/macro-habitat/bend) of effort were increased relative to the existing design (Table 6). Nonetheless, the probability of trend detection responded to increases in the number of bends (while maintaining the number of gear deployments) and, to a lesser degree, to increases in the number of gear deployments (while maintaining the number of bends; Figure 7). It is therefore more useful to maintain the configuration and level of effort of the existing design, as opposed to sampling more intensively at fewer bends. The power to detect increases in CPUE of juveniles was consistently greater (Table 6). Holding constant the number of gear deployments while increasing the number of bends increased the probability of rejecting the null hypothesis – Scenario 5: $1\text{-}\beta = 0.01, 0.05, 0.23$ (1% annual increase, 5% annual increase, 10% annual increase, respectively); Scenario 2: $1\text{-}\beta = 0.03, 0.13, 0.38$; Scenario 8: $1\text{-}\beta = 0.27, 0.65, 0.95$ (Figure 8). Maintaining the number of bends while increasing the number of gear deployments also lowered the Type II error rate – Scenario 1: $1\text{-}\beta = 0.01, 0.08, 0.29$; Scenario 2: $1\text{-}\beta = 0.03, 0.13, 0.38$; Scenario 3: $1\text{-}\beta = 0.14, 0.45, 0.76$ (Figure 8). Like before, decreasing the number of bends while simultaneously increasing the number of gear deployments (from 2 to 3) reduced power as compared to the existing design (Figure 8). In other words, the probability of rejecting the null hypothesis given Scenario 6 (1-β = 0.01, 0.05, 0.33) was less than Scenario 2 (1-β = 0.03, 0.13, 0.38). Insufficient captures prevented us from calculating power for adults (26 individuals; 11,283 gear deployments).

Middle Monitoring Area

The power to detect increases in CPUE of juveniles peaked asymptotically as the configuration of survey effort was increased relative to the existing design, but yielded diminishing returns with increases in the level of effort (compare Scenarios 8 and 9, Figures 9, 10). Holding constant the number of gear deployments while increasing the number of bends increased the probability of rejecting the null hypothesis – Scenario 5: 1-β = 0.05, 0.13, 0.31 (1% annual increase, 5% annual increase, 10% annual increase, respectively); Scenario 2: 1-β = 0.08, 0.28, 0.40; Scenario 8: 1-β = 0.41, 0.73, 0.98 (Figure 10). Maintaining the number of bends while increasing the number of gear deployments also lowered the Type II error rate – Scenario 1: 1-β = 0.08, 0.15, 0.36; Scenario 2: 1-β = 0.08, 0.28, 0.40; Scenario 3: 1-β = 0.17, 0.37, 0.57 (Figure 10). In contrast to the results from the Upper Monitoring Area, decreasing the number of bends (thereby improving logistics, less travel between sites) while increasing the number of gear deployments (from 2 to 3) did not reduce power as compared to the existing design (Figure 10). The probability of rejecting the null hypothesis given Scenario 6 (1-β = 0.14, 0.24, 0.41) was similar to that given Scenario 2 (1-β = 0.08, 0.28, 0.40). It may therefore more useful to sample more intensively at fewer bends, as opposed to maintaining the configuration and level of effort of the existing design. Insufficient captures prevented us from calculating power for age 1 and younger fish (9 individuals; 3,371 gear deployments) and adults (27 individuals; 3,371 gear deployments).

Lower Monitoring Area

The power to detect increases in CPUE of age 1 and younger fish was low, unless the configuration *and* level of effort were increased relative to the existing design (Table 6). Nonetheless, the probability of trend detection responded to increases in the number of bends (while maintaining the number of gear deployments) and, to a greater degree, to increases in the number of gear deployments (while maintaining the number of bends; Figure 11). For this size class, it may be more useful to sample more intensively at fewer bends, as opposed to maintaining the configuration and level of effort of the existing design. The power to detect increases in CPUE of juveniles was consistently greater (Table 6). Holding constant the number of gear deployments while increasing the number of bends increased the probability of rejecting the null hypothesis – Scenario 5: 1-β = 0.01, 0.09, 0.29 (1% annual increase, 5% annual increase, 10% annual increase, respectively); Scenario 2: 1-β = 0.03, 0.19, 0.44; Scenario 8: 1-β = 0.03, 0.20, 0.60 (Figure 12). Maintaining the number of bends while increasing the number of gear deployments also lowered the Type II error rate – Scenario 1: 1-β = 0.01, 0.04, 0.18; Scenario 2: 1-β = 0.03, 0.19, 0.44; Scenario 3: 1-β = 0.01, 0.20, 0.50 (Figure 12). Decreasing the number of bends while simultaneously increasing the number of gear deployments (from 2 to 3) reduced power as compared to the existing design (Figure 12). In other words, the probability of rejecting the null hypothesis given Scenario 6 ($1-\beta = 0.01, 0.11, 0.38$) was less than Scenario 2 ($1-\beta = 0.03, 0.19, 0.44$). It is therefore more useful to maintain the configuration and level of effort of the existing design, as opposed to sampling more intensively at fewer bends. The power to detect increases in CPUE of adults was slightly lower (Table 8). The Type II error rate declined with increases in the number of gear deployments (while maintaining the number of bends) and, to a lesser degree, to increasing the number of bends (while maintaining the number of gear deployments; Figures 11, 12). Similar to the result for young-of-the-year fish, it may be more useful to sample more intensively at fewer bends, as opposed to maintaining the configuration and level of effort of the existing design.

DISCUSSION

Results from the abundance analysis indicate a decreasing trend in relative abundance as you move from the Upper Monitoring Area downstream to the Middle and Lower Monitoring Areas. This result is most likely due to the Pallid Sturgeon Propagation and Population Augmentation Program, which utilizes six hatcheries located in the northern portion of the basin. In the Upper Monitoring Area, estimated coefficients for segment and year effects were dissimilar, indicating a strong, positive temporal trend, particularly in Segment 4. This positive trend was not observed elsewhere, as CPUE typically peaked in 2007 and 2008, followed by declines to the present.

- **CONCLUSION:** This analysis supports hypothesis 1.1A – there are differences between years in pallid sturgeon abundance, as well as hypothesis 1.2A – there are differences between segments in abundance, with abundance decreasing downstream. There is insufficient evidence to reject the null hypotheses 1.3 and 1.4; between year variation is large and not in a consistent direction.

Results from the gear selectivity analysis support the use of multiple gear types to ensure all size classes are adequately sampled. We found that six gear types (TN25S, GN14S, GN18S, GN41S, GN81S, and TLC2S) are ineffective at sampling individuals less than 250 mm in length. Conversely, one gear type (OT16S) is highly effective at sampling individuals of this length. Gear type TNS is the most effective at sampling the entire size distribution of pallid sturgeon, but will tend to miss the largest individuals.

- **RECOMMENDATION:** Maintain the use of multiple gear types to ensure all size classes are adequately sampled.

The power to detect increases in CPUE of age 1 and younger fish and adults was generally low (less than 0.80), unless the configuration (number of bends/segment/year) *and* level (number of gear deployments/macro-habitat/bend) of effort were increased relative to the existing design, and the annual increase was 10%. Nonetheless, the probability of trend detection responded to increases in the number of bends (while maintaining the number of gear deployments) and, most often to a lesser degree, to increases in the number of gear deployments (while maintaining the number of bends).

- **RECOMMENDATION:** To increase the probability of trend detection, allocate an increased number of gear deployments to more bends, with emphasis on the number of bends.

We found that trend detection is sensitive to the relative magnitude of the year effects. The nested design of the PSPAP points to one consistent, overwhelming theme determining statistical power – the influence of random effects due to river bend and subsample. Even small values of these variance components can overwhelm the regression coefficients in a model. If their magnitude is large, power will increase primarily with time; adding bends or gear deployments will likely make little improvement. The sensitivity of trend detection to variance components implies that one avenue for reducing their effect is to hypothesize and evaluate additional causal covariates. A second avenue may be identifying sections of the river with internally similar variance structures to improve trend detection at larger scales. Achieving this would require concerted effort to measure potentially controlling factors at relative large spatial scales, factors that may cause coherent variation, or might organize portions of the pallid sturgeon population into similar sub-populations (Urquhart et al. 1998). Reproductive and mortality data would also enhance trend detection with the added advantage of giving ecological information useful in devising a recovery response (see also Field et al. 2005). However, these data are more costly to collect than abundance data and may be prohibitively expensive. Nevertheless, the relative efficiency of these alternative data sources and optimal allocation of effort among them remains an issue for additional study.

- **RECOMMENDATION:** Hypothesize and evaluate additional causal covariates (e.g., discharge, acres shallow-water habitat, etc) linking pallid sturgeon abundance with environmental features.
- **RECOMMENDATION:** Identify sections of the river with internally similar variance structures to improve trend detection at larger scales.
- **RECOMMENDATION:** Collect reproductive and mortality data, which would enhance trend detection with the added advantage of giving ecological information useful in devising a recovery response.

