# POPULATION ABUNDANCE OF SPAWNING LAKE STURGEON IN THE ST. LAWRENCE RIVER

BY

JOHN D. SKELTON

B.S., SUNY COBLESKILL, 2022

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biology
in the Biology Department of the
State University of New York at Oneonta
2024

Approved by: ____________________________________
Dr. Daniel Stich
DATE
Associate Professor
State University of New York,
Oneonta, New York

____________________________________
Dr. Miranda Kearney
DATE
Assistant Professor
State University of New York,
Oneonta, New York

____________________________________
Dr. Alex Sotola
DATE
Assistant Professor
State University of New York,
Oneonta, New York

Contents

## Preface

### Objective

In this project, the primary goal is to analyze video footage of spawning Lake Sturgeon (*Acipenser fulvescens*) provided by the New York State Department of Environmental Conservation (NYSDEC) to empirically estimate the abundance of spawning Lake Sturgeon near the Iroquois Dam on two artificial spawning beds located in the St. Lawrence River and to determine detection probability of sturgeon. We expect that this process will help strengthen the detection of sturgeon on the spawning beds. We also expect that from 2011 to 2022, there will be fewer sturgeon each year on the spawning beds. This is because raw data collected throughout the years indicate a decrease in sturgeon observed. This project is relevant because it will assist the NYSDEC in determining the accuracy of their current method of population monitoring, and its effectiveness under differing abiotic conditions. Another significant impact of this project is that it will provide a robust estimate for the number of Lake Sturgeon coming to the spawning beds each year. This will help determine the long-term effectiveness of the artificial spawning beds. The primary products of this project will be open-sourced R code for replication either for this dataset or future datasets, as well as open-source simulation R code that will be used to validate analysis and inform management decisions. It is also expected that a final draft in the format of a draft manuscript will be submitted to NYSDEC.

### Background

Sturgeon are a very old group of fish, especially when compared to other more derived fish groups with fossil evidence suggesting that sturgeon evolved into the species we see today around the late Jurassic Period (Peterson et al. 2007). Drastic changes in environmental conditions such as overharvest and the installation of hydroelectric dams can affect these sensitive species particularly because they have longer life expectancies and mature much later in their life when compared to other fishes (Peterson et. al 2007; Environnement Illimité Inc 2010).

In modern times, human activities have caused major disturbances to sturgeon habitat. A particular threat that can be detrimental to sturgeon is hydroelectric dams (Environnement Illimité Inc. 2010). Dams are known to have significant effects on many fish species, specifically preventing migration and spawning, which can have a negative effect on migratory species (e.g. Caudill et. al 2013, Zydlewski et. al 2021). Sturgeon are largely dependent on large rocks and cobbles as spawning substrate (Peterson et. al 2007). If these substrates are not available they are unable to spawn. When dams are installed, the flows in dammed rivers are altered. This can cause suspended sediments and debris to fall out of the water column onto the spawning habitat. Additionally, with increased debris falling onto the spawning beds, eggs that have been laid can be buried and suffocate, killing the maturing embryos. Dams can also remove essential substrate constituents through scouring, or increase or decrease water depths to such a degree that spawning areas are no longer suitable, meaning sturgeon cannot reproduce in these areas. Because sturgeon need a long time to mature, reductions in reproductive success can have a proportionally larger impact on their populations than fishes with faster reproductive cycles or more generalized spawning habitat requirements.

The Lake Sturgeon (*Acipenser fulvescens*) is the only member of the *Acipenser* genus that spends its entire life cycle in freshwater (Peterson et al. 2007). Like other sturgeon species, they prefer to spawn on large gravel beds, with females releasing up to 680,000 eggs, which take around 8-14 days to hatch, fall between the spaces of the substrate and attach to the rocks due to their adheisive nature (Werner 2004; (Peterson et al. 2007). Lake Sturgeon Lake Sturgeon do not spawn in consecutive years, in fact it is typical for female sturgeon to have a four-year absence before returning to spawn again (Werner 2004). Thus, the negative impacts on population sizes from the reduction in reproductive success as a result of damming can be even more pronounced over the short temporal scales used by managers (e.g, 10 years) given the periodic nature of spawning in this species.

Lake Sturgeon are listed as Threatened by the New York State Department of Environmental Conservation (NYSDEC) (NYSDEC 2018). Currently there are seven extant populations of Lake Sturgeon in New York State managed by NYSDEC. All of these populations have undergone drastic reductions in numbers primarily because of overfishing and the installment of dams, but are now stable or slowly recovering (Peterson et al. 2007; NYSDEC 2018). The population of Lake Sturgeon in the St. Lawrence River actively spawn in the river; however, the New York Power Authority (NYPA) installed a series of dams that altered the spawning habitat in the river 1953-1959, making much of it inaccessible to Lake Sturgeon (Environnement Illimite Inc. 2010; Federal Energy Regulatory Commission [FERC] 2003). Largely due to recommendations from NYSDEC, NYPA laid large cobbles and stones into areas of the river where falling sediment and algal growth would not occur on these introduced rocks, essentially creating several artificial spawning beds for this Lake Sturgeon population. After the installation, the project was transferred to NYSDEC in 2008, who monitored spawning sturgeon on these artificial spawning beds by transect surveys using underwater cameras. In addition to capturing pictures of sturgeon, NYSDEC collected additional information including time of day, temperature, discharge, date, and geographic coordinates. Originally, sturgeon heavily used the spawning habitat provided by NYPA, but over time, it appears that Lake Sturgeon abundance on the beds has steadily decreased, causing concern for NYSDEC (Environnement Illimite Inc. 2010, NYSDEC 2021). It is unknown to what degree this pattern is attributable to observation error or underlying population trends or abiotic conditions. Furthermore, it is unknown how much field sampling effort is needed to accurately assess the status of the spawning population from year to year. These factors make understanding population trends in this population of Lake Sturgeon difficult at best. The primary focus of this study is to create statistical models of Lake Sturgeon abundance to address these specific unknowns. This project will assist the NYSDEC in making more informed management decisions in the future for this specific population. The approach applied to this population is broadly applicable and can be used in a multitude of situations regarding Lake Sturgeon management.

## Proposed Methods

### *Study Area*

The areas of study are adjacent to the Iroquois Dam on the St. Lawrence River (Figure 1). Sampling was performed by the New York State Department of Environmental Conservation (NYSDEC) between 2011 and 2022. Two spawning habitats were sampled, one upstream of the Iroquois Dam and the other downstream. Five 50-m transects were used on each spawning habitat, resulting in a total of ten transects. On each of these transects, underwater cameras were used to view adult sturgeon on spawning habitat. During this sampling process, NYSDEC also recorded abiotic factors including: water temperature, date, discharge and geographic locations of observed sturgeon. After sampling, NYSDEC then reviewed the recorded footage and identified how many sturgeon were observed on each transect.

*Data Processing*

The NYSDEC data will be processed and evaluated using the R programming language (R Core team 2022). Specifically, abundance of sturgeon will be estimated using Bayesian hierarchical models of abundance (e.g., Royle et al. 2014) that account for imperfect detection within the JAGS software package (Plummer 2003). The resulting framework will then be validated through simulation to determine sampling intensity or designs necessary for obtaining accurate and precise estimates of abundance.

**Literature Cited**

Caudill, C. C., Keefer, M..L., Clabough, T.S., Naughton, G.P., Burke, B.J., Peery, C.A. 2013. Indirect effects of impoundment on migrating fish: temperature gradients in fish ladders slow dam passage by adult Chinook salmon and steelhead. PloS one 8:e8556-e85586. Public Library of Science, United States.

Environnement Illimité Inc. 2010. Investigation of Lake Sturgeon spawning activities at Iroquois Dam on the St. Lawrence River in 2009. Prepared for the New York Power Authority, White Plains, NY.

New York State Department of Environmental Conservation (NYSDEC). 2018. Lake Sturgeon recovery plan 2018-2024. NYSDEC, Albany, NY. Available: https://www.dec.ny.gov/outdoor/111557.html (February 2022).

New York State Department of Environmental Conservation (NYSDEC). 2021. Lake Sturgeon population assessment report. NYSDEC, Albany, NY. Available:

https://www.dec.ny.gov/docs/fish_marine_pdf/lakesturgeonassessment.pdf (February 2022).

Peterson, D. L., Vecsei, P., Jennings, C. A. 2007. Ecology and biology of the Lake Sturgeon: a synthesis of current knowledge of a threatened North American Acipenseridae. Reviews in Fish Biology and Fisheries 17:59-76.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108-115.

R Core Team. 2022. R: a language environment for statistical computing. R Foundation for Statistical Programming. Vienna, Australia. Available: https://www.R-project.org/.

Werner, R. G. 2004. A field guide freshwater fishes of the northeast. 1st edition. Syracuse University Press.

Zydlewski, J., Stich, D.S., Roy, S., Bailey, M., Sheehan, T., Sprankle, K., 2021. What have we lost? Modeling dam impacts on American Shad populations throughout their native range. Frontiers in Marine science 8:https://www.frontiersin.org/article/10.3389/fmars.2021.734213.

