Making Sense of the Noise: Statistical Analysis of Environmental DNA Sampling for Invasive Asian Carp Monitoring Near the Great Lakes

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Engineering and Public Policy

Civil and Environmental Engineering

Jeffery W. Song

B.S., Environmental Engineering; B.S., Applied Mathematics & Statistics, Johns Hopkins University

Carnegie Mellon University

Pittsburgh, PA

May 2017

To my parents:

Young Kwon Song & Hye Sun Song

Acknowledgements

My Ph.D. was funded by the 2013-2014 Steinbrenner Institute Graduate Research Fellowship, the 2014-2017 National Science Foundation Graduate Research Fellowship (Grant No. DGE-1252522), and the Department of Engineering and Public Policy. I am extremely grateful for these funding sources and the financial support and freedom they gave me to complete my PhD studies and research.

First, I would like to express my deepest gratitude to my PhD advisor and committee chair, Dr. Mitchell Small, for his mentorship, patience and guidance these past five years. I would also like to thank my other committee members: Dr. Elizabeth Casman for her expertise, direction and motivation to complete my PhD this final year, Dr. Martin Schultz for his invaluable expertise, mentorship and help connecting me with the right people to move my research forward, and Dr. Jared Cohon, for his insightful questions about my research and helping me to broaden the scope of my dissertation.

I would alsolike to thank Dr. Emy Monroe, Dr. Katherine Bockrath, Dr. Maren Tuttle-lau, and Dr. Erica Mize at the U.S. Fish and Wildlife Service's Whitney Genetics Lab for their invaluable experimental and sampling data to parameterize my models. My research would not have been possible without their help.

I am incredibly grateful for the amazing EPP and CEE communities of talented professors, staff and students. I would like to especially thank Dr. Paul Fischbeck for getting me initially interested in this research topic, Dr. Jeanne VanBriesen, for her mentoring, feedback and advocacy for my research and Ms. Denise Caruso for her invaluable edits of my literature review.

I cannot thank enough my friends at the Upper Room Church in Pittsburgh. They have been my family away from home and I am so grateful for their support and love.

Finally, I would like to thank my family. I thank my grandfather, Dal Ho Song, for teaching me math when I was barely old enough to read and instilling in me a deep love and appreciation for learning. I thank my parents, Young Kwon Song and Hye Sun Song, for their life-long sacrifices, indomitable spirits and unconditional love.

Abstract

Sensitive and accurate detection methods are critical for monitoring and managing the spread of aquatic invasive species, such as invasive Silver Carp (SC; *Hypophthalmichthys molitrix*) and Bighead Carp (BH; *Hypophthalmichthys nobilis)* near the Great Lakes. A new detection tool called environmental DNA (eDNA) sampling, the collection and screening of water samples for the presence of the target species' DNA, promises improved detection sensitivity compared to conventional surveillance methods. However, the application of eDNA sampling for invasive species management has been challenging due to the potential of false positives, from detecting species' eDNA in the absence of live organisms. In this dissertation, I study the sources of error and uncertainty in eDNA sampling and develop statistical tools to show how eDNA sampling should be utilized for monitoring and managing invasive SC and BH in the United States.

In chapter 2, I investigate the environmental and hydrologic variables, e.g. reverse flow, that may be contributing to positive eDNA sampling results upstream of the electric fish dispersal barrier in the Chicago Area Waterway System (CAWS), where live SC are not expected to be present. I used a beta-binomial regression model, which showed that reverse flow volume across the barrier has a statistically significant positive relationship with the probability of SC eDNA detection upstream of the barrier from 2009 to 2012 while other covariates, such as water temperature, season, chlorophyll concentration, do not. This is a potential alternative explanation for why SC eDNA has been detected upstream of the barrier but intact SC have not.

In chapter 3, I develop and parameterize a statistical model to evaluate how changes made to the US Fish and Wildlife Service (USFWS)'s eDNA sampling protocols for invasive BH and SC monitoring from 2013 to 2015 have influenced their sensitivity. The model shows that changes to

the protocol have caused the sensitivity to fluctuate. Overall, when assuming that eDNA is randomly distributed, the sensitivity of the current protocol is higher for BH eDNA detection and similar for SC eDNA detection compared to the original protocol used from 2009-2012. When assuming that eDNA is clumped, the sensitivity of the current protocol is slightly higher for BH eDNA detection but worse for SC eDNA detection.

In chapter 4, I apply the model developed in chapter 3 to estimate the BH and SC eDNA concentration distributions in two pools of the Illinois River where BH and SC are considered to be present, one pool where they are absent, and upstream of the electric barrier in the CAWS given eDNA sampling data and knowledge of the eDNA sampling protocol used in 2014. The results show that the estimated mean eDNA concentrations in the Illinois River are highest in the invaded pools (La Grange; Marseilles) and are lower in the uninvaded pool (Brandon Road). The estimated eDNA concentrations in the CAWS are much lower compared to the concentrations in the Marseilles pool, which indicates that the few eDNA detections in the CAWS (3% of samples positive for SC and 0.4% samples positive for BH) do not signal the presence of live BH or SC. The model shows that >50% samples positive for BH or SC eDNA are needed to infer AC presence in the CAWS, i.e., that the estimated concentrations are similar to what is found in the Marseilles pool.

Finally, in chapter 5, I develop a decision tree model to evaluate the value of information that monitoring provides for making decisions about BH and SC prevention strategies near the Great Lakes. The optimal prevention strategy is dependent on prior beliefs about the expected damage of AC invasion, the probability of invasion, and whether or not BH and SC have already invaded the Great Lakes (which is informed by monitoring). Given no monitoring, the optimal strategy is

to stay with the status quo of operating electric barriers in the CAWS for low probabilities of invasion and low expected invasion costs. However, if the probability of invasion is greater than 30% and the cost of invasion is greater than $100 million a year, the optimal strategy changes to installing an additional barrier in the Brandon Road pool. Greater risk-aversion (i.e., aversion to monetary losses) causes less prevention (e.g., status quo instead of additional barriers) to be preferred. Given monitoring, the model shows that monitoring provides value for making this decision, only if the monitoring tool has perfect specificity (false positive rate = 0%).

Table of Contents

Acknowledgements iii

Abstract v

Table of Contents viii

List of Figures xii

List of Tables xx

1 Introduction 1

1.1 Motivation 1

1.2 Sources of error and uncertainty in eDNA sampling 6

1.2.1 Errors in eDNA sampling 6

1.2.2 Methodological errors in eDNA sampling 7

1.2.3 Uncertainty in the relationship between DNA and species' presence 10

1.3 Dissertation Overview 13

2 The Effect of Hydrology on Silver Carp eDNA Detection in the Chicago Area Waterway System 15

2.1 Background 15

2.2 Methods 17

2.2.1 Study Site and eDNA Sampling Results 17

2.2.2 Reverse Flow Volume 18

2.2.3 Additional Covariates 19

2.2.4 Beta-Binomial Regression Model 21

2.2.5 Model and Covariate Selection 23

2.3 Results 24
2.3.1 Binomial Regression Model 24
2.3.2 Beta-Binomial Regression Model 25
2.4 Discussion 28
3 Probabilistic Framework for Environmental DNA Sampling Design 31
3.1 Background 31
3.2 Methods and Data 33
3.2.1 Modeling framework 33
3.2.2 Evolution of the eDNA Sampling Protocol for Bighead and Silver Carp 38
3.2.3 Analyzing the influence of changing eDNA sampling protocols over time on BH and SC detection sensitivity 42
3.3 Results 43
3.3.1 Comparing the detection sensitivity of each year's eDNA sampling protocols 43
3.3.2 Analysis of future changes to sampling protocol 48
3.4 Discussion 51
4 Estimation of surface water eDNA concentrations based on statistical models of eDNA sampling protocols 54
4.1 Background 54
4.2 Methods 55
4.2.1 eDNA Study Site and Sampling Results 55
4.2.2 Bayesian Model Framework 57
4.2.3 MCMC model parameterization 63
4.2.4 Evaluating AC occupancy in the CAWS from eDNA sampling results 65

4.2.5 Using simulated eDNA sampling results to determine detection limits for inferring occupancy 66
4.3 Results 66
4.3.1 Posterior capture efficiency parameters 66
4.3.2 eDNA concentrations in the Illinois River and CAWS 66
4.3.3 Occupancy probability in the CAWS for 2014 69
4.3.4 Determining detection limits for inferring AC occupancy in the CAWS using simulated eDNA sampling results 70
4.4 Discussion 71
5 Value of Monitoring Information for Making Decisions about Asian Carp Prevention 74
5.1 Background 74
5.2 Methods and Data 75
5.2.1 Prevention strategies 75
5.2.2 Risk and Potential Impacts of Asian Carp 79
5.2.3 Simulating potential invasion fronts 79
5.2.4 Calculating expected costs and utility given the simulated invasion fronts 81
5.2.5 Decision modeling framework 83
5.2.6 Calculating the value of perfect information 84
5.2.7 Calculating the value of imperfect information 86
5.3 Results 89
5.3.1 Optimal prevention strategy given uncertainty about invasion 89
5.3.2 Influence of risk aversion on decision-making 90
5.3.3 Value of information 92

5.4 Discussion 94
6 Conclusions and Recommendations 96
6.1 Summary 96
6.2 Recommendations 98
6.2.1 Improving the method sensitivity of eDNA sampling protocols 98
6.2.2 Making accurate interpretations from eDNA sampling data 98
6.2.3 Implications for decision-making about AC prevention 99
6.2.4 Implications about AC monitoring in the CAWS and Illinois River 100
6.3 Future Work 100
References 103
Appendix A. Supporting Information for Chapter 2 113
A.1 R Code for Analysis 113
Appendix B. Supporting Information for Chapter 3 122
B.1 Fitting a distribution for the number of filters used per sample 122
B.2 Estimating the capture and extraction efficiencies from experimental data 124
B.3 Dilution assay data and results for both PCR and qPCR markers 127
B.4 Tornado diagrams showing influence of percentage changes to protocol parameters on the protocol's sensitivity 135
Appendix C. Supporting Information for Chapter 4 137
Appendix D. Supporting Information for Chapter 5 140

List of Figures

Figure 1-1. (a) Plan view and (b) elevation profile of the Illinois River from the Mississippi River to Lake Michigan with the current invasion front of BH and SC. The Illinois River has been broken up into pools by a series of lock and dam (L&D) systems. Each pool of the Illinois River is colored to denote the current state of Asian carp invasion: (red) established population, (yellow) some spawning, (green) live adults but no spawning, (pink) positive eDNA results but no adults. 3

Figure 2-1. Sites of positive Asian carp eDNA detection in the Chicago Area Waterway System (CAWS) connecting Lake Michigan to the Des Plaines River (a tributary of the Illinois River). The typical direction of flow is noted with black arrows. The location of the electric fish barrier, wastewater treatment plants, and three stream flow gages in the CAWS are also shown. The only live Asian carp observed upstream of the electric barrier was one bighead carp caught in Lake Calumet in 2010. 16

Figure 2-2. (a) Hydrograph from USGS stream gage 05536890 at Lemont, IL, (b) a hydrograph of only the reverse flows (flow towards Lake Michigan), and (c) the fraction of positive Silver Carp and Bighead Carp eDNA samples for all 191 eDNA sampling events in the CAWS from July 2009 to December 2012. 19

Figure 2-3. The 30-day moving average of reverse flow volume plotted against the fraction of positive Silver Carp eDNA samples for each sampling event from 2009 to 2012. The final

beta-binomial regression model fitted to the data is also plotted along with its 95% confidence level (dotted lines) 28

Figure 3-1. A schematic of the eDNA sampling process from sample collection to the determination of a positive sample. The circles in the center denote the probabilistic representation of the amount of DNA at each step and the likelihood of detecting the target DNA. The squares on the right detail the parameters of the sampling protocol that influence each step of the process. 34

Figure 3-2. The simulated per-sample detection sensitivity for (a) Bighead Carp and (b) Silver Carp eDNA as a function of mean eDNA concentration for each year's sampling protocol. It is assumed that the eDNA was randomly distributed (r = 100). The sensitivity of the 2009-2012 protocol is shown in gray in the other plots for comparison. The solid line is the mean value and the dashed lines represent the 95% credible intervals of the estimated detection sensitivity. 44

Figure 3-3. The simulated per-sample detection sensitivity for (a) Bighead Carp and (b) Silver Carp eDNA as a function of mean eDNA concentration for each year's sampling protocol. It is assumed that the eDNA was clumped (r = 0.3). The solid line is the mean value and the dashed lines represent the 95% credible intervals of the estimated detection sensitivity. 47

Figure 3-4. The minimum mean Silver Carp eDNA concentration in a water body that is detectable with per-sample detection sensitivity of 95% over a range of (a) sample volume, (b) capture efficiency, (c) extraction efficiency, (d) elution volume (e) reaction volume, and (f) number of replicates. The current sensitivity is noted with a circle at 41,700 copies/L for the 2015-2016 sampling protocol, which is a sample size of 0.25 liters, capture efficiency of 6.4%, extraction efficiency of 33%, elution volume of 200 µL, a reaction volume of 3 µL

and 8 replicates. For each graph, only the parameter of interest was changed while the other parameters were kept constant. It is assumed that the eDNA was randomly distributed and unclumped (r = 100). The solid line denotes the mean value while the dashed lines represent the 95% credible interval. 49

Figure 3-5. The minimum mean Silver Carp eDNA concentration in a water body that is detectable with per-sample detection sensitivity of 95% over a range of (a) sample volume, (b) capture efficiency, (c) extraction efficiency, (d) elution volume (e) reaction volume, and (f) number of replicates. The current sensitivity is noted with a circle at 1,320,000 copies/L for the 2015-2016 sampling protocol, which is a sample size of 0.25 liters, capture efficiency of 6.4%, extraction efficiency of 33%, elution volume of 200 µL, a reaction volume of 3 µL and 8 replicates. For each graph, only the parameter of interest was changed while the other parameters were kept constant. It is assumed that the eDNA was clumped (r = 0.3). The solid line denotes the mean value while the dashed lines represent the 95% credible interval. 51

Figure 4-1. The estimated mean concentration (μ_{ij}) of each marker (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) in the three pools of the Illinois River (La Grange, Marseilles, Brandon Road) sampled in Spring 2014 and in the CAWS, which was sampled in June 2014. The circles indicate the mean and the thick and thin lines show the 50% and 95% credible intervals, respectively. 67

Figure 4-2. The cumulative negative binomial density function of the number of copies of the AC-TM1 marker collected in a 2-L sample from the three pools of the Illinois River (La Grange, Marseilles, Brandon Road) and in the CAWS, given their respective means and the overall dispersion parameter, r, value of 0.18 (0.17-0.20). The 95% credible intervals are

shown as dotted lines. The x-axis for the number of copies is shown as log10(1 + copies) for interpretability 69
Figure 4-3. The probability of Asian carp occupancy in the CAWS (red – 1; yellow – 0) given the fraction of positive Silver Carp and Bighead Carp samples for a simulated sampling event of 228 samples. The actual historical results of the sampling event performed in June (SC = 7/228, BH = 1/228) is shown as a black dot. 71
Figure 5-1. Decision tree for analyzing the optimal preventative strategy (1. Status quo, 2. Additional barrier, 3. Hydrologic separation and 4. Stop prevention) given the prior probability of invasion at time = 0, p_0, and the NPV of each strategy for a simulated invasion front. 84
Figure 5-2. Decision tree for evaluating the optimal preventative strategies (1. Status quo, 2. Additional barrier, 3. Hydrologic separation and 4. Stop prevention) given a perfect monitoring method, the prior probability of invasion at time = 0, p_0, and the NPV of each strategy for a simulated invasion front. 85
Figure 5-3. Decision tree for evaluating the optimal preventative strategies (1. Status quo, 2. Additional barrier, 3. Hydrologic separation and 4. Stop prevention) given an imperfect monitoring method with sensitivity (sens), specificity (spec), the prior probability of invasion at time = 0, p_0, and the NPV of each strategy for a simulated invasion front. 87
Figure 5-4. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given risk neutrality, time frame of 50 years and a discount rate of 4%. The colors denote the optimal prevention strategy for each

combination of invasion costs and base probability of invasion. Red is the status quo, blue is additional barrier, and green is discontinue existing barriers..... 90
Figure 5-5. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given risk tolerances of (a) $1 billion and (b) $500 million. Red is the status quo, blue is additional barrier, and green is discontinue existing barriers. 91
Figure 5-6. Expected value of perfect information in units of million dollars as a function of the annual cost of invasion and the base probability of invasion, given risk neutrality, time frame of 50 years and a discount rate of 4%. 92
Figure 5-7. The expected value of imperfect information in units of million dollars as a function of (a) specificity assuming sensitivity is 100% and as a function of (b) sensitivity, assuming specificity is 100%. The calculation assumes risk neutrality, time frame of 50 years, a discount rate of 4%, base invasion probability of 50% and an annual invasion cost of $500 million. 93
Figure A-1 The Chicago Area Waterway System (CAWS) connecting Lake Michigan to the Des Plaines River (a tributary of the Illinois River). The names of Chicago's Metropolitan Water Reclamation District (MWRD) water quality monitoring locations are shown in the boxes

and are colored to match their corresponding eDNA monitoring location in the stream. The figure is adapted from [121]. 121
Figure B-1. Distribution of elution volume for filtration with the PW kit (2009-2012 protocol), filtration with the DN kit (2013-2014 protocol) and centrifugation with the DN kit (2015 protocol) 123
Figure B-2. Results of dilution assay for amplifying BH and SC DNA using the PCR markers BH-PCR and SC-PCR showing the proportion of replicates that are successfully amplified as a function of the expected number of target copies in the replicate and the curve for the fitted distribution of ψ for each marker. 132
Figure B-3. Results of dilution assay for sequencing BH and SC DNA using the markers BH-PCR and SC-PCR showing the proportion of replicates that are successfully sequenced as a function of the expected number of target copies in the replicate and the curve for the fitted distribution of ψ for each marker. 132
Figure B-4. Results of dilution assay for amplifying BH and SC DNA using the qPCR markers ACTM1, ACTM3, BHTM1, BHTM2, SCTM4, and SCTM5 showing the proportion of replicates that are successfully amplified as a function of the expected number of target copies in the replicate and the curve for the fitted distribution of ψ for each marker. 133
Figure B-5. Comparison of the probability of a positive sample for the PCR and qPCR methods given the number of copies in a replicate for Bighead Carp and Silver Carp eDNA detection. The solid line denotes the mean probability value and the dashed lines show the 95% credible interval. 134
Figure B-6. Tornado diagram showing the change in the minimum mean SC eDNA concentration detectable with probability 95% given a 10% decrease and increase for each of

the six sampling protocol parameters. The vertical axis is located at 41,700 copies/L, which is the sensitivity of the current 2015-2016 sampling protocol, i.e., a sample size of 0.25 liters, capture efficiency of 6%, extraction efficiency of 33%, elution volume of 200 µL, a reaction volume of 3 µL and 8 replicates. It is assumed that the eDNA was randomly distributed (r = 100). 135

Figure B-7. Tornado diagram showing the change in the minimum mean SC eDNA concentration detectable with probability 95% given a 10% decrease and increase for each of the six sampling protocol parameters. The vertical axis is located at 1,320,000 copies/L, which is the sensitivity of the current 2015-2016 sampling protocol, i.e., a sample size of 0.25 liters, capture efficiency of 6%, extraction efficiency of 33%, elution volume of 200 µL,

a reaction volume of 3 µL and 8 replicates. It is assumed that the eDNA was clumped (r = 0.3). 136

Figure C-1. A three-stage site occupancy model when interpreting eDNA sampling data 138

Figure C-2. The prior (grey) and posterior (black) distributions of the capture efficiency (ϕ_c) of the filtration method. 139

Figure C-3. The prior (grey) and posterior (black) distributions of the capture efficiency (ϕ_c) of the centrifugation method. 139

Figure D-1. Annual costs for implementing each of the prevention strategies over the 50 year time frame. The stop prevention strategy is not shown, since that will have zero preventative costs. 140

Figure D-2. The exponential utility function (Utility = $1 - e^{-NPV/R}$) for two risk tolerances (R), $1 billion and $500 million. 141

Figure D-3. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given risk neutrality, time frame of 100 years and a discount rate of 4%. The colors denote the optimal prevention strategy for each combination of invasion costs and base probability of invasion. Red is the status quo, blue is additional barrier and green is stop prevention. 142

Figure D-4. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given “environmental” risk tolerance of $1 billion. Red is the status quo, blue is additional barrier and green is stop prevention. ... 143

List of Tables

Table 1-1. The frequency of positive Bighead Carp and Silver Carp eDNA samples found in the Chicago Area Waterway System (CAWS) from 2009 to 2016 5

Table 2-1 Water quality and seasonal variables in the CAWS and their potential relationship with Asian Carp eDNA detection..... 20

Table 2-2. Regression model estimates for the full and final binomial regression model to assess the effect of reverse flow volume and other factors on SC eDNA detection in the CAWS from 2009 to 2012. The significant covariates (p-value <0.05) are bolded for each model.. 25

Table 2-3. Regression model estimates for the full and final beta-binomial regression model to assess the effect of reverse flow volume and other factors on SC eDNA detection in the CAWS from 2009 to 2012. The significant covariates (p-value <0.05) are bolded for each model..... 26

Table 3-1. The methods used in the Bighead and Silver Carp eDNA monitoring protocol as specified in the Quality Assurance Project Plan (QAPP) for the years 2009-2016. The major changes to the protocol each year are in bold text. 39

Table 3-2. The minimum mean eDNA concentration in the water body detectable with probability 95% after each change in the sampling protocol for both carp species, assuming that the eDNA was randomly distributed and unclumped (r = 100). The point value is the mean value and the values in parentheses represent the 95% credible interval..... 46

Table 3-3. The minimum mean eDNA concentration in the water body detectable with probability 95% after each change in the sampling protocol for both carp species, assuming

that the eDNA was clumped (r = 0.3). The point value is the mean value and the values in parentheses represent the 95% credible interval. 48

Table 4-1. The mean number of copies of each of the target markers in the eight qPCR replicates tested per sample (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) from each combination of sampling location (Marseilles pool, Brandon Road pool) and capture method (filtration or centrifugation). 25 samples were analyzed for each combination..... **Error! Bookmark not defined.**

Table 4-2. Priors for the sampling protocol parameters for the eDNA sampling performed in the Illinois River (La Grange, Marseilles and Brandon Road pool) and the CAWS in 2014..... 64

Table 5-1. Prevention strategies in the Illinois River and their cost, time frame, and preventative ability..... 76

Table 5-2. The equations to simulate the invasion front of AC from t = 0 to T for each of the four preventative strategies given the base probability of invasion, p_b. 81

Table 5-3. The net present value (NPV_i) calculation for each of the preventative strategies given the annual cost of invasion, C_I, the discount rate, d, and the invasion front I_t. 82

Table A-1. The number of positive Silver Carp and Bighead Carp eDNA samples for all sampling events from 2009-2012 in the CAWS upstream of the electric barrier. The date and location of each sampling event is also listed. Starting July 2010, an interagency group led by the US Army Corps of Engineers (USACE) took over eDNA sampling analysis from University of Notre Dame (UND)..... 115

Table A-2. The p-values of the β coefficient of the reverse flow volume covariate when fitting logit(p) in a beta-binomial model and the AIC of the model fit for a range of different time frames and types of moving averages of the reverse flow volume covariate: (a) 15, 30, 60

and 90-day; (b) simple and exponential. The exponential moving average is a weighted average using an alpha value of 2/(# of days+1). The best fitting covariate (30-day moving average) in terms of AIC is bolded. 120

Table A-3. The goodness of fit in terms of AIC for four different model types (Binomial, Beta-Binomial, Zero-Inflated Binomial, Zero-Inflated Beta-Binomial) using RevVol (30-day moving average of reverse flow volume) as the only covariate. The best fitting model (Beta-Binomial) in terms of AIC is bolded. 120

Table B-1 The posterior distributions of the seven unknown parameters (ϕc,1 – capture efficiency of centrifugation; ϕc,2 – capture efficiency of filtration; ϕe – extraction efficiency of IBI kit; λ1 – mean concentration of 103 dilution; r1 – dispersion parameter of 103 dilution; λ2 – mean concentration of 104 dilution; r2 – dispersion parameter of 104 dilution) fitted using the MCMC algorithm. 126

Table B-2. Correlation matrix of the seven unknown parameters ($\phi_{c,1}$ – capture efficiency of centrifugation; $\phi_{c,2}$ – capture efficiency of filtration; ϕ_e – extraction efficiency of IBI kit; λ_1 – mean concentration of 10^3 dilution; r_1 – dispersion parameter of 10^3 dilution; λ_2 – mean concentration of 10^4 dilution; r_2 – dispersion parameter of 10^4 dilution) fitted using the MCMC algorithm. 126

Table B-3. The frequency of successful PCR amplification for a series of dilutions of Bighead Carp and Silver Carp DNA. Data from Schultz and Lance (2015).[84] 128

Table B-4. The frequency of successful sequencing of the target DNA sequence for a series of dilutions of Bighead Carp and Silver Carp DNA. Data from Schultz and Lance (2015).[84] 129

Table B-5. The frequency of successful qPCR amplification of each marker for a series of dilutions of silver carp DNA (ACTM1, ACTM3, SCTM4, SCTM5) or bighead carp DNA

(BHTM1, BHTM2). These data were obtained from Farrington et al. 2015.2 The frequency is calculated 130
Table B-6. The posterior distribution of the probability that a single DNA copy amplifies or sequences from dilution assays testing the amplification of the PCR and qPCR markers and sequencing of the PCR markers over a range of target DNA concentrations 131
Table B-7. The frequency of positive Bighead Carp and Silver Carp eDNA samples found in all US waterways from 2013 to 2016 131
Table C-1. The number of positive samples of each target marker (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) for each combination of sampling location in the Illinois River(La Grange pool, Marseilles pool, Brandon Road pool) and capture method (filtration or centrifugation). 137
Table C-2. The estimated mean eDNA concentration distribution of each target marker (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) in each sampling location in the Illinois River (La Grange pool, Marseilles pool, Brandon Road pool) and the CAWS. The value in larger font is the mean and the values in the parentheses are the 95% credible intervals of the distribution. 137

1 Introduction

1.1 Motivation

In the last decade, environmental scientists have been screening water samples from lakes, rivers and oceans testing for the DNA of invasive fish and other aquatic species.[1–4] This process, known as environmental DNA (eDNA) sampling, is based on the premise that aquatic organisms leave traces of their DNA in the water as they move through a habitat, and the detection of DNA in a water sample should indicate the recent presence of the organism. This technology promises improved detection sensitivity and reduced cost compared to conventional monitoring methods, like netting and electrofishing.[5–9] However, eDNA sampling has also been shown to be able to detect trace amounts of eDNA from secondary sources, like fish slime on boat hulls, in the absence of live organisms.[10] This potential error in monitoring has caused considerable controversy, especially in the application of eDNA sampling for monitoring invasive Bighead Carp (BH; *Hypophthalmichthy nobilis*) and Silver Carp (SCC: *Hypophthalmichthy molitrix*) near the Great Lakes.

BH and SC, both of which are referred to as Asian carp (AC), are voracious high-volume filter-feeders that collect and consume plant and animal plankton.[11] Because of this characteristic, BH and SC were imported from Southeast Asia to the southern United States in the 1970s to clean algae in commercial fish tanks and retention ponds in wastewater treatment facilities.[12] They soon escaped via flooding into the Mississippi River, and are out-competing native planktivorous fish species.[13,14] SC are also notorious for their ability to leap three meters out of the water when startled, which has caused injuries to boaters and other recreational water users.[13,15] BH and SC are continuing to spread up the Illinois, Ohio and Upper Mississippi rivers and now threaten to

invade the Great Lakes, putting the region's fishing, recreational and tourism industries at risk.[15] The primary pathway of invasion into the Great Lakes is the Chicago Area Waterway System (CAWS), a series of man-made canals, which connects the Illinois River to Lake Michigan.

The current invasion status of Asian carp in the Illinois River is shown in Figure 1-1. [16] The Illinois River is broken up into a series of pools by lock and dam systems (L&D), which facilitate ship movement upstream (Figure 1-1b). In the lower Illinois River (Alton, La Grange and Peoria pools), Asian carp species are well-established with verified spawning. Further upstream, in the Starved Rock and Marseilles pools, small populations of adult Asian carp, juveniles and larvae have been caught. In the Dresden pool, about 47 miles downstream from Lake Michigan, adults have been caught, but no spawning has been spotted in this pool to-date. This is considered to be the leading edge of Asian carp invasion in the Illinois River, according to conventional monitoring methods. In 2002, the US Army Corps of Engineers (USACE) installed an electric fish dispersal barrier near the exit of the CAWS, about 37 miles downstream from Lake Michigan, to deter the upstream passage of AC, while maintaining the use of the CAWS for navigation and receipt of Chicago's storm water and sewage flows.

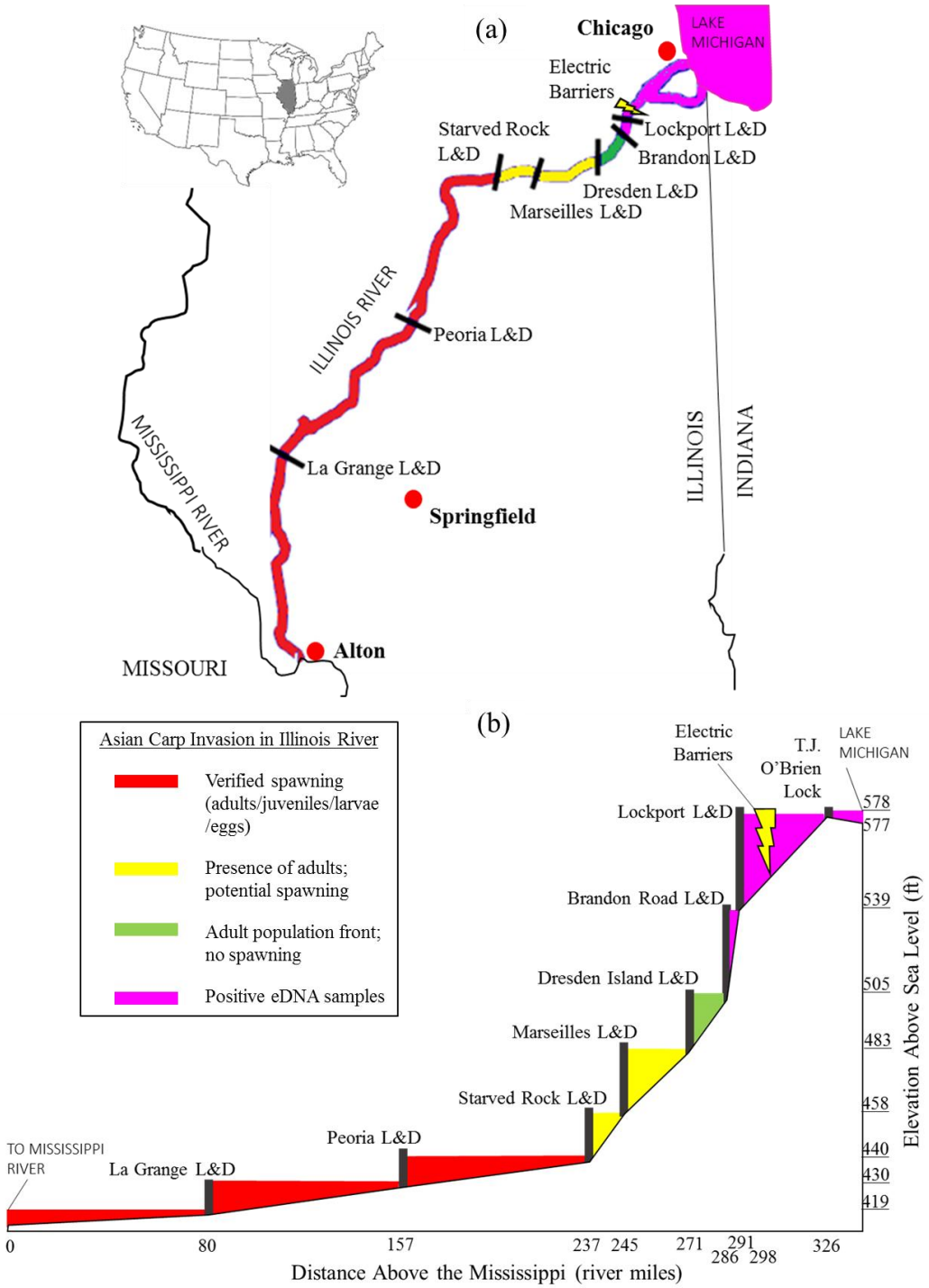

Figure 1-1. (a) Plan view and (b) elevation profile of the Illinois River from the Mississippi River to Lake Michigan with the current invasion front of BH and SC. The Illinois River has been broken up into pools by a series of lock and dam (L&D) systems. Each pool of the Illinois River is colored to denote the current state of Asian carp invasion: (red) established population, (yellow) some spawning, (green) live adults but no spawning, (pink) positive eDNA results but no adults.

Then, in 2009, researchers at the University of Notre Dame (UND), in consultation with USACE, developed an eDNA sampling protocol to be used as a novel detection tool for AC monitoring. Alarmingly, UND found positive samples of BH and SC eDNA throughout the Brandon Road pool, the Lockport pool below the barrier and throughout the CAWS well upstream of the electric barrier and far upstream of the conventionally-held invasion front.[17] The positive eDNA samples triggered intensive management efforts by USACE, most extremely, the repeated treatment of the CAWS with the fish toxin, rotenone. The first rotenone treatment in the CAWS near the electric barrier in December of 2009 found just one AC, an adult BH just downstream of the electric barrier.[18] A second rotenone treatment, applied in a stretch of the canal just six miles from Lake Michigan in May of 2010, yielded 100,000 pounds of dead fish comprising 40 species, but found no AC.[19]

In 2010, USACE took responsibility for eDNA sampling near the Great Lakes and continued to find positive samples of BH and SC eDNA upstream of the electric barrier in the CAWS, triggering multiple follow-up intensive capture efforts, which yielded zero AC (Table 1-1). To date, only one live AC (a 20-lb BH caught in Lake Calumet by commercial fisherman during regularly-scheduled conventional monitoring in June of 2010) has been seen or caught upstream of the barrier in the CAWS, despite thousands of hours of regular and triggered electrofishing and netting efforts.[20] In 2013, the U.S. Fish and Wildlife Service (USFWS) took responsibility for eDNA sampling from USACE and also continued to find positive samples of eDNA in the CAWS. However, positive eDNA sampling results in the CAWS no longer trigger follow-up management actions due to increased skepticism about the accuracy of the technique.[21]

Table 1-1. The frequency of positive Bighead Carp and Silver Carp eDNA samples found in the Chicago Area Waterway System (CAWS) from 2009 to 2016

Year (Laboratory)	eDNA samples from Chicago Area Waterway System (CAWS) [positive samples/total samples (%)]	
	Bighead Carp	Silver Carp
2009 (UND)	32/562 (5.7%)	15/562 (2.7%)
2010 (UND/USACE)	4/1182 (0.3%)	17/1182 (1.4%)
2011 (USACE)	24/2362 (0.0%)	34/2362 (1.4%)
2012 (USACE)	4/1194 (0.2%)	153/1194 (13.0%)
2013 (USFWS)	0/398 (0.0%)	21/398 (5.3%)
2014 (USFWS)	1/456 (0.2%)	30/456 (6.6%)
2015 (USFWS)	0/228 (0.0%)	0/228 (0.0%)
2016 (USFWS)	1/228 (0.4%)	1/228 (0.4%)

The discrepancy between the positive eDNA sampling results and the lack of live Asian carp captures upstream of the barrier in the CAWS raised the question: Are the positive eDNA samples false positives due to the detection of eDNA in the absence of AC (i.e., detecting eDNA from secondary sources) or are the subsequent conventional monitoring methods too insensitive — compared to eDNA sampling — to locate a small number of the target species in the CAWS?

Any effective invasive species monitoring tool requires both high species detection sensitivity and specificity. High sensitivity (the probability of detecting species given the species is present) is needed to be able to detect the invasive species when they are present at low densities. High specificity (the probability of not detecting species given the species is not present) is required to have high confidence in positive detections and reliably act upon that information. Conventional

monitoring methods, like electrofishing, suffer from poor detection sensitivity, especially for monitoring species at very low densities.[8,22] In contrast, eDNA sampling has been shown to have much higher sensitivity, but may suffer from poor specificity, due to the potential of false positives from detecting target eDNA in the absence of the target species.[7,8,10,23] The goal of this dissertation is to analyze the sources of error and uncertainty of eDNA sampling and evaluate the detection performance of eDNA sampling in order to help inform how this tool should be applied for invasive species monitoring.

1.2 Sources of error and uncertainty in eDNA sampling

1.2.1 Errors in eDNA sampling

There are two types of errors for any monitoring method: Type I error (false positive) and Type II error (false negative). A conventional monitoring method, like fishing, tests the null hypothesis that the target species is not present and the alternative hypothesis that the target species is present. Here, a false negative is defined as not detecting the target species when present (i.e., falsely accepting the null hypothesis), and a false positive is detecting the target species, when it is not present (i.e., falsely rejecting the null hypothesis).

However, eDNA sampling does not see or capture live organisms, but rather it detects their DNA, whose source is uncertain, e.g., excreta from live organisms, tissue from dead organisms, transfer from outside of system.[10,24,25] Therefore, eDNA sampling does not detect the presence of the target species but rather the presence of the target species' DNA. eDNA sampling tests the null hypothesis that the target species' DNA is not present and the alternative hypothesis that the target species' DNA is present. Thus, a false negative is a non-detection of target DNA when the DNA is present, and a false positive is a detection of target DNA when the DNA is not present. I define

these errors as the “methodological” errors of eDNA sampling. Further information is needed to infer whether or not a positive detection of target DNA indicates the presence of the target species.

In certain scenarios, eDNA sampling is applied where the species is known to be present (e.g., in controlled laboratory settings or independent observation of species abundance). Here a correct positive detection of eDNA could be assumed to be a detection of the presence of the organism, and the methodological sensitivity is equivalent to the sensitivity of detecting the presence of the species. However, in most scenarios of aquatic invasive species monitoring, the presence of the species is unknown or highly uncertain. Therefore, even if eDNA sampling has high method sensitivity (i.e., high probability of detecting the DNA when the DNA is present), it may lead to erroneous inferences about species presence, if it detects DNA present in the system from secondary sources, but the species is not actually present

The next section reviews what is currently known about the methodological errors in eDNA sampling and the relationship between eDNA presence and species presence in aquatic water bodies.

1.2.2 Methodological errors in eDNA sampling

1.2.2.1 Errors in sample collection, capture and extraction

The first step in eDNA sampling is water sample collection and capture. Here, error is introduced from the heterogeneous dispersion of eDNA in the water. eDNA may not be collected or captured, though present at the sampling location.

Typically, eDNA researchers have collected water at the sampling location using containers and transported them to the lab for capture of the DNA via filtration, centrifugation or precipitation. Each method has its own capture efficiency (i.e., the fraction of DNA in the sample that is

captured). Studies evaluating the methods differ significantly on which method is most efficient, which indicates that the methods may be influenced by the water and target species characteristics and thus the optimal method may be application dependent.[26–28]

After collection and capture, the eDNA is then extracted and purified to recover as much of the eDNA as possible from the filtrate, while removing potential inhibitory materials. Field samples often contain substances, such as humic acids from the biodegradation of organic matter, which can inhibit the detectability of DNA and thus produce false negatives.[29] To mitigate this risk, researchers have tested a variety of extraction kits and protocols, each of which has its own influence on the amount and quality of eDNA extracted for analysis.[30–33] These studies show that choosing the optimal extraction kit may also be dependent on the monitoring location and so pilot studies are recommended before selecting an extraction method.[24]

1.2.2.2 Errors in the detection of target eDNA sequence

The final step is to screen the extracted eDNA from the water sample for the presence of the species-specific DNA sequence using polymerase chain reaction (PCR).[17,34] Here, false negative errors can occur if there are too few initial molecules of target DNA present in the reaction, which lowers the likelihood for the designed primers to bind to the species-specific DNA sequence and amplify it.

Primers are short, created DNA fragments that contain the sequence complementary to the beginning of a species-specific DNA sequence (marker). The primers must be designed to be species-specific or they will attach to any and all DNA in the sample, causing cross-amplification (the amplification of other species' DNA) and thus false positives. One study showed that the presence of similar species' DNA (bull trout, brook trout, lake trout) can cause cross-amplification

or interference due to the similarity in their gene sequences. This work suggests that primers must be carefully designed to maximize the number of differences with non-target species' gene sequences.[35] However, developing primers with more mismatches often means designing longer primers, which can have trouble attaching to the target DNA, especially in environmental samples where the DNA may be heavily degraded.