Our power analysis also yielded insights regarding the trade-offs between allocating an increased number of gear deployments to fewer bends, as opposed to a fewer number of gear deployments to more bends. The extent to which within-bend (i.e., subsamples) data can be used to infer trends at larger scales depends on how well the subsamples represent the distribution of trends across multiple bends. A relevant general remark is that the number of bends is usually the most restrictive element in the design. Put another way, decreasing the number of bends (by 10%) while simultaneously increasing the number of gear deployments (from 2 to 3) often reduced power as compared to the existing design (Figures 7, 8, and 12). This is because trends over data grouped by bend are estimated at the bend scale. In some cases (see the following paragraph), increasing the number of subsamples can be effective because little bend-to-bend variation occurs. When among-bend variation is greater, adding subsamples within a bend produces little data of relevance for inference across the set of bends. A second contributor to the relative unimportance of the subsample variance component is that the variance of a mean (the standard error squared) is a monotonically decreasing function of sample size. As this feature is not shared with the bend-level variance components, the relative importance of the bend-level variance terms increases as the subsample size increases. This result is similar to that highlighted rather extensively by the hierarchical modeling literature and also described from a survey sampling perspective (see Urquhart et al. 1998; Raudenbush & Bryk 2002). Further discussion of the contributions of variance components at nested scales is provided by Snijders and Bosker (1999) and Hox (2002).

- **RECOMMENDATION:** Maintain, if not increase, the configuration (number of bends sampled/segment/year) of the existing design.

This result did not hold in all cases, however, as it is more useful to sample intensively at fewer bends in the Lower Monitoring Area (comparing Scenarios 2 and 6 – age 1 and younger and adults; comparing Scenarios 2 and 7 – juveniles, Figure 11). This result most likely arose from circumstantial differences in sampling across the monitoring areas, rather than being an exception to the generalization just discussed. Recall from previously that within each year and segment combination, a random sample of river bends is selected and, within each river bend, gear deployments are distributed across the macro-habitats present. The critical wording here is that gear deployments are distributed across the macro-habitats *present*. Thus, if a macro-habitat is not present, no gear deployments are made. In the Middle Monitoring Area, five macro-habitat types have been consistently sampled over the last five years. In the Lower Monitoring Area, there are nine macro-habitat types that have been consistently sampled. Recalling that we considered three levels of survey effort (1, 2, and 3 gear deployments per macro-habitat per bend), it becomes clear that two gear deployments/macro-habitat in the Upper Monitoring Area is not the same number of samples per bend as two gear deployments/macro-habitat in the Lower Monitoring Area. The point here is that results from one monitoring area are not directly comparable to results from any one of the other monitoring areas. In a previous work, Bryan et al. (2010) conducted an analytical power analysis of the PSPAP to determine the most efficient sampling (number of bends per segment) and sub-sampling (gear deployments per bend) effort in combination. The levels of subsample effort tested were fixed at 4, 8, 12, and 16 gear deployments per bend. The obvious advantage of this approach is the ability to compare monitoring areas ('zones' in Bryan et al. 2010) directly. For our purposes, however, we sought to match the analysis to the scenarios approved by the U.S. Army Corps of Engineers.

There were differences between the previous power analyses of the PSPAP conducted by Peery (2004), Bryan et al. (2010) and the analyses conducted here. Peery (2004) used a generalized linear model with a binomial error distribution. This logistic regression approach thus estimates the power to detect changes in occurrence, rather than relative abundance. Peery (2004) also did not consider variation among bends or among gear deployments within bends during data simulation. In addition, power was evaluated at $\alpha = 0.10$. These differences explain why Peery (2004) estimated higher power than the results included here. Bryan et al. (2010) developed an analytical method where the variance components were estimated from a linear mixed model. They used a standard ANOVA formulation with multiple fixed effects and nested random effects that resulted in an F-test for the main effect due to year. The null hypothesis was: $\mathrm{H}_0\colon \bar{\mu}_{year\ 1} = \cdots = \bar{\mu}_{year\ 10}$. The alternative hypothesis assumed at least one inequality. Thus, Bryan et al. (2010) determined if there existed a significant difference among any of the years sampled, while we compared the estimated yearly coefficients of 2011 and 2020. In addition, Bryan et al. (2010) examined compounding rates of decline (1, 5, and 10% per annum, over 10 years), while we considered rates of increase. Decreasing the population mean as they did, while keeping the variance fixed resulted in a lower non-centrality parameter. We simulated data assuming that the sample from each trawl was Negative Binomial distributed, which allowed the variance to increase as the mean increased. While our estimates of power are not optimistic, the estimates reported by Bryan et al. (2010) are conservative relative to the results presented here. We note that we found the analytical approach of Bryan et al. (2010) appealing so long as (1) simulations were not practical (i.e., a large number of scenarios), and (2) power calculations were required for similar scenarios and where the accuracy of a fraction of those calculations would be confirmed through simulation (for a discussion, see Gray & Burlew 2007). We adopted a simulation-based approach because (1) based on existing data, within-bend sample sizes tend to vary systematically (given the macro-habitats present) and (2) the flexibility to model additional covariates. We do acknowledge that for studies modeling multiple population effects on intercepts and/or slopes, simulations involving exponential families require substantially more computing power and time.

We limited our analysis to a rather narrow subset of possible design scenarios to provide a simple demonstration of the method and process for efficient survey design. We highlight the following limitations and offer suggestions for extending this work.

Estimates of power included here (simulation-based), as well as analytical estimates, require candidate values of sample and subsample variance components. Unfortunately, these are typically estimated imprecisely (Gray & Burlew 2007). This is because precision and negative-bias of group-level variance components improve with the number of levels (Hox 2002). Snijders and Bosker (1993) suggest estimating power over a range of variance values. For this report, time constraints prevented us from exploring in-depth the impact of bias and/or imprecision of variance component estimates, but doing so would be a worthwhile exercise.

We simulated data assuming that the year effects were compounding rates of increase. In reality, increases (or declines, for that matter) will not occur in a simple linear fashion, they will instead exhibit stochastic dynamics characteristic of natural populations. Detecting trends under these circumstances will be considerably more difficult (Field et al. 2005). Consequently, testing the generality of our results using a model that in some way includes population dynamics subject to environmental stochasticity would be a useful extension.

We presented the trend detection problem in a null-hypothesis framework because this remains the most commonly used approach. We have presented results based on the significance level $\alpha = 0.05$, but unless the cost of committing a Type I error is much greater than committing a Type II error, a larger value of α would be more appropriate. One solution is to set the levels of α and β such that their ratio is the same as the reciprocal ratio of the costs of the errors (Mapstone 1995) or, better yet, so that the overall expected cost of the errors is minimized (Field et al. 2004). We also draw attention to the fact that the expected cost of making a Type II error cannot be computed without specifying some prior expectation of the likelihood of the trend of interest (in our case, a pallid sturgeon population increase) (Field et al. 2004). This step is unfamiliar in the frequentist paradigm, but is a core component of Bayesian analysis, in which a prior expectation is combined with existing data to generate a posterior distribution for the focal parameter (Dorazio & Johnson 2003). A Bayesian approach is thus an attractive option for PSPAP – a set of decisions must be made about how to allocate resources between survey effort and management. These decisions will be made sequentially as the state of the system and the quality of information about it changes with time. Bayesian methods are becoming more accessible to ecologists and decision-makers in environmental management and we encourage members of PSPAP to investigate their application.

The need to combine a prior expectation with existing data has been recognized by some authors dealing with significance and power in frequentist statistics (e.g., Cascio & Zedeck 1983; Fox 2001).

In the frequentist paradigm, it is common practice to adhere to the so-called ‘five-eighty convention’ (Di Stefano 2003), in which the significance level, α, is fixed at 0.05 and a non-significant result is concluded if the ensuing statistical power is 0.8 or higher. Despite its popularity, a number of authors have noted how this may have disastrous consequences in the context of population or environmental management (see, for example, Gray 1990; Dayton 2001; Di Stefano 2001). The reason is that failing to detect a trend (a Type II error), particularly a negative trend, will result in inaction (Peterman 1990). On the other hand, erroneously concluding that there is a trend (a Type I error) will usually inflict short-term economic damage. Dayton (2001) therefore argued that the statistical ‘burden of proof’, which has been traditionally weighted against publishing spurious effects, should be flipped to ensure that potentially harmful or irreversible trends are not overlooked.

CONCLUSION

Demonstrating trends in survey data requires a long-term commitment to monitoring and consistent data collection. It may be too obvious to state that concentrating survey effort within a year is not adequate to generate the data necessary to evaluate trends across years, regardless of whether the interest is within segments, monitoring areas, or basin-wide. In the case of long-lived, rare species, trends of interest are likely to be of a magnitude of 1–3% per year or less (Urquhart et al. 1998). The objective of the PSPAP is timely detection of biologically significant trends; yet pallid sturgeon would most likely grow (or decline) at a lesser rate. On the other hand, population indices for fish can be so variable (see Fig’s 2-4) that a statistically significant “trend” of ±10% per year may occur over a period of only a few years due to factors other than a change in abundance. In that case, the threshold for statistical significance bears little relationship to biological significance. Thus, the greater challenge is not statistical power per se, but a better understanding of what is meaningful from a population perspective.

ACKNOWLEDGMENTS

This document was prepared in close collaboration with the United States Army Corps of Engineers, including Craig Fleming and Tim Welker. Yan Hong (MDMC) provided help with data access and formatting.

LITERATURE CITED

Arab, A., M. Wildhaber, C.K. Wikle, and C.N. Gentry. 2008. Zero-inflated modeling of fish catch per unit area resulting from multiple gears: application to channel catfish and shovelnose sturgeon in the Missouri River. North American Journal of Fisheries Management 28: 1044-1058.

Bates, D. and M. Maechler. 2009. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32. hyyp://CRAN.R-project.org/package=lme4.

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M. Henry, H. Stevens, and J.S. White. 2008. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution 24: 127-135.

Bryan, J.L., M.L. Wildhaber, D. Gladish, S. Holan, and M. Ellerseick. 2010. The power to detect trends in Missouri River fish populations within the Pallid Sturgeon Population Assessment Program: U.S. Geological Survey Open-File Report 2010-1020.