**Figures**

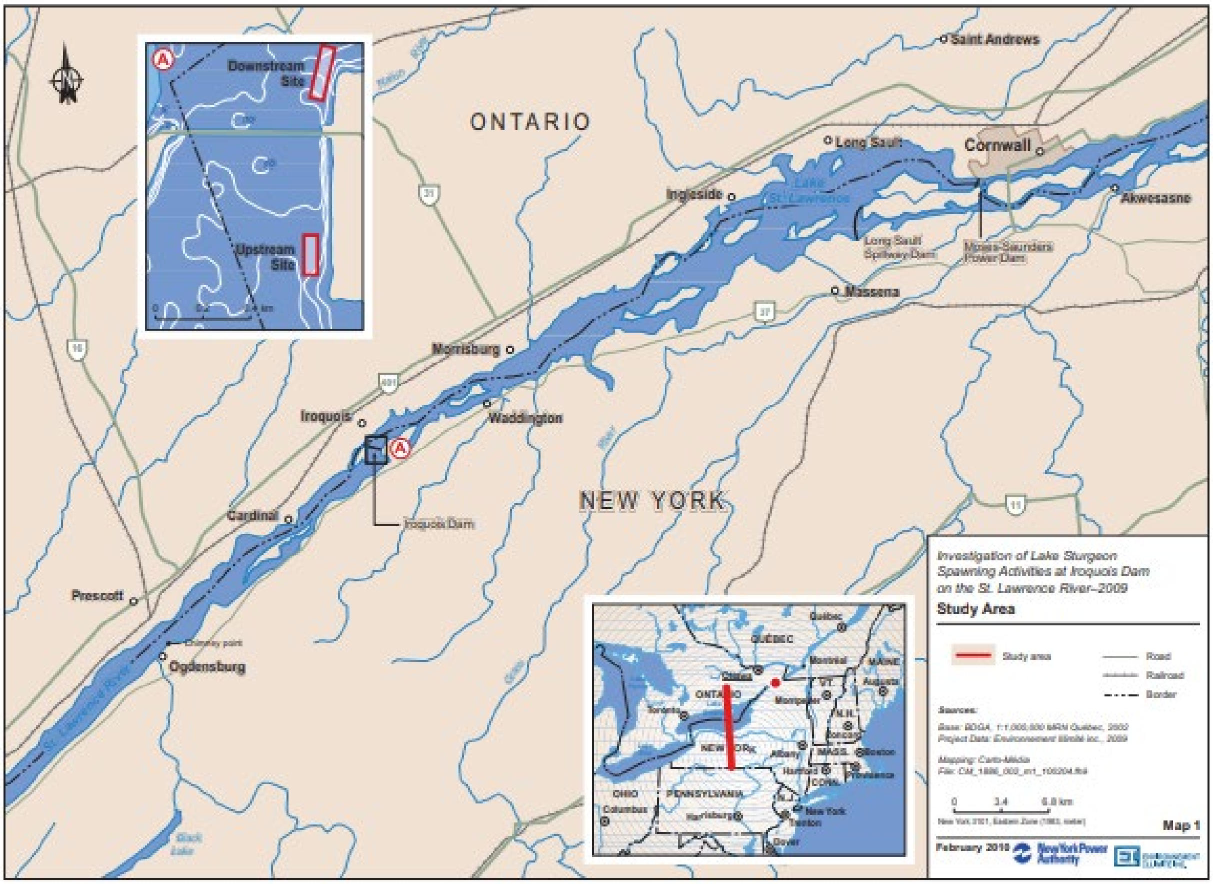


Figure 1. Map of the St. Lawrence River including the Iroquois Dam and the artificial spawning beds (indicated by the red rectangles in the inset) created both upstream and downstream of the dam (Environnement Illimité Inc. 2010).

## Chapter 1: Spawning abundance and dynamics of Lake Sturgeon in the upper St. Lawrence River near Iroquois Dam


### Abstract

The Lake Sturgeon *Acipenser fulvescens* is a large, long-lived freshwater fish that completes its entire lifecycle in freshwater, but migrates long distances. Human modifications to critical spawning habitats and migration corridors due to pollution and dams, coupled with high historical fishing mortality, have reduced most populations to all-time low abundances. In New York, USA, the species remains listed as threatened but recovery efforts including fishing moratoria, propagation and stocking, and habitat improvements have recovered 4 of the 7 remaining distinct population segments to interim targets for abundance. We assessed trends in Lake Sturgeon population abundance at two artificial spawning beds near Iroquois Dam for the aggregations on either side of the impassable barrier in the St. Lawrence River, North America. Fish were counted on 10 replicated video transects throughout the spawning season from 2011 through 2022 using underwater video. Counts of Lake Sturgeon in 30 m × 30 m grid cells along these transects were used to estimate population density at spawning beds upstream and downstream of the dam while accounting for imperfect detection within spatially explicit count models. Overall detection probability of the sturgeon was low (mean: 0.04, 95% CRI: 0.006-0.113) but was highest during the middle of the spawning season, reaching a maximum of 0.12 (0.10-0.13). Mean abundance was 10 (0-58) fish per 900 $m^2$ at the downstream bed and 5 (0-30) at the upstream bed, with a mean difference of 5 (4-6) fish per 900 $m^2$ across years. Periodicity in abundance was noted at the downstream bed. Abundance decreased slightly 2011-2023, with a mean population growth rate of -0.006 (-0.017 to -0.001) across beds. The approach we adopted for analysis could potentially be adapted to a wide range of data sources within Lake Sturgeon

monitoring programs, particularly if combined with auxiliary demographic information. Future studies could improve understanding and interpretation of trends from these models with information regarding immigration and emigration from the study area and recruitment to this population segment.

**Introduction**

The Lake Sturgeon *Acipenser fulvescens* is the only sturgeon in North America that completes its entire life cycle in fresh water (Peterson et al. 2007) and is among the largest and most unique freshwater fishes therein (Pollock et al. 2015). They can reach maximum lengths up to 3 m and may weigh up to 180 kg (Vélez-Espino and Koops 2009). Males generally mature at 12-15 years (Peterson et al. 2007) and may live up to 55 years (Colborne et al. 2021). Females take 18-27 years to reach sexual maturity (Peterson et al. 2007) and reach maximum ages of 80-150 years (Colborne et al. 2021). This later reproductive development confers increased investment in somatic growth (Beamish et al. 1996) and rapid growth through early life stages (Bruch et al. 2016), with high natural survival upon maturation (Colborne et al. 2021). This allows the species to achieve high reproductive output over their long lifespans; however, this also means that Lake Sturgeon are sensitive to anthropogenic perturbations (Peterson et al. 2007; Pollock et al. 2015). Recovery from population declines is therefore necessarily slower when compared to species with faster maturation times, with projected recovery timelines of 20-100 years (Vélez-Espino and Koops 2009).

Lake Sturgeon populations throughout most of their native range have undergone dramatic population reductions largely due to anthropogenic factors such as commercial fishing and loss or degradation of habitat (Pollock et al..2015; Bruch et al. 2016). The species is listed as threatened in most states and provinces (Bruch et al. 2016), including New York, USA where it has been listed as threatened since 1983 (NYSDEC 2018). Following closure of most Lake Sturgeon fisheries in Canada and the United States during the mid-20th century (e.g., Marsden and Langdon 2012), population fragmentation due to dams and habitat loss became the primary obstacles to recovery (Peterson et al. 2007). Dams, specifically, block spawning habitats historically used by Lake Sturgeon and modify spawning habitat proximal to spillways and powerhouses. As a wide-ranging species, effective connected river lengths of 700-1000 km may be required to limit population impacts (Auer 1996).

In New York State, out of 13 distinct sturgeon management units, 7 were identified for recovery actions (NYSDEC 2018) and have since been actively managed through fishery moratoria, hatchery supplementation, and habitat improvements (NYSDEC 2018). The current goals for recovery include 1) adult spawning targets of > 750 individuals per management unit with a lower 95% confidence interval of 500 fish and a minimum of 150 individuals in a given spawning aggregate in those management units with a lower 95% confidence interval exceeding 80 individuals, and 2) evidence of natural recruitment in 3 of 5 consecutive years during the most recent 20-year period (NYSDEC 2018). Four of the seven management units currently meet this goal, and all populations are currently considered to be at stable but low abundances (NYSDEC 2021).

The Upper St. Lawrence River management unit includes spawning aggregations in Black Lake, the Oswegatchie River, and the mainstem of the St. Lawrence River near Iroquois Dam (NYSDEC 2018, 2021). The management unit was recently determined to have reached target adult and juvenile recruitment objectives (NYSDEC 2021) associated with the NY Lake Sturgeon recovery plan (NYSDEC 2018). Collectively, spawning aggregations in the management unit total more than 750 adults, and reproduction occurs in all three spawning areas

(NYSDEC 2021). Spawning abundance appears to be most variable at artificial spawning beds created near Iroquois Dam (NYSDEC 2021). Preliminary monitoring of spawning Lake Sturgeon at these artificial spawning beds began in 2008 and included underwater video monitoring and egg and larval sampling. Early indications were that spawning adults used the beds extensively (Environnement Illimité Inc 2010). While Sturgeon eggs and larvae have also been observed there (Environnement Illimité Inc 2010), no juvenile recruitment has been observed in the areas proximal to spawning beds (NYSDEC 2021). Coarse population indices derived from maximum counts along video monitoring transects suggest that abundance has been highly variable and may be reduced in recent years (NYSDEC 2021). However, it is unknown whether this trend is biologically driven, or it is an artifact of the index used.

The goal of this study is to inform ongoing monitoring and management of Lake Sturgeon recovery and restoration activities. Our specific objectives were to 1) develop an index of annual abundance that accounts for imperfect detection, and 2) apply this approach to assess any long-term or spatial trends in Lake Sturgeon abundance near Iroquois Dam. To do this, we used a spatially explicit count model (N-mixture) to estimate abundance of Lake Sturgeon in 30 m x 30 m grids near the artificial spawning beds from 2011 through 2023.

## Methods

### *Study Area*

This study took place proximal to Iroquois Dam on the St. Lawrence River between Ontario, Canada and New York, USA. The study area included two, 30 m × 30 m artificial spawning beds: one downstream and one upstream of the Iroquois Dam. Both spawning beds were artificially created in 2007 (Environnement Illimité Inc 2010). They were formed primarily of variable sized gravel, with large boulders at the downstream ends, placed strategically to allow sturgeon to stage on the beds (Environnement Illimité Inc 2010).

*Field Sampling*

Field sampling was conducted during late March through June, from 2008 through 2022 around 9-12 m deep (NYSDEC 2021; (Environnement Illimité Inc 2010). Monitoring was initially performed by using a Aquavu Scout SRT underwater video camera (Environnement Illimité Inc 2010), and since 2012 with a more durable, stainless SplashCam Delta Vision (Ocean Systems Inc., Miami, USA). Five 100 m transects spaced 10 m apart were established on each bed and aligned with the current, with some variability among surveys and years. Each 100 m transect was divided into five segments. Two 30 m sections were set before and after the beds, 30 m was designated on the beds and about 5 m at the beginning and end of the transect to set and retrieve the camera used to capture sturgeon. Transects 1 and 5 were located off the sides of the spawning beds while transects 2 – 4 intersected the spawning beds and transect 3 was centered on the bed (Figure 2). The placement of the transects ensured the sample area would be larger than the beds being monitored. These transect locations were chosen to allow sturgeon abundance to be monitored on and off the spawning beds. Global positioning system (GPS) coordinates were used to locate and sample the transects for each sampling session. Video files were recorded for each transect for later review and analysis.