As the reaction progresses, the copies of the target DNA sequence that are generated become the new templates for replication, which sets off an exponential amplification of the target sequence. This process has made PCR a very sensitive detection tool, able to amplify just a few molecules of the target DNA. However, this high sensitivity creates the risk of contamination and false positives, since PCR can react to very small concentrations of target DNA spread inadvertently throughout the lab. Lab contamination can be contained through strict clean-lab protocols, decontamination procedures and separate labs for DNA extraction and DNA amplification, but it is not fool-proof.[36]

Currently, the most popular PCR method is real-time or quantitative PCR (qPCR). qPCR measures the exponential amplification throughout the PCR process, rather than just evaluating at the end via gel electrophoresis. This is done via the addition of sequence-specific DNA probes, which are small DNA molecules (like primers) that are complementary to the middle of the target sequence. They are labelled with a 'tag' that fluoresces (or lights up) only when the target sequence is assembled by DNA polymerase.

One benefit of qPCR, compared to conventional PCR, is that the initial quantity of target DNA in the sample can be estimated. The ability to reliably and accurately quantify the amount of initial eDNA in the sample is key for providing scientifically rigorous quantitative analysis of eDNA

concentration, degradation rate and production rate. However, there is significant measurement uncertainty in qPCR. Studies show that for the same sample, the estimated DNA concentrations can diverge significantly for different laboratories.[30,37] The authors of these studies suggest that the high variability is due to human error when developing standard curves and preparing the reactions (e.g., pipetting error).

1.2.3 Uncertainty in the relationship between DNA and species' presence

The previous discussion of methodological errors helps evaluate whether or not the target DNA is present or not, given eDNA sampling results. However, in order to make correct inferences about target species' presence from eDNA sampling results, i.e., whether a positive detection of the target species' eDNA indicates the recent presence of the live organism at the sampling location, decision-makers must understand the relationship between the presence of the target species' eDNA and the presence of the target species in a given water body. The next section reviews what is currently known about the manner and rate in which DNA enters the environment from aquatic organisms and how eDNA reacts to environmental conditions and variables.

1.2.3.1 Origin and production of eDNA

For macro-organisms, like fish, the primary source of eDNA is likely to be cells and tissues in the intestinal linings that are sloughed off and excreted with feces.[38] eDNA can also be discharged through bodily excretions.[17,34,39] Klymus found that increasing the amount of available food for Asian carp species in an aquarium increased the quantity of eDNA in the water, which suggests that eDNA production may be related with metabolism (i.e., fecal matter).[40] These sources link the presence of target eDNA in the water to the recent presence of the live organism. Researchers have uncovered correlation between eDNA concentration and organism abundance for a variety of

species (bluegill sunfish, common carp, Idaho salamanders, silver carp and bighead carp) in the laboratory and some field settings.[40–46] However, this relationship is not always found in the field.[47] Also, upon closer examination of the regression parameters, these are weak correlations with large variability.[48] This relationship between eDNA concentration and species biomass is also complicated by potential alternative sources, such as carcasses, ship hulls carrying fish slime and the feces of predators that feed on the target species, which have all been shown to contribute enough eDNA for detection.[10]

1.2.3.2 Temporal and spatial uncertainty

Another area of uncertainty is the transport, settling and persistence of eDNA in the water after it is released. These processes may mean that target eDNA is not necessarily proximate in space or time to an organism. The assumption of proximity can also lead to errors in interpretation. In a field study of eDNA transport, Deiner observed that eDNA from two lake dwelling invertebrate species was detected approximately 12 kilometers downstream from its lake habitat.[49] Another field study found that DNA from a caged trout was detected up to 2000 meters downstream.[50] These initial results show that eDNA can be transported significant distances and there is the possibility of 'false' positives from detecting transported eDNA far away from where the target organism is actually present. A recent study of eDNA transport in a mesocosm showed that eDNA does not follow the same transport dynamics as a conservative tracer and there is significant unexplained variability in how eDNA is transported downstream.[51]

eDNA can also settle and accumulate in lake and stream sediments. When Turner took sediment and water samples from an artificial pond stocked with common carp, he found much higher concentrations of its eDNA in sediment samples, compared to the water samples.[52] Also, eDNA

in the sediment was detected up to 130 days after the carp were removed, compared to 30 days in the water. Based on these results, he postulated that eDNA in the sediment may both degrade more slowly compared to eDNA in the water and accumulate. As a result, a release of eDNA from the sediment into the water could complicate inferences from eDNA sampling, since eDNA might be detected long after the organism has been present.

Finally, eDNA can persist and accumulate in the water. In most aquatic environments, DNA will decay as soon as it is shed, due to ultraviolet (UV) radiation from sunlight that breaks the bonds in the DNA molecules or via uptake by microbes in the water.[53] However, the speed of degradation is still not well understood. In recent freshwater laboratory experiments, researchers found that the degradation of eDNA to below the threshold of detectability occurred on a scale of days to months.[39,54–56] Researchers are interested in knowing the eDNA degradation rate, because a fast rate might mean that a positive detection of eDNA is more likely to be associated with the recent presence of the species, and not from picking up persistent DNA in the water from past populations or downstream transport.[53] Significant uncertainty still remains about the mechanisms and rate by which eDNA degrades in water bodies, which further complicates inferences from eDNA sampling.[56–58]

There is considerable uncertainty in the link between eDNA presence and the recent presence of the live organism. Unlike conventional surveillance methods, like fishing, where a positive detection involves physical capture of a live organism, eDNA sampling cannot conclusively determine the presence, abundance, age, gender or size of the target organism.

1.3 Dissertation Overview

In this dissertation, I describe the methods and results of four separate studies (Chapters 2 – 5) developing statistical and decision models to inform if and how eDNA sampling should be applied for BH and SC monitoring and management near the Great Lakes.

In chapter 2, I evaluate the relationship of positive SC eDNA samples upstream of the electric barrier in the CAWS from 2009-2012 to hydrologic and environmental covariates using a beta-binomial regression model. The purpose of this analysis is to evaluate if there are alternative factors that may be influencing or explaining the presence and detection of SC eDNA upstream of the electric barrier in the absence of live organisms.

In chapter 3, I develop a model of the eDNA sampling process to evaluate the sensitivity of the USFWS' eDNA sampling protocols from 2013 to 2015 as a function of the BH and SC eDNA concentration it can detect. This model is applied to determine if changes made to the eDNA sampling protocol from 2013 to 2015 impacted the sensitivity of the protocol over this time period, which protocol parameters are most influential on the sensitivity of the protocol, and how the statistical dispersion of eDNA in the water column influences the sensitivity of eDNA sampling.

In chapter 4, I apply the model developed in chapter 3 to estimate the eDNA concentration distributions (mean and dispersion) of BH and SC in the Illinois River and upstream of the electric barrier in the CAWS, from eDNA sampling results collected in 2014. Then, given the estimated concentration distributions, I evaluate whether the eDNA concentration in the CAWS in 2014 indicate BH and SC presence in the CAWS by calculating the similarity of these concentrations to concentrations where BH and SC are known to be present and absent in the Illinois River. My

model is then utilized to estimate the fraction of positive eDNA samples that is needed to infer BH and SC presence in the CAWS.

In chapter 5, I develop a decision tree model to evaluate the optimal preventative strategy for AC invasion in the CAWS, given uncertainty about the expected cost of invasion, the probability of invasion and the current state of invasion. Then, the model is applied to evaluate the value of information from monitoring for the current state of invasion given a range of perfect to imperfect (sensitivity and specificity less than one) monitoring methods.

2 The Effect of Hydrology on Silver Carp eDNA Detection in the Chicago Area Waterway System

2.1 Background

The Chicago Area Waterway System (CAWS) is the primary waterway of concern for possible transfer of invasive BH and SC into the Great Lakes.[59] The CAWS is a system of man-made canals that were built in the early 1900s for the purpose of carrying Chicago's sewage away from Lake Michigan into the Illinois River and to enable boat and barge traffic between these two water bodies. Currently, the CAWS transports 1,200 million gallons per day of Chicago's treated sewage effluent across the natural watershed boundary into the Mississippi River drainage system.[60] Due to the growing risk of invasive species transfer, the USACE installed an electric fish dispersal barrier at the exit of the CAWS to prevent the upstream passage of invasive BH and SC, while still maintaining the use of the CAWS for Chicago's water quality management and cargo shipping.

However, from 2009 – 2012, eDNA sampling in the CAWS found positive samples of BH and SC eDNA throughout the CAWS, indicating that invasive BH and SC organisms may have bypassed the barrier and invaded the CAWS (Figure 2-1).[3,23,61] These positive eDNA results triggered intensive management and monitoring upstream of the electric barrier; however, only one live BH adult (captured in June 2010) and no SC have been found to date. This raised the possibility that the positive samples of BH and SC eDNA in the CAWS may be due to the presence and detection of allochthonous eDNA, i.e., eDNA brought in from elsewhere by stream flow, predator feces, boats, fish carcasses, and/or sediment.[10,24,62,63]

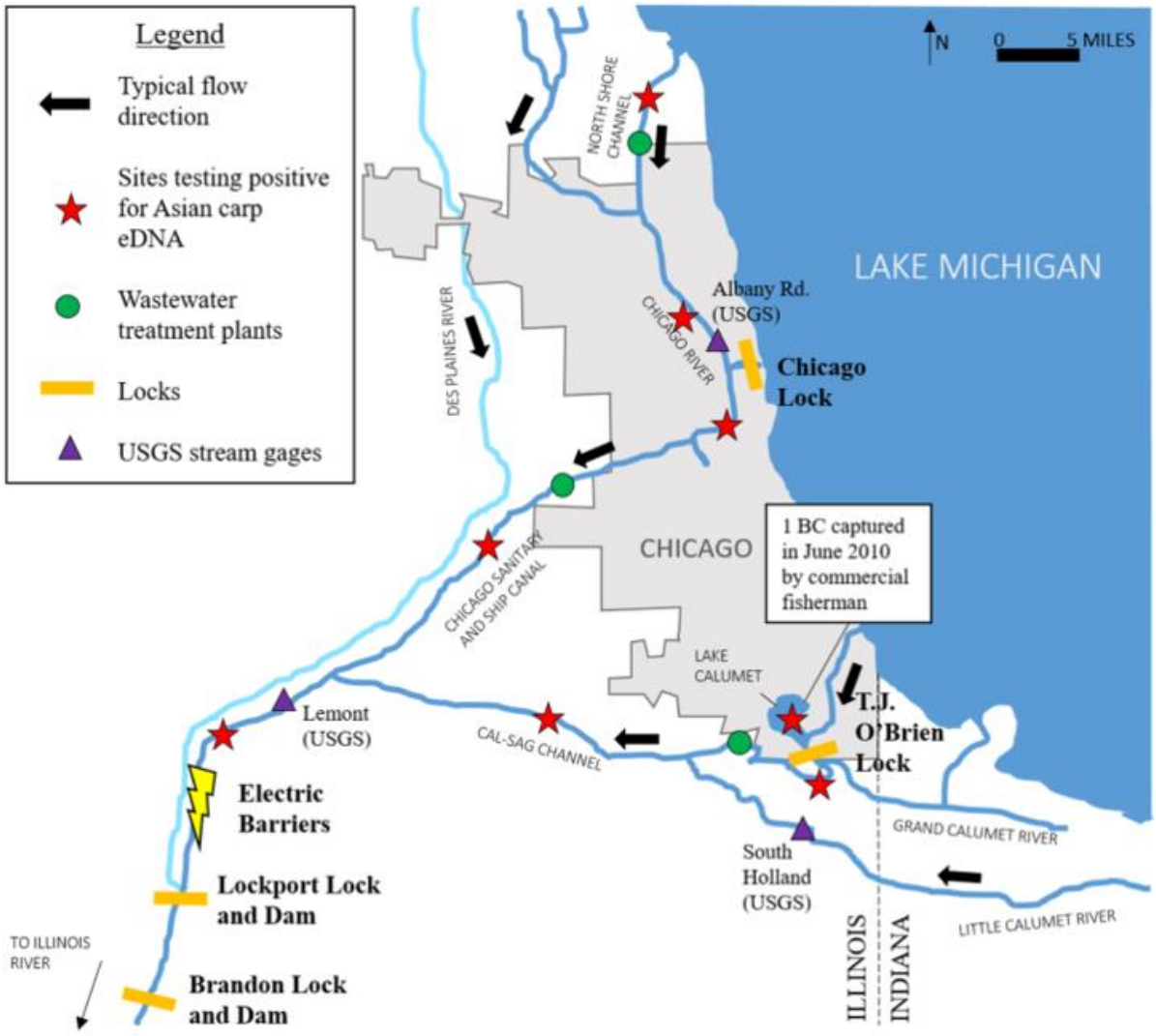

Figure 2-1. Sites of positive Asian carp eDNA detection in the Chicago Area Waterway System (CAWS) connecting Lake Michigan to the Des Plaines River (a tributary of the Illinois River). The typical direction of flow is noted with black arrows. The location of the electric fish barrier, wastewater treatment plants, and three stream flow gages in the CAWS are also shown. The only live Asian carp observed upstream of the electric barrier was one bighead carp caught in Lake Calumet in 2010.

One possible mode of transport of allochthonous eDNA from the Illinois river into the CAWS is simple advection during periods of "reverse flow". By reverse flow I mean water flowing from the Illinois River towards Lake Michigan, instead of away from it. This primarily occurs during dry weather conditions, when the flow of water out of the CAWS is slowed to meet water level regulations for the lake. This causes the water in the canal system to pond. Strong surface winds can then move the water in the lakeward direction. At these low flow periods, wastewater effluent

discharges (see Figure 2-1) can cause density currents, which can further contribute to the reverse flow phenomenon.[60] In this chapter, I determine the influence of environmental variables in the CAWS, including reverse flow events, on AC eDNA detection in the CAWS, by regressing the probability of detection against these variables.

A number of other environmental conditions have also been associated with the probability of detecting eDNA.[42,57,64–66] Prior studies that have examined possible relationships between environmental variables and eDNA sampling results have used a binomial (logistic) regression model for the probability of a positive eDNA detection, which assumes that the eDNA samples are independent and identically distributed.[9,67,68] I argue that the independence assumption probably does not hold for eDNA sampling, because eDNA is likely to be clumped in cells and tissues rather than randomly distributed in the environment.[69] Therefore, correlation is likely to occur between the sampling results in a sampling event. Ignoring such correlation in logistic regression type models will underestimate the standard errors of the parameter estimates, and covariates may be falsely declared to be statistically significant. Therefore, I use a beta-binomial regression model to incorporate this additional correlation.[70]

2.2 Methods

2.2.1 Study Site and eDNA Sampling Results

eDNA sampling for BH and SC detection was performed in the CAWS from 2009 to 2012. On average, 28 samples were collected per sampling event. Most of the eDNA sampling events were located near Lake Michigan in the North Shore Channel, the Chicago River, and Lake Calumet/Little Calumet River (Figure 2-1). From 2009 to 2012, there were 191 eDNA sampling events in these areas comprising 5281 total samples (Table A-1). The sampling was performed by

University of Notre Dame researchers from 2009 to May 2010 and then by an interagency group led by the US Army Corps of Engineers (USACE) from July 2010 to 2012. I selected the time frame of 2009-2012 for this analysis because changes to the eDNA sampling protocol after 2012 affected its sensitivity and specificity [71] The samples were taken mainly during the months of April through October, because the CAWS often freezes in the winter. In total, from 2009 to 2012, 219 out of 5281 (~4%) samples were positive for silver carp eDNA, and 72 of the 191 sampling events (38%) had at least one positive sample of silver carp eDNA. The majority of the positive SC eDNA sampling events occurred in 2012 (37 out of 58 events). I analyzed just the SC eDNA sampling results because there were too few positive BH eDNA sampling events (10/191, 5%).

2.2.2 Reverse Flow Volume

The reverse flow volume into the CAWS prior to each eDNA sampling event was calculated using the 10-minute flow data from the USGS stream gage at Lemont, IL (USGS 05536890), the closest gage upstream from the electric fish dispersal barrier near Romeoville, IL (see Figure 2-1). Figure 2-2a shows the full hydrograph from 2009-2012 at Lemont, IL, while Figure 2-2b presents the hydrograph for only the reverse flow periods. The corresponding eDNA sampling results are shown in Figure 2-2c. In 2009, there were just 48 10-minute periods of reverse flow moving a total of 34 million gallons (MG) in the upstream direction. In 2010, there were 125 10-minute periods, accounting for 80 MG of reverse flow volume, and in 2011, 287 periods for a total of 239 MG of reverse flow volume. In 2012, there was a noticeable increase in reverse flow with a total of 1077 periods corresponding to a total of 929 MG. As shown in Figure 2-2c, 2012 was also the year with the most frequent SC eDNA detection.

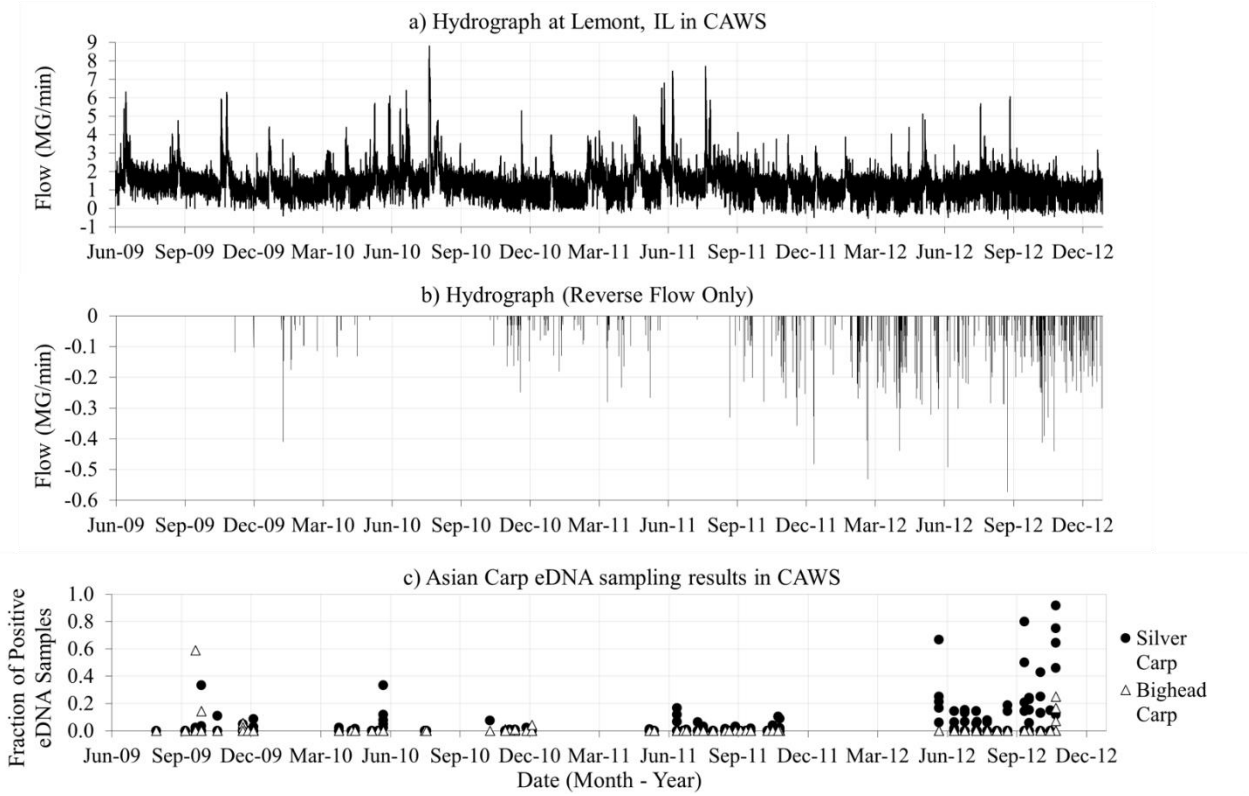

Figure 2-2. (a) Hydrograph from USGS stream gage 05536890 at Lemont, IL, (b) a hydrograph of only the reverse flows (flow towards Lake Michigan), and (c) the fraction of positive Silver Carp and Bighead Carp eDNA samples for all 191 eDNA sampling events in the CAWS from July 2009 to December 2012.

Because of the irregular nature of the flow reversals in the CAWS above the Lemont gage, I took a moving average of the reverse flow volume prior to each eDNA sampling event rather than just the reverse flow volume on the day of the sampling event. For each eDNA sampling event, I calculated the 30-day average of the daily reverse flow volume, which is the average daily reverse flow volume over the preceding 30 days from the sampling date. Other functional forms and averages were tested (e.g., 14-day, 60-day, 90-day, etc.), but the 30-day average provided the best fit (Table A-2).

2.2.3 Additional Covariates

Table 2-1 lists all of the regression variables, including reverse flow volume, that were analyzed for their relationship with eDNA detection in the CAWS. For the water level in the CAWS, I

obtained gage height data at the USGS stream gage in the Chicago River at Albany Rd (USGS 05536105). This represents the influence of dilution due to stream volume. I obtained precipitation data from the USGS gage in the Little Calumet River at South Holland, IL (USGS 05536290). For these two hydrologic variables, I used the 30-day moving average to be consistent with the reverse flow variable.

Table 2-1 Water quality and seasonal variables in the CAWS and their potential relationship with Asian Carp eDNA detection

Variable	Units	Possible relationship with eDNA detection in CAWS
Reverse Flow Volume	30-day moving average of daily reverse flow volume [million gallons]	Possible external source of genetic material into system. May also indicate less dilution of eDNA by Lake Michigan water.
Gage Height	30-day moving average of daily average gage height [feet]	Higher stream levels may dilute eDNA making it more difficult to detect.
Precipitation	30-day moving average of daily precipitation [inches]	Higher precipitation may dilute eDNA making it more difficult to detect.
Season	Spring, Summer, Fall	AC spawning is triggered by high flow (typically in the Spring).
Temperature	Surface water temperature [°Celsius]	Influences desirability of habitat and signals spawning time. Influences eDNA degradation rate. [57]
pH	none	Can stress fish if too high or low. Influences eDNA degradation rate. [65]
Dissolved Oxygen	Dissolved oxygen concentration [mg/L]	Can stress fish if too low.
Chlorophyll	Chlorophyll-a concentration [µg/L]	Indicates presence of algae which is a food source. Influences eDNA degradation rate. [65]

Other water quality variables, such as pH, water temperature, dissolved oxygen and chlorophyll-a concentration, have been shown in past studies to influence eDNA concentration or degradation rate.[42,57,65,66] To assess the influence of these variables, I obtained ambient water quality data in the CAWS from the Metropolitan Water Reclamation District of Greater Chicago's Ambient

Water Quality Monitoring (AWQM) Program. For each eDNA sampling event, I selected the water quality data from the monitoring location that was closest in space and time to the location and date of the sampling event (Figure A-1). I also considered the season, which has been shown to influence Asian carp spawning timing and movement.[72,73] Season was modeled using a categorical variable for spring, summer and fall.

2.2.4 Beta-Binomial Regression Model

Environmental DNA sampling results are typically reported as the number of positive and negative samples in a group of samples. The standard statistical model for such a collection of binary observations is the binomial regression (BR) model. For the BR model, m_{ij} is a binary random variable representing the result of sample j (j = 1,2, …. n_i) in sampling event i, which has a value of 1 if positive, and 0 if negative. y_i is the sum of the n_i binary variables, m_{ij}, which represents the number of positive eDNA samples that occurred in sampling event i. If I assume that each of the samples is independent and identically distributed, then y_i can be modeled with a binomial distribution with probability of a positive sample (p_i) and the number of samples (n_i). The mean of this distribution is $n_i p_i$ and the variance of the distribution is $n_i p_i(1-p_i)$.

$$y_i \sim Binomial(n_i, p_i) \quad (2\text{-}1)$$

For each sampling event i, there is also a set of related covariates that may influence the probability of detection, p_i. The probability of a positive sample, p_i, can be modeled as a linear function of the explanatory variables using a logit link, as shown in equation 2-2. These covariate values are represented by the vector $\boldsymbol{x_i} = (x_{i1},\ldots x_{ic})$ for the c covariates in the model. The coefficients in vector $\boldsymbol{\beta}$ represent the influence of each covariate on the log odds of a positive sample.

$$log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x_i}\,\boldsymbol{\beta} \qquad (2\text{-}2)$$

However, for modeling eDNA sampling, I may not want to assume that the samples are independent and identically distributed. Therefore, I need an additional parameter to account for the possible correlation between the individual samples in a given sampling event. The beta-binomial (BB) regression model extends the BR model by including a scale parameter ρ, which accounts for this correlation. The distribution of the BB model is very similar to the BR model, where y_i follows a binomial distribution, conditional on the probability of a positive sample (p_i*) and the number of samples (n_i) in each sampling event. However, the BB model gives added flexibility by considering p_i* as a random variable that also varies between the individual samples in each of the sampling events. Equation 2-3 shows the model with y_i distributed binomially conditional on p_i^*, which follows a beta distribution including the parameter ρ to model the extra-variation.

$$y_i \sim Binomial(n_i, p_i^*)$$

$$p_i^* \sim Beta\left(\frac{p_i}{\rho}, \frac{1-p_i}{\rho}\right) \qquad (2\text{-}3)$$

The mean of this distribution is $n_i p_i$, as before, but the variance of the distribution is $n_i p_i (1-p_i)\left\{1 + (n_i - 1)\frac{\rho}{1+\rho}\right\}$. The term $\left\{1 + (n_i - 1)\frac{\rho}{1+\rho}\right\}$ acts as a multiplier for the binomial variance. The greater the value for ρ, the larger the variance. If the value is zero, then the multiplier value is 1 and the model will converge to a binomial model. I use a logit link for p_i (see equation 2-2) and an additional log link for ρ *(log(ρ)*= γ_0). I assumed that the ρ value was constant and independent of the variables.

The regression parameter values for each model were fit using maximum likelihood estimation to choose the parameters, $\boldsymbol{\beta}$ and γ_0, for which the observed sampling results ($\boldsymbol{y_i}$'s and $\boldsymbol{n_i}$*'s*) and the covariate values, $\boldsymbol{x}$, were most likely to occur.[74] I also checked to see if a zero-inflated model (i.e., a distribution that allows for frequent zero-valued observatiosn) was a better fit for the data, due to the high number of sampling events with zero positive samples. The zero-inflated adjustment works via a mixture model and an additional parameter, p_0, where y_i is zero with probability p_0 and y_i is binomial or beta-binomial distributed with probability 1- p_0. [75] The beta-binomial regression model out-performed the zero-inflated beta-binomial regression model (Table A-3).

2.2.5 Model and Covariate Selection

I first fit the full BR model with all covariates, and then fit a final BR model using only the covariates that had a statistically significant relationship (p-value < 0.05) in the full BR model. I repeated this for the BB model to see if the BB model out-performed the BR model. I used Akaike's Information Criterion (AIC) to compare the overall fits of all four models. AIC was calculated as $2k - 2\ln(L)$, where k is the number of regression parameters including the intercept and L is the maximized value of the likelihood function.[76] A lower AIC indicates a stronger fit. I also tested the covariates for co-linearity by evaluating their Variance Inflation Factor (VIF), which measures how much the variance of the estimated regression coefficient is inflated as compared to when the variables are not linearly related. There was no significant co-linearity identified. All of the models were analyzed using R 3.2.2.[77] The R code and dataset used for running the analyses can be found in Appendix A. .

2.3 Results

2.3.1 Binomial Regression Model

Table 2-2 displays the coefficient estimates for the full and final binomial regression models. In the full binomial regression model, the variables Reverse Flow Volume (p-value = $5x10^{-20}$), Precipitation (p-value = 0.026), Temperature (p-value = 0.050), pH (p-value = 0.003) and Chlorophyll (p-value = 0.003) had a significant relationship with SC eDNA detection in the CAWS at a significance level of ≤0.05. The variables Gage Height, Season and Dissolved Oxygen were not statistically significant. The variables Reverse Flow Volume, Temperature and pH had a positive relationship with SC eDNA detection, while Chlorophyll and Precipitation had negative relationships. In the final binomial regression model, only these five variables were included, which slightly improved the overall fit of the model.

Table 2-2. Regression model estimates for the full and final binomial regression model to assess the effect of reverse flow volume and other factors on SC eDNA detection in the CAWS from 2009 to 2012. The significant covariates (p-value <0.05) are bolded for each model.

Variable	Category	Results for the binomial regression model					
		Full model			Final model		
		Coefficient	SE	p-value	Coefficient	SE	p-value
logit(p)							
Intercept		**-10.1**	**1.31**	**7.7E-13**	**-8.91**	**1.11**	**1.42E-13**
Reverse Flow Volume		**0.956**	**0.092**	**5.0E-20**	**0.832**	**0.067**	**3.1E-26**
Gage Height		0.449	0.324	0.168			
Precipitation		**-5.15**	**2.29**	**0.026**	-2.65	1.80	0.142
Season	Fall						
	Spring	-0.243	0.217	0.265			
	Summer	0.389	0.222	0.082			
Dissolved Oxygen		0.084	0.084	0.322			
Temperature		**0.050**	**0.025**	**0.050**	**0.057**	**0.018**	**0.002**
pH		**0.579**	**0.196**	**0.003**	**0.574**	**0.160**	**0.004**
Chlorophyll		**-0.116**	**0.039**	**0.003**	**-0.111**	**0.037**	**0.004**
Goodness of Fit							
Number of parameters (k)			10			6	
AIC			588			587	

2.3.2 Beta-Binomial Regression Model

Table 2-3 shows the coefficient estimates for the full and final beta-binomial regression models. In contrast to the binomial regression models, only the variable Reverse Flow Volume (p-value = 1.2×10^{-8}) has a significant relationship with SC eDNA detection at a significance level of ≤0.05 in the full BB model. The variables Precipitation, Temperature, pH and Chlorophyll that were found to be significant in the binomial regression model are no longer significant in the beta-binomial regression model. The AIC of the full beta-binomial regression model is much lower than the AIC of the full binomial regression model (484 vs. 587), showing that the beta-binomial regression

model in general is a stronger fit for modeling the probability of SC eDNA detection. Also, the scale parameters of both the full and final beta-binomial regression models are statistically significant, which shows that there is extra-binomial variation in the data, which is captured in the beta-binomial regression model.

Table 2-3. Regression model estimates for the full and final beta-binomial regression model to assess the effect of reverse flow volume and other factors on SC eDNA detection in the CAWS from 2009 to 2012. The significant covariates (p-value <0.05) are bolded for each model.

Variable	Category	Results for the beta-binomial regression model					
		Full model			Final model		
		Coefficient	SE	p-value	Coefficient	SE	p-value
logit(p)							
Intercept		**-6.37**	**2.03**	**0.002**	**-3.67**	**0.176**	**5.9E-51**
Reverse Flow Volume		**0.784**	**0.131**	**1.2E-8**	**0.764**	**0.097**	**2.9E-13**
Gage Height		0.230	0.494	0.643			
Precipitation		-2.94	3.19	0.357			
Season	Fall						
	Spring	-0.017	0.366	0.963			
	Summer	0.045	0.336	0.893			
Dissolved Oxygen		0.010	0.114	0.930			
Temperature		0.034	0.035	0.330			
pH		0.273	0.310	0.381			
Chlorophyll		-0.074	0.078	0.342			
ln(ρ)							
Intercept		**-2.27**	**0.162**	**4.5E-31**	**-2.03**	**0.156**	**5.5E-28**
Goodness of Fit							
Number of parameters (k)			11			3	
AIC			493			484	

For the final beta-binomial regression model, only Reverse Flow Volume (RevVol) is included. This model has the lowest AIC of the four tested models and is thus the strongest fit and preferred

model. The final beta-binomial regression model equation for the probability of a positive SC eDNA sample is logit(p_i)= -3.7 + 0.76*RevVol. The shape parameter is ln(ρ) = -2.03 or ρ = 0.131. Figure 4 shows a scatter plot of the explanatory variable, RevVol, against the fraction of positive SC eDNA samples for all 191 sampling events with the final beta-binomial regression model equation plotted over it. The model shows that given that the 30-day average of reverse flow volume is zero, the mean probability of a positive sample is 0.025. (Figure 2-3) The 95% confidence interval of the probability estimate is from 0 to 0.131. For a 30-day average of reverse flow volume of 1 million gallons, the mean probability of a positive sample is 0.052 (CI: 0, 0.212). At 4 million gallons, the mean probability of a positive sample is 0.35 (CI: 0.107, 0.641). Thus, a positive detection should be less likely (though arguably more alarming) during periods of low reverse flow.

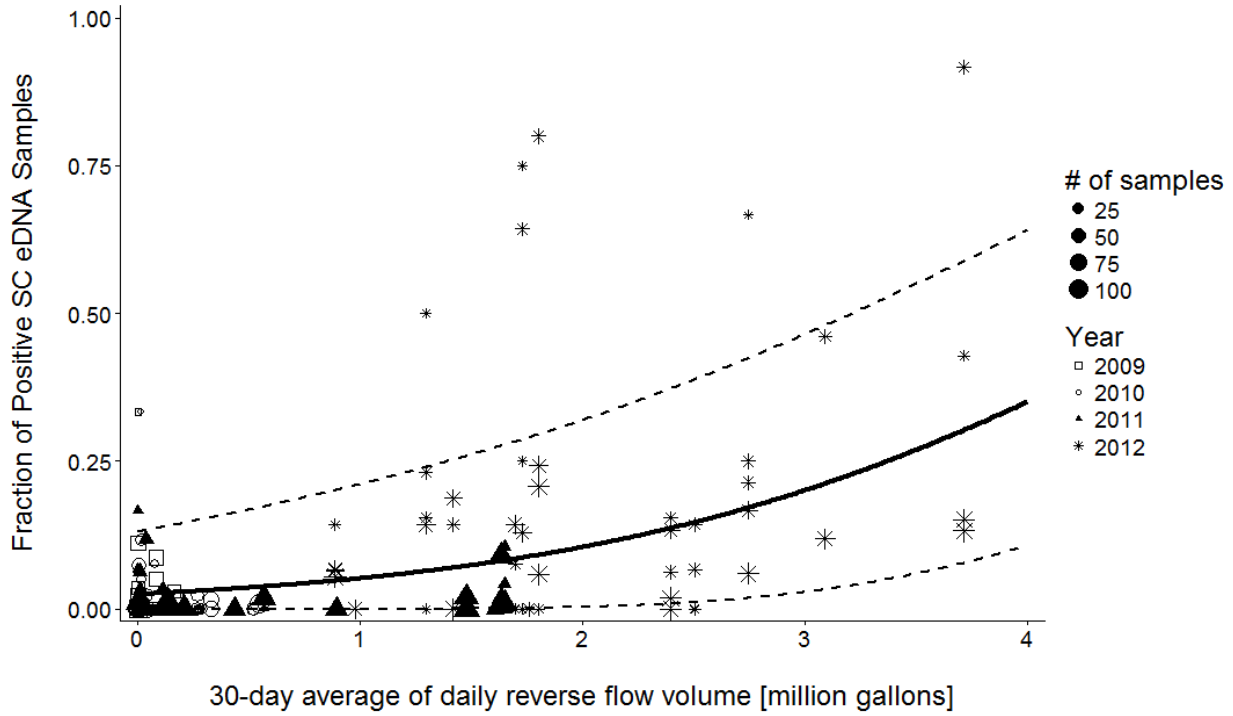

Figure 2-3. The 30-day moving average of reverse flow volume plotted against the fraction of positive Silver Carp eDNA samples for each sampling event from 2009 to 2012. The final beta-binomial regression model fitted to the data is also plotted along with its 95% confidence level (dotted lines)

2.4 Discussion

In this chapter, I found that reverse flow volume into the CAWS has a significantly positive relationship with silver carp eDNA detection from 2009-2012. One explanation for this relationship could be that flow reversals are carrying eDNA from below the barrier into the CAWS, increasing the likelihood of a positive eDNA detection. This possibility was considered by the experts of the Environmental DNA Calibration Study (ECALS), who concluded that flow reversals could bring eDNA into the CAWS, but would have an influence only in the half-mile above the electric barrier and not near Lake Michigan where the majority of eDNA samples are taken.[78] I agree with this conclusion, as I found that the periods of flow reversals are infrequent and would be unlikely to carry eDNA from below the electric fish barrier 25 miles to the eDNA sampling locations near Lake Michigan. A more likely explanation of this relationship is that the amount of

reverse flow volume is negatively correlated with the amount of dilution by Lake Michigan water of eDNA already present in the CAWS. This implies that there is a minor, sporadic source of AC eDNA in the CAWS that is normally diluted to below detection levels by typical flows.

The positive relationship of eDNA detection probability with reverse flow volume could provide an alternative explanation for the jump in positive eDNA samples in 2012 other than an increase in live carp presence, which was not corroborated by conventional monitoring and management efforts. In 2012, thousands of man-hours were devoted to electrofishing and netting in the CAWS, yielding no live SC or BH, despite the increase in eDNA detection.[20] One reason for this could be that the increased reverse flow volume in the system could have made the small amounts of eDNA from external sources (e.g., boat hulls, sewage effluents, etc.) easier to detect in the absence of live SC organisms. My expectation is that a statistical model like this could help differentiate between a "signal" caused by an actual invasion of Asian carp into the CAWS and "noise" due to changes in the stream characteristics of the system. Specifically, a large spike in positive samples during periods with no flow reversals in the system is much more worrisome than a large number of positive samples during periods with high reverse flow volume.

The beta-binomial regression model gave a superior fit compared to the binomial regression model for eDNA sampling results. This supports the assertion that eDNA sampling may not be an independent process and that correlation between the samples must be captured in any statistical model of eDNA sampling. The physical explanation of this correlation between the samples is that eDNA is most likely not distributed randomly in the environment, but in clumps, e.g. in cells and tissue fragments.[69] Most statistical models of the probability of eDNA detection and the covariates that influence it utilize a binomial or logistic regression model framework. [9,68] In this chapter, I

found that covariates, like temperature and pH, which were determined to have a statistically significant relationship with eDNA detection probability using a binomial regression model, were not significant in the beta-binomial regression model. I recommend that future statistical models for eDNA sampling consider the beta-binomial regression model when calculating the probability of a positive sample, as a way to account for this clumping, and improve our understanding of covariate influence on eDNA detection.

3 Probabilistic Framework for Environmental DNA Sampling Design

3.1 Background

One of the advantages of eDNA sampling compared to conventional survey techniques, such as electrofishing and netting, is its high detection sensitivity.[7,8,17,22,68,79–81] However, despite its high sensitivity, eDNA sampling can still suffer from false negative errors, which should be minimized when designing robust eDNA sampling protocols for aquatic invasive species detection and management.[82,83] Furthermore, as eDNA analytical techniques are rapidly developing, it is not uncommon for analytical protocols to be modified during the course of a long-term study. Changes in analytical protocols can alter their sensitivity, though this is not typically corrected for or even acknowledged.

As discussed in Chapter 1, eDNA sampling protocols involve multiple steps of analysis and false negative errors can arise at each of the steps.[82] For example, during water sample collection, false negatives can occur if target DNA is not collected in the sample for subsequent laboratory analysis.[84,85] During the capture and extraction processes, DNA can be lost or diluted.[27,28,30,31,58,86] During the polymerase chain reaction (PCR) DNA amplification process, a false negative can occur if the primers do not bind to the target DNA sequences in the sample.[30,87,88]

Despite these potential sources of error, many eDNA sampling protocol design choices are made on an ad-hoc basis (i.e., ease of use, availability of equipment) rather than from the results of pilot studies or other prior analyses.[24,89] I present a modeling framework to help eDNA practitioners understand how each step of a typical species-specific eDNA sampling protocol influences the protocol's overall sensitivity. This model is composed of a series of mathematical and probabilistic equations that characterize an eDNA sampling protocol to estimate the protocol's sensitivity given

a range of target eDNA concentrations. This model can then be applied to screen past or proposed changes to an eDNA sampling protocol for their influence on the protocol's sensitivity.

Unlike other survey methods, eDNA sampling does not see or capture live organisms, but rather it detects their DNA, whose origin is uncertain, e.g., excreta from live organisms, tissue from dead organisms, transfer from outside of the system.[10,24,25] Therefore, though a positive sample of the target eDNA may indicate the presence of a living target organism, it is not conclusive. To compare and evaluate the sensitivity of eDNA sampling protocols in the absence of information about target species presence, I model the sensitivity of eDNA sampling protocols as a function of the concentration of target DNA it can detect, namely, the minimum amount of target DNA that is detectable by the eDNA sampling protocol with probability 95%.[84]

The model developed in this chapter extends the model of eDNA sampling developed by Schultz and Lance and incorporates new understanding about eDNA and new developments in eDNA sampling technology.[84] I model water sample collection using a negative binomial distribution instead of a Poisson distribution, which enables the modeling of a clumped or random distribution of eDNA in the water column. I use logistic normal distributions to model the variation in the likelihood of successful PCR and quantitative PCR (qPCR) amplification.[69] I also include the eDNA capture step in this model, which has been shown to cause additional losses in DNA yield.[27] This model is then applied, retrospectively, to investigate the evolution of the US Fish and Wildlife Service (USFWS)'s BH and SC eDNA sampling protocols from 2013 to 2015 and evaluate how these changes may have influenced the protocol's overall detection sensitivity. The model is also used to explore what changes to the current eDNA sampling protocol would have the most potential impact on the protocol's detection sensitivity.