Burnham, K.P. and Anderson, D.R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach. 2d ed. New York: Springer-Verlag.

Cascio, W.F. and S. Zedeck. 1983. Open a new window in rational research planning: adjust alpha to maximize statistical power. Personnel Psychology 36: 517-526.

Cohen, J. 1962. The statistical power of abnormal-social psychological research: a review. Journal of Abnormal and Social Psychology 65 (3): 145-153.

Cohen, J. 1969. Statistical Power Analysis for the Behavioral Sciences. 1st Edition, Lawrence Erlbaum Associates, Hillsdale.

Dayton, P.K. 2001. Reversal of the burden of proof in fisheries management. Science 279: 821-822.

Di Stefano, J. 2001. Power analysis and sustainable forest management. Forest Ecology and Management 154: 141-153.

Di Stefano, J. 2003. How much power is enough? Against the development of arbitrary convention in statistical power calculations. Functional Ecology 17: 707-709.

Dorazio, R.M. and F.A. Johnson. 2003. Bayesian inference and decision theory – a framework for decision making in natural resource management. Ecological Applications 13: 556-563.

Elith, J. and J. Leathwick. 2009. The contribution of species distribution modelling to conservation prioritization. Pages 70-93 in A. Moilanen, K.A. Wilson, and H. Possingham (eds.) Spatial Conservation Prioritization: Quantitative Methods and Computational Tools. Oxford University Press Inc., New York.

Field, S.A., A.J. Tyre, N. Jonzen, J.R. Rhodes, and H. Possingham. 2004. Minimizing the cost of environmental management by optimizing statistical thresholds. Ecology Letters 7: 669-675.

Field, S.A., A.J. Tyre, and H. Possingham. 2005. Optimizing allocation of monitoring effort under economic and observational constraints. Journal of Wildlife Management 69 (2): 473-482.

Fox, D.R. 2001. Environmental power analysis – a new perspective. Environmetrics 12: 437-449.

Gray, J. 1990. Statistics and the precautionary principle. Marine Pollution Bulletin 21: 174-176.

Gray, B.R. and M.M. Burlew. 2007. Estimating trend precision and power to detect trends across grouped count data. Ecology 88(9): 2364-2372.

Hojsgaard, S. and U. Halekoh. 2011. doBy: Groupwise summary statistics, general linear contrasts, LSMEANS (least-squares-means), and other utilities. R package version 4.4.0.

Hox, J. 2002. Multilevel analysis: techniques and applications. Lawrence Erlbaum, Mahwah, New Jersey.

Mapstone, B. 1995. Scalable decision rules for environmental impact studies: effects size, type 1, and type 2 errors. Ecological Applications 5: 410-410.

Peterman, R. 1990. The importance of reporting statistical power: the forest decline and acidic deposition example. Ecology 71: 2024-2027.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar and the R Core Team. 2009. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-96.

Peery, M.Z. 2004. Power to detect trends in pallid and shovelnose sturgeon populations in the Missouri River. Sustainable Ecosystems Institute, Portland, Oregon.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available from URL: http://www.r-project.org/foundation.

Raudenbush, S.W. and A. S. Bryk. 2002. Hierarchical linear models: applications and data analysis methods. Second Edition. Sage, Thousand Oaks, California.

Snijders, T.A.B. and R. J. Bosker. 1993. Standard errors and sample sizes for two-level research. Journal of Educational Statistics 18: 237-259.

Snijders, T.A.B. and R. J. Bosker. 1999. Multilevel analysis. Sage Publications, London.

Szabo, J.K., P.D. Davy, M.J. Hooper, & L.B. Astheimer. 2009. Predicting avian distributions to evaluate spatiotemporal overlap with locust control operations in eastern Australia. Ecological Applications 19(8): 2026-2037.

Tierney, L., A.J. Rossini, N. Li, and H. Sevcikova. 2011. snow: Simple Network of Workstations. R package version 0.3-3.

Urquhart, N.S., S.G. Paulsen, and D.P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications 8: 246-257.

Welker, T.L. and M.R. Drobish. (eds). 2010. Missouri River standard operating procedures for sampling and data collection. Vol. 1.5, U.S. Army Corps of Engineers, Omaha District, Yankton, SD.

Whittingham, M.J., P.A. Stevens, R.B. Bradbury, and R.P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behavior. Journal of Animal Ecology 75: 1182-1189.

Table 1. Model set and model selection results. Models with ΔAIC values < 4 are shown in bold text.

	Upper [A]		Middle [B]		Lower [C]	
Model*	ΔAIC	Weight	ΔAIC	Weight	ΔAIC	Weight
Gear	265.05	1.47 E-58	8.27	1.20 E-02	109.81	7.15 E-25
Gear + Segment × Year	130.84	2.04 E-29	7.15	2.10 E-02	13.22	6.75 E-04
Gear + Segment × Year + Month	**2.58**	**1.46 E-01**	4.59	7.54 E-02	14.08	4.39 E-04
Gear + Segment × Year + Mesohabitat	128.70	5.95 E-29	9.17	7.63 E-03	5.01	4.09 E-02
Gear + Segment × Year + Temperature	89.69	1.77 E-20	9.05	8.11 E-03	**0.40**	**4.11 E-01**
Gear + Segment × Year + Month + Mesohabitat	**0.00**	**5.28 E-01**	6.63	2.73 E-02	5.55	3.12 E-02
Gear + Segment × Year + Month + Temperature	**3.29**	**1.02 E-01**	**0.00**	**7.50 E-01**	9.29	4.82 E-03
Gear + Segment × Year + Mesohabitat × Temperature	89.04	2.44 E-20	13.11	1.07 E-03	**0.00**	**5.01 E-01**
Gear + Segment × Year + Month + Mesohabitat × Temperature	**1.71**	**2.25 E-01**	4.07	9.79 E-02	7.85	9.91 E-03

* Fixed effects shown only: all models included the random effect (1|Bend) + (1|SubSample) and the offset (SetEffort).

[A] Segments 2,3,4

[B] Segments 5,6

[C] Segments 7,8,9,10,13,14

Table 2. Sampling effort (# Bends/Segment/Year - # Gear deployments/macro-habitat) by Scenario. Scenario 2 represents the existing design.

Segment	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Scenario 9
2	12-1	12-2	12-3	11-1	11-2	11-3	13-1	13-2	13-3
3	21-1	21-2	21-3	19-1	19-2	19-3	23-1	23-2	23-3
4	12-1	12-2	12-3	11-1	11-2	11-3	13-1	13-2	13-3
5-6	10-1	10-2	10-3	9-1	9-2	9-3	11-1	11-2	11-3
7	12-1	12-2	12-3	11-1	11-2	11-3	13-1	13-2	13-3
8	15-1	15-2	15-3	14-1	14-2	14-3	16-1	16-2	16-3
9	20-1	20-2	20-3	18-1	18-2	18-3	22-1	22-2	22-3
10	10-1	10-2	10-3	9-1	9-2	9-3	11-1	11-2	11-3
13	11-1	11-2	11-3	10-1	10-2	10-3	12-1	12-2	12-3
14	14-1	14-2	14-3	13-1	13-2	13-3	15-1	15-2	15-3

Table 3. Parameter estimates from the AICc best model-Upper Monitoring Area. The intercept is the log(Set Effort) for gear type MFS fished in 2006 in Segment 2. All other estimates are coded as deviations from the intercept on a log scale. Rows in bold indicate significance at the α =0.05 level.

Fixed Effect	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-19.94	382.92	-0.05	0.96
Gear: OT16S	3.79	3.56	1.06	0.29
Gear: TLC1S	5.78	3.56	1.63	0.10
Gear: TNS	3.98	3.56	1.12	0.26
Segment 3	**1.16**	**0.36**	**3.19**	**0.00**
Segment 4	**0.89**	**0.39**	**2.26**	**0.02**
Year: 2007	0.40	0.41	0.96	0.34
Year: 2008	-0.16	0.50	-0.31	0.76
Year: 2009	0.62	0.41	1.53	0.13
Year: 2010	**1.02**	**0.38**	**2.70**	**0.01**
Mesohabitat: CHNB	1.07	3.56	0.30	0.76
Mesohabitat: ITIP	1.39	3.56	0.39	0.70
Mesohabitat: TLWG	2.93	3.61	0.81	0.42
Segment 3: 2007	-0.41	0.46	-0.90	0.37
Segment 4: 2007	-0.29	0.49	-0.60	0.55
Segment 3: 2008	0.51	0.53	0.97	0.33
Segment 4: 2008	0.74	0.54	1.37	0.17
Segment 3: 2009	-0.36	0.44	-0.81	0.42
Segment 4: 2009	0.74	0.46	1.62	0.11
Segment 3: 2010	-0.64	0.41	-1.57	0.12
Segment 4: 2010	**0.92**	**0.42**	**2.20**	**0.03**

Table 4. Parameter estimates from the AICc best model-Middle Monitoring Area. The intercept is the log(Set Effort) for gear type GN14S fished in 2005 in Segment 5. All other details as in Table 3.