*Video analysis*

During filming, the underwater video camera and GPS unit were connected to a digital video recorder, either a Sony DCR-SR80 digital camera (Sony Corporation, Tokyo, Japan), or later a Blackmagic Design recording monitor (Blackmagicdesign, Port Melbourne, Australia). A GPS video overlay was used with recorded GPS coordinates to obtain georeferenced footage (Environnement Illimité Inc. 2010). The footage was later reviewed (Figure 3), and the number of Lake Sturgeon observed on each transect was counted. While reviewing the footage other measurements including temperature, date, transect number, GPS coordinates of each observed sturgeon, and presence/absence of sturgeon on the installed substrate were also recorded into a Microsoft Excel file each year and combined for analysis. The transect coordinates were joined with spatial data from fish observations, and a $0.25 km^2$ spatial grid made of 30 m × 30 m cells covering the study area to identify grid cells that were surveyed but had zero observed Lake Sturgeon. The size of 30 m × 30 m was chosen because it provided the maximum resolution that still allowed for replication within sites and surveys each year. Multiple observations of Lake Sturgeon within grid cells along a single transect (identified by video file) were summed because double counting of individuals was highly unlikely at that spatial scale.

*Data analysis*

We used a spatially explicit N-mixture model (Royle 2004) to estimate abundance of Lake Sturgeon in 30 m × 30 m grid cells that correlated with the monitored transects upstream and downstream of Iroquois Dam following approaches similar to those used for the spatially explicit occupancy models described by Rushing et al. (2019). We incorporated a generalized additive model (GAM) within the abundance model to estimate complex, nonlinear spatial patterns in Lake Sturgeon abundance as related to artificial spawning beds and to account for the fact that abundances of Lake Sturgeon in 30 m × 30 m grid cells were spatially correlated.

The number of Lake Sturgeon observed per grid cell ($i$) on each day ($t$) in each year ($j$) was summed within transects to result in a total count per cell for each transect. We assumed that the number ofLake Sturgeon observed ($y_{i,t,j}$) was the outcome of a binomial distribution based on the probability of observing individual fish on each day ($p_t$) and the true abundance of Lake Sturgeon within that grid cell each year ($N_{i,j}$):

$$y_{i,t,j} \sim \text{Binomial}(p_t, N_{i,j}).$$

Abundance in each year was modeled as the outcome of a Poisson count process, with a mean of $\mu_{N_{i,j}}$ that was modeled on the $\log_e$ scale as a function of a spatial GAM using 1-15 knots (bends) (k) with shared priors on basis function ($g$) coefficients (β):

$$N_{i,j} \sim \text{Poisson}\left(\mu_{N_{i,j}}\right),$$

$$\log\left(\mu_{N_{i,j}}\right) \sim f(\text{Easting}_i, \text{Northing}_i),$$

$$\text{f}\left(\text{Easting}_\text{i}, \text{Northing}_\text{i}\right) \sim \sum_{k=1}^{K} g_k\,(\text{Easting}_i, \text{Northing}_i)\beta_k,$$

assuming a multivariate normal prior for all $\beta_k$ with a mean of zero and a shared variance of $\sigma_\beta^2$:

$$\beta_k \sim \text{Mvnormal}\left(0, \sigma_\beta^2\right).$$

We applied a Bayesian penalization term (λ) to the smoothing terms to avoid overfitting. We did this using the generalized thin-plate spline method (Duchhon 1977; Wood 2003) for the smoothing basis in the *mgcv* package for R version 4.2.2 (R Core Team 2022). We derived the prior for the variance of $\beta_1$ ($\sigma^2_{\beta_1}$ = 0.95) from the grid cell coordinates, and specified λ according to a minimally informative gamma prior:

$$\lambda \sim \mathrm{Gamma}(0.05, 0.005).$$

We used a hierarchical prior on *p* to allow sharing of information within day of year across sites and years. The mean detection probability ($\mu_p$) was specified using minimally informative, normal hyperprior on the logit scale with a mean of zero and a standard deviation of one:

$$\mathrm{logit}(\mu_p) \sim \mathrm{Normal}(0, 1),$$

and the global variance for *p* ($\sigma^2_p$) was drawn from a uniform prior between zero and 10. Detection probability was allowed to vary among days and fixed across years to capture seasonal aggregation on spawning beds. Daily detection probabilities were drawn from a logit-scale, normal prior:

$$\mathrm{logit}(p_t) \sim \mathrm{Normal}(\mathrm{logit}(\mu_p), \sigma^2_p).$$

All model parameters and derived quantities were estimated with Markov Chain Monte Carlo (MCMC) methods using the Gibbs sampler in JAGS (Plummer 2003) through the *R2Jags* package (Su and Yajima 2021) in R (R Core Team 2022). We ran 50,000 iterations for each of 3 chains, with a burn-in period of 25,000 iterations, and we retained every 5$^{th}$ sample, for a total of 15,000 samples from the posterior distribution. We confirmed that all Markov Chains converged with $\hat{r} < 1.1$ (Gelman and Rubin 1992), and bulk effective sample sizes of several hundred samples per chain indicated a sufficient number of independent samples from each posterior (Kruschke 2010). Goodness-of-fit tests showed no evidence of lack of fit, with a Bayesian p-value of 0.486 (Figure S1). Summary statistics and derived quantities were calculated for each iteration of each Markov Chain to incorporate biological and computational uncertainty into population estimates.

We derived summary statistics for $N$ as 1) mean abundance ($N$) within grid cell $i$ each year $j$ ($N_{i,j}$), 2) the sum of abundance within beds (upstream or downstream) each year, and 3) the sum of abundance across spawning beds each year. We derived a population growth rate ($r$) based on change in $N_{i,j}$ from 2011 through 2022. Each of our indices of $N$ assumes that individuals were not double counted within grid cells, with the potential for compounding errors through summation in the case of the third. The true degree of double counting in surveys is unknown, although precautions were taken in survey design, execution, and post-hoc aggregation of counts to minimize potential for violation. While this warrants caution in interpreting magnitude, we suspect that each of the indices may be useful for population monitoring because trends in abundance are not expected to vary unless double counting also varies systematically across years, which is unlikely. We did not include other abiotic factors.

## Results

From 2011 through 2023, more Lake Sturgeon were generally observed on the downstream spawning bed compared to the upstream bed, and spatial distribution of fish, while consistently highest near spawning beds, was variable from year to year (Figure 2). Detection probability of individual Lake Sturgeon was generally low, with an overall mean of 0.040 (0.006-0.113). However, detection varied seasonally and was highest (0.116, 0.102-0.1131) during the middle of the year (May-June) across sites and years (Figure 3).

Mean abundance was 10 (0-58) fish per 900 $m^2$ at the downstream bed and 5 (0-30) at the upstream bed, with a mean difference of 5 (4-6) fish per 900 $m^2$ across all years and grid cells (Figure 4). Exclusion of zero from the estimate of the difference indicates that it is statistically significant. However, abundance varied annually within beds. The sum of abundance within sites was also consistently higher at the downstream, although it varied among years (Figure 4). Abundance was highest at the downstream bed in 2012 (533, 389-726) and highest at the upstream bed in 2015 (386, 233-601), with a mean difference of 113 (82-147) fish per bed across years. The sum of abundance across all grid cells and beds generally decreased from 2011 (761, 484-1145) through 2022 (712, 420-1127), but was highest in 2015 (857, 552-1276) and lowest in 2017 (445, 161-845). We observed cyclical patterns in each abundance index for Lake Sturgeon at the downstream bed that were muted but apparent in estimates of total abundance across all sites (Figures 2 and 4).

Lake Sturgeon abundance decreased overall from 2011 to 2022 despite interannual variability according to all abundance metrics. This corresponded to an average $r$ of -0.006 (-0.017— -0.001) across beds within this DPS. This decrease was more pronounced at the upstream bed (-0.01, -0.03– -0.002) than at the downstream bed (-0.0019, -0.0011– -0.0009)

## Discussion

We successfully adapted spatially explicit occupancy models (Rushing et al. 2019) to count data to estimate multiple abundance indices for Lake Sturgeon at two artificial spawning beds while accounting for imperfect detection of individuals. Within beds, abundance exhibited spatial and temporal variability. Overall, abundance on spawning beds decreased slightly from 2011 through 2022. The approach used in this study has potential to be useful for other monitoring programs and may be adaptable to a variety of index data.

Lake Sturgeon abundance was highest in 30 m × 30 m grid cells immediately proximal to artificial spawning beds at both upstream and downstream locations. Maximum densities of Lake Sturgeon exceeded 100 fish per 900 $m^2$. At the downstream bed, the highest densities were observed early in the time period, whereas at the upstream bed the maximum density was observed in 2015. It is notable that the high density estimated in 2015 at the upstream site may be due to low detection based on low counts, whereby estimates of density would approximate the global mean given the hierarchical nature of the model used.

Abundance and detectability of Lake Sturgeon varied over time at different scales. Probability of detection varied seasonally across grid cells and years coincident with aggregation during the spawning season. Individual detection was lowest early in the season, becoming highest at the peak of the run, and then decreasing again at the end of the run. On an interannual basis, abundance exhibited semi-regular cycles within the downstream bed but muted or no such trend at the upstream bed. Despite this interannual variability, the population appears to have decreased slightly over the 10-year study period.

The approach we adapted from Rushing et al. (2019) may have broad applications for a variety of index data that are collected in fisheries monitoring, but also requires validation of multiple assumptions for future use. Two key assumptions of unmarked population analysis methods are that individuals are not double-counted and that individual variability for detection does not vary within primary periods (years in the case of this study). We attempted to minimize potential for double counting through study design (i.e., transect spacing and conservative inclusion of individuals in counts), however it is still possible that fish were double counted to an unknown degree. In this case, our estimates could be considered an index of maximum density at the various spatial and temporal scales considered. However, since double counting is unlikely to vary systematically across years, we think the trends in abundance and population estimates are unlikely to be systematically biased. Likewise, the range of data used for this study corresponded to the start of the spawning run to avoid inclusion of dates when fish were not available for detection. Raw counts were generally highest near the middle of the annual spawning season, as was individual detection probability. In this case, the detection probability may be more appropriately thought of as a joint probability of detection and use. While this heterogeneity would result in underestimates of true abundance, we would also suspect that interannual trends to be unbiased since inter-annual variability in the general shape of spawner counts was consistent across years.