3.2 Methods and Data

3.2.1 Modeling framework

Figure 3-1 shows the five steps of a typical species-specific eDNA sampling protocol modeled in this paper: (1) collection of eDNA from the water sample; (2) capture, extraction and purification of the eDNA into a concentrated elution; (3) preparing the reaction wells for the PCR or qPCR assay; (4) amplification of the target DNA marker(s) and (5) the determination of a positive sample. The figure also details the key sampling protocol variables that parameterize each step of the model. Given the parameters of the eDNA sampling protocol and the concentration distribution of target eDNA in the water, the model simulates the probabilistic distribution of the number of copies of the target DNA present after each step of the eDNA sampling process and then calculates the probability that the DNA is detected in the PCR or qPCR replicates in order to confirm a positive sample.

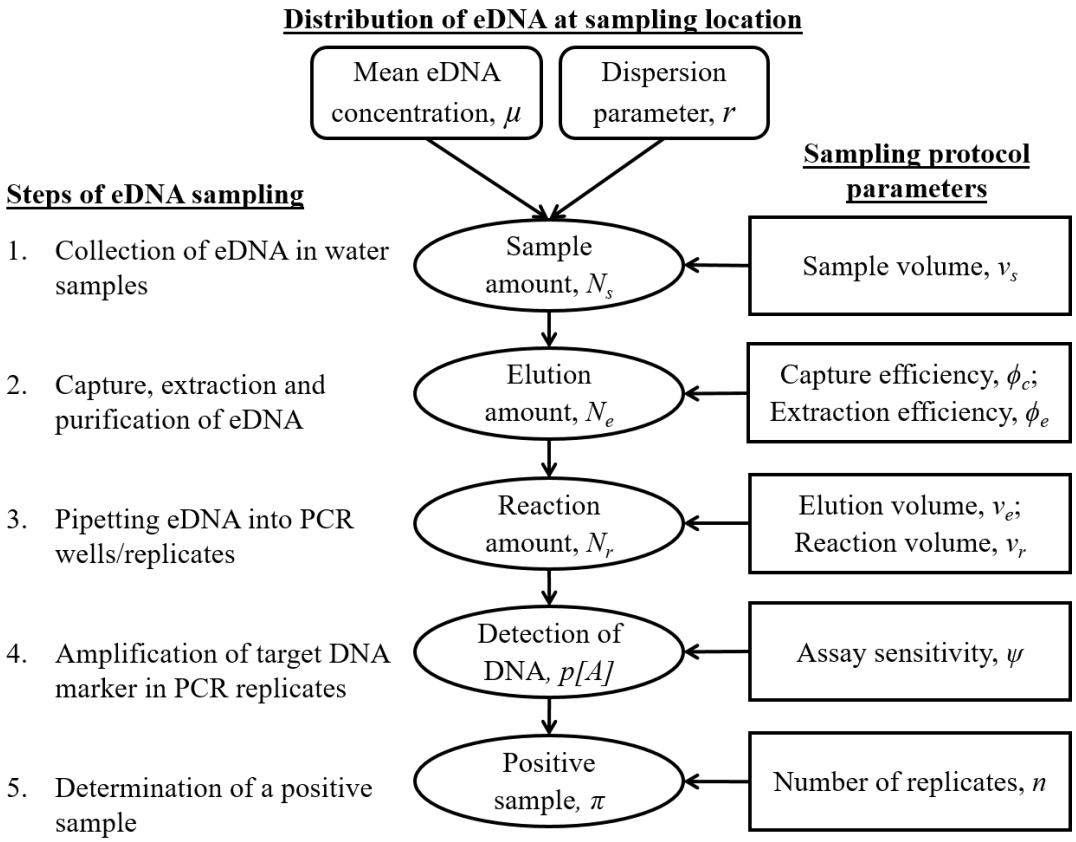

Figure 3-1. A schematic of the eDNA sampling process from sample collection to the determination of a positive sample. The circles in the center denote the probabilistic representation of the amount of DNA at each step and the likelihood of detecting the target DNA. The squares on the right detail the parameters of the sampling protocol that influence each step of the process.

3.2.1.1 Modeling the eDNA distribution in the collected water sample

For most aquatic macroorganisms, eDNA is likely to be present in the water within cells and tissues, rather than as individual DNA molecules.[38] I capture the possible distributions of eDNA from random to clumped using the two-parameter negative binomial distribution.[90] Assuming that the target eDNA is distributed in the water body with mean concentration (μ) and dispersion parameter (r), I model the number of DNA molecules collected in a sample (N_s) with sample volume (v_s) as a draw from a negative binomial distribution with mean μv_s.

$$N_s \sim NegBinom(\mu v_s, r) \quad (3\text{-}1)$$

The parameter, *r*, defines the statistical dispersion of the DNA in the water. As *r* approaches zero, the distribution of DNA becomes more clumped, while as *r* increases to infinity, the distribution converges to a Poisson distribution, indicating that the DNA is randomly distributed.

3.2.1.2 Modeling the number of eDNA copies after capture and extraction steps

The eDNA collected in the sample (N_s) is then captured and extracted to isolate the eDNA and remove potentially inhibitory substances before PCR analysis. Both steps can cause some losses in DNA yield. Different capture techniques (e.g., filtration or centrifugation) have different capture efficiencies (ϕ_c: the fraction of DNA captured from the sample) and different DNA extraction kits and protocols have different extraction efficiencies (ϕ_e: the fraction of DNA successfully extracted and isolated from the filtrate or centrifuged material).[27,28,30,31,58]

Assuming that each DNA copy in N_s has the same independent probability ϕ_c of being captured, I model the number of DNA copies captured from the sample (N_c) as a binomial distribution with probability ϕ_c and N_s potential copies to be captured.

$$N_c \sim Binomial(\phi_c, N_s) \quad (3\text{-}2)$$

Since N_s is a negative binomial distribution (see equation 3-2), N_c is a compound binomial-negative binomial distribution, which can be reduced to a negative binomial distribution with mean $\mu v_s \phi_c$ and dispersion parameter r.[91] Using the same assumption for the subsequent extraction step, I model the amount of DNA in the eluate (N_e), after accounting for collection, capture and extraction, as a draw from a negative binomial distribution with mean $\mu v_s \phi_c \phi_e$ and dispersion parameter r.

$$N_e \sim NegBinom(\mu v_s \phi_c \phi_e, r) \quad (3\text{-}3)$$

The sensitivity of the protocol at this stage, i.e. the probability that DNA is successfully collected, captured and extracted (θ), is one minus the probability that no DNA is present in the eluate. This equation is used later to calculate the overall sensitivity of the protocol.

$$\theta = 1 - P[N_e = 0] = 1 - \left(\frac{r}{r+\mu v_s \phi_c \phi_e}\right)^r \quad (3\text{-}4)$$

3.2.1.3 Modeling the number of eDNA copies per PCR well

The DNA in the eluate (N_e) is diluted by an elution volume (v_e) and analyzed using polymerase chain reaction (PCR) to detect for the presence of the species-specific DNA sequence or marker.[17,34] This is done by placing aliquots of the eluate into multiple replicates or wells of the PCR-plates. Assuming that the eluate is well-mixed, the amount of DNA in each replicate (N_r) can be modeled using a Poisson distribution with mean $N_e f$, where the variable $f = v_r/v_e$ is the fraction of the elution volume (v_e) used per reaction well (v_r).

$$N_r \sim Poisson(N_e f) \quad (3\text{-}5)$$

In this model, I assume that there is perfect specificity, i.e. there is zero chance of a positive PCR replicate if there is no target DNA present in the eluate. Therefore, I only need to model the probability of a positive amplification, assuming that target DNA is present in the eluate. Assuming that $N_e > 0$, the negative binomial distribution for N_e (Equation 3-3) can be approximated as a gamma distribution, where the shape term is r and the scale term is $r/(\mu v_s \phi_c \phi_e)$. Therefore, given that $N_e > 0$, the amount of DNA in each replicate, N_r, is now a Poisson distribution with a gamma-distributed mean $N_e f$, which reduces to a negative binomial distribution with mean $\mu v_s \phi_c \phi_e f$ and dispersion parameter r.[92]

$$N_r \sim NegBinom(\mu v_s \phi_c \phi_e f, r) \quad (3\text{-}6)$$

3.2.1.4 Modeling the probability of amplification of eDNA in PCR replicates

The probability of successful PCR amplification in each of the wells depends on the number of copies of the target marker present in the replicate for the reaction to successfully occur. Assuming that each target DNA copy in the replicate has an equal chance of amplification (ψ), I model the number of copies that amplify (N_a) as a binomial distribution with probability (ψ) and N_r possible copies for amplification.

$$N_a \sim Binomial(\psi, N_r) \quad (3\text{-}7)$$

Since N_r is a negative binomial distribution (Equation 3-6), N_a can be modeled as a compound binomial-negative binomial distribution, which again is reduced to a negative binomial distribution with mean $\mu v_s \phi_c \phi_e f \psi$ and dispersion parameter r.

$$N_a \sim NegBinom(\mu v_s \phi_c \phi_e f \psi, r) \quad (3\text{-}8)$$

Then, I model the probability of successful amplification in the replicate ($p[A]$) as the probability that at least one copy amplifies or one minus the probability that zero copies amplify.

$$p[A] = 1 - p[N_a = 0] = 1 - \left(\frac{r}{r + \mu v_s \phi_c \phi_e f \psi}\right)^r \quad (3\text{-}9)$$

3.2.1.5 Modeling the determination of a positive sample

Given the probability of amplification in each replicate, the probability of a positive sample depends on the number of replicates (n) that are tested per sample and how many positive replicates are needed. Typically, a sample is determined to be positive, if at least one of the replicates successfully amplifies. For this case, the probability of a positive sample (π) would be one minus the probability that none of the replicates successfully amplify.

$$\pi \;=\; 1-(1-p[A])^{n} \qquad (3\text{-}10)$$

To calculate the overall detection sensitivity of the protocol (Ω), I combine the probability that the protocol successfully collects, captures and extracts the target DNA present in the water body (θ) and the probability that the target DNA is successfully amplified in at least one replicate given that the DNA is present in the eluate (π).

$$\Omega \;=\; \theta * \pi = \left[1-\left(\frac{r}{r+\mu v_s \phi_c \phi_e}\right)^{r}\right] * \left[1-\left(\frac{r}{r+\mu v_s \phi_c \phi_e f \psi}\right)^{rn}\right] \qquad (3\text{-}11)$$

In summary, the model evaluates the detection sensitivity (Ω) conditioned on the parameters of the sampling protocol (v_s, ϕ_c, ϕ_e, v_e, v_r, ψ, n) and the distribution of eDNA at the sampling location (μ, r).

3.2.2 Evolution of the eDNA Sampling Protocol for Bighead and Silver Carp

I use this model to evaluate how changes to the eDNA sampling protocol for BH and SC detection from 2013 to 2015 have impacted the protocol's sensitivity. The original protocol (used from 2009 to 2012), described in Jerde et al., was initially developed and used by the University of Notre Dame (UND) in 2009 and later applied by the US Army Corps of Engineers (USACE) for BH and SC eDNA sampling from 2010 to 2012.[17] In 2013, the USFWS assumed responsibility for BH and SC eDNA monitoring and began modifying the protocol each year. The procedures and modifications of the BH and SC eDNA protocols for each year are found in Table 3-1.[93]

Table 3-1. The methods used in the Bighead and Silver Carp eDNA monitoring protocol as specified in the Quality Assurance Project Plan (QAPP) for the years 2009-2016. The major changes to the protocol each year are in bold text.

Sampling protocol parameters	Year (Lab)			
	2009-2012 (UND/ USACE)	2013 (USFWS)	2014 (USWFS)	2015-2016 (USFWS)
Sample volume	2-L bottle	2-L bottle	2-L bottle	**Five 50-mL bottle**
Capture method	Filtration (1.5 µm glass filter)	Filtration (1.5 µm glass filter)	Filtration (1.5 µm glass filter)	**Centrifugation**
Extraction method	MoBio® PowerWater	**Qiagen DNeasy Blood & Tissue**	Qiagen DNeasy Blood & Tissue	**IBI Scientific gMAX mini**
Elution volume	100 µL	**If 8 or fewer filters, then 200 µL for each filter Else 100 µL each**	If 8 or fewer filters, then 200 µL for each filter Else 100 µL each	200 µL
Reaction volume	1 µL	1 µL	**3 µL**	3 µL
Type of assay and markers	PCR SC-PCR BH-PCR	PCR SC-PCR BH-PCR	**qPCR** **ACTM1/ACTM3 SCTM4/SCTM5 BHTM1/BHTM2**	qPCR ACTM1/ACTM3 SCTM4/SCTM5 BHTM1/BHTM2
Number of replicates	8	8	8	8
Determining a positive sample	The species-specific marker amplifies and is successfully sequenced in at least one replicate	The species-specific marker amplifies and is successfully sequenced in at least one replicate.	**Two AC markers and at least one of the species-specific markers amplify in at least one replicate (does not have to be same one)**	Two AC markers and at least one of the species- specific markers amplify in at least one replicate (does not have to be same one)

In the original protocol, UND and USACE took samples using a two-liter bottle, and filtered the water through a 1.5 µm glass filter. The DNA was extracted using the MoBio® PowerWater DNA

Isolation Kit (PW) and eluted into 100 µL. Then, 1 µL of the elution was placed into each of 8 replicates test wells. A PCR assay with species-specific markers and primers was used to analyze the replicates for the presence of BH and/or SC DNA.[17] If there was successful amplification, the amplified DNA was sequenced to ensure that the correct species-specific marker had been found.[94]

In 2013, the USFWS modified both the extraction method and the elution volume. The change in extraction method was motivated by laboratory experiments showing that the Qiagen DNeasy Blood and Tissue Kit (DN) yielded more DNA than PW.[30] I estimated the extraction efficiency of the DN kit and its analogous IBI Scientific gMAX mini kit (used in 2015-2016) from experimental data (see Appendix B.2 for more details) using a logistic normal distribution with a mean efficiency of 33% and standard deviation of 3%. For comparison, previous studies of protocol sensitivity represented the extraction efficiency of the previously-used PW kit using a triangular distribution with a median of 15% and a range of 0 to 30%.[84] There was also an increase in the elution volume (v_e), which was initially 100 µL for every sample. In 2013 and 2014, the elution volume varied depending on the number of filters used to process the sample. If 8 or fewer filters were used, then each of the filters was processed and eluted with 200 µL before being pooled, and if 9 or more filters were used, then each of the filters was eluted with 100 µL before being pooled.[93] The elution volume was modeled as a discrete probability distribution using this rule and actual records of the number of filters used per sample by USFWS (see Figure B-1). The distribution has a mean of 688 µL with a standard deviation of 340 µL.

For the initial PCR assay and PCR markers (SC-PCR, BH-PCR) used from 2009 to 2013, a sample was determined to be positive if the species-specific marker amplified and was successfully sequenced in at least one replicate. The likelihood of amplifying or sequencing a single copy of

each PCR marker (ψ) was modeled as a logistic normal distribution and fit using experimental data (see Table B-3 and Table B-4). For the BH-PCR and SC-PCR markers, 100% mean amplification probability is achieved at around 10 target DNA copies in the replicate (Figure B-2). To achieve 100% sequencing probability, ~100 copies are needed in the replicate for the BH-PCR marker and ~10 copies are needed for the SC-PCR marker.

To include the additional sequencing step in the PCR assay for BH and SC eDNA analysis, I adapted Equation 3-10 in the model to be the probability that at least one replicate amplifies the target species-specific marker ($p[A]$) and is successfully sequenced ($p[S]$), or one minus the probability that none of the replicates do both. Thus, the probability of a positive sample for the PCR assay given that target DNA is present in the eluate (π) is modeled as shown below in Equation 3-12.

$$\pi = 1 - (1 - p[A_{SCPCR}] * p[S_{SCPCR}])^n \quad (3\text{-}12)$$

In 2014, a qPCR assay was introduced along with a new suite of species-specific markers (primers and probes).[87,95] Because of the genetic similarity between BH and SC, two markers were designed to detect both species (ACTM1, ACTM3) and four additional markers were designed to be species-specific to BH (BHTM1, BHTM2) and SC (SCTM4, SCTM5). To model the likelihood of amplifying or sequencing a single copy (ψ) for the new markers, I fit new experimental data testing these new markers to logistic normal distributions (Table B-5).[87] For the six qPCR markers, 100% mean amplification probability is achieved between 3-10 copies per replicate (Figure B-4).

Along with the change to the qPCR assay, the reaction volume (v_r) increased from 1 to 3 µL and the criterion for determining a positive sample changed. Subsequently, both Asian carp-specific markers and at least one of the two species-specific markers had to amplify in at least one of the

replicates. These did not have to occur in the same replicate. Equation 3-13 shows the probability of a positive sample for the qPCR assay given that target DNA is present in the eluate (π), modifying Equation 3-10 of the model.

$$\pi = \{1 - (1 - \ p[A_{ACTM1}])^n\} * \{1 - (1 - \ p[A_{ACTM3}])^n\} * \{1 - [(1 - \ p[A_{SCTM4}]) * (1 - p[A_{SCTM5}])]^n\} \quad (3\text{-}13)$$

In 2015, USFWS changed the DNA capture method and the sample volume (v_s) from filtering a 2-L water sample using a 1.5 µm filter to centrifuging and combining the pelleted material from five 50-mL tubes (250-mL total). Estimated capture efficiencies (ϕ_c) for the initial filtration method and current centrifugation method were modeled as logistic normal distributions using experimental data (see Appendix B.2). The estimated capture efficiency for filtration is 3.7% with a standard deviation of 1.1% and the estimated capture efficiency for centrifugation is 6.4% with a standard deviation of 1.8%. Also, the elution volume (v_e) was decreased and standardized to 200 µL per sample.

3.2.3 Analyzing the influence of changing eDNA sampling protocols over time on BH and SC detection sensitivity

Given each year's sampling protocol parameters, I used the model to simulate each step of the eDNA sampling process to calculate the protocol's overall detection sensitivity ($\Omega = \theta * \pi$) over a range of mean eDNA concentrations (μ) for both species. Because of uncertainty in the *r* parameter, I varied the *r* parameter parametrically to simulate both a random and a clumped distribution of eDNA in the water column. The random distribution was simulated by specifying $r = 100$, which approximates a Poisson distribution. The clumped distribution was simulated by specifying $r = 0.3$, which is the value of *r* estimated by Furlan et al. for eDNA from the *Oriental*

weatherloach. I then used the model results to calculate and compare the minimum mean concentration that is detectable with probability 95% by each of the eDNA sampling protocols. A lower value of this number reflects a greater ability to detect lower concentrations and thus a higher overall detection sensitivity of the protocol. The model is also used to explore how future changes to different parameters of the sampling protocol, such as the sample volume, elution volume, capture and extraction efficiencies, the number of replicates, and reaction volume, would affect the detection sensitivity of USFWS' current BH and SC protocol.

I used Monte Carlo simulations to perform the sensitivity calculations with 10000 sample iterations. This was done to model the uncertainty in the overall sensitivity calculations resulting from uncertainties in the values of model parameters. All of the models were run using R 3.2.2.[77]

3.3 Results

3.3.1 Comparing the detection sensitivity of each year's eDNA sampling protocols

I compare the sensitivity of each protocol change to the original protocol in Figure 3-2. It shows the overall probability of a positive sample (Ω) for each year's sampling protocol, as a function of the mean target eDNA concentration (μ) for both species, when the dispersion parameter r is set equal to 100 to simulate a random distribution of eDNA in the water column. The curves for the original 2009-2012 protocol are shown in gray. The curves show that the sensitivity of the sampling protocols fluctuated over the years. In 2013, there is a decrease in the sensitivity of the protocol, as seen by the shifting of the curves to the right. In 2014, the curves shift to the left showing a significant improvement in sensitivity compared to the original protocol. This shift is larger for BH compared to SC. Then, in 2015, the curves shift slightly back to the right for both species, indicating a loss of sensitivity. For BH, the sensitivity of the protocol in 2015 is better

than the original protocol, while for SC, the sensitivity is comparable. The protocol with the highest sensitivity was that used in 2014.

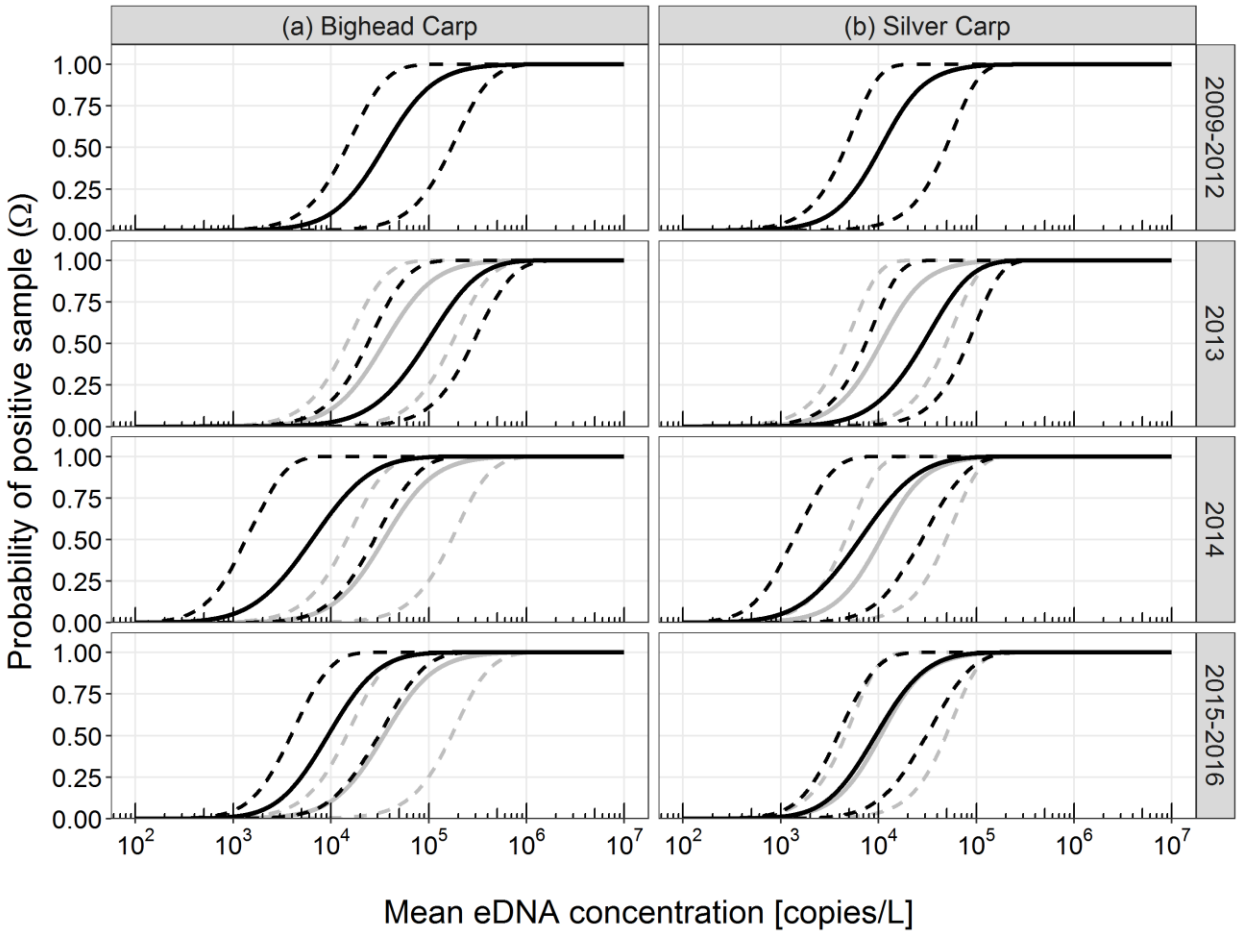

Figure 3-2. The simulated per-sample detection sensitivity for (a) Bighead Carp and (b) Silver Carp eDNA as a function of mean eDNA concentration for each year's sampling protocol. It is assumed that the eDNA was randomly distributed (r = 100). The sensitivity of the 2009-2012 protocol is shown in gray in the other plots for comparison. The solid line is the mean value and the dashed lines represent the 95% credible intervals of the estimated detection sensitivity.

I also quantify the sensitivity of each year's eDNA sampling protocols as the minimum mean eDNA concentration in the water body detectable with probability 95%. Table 3-2 shows this minimum concentration after each specific change made to the protocols in each year. From 2012 to 2013, the overall loss of sensitivity (i.e., an increase in the minimum mean concentration detectable with probability 95%) from 47,900 to 110,000 copies/L for SC and 182,000 to 424,000

copies/L for BH was primarily due to the larger elution volume, which caused additional dilution of the eDNA. This overwhelmed the improvement in efficiency from the change in extraction kits. From 2013 to 2014, the qPCR assay and new markers improved the sensitivity for detecting BH to approach the sensitivity for detecting SC at 114,000 copies/L. However, there was very little improvement in the sensitivity for detecting SC with the new markers. The increase in reaction volume in 2014 also improved the sensitivity of both BH and SC to 38,000 copies/L. From 2014 to 2015, the gains in sensitivity due to the reduction in elution volume (38,000 copies/L to 9,300 copies/L) and the switch to centrifugation (9,300 copies/L to 5,400 copies/L) was not enough to compensate for the loss in sensitivity caused by the smaller sample volume (5,400 copies/L to 43,500 copies/L). Comparing the detection sensitivity of each year's eDNA sampling protocols assuming $r = 0.3$

Table 3-2. The minimum mean eDNA concentration in the water body detectable with probability 95% after each change in the sampling protocol for both carp species, assuming that the eDNA was randomly distributed and unclumped (r = 100). The point value is the mean value and the values in parentheses represent the 95% credible interval.

Year	Change in sampling protocol	Minimum mean eDNA concentration detectable with probability 95% [copies/L]	
		Bighead Carp	Silver Carp
2009-2012	Initial CAWS field sampling protocol	182,000 (44,000 – 527,000)	47,900 (11,000 –120,000)
2013	PowerWater® Kit replaced with DNEasy Kit	51,800 (25,000 – 89,000)	12,900 (6,300 – 21,000)
	Elution volume increased from 100 µL to ~700 µL	424,000 (74,000 – 875,000)	110,000 (18,000 – 198,000)
2014	cPCR replaced with qPCR and introduced new markers	114,000 (12,000 – 296,000)	114,000 (12,000 – 296,000)
	PCR reaction volume increased from 1 to 3 µL	38,000 (4,000– 99,000)	38,000 (4,000– 99,000)
2015-2016	Elution volume reduced from ~750 µL to 200 µL	9,300 (2,500– 20,000)	9,300 (2,500– 20,000)
	Filtration replaced with centrifugation	5,400 (1,500 – 14,000)	5,400 (1,640 – 14,000)
	Sample volume reduced from 2 L to 0.25 L	43,500 (12,000 – 112,000)	43,500 (11,500 – 112,000)

When I assume the DNA is clumped (r = 0.3), the fluctuation in the sensitivity of the protocol over time stays relatively the same (Figure 3-3). However, in comparison to the estimates of detection sensitivity when assuming the DNA is randomly distributed (r = 100), the curves are shifted and stretched to the right, meaning that larger mean concentrations of target DNA are needed to achieve the same detection probability when the DNA is assumed to be clumped.[69]

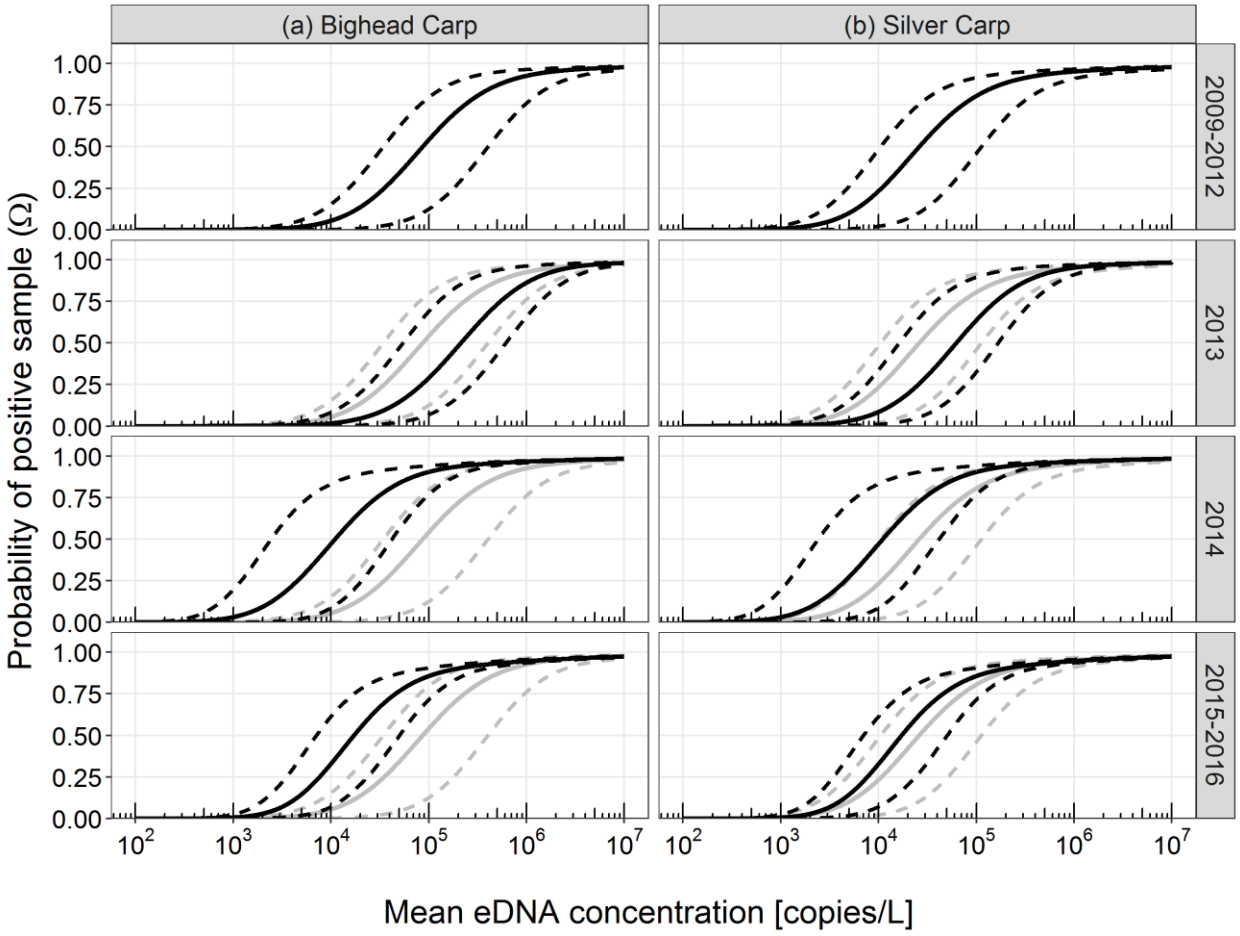

Figure 3-3. The simulated per-sample detection sensitivity for (a) Bighead Carp and (b) Silver Carp eDNA as a function of mean eDNA concentration for each year's sampling protocol. It is assumed that the eDNA was clumped (r = 0.3). The solid line is the mean value and the dashed lines represent the 95% credible intervals of the estimated detection sensitivity.

This can also be seen in the minimum mean eDNA concentration in the water body detectable with probability 95% after each specific change made to the protocols, shown in Table 3-3. The mean eDNA concentrations needed for detection are ~10x larger when assuming a clumped distribution rather than a random distribution of eDNA. There are a few more noticeable changes. The loss in sensitivity from 2014 to the current 2015-2016 protocol is much more severe, in fact, causing the sensitivity for SC detection to drop lower than the original 2009-2012 protocol.

Table 3-3. The minimum mean eDNA concentration in the water body detectable with probability 95% after each change in the sampling protocol for both carp species, assuming that the eDNA was clumped (r = 0.3). The point value is the mean value and the values in parentheses represent the 95% credible interval.

Year	Change in sampling protocol	Minimum mean eDNA concentration detectable with probability 95% [copies/L]	
		Bighead Carp	Silver Carp
2009-2012	Initial CAWS field sampling protocol	1,770,000 (531,000 – 6,200,000)	935,000 (316,000 – 3,540,000)
2013	PowerWater® Kit replaced with DNEasy Kit	574,000 (309,000 – 986,000)	327,000 (186,000 – 566,000)
	Elution volume increased from 100 µL to ~700 µL	2,650,000 (676,000 – 5,700,000)	948,000 (305,000 – 1,910,000)
2014	cPCR replaced with qPCR and introduced new markers	565,000 (191,000 –1,220,000)	565,000 (191,000 –1,220,000)
	PCR reaction volume increased from 1 to 3 µL	341,000 (171,000 – 650,000)	341,000 (171,000 – 650,000)
2015-2016	Elution volume reduced from ~750 µL to 200 µL	281,000 (161,000 – 486,000)	281,000 (161,000 – 486,000)
	Filtration replaced with centrifugation	163,000 (89,000 – 288,000)	163,000 (89,000 – 288,000)
	Sample volume reduced from 2 L to 0.25 L	1,310,000 (714,000 – 2,310,000)	1,310,000 (714,000 – 2,310,000)

3.3.2 Analysis of future changes to sampling protocol

The model is also used to analyze how changes to the current BH and SC eDNA sampling protocol could improve the overall sensitivity for future eDNA sampling protocols. Figure 3-4 describes how modifications to six of the eDNA sampling protocol parameters (sample volume, capture efficiency, extraction efficiency, elution volume, reaction volume, and the number of replicates per sample) would change the per-sample detection sensitivity of the SC eDNA sampling protocol, assuming that the DNA is randomly dispersed (r = 100). The figure plots the mean SC eDNA concentration detectable with 95% probability using the current eDNA sampling protocol for a range of different values of each parameter holding all other values constant. The circles on each

plot mark the sensitivity of the current protocol for SC eDNA sampling at 41,700 copies/L, as shown in Table 3-2.

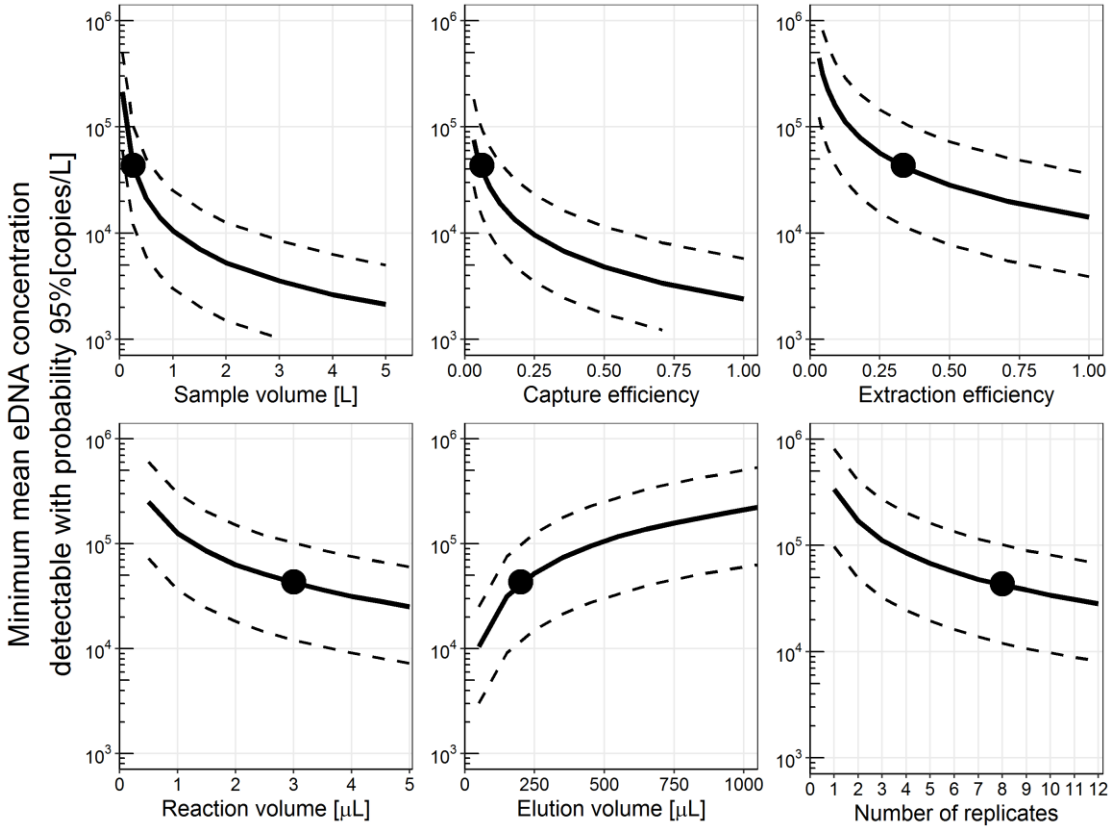

Figure 3-4. The minimum mean Silver Carp eDNA concentration in a water body that is detectable with per-sample detection sensitivity of 95% over a range of (a) sample volume, (b) capture efficiency, (c) extraction efficiency, (d) elution volume (e) reaction volume, and (f) number of replicates. The current sensitivity is noted with a circle at 41,700 copies/L for the 2015-2016 sampling protocol, which is a sample size of 0.25 liters, capture efficiency of 6.4%, extraction efficiency of 33%, elution volume of 200 µL, a reaction volume of 3 µL and 8 replicates. For each graph, only the parameter of interest was changed while the other parameters were kept constant. It is assumed that the eDNA was randomly distributed and unclumped (r = 100). The solid line denotes the mean value while the dashed lines represent the 95% credible interval.

Larger sample volumes, higher capture and extraction efficiencies, larger reaction volumes, smaller elution volumes and higher number of replicates will improve overall sensitivity. The asymptotes of the plots capture the maximum potential improvement in sensitivity for each of the parameters for the current protocol. For extraction efficiency, reaction volume and the number of replicates, the maximum achievable sensitivity levels out at around 20,000 copies/L. By

comparison, for sample volume and capture efficiency, the curve levels out at around 2,000 copies/L. When comparing percentage changes to each protocol parameter, the model shows that for most of the parameter values, the same percentage change will have the same impact on the overall sensitivity, e.g., a 10% increase in sample volume has the same mathematical effect as a 10% increase in capture efficiency (Figure B-6).

However, if the eDNA is assumed to be clumped (r = 0.3), the parameters that are upstream or earlier in the process (e.g., sample volume) have much greater effect than those parameters that are downstream or later in the process (e.g., reaction volume and elution volume), which have little effect (Figure 3-5). This is also seen when analyzing percentage changes to each of the parameters, which shows that sample volume, capture efficiency and extraction efficiency have far greater influence than reaction volume, elution volume or number of replicates on the sensitivity of the protocol, when the eDNA is assumed to be clumped (Figure B-7).

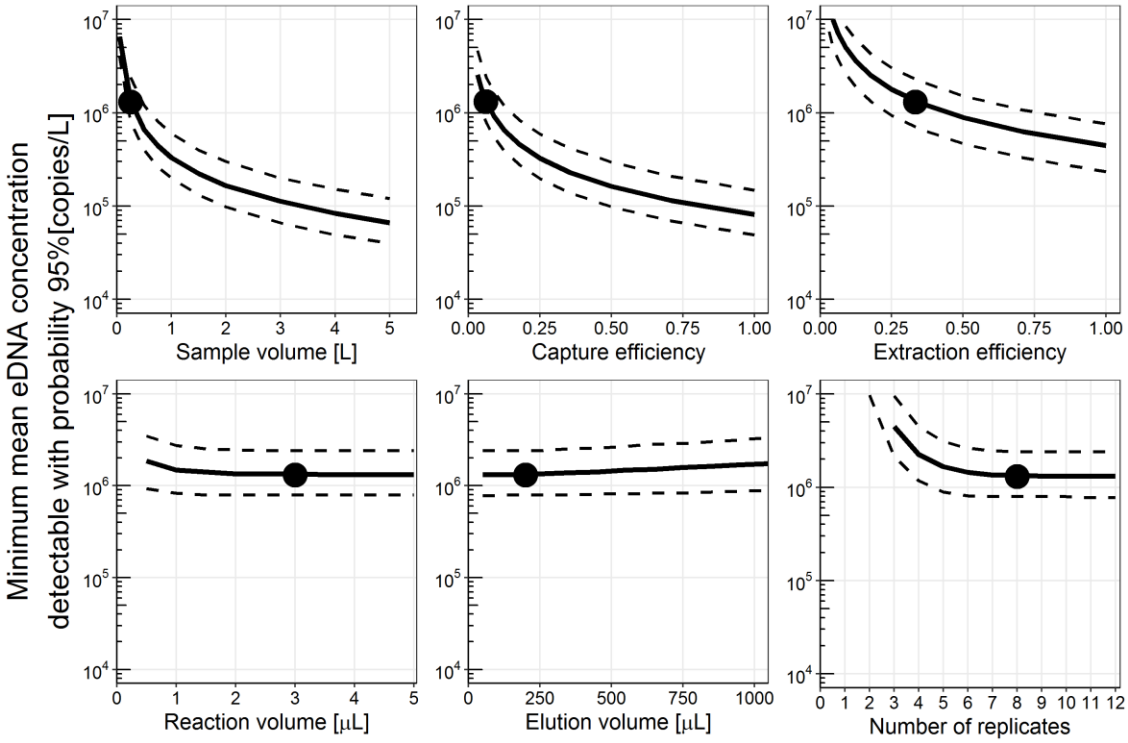

Figure 3-5. The minimum mean Silver Carp eDNA concentration in a water body that is detectable with per-sample detection sensitivity of 95% over a range of (a) sample volume, (b) capture efficiency, (c) extraction efficiency, (d) elution volume (e) reaction volume, and (f) number of replicates. The current sensitivity is noted with a circle at 1,320,000 copies/L for the 2015-2016 sampling protocol, which is a sample size of 0.25 liters, capture efficiency of 6.4%, extraction efficiency of 33%, elution volume of 200 µL, a reaction volume of 3 µL and 8 replicates. For each graph, only the parameter of interest was changed while the other parameters were kept constant. It is assumed that the eDNA was clumped (r = 0.3). The solid line denotes the mean value while the dashed lines represent the 95% credible interval.