Fixed Effect	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-19.71	2557.00	-0.01	0.99
Gear: GN41S	0.00	0.21	0.01	1.00
Gear: MFS	-17.95	839.90	-0.02	0.98
Gear: OT16S	**-1.14**	**0.49**	**-2.35**	**0.02**
Gear: TLC1S	**3.26**	**0.26**	**12.76**	**0.00**
Gear: TNS	-0.66	0.50	-1.33	0.18
Segment 6	-0.35	0.89	-0.39	0.70
Year: 2006	-0.23	0.69	-0.33	0.74
Year: 2007	**1.09**	**0.54**	**2.01**	**0.04**
Year: 2008	0.19	0.56	0.34	0.74
Year: 2009	0.28	0.55	0.52	0.61
Year: 2010	-0.10	0.55	-0.19	0.85
Temperature	**0.07**	**0.03**	**2.54**	**0.01**
Segment 6:2006	0.94	1.03	0.92	0.36
Segment 6:2007	0.88	0.91	0.97	0.33
Segment 6:2008	**2.07**	**0.93**	**2.23**	**0.03**
Segment 6:2009	1.11	0.91	1.21	0.23
Segment 6:2010	1.18	0.92	1.28	0.20

Table 5. Parameter estimates from the AICc best model-Lower Monitoring Area. The intercept is the log(Set Effort) for gear type GN14S fished in 2005 in Segment 7. All other details as in Table 3.

Fixed Effect	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	**-6.68**	**2.62**	**-2.55**	**0.01**
Gear: GN18S	**1.77**	**0.45**	**3.98**	**0.00**
Gear: GN41S	0.51	0.55	0.94	0.35
Gear: GN81S	**1.94**	**0.44**	**4.36**	**0.00**
Gear: MFS	-15.78	393.55	-0.04	0.97
Gear: OT16S	0.50	0.46	1.09	0.28
Gear: TLC2S	**3.99**	**0.46**	**8.73**	**0.00**
Gear: TN25S	**1.21**	**0.53**	**2.27**	**0.02**
Gear: TNS	**1.27**	**0.45**	**2.81**	**0.00**
Segment 8	0.13	0.68	0.19	0.85
Segment 9	-0.04	0.69	-0.05	0.96
Segment 10	0.04	0.81	0.05	0.96
Segment 13	0.32	0.67	0.47	0.64
Segment 14	-0.58	0.72	-0.80	0.43
Year: 2006	-0.59	0.90	-0.65	0.51
Year: 2007	**2.18**	**0.70**	**3.13**	**0.00**
Year: 2008	**1.87**	**0.70**	**2.66**	**0.01**
Year: 2009	**1.53**	**0.71**	**2.14**	**0.03**
Year: 2010	0.51	0.79	0.64	0.52
Mesohabitat: CHNB	-0.01	2.50	-0.01	1.00
Mesohabitat: ITIP	-0.10	2.71	-0.04	0.97
Mesohabitat: POOL	0.41	2.50	0.17	0.87
Mesohabitat: TLWG	-1.41	5.73	-0.25	0.81
Temperature	0.04	0.14	0.30	0.76
Segment 8: 2006	0.22	0.95	0.23	0.82
Segment 9: 2006	1.04	0.93	1.12	0.26
Segment 10: 2006	1.22	1.04	1.17	0.24
Segment 13: 2006	-0.18	0.98	-0.18	0.86
Segment 14: 2006	0.26	1.01	0.26	0.79
Segment 8: 2007	-0.77	0.72	-1.07	0.29
Segment 9: 2007	-1.22	0.73	-1.68	0.09
Segment 10: 2007	-1.64	0.87	-1.87	0.06
Segment 13: 2007	**-2.73**	**0.76**	**-3.59**	**0.00**
Segment 14: 2007	-1.61	0.81	-1.99	0.05
Segment 8: 2008	-0.34	0.74	-0.47	0.64
Segment 9: 2008	-0.67	0.75	-0.90	0.37
Segment 10: 2008	-1.40	0.90	-1.55	0.12
Segment 13: 2008	**-1.95**	**0.80**	**-2.43**	**0.01**
Segment 14: 2008	-1.53	0.91	-1.69	0.09
Segment 8: 2009	-0.36	0.75	-0.47	0.64
Segment 9: 2009	-0.55	0.76	-0.73	0.46
Segment 10: 2009	-1.04	0.90	-1.16	0.25
Segment 13: 2009	**-2.21**	**0.84**	**-2.61**	**0.01**
Segment 14: 2009	-0.74	0.85	-0.87	0.38
Segment 8: 2010	0.35	0.82	0.43	0.67
Segment 9: 2010	-0.43	0.82	-0.53	0.60
Segment 10: 2010	-0.88	0.94	-0.93	0.35
Segment 13: 2010	-1.31	0.82	-1.60	0.11
Segment 14: 2010	-1.31	0.90	-1.46	0.14
Mesohabitat CHNB: Temp.	-0.01	0.14	-0.09	0.93
Mesohabitat ITIP: Temp.	-0.02	0.15	-0.13	0.90
Mesohabitat POOL: Temp	-0.01	0.14	-0.10	0.92
Mesohabitat TLWG: Temp	0.10	0.24	0.43	0.67

Table 6. The proportion of replicates (out of 100) where the resulting t-statistic ≥ 1.96. MA = Monitoring Area; YOY = young-of-year fish; JUV = juveniles; AD = adults; * = insufficient catch data. See Table 2 for Scenario descriptions.

MA-Size Class-% Annual Increase	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Scenario 9
Upper-YOY-1%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Upper-YOY-5%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
Upper-YOY-10%	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.35
Upper-JUV-1%	0.01	0.03	0.14	0.01	0.01	0.01	0.16	0.27	0.56
Upper-JUV-5%	0.08	0.13	0.45	0.03	0.05	0.05	0.24	0.65	0.92
Upper-JUV-10%	0.29	0.38	0.76	0.18	0.23	0.33	0.57	0.95	0.99
Upper-AD-1%	*	*	*	*	*	*	*	*	*
Upper-AD-5%	*	*	*	*	*	*	*	*	*
Upper-AD-10%	*	*	*	*	*	*	*	*	*
Middle-YOY-1%	*	*	*	*	*	*	*	*	*
Middle-YOY-5%	*	*	*	*	*	*	*	*	*
Middle-YOY-10%	*	*	*	*	*	*	*	*	*
Middle-JUV-1%	0.08	0.08	0.17	0.00	0.05	0.14	0.20	0.41	0.62
Middle-JUV-5%	0.15	0.28	0.37	0.02	0.13	0.24	0.41	0.73	0.74
Middle-JUV-10%	0.36	0.40	0.57	0.06	0.31	0.41	0.68	0.98	0.89
Middle-AD-1%	*	*	*	*	*	*	*	*	*
Middle-AD-5%	*	*	*	*	*	*	*	*	*
Middle-AD-10%	*	*	*	*	*	*	*	*	*
Lower-YOY-1%	0.00	0.00	0.01	0.01	0.00	0.03	0.05	0.05	0.00
Lower-YOY-5%	0.01	0.01	0.08	0.01	0.01	0.06	0.08	0.03	0.08
Lower-YOY-10%	0.09	0.05	0.14	0.09	0.05	0.13	0.09	0.14	0.36
Lower-JUV-1%	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.03	0.16
Lower-JUV-5%	0.04	0.19	0.20	0.03	0.09	0.11	0.05	0.20	0.58
Lower-JUV-10%	0.18	0.44	0.50	0.14	0.29	0.38	0.31	0.60	0.88
Lower-AD-1%	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Lower-AD-5%	0.01	0.01	0.10	0.00	0.01	0.10	0.01	0.04	0.05
Lower-AD-10%	0.03	0.21	0.33	0.05	0.10	0.15	0.13	0.38	0.13

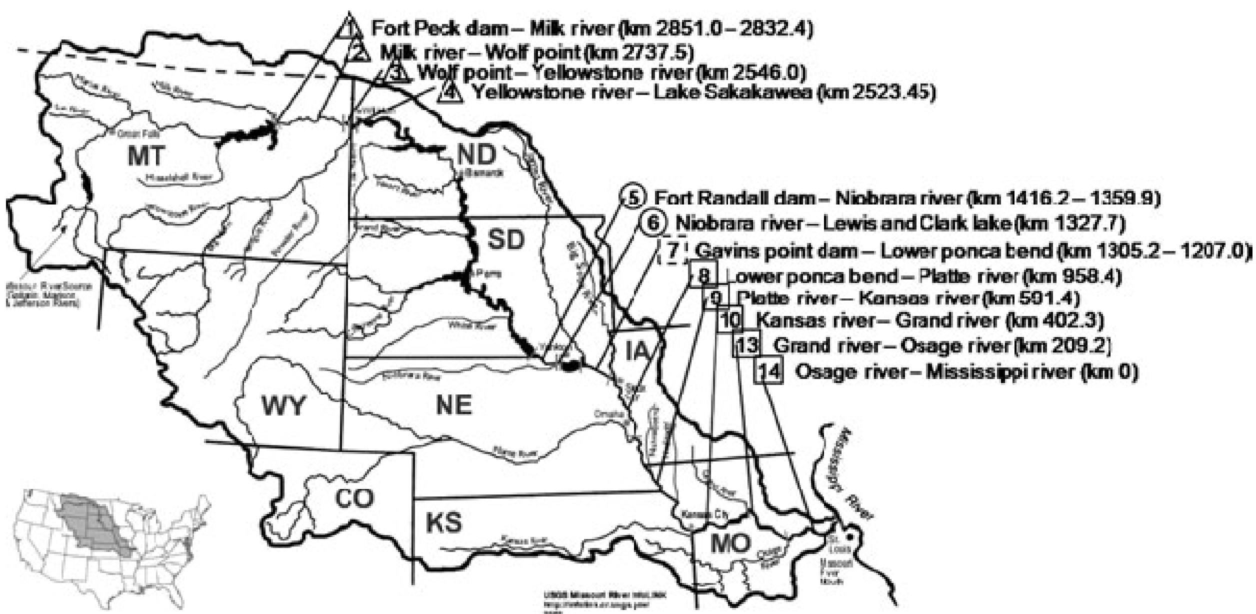

Figure 1. The Missouri River Basin. Segments are numbered and monitoring areas are represented as follows: triangles represent the Upper Monitoring Area; circles represent the Middle Monitoring Area; squares represent the Lower Monitoring Area.