Future fishery monitoring efforts that employ spatially structured survey and analysis methods for sturgeons (e.g., Sweka et al. 2007; Pendleton and Adams 2021) could benefit from incorporation of other demographic information to improve utility and of estimates from these methods. Incorporation of design considerations can reduce the likelihood of double counting within replicate sampling events. Additional demographic information from telemetry studies could inform seasonal patterns in use and availability (e.g., Withers et al. 2019; Kazyak et al. 2020).


## Acknowledgments

This work was funded by New York State Department of Environmental Conservation (NYSDEC), the St. Lawrence River Research and Education Fund (SLRREF) through Award Number 94410-1175387 to SUNY Oneonta, with additional support provided by SUNY Oneonta.


## Data Availability Statement

All data and code necessary to reproduce this analysis are available through GitHub: https://github.com/danStich/lake-sturgeon-spawning.

## References


Auer, N. A. 1996. Importance of habitat and migration to sturgeons with emphasis on Lake Sturgeon. Canadian Journal of Fisheries and Aquatic Sciences 53(S1):152–160.

Bruch, R. M., T. J. Haxton, R. Koenigs, A. Welsh, and S. J. Kerr. 2016. Status of Lake Sturgeon (*Acipenser fulvescens* Rafinesque 1817) in North America. Journal of Applied Ichthyology 32(S1):162–190.

Colborne, S. F., T. A. Hayden, C. M. Holbrook, D. W. Hondorp, and C. C. Krueger. 2021. Lake Sturgeon (*Acipenser fulvescens*) annual adult survival estimated from acoustic telemetry. Journal of Great Lakes Research 47(6):1814–1823.

Duchon, J. 1977. Splines minimizing rotation-invariant semi-norms in Solobev spaces. Pages 85-100 in W. Shemp and K. Zeller, editors. Construction theory of functions of several variables. Springer, Berlin.

Environnement Illimité Inc. 2010. Investigation of Lake Sturgeon Spawning Activities at Iroquois Dam on the St. Lawerence River in 2009. Prepared for the New York Power Authority, White Plains, NY.

Federal Energy Regulatory Commission (FERC). 2003. Order Approving Settlement Agreements, Dismissing Complaint, And Issuing New License, P-2000-036 and EL03-224-000. FERC, Washington, DC.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457-511.

Kazyak, D., A. M. Flowers, N. J. Hostetter, J. A. Madsen, M. W. Breece, A. Higgs, L. M. Brown, A. Royle, and D. A. Fox 2020. Integrating side-scan sonar and acoustic telemetry to estimate the annual spawning run size of Atlantic sturgeon in the Hudson River. Canadian Journal of Fisheries and Aquatic Sciences 77:1038-1048.

Kruschke, J. 2010. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and WinBUGS. Academic Press, San Diego, California. 672 pp.

Marsden, J. E., and R. W. Langdon. 2012. The history and future of Lake Champlain’s fishes and fisheries. Journal of Great Lakes Research 38:19–34.

New York State Department of Environmental Conservation. 2018. Lake Sturgeon recovery plan 2018 - 2024. NYSDEC, Albany, NY. Available: https://www.dec.ny.gov/outdoor/111557.html (February 2022).

New York State Department of Environmental Conservation. 2021. Lake Sturgeon population assessment report. NYSDEC, Albany, NY. Available: https://www.dec.ny.gov/docs/fish_marine_pdf/lakesturgeonassessment.pdf (February 2022).

Pendleton, R. M., and R. D. Adams. 2021. Long-term trends in juvenile Atlantic Sturgeon abundance may signal recovery in the Hudson River, New York, USA. North American Journal of Fisheries Management 44:1170 -1181.

Peterson, D. L., P. Vecsei, and C. A. Jennings. 2007. Ecology and biology of the Lake Sturgeon: a synthesis of current knowledge of a threatened North American Acipenseridae. Reviews in Fish Biology and Fisheries 17(1):59–76.

Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, and A. Zeileis, editors. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria.

Pollock, M., M. Carr, N. Kreitals, and I. Phillips. 2015. Review of a species in peril: what we do not know about Lake Sturgeon may kill them. Environmental Reviews 23:30–43.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (August 2023).

Su, Y.-S. and M. Yajima. 2021. R2jags: Using R to run 'JAGS', version 0.7-1. URL: https://CRAN.R-project.org/package=R2jags (August 2023).

Sweka, J. A., J. Mohler, M. J. Millard, T. Kehler, A. Kahnle, K. Hattala, G. Kenney, and A. Higgs. 2007. Juvenile Atlantic Sturgeon habitat use in Newburgh and Haverstraw Bays of the Hudson River: implications for population monitoring. North American Journal of Fisheries Management 27:1058-1067.

Vélez-Espino, L. A., and M. A. Koops. 2009. Recovery Potential Assessment for Lake Sturgeon in Canadian Designatable Units. North American Journal of Fisheries Management 29(4):1065–1090.

Withers, J. L., D. Einhouse, M. Clancy, R. Neuenhoff, and J. Sweka. 2019. Integrating acoustic telemetry into a mark-recapture model to improve catchability parameters and abundance estimates of Lake Sturgeon in eastern Lake Erie. North American Journal of Fisheries Management 39:913-920.

Wood, S. N. 2003. Thin plate regression splines. Journal of the Royal Statistical Society B 65:95-114.

**Figures**

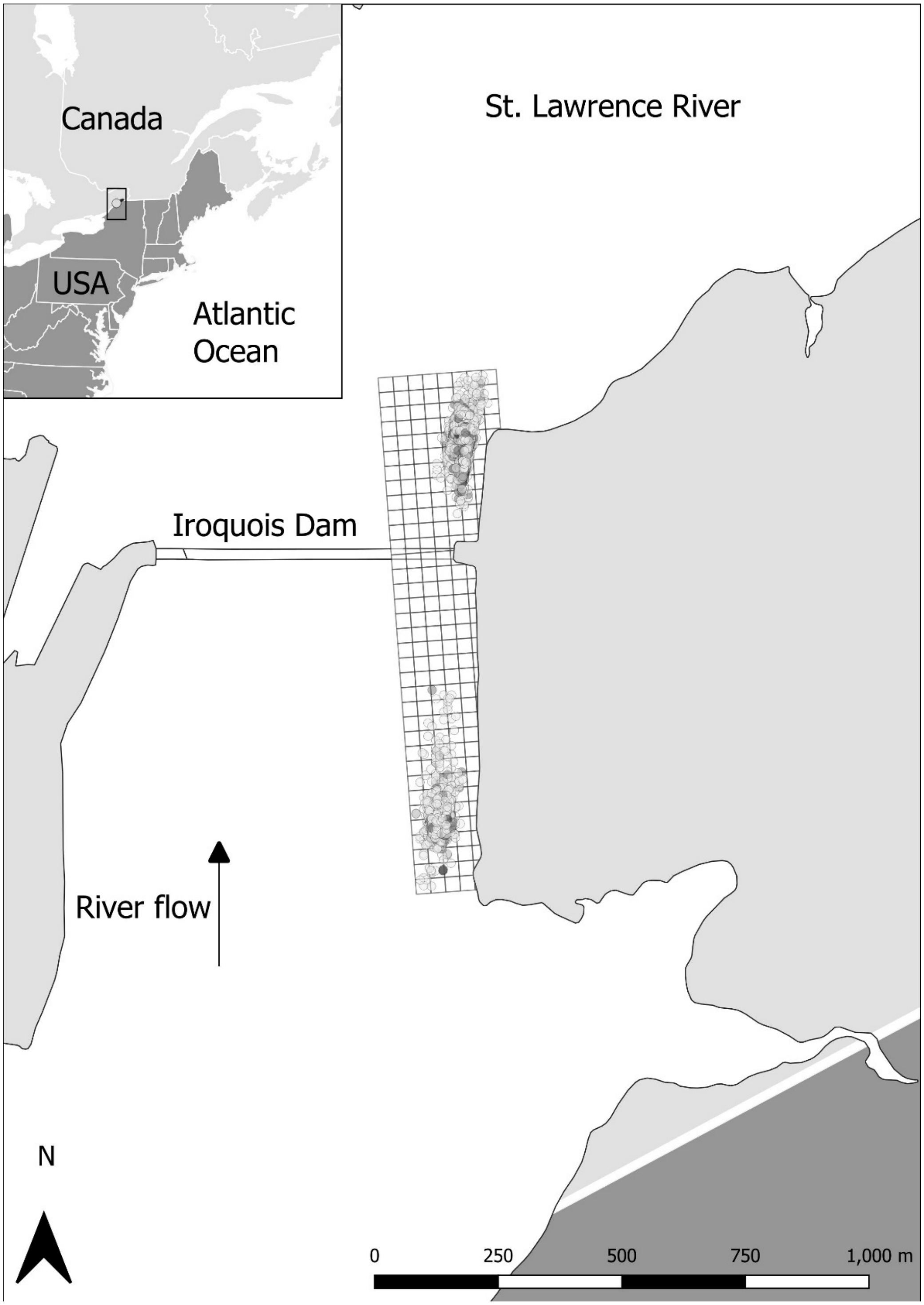


Figure 1. Map of the study system showing location in North America (inset) and location relative to Iroquois Dam in the St. Lawrence River, NY.

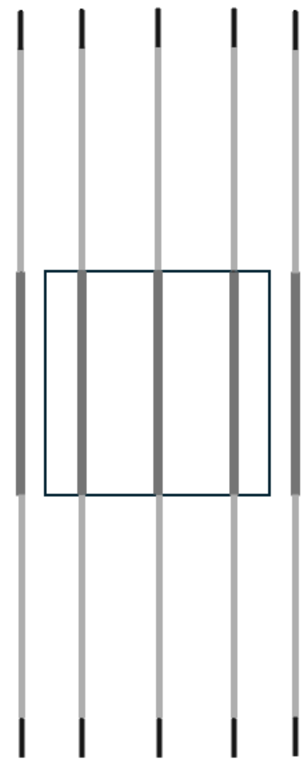

Figure 2. Schematic of the transects designed for the artificial spawning beds. Each transect (100m) was broken up into several sections: two 30m sections before and after the beds (light gray, one 30m section covering the length of the spawning beds (dark gray), and two 5m sections for camera installment and retrieval (black).