3.4 Discussion

This chapter describes a model that can be utilized to evaluate the sensitivity of an eDNA sampling protocol and to estimate the gains or losses in sensitivity from past and future modifications to the protocol. Applied to the example of invasive BH and SC eDNA sampling by USFWS near the Great Lakes, the model quantified how the changes made to the eDNA sampling protocols from 2013 to 2015 have both improved and impaired the protocol's overall sensitivity. Assuming that the eDNA is randomly distributed (r = 100), the changes to the protocol greatly improved the sensitivity for BH eDNA detection and slightly improved the sensitivity for SC eDNA detection relative to the original protocol. If the eDNA is clumped (r = 0.3), the sensitivity for BH eDNA

detection has improved, but the sensitivity for SC eDNA detection has worsened. The most sensitive protocol in either scenario was the one in effect in 2014.

The overall loss in sensitivity for the current sampling protocol in comparison to the most sensitive 2014 protocol was primarily caused by the decrease in sample volume from 2 L to 0.25 L. This model shows that the protocol parameters upstream of the eDNA sampling process will influence and limit the amount of target DNA that is available for the following PCR analysis step, especially if the eDNA is clumped. Most research in improving the sensitivity of eDNA sampling protocols has focused on the development of more sensitive PCR-based detection platforms, like droplet-digital PCR (ddPCR) and laser transmission spectroscopy (LTS), which are potentially able to detect down to one copy of the target DNA.[81,96–98] This model shows that these new PCR-based methods will be affected by the sampling protocol design choices earlier in the process.[58] eDNA practitioners interested in maximizing the sensitivity of their eDNA survey should look at the entirety of their eDNA sampling protocol.

Also, the model shows that the eDNA distribution impacts which parameters are most influential on the sensitivity of the protocol. In the model equations, the upstream protocol parameters (sample volume, capture efficiency, extraction efficiency) have influence on the probability of collecting, capturing and extracting eDNA (Equation 3-4) and the probability of amplification (Equation 3-9), while the parameters later in the process (reaction volume, elution volume, number of replicates) only influence the probability of amplification. However, when the eDNA is randomly distributed ($r = 100$), the probability of collecting eDNA is very likely to be one at high eDNA concentrations, regardless of the protocol parameter values. Thus, if eDNA is randomly distributed, the overall sensitivity is influenced solely by the probability of amplification, where

all of the protocol parameters have similar influence. In comparison, if the eDNA is clumped, θ is typically less than 1, even at high mean concentrations. For example, given the current protocol parameter values, when μ = 100,000 copies/L, the value of θ is 1 when r = 100, but 0.89 when r = 0.3 Therefore, if eDNA is clumped, changes to the upstream protocol parameters have significantly more influence on the overall sensitivity of the protocol than changes to the downstream protocol parameters.

This model also shows the importance of including the capture and extraction steps when quantifying and evaluating the sensitivity of eDNA surveys.[27,31] In Schultz and Lance's work evaluating the sensitivity of the original 2009-2012 BH and SC protocol, they estimate a minimum mean concentration of ~4,200 copies/L for BH and ~1,100 copies/L for SC to achieve a 95% per-sample detection probability.[84] In comparison, my updated model estimates the minimum mean concentration at ~176,000 copies/L for BH and 46,200 copies/L for SC. This large decrease in the sensitivity is caused by including the losses from the filtration step (capture efficiency = ~3.7%), which was not included in Schultz and Lance's model. Another model in the literature (Furlan et al.'s for evaluating the sensitivity of eDNA sampling for *Oriental weatherloach*) does not include the losses from either the capture or extraction processes and thus estimates a minimum mean concentration of just ~60 copies/L to achieve a 95% per-sample detection probability.[69] My model shows that the capture and extraction steps cause significant losses in eDNA yield, and thus must be included when evaluating and comparing the sensitivity of different eDNA sampling protocols.[27,28,31]

4 Estimation of surface water eDNA concentrations based on statistical models of eDNA sampling protocols

4.1 Background

Chapter 3 answered the question of how the sampling and analytical protocol for eDNA sampling affects the ability to detect the target eDNA when present. However, this does not address the question of whether positive BH or SC eDNA samples indicate the presence of Asian carp in the CAWS and other sampling locations near the Great Lakes. Unlike other surveillance tools, eDNA sampling does not capture or spot live organisms. Therefore, when species presence is unknown or unlikely, e.g. in most invasive species monitoring situations, it is difficult to interpret whether a positive detection of the target species' DNA is due to live organismal presence or due to alternate vectors of target eDNA, e.g. boat hulls carrying fish slime, stream flow, etc., that are detectable in the absence of the species.[99]

Currently, the most commonly-used statistical model for interpreting species presence from eDNA sampling data is the site occupancy model, which estimates the probability that the target species is present at the monitoring site (occupancy probability) based on the sensitivity of the eDNA sampling method and the eDNA sampling results.[67,100–102] The most common version of this model is the three-stage site occupancy model for eDNA sampling, which estimates the probability of occupancy of the target species by accounting for both the probability the sample collects the eDNA when present at the sampling location (availability probability) and the probability it successfully detects the eDNA when collected (detection probability).

However, this modeling framework implicitly assumes that if the target species' DNA is present and detected at the sampling location, then the target species must then be present. I propose that

this is an incorrect assumption when interpreting eDNA sampling results, because of the aforementioned uncertainty about the origin of eDNA. In this chapter, I use the statistical model (from Chapter 3) to interpret eDNA sampling results not as a function of target species presence but as a function of the concentration of target species' eDNA and the ability of the eDNA sampling protocol to detect that eDNA concentration.[84,103] This model estimates the concentration distribution of the target species' eDNA in the sampled water body from the eDNA sampling results, rather than species' occupancy.[103]

I use this model to analyze the U.S. Fish and Wildlife Service (USFWS)'s AC eDNA sampling in the Illinois River below the electric barrier and the CAWS upstream of the electric barrier in 2014. The model is used to estimate the AC eDNA concentration distributions in the Illinois River downstream of the electric barrier in the La Grange, Marseilles and Brandon Road pools and in the CAWS upstream of the electric barrier. BH and SC are known to be present in the La Grange and Marseilles pool but absent in the Brandon Road pool. Then, I estimate whether BH and SC are present or absent in the CAWS by calculating the similarity of the estimated eDNA concentration distributions in the CAWS to the eDNA concentration distributions in the pools below the barrier with known AC presence (Marseilles) and absence (Brandon Road).

4.2 Methods

4.2.1 eDNA Study Site and Sampling Results

In Spring of 2014, US Fish and Wildlife Service (USFWS) collected water samples at three sampling locations below the barrier in the Illinois Waterway to test for the presence of AC eDNA: (1) La Grange pool (141 to 218 miles downstream from the electric barrier), (2) Marseilles pool (29 to 53 miles downstream from the electric barrier), where adult Asian carp have been caught

and some spawning has been observed; and (3) Brandon pool (7 to 12 miles downstream from the electric barrier) where there have been no live captures of either BH or SC.[21]

At each location, USFWS collected 25 250-mL samples, which were processed using centrifugation, and 25 2-L samples, which were processed using filtration. The DNA captured from each sample by centrifugation or filtration was then extracted using the Qiagen DNEasy Blood & Tissue Kit and analyzed using quantitative PCR (qPCR) to quantify the total number of copies of the target eDNA in the sample. USFWS tested for the presence and quantity of six Asian carp-specific DNA sequences (markers) in each sample. Two of the markers are specific to both BH and SC (AC-TM1 and AC-TM3), two of the markers are specific to just BH (BH-TM1 and BH-TM2) and two markers are specific to just SC (SC-TM4 and SC-TM5. Details about the six species-specific markers and associated primers used for AC eDNA detection using qPCR can be found in Farrington et al.[104] The average copy number of each marker per sample collected from these three pools are found in Table 4-1.

Table 4-1. The mean number of copies of each of the target markers in the eight qPCR replicates tested per sample (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) from each combination of sampling location (Marseilles pool, Brandon Road pool) and capture method (filtration or centrifugation). 25 samples were analyzed for each combination.

Sampling location	Capture method	Mean number of copies of the target marker					
		AC-TM1	AC-TM3	BH-TM1	BH-TM2	SC-TM4	SC-TM5
La Grange pool	Filtration	180	71	17	35	41	224
	Centrifugation	600	470	20	28	330	590
Marseilles pool	Filtration	50	40	18	13	19	30
	Centrifugation	23	17	2.3	3.0	17	25
Brandon Road pool	Filtration	0.33	0.23	0	0	0	0
	Centrifugation	5.0	6.0	2.2	4.4	1.4	1.9

In June 2014, USFWS collected and analyzed 228 samples from the CAWS to test for the presence of AC eDNA upstream of the electric barrier near Lake Michigan. These samples were processed via filtration and the DNA was extracted using the DN kit. Then, the extract was analyzed using qPCR. However, these samples taken in the CAWS are analyzed and reported only as the number of samples that are positive and negative for BH or SC eDNA, rather than the quantitative copy numbers of the six species-specific markers tested in the qPCR replicates. A sample is determined to be positive for BH, if both AC-specific markers and at least one BH-specific marker is present in the sample, and a sample is determined to be positive for SC, if both AC-specific markers and at least one SC-specific marker is present in the sample. One out of the 228 samples tested positive for BH and 7 out of 228 samples tested positive for SC.

4.2.2 Bayesian Model Framework

To estimate the concentration distributions of the six species-specific DNA markers at a given sampling location from the eDNA sampling data, I model the expected number of copies of each

marker that is present after each step of the eDNA sampling process (water sample collection, DNA capture and extraction into elution, preparation of qPCR reaction wells and amplification), given the initial eDNA concentration at the sampling location. The choice of sampling protocol variables (e.g., centrifugation or filtration) influence how much DNA is present after each step of the protocol and ultimately the amount of DNA that is detected and/or quantified in each of the qPCR replicate wells. The following model framework descriptions are based on the model described in Chapter 3.

4.2.2.1 Estimating eDNA concentration distributions in the Illinois River using qPCR copy number data

Let's assume that for a given sampling location *j*, there is a concentration distribution of each target marker *i* with mean concentration, μ_{ij}, and dispersion parameter, *r*. I assumed that the eDNA would be distributed similarly for each target marker and at each sampling location. I model the number of copies of each marker *i* collected in each sample *k* from sampling location *j*, as a two-parameter negative binomial random variable, which depends on the concentration distribution of the marker (μ_{ij}, *r*) and the sample volume (v_s).

$$a_{ijk} \sim NegBinom(\mu_{ij} v_s, r) \quad (4\text{-}1)$$

Then, the DNA collected in each sample is captured and extracted and placed in an elution volume. This depends on the capture efficiency (ϕ_c) and the extraction efficiency (ϕ_e). The number of copies of each marker *i* collected in each sample *k* from sampling location *j* after capture and extraction (e_{ijk}) can also be modeled as a negative binomial random variable, with mean $\mu_{ij} v_s \phi_c \phi_e$ and dispersion parameter *r* (see section 3.2.1.2).

$$e_{ijk} \sim NegBinom(\mu_{ij} v_s \phi_c \phi_e, r) \quad (4\text{-}2)$$

Then, qPCR replicates are prepared using the extracted DNA. I assume that the extracted DNA in the eluate is well-mixed, and so the number of copies of each marker *i* in sample *k* delivered to each qPCR replicate *l* (x_{ijkl}) can be modeled as a Poisson random variable with mean $e_{ijk}v_r/v_e$, where e_{ijk} is the number of copies in the eluate, v_e is the elution volume and v_r is the volume of elution used per replicate.

$$x_{ijkl} \sim Poisson(e_{ijk}v_r/v_e) \quad (4\text{-}3)$$

For analyzing the quantitative copy number data from the eDNA sampling in the Illinois River, I am given the total number of copies of each marker in the sample across the eight replicates tested per sample (y_{ijk}). This is modeled as the sum of eight x_{ijkl}, which is a Poisson random variable with mean $e_{ijk}v_r/v_e*8$.

$$y_{ijk} \sim Poisson(e_{ijk}v_r/v_e * 8) \quad (4\text{-}4)$$

I use equations 4-2 and 4-4 to infer the mean concentration (μ_{ij}) and dispersion (r) of each marker *i* at each sampling location *j*, given the sampling data (y_{ijk}) for each marker *i* in sample *j* collected at sampling pool *k* and the sampling protocol parameters (sample volume (v_s), capture efficiency (ϕ_c), extraction efficiency (ϕ_e), elution volume (v_e), and reaction volume (v_r)),

4.2.2.2 Estimating eDNA concentration distributions in the CAWS from positive/negative data

This modeling framework is then adapted to estimate the eDNA concentration distribution, if the eDNA sampling results are simply reported as either positive or negative for the target species' DNA, like in the CAWS.

First, I model the probability that target marker *i* from sampling location *j* in sample *k* is available for qPCR analysis (θ_{ijk}). This probability (equation 4-2) is one minus the probability that no target

marker copies are present in the elution after capture and extraction, which is dependent on the concentration distribution of target marker i in sampling location j (μ_{ij}, r), the sample volume (v_s) and the capture (ϕ_c) and extraction efficiencies (ϕ_e). This is similar to the availability probability of DNA used in the site-occupancy modeling framework, but is modeled explicitly as a function of the concentration distribution of the target marker in the sampling location (μ_{ij}, r) and the protocol parameter values (v_s, ϕ_c, ϕ_e).

$$\theta_{ijk} = 1 - P\left[e_{ijk} = 0\right] = 1 - \left(\frac{r}{r+\mu_{ij}v_s\phi_c\phi_e}\right)^r \quad (4\text{-}5)$$

The next step of the eDNA process is the likelihood that the target species' DNA is successfully detected using qPCR given that DNA is available for analysis. For this step, I model the amount of DNA that is delivered to each of the qPCR replicates. Assuming that DNA is present in the eluate ($e_{ijk} > 0$), I approximate e_{ijk}, a negative binomial random variable, as a gamma random variable. Then, given that the DNA in the eluate is well-mixed, the number of copies of each marker i collected at sampling pool j in sample k delivered to each replicate l (x_{ijkl}) is a Poisson random variable with a gamma-distributed mean (e_{ijk}), which reduces to a negative binomial distribution with mean $\mu_{ij}v_s\phi_c\phi_e v_r/v_e$ and dispersion parameter r.

$$x_{ijkl} \sim NegBinom(\mu_{ij}v_s\phi_c\phi_e v_r/v_e, r) \quad (4\text{-}6)$$

Then, assuming that each copy of the target marker i in the replicate has an equal probability of amplification (ψ_i), I model the total number of copies that undergo amplification in each replicate as a binomial distribution with probability ψ_i and x_{ijkl} possible copies for amplification. Since x_{ijkl} is a negative binomial distribution, the number of copies of target marker i that amplify in each

replicate (n_{ijkl}) is again a compound binomial-negative binomial distribution, which reduces to a negative binomial distribution with mean $\mu_{ij}v_s\phi_c\phi_e v_r\psi_i/v_e$ and dispersion parameter r.

$$n_{ijkl} \sim NegBinom(\mu_{ij}v_s\phi_c\phi_e v_r\psi_i/v_e, r) \quad (4\text{-}7)$$

Therefore, the probability of successful amplification of target marker i from sampling location j in each replicate l of sample k (p_{ijkl}) is the probability that at least one copy of the target marker amplifies or one minus the probability that zero copies amplify).

$$p_{ijkl} = 1 - P[n_{ijkl} = 0] = 1 - \left(\frac{r}{r+\mu_{ij}v_s\phi_c\phi_e v_r\psi_i/v_e}\right)^r \quad (4\text{-}8)$$

For the AC eDNA sampling protocol in the CAWS, eight replicates are tested per sample and a sample is determined to be positive for the presence of the target marker if at least one of the replicates of the sample is positive for the target marker. Therefore, the probability of a positive detection of marker i of sample k collected from sampling location j (π_{ijk}) is one minus the probability that none of the eight replicates successfully detect the target marker.

$$\pi_{ijk} = 1 - (1 - p_{ijkl})^8 = 1 - \left(\left(\frac{r}{r+\mu_{ij}v_s\phi_c\phi_e v_r/v_e\psi_i}\right)^r\right)^8 \quad (4\text{-}9)$$

The overall probability of a positive sample (Ω_{ijk}) for the presence of target marker i given the eDNA concentration distribution (μ_{ij}, r) and the sampling protocol parameters is the probability that eDNA is available in the elution (θ_{ijk}) multiplied by the probability that the target marker is successfully detected in at least one replicate (π_{ijk}).

$$\Omega_{ijk} = \theta_{ijk} * \pi_{ijk} = \left[1 - \left(\frac{r}{r+\mu_{ij}v_s\phi_c\phi_e}\right)^r\right] * \left[1 - \left(\frac{r}{r+\mu_{ij}v_s\phi_c\phi_e v_r/v_e\psi_i}\right)^{8r}\right] \quad (4\text{-}10)$$

For AC eDNA sampling protocol in the CAWS, a sample is determined to be positive or negative for BH or SC based on the successful detection of multiple markers not just one species-specific marker. Specifically, a sample is positive for BH eDNA or SC eDNA if the sample amplifies both of the AC-TM1 and AC-TM3 markers, as well as at least one of the species-specific markers respectively, BH-TM1/BH-TM2and SC-TM4/SC-TM5, in at least one of the replicates. The amplifications do not need to occur in the same replicate. Assuming that the marker indices *i* are numbered as follows (1 – ACTM1, 2 – ACTM2, 3- BHTM1, 4 – BHTM2, 5 – SCTM4, 6 – SCTM5), the overall probability of a positive eDNA sample for each of the species ($\Omega_{BH,jk}$, $\Omega_{SC,jk}$) in sample *k* from sampling location *j* is:

$$\Omega_{BH,jk} = \Omega_{1jk} * \Omega_{2jk} * \left(1 - \left(1 - \Omega_{3jk}\right) * \left(1 - \Omega_{4jk}\right)\right)$$

$$\Omega_{SC,jk} = \Omega_{1jk} * \Omega_{2jk} * \left(1 - \left(1 - \Omega_{5jk}\right) * \left(1 - \Omega_{6jk}\right)\right) \quad (4\text{-}11)$$

Finally, these probabilities of a positive sample for BH and SC eDNA, $\Omega_{BH,jk,}$ or $\Omega_{SC,jk}$, are realized as positive (y = 1) or negative (y = 0) samples using Bernoulli random variables.

$$y_{BH,jk} \sim Bernoulli(\Omega_{BH,jk})$$

$$y_{SC,jk} \sim Bernoulli(\Omega_{SC,jk}) \quad (4\text{-}12)$$

Therefore, given the number of positive and negative samples of BH and SC eDNA at each sampling location *j* ($y_{BH,jk}$, $y_{SC,jk}$), the parameters of the sampling protocol (sample volume (v_s), capture efficiency (ϕ_c), extraction efficiency (ϕ_e), elution volume (v_e), and reaction volume (v_r)) and amplification probabilities of each marker (ψ_i), I use equations 4-10 to 4-12 to calculate the mean concentration (μ_i) and dispersion (r) of each marker *i* in sampling location *j* given the eDNA

sampling results at each sampling location using a Markov Chain Monte Carlo (MCMC) simulation method.

4.2.3 MCMC model parameterization

MCMC is a Bayesian inference method for estimating the posterior probability distributions of model parameters, in this case, the mean eDNA concentration and dispersion at each sampling location and the sampling protocol parameters. MCMC works by simulating observations from the posterior probability distribution, obtained via Bayes' rule, $p(\boldsymbol{\theta}|x) = \frac{p(x|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(x|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}$, where $p(\boldsymbol{\theta}|x)$ is the posterior distribution of the model parameters $\boldsymbol{\theta}$ given the data x, $p(\boldsymbol{\theta})$ is the prior distribution of the model parameters and $p(x|\boldsymbol{\theta})$ is the likelihood of the data. The idea behind MCMC is to estimate the probability of $p(\boldsymbol{\theta}|x)$, assuming that generating a large random sample from the probability density function is similar to knowing the exact form of the distribution.[105]

One MCMC method is the Gibbs sampling algorithm, which sequentially samples from each model variable conditional on the other variables $p(\theta_i|\ \theta_{j\neq i},x)$ to produce a Markov chain, where the stationary distribution is the joint posterior distribution $p(\boldsymbol{\theta}|x)$.[106] Gibbs sampling is particularly helpful for sampling the posterior distribution of a Bayesian model that is defined by a collection of conditional distributions (which is the case for this model). Then, by running the Markov chain to equilibrium (after a "burn-in" period), the samples approximate the distribution of $p(\boldsymbol{\theta}|x)$.

The prior model parameter values for the eDNA sampling protocol variables applied in 2014 for sampling in the Illinois River and the CAWS are shown in Table 4-2. Estimates of these values were determined from analysis of the official protocol listed in the Quality Assurance Protection Plan (QAPP) and data from laboratory experiments performed by USFWS' Whitney Genetics Lab,

as described in Chapter 3. The capture and extraction efficiencies are modeled using logit distributions in order to capture the uncertainty in these estimates. These estimated distributions for the efficiency parameters are included as priors in the model. The amplification probabilities for the six species-specific markers (ψ_i) can be found in Table B-6.

Table 4-2. Priors for the sampling protocol parameters for the eDNA sampling performed in the Illinois River (La Grange, Marseilles and Brandon Road pool) and the CAWS in 2014

Sampling protocol parameter	Sampling Location (*Capture technique*)		
	Illinois River (*Filtration*)	Illinois River (*Centrifugation*)	CAWS (*Filtration*)
Sample volume, v_s	2 L	0.25 L	2 L
Capture efficiency, ϕ_c mean (95% CI)	3.5% (2.0% - 6.2%)	6.1% (3.5% - 10%)	3.5% (2.0% - 6.2%)
Extraction efficiency, ϕ_e mean (95% CI)	33% (28% - 39%)	33% (28% - 39%)	33% (28% - 39%)
Elution volume, v_e	~700 µL	200 µL	~700 µL
Reaction volume, v_r	3 µL	3 µL	3 µL

For the model, I assumed that there would be a different mean concentration (μ_{ij}) of each marker ($i = 1,\ldots, 6$) at each of the four sampling locations (j = 1 (La Grange), 2 (Marseilles), 3 (Brandon Road), 4 (CAWS)) and an overall value for the dispersion parameter (r) for all markers and sampling locations, assuming that each marker is distributed similarly at all locations.

The model was run using the Just Another Gibbs Sampler (JAGS) program in the R programming language.[77] I specified an uninformative prior for estimating the mean concentration (μ_{ij}) for each marker i, sampling location j, assuming that $\log(\mu_{ij})$ values are drawn from the same normal distribution. The mean of this distribution was given a flat normal prior with mean 0 and variance 10000, while the standard deviation was given a flat gamma prior with scale and shape both 0.001. The dispersion parameter (r) was given a flat gamma prior with scale and shape both 0.001. I used

three chains with 100,000 iterations, discarding the first 50,000 steps as burn-in, and checked for convergence using the Gelman-Rubin statistic.[107]

4.2.4 Evaluating AC occupancy in the CAWS from eDNA sampling results

The probability that AC are present in the CAWS (occupancy probability) is evaluated by calculating the probability that the estimated eDNA concentration of each marker in the CAWS are similar to the estimated eDNA concentration of each marker in the sampling locations where AC are present (Marseilles pool) or absent (Brandon Road pool). The occupancy probability defines the probability that both AC species are present in the CAWS (like in the Marseilles pool), and not each species individually. This is calculated via a classification algorithm described below.

Let's define Φ_O as the probability of AC presence in the CAWS. The classification algorithm takes a random draw (x_i) from each marker's estimated mean concentration distribution in the CAWS (μ_{i4}) and determines the likelihood of that concentration value in the mean concentration distribution in Marseilles pool ($P(x_i|\mu_{i2})$) and the concentration distribution in Brandon pool ($P(x_i|\mu_{i3})$). The lognormal distribution is used to model the concentration distributions. Then, the product of the likelihoods of being in each of the pools is calculated to determine the total likelihood of presence ($\prod_{i=1}^{6} P(x_i|\mu_{i1})$) and absence ($\prod_{i=1}^{6} P(x_i|\mu_{i2})$). Finally, the probability that AC are present in the CAWS (Φ_O) is determined by dividing the total likelihood of being present by the sum of the two likelihoods: $\Phi_O = \frac{\prod_{i=1}^{6} P(x_i|\mu_{i1})}{\prod_{i=1}^{6} P(x_i|\mu_{i1}) + \prod_{i=1}^{6} P(x_i|\mu_{i2})}$. This process is repeated for 50,000 draws to get a distribution for the estimate of Φ_O.

4.2.5 Using simulated eDNA sampling results to determine detection limits for inferring occupancy

This algorithm is then applied given a range of hypothetical eDNA sampling results in the CAWS to determine the fraction of positive samples that are required to strongly indicate AC presence or occupancy in the CAWS. This is done by simulating eDNA sampling events (assuming there are 228 samples per event) with a range of positive BH and SC eDNA samples from 0 to 100% positive samples. Then, the mean eDNA concentrations of each marker are estimated for each hypothetical sampling event given the 2014 sampling protocol and 'classified' using the algorithm described in 4.2.4 to determine a probability that AC are present in the CAWS given each hypothetical sampling event.

4.3 Results

4.3.1 Posterior capture efficiency parameters

The posterior distributions of the capture efficiency of filtration (Figure C-2) and centrifugation (Figure C-3) were updated based on the eDNA sampling results in the Illinois River. The estimated posterior distribution for the capture efficiency for centrifugation increased from 6.1% (3.5% - 10%) to 10% (6.7% - 15%) and the posterior distribution for the capture efficiency for filtration decreases from the prior distribution of 3.5% (2.0% - 6.2%) to 2.1% (1.4% - 3.2%).

4.3.2 eDNA concentrations in the Illinois River and CAWS

The estimated posterior distributions of the mean concentration of each target marker (μ_{ij}) at each of the sampling locations are shown in Figure 4-1. Overall, the estimated mean concentrations are largest in the La Grange pool and Marseilles pool, where AC are present, and are lower where AC are not present (Brandon Road pool and the CAWS). For the AC-TM1 marker, the mean

concentration in the La Grange pool is 4,030,000 (95% credible interval: 1,840,000 – 8.340,000) copies/L and in the Marseilles pool, it is 442,000 (215,000 – 870,000) copies/L. In the Brandon Road pool, this decreases to 19.800 (9,490 – 39,200) copies/L and it is 9,420 (1,020 – 78,900) copies/L in the CAWS. All of the estimated posterior distributions for the mean concentration values for the six target markers in the three locations can be found in Table C-2. The variability of the mean eDNA concentration of each target marker is much larger in the CAWS compared to the other pools.

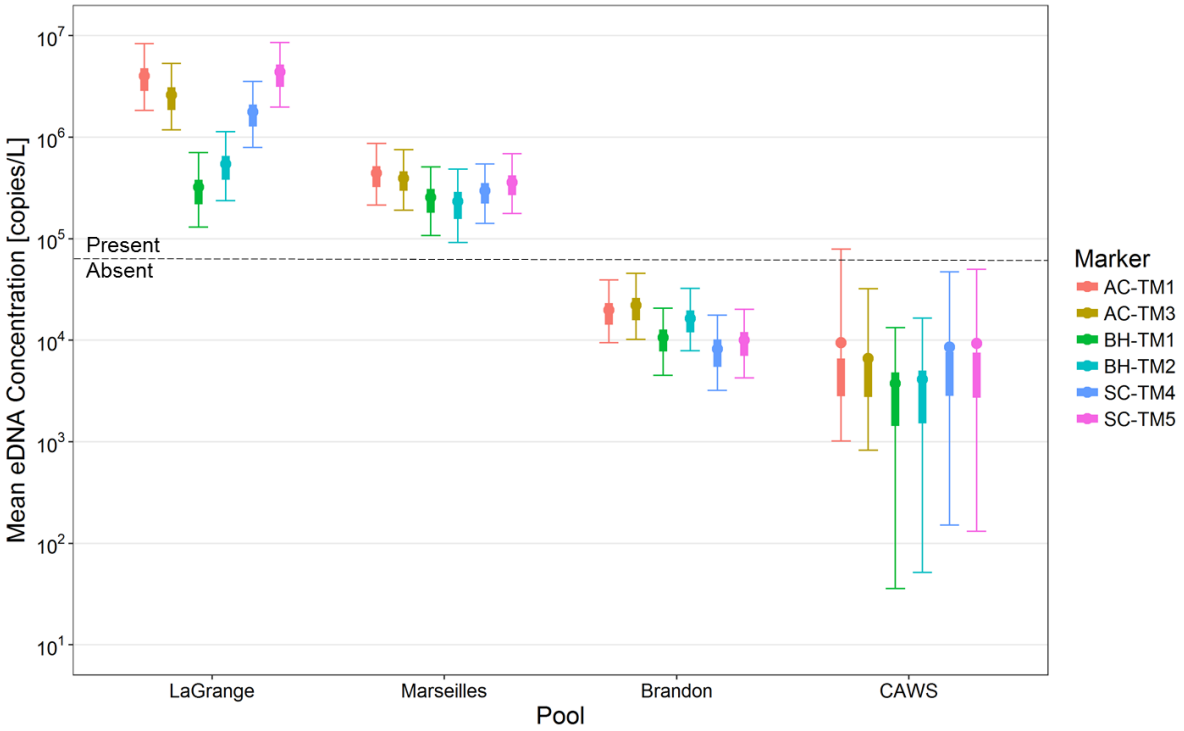

Figure 4-1. The estimated mean concentration (μ_{ij}) of each marker (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) in the three pools of the Illinois River (La Grange, Marseilles, Brandon Road) sampled in Spring 2014 and in the CAWS, which was sampled in June 2014. The circles indicate the mean and the thick and thin lines show the 50% and 95% credible intervals, respectively.

The posterior value for the dispersion parameter, *r*, for the three sampling locations and six markers was estimated to be 0.18 (0.16 – 0.20). The cumulative negative binomial density function of the number of copies of the AC-TM1 marker at each sampling location, given the mean concentration

values and the dispersion parameter in Figure 4-2. The figure shows that the estimated concentration distributions are highly dispersed and clumped, and so there is a likelihood of not collecting DNA in a sample (i.e., a false negative), even at high concentrations. The probability of not collecting the AC-TM1 marker in a given sample at each location can be seen where each curve meets the y-axis. For example, in the La Grange pool, where the mean concentration of the AC-TM1 marker is $4x10^6$ copies/L, there is a 4.2% (3.0 to 5.5%) chance that no copies are collected in a 2-L sample. In the Marseilles pool, this probability is 6.2% (4.6 to 7.8%). %). In the Brandon Road pool, this probability decreases to 11% (95% CI: 8.5 to 13%), and in the CAWS, it is 15% (9.2 to 22%).

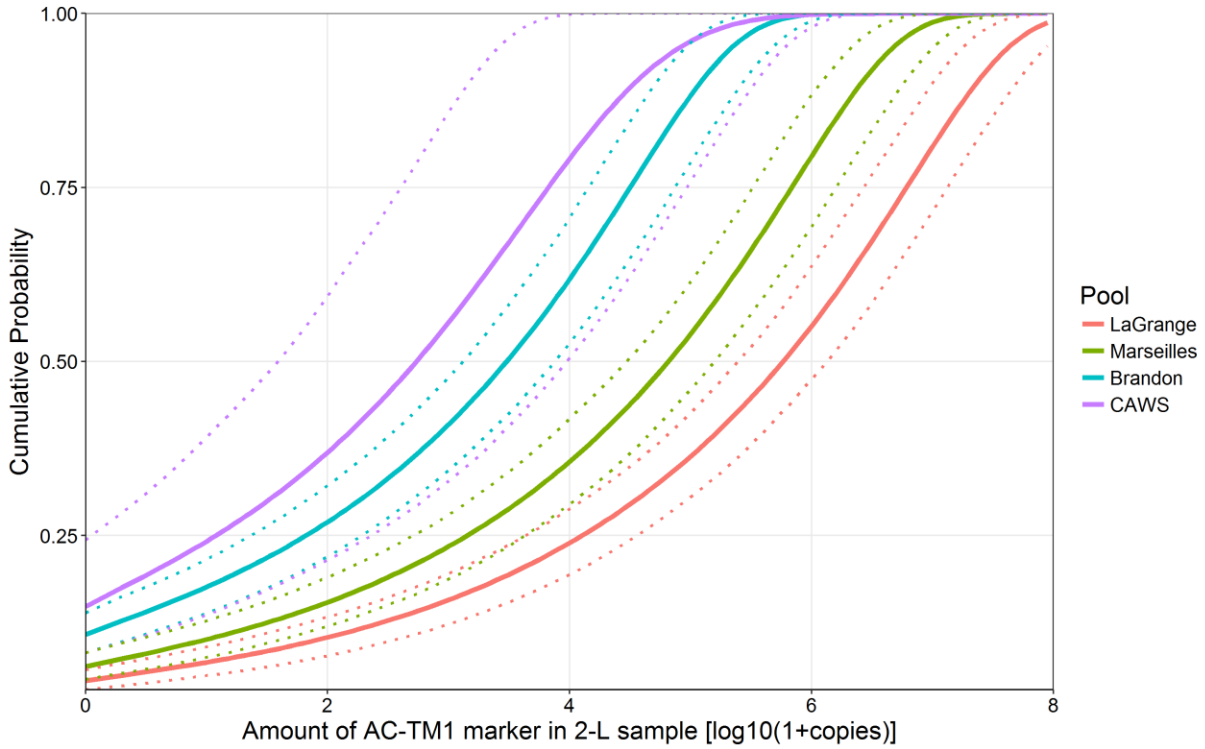

Figure 4-2. The cumulative negative binomial density function of the number of copies of the AC-TM1 marker collected in a 2-L sample from the three pools of the Illinois River (La Grange, Marseilles, Brandon Road) and in the CAWS, given their respective means and the overall dispersion parameter, *r*, value of 0.18 (0.17-0.20). The 95% credible intervals are shown as dotted lines. The x-axis for the number of copies is shown as log10(1 + copies) for interpretability

4.3.3 Occupancy probability in the CAWS for 2014

The posterior mean concentration distributions for each target marker of AC in the Marseilles pool, Brandon Road pool and the CAWS are applied to determine the likelihood of AC occupancy in the CAWS. I did this by using the algorithm described in 4.2.4 to calculate the similarity of the mean concentration distributions in the CAWS to the distributions in the Marseilles pool where AC have invaded. Using this method, the estimated occupancy probability in the CAWS given the June 2014 sampling results (7 out of 228 samples positive for SC eDNA and 1 out of 228 samples positive for BH eDNA) was 0%.

4.3.4 Determining detection limits for inferring AC occupancy in the CAWS using simulated eDNA sampling results

I then evaluated the probability of inferring AC occupancy in the CAWS given a range of hypothetical BH and SC eDNA sampling results. The mean probability of AC occupancy as a function of the fraction of positive BH and SC eDNA samples in a hypothetical sampling event of 228 total samples is shown in Figure 4-3. The figure shows that when the fraction of positive samples is between 0 and ~33% (76 positive sample out of 228 total samples) for BH or SC, the occupancy probability is ~ 0%, i.e. the estimated posterior distributions of the mean eDNA concentrations are very similar to the lower concentrations in Brandon Road pool where AC are absent compared to the higher concentrations in Marseilles pool where AC are present. As the fraction of positive samples for either species approaches one, the mean occupancy probability goes to one as well, since the estimated posterior distributions will be larger and more similar to the concentrations where AC are present. However, there is a transition zone (indicated by the orange on the figure), where it is uncertain whether the estimated distributions from the eDNA results indicate a strong signal of presence (i.e., similarity to concentrations in Marseilles pool) or noise (i.e., similarity to concentrations in Brandon Road pool).

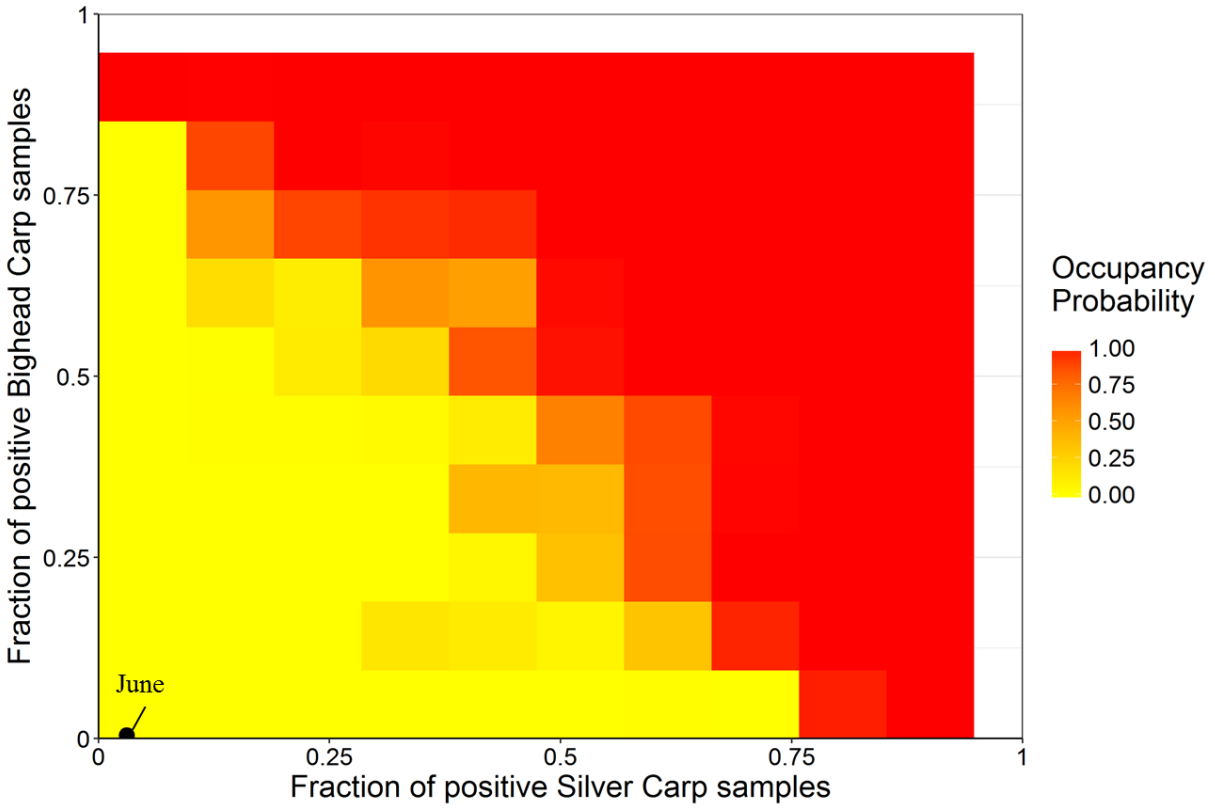

Figure 4-3. The probability of Asian carp occupancy in the CAWS (red – 1; yellow – 0) given the fraction of positive Silver Carp and Bighead Carp samples for a simulated sampling event of 228 samples. The actual historical results of the sampling event performed in June (SC = 7/228, BH = 1/228) is shown as a black dot.

4.4 Discussion

In this chapter, I applied the statistical model from chapter 3 to estimate the mean concentration and dispersion of AC-specific DNA markers from both qPCR copy number data in the Illinois River and positive/absence data in the CAWS. In the Illinois River, the model shows that the concentration distributions are highest downstream in the Illinois River (La Grange and Marseilles pools) and decrease moving further upstream towards Lake Michigan (Brandon Road pool). The few positive detections of BH and SC eDNA upstream of the electric fish dispersal barrier in the CAWS (which should be preventing the upstream invasion of BH and SC) indicate eDNA concentration distributions that are very similar to the distributions found in the Brandon Road pool.

The model estimates a fairly high mean concentration of AC eDNA markers (~10^4 copies/L) in the Brandon Road pool and the CAWS, which are both upstream of the conventionally-held invasion front. One interpretation for this is that eDNA sampling is more sensitive than the conventional methods, and thus the positive eDNA sampling results indicate that AC have invaded both Brandon Road and the CAWS. However, intensive fishing and capture efforts throughout the decade have found no AC in the Brandon Road pool and only one live BH carp in the CAWS upstream of the electric barrier. I conclude that the positive eDNA sampling results in the Brandon Road pool and the CAWS are detections of trace, background or 'noisy' eDNA brought into these systems, unrelated to the recent, live presence of the target species. The high sensitivity of eDNA sampling may actually be causing erroneous detections, due to the ability to detect small concentrations of target DNA.