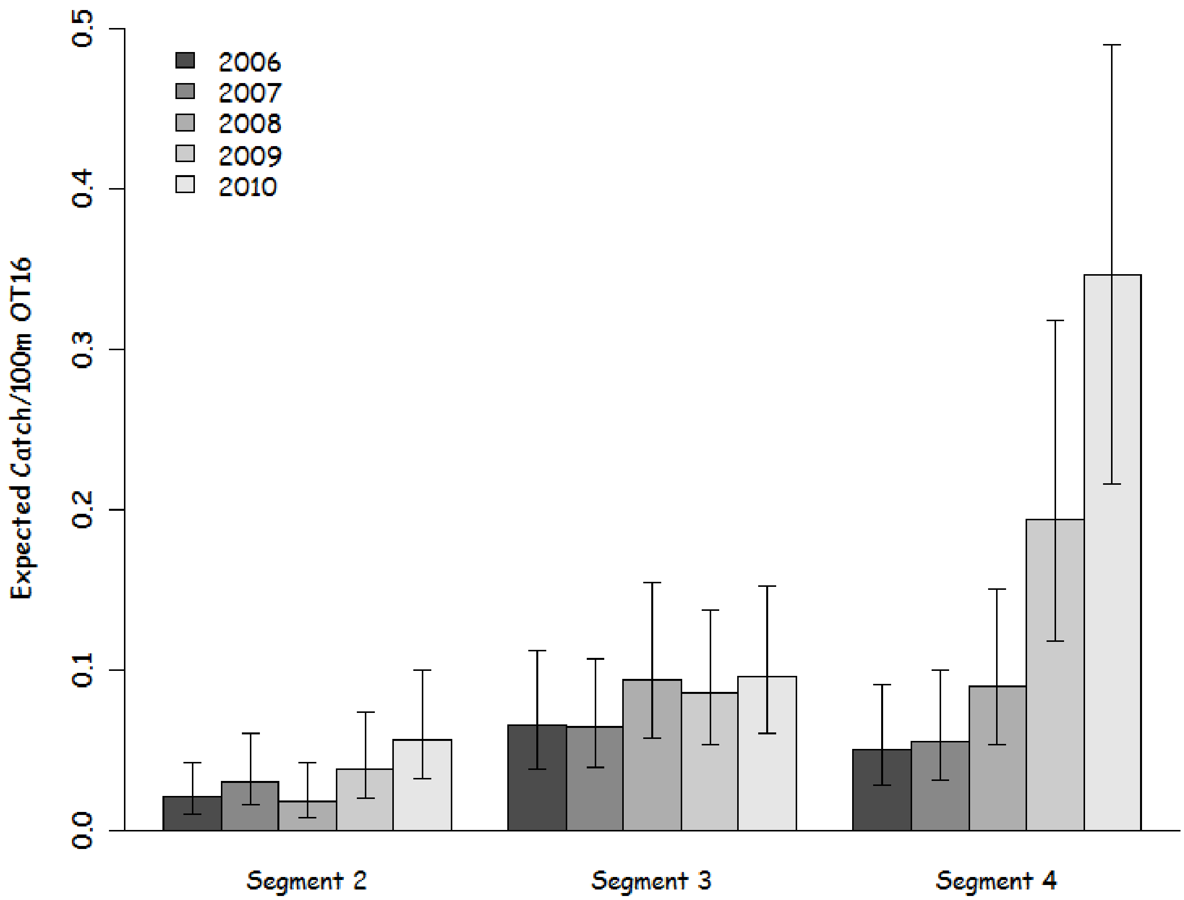

Figure 2. Upper Monitoring Area: predicted catch of pallid sturgeon. Error bars are approximate 95% confidence intervals.

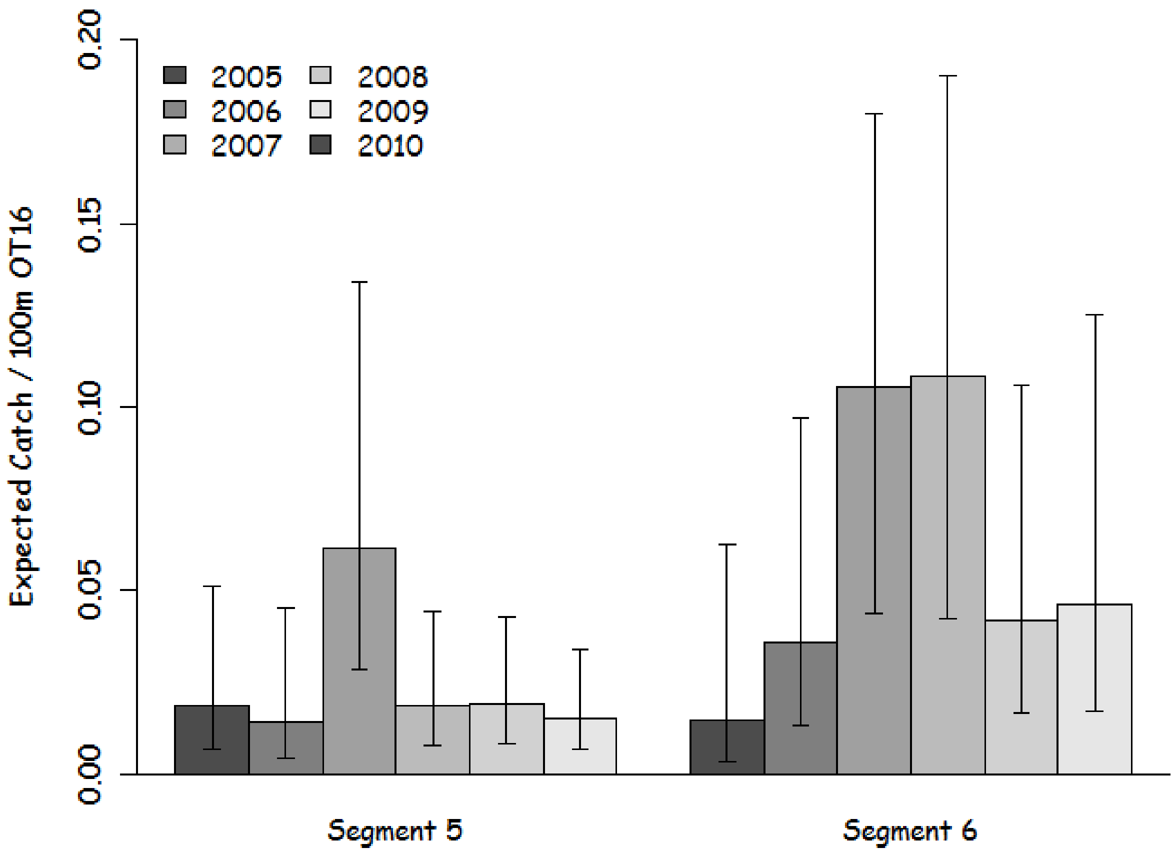

Figure 3. Middle Monitoring Area: predicted catch of pallid sturgeon.
Error bars are approximate 95% confidence intervals.