Figure 3. A still frame of recorded transect footage. In this footage, qoordinates (lattitude & longitude), time, date, speed of camera, and Lake Sturgeon can be observed.

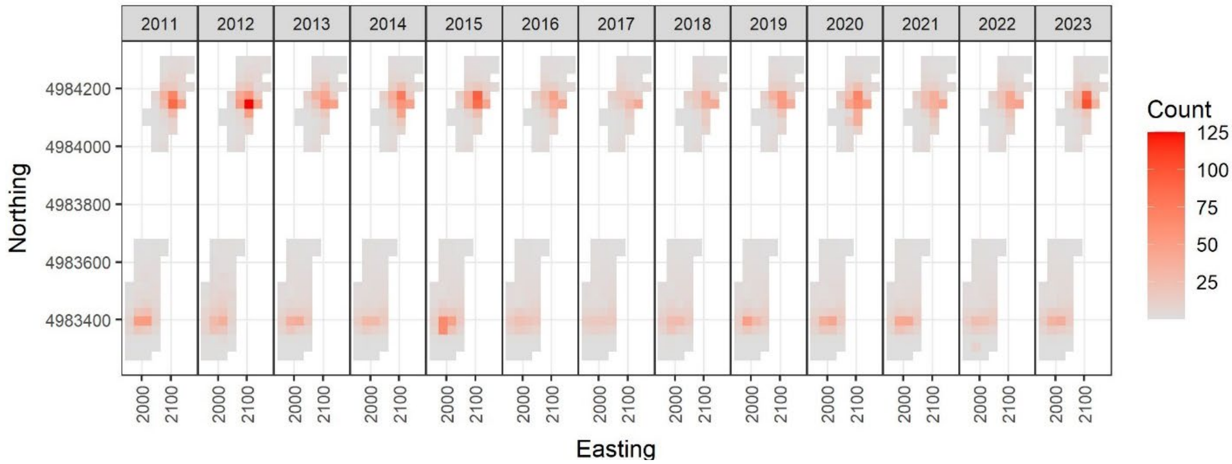


Figure 4. Mean Lake Sturgeon *Acipenser fulvescens* densities near the artificial upstream spawning beds (lower) and downstream spawning beds (upper) proximal to Iroquois Dam in the St. Lawrence River, North America, from 2011 to 2022 in 30 m × 30 m grid cells.

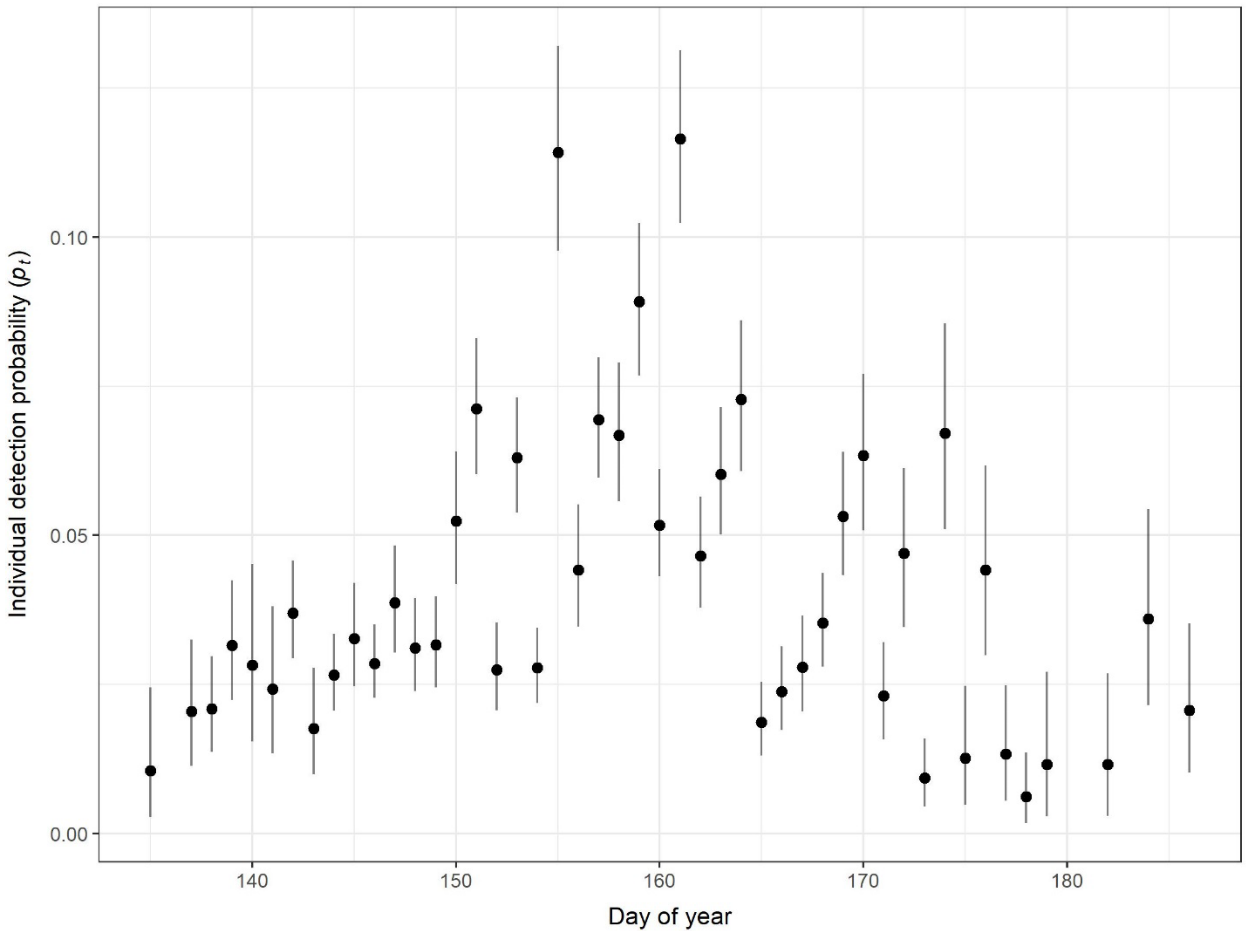


Figure 5. Individual detection probability of Lake Sturgeon *Acipenser fulvescens* at artificial spawning beds Iroquois Dam in the St. Lawrence River within days of year across sites and beds 2011-2023. Points represent predicted posterior mean and vertical error bars are 95% credible intervals.

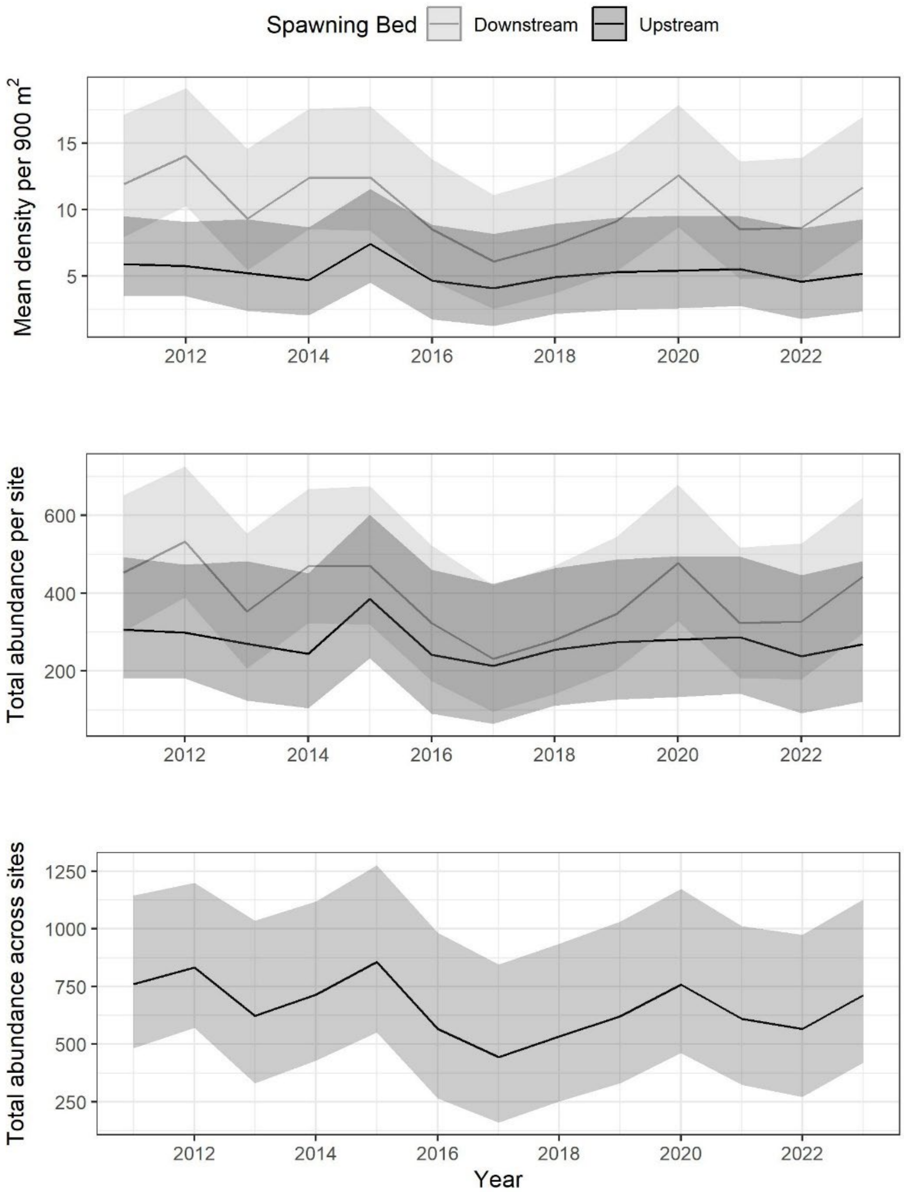


Figure 6. Abundance indices for Lake Sturgeon *Acipenser fulvescens* near artificial spawning beds upstream and downstream of Iroquois Dam in the St. Lawrence River, North American each year, showing maximum abundance per 900 $m^2$ within beds and years (top), sum of abundances within beds and years (middle), and sum of abundances across both beds within years (bottom). Solid lines are posterior predictive means and ribbons are 95% credible intervals.