The results show that there is significant clumping (r = 0.18) of AC eDNA, which affirms the hypothesis that eDNA is most likely found in a clumped distribution in the water column rather than being randomly-distributed.[103] Therefore, a negative binomial distribution should be used when modeling eDNA concentration compared to the Poisson distribution. The clumped distribution also helps explain why in the La Grange pool, there were a few negative samples, despite the large estimated mean concentrations for the AC markers (~10^6 to 10^7 copies/L) (Table C-1).

The method of deriving AC eDNA concentrations described in this chapter differs from the method used by Schultz et al., which estimated that the AC eDNA concentrations in the CAWS are ~ 100 copies/L.[108] In comparison, my model estimates that the eDNA concentrations in the CAWS are 1,000 to 10,000 copies/L, which is a 10 to 100 fold increase. The key differences are that my model

includes the additional losses due to the filtration step and the use of a negative binomial distribution rather than a Poisson distribution. Including the low capture efficiency of filtration (~4%) means that larger eDNA concentrations are estimated at the sampling location for similar numbers of positive eDNA samples. Also, modeling the eDNA concentration distribution as a negative binomial distribution with the additional dispersion parameter means that the estimated mean eDNA concentrations can be larger if eDNA is clumped. For example, when using a Poisson distribution (i.e., no *r* parameter) in my model rather than a negative binomial distribution, the estimated mean eDNA concentration values for BH and SC drop to 200 to 2,000 copies/L.

5 Value of Monitoring Information for Making Decisions about Asian Carp Prevention

5.1 Background

Since 2002, the U.S. Army Corps of Engineers (USACE) has been managing a series of electric fish dispersal barriers in the CAWS to prevent the upstream passage of invasive AC and other invasive fish species into the Great Lakes. The leading edge of AC invasion is currently in the Dresden Island pool, 18 miles south (downstream) of the barrier.[109] In recent years, however, the efficacy of the electric barrier as the solution for long-term AC prevention has been called into question, due to the approaching invasion front and the detection of AC eDNA upstream of the electric barrier.[110]

To improve the efficacy of their AC prevention efforts, the USACE is conducting a study to consider building additional prevention structures (e.g., another electric barrier) in the Brandon Road pool, just upstream of the invasion front. The decision of whether or not to invest in more preventative measures will be dependent on the prior belief about the current invasion status of AC (i.e., whether or not AC have invaded the CAWS), which is informed by the monitoring data.

Studies show that early detection of invasive species allows for more cost-effective invasive species management decisions.[111] Conventional monitoring tools, like electrofishing and netting, have poor sensitivity at low fish densities and thus may not be detecting the early invasion of AC past the barriers.[112] In contrast, eDNA sampling is highly sensitive, able to detect trace amounts of DNA, but, as shown previously in this dissertation, may not be very specific, since a positive detection of DNA may occur in the absence of live organisms. In this chapter, I develop a decision tree model to help evaluate the optimal strategy for AC prevention from the available strategies,

and to determine if available monitoring tools (like eDNA sampling and electrofishing) offer any useful information for making this decision.

Decision trees are a way of representing the alternatives available to the decision maker, the uncertainties involved in the decision and the measure of evaluating how well the decision will achieve the desire objective.[113] For this decision-making scenario, the alternatives are the preventative options, the key uncertainties are the current and future probability of AC invasion and the decision rule is choosing the prevention strategy that minimizes the expected costs, which includes the prevention costs of the chosen strategy and the expected invasion damages that AC pose to the Great Lakes ecosystem and related industries given that prevention fails.

The costs are measured using the net present value (NPV) formula to calculate the present value of the future cost flows over a specified timeframe at a given discount rate. I also consider the influence of risk aversion (i.e., aversion to monetary loss) on decision-making using an exponential utility function.[114] Then, the decision tree model is applied to evaluate the value of information (VOI) provided by monitoring for the current invasion status of AC upstream of the electric barrier in the CAWS before making the decision. This analysis is done for a set of hypothetical monitoring methods with detection sensitivities and specificities ranging from zero to one to assess how false positives and false negatives influence the VOI of monitoring for AC invasion.

5.2 Methods and Data

5.2.1 Prevention strategies

Table 5-1 shows a list of the possible prevention strategies in the CAWS that are assessed in this chapter. The strategies include (1) maintaining the status quo (electric barriers at Lockport pool), (2) building an additional prevention structure at Brandon Road Lock and Dam, (3) hydrologic

separation of the CAWS and (4) to stop operating the electric barriers. The estimated costs of each strategy comes from the Great Lakes Mississippi River Interbasin Report, which is a comprehensive study of the potential technologies that could be utilized to prevent aquatic invasive species transfer between the basins, and the current Asian Carp Action Plan.[115,116] The preventative ability of each technology was determined from an expert elicitation study, which asked 11 experts on aquatic invasive species to estimate the "percent effectiveness of each prevention action to keep Asian carps from ever establishing in Lake Michigan or its tributaries if one and only one of the following actions is implemented."[117]

Table 5-1. Prevention strategies in the Illinois River and their cost, time frame, and preventative ability

Prevention strategy	Probability of prevention (p_{prev})	Implementation time frame	Annual Preventative Costs
#1: Status quo (electric barriers in the CAWS)	92%	Zero years	$12 million a year (O&M)
#2: Build additional barrier at Brandon Road pool	95%	Three years	$52 million (capital) $12 million a year (O&M)
#3: Lakeside hydrologic separation	99%	25 years	$18 billion (capital) $85 million a year (O&M)
#4: Stop operating electric barriers	0%	Zero years	$0

5.2.1.1 Status Quo: Electric Barriers

Currently, there are two electric barriers in operation in the CAWS: Barrier IIA and Barrier IIB, which were placed into operation in 2009 and 2011 respectively. The first demonstration barrier installed in 2002 is being replaced by Barrier I, which began construction in 2016 and will be complete by 2018. The electric barriers are steel electrodes secured to the bottom of the canal. These electrodes are connected to a control building, with generators which generate a DC pulse

through the electrodes which creates an electric filed in the water that discourages fish from crossing.[118]

I set the prevention probability of the status quo strategy to 92%, which is the median estimate from the expert elicitation of the electric barrier's preventative ability.[117] There are multiple avenues of failure of the barriers. Studies have shown that boat passage through the system of electric barriers can disrupt the electric field and allow fish to pass through along with the barge.[118] The electric barriers must also be shut down occasionally for maintenance purposes, which could be opportunities for fish passage.[119] Times of extreme flooding could cause the water level to rise above the electric field allowing for undeterred movement.[119] The current cost of operating and maintaining the electric barriers is ~$12 million a year, which comes from the actual operating budget.[116]

5.2.1.2 Install additional barrier at Brandon Road pool

Brandon Road pool has been identified as a potential location for an additional barrier, because the pool is located downstream of the confluence of the CAWS and the Des Plaines River (a potential invasion pathway) and there is a 25-foot drop in water elevation from the downstream side of the lower dam to the pool level at the upstream side, which is expected to limit upstream transfer. This has been corroborated by ongoing monitoring efforts, which have found no AC in the Brandon Road pool.[116] USACE is exploring the installation of a non-physical barrier, like electric, acoustic or CO_2 barriers, in the pool.[16,120] For the purpose of this analysis, it is assumed that this strategy will be an additional electric barrier. GLMRIS estimated that the initial capital cost of installing an electric barrier system at a new location would be $52 million, the operation and maintenance cost would be ~$12 million a year and that it would take 3 years to implement.[121]

The combined preventative ability of existing and proposed electric barriers is assumed to be 95%, which is the high estimate of the electric barriers' preventative ability.[117]

5.2.1.3 Full Hydrologic Separation

Hydrologic separation is the complete closure of the CAWS, which effectively removes the canal system as a route for invasion by Asian carp or other aquatic invasive species. The lakeside hydrologic separation proposal described in GLMRIS involves building four physical barriers throughout the CAWS that would sever the aquatic connection. The barriers are designed to prevent any flooding over the barriers up to the 500-year storm.[121] However, this also necessitates building additional reservoirs, water treatment plants and tunnel systems, to manage Chicago's water supply, sewage effluent, water quality and flooding issues, which are currently managed using the CAWS.

The capital cost of the barriers plus the additional infrastructure is $18 billion, over 25 years of implementation (Figure D-1). Once built, there is an estimated operation and maintenance cost of $85 million a year, which is primarily due to operating the additional infrastructure. The physical barriers are built at the end of the construction timeline, once all of the water quality and flood-risk mitigation infrastructure is installed. Thus, the preventative ability of this strategy is not realized until year 25. I assume that the current status quo (electric barriers) persists until the physical barriers are complete. Experts estimate the median preventative ability of hydrologic separation to be 99%.[117] Potential failures include flooding over the barriers as well as human-mediated transfer of AC across the physical barriers.

5.2.1.4 Stop prevention

The final proposed prevention strategy would be to cease operation of the electric barriers. The benefit of this plan would be the cost savings from not operating the barriers. However, the preventative ability would now be 0%. I assumed that this decision to turn off the barriers could be implemented immediately.

5.2.2 Risk and Potential Impacts of Asian Carp

For AC invasion to occur in the Great Lakes, AC must be introduced into the system (e.g., invade past the barriers) and be established in the region (i.e., survive and reproduce successfully). Studies have shown that Asian carp have a high likelihood of establishing in the Great Lakes, if they are introduced.[122–125] Once established, Asian carp species are likely to cause negative economic impacts on the Great Lakes' fishing industry by out-competing native aquatic sport fish species.[13,125,126] AC could also modify the local food web, which could have harmful ecological effects, causing long-term economic damage due to the lack of ecological diversity and health of the Great Lakes. However, the severity of the economic and ecological impacts in the Great Lakes caused by AC is very uncertain. Studies have shown that Asian carp are likely to be established in the Great Lakes once invaded.[122,123,125,127] However, their impact on native fish species could range from causing lower fitness due to competition or helping species that can feed on Asian carp larvae.[125] For the purpose of this analysis, I consider a wide range of estimated damages from $0 to $1 billion a year.

5.2.3 Simulating potential invasion fronts

The key uncertainty for this decision is whether or not the Great Lakes is already invaded by Asian carp and if not currently invaded, when the invasion will occur. In this chapter, I assume that the

probability of invasion is equivalent to the probability of AC being introduced and establishing successfully into the Great Lakes.

I model the invasion status of the Great Lakes (I_t) at each time step from $t = 0$ to T years. The invasion status at each time step is either 0 (not invaded) or 1 (invaded). I assumed that at each time step t, if it is invaded in the previous time period, $t - 1$, ($I_{t-1} = 1$), it will stay invaded in the following time period, i.e., the probability of moving from state 1 to state 1 equals one and the probability of moving from state 1 to state 0 is zero. If it is not invaded in the previous time period $t - 1$ ($I_{t-1} = 0$), the probability it will be invaded in this time period t, depends on the base likelihood of invasion (p_b) multiplied by the probability of the preventative measure failing to prevent the invasion that year ($1 - p_{prev,t}$). The base probability of invasion (p_b) is the probability of invasion each year given no prevention. The prevention probability ($p_{prev,t}$) has a time index, because for some of the strategies (e.g., hydrologic separation and additional barrier), the preventative probability will change over time.

Given the choice of prevention strategy, I generate a simulated invasion front I_t from $t = 0$ to T using the following steps:

1. Assume an initial state of I_0
2. For $t = 1$ to T,
    - If $I_{t-1} = 0$, generate I_t by drawing from a uniform random variable from 0 to 1 (U) and setting $I_t = 1$ if $U < p_t$ and 1 if $U > p_t$, i.e. it becomes invaded with probability $p_t = (p_b)*(1 - p_{prev,t})$.
    - If $I_{t-1} = 1$, then $I_t = 1$

The invasion front depends on the $p_{prev,t}$ of the chosen prevention strategy, the base probability of invasion, p_b and the initial state of invasion I_0. The equations for calculating the invasion front for each of the prevention strategies are shown in Table 5-2.

Table 5-2. The equations to simulate the invasion front of AC from t = 0 to T for each of the four preventative strategies given the base probability of invasion, p_b.

Preventative Option	I_t for t = 1 to T	
#1: Status Quo	$If\ I_{t-1} = 0,$	$I_t = \begin{cases} 1\ if\ U_t \leq p_b * .08 \\ 0\ if\ U_t > p_b * .08 \end{cases}$
	$If\ I_{t-1} = 1,$	$I_t = 1$
#2: Additional control structures	$If\ I_{t-1} = 0,$	$I_t = \begin{cases} 1\ if\ U_t \leq p_b * .08 \\ 0\ if\ U_t > p_b * .08 \end{cases}\ for\ t \leq 3$
		$I_t = \begin{cases} 1\ if\ U_t \leq p_b * .05 \\ 0\ if\ U_t > p_b * .05 \end{cases}\ for\ t > 3$
	$If\ I_{t-1} = 1,$	$I_t = 1$
#3: Hydrologic separation	$If\ I_{t-1} = 0,$	$I_t = \begin{cases} 1\ if\ U_t \leq p_b * .08 \\ 0\ if\ U_t > p_b * .08 \end{cases}\ for\ t \leq 25$
		$I_t = \begin{cases} 1\ if\ U_t \leq p_b * .01 \\ 0\ if\ U_t > p_b * .01 \end{cases}\ for\ t > 25$
	$If\ I_{t-1} = 1,$	$I_t = 1$
#4: Stop prevention	$If\ I_{t-1} = 0,$	$I_t = \begin{cases} 1\ if\ U_t \leq p_b \\ 0\ if\ U_t \leq p_b \end{cases}$
	$If\ I_{t-1} = 1,$	$I_t = 1$

5.2.4 Calculating expected costs and utility given the simulated invasion fronts

The optimal preventative strategy is chosen based on which strategy maximizes the expected net present value of benefits minus costs. In this analysis, I assume there are no benefits and just consider the prevention costs from implementing the strategy and the uncertain invasion costs, which depend on the simulated invasion fronts, and the expected cost of invasion, C_I. I also assume

that the decision is made at $t = 0$ and the construction and operation costs of implementing the chosen prevention strategy is irreversible. The equations for calculating the NPV of each strategy given the simulated invasion forecast (I_t), the cost of invasion, C_I, and the prevention costs for each strategy are shown in Table 5-3. In the NPV calculations, d represents the discount rate. As seen in the model equations, the annual invasion costs (C_I) are not incurred unless the invasion has occurred ($I_t = 1$).

Table 5-3. The net present value (NPV_i) calculation for each of the preventative strategies given the annual cost of invasion, C_I, the discount rate, d, and the invasion front I_t.

Preventative Option	Net Present Value of Expected Invasion Costs + Prevention Costs [$ million] (NPV_i)
#1: Status Quo	$\sum_{t=1}^{T} \frac{-C_I * I_t - 12}{(1+d)^t}$
#2: Additional control structures	$-52 + \sum_{t=1}^{3} \frac{-C_I * I_t - 12}{(1+d)^t} + \sum_{t=4}^{T} \frac{-C_I * I_t - 24}{(1+d)^t}$
#3: Hydrologic separation	$\sum_{t=1}^{25} \frac{-C_I * I_t - 12}{(1+d)^t} + \sum_{t=1}^{20} \frac{-700}{(1+d)^t} + \sum_{t=21}^{23} \frac{-800}{(1+d)^t} + \sum_{t=24}^{25} \frac{-1100}{(1+d)^t} + \sum_{t=26}^{T} \frac{-C_I * I_t - 85}{(1+d)^t}$
#4: Stop prevention	$\sum_{t=1}^{T} \frac{-C_I * I_t}{(1+d)^t}$

The utility (U) of each strategy is also calculated using an exponential utility function, as shown in equation 5-1. This equation converts the NPV of the costs into a utility value. The optimal prevention strategy is the one that maximizes expected utility. In the exponential utility function,

R represents the risk tolerance of the decision-maker, which is defined as the decision maker's willingness to accept risk.[128] Lower values of *R* indicates greater risk aversion (i.e., less willing to accept risk) and is measured by valuing each unit loss of NPV to be exponentially worse (Figure D-2).

$$U = 1 - e^{-NPV/R} \quad (5\text{-}1)$$

This formulation of risk aversion is a monetary one in a decision-analytic framework. The decision-maker is averse to all monetary losses, whether it is due to money spent on prevention or due to damages from invasion. However, the decision-maker may be worried solely about the risk of damage caused by AC and is willing to take on the risk of prevention to avoid those damages. To evaluate this, I also consider an "environmental" risk averse decision-maker, who is averse to expected invasion damages but neutral about the prevention costs.

5.2.5 Decision modeling framework

The decision tree used to solve the decision problem is shown in Figure 5-1. The optimal strategy is the one that maximizes the expected NPV (or utility), which depends on the uncertainty about the initial state of invasion (I_0). Given no additional information about the current state of invasion at time = 0, I assume that the prior probability of invasion that $I_0 = 1$ is the probability of invasion given the current preventative probability and belief about the base probability of invasion: $p_0 = p_b*(1 - .92)$. $NPV_{i,I0=0}$ and $NPV_{i,I0=1}$ are the NPV values for each of the four preventative strategies, assuming that the initial state is either $I_0 = 0$ and $I_0 = 1$, given a value of p_b and C_I. Equation 5-2 shows the expected value of the optimal decision given no information.

$$EV_N = \max_i \left(.08p_b * NPV_{i,I_0=1} + (1 - .08p_b) * NPV_{i,I_0=0}\right) \quad (5\text{-}2)$$

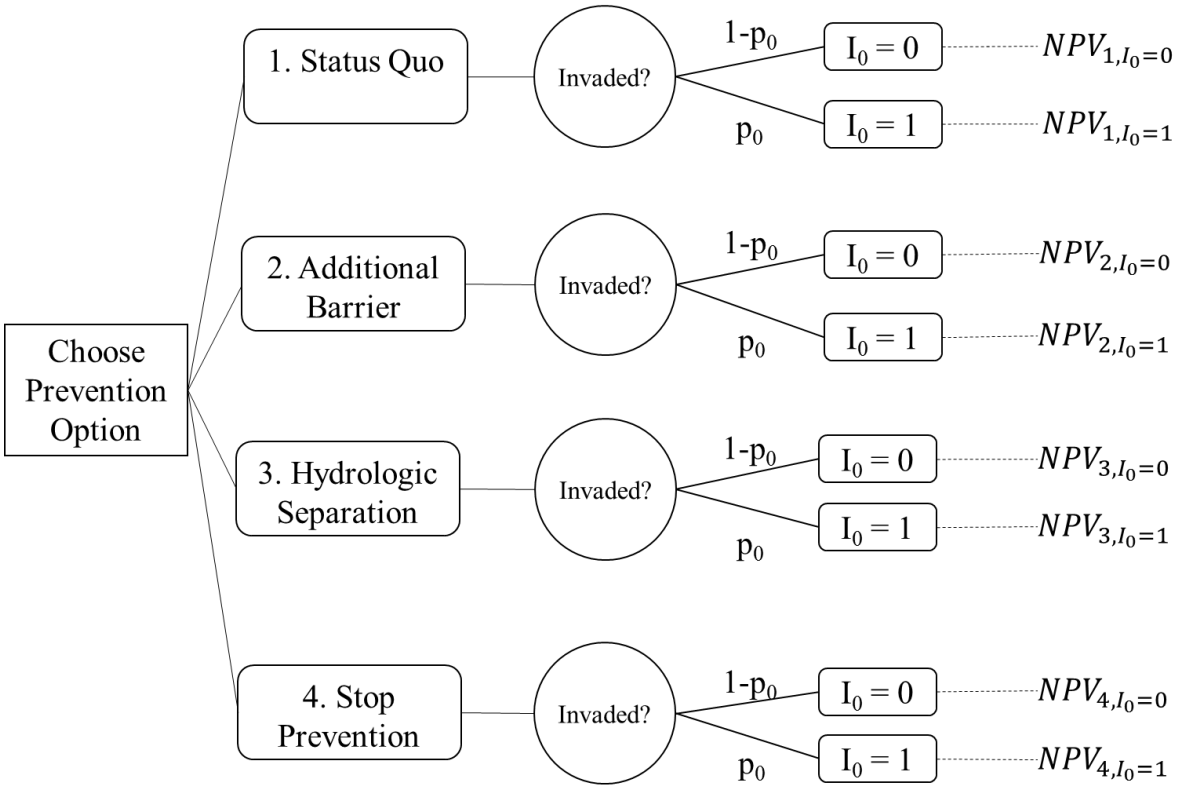

Figure 5-1. Decision tree for analyzing the optimal preventative strategy (1. Status quo, 2. Additional barrier, 3. Hydrologic separation and 4. Stop prevention) given the prior probability of invasion at time = 0, p_0, and the NPV of each strategy for a simulated invasion front.

5.2.6 Calculating the value of perfect information

To calculate the expected value of perfect information (EVPI), I adapt the decision tree to show the monitoring event, which occurs before the decision is made. A figure of the decision tree with a perfect monitoring method is shown in Figure 5-2

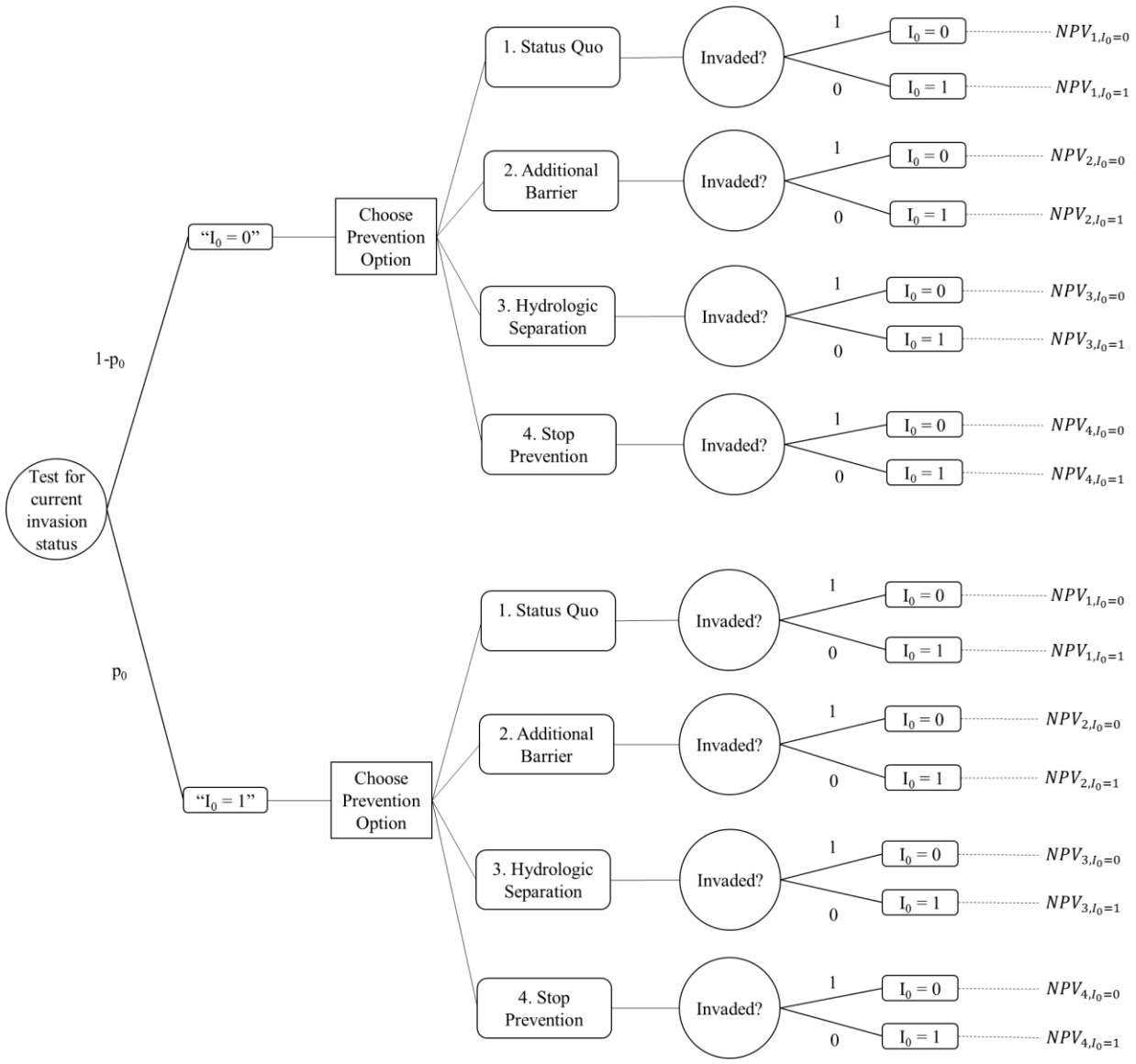

Figure 5-2. Decision tree for evaluating the optimal preventative strategies (1. Status quo, 2. Additional barrier, 3. Hydrologic separation and 4. Stop prevention) given a perfect monitoring method, the prior probability of invasion at time = 0, p_0, and the NPV of each strategy for a simulated invasion front.

The expected value with perfect information is calculated by choosing the preventative strategy that maximizes NPV given that the decision-maker knows $I_0 = 1$, and the preventative strategy that maximizes NPV given $I_0 = 0$, and record the expected value of those 'optimal' decisions. In the scenario of perfect monitoring (i.e., 100% sensitivity and specificity, the probability that monitoring indicates invasion will be equal to the prior probability of invasion, $p_0 = .08p_b$.

$$EV_P = (1 - .08p_b) * \max_i\left(NPV_{i,I_0=0}\right) + (.08p_b) * \max_i\left(NPV_{i,I_0=1}\right) \quad (5\text{-}3)$$

To calculate the EVPI, I simulate 10,000 iterations of potential invasion fronts and get the EV_N and EV_P of each iteration. The EVPI is the expected value gained from conclusively knowing the initial invasion status (I_0) before making the decision. EVPI is calculated by taking the average of the EV_P values minus the average of the EV_N values for the 10,000 iterations. The simulations and EVPI calculations are performed over a range of potential values for p_b and C_I.

5.2.7 Calculating the value of imperfect information

The expected value of imperfect information (EVII) is the expected value gained from imperfect monitoring about what I_0 is before making the decision. A figure of the decision tree given imperfect monitoring information is shown in Figure 5-3.

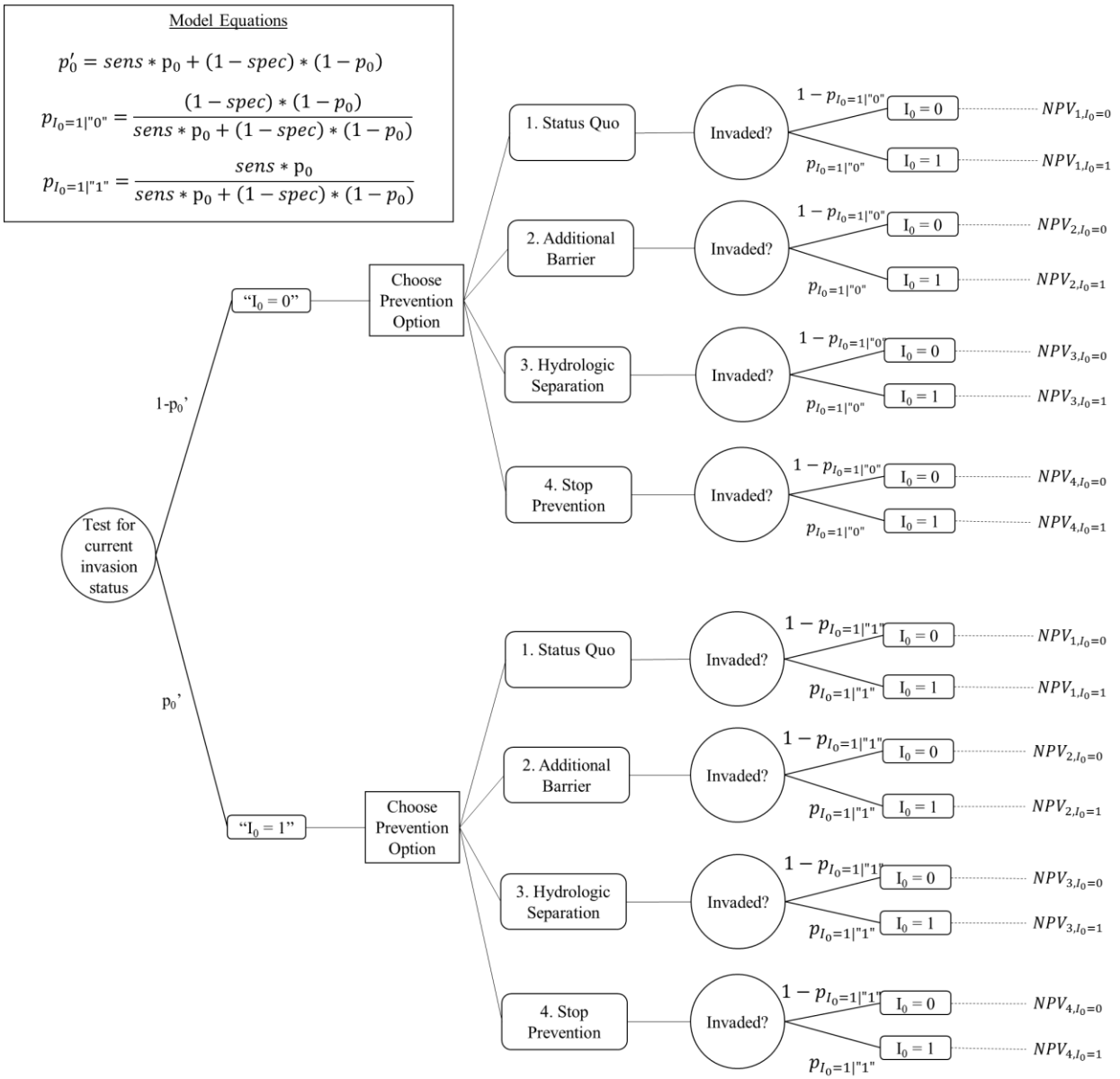

Figure 5-3. Decision tree for evaluating the optimal preventative strategies (1. Status quo, 2. Additional barrier, 3. Hydrologic separation and 4. Stop prevention) given an imperfect monitoring method with sensitivity (sens), specificity (spec), the prior probability of invasion at time = 0, p_0, and the NPV of each strategy for a simulated invasion front.

Given a sensitivity (sens) and specificity (spec) value of a hypothetical monitoring method, the probability of indicating invasion (p') is the true positive rate (sensitivity * prior probability of invasion) plus the false positive rate (1 – specificity)* (1 – prior probability of invasion).

$$p' = sens * (.08p_b) + (1 - spec) * (1 - .08p_b) \quad (5\text{-}4)$$

Then, given that the monitoring method indicates either invasion or non-invasion for I_0, the decision-maker makes the optimal decision. However, unlike the previous scenario, the updated likelihood of invasion or non-invasion given the monitoring information is not conclusive due to the errors in the monitoring method. The probability that it is actually invaded ($I_0 = 1$) given that the monitoring method detected an invasion is:

$$p_{I_0=1|"1"} = \frac{sens*.08p_b}{sens*.08p_b+(1-spec)*(1-.08p_b)} \quad (5\text{-}5)$$

The likelihood that it is actually invaded ($I_0 = 1$) given that the method did not detect an invasion is

$$p_{I_0=1|"0"} = \frac{(1-spec)*(1-.08p_b)}{sens*.08p_b+(1-spec)*(1-.08p_b)} \quad (5\text{-}6)$$

Therefore, the expected value of the decision given imperfect information is shown in Equation 5-7, which is determined by the optimal decisions given the monitoring method yields 0 or 1 and uncertainty about that information at each decision point.

$$EV_I = p' * \max_i\left(p_{I_0=1|"1"} * NPV_{i,I_0=1} + \left(1 - p_{I_0=1|"1"}\right) * NPV_{i,I_0=0}\right) + (1 - p') * \max_i\left(p_{I_0=1|"0"} * NPV_{i,I_0=1} + (1 - p_{I_0=1|"0"}) * NPV_{i,I_0=0}\right) \quad (5\text{-}7)$$

Similar to the EVPI calculation, the EVII is calculated by taking the average of the EV_I minus the average of the EV_N for the 10,000 iterations.

All of the decision model calculations were analyzed using R 3.2.2.[77] Other key assumptions for the analysis include a time frame, *T*, of 50 years and a discount rate of 4%.

5.3 Results

5.3.1 Optimal prevention strategy given uncertainty about invasion

The optimal prevention strategy given no monitoring information as a function of p_b and C_I, is shown in Figure 5-4. The base probability of invasion, p_b, is on the y-axis from 0 to 1 and the annual cost of invasion, C_I, is the x-axis from 0 to $1 billion. For p_b greater than 0.30 and C_I greater than $100 million, the optimal strategy is additional prevention. If p_b is less than 0.30, then the optimal strategy is to stay with the status quo. Increasing probability of invasion leads to greater need for prevention. When either p_b or C_I is zero or near-zero, then the optimal strategy is to stop prevention, since the status quo prevention costs are unnecessary. Hydrologic separation is never chosen as an optimal decision under these assumptions. The optimal prevention strategies are similar when assuming a longer time frame of 100 years, except for that status quo is now preferable to stop prevention at very low invasion probabilities (Figure D-3).

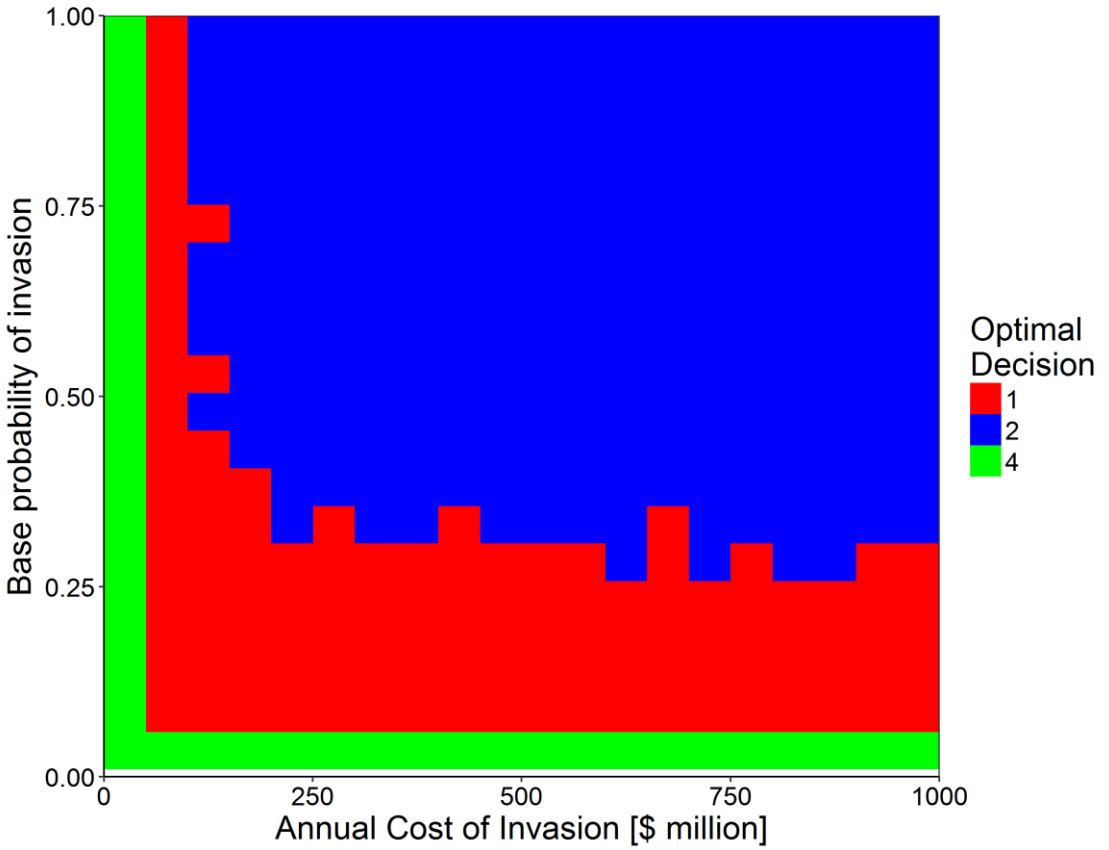

Figure 5-4. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given risk neutrality, time frame of 50 years and a discount rate of 4%. The colors denote the optimal prevention strategy for each combination of invasion costs and base probability of invasion. Red is the status quo, blue is additional barrier, and green is discontinue existing barriers.

5.3.2 Influence of risk aversion on decision-making

Figure 5-5a shows that if the decision-maker is risk averse to monetary losses (R = \$1 billion), then the status quo strategy becomes more preferred to additional prevention for many of the possible combinations of p_b and C_I, except for when annual invasion costs are \$100 to \$250 million a year and there is a high probability of invasion (>50%). Also, stopping prevention is now preferred at very high annual costs and low invasion probabilities. This trend continues for higher risk aversion (R = \$500 million), where the status quo and stopping prevention are the dominant strategies (Figure 5-5b). If the decision-maker is only averse to the risk of damages from invasion, then the optimal prevention strategies look similar to the case of risk neutrality (Figure D-4).

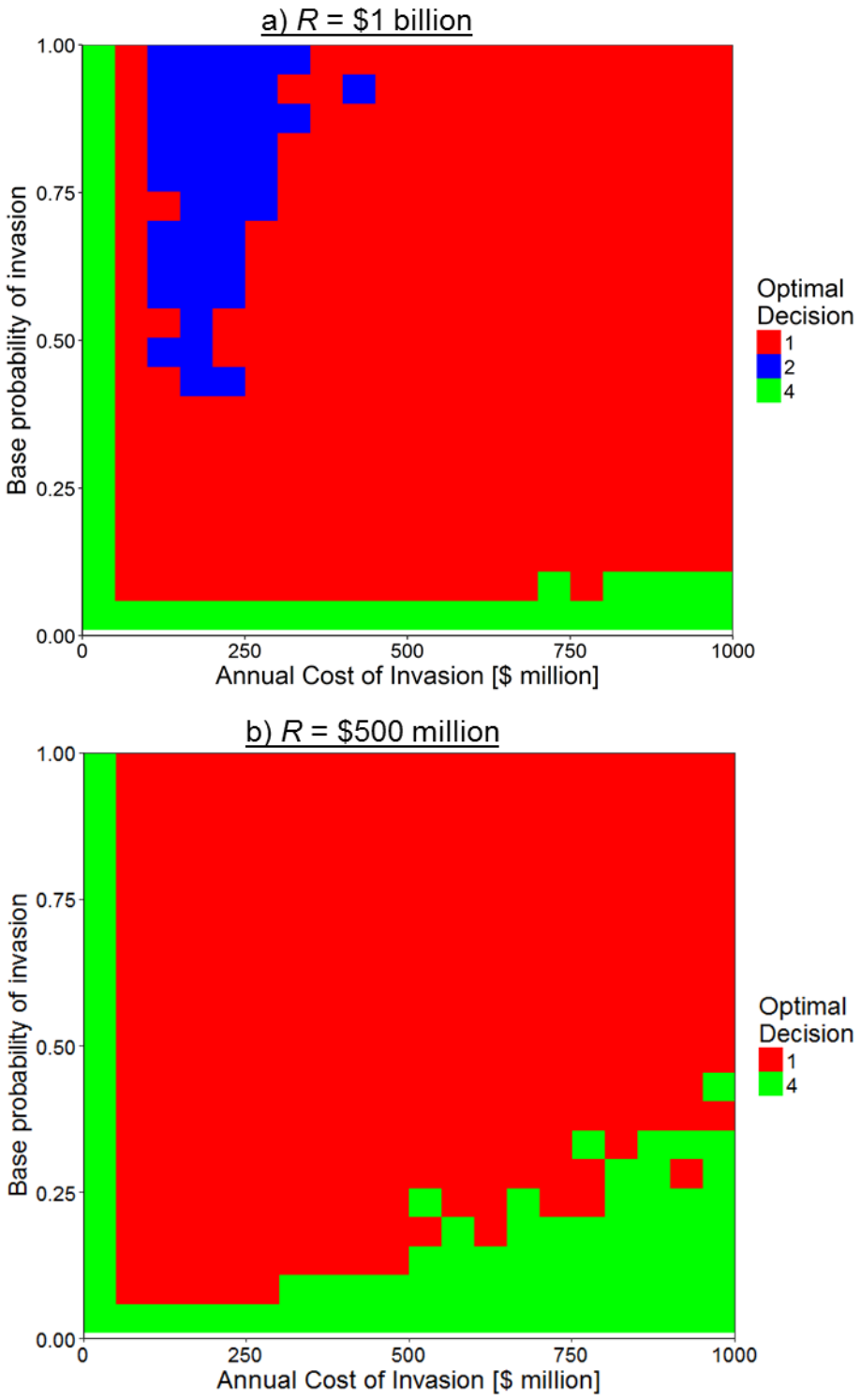

Figure 5-5. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given risk tolerances of (a) $1 billion and (b) $500 million. Red is the status quo, blue is additional barrier, and green is discontinue existing barriers.