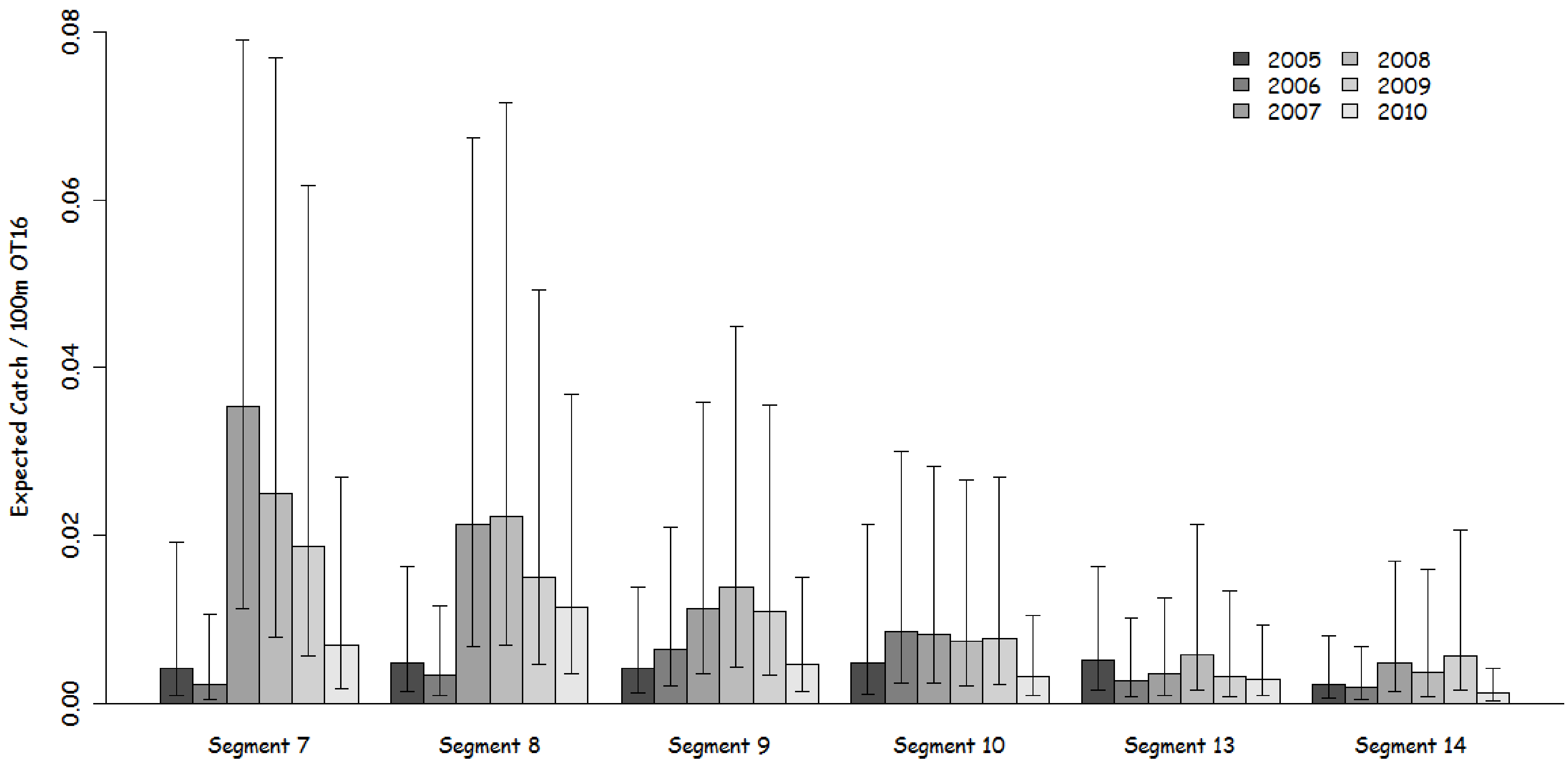

Figure 4. Lower Monitoring Area: expected catch of pallid sturgeon.
Error bars are approximate 95% confidence intervals.

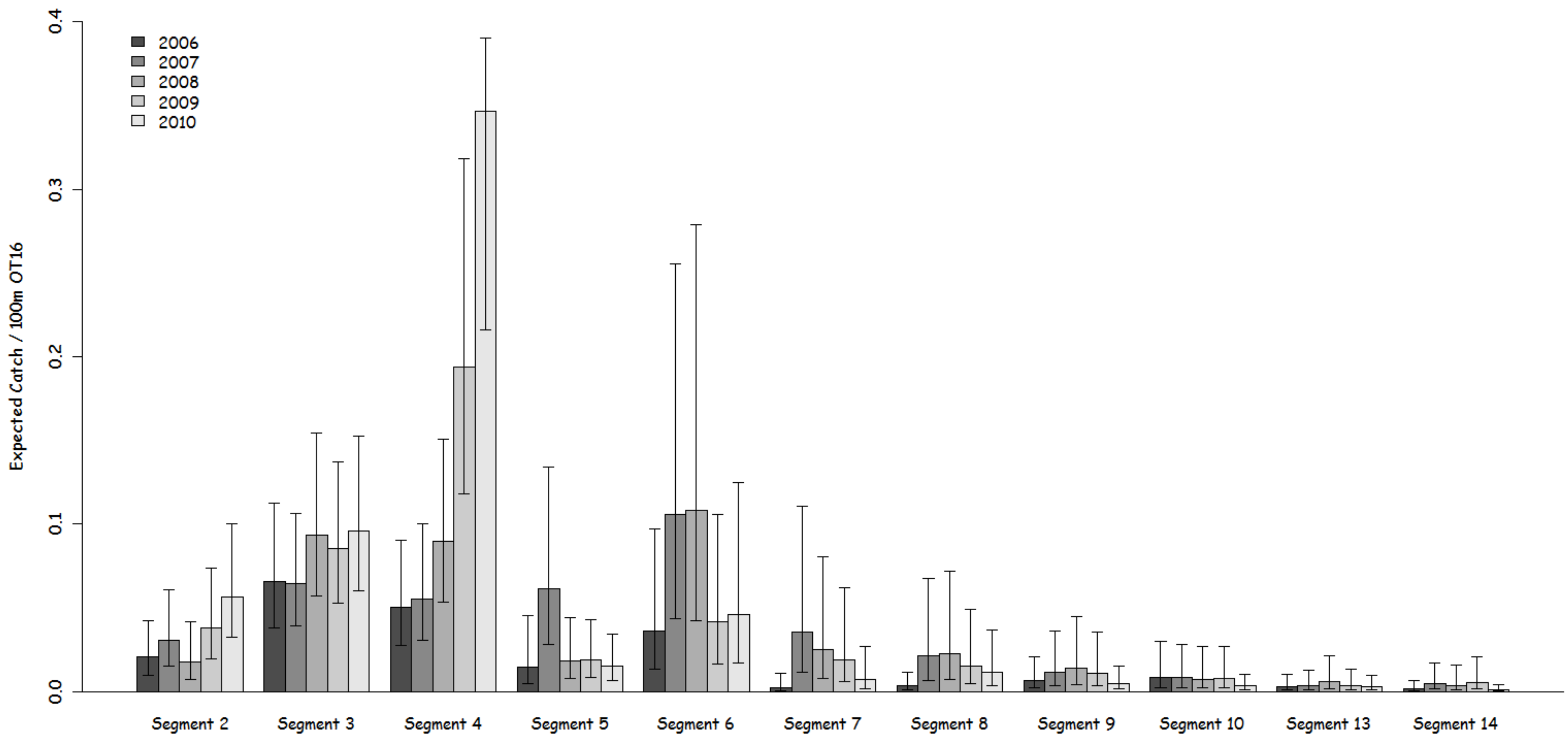

Figure 5. All sampling segments: expected catch of pallid sturgeon.
Error bars are approximate 95% confidence intervals.

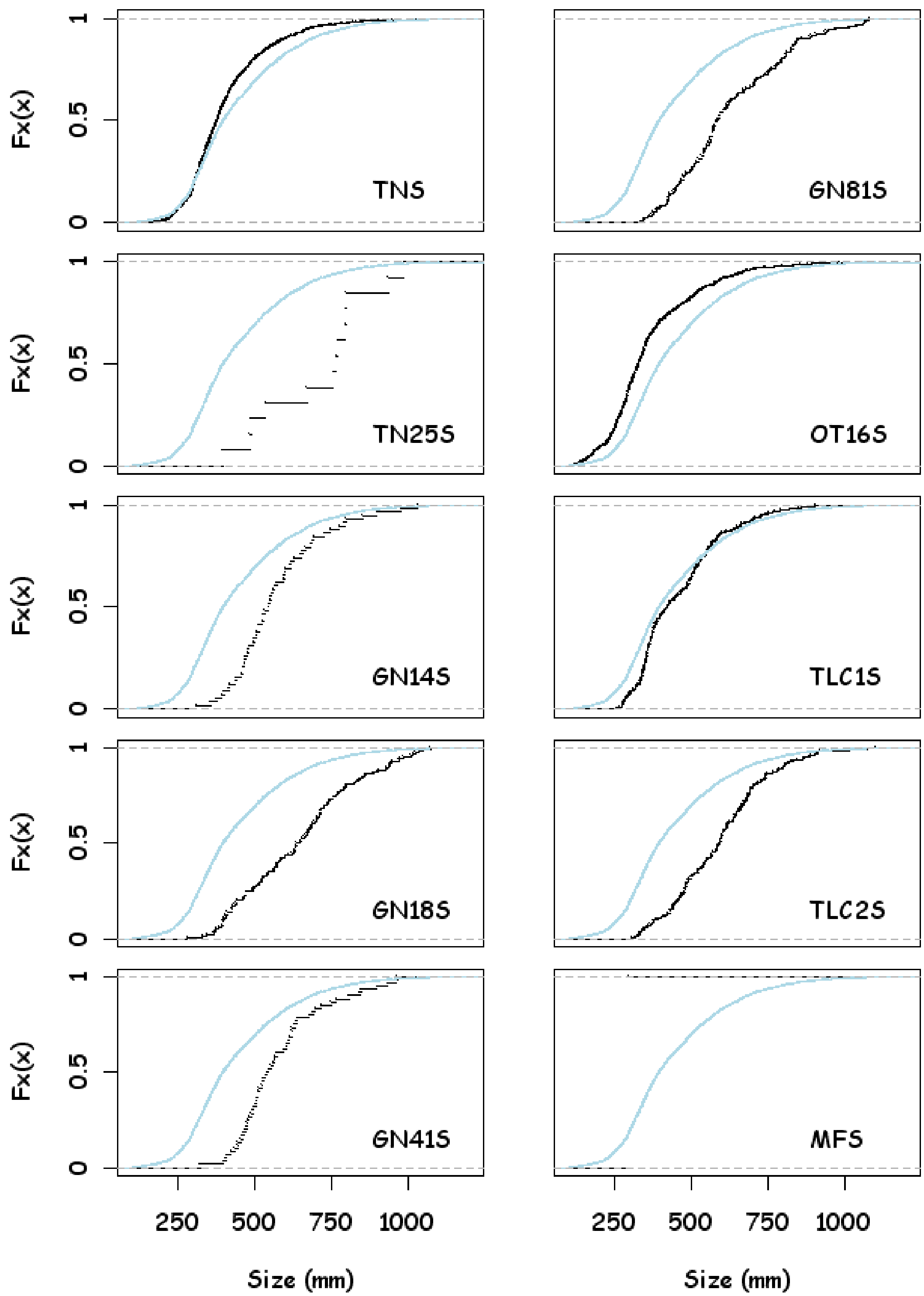

Figure 6. Comparisons of the size CDF for all gears utilized in the Pallid Sturgeon Population Assessment Program. In each plot, the size CDF for all gears is indicated by the blue curve.