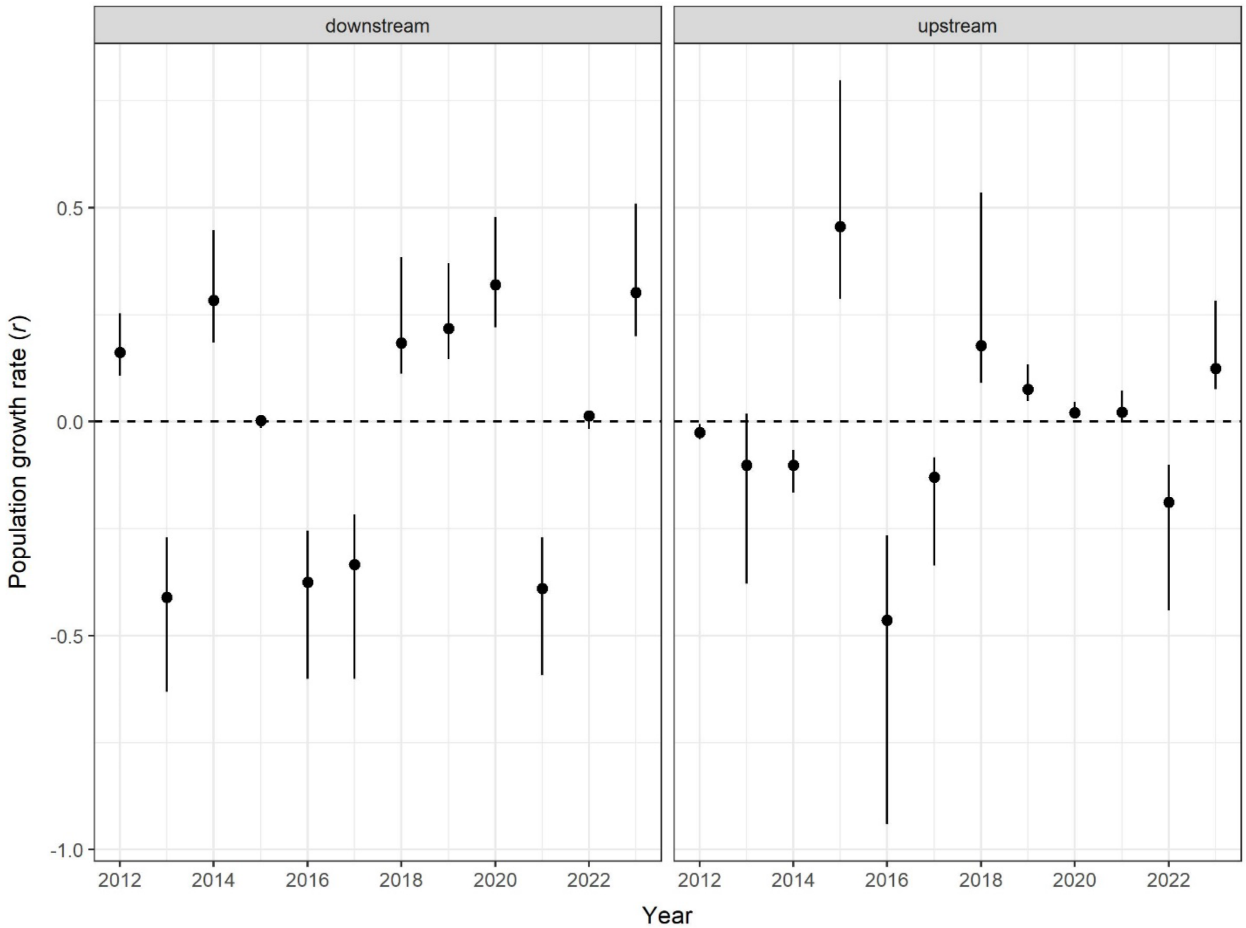


Figure 7. Plot of changes to population growth rate of lake lturgeon *Acipenser fulvescens* at artificial spawning beds upstream and downstream of Iroquois Dam in the St. Lawrence River, North America from 2011 through 2022.

**Supplemental Information**

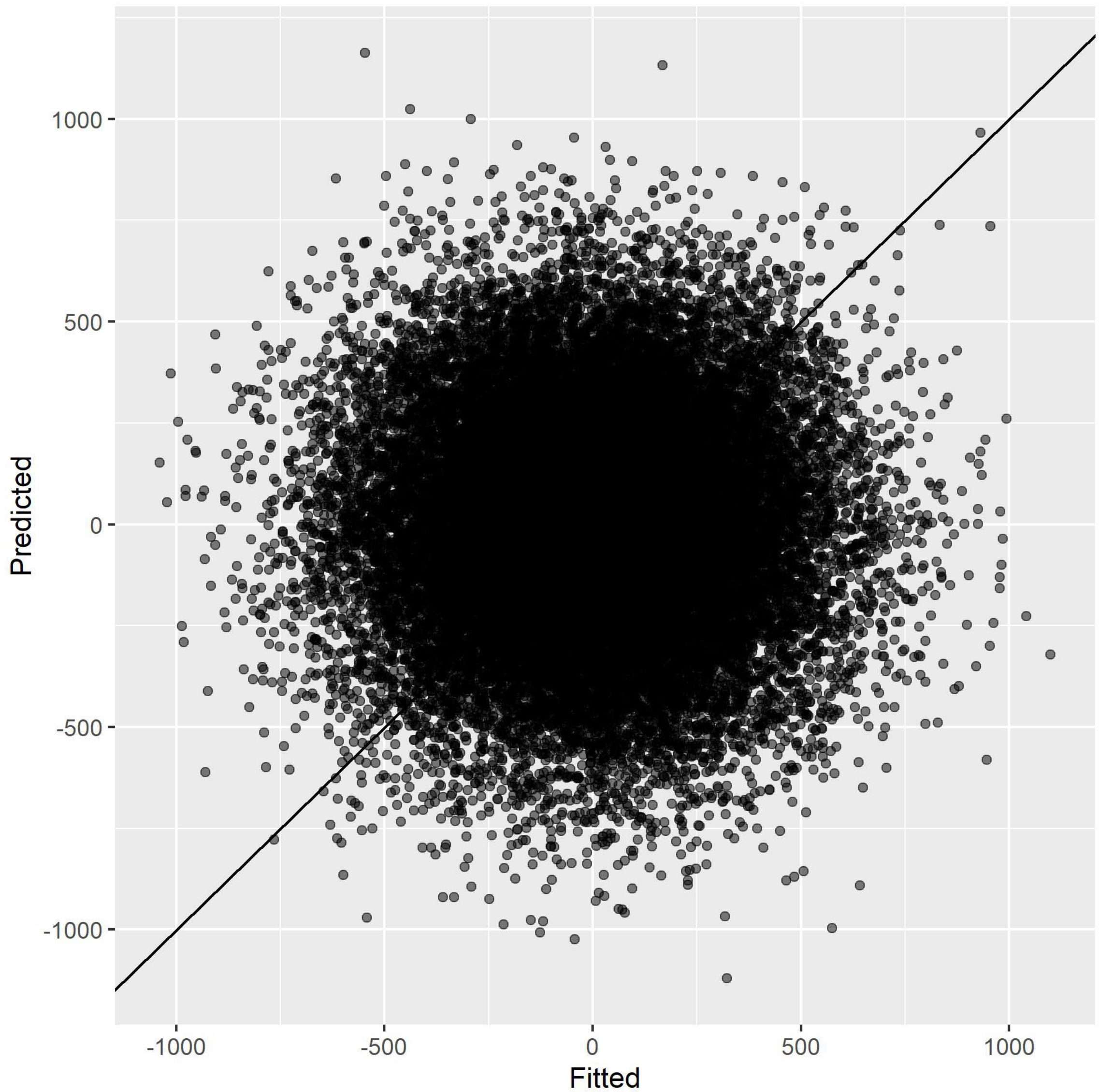


Figure S1. Plot of residuals for the fitted data ($y_{i,\ t,\ j} - N_{i,j}p_t$) compared to residuals of the posterior predicted fit ($E(y_{i,\ t,\ j} \mid \theta)$ - $N_{i,j}p_t$). The black line is a 1:1 line for comparison, and the Bayesian p-value (probability that Predicted > Fitted) was 0.501, suggesting little evidence for lack of fit.

## Chapter 2: Simulation Validation of N-mixture Analysis of Video Recorded Spawning Lake Sturgeon

### Introduction

Conservation and species management depends on accurate population size estimation. Often, management decisions are informed by models made from limited datasets collected from single populations. This means there is large room for error in the creation of these models. Additionally, while models designed for specific populations are becoming more complex, advances in computing make it easier to create models without a strong foundation in modeling. Due to this, advanced models are created faster than they are tested for accuracy (Martis 2006). If the model is inaccurate, it can result in poor management decisions for the target population on which the model was based (Martis 2006). Without the use of simulation-based validation, there is no way of knowing if a model truly reflects the population being studied or whether it provides accurate predictions.

One useful method of modeling populations relies on N-mixture models. These models are used to estimate a specific population's abundance while accounting for imperfect detection (Rushing et al. 2019) and are useful when survey methods do not involve marking or tagging (Madsen and Royle 2023). N-mixture models play a key role in ecological studies because they can help determine accurate population abundances, even when the population size is known to fluctuate over time (Madsen and Royle 2023). Since most populations do fluctuate, it is important for managers to be able to address this issue. In addition to using N-Mixture Models, generalized additive models (GAM) are often utilized. Study populations are affected in their abundance and distribution by many environmental factors (Rushing et al. 2019). GAMs allow modelers to incorporate multiple factors into the analysis, including latent spatial patterns in abundance among sites. Without including GAMs in within N-mixture models, nonlinear relationships such as spatial patters would be missed by linear predictors (Rushing et al. 2019).

We used this combined data analysis to study sturgeon populations in NYS that were observed using an underwater video camera that followed predetermined transects, with unmarked fish. On these video surveys, it is also likely that the field team experienced imperfect detection of the spawning population due to factors like water clarity and errors involving the underwater camera. While the advantages of N-mixture models and GAM~~'~~s are clear, it is still important to run simulation-based validation on these models to make predictions over time and to test the effectiveness of current sampling procedures. Simulations create hypothetical scenarios the model can interact with as many times as the sampler desires. This shows how helpful the model is over time and in different situations.