5.3.3 Value of information

The expected value of perfect information (EVPI) as a function of p_b and C_I, is shown in Figure 5-6. The figure shows that there is value of information for making this decision if the monitoring method is perfect (i.e., specificity and sensitivity equal to 1). The EVPI ranges from 0 to $71 million over the 50 year time frame over the range of p_b and C_I. The value is highest at high invasion probability and invasion costs and goes to zero when the base probability is low.

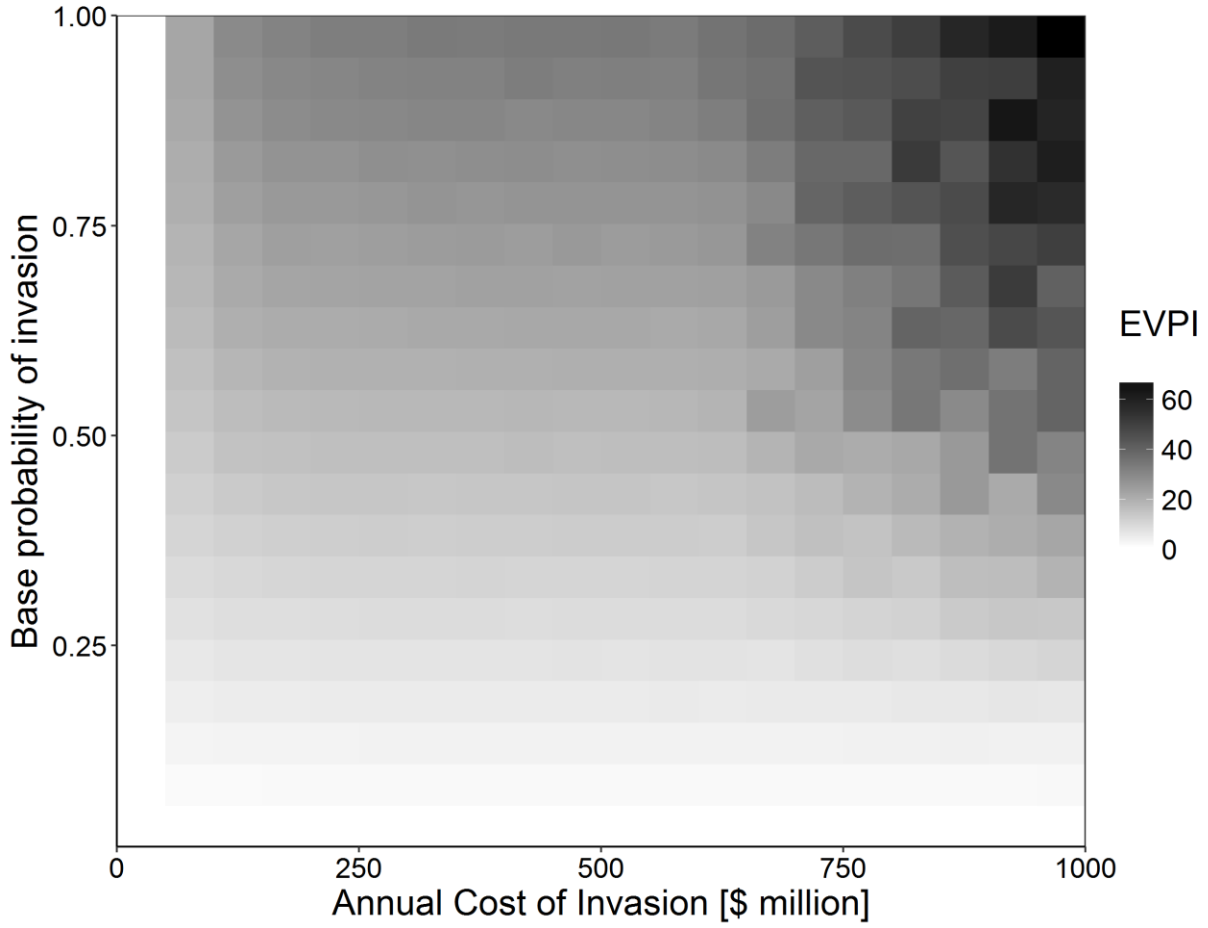

Figure 5-6. Expected value of perfect information in units of million dollars as a function of the annual cost of invasion and the base probability of invasion, given risk neutrality, time frame of 50 years and a discount rate of 4%.

Then, the value of information was evaluated given an imperfect monitoring method (i.e., sensitivity and specificity less than 1). This was done for the scenario where p_b is 0.50 and C_I is

$500 million. The EVPI for this scenario is $16 million. Figure 5-7 shows the expected value of imperfect information (EVII) for this scenario, given information from a hypothetical imperfect monitoring method with specificity and sensitivity ranging from 0 to 1. The results show that the value of monitoring information is primarily due to perfect specificity. When specificity drops below 1, the value of monitoring information goes straight to 0. Even at perfect sensitivity, if the specificity is less than one, there is no added value of information. In comparison, at perfect specificity, the value of information decreases linearly from the EVPI of $17 million to $0, as the sensitivity decreases from 1 to 0.

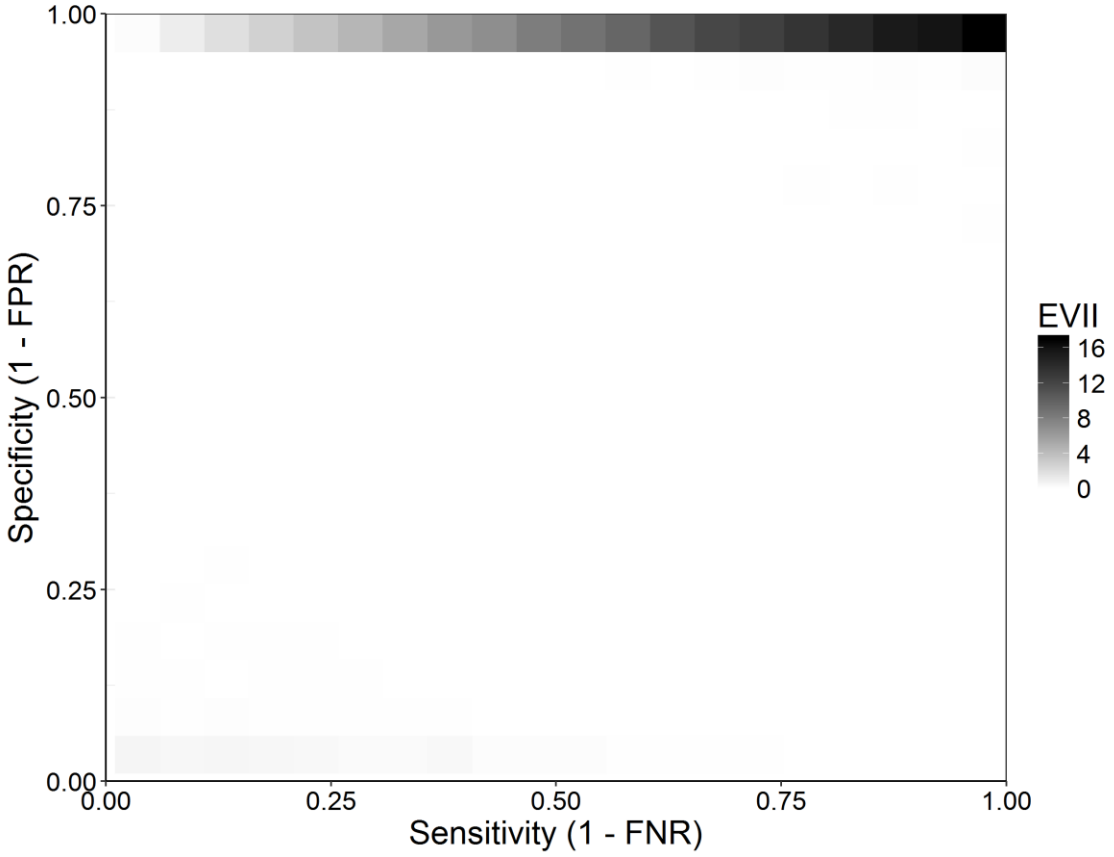

Figure 5-7. The expected value of imperfect information in units of million dollars as a function of (a) specificity assuming sensitivity is 100% and as a function of (b) sensitivity, assuming specificity is 100%. The calculation assumes risk neutrality, time frame of 50 years, a discount rate of 4%, base invasion probability of 50% and an annual invasion cost of $500 million.

5.4 Discussion

This chapter outlines a framework of a decision tree model to evaluate the optimal strategies for preventing Asian carp (AC) passage into the Great Lakes, given uncertainty about the probability of invasion and the annual cost of invasion. This model recommends that additional barriers be installed if there is a moderate baseline probability of invasion and annual invasion costs. At lower values, the status quo is preferred and hydrologic separation is never preferred.

There are many advocates for hydrologic separation of the CAWS, who argue that separation is the only way to ensure that invasive species like AC do not invade the Great Lakes.[59] However, this model shows that this option may not offer the benefits as publicized. If invasion is imminent (i.e., high p_b), then the invasion is likely to occur during the 25-year implementation time before the stricter prevention can take place. However, if there is time to implement separation, (i.e., low p_b), then the other less costly prevention options like staying at the status quo will offer very high levels of prevention at much lower cost. Also, there are additional indirect costs of hydrologic separation to other industries, like cargo shipping, that are not included in this analysis, which must be accounted for when choosing this strategy. This model suggests that hydrologic separation is unlikely to be the optimal strategy for AC prevention.

Greater aversion to monetary risks makes the decision-maker prefer strategies with less prevention. This is because risk-averse decision-makers are averse to the 'risky' costs of additional prevention, which may fail. This has also been seen in studies showing that invasive species managers prefer the known costs of control over the risky costs of prevention, in the management of zebra mussels [129] However, assuming that the decision maker is primarily averse to the risk of invasion and incurring environmental damages (and not the prevention costs), the decision maker will behave

much like a monetary risk-neutral decision maker. Here, the decision-maker is neutral to the cost of prevention and averse to environmental damages and therefore is willing to invest in more prevention in the form of an additional electric barrier. However, even under this scenario, hydrologic separation is not an optimal strategy.

The model shows that there is added value from monitoring, but the VOI requires perfect specificity. If it is known conclusively that AC have invaded (I_o= 1), then the decision-maker can switch to stop prevention and avoid the wasted preventative costs. However, if specificity is less than one, then there is a chance that there are false positives and thus the decision-maker cannot stop prevention with confidence given a positive detection and thus the decision-maker has to make the decision given uncertainty about whether or not it is actually invaded, like before.

6 Conclusions and Recommendations

In this chapter, I summarize the findings from the four studies in this dissertation, discuss their policy implications and potential future work.

6.1 Summary

Chapter 2 describes a beta-binomial regression model which analyzed the SC eDNA detections upstream of the electric barrier from 2009-2012. The model found that the detection probability of SC eDNA had a significantly positive relationship with reverse flow into the CAWS, while other environmental covariates, such as water temperature, gage height, did not. This is significant because it provides a possible alternative explanation for why SC eDNA has been detected upstream of the barrier but intact silver carp have not, i.e., that during reverse flow events, existing background eDNA is diluted to a lesser degree than during normal flow conditions. It also found that the beta-binomial model was a much stronger fit for modeling the probability of eDNA detection than the binomial model, presumably because eDNA is not randomly dispersed

In chapter 3, I developed a model to evaluate the sensitivity of eDNA sampling protocols and applied it to the example of BH and SC eDNA sampling by US Fish and Wildlife Service from 2013 to 2015. The model shows that the sensitivity of the protocols has fluctuated from 2013 to 2015 as changes were made to the protocol. Changes to a more efficient extraction method in 2013, a new qPCR platform in 2014 and a more efficient capture method in 2015 improved the sensitivity, while changes to a larger elution volume in 2013 and a smaller sample volume in 2015 reduced the sensitivity. Overall, when assuming that eDNA is randomly distributed, the sensitivity of the current protocol is higher for BH eDNA detection and similar for SC eDNA detection compared to the original protocol used from 2009-2012. When assuming that eDNA is clumped,

the sensitivity of the current protocol is slightly higher for BH eDNA detection but worse for SC eDNA detection. The model shows that if the eDNA is clumped, an assumption supported by modeling results of the previous chapter, then the sampling protocol parameters earlier in the process have a greater influence on the sensitivity than the parameters later in the process.

In chapter 4, I apply the model developed in chapter 3 to estimate the BH and SC eDNA concentrations in sampling locations in the Illinois River and the CAWS. The model shows that the concentrations are highest in the pools furthest downstream (La Grange and Marseilles) where AC are known to be invaded, and are lower in the locations further upstream (Brandon Road pool and the CAWS), where AC have not invaded. The model also shows that the BH and SC eDNA at these locations is highly clumped. These concentrations are used to show that the eDNA concentrations in the CAWS are much lower than the concentrations in Marseilles pool and are very similar to the concentrations in Brandon Road pool. This indicates that the likelihood of AC occupancy in the CAWS is currently zero. The model is then applied to determine what fraction of positive eDNA samples in the CAWS would indicate concentrations similar to Marseilles pool and thus indicate species presence. The model shows that >30% positive samples are needed to indicate presence and >50% are needed to infer 100% similarity to the Marseilles pool.

In chapter 5, I develop a decision model to evaluate the value of information that monitoring provides for making decisions about BH and SC prevention into the Great Lakes. The optimal preventative strategy is dependent on prior beliefs about the expected damage of AC invasion, the probability of invasion, and whether or not BH and SC have already invaded the Great Lakes. Given no monitoring, the optimal strategy is to stay with the status quo of operating electric barriers in the CAWS for low probabilities of invasion and low expected invasion costs. However,

when the probability of invasion is greater than 30% and the cost of invasion is greater than $100 million a year, the optimal strategy changes to installing an additional barrier in the Brandon Road pool. Greater risk-aversion causes less prevention (e.g., status quo instead of additional barriers) to be more preferred. Over a range of sensitivity and specificity values for a hypothetical monitoring method, the model shows that information provides value only if the method has perfect specificity (false positive rate = 0%). This condition is difficult to meet by eDNA sampling, which suffers from significant background noise.

6.2 Recommendations

6.2.1 Improving the method sensitivity of eDNA sampling protocols

Chapters 4 shows that the AC eDNA in the CAWS and the Illinois River is highly clumped. Thus, the method sensitivity of the AC eDNA sampling protocol is more influenced by the protocol steps that are upstream in the process (e.g., sample volume) (Chapter 3). This suggests that maximizing the method sensitivity of eDNA sampling protocols should focus on higher sample volumes, higher capture efficiencies and extraction efficiencies when designing future protocols. The current sampling protocol (which began to be used in 2015), has the lowest sensitivity for SC eDNA detection out of all the protocols (when assuming a clumped distribution of eDNA), primarily because of the low sample volume (0.25 L). This is somewhat reflected in the BH and SC eDNA sampling results across the country, where SC eDNA detections dropped off from ~1% in 2013/2014 to <0.05% in 2015/2016 (Table B-7).

6.2.2 Making accurate interpretations from eDNA sampling data

This dissertation also suggests that there may be a low-level of background eDNA that is detectable in the absence of live AC organisms. I propose that federal agencies interested in applying eDNA

sampling for invasive Asian carp detection perform regular and consistent sampling across the gradient of AC invasion in the Illinois river, starting from locations with heavy AC presence and into the CAWS and in Lake Michigan. The purpose would be to establish a background level of eDNA concentrations where Asian carp are absent and the eDNA concentration where Asian carp are present. Thus, if the concentration gradient along the Illinois River changes spatially, e.g. the concentrations in the Brandon Road pool are now similar to the concentration distributions in the Marseilles pool, then this could also indicate new movement of AC.

I also recommend that the eDNA sampling results should be reported as copy numbers rather than just as positive/negative. Chapter 4 showed that analysis using results with copy numbers allows for concentration estimates with lower variability compared to concentration estimates from using positive/negative results. Under the current protocol, a detection with 10000 copies and a sample with 100 copies is reported identically as one positive sample. However, it is expected that a positive sample with 10,000 copies will be more meaningful (i.e., indicate AC presence) compared to the sample with 100 copies.

6.2.3 Implications for decision-making about AC prevention

The decision analysis in Chapter 5 suggests that there may be benefit to installing an additional barrier in the Brandon Road pool. Installing an additional barrier is shown in chapter 5 to be a cost-effective way at improving prevention, especially if there is a high likelihood of invasion and damage from AC. However, the model does show that increased risk aversion of decision-makers favors staying with the status quo, since an additional barrier is not fool-proof and may be a wasteful expenditure. The model also shows that hydrologic separation of the CAWS is not a viable solution under almost every assumption about invasion costs and probability, due to the

high costs and 25-year implementation time. If invasion is imminent, then implementing hydrologic separation will be ineffective and wasteful. However, if the probability of invasion is very low (thus making hydrologic separation potentially effective), then adding another electric barrier will also be highly effective at much lower cost.

6.2.4 Implications about AC monitoring in the CAWS and Illinois River

The study in Chapter 5 shows that monitoring methods provide value for making decisions about AC prevention only if the specificity of the monitoring method is 1 (i.e., that the false positive rate is ~0%). This suggests that decision-makers should prefer monitoring methods with very high specificity but low sensitivity (e.g., fishing and netting) compared to tools with high sensitivity but less than perfect specificity (e.g., eDNA sampling). This has already been seen in recent years where invasive AC managers have been relying increasingly on conventional surveillance tools, like electrofishing, traps and netting, and have significantly reduced the number of eDNA sampling events in the CAWS and the Illinois River.[116] However, the goal of developing an effective invasive species monitoring strategy should seek to attain high sensitivity and high specificity. I suggest that the models developed in this dissertation could be used to minimize the risk of false positives and to utilize the high sensitivity of eDNA sampling more effectively.

6.3 Future Work

This dissertation assesses the false negatives of eDNA sampling stemming from the sampling protocol and the false positives from detecting target eDNA in the absence of the target organism.[82] This work assumed that the probability of falsely detecting non-target DNA in the PCR or qPCR analysis is zero when the target DNA is not present. The modeling framework proposed in Chapter 3 could be extended to include a probability of falsely amplifying non-target DNA. Studies show

that one way to mitigate this false positive rate would be to analyze multiple replicates per sample and require multiple positive replicates before determining a positive sample.[130–132] The current protocol for BH and SC eDNA sampling is vulnerable to these method false positives, since only one replicate out of eight needs to be positive for each marker to determine a positive sample. I recommend future studies on the probability of these false detections in the laboratory and statistical models to help inform changes to the protocol to minimize this risk, e.g. requiring more than one positive replicate.

Also, throughout this work, I modeled the CAWS as a single location. However, in reality, the CAWS is a series of canals with complex hydrology.[60] Future work should integrate my updated statistical model of eDNA sampling with hydrodynamic models of the CAWS, invasive fish movement, alternative shedding rates, eDNA shedding rates and degradation rates, similar to the work performed by Schultz et al.[108] This would allow for a stronger assessment of the hydrologic influences on eDNA concentration and the relationship between eDNA detection and potential fish presence in the CAWS.

In chapter 5, I developed a decision tree model considering only a one-time decision about prevention and a one-time monitoring event. It is more realistic that decision-makers will be making yearly decisions and performing yearly monitoring events. The decision model could be extended to consider a stochastic decision-making process.[129] Also, the probabilities of invasion, preventative ability, and costs are considered as deterministic values, rather than as distributions. Future work should characterize these values as distributions in order to incorporate more of the uncertainty of these variables into the decision model.

This model also assumes that AC can only invade the Great Lakes if the preventative measure fails. However, there are other pathways, e.g. human-mediated transport, which could cause AC invasion though the preventative method is successful at preventing AC passage through the CAWS into the Great Lakes. These pathways are very unlikely, due to the illegal nature of transporting live AC, but are possible and may make costly prevention measures less desirable due to its ineffectiveness at stopping these pathways. Also, in my model I assumed that the decision-makers have no ability to control and manage the expected damages once the invasion occurs. Future work on the model could include the choice of control strategies in addition to prevention strategies for minimizing the risk of invasion and the environmental damages.[129,133] Finally, the decision model only considers the objective of minimizing preventative costs and expected invasion costs. However, these prevention strategies also have direct and indirect impacts on the environment (e.g., water quality) and other industries (e.g., cargo shipping). Future models could include the balancing of these competing objectives and differing stakeholder values.

References

(1) Ficetola, G. F.; Miaud, C.; Pompanon, F.; Taberlet, P. Species detection using environmental DNA from water samples. *Biol. Lett.* **2008**, *4* (4), 423–425.

(2) Thomsen, P. F.; Kielgast, J.; Iversen, L. L.; Møller, P. R.; Rasmussen, M.; Willerslev, E. Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLoS ONE* **2012**, *7* (8), e41732.

(3) Jerde, C. L.; Chadderton, W. L.; Mahon, A. R.; Renshaw, M. A.; Corush, J.; Budny, M. L.; Mysorekar, S.; Lodge, D. M. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Can. J. Fish. Aquat. Sci.* **2013**, *70* (4), 522–526.

(4) Baldigo, B. P.; Sporn, L. A.; George, S. D.; Ball, J. A. Efficacy of Environmental DNA to Detect and Quantify Brook Trout Populations in Headwater Streams of the Adirondack Mountains, New York. *Trans. Am. Fish. Soc.* **2017**, *146* (1), 99–111.

(5) Dejean, T.; Valentini, A.; Miquel, C.; Taberlet, P.; Bellemain, E.; Miaud, C. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog Lithobates catesbeianus. *J. Appl. Ecol.* **2012**, *49* (4), 953–959.

(6) Pilliod, D. S.; Goldberg, C. S.; Arkle, R. S.; Waits, L. P. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Can. J. Fish. Aquat. Sci.* **2013**, *70* (8), 1123–1130.

(7) Smart, A. S.; Tingley, R.; Weeks, A. R.; van Rooyen, A. R.; McCarthy, M. A. Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. *Ecol. Appl.* **2015**, *25* (7), 1944–1952.

(8) Wilcox, T. M.; McKelvey, K. S.; Young, M. K.; Sepulveda, A. J.; Shepard, B. B.; Jane, S. F.; Whiteley, A. R.; Lowe, W. H.; Schwartz, M. K. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. *Biol. Conserv.* **2016**, *194*, 209–216.

(9) Schmelzle, M. C.; Kinziger, A. P. Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Mol. Ecol. Resour.* **2016**, *16* (4), 895–908.

(10) Merkes, C. M.; McCalla, S. G.; Jensen, N. R.; Gaikowski, M. P.; Amberg, J. J. Persistence of DNA in Carcasses, Slime and Avian Feces May Affect Interpretation of Environmental DNA Data. *PLoS ONE* **2014**, *9* (11), e113346.

(11) Radke, R. J.; Kahl, U. Effects of a filter-feeding fish [silver carp, Hypophthalmichthys molitrix (Val.)] on phyto-and zooplankton in a mesotrophic reservoir: results from an enclosure experiment. *Freshw. Biol.* **2002**, *47* (12), 2337–2344.

(12) Freeze, M.; Henderson, S. Distribution and Status of the Bighead Carp and Silver Carp in Arkansas. *North Am. J. Fish. Manag.* **1982**, *2* (2), 197–200.

(13) Irons, K. S.; Sass, G. G.; McClelland, M. A.; Stafford, J. D. Reduced condition factor of two native fish species coincident with invasion of non-native Asian carps in the Illinois River, U.S.A. Is this evidence for competition and reduced fitness? *J. Fish Biol.* **2007**, *71*, 258–273.

(14) Sampson, S. J.; Chick, J. H.; Pegg, M. A. Diet overlap among two Asian carp and three native fishes in backwater lakes on the Illinois and Mississippi rivers. *Biol. Invasions* **2009**, *11* (3), 483–496.

(15) Buck, E. H.; Upton, H. F.; Stern, C. V.; Nicols, J. E. *Asian carp and the Great Lakes region*; Congressional Research Service, 2014.

(16) 2016 Asian Carp Action Plan. Asian Carp Regional Coordinating Committee (ACRCC) 2016.

(17) Jerde, C. L.; Mahon, A. R.; Chadderton, W. L.; Lodge, D. M. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conserv. Lett.* **2011**, *4* (2), 150–157.

(18) Egan, D. A High-Tech Hunt for Asian Carp - In Hunt for Asian Carp, Scientists Find DNA - and Controversy. *Milwaukee-Wis. J. Sentin.* **2012**.

(19) Egan, D. A High-Tech Hunt for Asian Carp - Chicago River Becomes Battleground Test Lab. *Milwaukee-Wis. J. Sentin.* **2012**.

(20) MRWG. Monitoring and Response Plan for Asian Carp in the Upper Illinois River and Chicago Area Waterway System. Asian Carp Regional Coordinating Committee (ACRCC) May 2013.

(21) *Quality Assurance Project Plan eDNA Monitoring of Bighead and Silver Carps*; U.S. Fish and Wildlife Service, 2015.

(22) Wilson, C.; Wright, E.; Bronnenhuber, J.; MacDonald, F.; Belore, M.; Locke, B. Tracking ghosts: combined electrofishing and environmental DNA surveillance efforts for Asian carps in Ontario waters of Lake Erie. *Manag. Biol. Invasions* **2014**, *5* (3), 225–231.

(23) Mahon, A. R.; Jerde, C. L.; Galaska, M.; Bergner, J. L.; Chadderton, W. L.; Lodge, D. M.; Hunter, M. E.; Nico, L. G. Validation of eDNA Surveillance Sensitivity for Detection of Asian Carps in Controlled and Field Experiments. *PLoS ONE* **2013**, *8* (3), e58316.

(24) Goldberg, C. S.; Turner, C. R.; Deiner, K.; Klymus, K. E.; Thomsen, P. F.; Murphy, M. A.; Spear, S. F.; McKee, A.; Oyler-McCance, S. J.; Cornman, R. S.; et al. Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol. Evol.* **2016**.

(25) Roussel, J.-M.; Paillisson, J.-M.; Tréguier, A.; Petit, E. The downside of eDNA as a survey tool in water bodies. *J. Appl. Ecol.* **2015**, n/a-n/a.

(26) Deiner, K.; Walser, J.-C.; Mächler, E.; Altermatt, F. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biol. Conserv.* **2015**, *183*, 53–63.

(27) Eichmiller, J. J.; Miller, L. M.; Sorensen, P. W. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Mol. Ecol. Resour.* **2016**, *16* (1), 56–68.

(28) Minamoto, T.; Naka, T.; Moji, K.; Maruyama, A. Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. *Limnology* **2016**, *17* (1), 23–32.

(29) McKee, A. M.; Spear, S. F.; Pierson, T. W. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol. Conserv.* **2015**, *183*, 70–76.

(30) Amberg, J. J.; Grace McCalla, S.; Monroe, E.; Lance, R.; Baerwaldt, K.; Gaikowski, M. P. Improving efficiency and reliability of environmental DNA analysis for silver carp. *J. Gt. Lakes Res.* **2015**, *41* (2), 367–373.

(31) Deiner, K.; Walser, J.-C.; Mächler, E.; Altermatt, F. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biol. Conserv.* **2015**, *183*, 53–63.

(32) Takahara, T.; Minamoto, T.; Doi, H. Effects of sample processing on the detection rate of environmental DNA from the Common Carp (Cyprinus carpio). *Biol. Conserv.* **2014**.

(33) Turner, C. R.; Miller, D. J.; Coyne, K. J.; Corush, J. Improved Methods for Capture, Extraction, and Quantitative Assay of Environmental DNA from Asian Bigheaded Carp (Hypophthalmichthys spp.). *PLoS ONE* **2014**, *9* (12), e114329.

(34) Ficetola, G. F.; Miaud, C.; Pompanon, F.; Taberlet, P. Species detection using environmental DNA from water samples. *Biol. Lett.* **2008**, *4* (4), 423–425.

(35) Wilcox, T. M.; McKelvey, K. S.; Young, M. K.; Jane, S. F.; Lowe, W. H.; Whiteley, A. R.; Schwartz, M. K. Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity. *PLoS ONE* **2013**, *8* (3), e59520.

(36) Champlot, S.; Berthelot, C.; Pruvost, M.; Bennett, E. A.; Grange, T.; Geigl, E.-M. An Efficient Multistrategy DNA Decontamination Procedure of PCR Reagents for Hypersensitive PCR Applications. *PLoS ONE* **2010**, *5* (9), e13042.

(37) Donald, C. E.; Qureshi, F.; Burns, M. J.; Holden, M. J.; Blasic, J. R.; Woolford, A. J. An inter-platform repeatability study investigating real-time amplification of plasmid DNA. *BMC Biotechnol.* **2005**, *5* (1), 15.

(38) Barnes, M. A.; Turner, C. R. The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.* **2015**.

(39) Thomsen, P. F.; Kielgast, J.; Iversen, L. L.; Wiuf, C.; Rasmussen, M.; Gilbert, M. T. P.; Orlando, L.; Willerslev, E. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* **2012**, *21* (11), 2565–2573.

(40) Klymus, K. E.; Richter, C. A.; Chapman, D. C.; Paukert, C. Quantification of eDNA shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp Hypophthalmichthys molitrix. *Biol. Conserv.* **2014**.

(41) Maruyama, A.; Nakamura, K.; Yamanaka, H.; Kondoh, M.; Minamoto, T. The release rate of environmental DNA from juvenile and adult fish. *PloS One* **2014**, *9* (12), e114639.

(42) Takahara, T.; Minamoto, T.; Yamanaka, H.; Doi, H.; Kawabata, Z. 'ichiro. Estimation of Fish Biomass Using Environmental DNA. *PLoS ONE* **2012**, *7* (4), e35868.

(43) Doi, H.; Uchii, K.; Takahara, T.; Matsuhashi, S.; Yamanaka, H.; Minamoto, T. Use of Droplet Digital PCR for Estimation of Fish Abundance and Biomass in Environmental DNA Surveys. *PLOS ONE* **2015**, *10* (3), e0122763.

(44) Pilliod, D. S.; Goldberg, C. S.; Arkle, R. S.; Waits, L. P. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Can. J. Fish. Aquat. Sci.* **2013**, *70* (8), 1123–1130.

(45) Lacoursière-Roussel, A.; Rosabal, M.; Bernatchez, L. Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Mol. Ecol. Resour.* **2016**, n/a–n/a.

(46) Lacoursière-Roussel, A.; Côté, G.; Leclerc, V.; Bernatchez, L. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *J. Appl. Ecol.* **2015**, n/a–n/a.

(47) Spear, S. F.; Groves, J. D.; Williams, L. A.; Waits, L. P. Using environmental DNA methods to improve detectability in a hellbender (Cryptobranchus alleganiensis) monitoring program. *Biol. Conserv.* **2014**.

(48) Iversen, L. L.; Kielgast, J.; Sand-Jensen, K. Monitoring of animal abundance by environmental DNA — An increasingly obscure perspective: A reply to Klymus et al., 2015. *Biol. Conserv.* **2015**.

(49) Deiner, K.; Altermatt, F. Transport distance of invertebrate environmental DNA in a natural river. *PloS One* **2014**, *9* (2), e88786.

(50) Jane, S. F.; Wilcox, T. M.; McKelvey, K. S.; Young, M. K.; Schwartz, M. K.; Lowe, W. H.; Letcher, B. H.; Whiteley, A. R. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol. Ecol. Resour.* **2014**, n/a-n/a.

(51) Jerde, C. L.; Olds, B. P.; Shogren, A. J.; Andruszkiewicz, E. A.; Mahon, A. R.; Bolster, D.; Tank, J. L. Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environ. Sci. Technol.* **2016**, *50* (16), 8770–8779.

(52) Turner, C. R.; Uy, K. L.; Everhart, R. C. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biol. Conserv.* **2014**.

(53) Barnes, M. A.; Turner, C. R.; Jerde, C. L.; Renshaw, M. A.; Chadderton, W. L.; Lodge, D. M. Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environ. Sci. Technol.* **2014**, *48* (3), 1819–1827.

(54) Dejean, T.; Valentini, A.; Duparc, A.; Pellier-Cuit, S.; Pompanon, F.; Taberlet, P.; Miaud, C. Persistence of Environmental DNA in Freshwater Ecosystems. *PLoS ONE* **2011**, *6* (8), e23398.

(55) Pilliod, D. S.; Goldberg, C. S.; Arkle, R. S.; Waits, L. P. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Mol. Ecol. Resour.* **2014**, *14* (1), 109–116.

(56) Strickler, K. M.; Fremier, A. K.; Goldberg, C. S. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biol. Conserv.* **2015**, *183*, 85–92.

(57) Eichmiller, J. J.; Best, S. E.; Sorensen, P. W. Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water. *Environ. Sci. Technol.* **2016**.

(58) Piggott, M. P. Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecol. Evol.* **2016**, n/a-n/a.

(59) Rasmussen, J. L.; Regier, H. A.; Sparks, R. E.; Taylor, W. W. Dividing the waters: The case for hydrologic separation of the North American Great Lakes and Mississippi River Basins. *J. Gt. Lakes Res.* **2011**, *37* (3), 588–592.

(60) Duncker, J. J.; Johnson, K. K. *Hydrology of and Current Monitoring Issues for the Chicago Area Waterway System, Northeastern Illinois*; U.S. Geological Survey Scientific Investigations Report 2015–5115; 2015; p 48.

(61) Jerde, C. L.; Mahon, A. R.; Chadderton, W. L.; Lodge, D. M. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conserv. Lett.* **2011**, *4* (2), 150–157.

(62) Turner, C. R.; Uy, K. L.; Everhart, R. C. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biol. Conserv.* **2014**.

(63) Jane, S. F.; Wilcox, T. M.; McKelvey, K. S.; Young, M. K.; Schwartz, M. K.; Lowe, W. H.; Letcher, B. H.; Whiteley, A. R. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol. Ecol. Resour.* **2015**, *15* (1), 216–227.

(64) Pilliod, D. S.; Goldberg, C. S.; Arkle, R. S.; Waits, L. P. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Mol. Ecol. Resour.* **2014**, *14* (1), 109–116.

(65) Barnes, M. A.; Turner, C. R.; Jerde, C. L.; Renshaw, M. A.; Chadderton, W. L.; Lodge, D. M. Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environ. Sci. Technol.* **2014**, *48* (3), 1819–1827.

(66) Strickler, K. M.; Fremier, A. K.; Goldberg, C. S. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biol. Conserv.* **2015**, *183*, 85–92.

(67) Schmidt, B. R.; Kéry, M.; Ursenbacher, S.; Hyman, O. J.; Collins, J. P. Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods Ecol. Evol.* **2013**, *4* (7), 646–653.

(68) Dougherty, M. M.; Larson, E. R.; Renshaw, M. A.; Gantz, C. A.; Egan, S. P.; Erickson, D. M.; Lodge, D. M. Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *J. Appl. Ecol.* **2016**, *53* (3), 722–732.

(69) Furlan, E. M.; Gleeson, D.; Hardy, C. M.; Duncan, R. P. A framework for estimating the sensitivity of eDNA surveys. *Mol. Ecol. Resour.* **2016**, *16* (3), 641–654.

(70) Gange, S. J.; Munoz, A.; Saez, M.; Alonso, J. Use of the Beta-Binomial Distribution to Model the Effect of Policy Changes on Appropriateness of Hospital Stays. *Appl. Stat.* **1996**, *45* (3), 371.

(71) Schultz, M. T.; Lance, R. F. Modeling the Sensitivity of Field Surveys for Detection of Environmental DNA (eDNA). *PLOS ONE* **2015**, *10* (10), e0141503.

(72) Erickson, R. A.; Rees, C. B.; Coulter, A. A.; Merkes, C. M.; McCalla, S. G.; Touzinsky, K. F.; Walleser, L.; Goforth, R. R.; Amberg, J. J. Detecting the movement and spawning activity of bigheaded carps with environmental DNA. *Mol. Ecol. Resour.* **2016**.

(73) Deters, J. E.; Chapman, D. C.; McElroy, B. Location and timing of Asian carp spawning in the Lower Missouri River. *Environ. Biol. Fishes* **2013**, *96* (5), 617–629.

(74) Czepiel, S. A. Maximum likelihood estimation of logistic regression models: theory and implementation. *Available Czep Netstatmlelr Pdf* **2002**.

(75) Welsh, A. H.; Cunningham, R. B.; Donnelly, C. F.; Linedenmayer, D. B. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecol. Model.* **1996**, No. 88, 297–308.

(76) Burnham, K. P. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol. Methods Res.* **2004**, *33* (2), 261–304.

(77) R Development Core Team. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.

(78) Environmental DNA Calibration Study (ECALS). Interim Technical Review Report. February 2013.

(79) Dejean, T.; Valentini, A.; Miquel, C.; Taberlet, P.; Bellemain, E.; Miaud, C. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog Lithobates catesbeianus. *J. Appl. Ecol.* **2012**, *49* (4), 953–959.

(80) Takahara, T.; Minamoto, T.; Doi, H. Using Environmental DNA to Estimate the Distribution of an Invasive Fish Species in Ponds. *PLoS ONE* **2013**, *8* (2), e56584.

(81) Simmons, M.; Tucker, A.; Chadderton, W. L.; Jerde, C. L.; Mahon, A. R.; Taylor, E. Active and passive environmental DNA surveillance of aquatic invasive species. *Can. J. Fish. Aquat. Sci.* **2016**, *73* (1), 76–83.

(82) Darling, J. A.; Mahon, A. R. From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environ. Res.* **2011**, *111* (7), 978–988.

(83) Rees, H. C.; Maddison, B. C.; Middleditch, D. J.; Patmore, J. R. M.; Gough, K. C. The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* **2014**, n/a-n/a.

(84) Schultz, M. T.; Lance, R. F. Modeling the Sensitivity of Field Surveys for Detection of Environmental DNA (eDNA). *PLOS ONE* **2015**, *10* (10), e0141503.

(85) Mächler, E.; Deiner, K.; Spahn, F.; Altermatt, F. Fishing in the Water: Effect of Sampled Water Volume on Environmental DNA-Based Detection of Macroinvertebrates. *Environ. Sci. Technol.* **2016**, *50* (1), 305–312.

(86) McKee, A. M.; Spear, S. F.; Pierson, T. W. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol. Conserv.* **2015**, *183*, 70–76.

(87) Farrington, H. L.; Edwards, C. E.; Guan, X.; Carr, M. R.; Baerwaldt, K.; Lance, R. F. Mitochondrial Genome Sequencing and Development of Genetic Markers for the Detection of DNA of Invasive Bighead and Silver Carp (Hypophthalmichthys nobilis and H. molitrix) in Environmental Water Samples from the United States. *PLOS ONE* **2015**, *10* (2), e0117803.

(88) Davy, C. M.; Kidd, A. G.; Wilson, C. C. Development and Validation of Environmental DNA (eDNA) Markers for Detection of Freshwater Turtles. *PloS One* **2015**, *10* (7), e0130965.

(89) Rees, H. C.; Maddison, B. C.; Middleditch, D. J.; Patmore, J. R. M.; Gough, K. C. REVIEW: The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* **2014**, *51* (5), 1450–1459.

(90) White, G. C.; Bennetts, R. E. Analysis of frequency count data using the negative binomial distribution. *Ecology* **1996**, *77* (8), 2549–2557.

(91) Melbourne, B. A.; Hastings, A. Extinction risk depends strongly on factors contributing to stochasticity. *Nature* **2008**, *454* (7200), 100–103.

(92) Hilbe, J. M. Modeling Count Data. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp 836–839.

(93) Quality Assurance Project Plan eDNA Monitoring of Bighead and Silver Carps. U.S. Fish and Wildlife Service 2015.

(94) Sanger, F.; Coulson, A. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Sel. Pap. Frederick Sanger Comment.* **1996**, *94*, 382.

(95) Lance, R. F.; Farrington, H. L.; Edwards, C. E.; Guan, X.; Carr, M. R.; Baerwaldt, K. *Numerous new mitogenomic sequences and multiple new environmental DNA markers for invasive bighead and silver carp (Hypophthalmichthys nobilis and H. molitrix) populations in North America*; 3624; 2014.

(96) Doi, H.; Takahara, T.; Minamoto, T.; Matsuhashi, S.; Uchii, K.; Yamanaka, H. Droplet Digital Polymerase Chain Reaction (PCR) Outperforms Real-Time PCR in the Detection of Environmental DNA from an Invasive Fish Species. *Environ. Sci. Technol.* **2015**, *49* (9), 5601–5608.

(97) Egan, S. P.; Barnes, M. A.; Hwang, C.-T.; Mahon, A. R.; Feder, J. L.; Ruggiero, S. T.; Tanner, C. E.; Lodge, D. M. Rapid Invasive Species Detection by Combining Environmental DNA with Light Transmission Spectroscopy: eDNA-LTS species detection. *Conserv. Lett.* **2013**, *6* (6), 402–409.

(98) Nathan, L. M.; Simmons, M.; Wegleitner, B. J.; Jerde, C. L.; Mahon, A. R. Quantifying Environmental DNA Signals for Aquatic Invasive Species Across Multiple Detection Platforms. *Environ. Sci. Technol.* **2014**, *48* (21), 12800–12806.

(99) Merkes, C. M.; McCalla, S. G.; Jensen, N. R.; Gaikowski, M. P.; Amberg, J. J. Persistence of DNA in Carcasses, Slime and Avian Feces May Affect Interpretation of Environmental DNA Data. *PLoS ONE* **2014**, *9* (11), e113346.

