Optimization and validation of Chinook salmon eDNA detection in the San Francisco Estuary

By

THIAGO MISSFELDT SANCHES

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Andrea Schreier, Chair

_____

Brett Harvey

_____

Michael Miller

Committee in Charge

2022

Abstract

To monitor the health of Chinook Salmon populations in California's Central Valley, it's necessary to measure and evaluate all stages of their life cycle. The outmigration journey of the pre-smolt juvenile Chinook is one of the least understood life stages, yet it's also one of the most dangerous for the species. Rising water temperatures, drought, invasive predatory species and degradation of the marshland habitats in the San Francisco Bay-Delta (SFBD) created an unsustainable environment where most juvenile Chinook never make it to the Pacific Ocean. Although multiple marsh restoration projects are being developed to lessen the anthropogenic impact, the Chinook Salmon populations are still declining. An increased focus in monitoring the habitat use of pre-smolt Chinook salmon in the shallow-water marsh habitat of the upper SFBD can provide an invaluable insight into the best management strategies to ensure the survival of the species. Due to complications in monitoring the marsh habitats with conventional methods, I evaluated the use of environmental DNA (eDNA) as an alternative indicator of Chinook salmon presence. For the best use of this novel technology in the marsh conditions, we validated and optimized the performance of a Chinook salmon eDNA assay for use in an estuarine environment. Then we compared the effectiveness of eDNA detection to trawling, measuring the sensitivity of each method as well as identifying the biases, optimal working conditions, advantages, and disadvantages of these survey methods for mapping habitat use of the pre-smolt Chinook salmon.

eDNA is a stable molecule that can persist in the environment for long periods. Since all living creatures release eDNA into the environment, almost any environmental sample contains eDNA. The eDNA can reflect current or historical distributions of the species living in the environment sampled, with environmental conditions such as eDNA transport, eDNA degradation,

sedimentation and PCR inhibitors regulating the shape and duration of the eDNA plumes produced by our target species. Conducting a single species eDNA assay in an aquatic environment consists of four main steps: water filtration, DNA extraction, PCR inhibitor removal and qPCR amplification. In the filtration step a fixed amount of water, normally 1 L per sample, is filtered through a material which will capture the eDNA while the flowthrough is discarded. The filter pore size is the factor that determines the state of the eDNA that will be measured. Larger pores prioritize larger particles such as scales and eggs while smaller pores favor individual cells, subcellular structures and free eDNA. Then the extraction step isolates the eDNA from the filter, and the yield of this step largely influences the sensitivity of the assay. The PCR Inhibitor removal step works to remove any possible contaminants that might impact the amplification of eDNA; therefore, this step is essential to provide reliability of the assay across different environments. Last, the qPCR step amplifies the eDNA to a detectable amount of fluorescence.

Marsh habitat is a challenging environment for eDNA surveys because it is associated with an elevated turbidity, which may clog filters and elevate presence of qPCR inhibitors. The first objective of my dissertation was to optimize a Chinook salmon qPCR assay to detect Chinook eDNA in the marshy and estuarine conditions of the SFBD. As the goal for the study was to implement the assay on a large scale, I not only focused on the sensitivity of the assay, but also accounted for the time and costs associated with sampling. I tested a total of 27 combinations of filters, extraction, and inhibitor removal methods. For the conditions of the SFBD we opted for a protocol using glass fiber filters, magnetic bead DNA extraction, and an extra step for PCR inhibitor removal, to best achieve our goals of a large scale eDNA monitoring system for juvenile Chinook salmon in the Upper SFBD.

The second chapter of my dissertation determined how best to interpret the meaning of an eDNA positive or negative detection. To do so, I needed to evaluate the dispersion of the eDNA once it is released to better characterize the distribution of the juvenile Chinook salmon. In the literature, eDNA particles have initially been described as fine particulate organic matter (FPOM) particles and simulated as such with hydrological models. For all hydrological models for eDNA, three factors are essential: eDNA production, eDNA degradation and eDNA transport. In general, eDNA production is considered proportional to the mass of the target species, although more recent papers suggest that eDNA production is more correlated with surface area in the case of fishes. Stress, death, and environmental conditions may also affect the release of eDNA. Meanwhile eDNA degradation is mostly dependent on the environment with the bacterial activity dictating most of the eDNA destruction. eDNA can also become undetectable due to precipitation and binding to molecules that inhibit the amplification step of the assay. eDNA transport mostly is dictated by advection in the case of marine and lotic water bodies while in lentic systems turbulent diffusion is the main factor for eDNA dispersion.