While models can be extremely useful tools for species management, they also can be detrimental to a population if they make incorrect predictions. Due to the limited data that make these models, there is a clear need to check their accuracy. Simulation-based validation is a method that can be used to ensure a model is performing as it should.

**Methods**

We used detection probabilities from the St. Lawrence River survey described in Chapter 1 to determine the precision and accuracy of the estimates generated from our N-Mixture model using data simulation. We used the detection probabilities from the model of Lake Sturgeon populations around the Iroquois Dam to generate random, known observations. These observations were made to represent varying samples and differing numbers of individuals to determine the accuracy and precision of the model. For simulation, a hypothetical population (*N*) and the probability of detecting individuals of that population (*p*) were randomly selected. Then counts were randomly simulated in each of the previously created grid cells associated with the field survey transects.

The N-mixture model from Chapter One was then used to estimate population abundance ($\widehat{N}$) and a detection probability ($\hat{p}$) for the simulated data. Estimates of $\widehat{N}$ and $\hat{p}$ were then compared to the known values of N and P used to simulate the data set. This would determine how accurate the estimates generated by the model were compared to a known N and p. This procedure was then repeated 100 times (Figure 1). The simulation code was made using the programming language R (R Core Team 2023). N-mixture models were created in JAGS (Plummer 2003) and analyzed through the R2Jags package (Su and Yajima 2024). All code used is publicly available on GitHub (see Data Availability Statement).

**Results**

Our simulation indicated that estimated abundance and detection probability were relatively unbiased compared to “known” values used in simulation. Because the known values used for simulation were based on empirical estimation and sampling design, this suggests that estimates from the empirical analysis are unbiased. The mean bias in estimated abundance from the simulation-based validation was -0.065 (95% credible interval [CRI] = -0.867-1.222). While the abundance bias was close to zero, there was a slight underestimation of individuals on average (Figure 1). Because the counts were near zero, the most extreme errors in estimation, however, occurred in a positive direction (i.e., overestimation).

Estimated detection probability was relative unbiased, with a mean bias of -0.003 (95% CRI = -0.012-0.005; Figure 2). While overall bias was <1%, is indicated our model slightly underestimated detection probability. This underestimation however, was close to zero meaning that the underestimation of detection probability was minute.

We noted an inverse relationship in between bias estimated detection probability ($\hat{p}$) and bias in estimated abundance ($\widehat{N}$). The mean Pearson correlation coefficient ($r$) for the relationship was $r$ = -0.80. As the bias of $\hat{p}$ increased, the bias of $\widehat{N}$ decreased (Figure 3).

**Discussion**

The objective of this thesis was to determine the population abundance and detection probability of Lake Sturgeon based on video-recorded observations near the Iroquois Dam. Because our model was created for a novel application to sturgeon data, we used a simulation-estimation approach to validate the statistical model based on recommendations about best practices (Power 1993; Zurell et al. 2010; Augusiak et al. 2014; Bellier et al. 2016). This represents an important, but often overlooked, aspect of developing novel statistical models (Rykiel Jr.1996)). Issues associated with model validity are particularly pervasive in development and application of N-mixture count models (Kéry 2004; Link et al. 2018; Barker et al. 2018), particularly when detection probability is low (Couturier et al. 2013; Dennis et al. 2014).(Couturier et al. 2013; Dennis et al. 2015) Based on simulations, the statistical model used to estimate Lake Sturgeon abundance appears to be relatively unbiased with respect to both abundance and detection probability.

Estimated abundance and detection probability were relatively unbiased according to the results of our simulation study. Overall our modeling framework underestimated the mean number of individuals in 30 m × 30 m grid cells by about 0.06 individuals on average under the current sampling design used in the empirical study. The model also slightly underestimated detection probability, but by an average of only 0.003, meaning there was also very little bias in our model regarding detection. The 95% credible intervals for bias in both parameters indicated broad overlap with zero and no statistically significant bias. This also means that this model is broadly applicable to other populations and projects that use a variety of monitoring methods.

The primary goal of this thesis was to create a model to estimate the abundance and detection probability of Lake Sturgeon using video capture data. Statistical validation of the model used for estimating abundance provides confidence in trends observed in empirical analysis. This model will continue to be used by the New York State Department of Environmental Conservation (NYSDEC) as monitoring of this population continues. Because this model accurately reflects sturgeon life history, it can be used as a framework for studying Lake Sturgeon as well as being beneficial in the design of future studies about this or other species.

**References**

Augusiak, J., P. J. Van den Brink, and V. Grimm. 2014. Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach. Ecological Modelling 280:117–128.

Barker, R. J., M. R. Schofield, W. A. Link, and J. R. Sauer. 2018. On the Reliability of N-mixture Models for Count Data. Biometrics 74(1):369–377.

Bellier, E., M. Kéry, and M. Schaub. 2016. Simulation-based assessment of dynamic N-mixture models in the presence of density dependence and environmental stochasticity. Methods in Ecology and Evolution 7(9):1029–1040.

Couturier, T., M. Cheylan, A. Bertolero, G. Astruc, and A. Besnard. 2013. Estimating abundance and population trends when detection is low and highly variable: A comparison of three methods for the Hermann's tortoise. The Journal of Wildlife Management 77(3):454–462.

Dennis, E. B., B. J. T. Morgan, and M. S. Ridout. 2015. Computational Aspects of N-Mixture Models. Biometrics 71(1):237–246.

Kéry, M. 2004. Extinction Rate Estimates for Plant Populations in Revisitation Studies: Importance of Detectability. Conservation Biology 18(2):570–574.

Link, W., M. R. Schofield, B. J. Richard, and J. R. Sauer. 2018. On the robustness of N-mixture models. Ecology 99(7):1547–1551.

Madsen, L., and J. A. Royle. 2023. A review of N-mixture models. WIREs Computational Statistics 15(6):e1625.

Martis, M. S. 2006. Validation of Simulation Based Models: A Theoretical Out. Electronic Journal of Business Research Methods 4(1):39–45.

Power, M. 1993. The predictive validation of ecological and environmental models. Ecological Modelling 68(1):33–50.

Rykiel Jr., E. J. 1996. Testing ecological models: the meaning of validation. Ecological Modelling 90(3):229–244.

Rushing, C. S., J. A. Royle, D. J. Ziolkowski, and K. L. Pardieck. 2019. Modeling spatially and temporally complex range dynamics when detection is imperfect. Scientific Reports 9(1):12805.

Zurell, D., U. Berger, J. S. Cabral, F. Jeltsch, C. N. Meynard, T. Münkemüller, N. Nehrbass, J. Pagel, B. Reineking, B. Schröder, and V. Grimm. 2010. The virtual ecologist approach: simulating data and observers. Oikos 119(4):622–635.

**Figures**

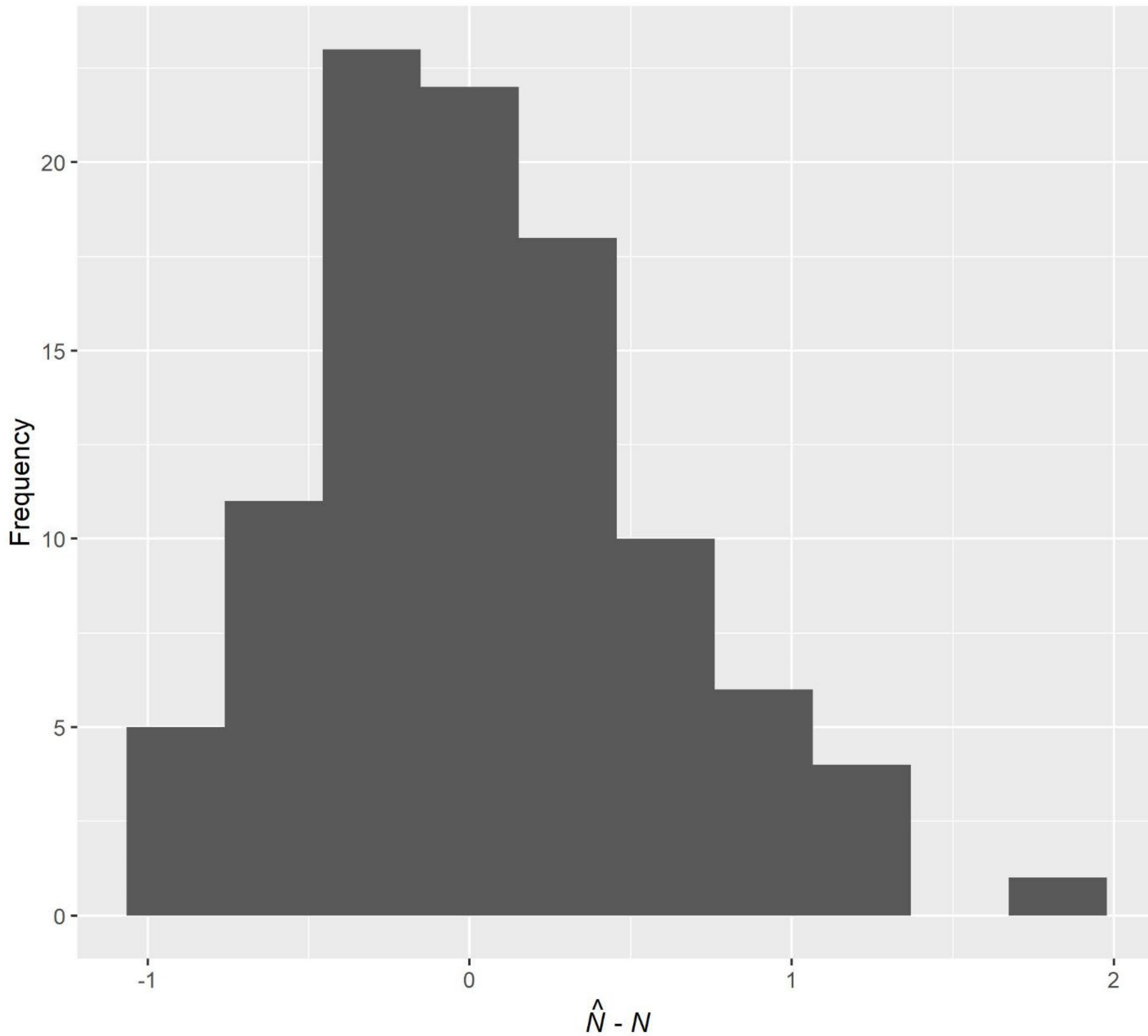


Figure 1. Histogram showing distribution of error calculated as the difference between estimated abundance ($\widehat{N}$) and known abundance used for simulating test data ($N$).