(100) Willoughby, J. R.; Wijayawardena, B. K.; Sundaram, M.; Swihart, R. K.; DeWoody, J. A. The importance of including imperfect detection models in eDNA experimental design. *Mol. Ecol. Resour.* **2016**.

(101) Hunter, M. E.; Oyler-McCance, S. J.; Dorazio, R. M.; Fike, J. A.; Smith, B. J.; Hunter, C. T.; Reed, R. N.; Hart, K. M. Environmental DNA (eDNA) Sampling Improves Occurrence and Detection Estimates of Invasive Burmese Pythons. *PLOS ONE* **2015**, *10* (4), e0121655.

(102) MacKenzie, D. I.; Nichols, J. D.; Lachman, G. B.; Droege, S.; Andrew Royle, J.; Langtimm, C. A. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **2002**, *83* (8), 2248–2255.

(103) Furlan, E. M.; Gleeson, D.; Hardy, C. M.; Duncan, R. P. A framework for estimating the sensitivity of eDNA surveys. *Mol. Ecol. Resour.* **2015**, n/a-n/a.

(104) Farrington, H. L.; Edwards, C. E.; Guan, X.; Carr, M. R.; Baerwaldt, K.; Lance, R. F. Mitochondrial Genome Sequencing and Development of Genetic Markers for the Detection of DNA of Invasive Bighead and Silver Carp (Hypophthalmichthys nobilis and H. molitrix) in Environmental Water Samples from the United States. *PLOS ONE* **2015**, *10* (2), e0117803.

(105) Tierney, L. Markov chains for exploring posterior distributions. *Ann. Stat.* **1994**, 1701–1728.

(106) Geman, S.; Geman, D. Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, No. 6, 721–741.

(107) Brooks, S. P.; Gelman, A. General Methods for Monitoring Convergence of Iterative Simulations. *J. Comput. Graph. Stat.* **1998**, *7* (4), 434–455.

(108) Schultz, M. T.; Cerco, C. F.; Skahill, B. E.; Lance, R. F.; Noel, M. R.; DiJoseph, P. K.; Smith, D. L.; Guilfoyle, M. P. *A Probabilistic Analysis of Environmental DNA Monitoring Results in the Chicago Area Waterway System*; Engineer Research and Development Center United States Army Corps of Engineers, 2014.

(109) Monitoring and Response Plan for Asian Carp in the Upper Illinois River and Chicago Area Waterway System. Asian Carp Regional Coordinating Committee (ACRCC) 2016.

(110) Egan, D. A High-Tech Hunt for Asian Carp - Fish Barrier vs. Carp DNA: What to Believe? *Milwaukee-Wis. J. Sentin.* **2012**.

(111) Vander Zanden, M. J.; Hansen, G. J. A.; Higgins, S. N.; Kornis, M. S. A pound of prevention, plus a pound of cure: Early detection and eradication of invasive species in the Laurentian Great Lakes. *J. Gt. Lakes Res.* **2010**, *36* (1), 199–205.

(112) Collins, S. F.; Butler, S. E.; Diana, M. J.; Wahl, D. H. Catch Rates and Cost Effectiveness of Entrapment Gears for Asian Carp: A Comparison of Pound Nets, Hoop Nets, and Fyke Nets in Backwater Lakes of the Illinois River. *North Am. J. Fish. Manag.* **2015**, *35* (6), 1219–1225.

(113) Clemen, R. T.; Reilly, T. *Making hard decisions with DecisionTools*; Cengage Learning, 2013.

(114) Pratt, J. W.; Zeckhauser, R. J. Proper Risk Aversion. *Econometrica* **1987**, *55* (1), 143.

(115) GLMRIS. Summary of the Great Lakes and Mississippi River Interbasin Study Report.

(116) MRWG. 2017 Asian Carp Action Plan.

(117) Wittmann, M. E.; Cooke, R. M.; Rothlisberger, J. D.; Lodge, D. M. Using Structured Expert Judgment to Assess Invasive Species Prevention: Asian Carp and the Mississippi—Great Lakes Hydrologic Connection. *Environ. Sci. Technol.* **2014**, *48* (4), 2150–2156.

(118) Parker, A. D.; Glover, D. C.; Finney, S. T.; Rogers, P. B.; Stewart, J. G.; Simmonds, R. L. Direct observations of fish incapacitation rates at a large electrical fish barrier in the Chicago Sanitary and Ship Canal. *J. Gt. Lakes Res.* **2015**, *41* (2), 396–404.

(119) Schwar, M. T. What Are the Odds? Factors influencing potential Asian carp migration past the CSSC fish barrier. *World Environ. Water Resour. Congr. 2012 Crossing Boundaries* **2012**, 1746–1752.

(120) Cupp, A. R.; Erickson, R. A.; Fredricks, K.; Swyers, N.; Hatton, T.; Amberg, J. Responses of invasive silver and bighead carp to a carbon dioxide barrier in outdoor ponds. *Can. J. Fish. Aquat. Sci.* **2016**.

(121) *Great Lakes and Mississippi Interbasin Study*; US Army Corps of Engineers.

(122) Cooke, S. L.; Hill, W. R. Can filter-feeding Asian carp invade the Laurentian Great Lakes? A bioenergetic modelling exercise. *Freshw. Biol.* **2010**, *55* (10), 2138–2152.

(123) Kocovsky, P. M.; Chapman, D. C.; McKenna, J. E. Thermal and hydrologic suitability of Lake Erie and its major tributaries for spawning of Asian carps. *J. Gt. Lakes Res.* **2012**, *38* (1), 159–166.

(124) Lauber, T. B.; Stedman, R. C.; Connelly, N. A.; Rudstam, L. G.; Ready, R. C.; Poe, G. L.; Bunnell, D. B.; Höök, T. O.; Koops, M. A.; Ludsin, S. A.; et al. Using Scenarios to Assess Possible Future Impacts of Invasive Species in the Laurentian Great Lakes. *North Am. J. Fish. Manag.* **2016**, *36* (6), 1292–1307.

(125) Zhang, H.; Rutherford, E. S.; Mason, D. M.; Breck, J. T.; Wittmann, M. E.; Cooke, R. M.; Lodge, D. M.; Rothlisberger, J. D.; Zhu, X.; Johnson, T. B. Forecasting the Impacts of Silver and Bighead Carp on the Lake Erie Food Web. *Trans. Am. Fish. Soc.* **2016**, *145* (1), 136–162.

(126) DeGrandchamp, K. L.; Garvey, J. E.; Colombo, R. E. Movement and Habitat Selection by Invasive Asian Carps in a Large River. *Trans. Am. Fish. Soc.* **2008**, *137* (1), 45–56.

(127) Chen, P.; Wiley, E. O.; Mcnyset, K. M. Ecological niche modeling as a predictive tool: silver and bighead carps in North America. *Biol. Invasions* **2006**, *9* (1), 43–51.

(128) Howard, R. A. Decision analysis: practice and promise. *Manag. Sci.* **1988**, *34* (6), 679–695.

(129) Finnoff, D.; Shogren, J. F.; Leung, B.; Lodge, D. Take a risk: Preferring prevention over control of biological invaders. *Ecol. Econ.* **2007**, *62* (2), 216–222.

(130) Chambert, T.; Miller, D. A.; Nichols, J. D. Modeling false positive detections in species occurrence data under different study designs. *Ecology* **2015**, *96* (2), 332–339.

(131) Ficetola, G. F.; Taberlet, P.; Coissac, E. How to limit false positives in environmental DNA and metabarcoding? *Mol. Ecol. Resour.* **2016**, *16* (3), 604–607.

(132) Lahoz-Monfort, J. J.; Guillera-Arroita, G.; Tingley, R. Statistical approaches to account for false-positive errors in environmental DNA samples. *Mol. Ecol. Resour.* **2016**, *16* (3), 673–685.

(133) Leung, B.; Lodge, D. M.; Finnoff, D.; Shogren, J. F.; Lewis, M. A.; Lamberti, G. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. *Proc. R. Soc. Lond. B Biol. Sci.* **2002**, *269* (1508), 2407–2413.

Appendix A. Supporting Information for Chapter 2

A.1 R Code for Analysis

This is the code used to run the binomial and beta-binomial regression modeling analysis shown in Chapter 2. It also includes the code used to create the figures.

```
################################################
#RCode for The Effect of Streamflow Conditions on
#Silver Carp eDNA Detection in the CAWS
################################################

#install and load necessary packages
install.packages(c("gamlss","ggplot2","usdm"))
library(gamlss)
library(ggplot2)
library(usdm)

#load data table
df<-read.table("RData.csv",header=TRUE,sep=",")
n<-nrow(df)

#1. test covariates for colinearity
df_var<-subset(df,select=-c(1:4,12)) #remove the detection data
vifstep(df_var,th=10)

#2. Binomial regression will all covariates
df_test<-subset(df,select=-c(3:4,15))
#Full Model
reg_b<-gamlss(cbind(SC_Y,SC_N-SC_Y)~.,data=df_test,family = BI)
summary(reg_b)

reg_qb<-gamlss(cbind(SC_Y,SC_N-SC_Y)~.,data=df_test,family = BI)
summary(reg_qb)

#Final Model
reg_b2<-gamlss(cbind(SC_Y,SC_N-SC_Y)~sma_30_vol_MG+precip+Temp+pH+Chloroph,data=df_test,family =
BI)
summary(reg_b2)

#3. Beta-Binomial regression will all covariates
#Full Model
reg_bb<-gamlss(cbind(SC_Y,SC_N-SC_Y)~.,data=df_test,family = BB)
summary(reg_bb)

#Final Model
reg_bb2<-gamlss(cbind(SC_Y,SC_N-SC_Y)~sma_30_vol_MG,data=df_test,family = BB)
summary(reg_bb2)

#3. Make Figure 4
```

```
attach(df)
Year<-factor(df$Year)
#make data frame for plotting
df_plot<-data.frame(SC_Y=SC_Y,SC_N=SC_N,sma_30_vol_MG=sma_30_vol_MG,Year=Year)
detach(df)

#regress the final BB model
reg = gamlss(cbind(SC_Y,SC_N-SC_Y)~sma_30_vol_MG,sigma.formula=~1,data=df_plot,family =
BB,trace=FALSE)

#predict the mean and 95% confidence intervals using BB model
d_p<-data.frame("sma_30_vol_MG"=seq(0,5,.01))
mu_p<-predict(reg,what="mu",type="response",newdata=d_p)
sigma_p<-predict(reg,newdata=d_p,what="sigma",type="response")
d_p$mu_p<-mu_p
d_p$sigma_p<-sigma_p #not variance
var_p <- mu_p*(1-mu_p)*(sigma_p/(sigma_p+1))

#get alpha/beta of beta distribution
estBetaParams <- function(mu, var) {
  alpha <- ((1 - mu) / var - 1 / mu) * mu ^ 2
  beta <- alpha * (1 / mu - 1)
  return(params = list(alpha = alpha, beta = beta))
}
params = estBetaParams(mu_p,var_p)
d_p$alpha_p<-params$alpha
d_p$beta_p<-params$beta

#calculate the 95% CI at each point
d_p$lower<-qbeta(.025,params$alpha,params$beta)
d_p$upper<-qbeta(.975,params$alpha,params$beta)

#plot
ggplot(data=df_plot)+
  #plot the points
  geom_point(aes(x=sma_30_vol_MG,y=SC_Y/SC_N,size = SC_N,shape=Year)) +
  scale_shape_manual(values=c(0,1,17,8))+
  #plot the fitted line & 95% CI
  geom_line(data=d_p,aes(x=sma_30_vol_MG,y=mu_p),size = 1.5) +
  geom_line(data=d_p,aes(x=sma_30_vol_MG,y=lower),linetype = "dashed",size = 1) +
  geom_line(data=d_p,aes(x=sma_30_vol_MG,y=upper),linetype = "dashed",size = 1) +

  #eliminates background, gridlines, and chart border
  theme_bw()+
  theme(
    plot.background = element_blank()
    ,panel.grid.major = element_blank()
    ,panel.grid.minor = element_blank()) +

  #draws x and y axis line
  theme(axis.line.x = element_line(colour="black"),
        axis.line.y = element_line(colour="black")) +
  theme(text=element_text(size=18,colour="black"),
        axis.text = element_text(colour="black"),
```

```
    legend.key=element_blank())+
scale_y_continuous(limits = c(0,1),breaks=c(0,.25,.5,.75,1),expand=c(0,.02))+
scale_x_continuous(limits=c(0,4),expand=c(.02,0))+
ylab("Fraction of Positive SC eDNA Samples\n") +
xlab("\n30-day average of daily reverse flow volume [million gallons]") +
labs(shape = "Year",size = "# of samples")
```

Table A-1. The number of positive Silver Carp and Bighead Carp eDNA samples for all sampling events from 2009-2012 in the CAWS upstream of the electric barrier. The date and location of each sampling event is also listed. Starting July 2010, an interagency group led by the US Army Corps of Engineers (USACE) took over eDNA sampling analysis from University of Notre Dame (UND)

Date of Sample	Laboratory	Location	Silver Carp eDNA		Bighead Carp eDNA	
			Total # of samples	# of positive samples	Total # of samples	# of positive samples
8/3/2009	UND	CR5	25	0	25	0
9/10/2009	UND	CR1	13	0	13	0
9/10/2009	UND	CR2	7	0	7	0
9/23/2009	UND	CRA	14	0	14	0
9/23/2009	UND	CRC	27	0	27	0
9/23/2009	UND	CRD	44	1	44	26
9/10/2009	UND	CRM	73	0	73	0
10/22/2009	UND	CR1	7	0	7	0
10/1/2009	UND	CR3	3	1	3	0
10/1/2009	UND	CR4	38	0	38	0
10/1/2009	UND	CR5	28	1	28	0
10/1/2009	UND	CRE	21	0	21	3
10/22/2009	UND	NSC	45	5	45	0
11/24/2009	UND	CLK	3	0	3	0
11/24/2009	UND	CR5	1	0	1	0
11/24/2009	UND	CRA	11	0	11	0
11/24/2009	UND	CRB	3	0	3	0
11/24/2009	UND	CRC	13	0	13	0
11/24/2009	UND	CRD	34	0	34	2
11/24/2009	UND	CRE	40	2	40	1
11/24/2009	UND	LKC	3	0	3	0
12/8/2009	UND	CRA	46	4	46	0
12/8/2009	UND	CRC	5	0	5	0
12/8/2009	UND	CRD	33	1	33	0
12/1/2009	UND	CRM	11	0	11	0
12/8/2009	UND	LKC	14	0	14	0
3/30/2010	UND	CRA	39	0	39	0
3/30/2010	UND	CRB	2	0	2	0
3/30/2010	UND	CRC	7	0	7	0

3/30/2010	UND	CRD	43	1	43	0
3/30/2010	UND	LKC	14	0	14	0
4/20/2010	UND	CR1	20	0	20	0
4/15/2010	UND	CRC	6	0	6	0
4/15/2010	UND	CRD	61	0	61	0
4/20/2010	UND	NSC	67	1	67	0
5/27/2010	UND	BHR	13	1	13	0
5/27/2010	UND	CR2	45	1	45	0
5/27/2010	UND	CR3	34	4	34	0
5/20/2010	UND	CRC	9	0	9	0
5/20/2010	UND	CRD	35	0	35	0
5/27/2010	UND	CRM	20	1	20	0
5/27/2010	UND	MXZ	3	1	3	0
5/12/2010	UND	NSC	58	0	58	0
7/22/2010	USACE	CLK	24	0	24	0
7/20/2010	USACE	CRA	85	0	85	0
7/20/2010	USACE	CRB	5	0	5	0
7/22/2010	USACE	CRB	3	0	3	0
7/20/2010	USACE	CRC	8	0	8	0
7/22/2010	USACE	LKC	68	0	68	0
10/13/2010	USACE	CR5	54	4	54	0
11/2/2010	USACE	BHR	1	0	1	0
11/15/2010	USACE	CR1	4	0	4	0
11/2/2010	USACE	CR2	81	0	81	1
11/30/2010	USACE	CR5	45	1	45	0
11/8/2010	USACE	CRC	13	0	13	0
11/8/2010	USACE	CRD	96	1	96	0
11/8/2010	USACE	CRE	1	0	1	0
11/2/2010	USACE	CRM	27	0	27	0
11/2/2010	USACE	MXZ	5	0	5	0
11/15/2010	USACE	NSC	110	1	110	1
12/7/2010	USACE	CR4	29	0	29	0
12/7/2010	USACE	CR5	47	0	47	2
5/16/2011	USACE	CR1	3	0	3	0
5/10/2011	USACE	CR2	79	1	79	0
5/10/2011	USACE	CRM	24	0	24	0
5/10/2011	USACE	MXZ	11	0	11	0
5/16/2011	USACE	NSC	111	0	111	0
6/23/2011	USACE	BHR	2	0	2	0
6/15/2011	USACE	CLK	34	4	34	0
6/27/2011	USACE	CR1	9	0	9	0
6/23/2011	USACE	CR2	79	0	79	0
6/15/2011	USACE	CRA	3	0	3	0

6/15/2011	USACE	CRB	12	2	12	0
6/15/2011	USACE	CRC	51	0	51	0
6/23/2011	USACE	CRM	24	0	24	0
6/15/2011	USACE	LKC	15	1	15	0
6/23/2011	USACE	MXZ	9	0	9	0
6/27/2011	USACE	NSC	105	1	105	0
7/12/2011	USACE	CLK	11	0	11	0
7/19/2011	USACE	CLK	11	0	11	0
7/19/2011	USACE	CRA	2	0	2	0
7/12/2011	USACE	CRB	5	0	5	0
7/19/2011	USACE	CRB	4	0	4	0
7/12/2011	USACE	CRC	13	0	13	0
7/19/2011	USACE	CRC	15	0	15	0
7/12/2011	USACE	CRD	41	0	41	0
7/19/2011	USACE	CRD	50	1	50	0
7/12/2011	USACE	LKC	32	2	32	0
7/19/2011	USACE	LKC	32	1	32	0
8/17/2011	USACE	BHR	2	0	2	0
8/1/2011	USACE	CLK	13	0	13	0
8/30/2011	USACE	CLK	14	0	14	0
8/22/2011	USACE	CR1	6	0	6	0
8/17/2011	USACE	CR2	73	1	73	0
8/17/2011	USACE	CR3	7	0	7	0
8/1/2011	USACE	CRB	4	0	4	0
8/30/2011	USACE	CRB	4	0	4	0
8/1/2011	USACE	CRC	8	0	8	0
8/30/2011	USACE	CRC	14	0	14	0
8/30/2011	USACE	CRD	51	0	51	0
8/17/2011	USACE	CRM	23	0	23	0
8/1/2011	USACE	LKC	32	0	32	0
8/30/2011	USACE	LKC	31	1	31	0
8/17/2011	USACE	MXZ	9	0	9	0
8/22/2011	USACE	NSC	108	0	108	0
9/19/2011	USACE	CR1	8	0	8	0
9/13/2011	USACE	CR2	79	0	79	0
9/6/2011	USACE	CR5	57	0	57	0
9/13/2011	USACE	CRM	24	0	24	0
9/13/2011	USACE	MXZ	8	0	8	0
9/19/2011	USACE	NSC	106	2	106	0
10/11/2011	USACE	CLK	14	0	14	0
10/27/2011	USACE	CLK	14	0	14	0
10/25/2011	USACE	CR1	3	0	3	0
10/18/2011	USACE	CR2	79	0	79	0

10/25/2011	USACE	CR2	65	0	65	0
10/25/2011	USACE	CR3	3	0	3	0
10/26/2011	USACE	CR4	124	0	124	0
10/26/2011	USACE	CR5	104	2	104	0
10/27/2011	USACE	CR5	1	0	1	0
10/11/2011	USACE	CRB	13	0	13	0
10/27/2011	USACE	CRB	5	0	5	0
10/11/2011	USACE	CRC	5	0	5	0
10/27/2011	USACE	CRC	15	0	15	0
10/11/2011	USACE	CRD	51	0	51	0
10/27/2011	USACE	CRD	51	1	51	0
10/27/2011	USACE	CRE	113	10	113	0
10/18/2011	USACE	CRM	24	1	24	0
10/25/2011	USACE	CRM	19	2	19	0
10/11/2011	USACE	LKC	30	0	30	0
10/27/2011	USACE	LKC	31	0	31	0
10/18/2011	USACE	MXZ	11	0	11	0
10/25/2011	USACE	MXZ	8	0	8	0
10/25/2011	USACE	NSC	111	1	111	0
5/22/2012	USACE	CLK	16	4	16	0
5/22/2012	USACE	CRB	3	2	3	0
5/22/2012	USACE	CRC	14	3	14	0
5/22/2012	USACE	CRD	50	3	50	0
5/22/2012	USACE	LKC	30	5	30	0
6/11/2012	USACE	CLK	14	2	14	0
6/25/2012	USACE	CLK	15	1	15	0
6/11/2012	USACE	CR1	4	0	4	0
6/11/2012	USACE	CRB	5	0	5	0
6/25/2012	USACE	CRB	4	0	4	0
6/11/2012	USACE	CRC	7	0	7	0
6/25/2012	USACE	CRC	13	2	13	0
6/25/2012	USACE	CRD	52	0	52	0
6/11/2012	USACE	LKC	16	1	16	0
6/25/2012	USACE	LKC	30	4	30	0
6/11/2012	USACE	NSC	53	1	53	0
7/11/2012	USACE	CLK	6	0	6	0
7/24/2012	USACE	CLK	13	1	13	0
7/10/2012	USACE	CR1	2	0	2	0
7/10/2012	USACE	CR2	35	5	35	0
7/24/2012	USACE	CRB	4	0	4	0
7/11/2012	USACE	CRC	7	0	7	0
7/24/2012	USACE	CRC	8	0	8	0
7/10/2012	USACE	CRM	15	0	15	0

7/11/2012	USACE	LKC	30	2	30	0
7/24/2012	USACE	LKC	31	2	31	0
7/10/2012	USACE	MXZ	7	1	7	0
7/10/2012	USACE	NSC	55	3	55	0
8/20/2012	USACE	CLK	14	0	14	0
8/6/2012	USACE	CR2	35	0	35	0
8/20/2012	USACE	CRB	3	0	3	0
8/20/2012	USACE	CRC	14	2	14	0
8/20/2012	USACE	CRD	51	0	51	0
8/6/2012	USACE	CRM	15	0	15	0
8/20/2012	USACE	LKC	32	6	32	0
8/6/2012	USACE	MXZ	7	0	7	0
9/17/2012	USACE	CLK	13	2	13	0
9/11/2012	USACE	CR1	4	2	4	0
9/11/2012	USACE	CR2	35	5	35	0
9/17/2012	USACE	CRB	4	0	4	0
9/17/2012	USACE	CRC	13	3	13	0
9/17/2012	USACE	CRD	51	3	51	0
9/11/2012	USACE	CRM	15	12	15	0
9/17/2012	USACE	LKC	33	8	33	0
9/11/2012	USACE	MXZ	7	0	7	0
9/11/2012	USACE	NSC	53	11	53	0
10/22/2012	USACE	CLK	14	9	14	1
10/2/2012	USACE	CR1	4	1	4	0
10/15/2012	USACE	CR1	4	0	4	0
10/2/2012	USACE	CR2	31	4	35	0
10/22/2012	USACE	CRB	4	3	4	1
10/22/2012	USACE	CRC	13	6	13	0
10/22/2012	USACE	CRD	42	5	42	0
10/2/2012	USACE	CRM*	-	-	15	0
10/22/2012	USACE	LKC	12	11	12	2
10/2/2012	USACE	MXZ	7	3	7	0
10/2/2012	USACE	NSC	53	7	53	0
10/15/2012	USACE	NSC	53	8	53	0

Table A-2. The p-values of the β coefficient of the reverse flow volume covariate when fitting logit(p) in a beta-binomial model and the AIC of the model fit for a range of different time frames and types of moving averages of the reverse flow volume covariate: (a) 15, 30, 60 and 90-day; (b) simple and exponential. The exponential moving average is a weighted average using an alpha value of 2/(# of days+1). The best fitting covariate (30-day moving average) in terms of AIC is bolded.

Variable (different averages of daily reverse flow volume)	p-value of β	AIC
30-day moving average	**2.9E-13**	**484**
60-day moving average	2.8E-10	490
90-day exponential moving average	7.7E-10	490
60-day exponential moving average	3.1E-10	491
30-day exponential moving average	6.3E-09	498
90-day moving average	4.8E-07	500
15-day moving average	3.3E-06	509
15-day exponential moving average	1.5E-05	510

Table A-3. The goodness of fit in terms of AIC for four different model types (Binomial, Beta-Binomial, Zero-Inflated Binomial, Zero-Inflated Beta-Binomial) using RevVol (30-day moving average of reverse flow volume) as the only covariate. The best fitting model (Beta-Binomial) in terms of AIC is bolded.

Model (with just RevVol as the variable)	AIC
Beta-Binomial	**484**
Zero-Inflated Beta-Binomial	486
Zero-Inflated Binomial	590
Binomial	630

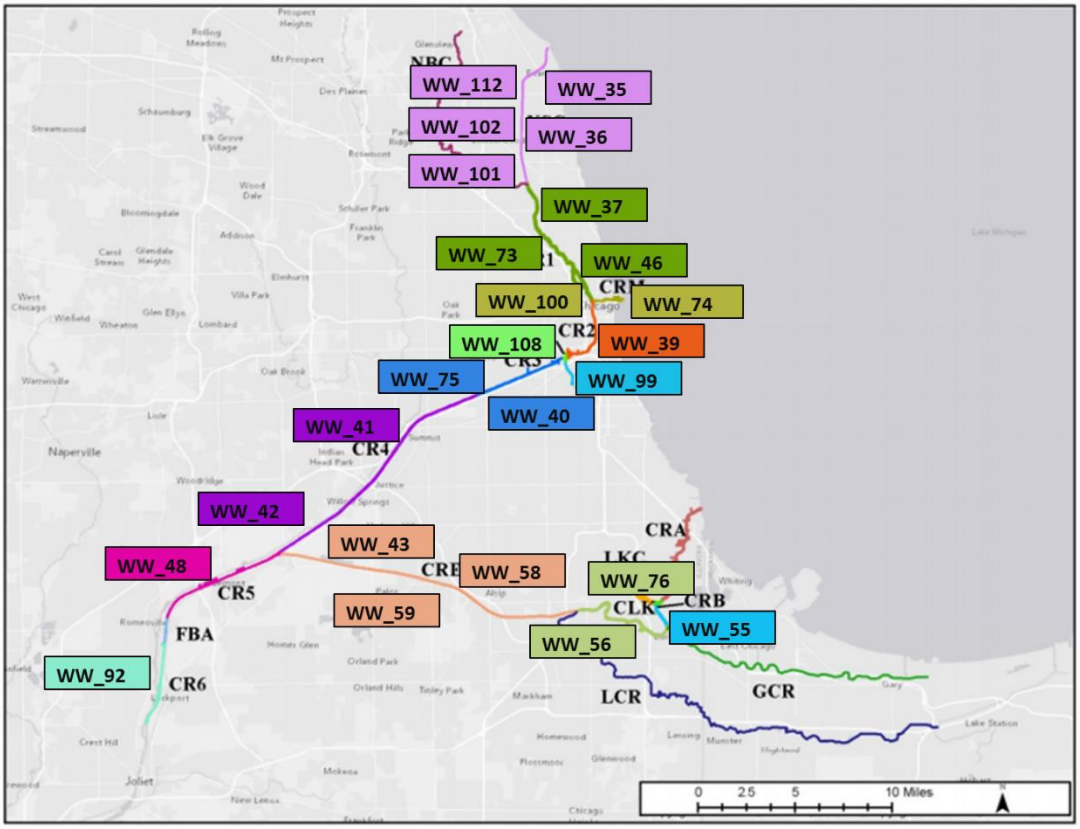

Figure A-1 The Chicago Area Waterway System (CAWS) connecting Lake Michigan to the Des Plaines River (a tributary of the Illinois River). The names of Chicago's Metropolitan Water Reclamation District (MWRD) water quality monitoring locations are shown in the boxes and are colored to match their corresponding eDNA monitoring location in the stream. The figure is adapted from [121].

Appendix B. Supporting Information for Chapter 3

B.1 Fitting a distribution for the number of filters used per sample

In 2014, USFWS' Whitney Genetics Lab (WGL) processed 6144 water samples using filtration. The number of filters used per sample was carefully documented. I use a zero-truncated or positive Poisson random variable to model the number of filters used per sample. The positive Poisson distribution is the Poisson distribution, given that the value of the random variable must be greater than zero. This distribution was used, because the number of filters used per sample must be at least one. The probability mass function of the positive Poisson distribution ($g(k;\lambda)$) is derived from a standard Poisson distribution ($f(k;\lambda)$), as shown here:

$$g(k;\lambda) = P(X = k|X > 0) = \frac{f(k;\lambda)}{1-f(0;\lambda)} = \frac{\lambda^k e^{-\lambda}}{k!(1-e^{-\lambda})} \quad \text{(B-1)}$$

The lambda parameter or mean of the distribution is 3.4 filters per sample, which is very close to the rule of thumb stated in the QAPP of using 3 filters per sample.

Then, using this distribution and the protocol that if 8 or fewer filters were used, then each of the filters were eluted into 200 µL and if 9 or more filters were used, then each of the filters were eluted into 100 µL, I can simulate the distribution of elution volume per sample for the filtration process. This distribution is shown in Figure A1. The expected value of the elution volume under these assumptions for filtration is 680 µL with a standard deviation of 340 µL. This is larger than the elution volume for the centrifugation process (began in 2015), which is always 200 µL for each sample and the elution volume for the filtration process used by USACE (from 2009-2012).

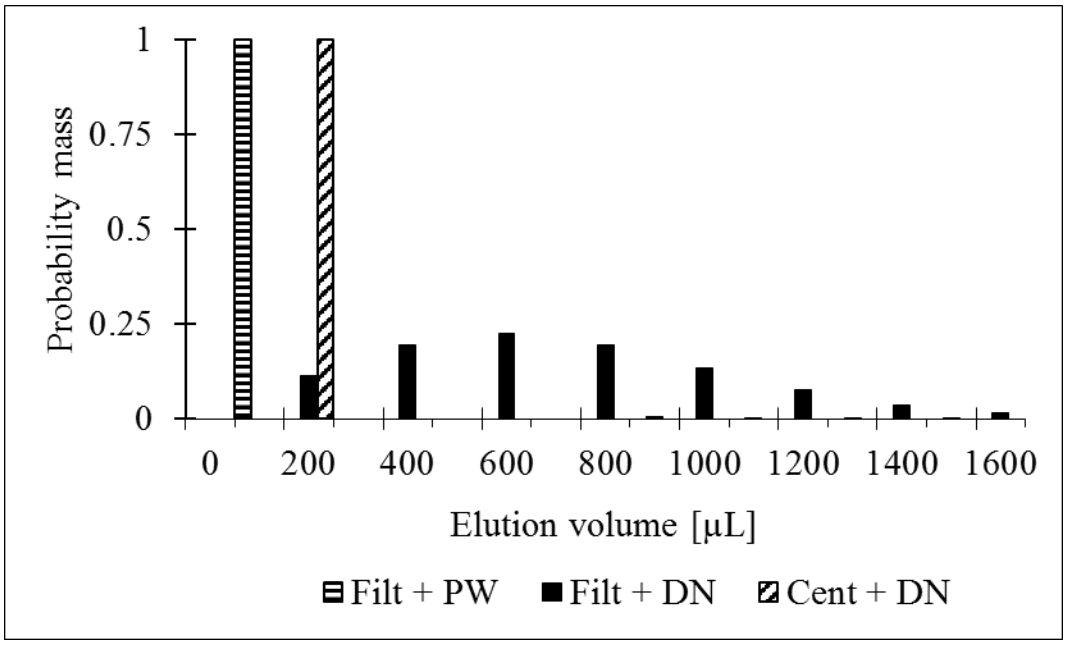

Figure B-1. Distribution of elution volume for filtration with the PW kit (2009-2012 protocol), filtration with the DN kit (2013-2014 protocol) and centrifugation with the DN kit (2015 protocol)

B.2 Estimating the capture and extraction efficiencies from experimental data

In 2016, WGL performed experiments to quantify the extraction efficiency of the IBI Scientific gMax mini (IBI) kit, which is a copy of the Qiagen DNEasy (DN) kit, and the capture efficiencies of filtration and centrifugation.

- First, WGL prepared initial cell dilutions of 10^3 and 10^4 copies/µL. To confirm the amount of DNA copies in each dilution, WGL quantified the number of copies in 32 qPCR replicates of each dilution (x_{ij}). I modeled the amount of DNA delivered to each replicate as a negative binomial random variable with mean $\lambda_i v$ where λ_i is the mean concentration of each dilution i, v is the volume of cell dilution tested (3 µL) and r_i measures the dispersion of the DNA in the dilution for each dilution.

$$x_{ij} \sim NegBinom(\lambda_i v, r_i) \quad \text{(B-2)}$$

- To determine the extraction efficiency of the IBI kit, WGL pipetted 10 µL of the 10^3 copies/µL dilution into an extraction tube for the IBI kit 30 times. 8 qPCR replicates were prepared post-extraction and the amount of DNA in each replicate was quantified. I model the amount of DNA extracted per sample (E_i) as a negative binomial random variable with mean $\lambda_1 v \phi_e$ and dispersion parameter r, where v = 10 µL and ϕ_e is the unknown extraction efficiency of the IBI kit. Then, I model the amount of DNA in each qPCR replicate j from extraction tube i, y_{ij}, as a Poisson random variable with mean $E_i f$, where f = 3 µL/200 µL.

$$E_i \sim NegBinom(\lambda_1 v \phi_e, r_1)$$

$$y_{ij} \sim Poisson(E_i f) \quad \text{(B-3)}$$

- To determine the capture efficiencies of filtration and centrifugation, WGL delivered 10 μL of the 10^4 copies/μL dilution into a 2-L sample for filtration and 5 50-mL tubes for centrifugation. After the capture process, the samples were extracted using the IBI kit and eluted into 200 μL. Eight replicates were prepared for each sample and the amount of DNA in each replicate was quantified using qPCR. I model the amount of DNA in each sample *i* after capture by the given method *j* (j = 1 for centrifugation and 2 for filtration) and extraction via the IBI kit (C_{ij}) using a negative binomial random variable with mean $\lambda_2 v \phi_{c,j} \phi_e$, where v = 10 μL, $\phi_{c,j}$ is unknown and ϕ_e is known from the previous experiment. Then, I model the amount of DNA in each qPCR replicate *k*, z_{ijk}, as a Poisson random variable with mean $C_{ij}f$, where, f = 3 μL/200 μL.

$$C_{ij} \sim NegBinom(\lambda_2 v \phi_{c,j} \phi_e, r_2)$$

$$z_{ijk} \sim Poisson(C_{ij} f) \quad \text{(B-4)}$$

I analyzed the data for all three experiments in one run using JAGS in R. The parameters ($\lambda_1, \lambda_2, r_1, r_2, \phi_e, \phi_{c,1}$ and $\phi_{c,2}$) were fit given the results from the three experiments (x_{ij}, y_{ij}, z_{ijk}). I assumed a gamma distribution for λ_1 and λ_2, with an uninformative prior (shape = .001, rate = .001) and a wide, uninformative uniform prior for r_1 and r_2. I assumed a logistic normal distribution for each of the efficiency terms, ϕ_e, $\phi_{c,1}$ and $\phi_{c,2}$, with an uninformative normal prior for their mean (mean = 0, precision = .0001) and an uninformative gamma prior for their precision (shape = .001, rate = .001). The posterior distributions of each parameter can be found in **Table** B-1 and the correlation matrix of these parameters can be found in Table B-2.

Table B-1 The posterior distributions of the seven unknown parameters (ϕc,1 – capture efficiency of centrifugation; ϕc,2 – capture efficiency of filtration; ϕe – extraction efficiency of IBI kit; λ1 – mean concentration of 103 dilution; r1 – dispersion parameter of 103 dilution; λ2 – mean concentration of 104 dilution; r2 – dispersion parameter of 104 dilution) fitted using the MCMC algorithm.

Parameter	Mean	Standard deviation	95% Confidence Interval	
			Lower (2.5%)	Upper (97.5%)
$\phi_{c,1}$	0.064	0.018	0.035	0.11
$\phi_{c,2}$	0.037	0.011	0.020	0.062
ϕ_e	0.33	0.028	0.28	0.39
λ_1	1520	87	1360	1700
λ_2	18100	349	12500	26200
r_1	10	1.9	6.9	14
r_2	0.90	0.13	0.68	1.1

Table B-2. Correlation matrix of the seven unknown parameters ($\phi_{c,1}$ – capture efficiency of centrifugation; $\phi_{c,2}$ – capture efficiency of filtration; ϕ_e – extraction efficiency of IBI kit; λ_1 – mean concentration of 10^3 dilution; r_1 – dispersion parameter of 10^3 dilution; λ_2 – mean concentration of 10^4 dilution; r_2 – dispersion parameter of 10^4 dilution) fitted using the MCMC algorithm.

Correlation Matrix	$\phi_{c,1}$	$\phi_{c,2}$	ϕ_e	λ_1	λ_2	r_1	r_2
$\phi_{c,1}$	1	0.5	-0.28	0.19	-0.61	-0.0087	-0.010
$\phi_{c,2}$	-	1	-0.26	0.18	-0.62	-0.0033	-0.019
ϕ_e	-	-	1	-0.67	-0.006	-0.006	0.011
λ_1	-	-	-	1	0.006	-0.0059	0.011
λ_2	-	-	-	-	1	0.013	-0.028
r_1	-	-	-	-	-	1	-0.015
r_2	-	-	-	-	-	-	1

B.3 Dilution assay data and results for both PCR and qPCR markers

I estimated the probability of amplification or sequencing one copy of the PCR and qPCR markers, ψ, using experimental data on the proportion of replicates that undergo successful amplification or sequencing over a series of dilutions of initial target DNA copy amounts (Table B-3; Table B-4; Table B-5).[84,87]

I modeled ψ for each marker/reaction by assuming that the number of target marker copies in each of the replicates is distributed as a Poisson random variable with mean equal to the expected number of target marker copies in the dilution. The probability that at least one copy amplifies (p) in the replicate, i.e. the probability of a positive replicate, is one minus the probability that none of the copies in the replicate amplifies or is successfully sequenced. Then, given the number of positive replicates (n_{pos}) out of the number of tested replicates (n_{rep}), I can estimate the probability of amplifying or sequencing a single copy (ψ)

$$N_r \sim Poisson(\lambda)$$

$$p = 1 - (1 - \psi)^{N_r}$$

$$n_{pos} = Binomial(n_{rep}, p) \quad \text{(B-5)}$$

I fit ψ for each marker and reaction using JAGS in R. For each marker/reaction, I specified an uninformative logistic normal prior for ψ (mean = 0, variance = 100000, both on the logit scale).