I built a simplified one-dimensional riverine eDNA transport model to estimate the detection radius of an individual eDNA source and evaluate the effects of the degradation rate, eDNA diffusion coefficient and river advection onto the distance that the eDNA particles can travel before degrading or settling in the bottom of the river. I identified that the main driver of eDNA dispersion is in most cases the advection of the system, which can greatly influence the transport of eDNA. Due to this effect the model can be used to estimate the effective part of the river that was measured for presence or absence of Chinook salmon with the eDNA assay. The modeling also suggests that transect sampling can be a valuable alternative to discrete, or point, sampling in order to increase the repeatability of detection and increase the sensitivity of the assay, with

medium sized transects (250-500 m) providing most of the benefits of transect sampling. The model also uses a Bayesian Monte Carlo Markov chain to estimate degradation rates, production rates and diffusion coefficients for further cage studies.

Between the years 2018 and 2021, I collaborated with the California Department of Water Resources on a study that performed pairwise sampling of eDNA and trawling to compare the detection rate of both methods to estimate their biases. Water conditions were also measured to describe the sampled environments. I observed that the eDNA survey had a higher detection rate than the trawling survey, with a 45% detection rate compared to 13% detection rate for trawling. eDNA detection was present in a broad set of environments while trawling detections were confined mostly to a narrow set of salinity and dissolved oxygen conditions. Trawling detections were correlated with shoals in the main corridor of Suisun Marsh while detection was minimal in the upper marshlands. eDNA positivity was similar throughout the sampled years while trawl detection occurred mostly in the wet year of 2019. Trawling seems more efficient when river discharge is high, possibly caused by a higher movement or by a higher number of juvenile Chinook throughout the system. My results suggest that trawling surveys may be hampered by dense vegetation and muddy substrates found in the upper marshlands, while eDNA is a promising alternative survey method for these conditions. With the increasing frequency of drought and reduced flows in the SFBD, we expect that eDNA surveys will be a valuable strategy to monitor the Chinook salmon population as well as measure the efficiency of conservation efforts in the region. Although the eDNA detection rate can be a better indicator of Chinook salmon habitat use, trawling and other fish capture survey methods can generate additional information about the health of populations with measures such as size, weight, diet, age, and origin. Therefore, eDNA and trawling provide a complementary view of the juvenile Chinook salmon population. We

recommend that managers conduct eDNA surveys to estimate the habitat use and occupancy rates and use targeted trawling surveys to confirm Chinook presence and provide information that cannot be obtained from eDNA alone.

Table of Contents

Optimizing an eDNA protocol for estuarine environments: balancing sensitivity, cost, and time 1