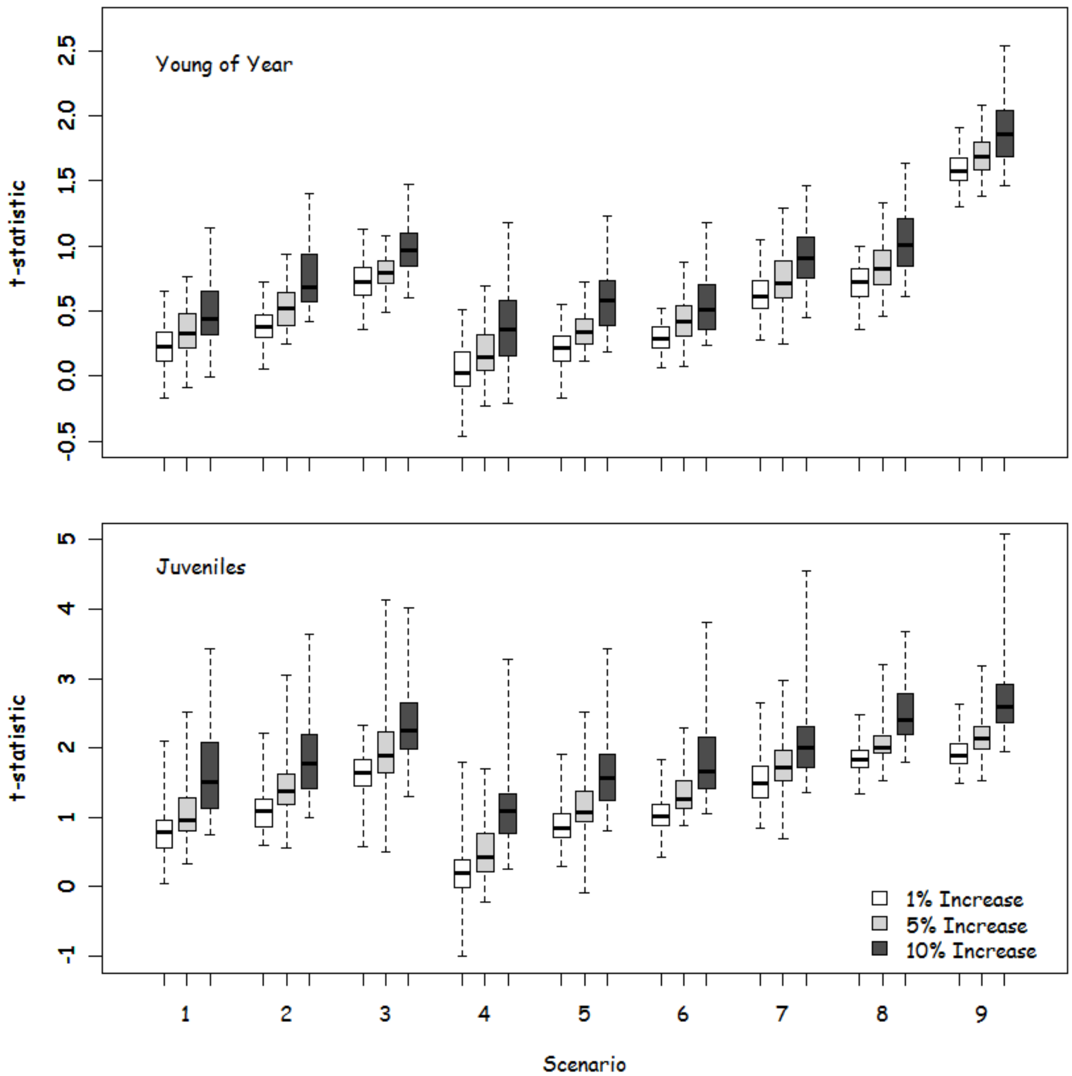

Figure 7. Boxplots of the *t*-statistics for the simulated 2011 and 2020 yearly coefficients in the Upper Monitoring Area (Segments 2, 3, 4). Three compounding rates of increase are shown. Bold lines indicate the median of 100 simulations, boxes are the inter-quartile range (25^{th} -75^{th} percentile), and the whiskers indicate the range of the data. See Table 2 for Scenario descriptions.

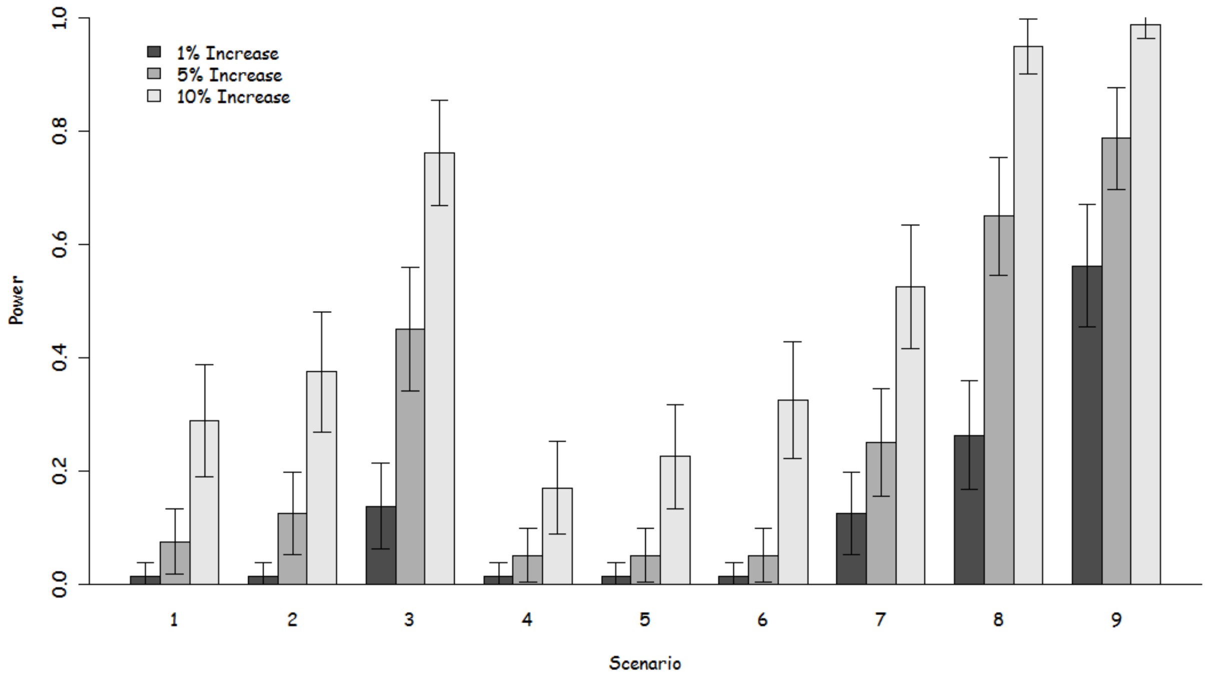

Figure 8. Power (Pr [t > 1.96]) to detect changes in relative abundance in juvenile fish over a ten-year period (Upper Monitoring Area). Three compounding rates of increase are shown and results are from 100 simulations. Error bars are binomial proportion confidence intervals. See Table 2 for Scenario descriptions.

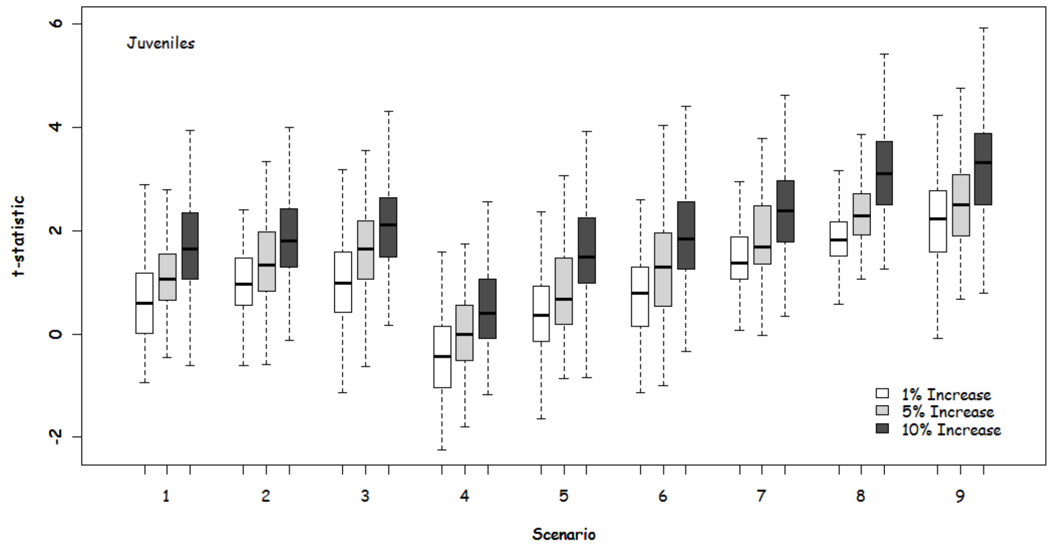

Figure 9. Boxplot of the *t*-statistics for the simulated 2011 and 2020 yearly coefficients for the Middle Monitoring Area (Segments 5 and 6). All other details as in Fig 7.