Table B-3. The frequency of successful PCR amplification for a series of dilutions of Bighead Carp and Silver Carp DNA. Data from Schultz and Lance (2015).[84]

Estimated copy number	Frequency of successful amplification	
	Silver Carp	Bighead Carp
1	0.28	0.22
2	0.3	0.7
3	0.3	0.8
4	0.57	0.8
5	0.85	0.82
6	0.63	0.93
7	-	0.97
8	0.9	0.97
9	0.93	1
10	0.98	1
11	1	1
12	0.97	1
13	0.97	1
14	1	1
15	1	1
50	1	1
100	1	1
200	1	1
500	1	1
1000	1	1

Table B-4. The frequency of successful sequencing of the target DNA sequence for a series of dilutions of Bighead Carp and Silver Carp DNA. Data from Schultz and Lance (2015).[84]

Estimated copy number	Frequency of successful sequencing	
	Silver Carp	Bighead Carp
1	1	0.00
2	0.89	0.10
3	1	0.00
4	0.94	0.24
5	0.95	0.14
6	0.95	0.21
7	1	0.45
8	0.89	0.13
9	0.54	-
10	1	-
11	0.9	-
12	0.97	-
13	0.97	-
14	1	-
15	1	0.83
25	-	0.45
50	-	0.70
100	-	0.93

Table B-5. The frequency of successful qPCR amplification of each marker for a series of dilutions of silver carp DNA (ACTM1, ACTM3, SCTM4, SCTM5) or bighead carp DNA (BHTM1, BHTM2). These data were obtained from Farrington et al. 2015.2 The frequency is calculated

Estimated Copy Number	Frequency of Successful Amplification				Estimated Copy Number	Frequency of Successful Amplification	
	ACTM1	ACTM3	SCTM4	SCTM5		BHTM1	BHTM2
1.6	0.5	0	0.25	0.5	1.1	0.25	0
3.4	0.5	0.75	1	0.75	2.2	1	0.75
46	1	1	1	1	16	1	1
508	1	1	1	1	230	1	1
6051	1	1	1	1	2193	1	1

Table B-6. The posterior distribution of the probability that a single DNA copy amplifies or sequences from dilution assays testing the amplification of the PCR and qPCR markers and sequencing of the PCR markers over a range of target DNA concentrations

Type of Assay	Type of Reaction	Marker Name	Probability of amplifying or sequencing a single target DNA copy [ψ], mean (5% - 95%)
PCR	Amplification	BH-PCR	0.34 (0.28 – 0.41)
		SC-PCR	0.28 (0.22 – 0.34)
	Sequencing	BH-PCR	0.04 (0.03 – 0.05)
		SC-PCR	0.36 (0.27 – 0.46)
qPCR	Amplification	ACTM1	0.27 (0.07 – 0.65)
		ACTM3	0.27 (0.07 – 0.61)
		SCTM4	0.42 (0.13 – 0.78)
		SCTM5	0.36 (0.11 – 0.74)
		BHTM1	0.47 (0.17 – 0.84)
		BHTM2	0.52 (0.11 – 0.98)

Table B-7. The frequency of positive Bighead Carp and Silver Carp eDNA samples found in all US waterways from 2013 to 2016

Year (Laboratory)	eDNA samples in all US waterways [positive samples/total samples (%)]	
	Bighead Carp	Silver Carp
2013 (USFWS)	0/2083 (0.0%)	24/2083 (1.1%)
2014 (USFWS)	7/5714 (0.12%)	34/5714 (0.60%)
2015 (USFWS)	15/6109 (0.25%)	3/6109 (0.05%)
2016 (USFWS)	1/5977 (0.02%)	1/5977 (0.02%)

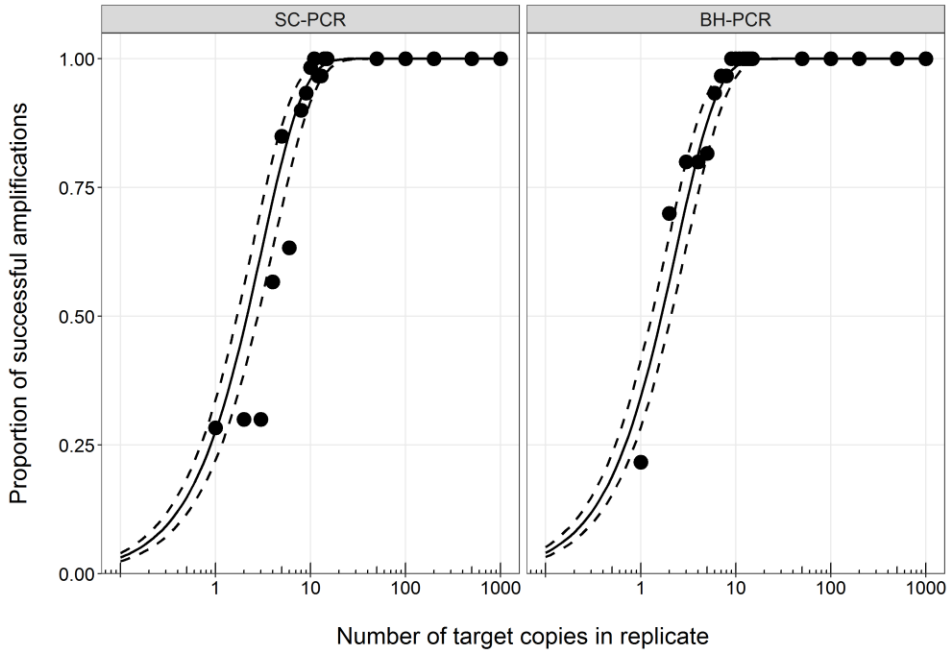

Figure B-2. Results of dilution assay for amplifying BH and SC DNA using the PCR markers BH-PCR and SC-PCR showing the proportion of replicates that are successfully amplified as a function of the expected number of target copies in the replicate and the curve for the fitted distribution of ψ for each marker.

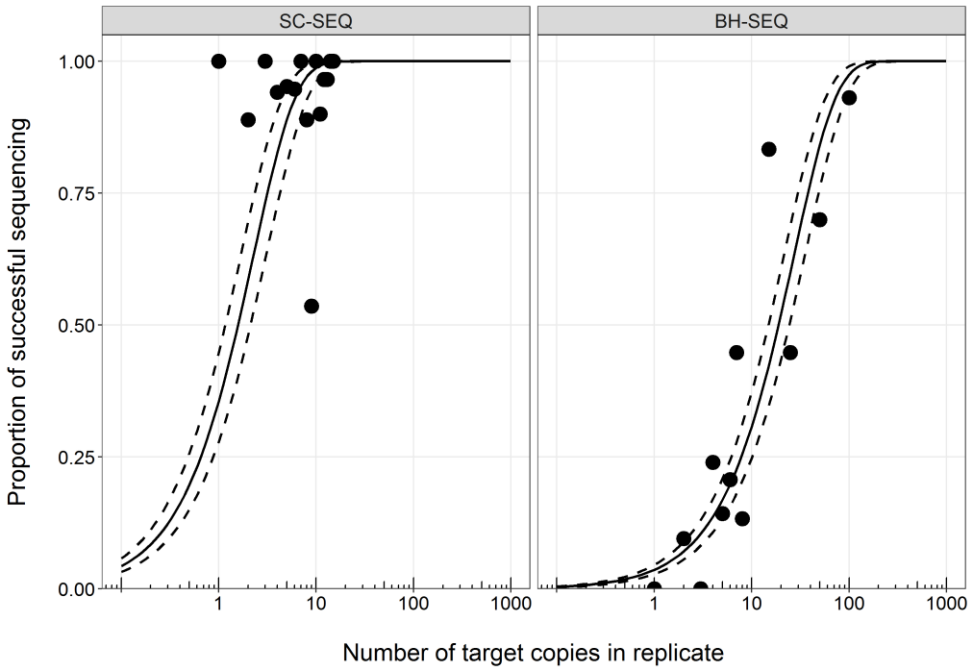

Figure B-3. Results of dilution assay for sequencing BH and SC DNA using the markers BH-PCR and SC-PCR showing the proportion of replicates that are successfully sequenced as a function of the expected number of target copies in the replicate and the curve for the fitted distribution of ψ for each marker.

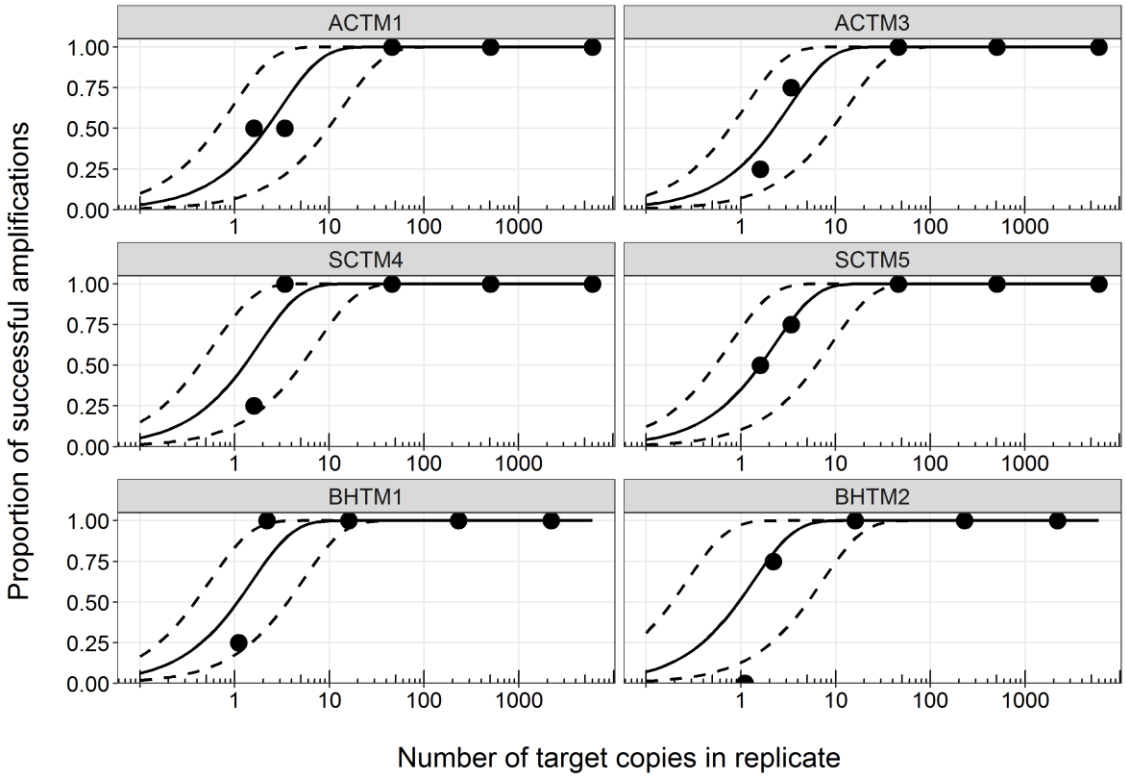

Figure B-4. Results of dilution assay for amplifying BH and SC DNA using the qPCR markers ACTM1, ACTM3, BHTM1, BHTM2, SCTM4, and SCTM5 showing the proportion of replicates that are successfully amplified as a function of the expected number of target copies in the replicate and the curve for the fitted distribution of ψ for each marker.

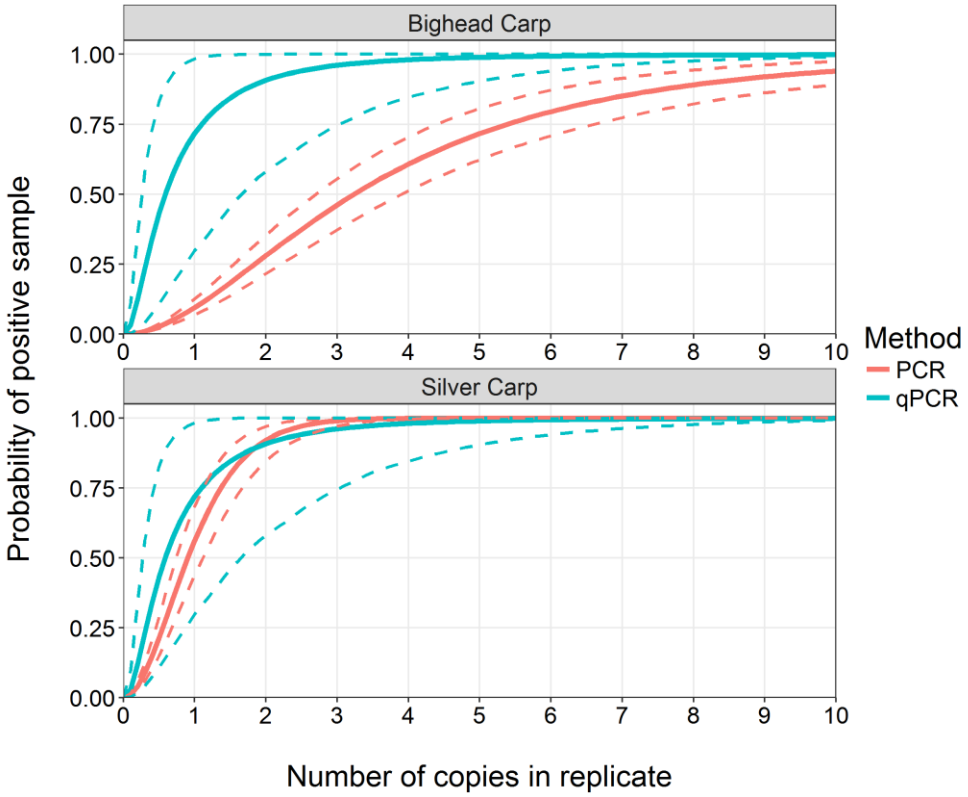

Figure B-5. Comparison of the probability of a positive sample for the PCR and qPCR methods given the number of copies in a replicate for Bighead Carp and Silver Carp eDNA detection. The solid line denotes the mean probability value and the dashed lines show the 95% credible interval.

B.4 Tornado diagrams showing influence of percentage changes to protocol parameters on the protocol's sensitivity

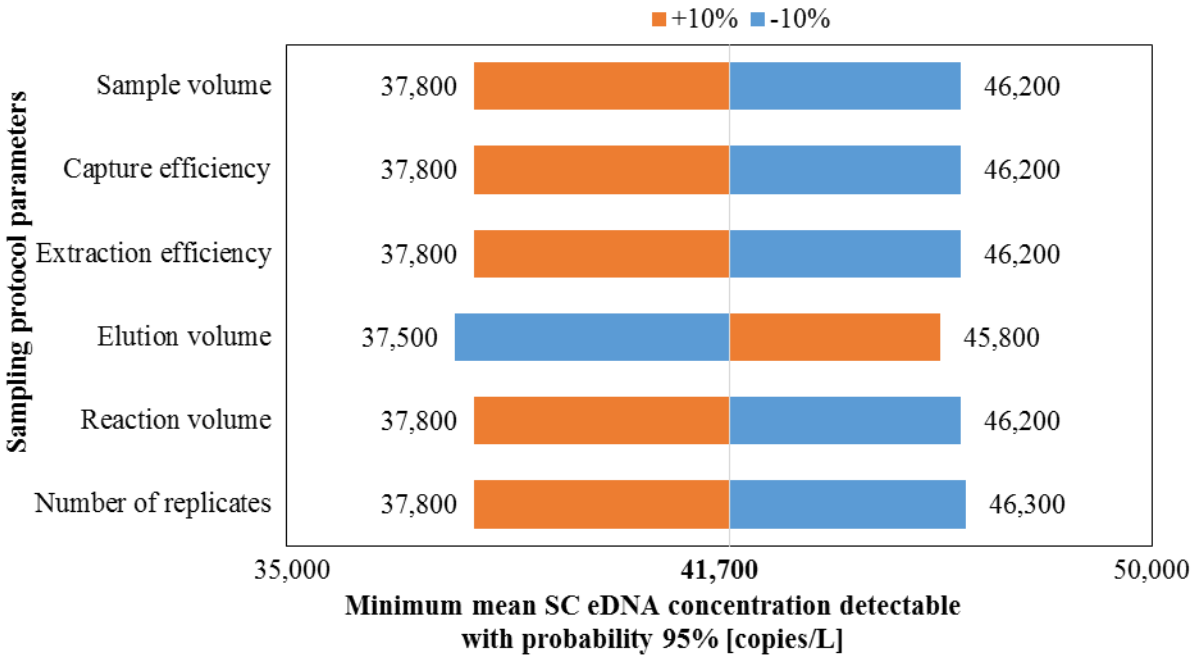

Figure B-6. Tornado diagram showing the change in the minimum mean SC eDNA concentration detectable with probability 95% given a 10% decrease and increase for each of the six sampling protocol parameters. The vertical axis is located at 41,700 copies/L, which is the sensitivity of the current 2015-2016 sampling protocol, i.e., a sample size of 0.25 liters, capture efficiency of 6%, extraction efficiency of 33%, elution volume of 200 µL, a reaction volume of 3 µL and 8 replicates. It is assumed that the eDNA was randomly distributed (r = 100).

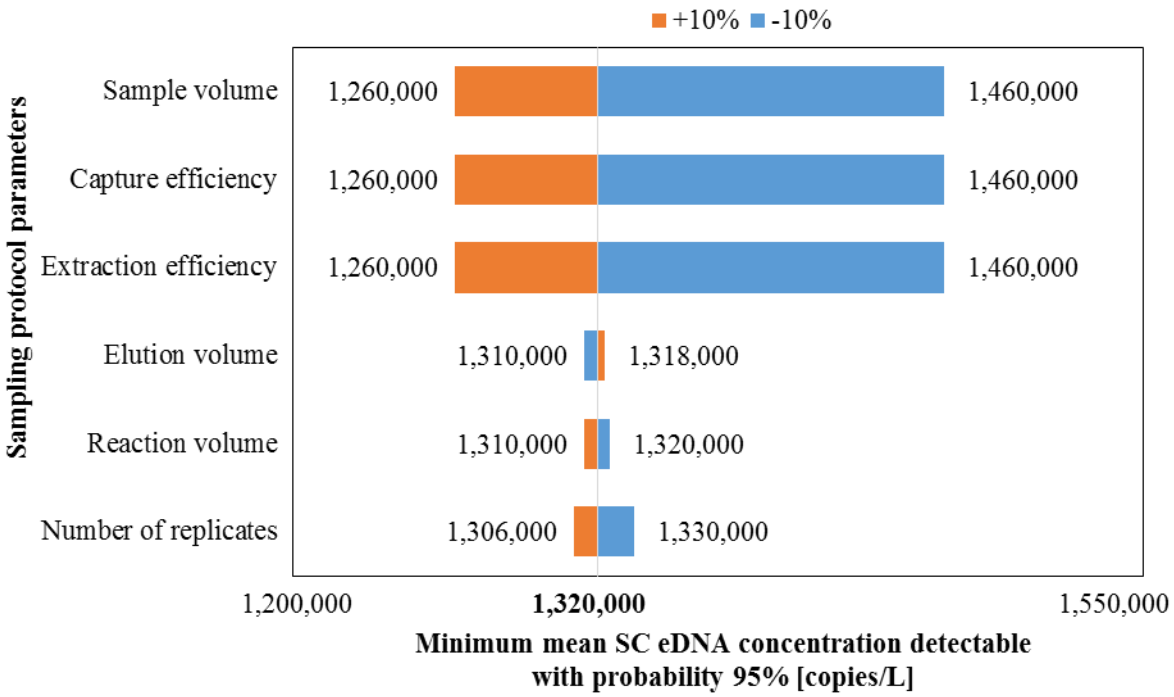

Figure B-7. Tornado diagram showing the change in the minimum mean SC eDNA concentration detectable with probability 95% given a 10% decrease and increase for each of the six sampling protocol parameters. The vertical axis is located at 1,320,000 copies/L, which is the sensitivity of the current 2015-2016 sampling protocol, i.e., a sample size of 0.25 liters, capture efficiency of 6%, extraction efficiency of 33%, elution volume of 200 µL, a reaction volume of 3 µL and 8 replicates. It is assumed that the eDNA was clumped (r = 0.3).

Appendix C. Supporting Information for Chapter 4

Table C-1. The number of positive samples of each target marker (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) for each combination of sampling location in the Illinois River(La Grange pool, Marseilles pool, Brandon Road pool) and capture method (filtration or centrifugation).

Sampling location	Capture method	Number of positive samples for target marker _____ out of 25 total samples					
		AC-TM1	AC-TM3	BH-TM1	BH-TM2	SC-TM4	SC-TM5
La Grange pool	Filtration	19	19	19	19	19	19
	Centrifugation	23	23	23	22	23	23
Marseilles pool	Filtration	23	22	20	21	23	23
	Centrifugation	21	20	11	13	17	19
Brandon Road pool	Filtration	1	1	0	0	0	0
	Centrifugation	5	6	2	2	2	2

Table C-2. The estimated mean eDNA concentration distribution of each target marker (AC-TM1, AC-TM3, BH-TM1, BH-TM2, SC-TM4, SC-TM5) in each sampling location in the Illinois River (La Grange pool, Marseilles pool, Brandon Road pool) and the CAWS. The value in larger font is the mean and the values in the parentheses are the 95% credible intervals of the distribution.

Target marker	Mean eDNA concentration [copies/L]			
	La Grange pool	Marseilles pool	Brandon Road pool	CAWS
AC-TM1	4,030,000 (1,840,000 – 8,340,000)	442,000 (215,000 – 870,000)	19,800 (9,490 – 39,200)	9,420 (1,020 – 78,900)
AC-TM3	2,600,000 (1,180,000 – 5,330,000)	393,000 (192,000 – 755,000)	22,100 (4,510 – 20,600)	6,610 (828 – 32,300)
BH-TM1	321,000 (130,000 – 704,000)	256,000 (108,000 – 508,000)	10,600 (7,850 – 32,400)	3,770 (35 – 13,300)
BH-TM2	545,000 (237,000 – 1,130,000)	233,000 (91,500 – 483,000)	16,400 (7,850 – 32,400)	4,100 (52 – 16,500)
SC-TM4	1,780,000 (794,000 – 3,540,000)	296,000 (141,000 – 546,000)	8,230 (3,200 – 17,600)	8,540 (152 – 47,300)
SC-TM5	4,380,000 (1,970,000 – 8,570,000)	360,000 (177,000 – 688,000)	10,000 (4,260 – 20,100)	9,330 (131 – 49,900)

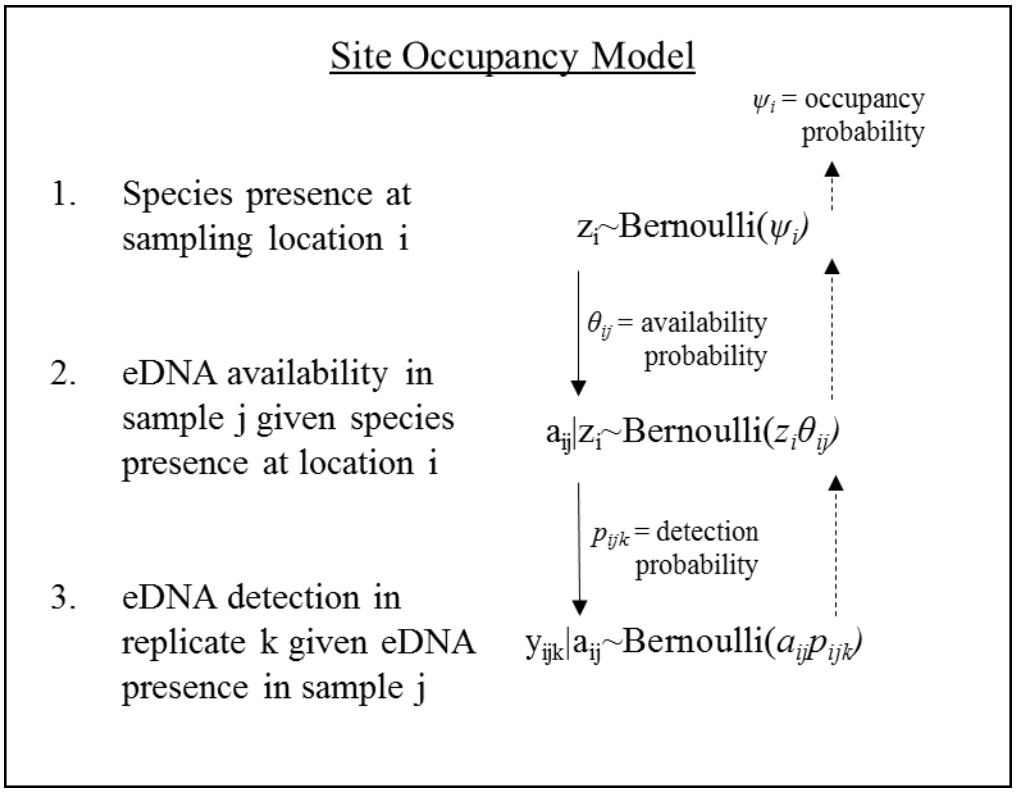

Figure C-1. A three-stage site occupancy model when interpreting eDNA sampling data

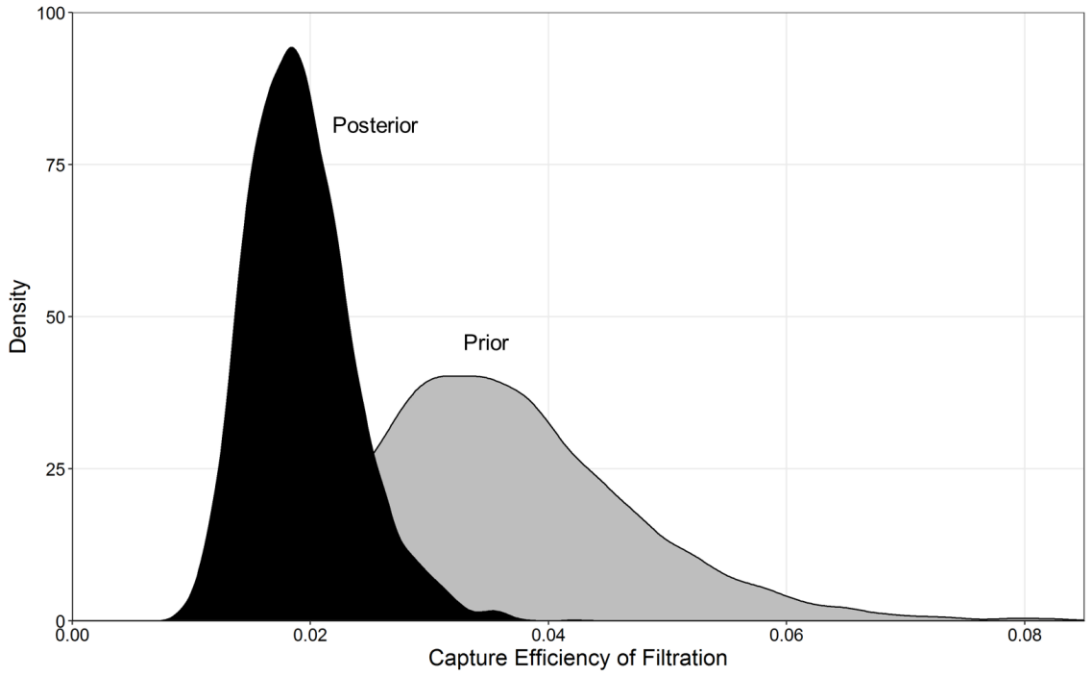

Figure C-2. The prior (grey) and posterior (black) distributions of the capture efficiency (ϕ_c) of the filtration method.

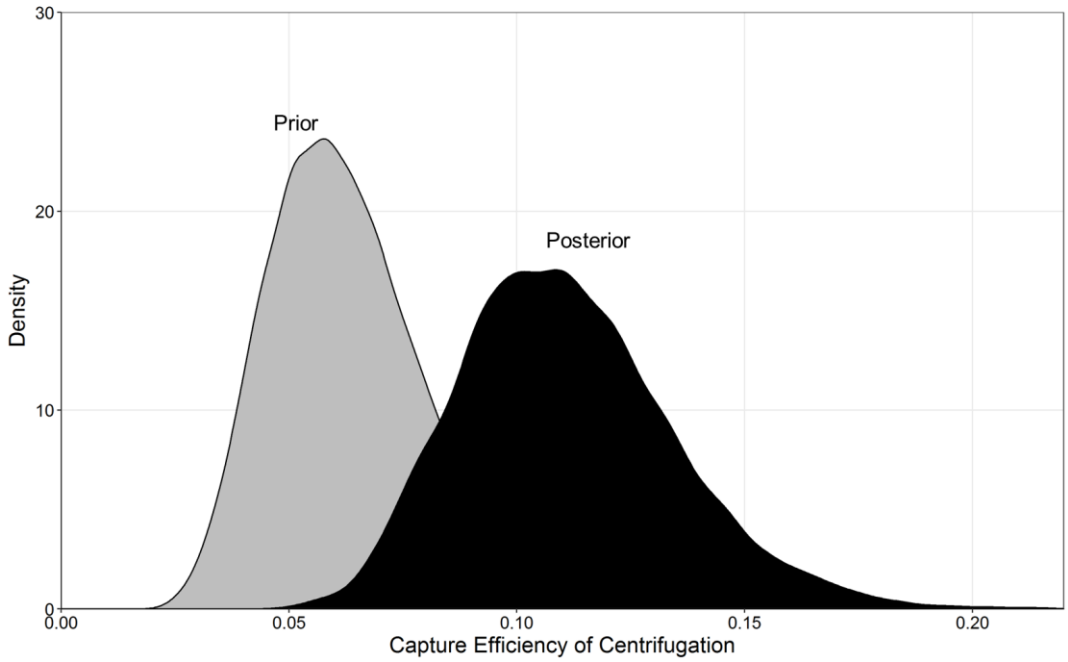

Figure C-3. The prior (grey) and posterior (black) distributions of the capture efficiency (ϕ_c) of the centrifugation method.

Appendix D. Supporting Information for Chapter 5

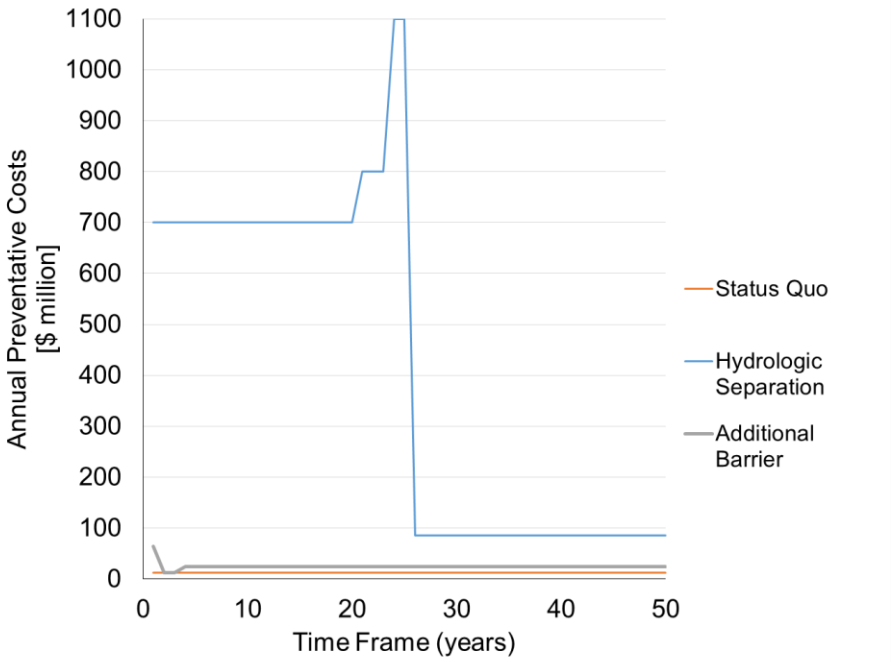

Figure D-1. Annual costs for implementing each of the prevention strategies over the 50 year time frame. The stop prevention strategy is not shown, since that will have zero preventative costs.

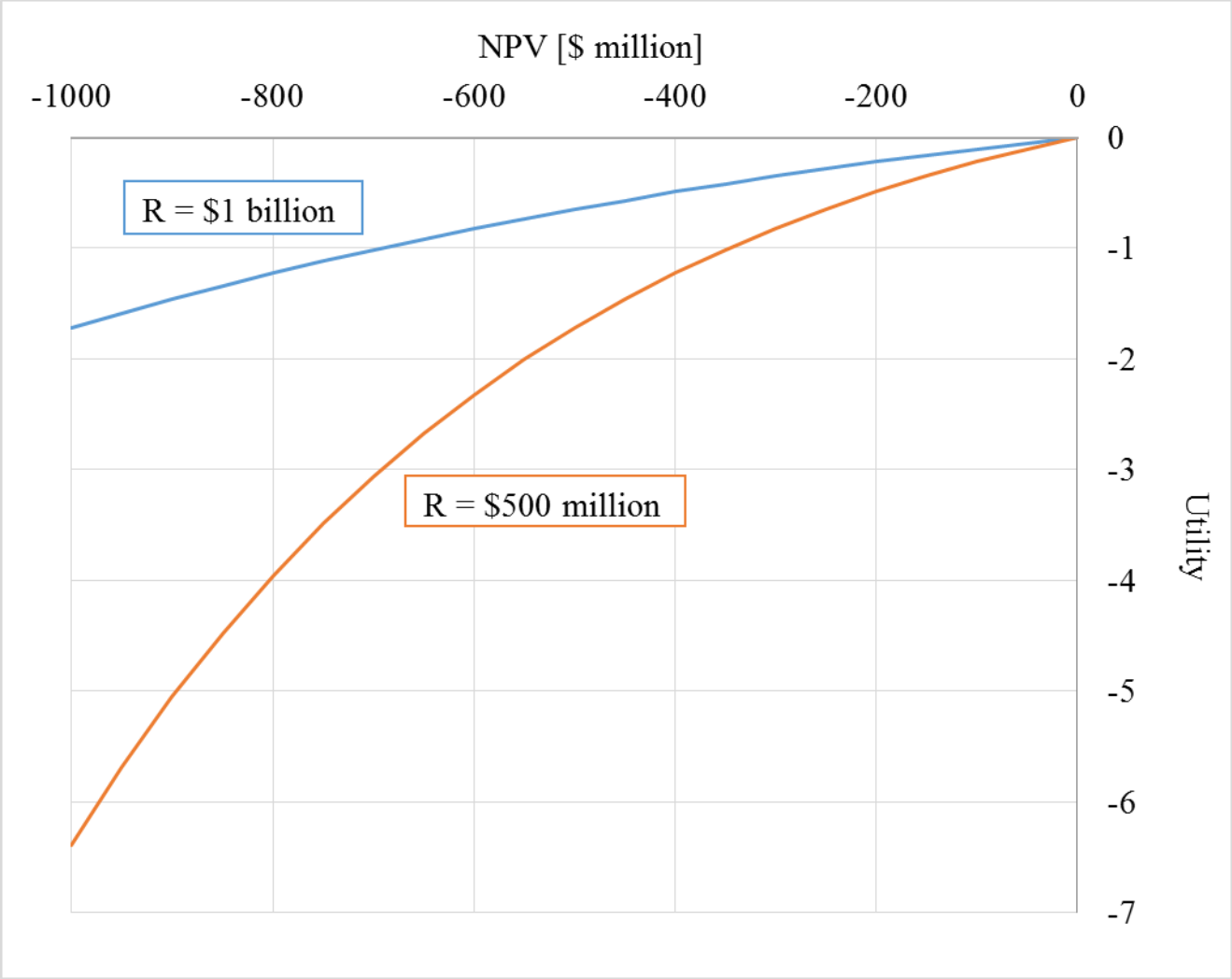

Figure D-2. The exponential utility function (Utility = $1 - e^{-NPV/R}$) for two risk tolerances (R), $1 billion and $500 million.

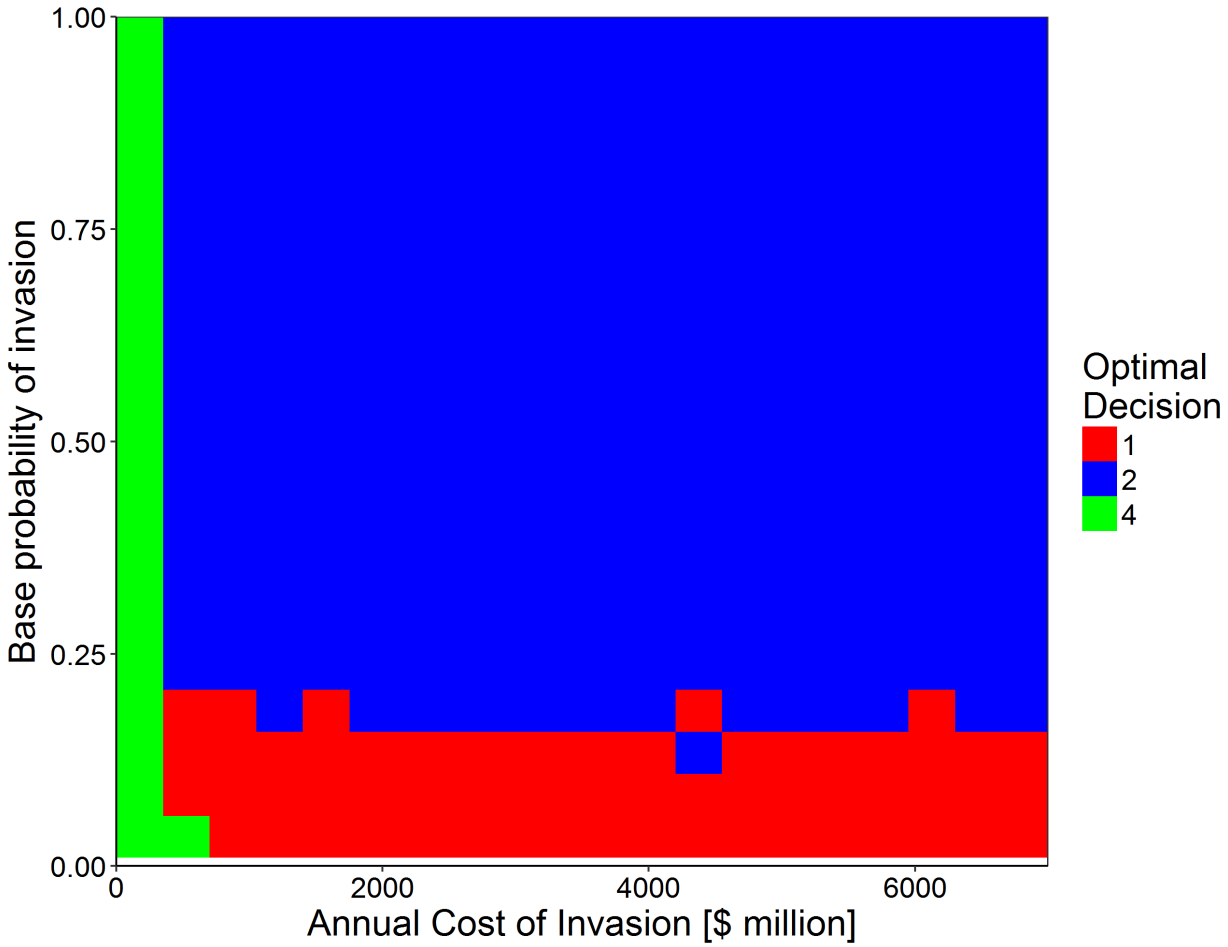

Figure D-3. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given risk neutrality, time frame of 100 years and a discount rate of 4%. The colors denote the optimal prevention strategy for each combination of invasion costs and base probability of invasion. Red is the status quo, blue is additional barrier and green is stop prevention.

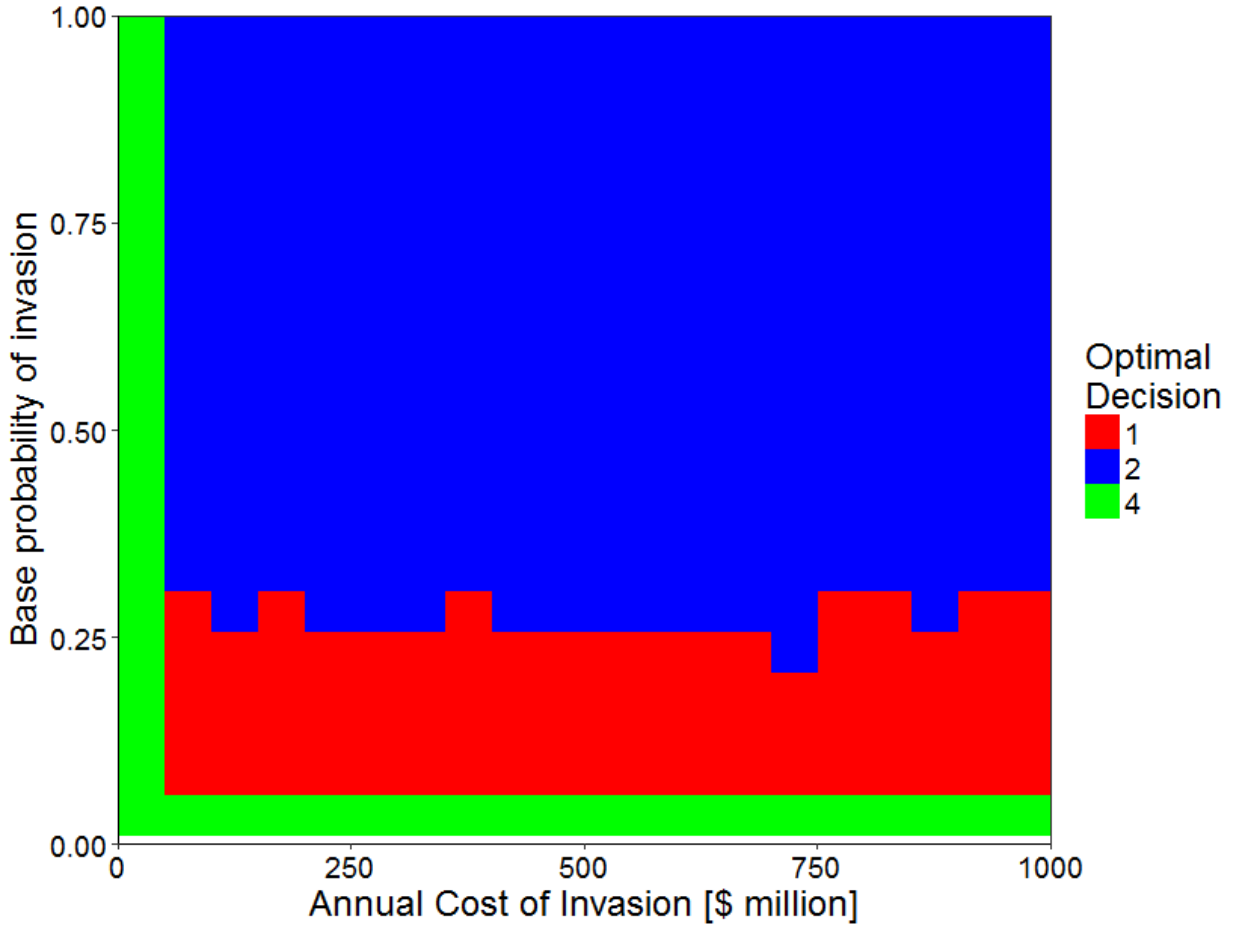

Figure D-4. Optimal prevention strategy as a function of the annual cost of invasion on the x-axis and the base probability of invasion on the y-axis, given "environmental" risk tolerance of $1 billion. Red is the status quo, blue is additional barrier and green is stop prevention.