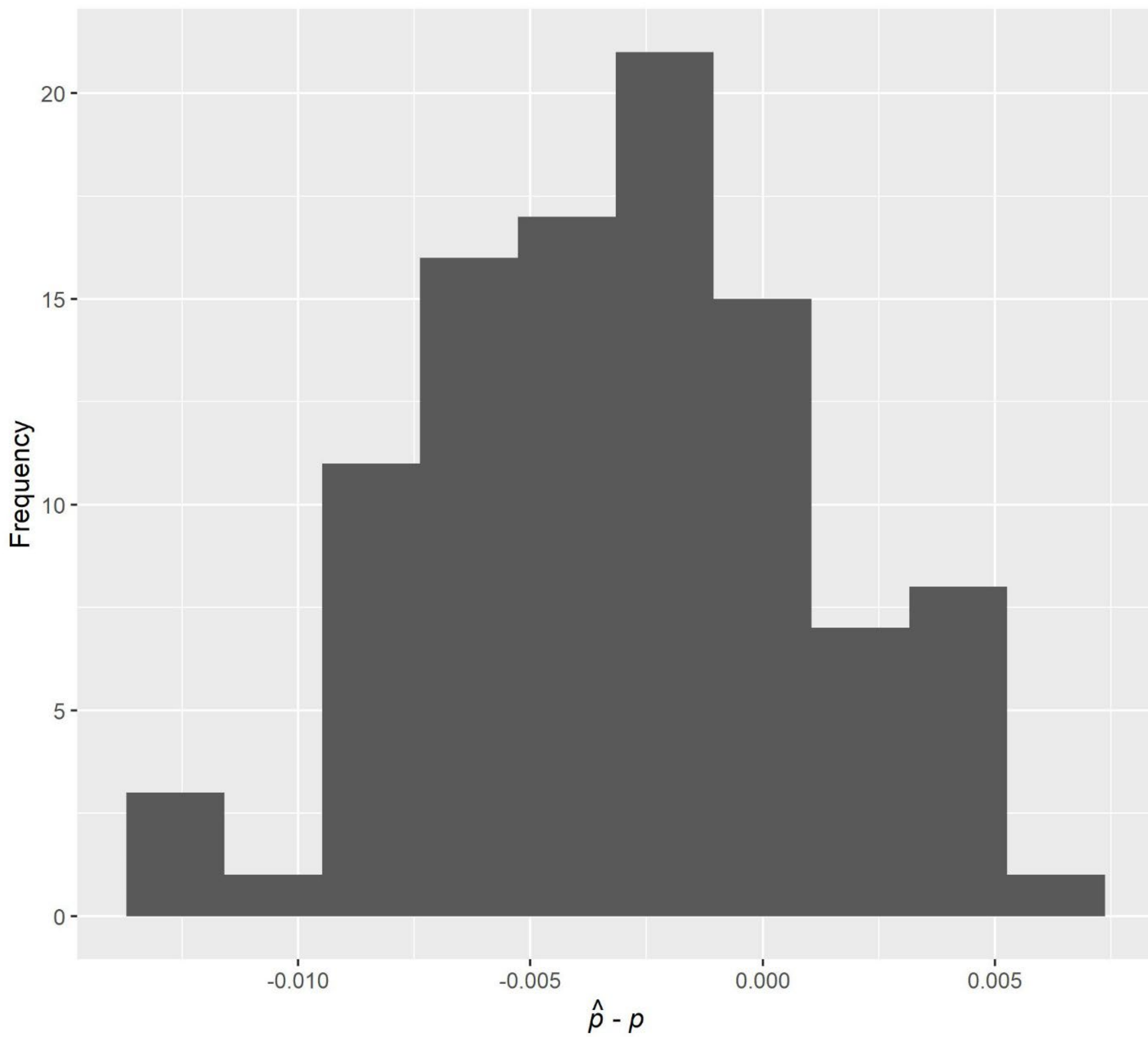


Figure 2. Histogram showing distribution of error calculated as the difference between estimated detection probability ($\hat{p}$) and known detection probability used for simulating test data ($p$).

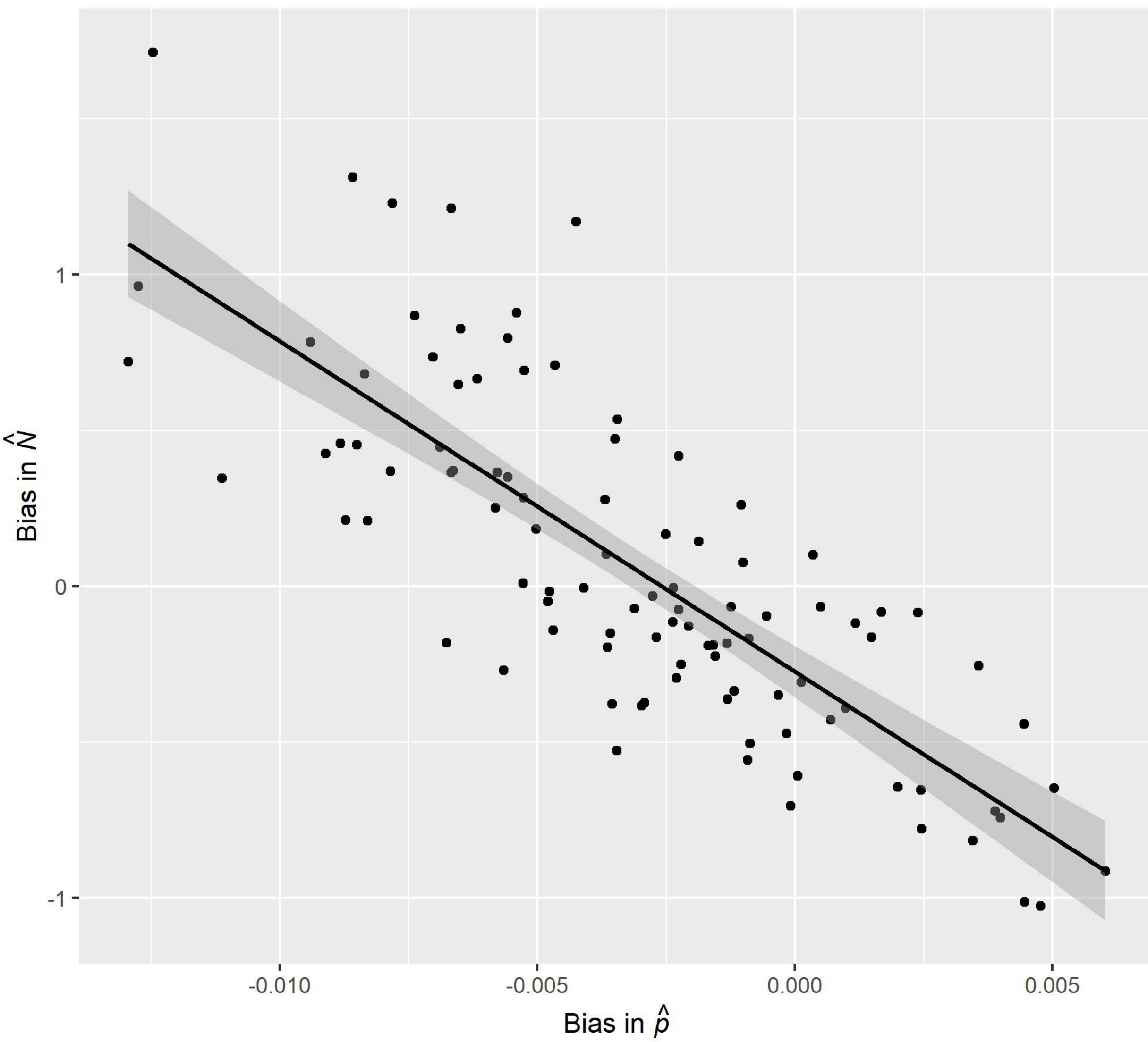


Figure 3. Scatterplot with linear regression relationship shown to help visualize correlation in bias of estimated abundance ($\widehat{N}$) compared to bias in estimated detection probability ($\hat{p}$).

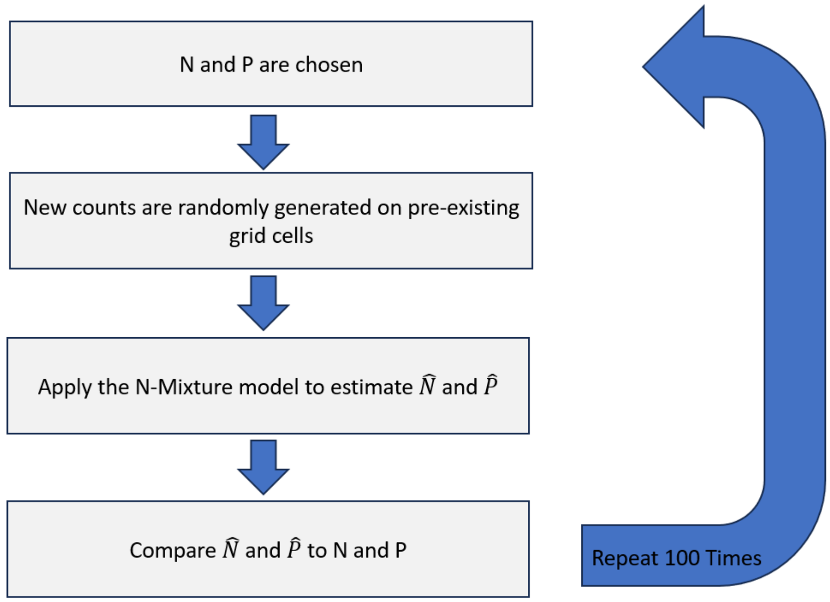


Figure 1: A visual representation of the process used for data simulation. This process was performed using the programing software R.