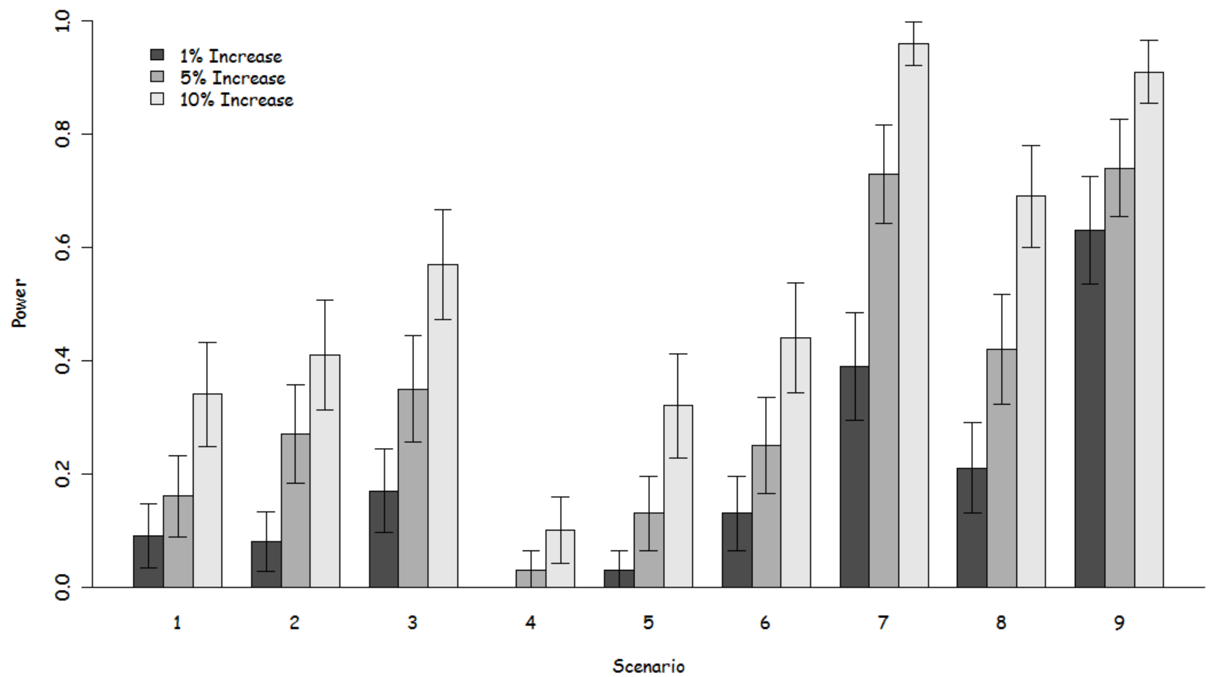

Figure 10. Power (Pr [t > 1.96]) to detect changes in relative abundance in juvenile fish over a ten-year period (Middle Monitoring Area). All other details as in Figure 8.

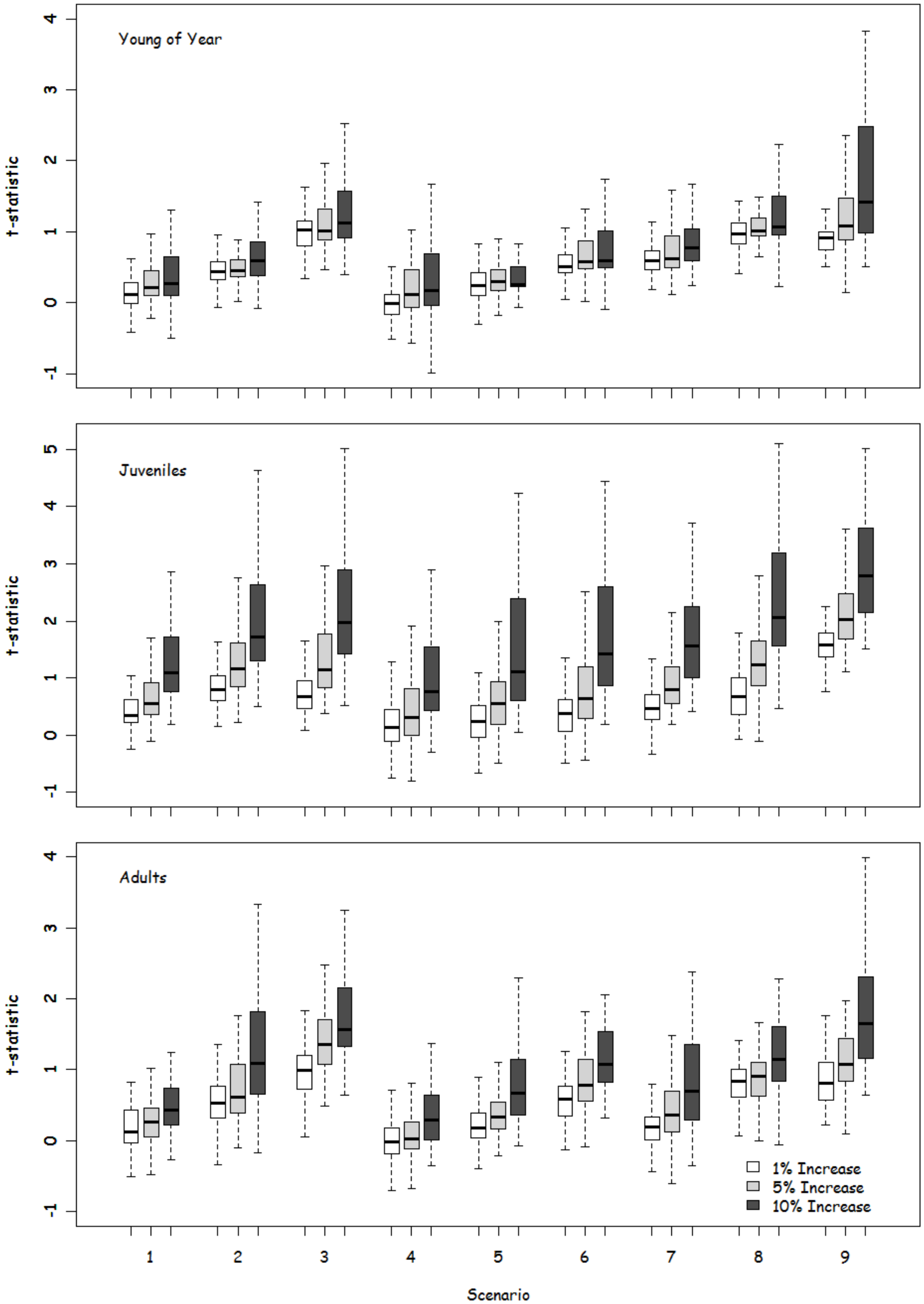

Figure 11. Boxplots of the *t*-statistics for the simulated 2011 and 2020 yearly coefficients for the Lower Monitoring Area. All other details as in Fig 7.

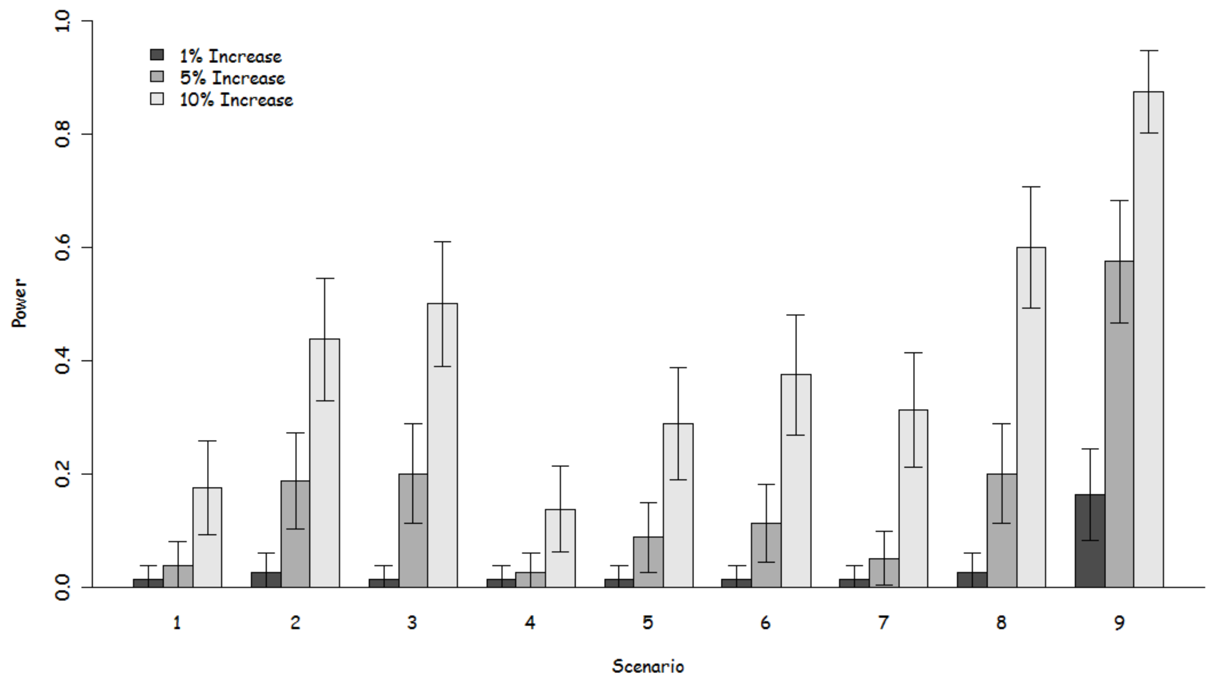

Figure 12. Power (Pr [t > 1.96]) to detect changes in relative abundance in juvenile fish over a ten-year period (Lower Monitoring Area). All other details as in Figure 8.