Abstract 1

Author Summary 2

Introduction 3

Methods 6

Ethics statement 6

Accounting for cost and time in experimental design 6

Experimental Design 7

Sampling 8

Estimation of average input eDNA 9

Filtration 10

Extractions 10

Secondary Inhibitor removal 12

qPCR amplification oligos 13

Data analysis 14

Results 14

Effect of filter type on DNA extraction yield 14

DNA retention and filtration time per filter type 16

Comparison between DNA extraction protocols 18

Importance of each protocol step on total DNA yield 20

Conditions in which the use of SIR is necessary 21

Discussion 22

References 28

Supporting information 30

eDNA transport and novel sampling opportunities 31

Abstract 31

Introduction 32

Materials and Methods 35

Overview 35

Definition of transect sampling and discrete sampling 37

eDNA production and degradation 38
eDNA transport 38
[eDNA] spatial distribution 38
Sampled eDNA 39
Sampled eDNA from multiple sources 41
Model assumptions 41
Cases of interest 43
Generating random fish distributions 44
Average detection rate for randomly distributed eDNA sources 45
Estimation of effective sampled distance (ESD) 45
Conversion from sample eDNA copy number to probability of detection 46
Data analysis 46
Results 46
Effects of each environmental parameter on the [eDNA] distribution 46
eDNA distribution reflects target species distribution in low advection systems 48
Transect sampling increases Effective Sampling Distance 50
Advection and transect length interact to define the ESD of a single source 52
Presence of multiple sources increases the advantage of transect sampling over discrete sampling 54
Differences between transect and discrete sampling are attenuated in high advection situations 56
Longer transects have diminishing returns 56
Standard deviation of the detection probability is smaller for long transects 56
Discussion 59
References 64
Supplementary 2: Fitting the eDNA dispersion model to published data 68
Comparative analysis of environmental DNA and trawling surveys for juvenile Chinook salmon in the San Francisco Bay-Delta 70
Abstract 70
Introduction 71
Methods 77
Study Area 77
Sample site selection 78
Trawling and water quality sampling 80
eDNA sampling 80

eDNA extraction and amplification 82
Data analysis 82
Results 84
eDNA showed higher positivity rates for all sampled years and the difference in positivity is accentuated in dry years. 84
Spatial distribution of positive sites for eDNA and trawl 86
eDNA has higher detectability in all measured morphological and hydrological conditions and the detection gap is most pronounced in cut banks and terminals 87
Salinity is the most statistically significant water quality parameter that differentiates eDNA and trawling detections 89
Logistic regression model 91
Discussion 93
eDNA has higher detection rate than trawling 94
Terminal channels and cut banks strongly impacted eDNA versus trawl detectability 95
Salinity had the strongest bias in the eDNA and trawl detectability 96
eDNA shown little to no cross-contamination between sites 97
Effects of eDNA - trawl detection imbalance on model predictions 98
Integrating eDNA and conventional sampling 99
Future of eDNA surveys in the San Francisco Bay-Delta 101
References 102
Supplementary 3: alternative models 104
Logistic CAR model 104
Feature interaction model 106
Supplementary 4: Model Comparison 107

Optimizing an eDNA protocol for estuarine environments: balancing sensitivity, cost, and time

Abstract

Environmental DNA (eDNA) analysis has gained traction as a precise and cost-effective method for species and waterways management. To date, publications on eDNA protocol optimization have focused primarily on DNA yield. Therefore, it has not been possible to evaluate the cost and speed of specific components of the eDNA protocol, such as water filtration and DNA extraction method when designing or choosing an eDNA protocol. At the same time, these two parameters are essential for the experimental design of a project. Here we evaluate and rank 27 different eDNA protocols in the context of Chinook salmon (*Oncorhynchus tshawytscha*) eDNA detection in an estuarine environment. We present a comprehensive evaluation of multiple eDNA protocol parameters, balancing time, cost and DNA yield. We collected samples composed of 500 mL estuarine water from Deverton Slough (38°11'16.7"N 121°58'34.5"W) and 500 mL from tank water containing 1.3 juvenile Chinook Salmon per liter. Then, we compared extraction methods, filter types, use of inhibitor removal kit for DNA yield, processing time, and protocol cost. Lastly, we used an MCMC algorithm together with machine learning to understand the DNA yield of each step of the protocol as well as the interactions between those steps. Glass fiber filtration was to be the most resilient to high turbidites, filtering the samples in 2.32 ± 0.08 min instead of 14.16 ± 1.86 min and 6.72 ± 1.99 min for nitrocellulose and paper filter N1, respectively. The filtration DNA yield percentages for paper filter N1, glass fiber, and nitrocellulose were 0.00045 ± 0.00013,

0.00107 ± 0.00013, 0.00172 ± 0.00013. The DNA extraction yield percentage for QIagen, dipstick, NaOH, magnetic beads, and direct dipstick ranged from 0.047 ± 0.0388 to 0.475 ± 0.0357. For estuarine waters, which are challenging for eDNA studies due to high turbidity, variable salinity, and the presence of PCR inhibitors, we found that a protocol combining glass filters, magnetic beads, and an extra step for PCR inhibitor removal, is the method that best balances time, cost, and yield. In addition, we provide a generalized decision tree for determining the optimal eDNA protocol for other studies in aquatic systems. Our findings should be applicable to most aquatic environments and provide a clear guide for determining which eDNA protocol should be used under different study constraints.

Author Summary

The use of environmental DNA (eDNA) analysis for monitoring wildlife has steadily grown in recent years. Though, due to differences in the ecology of the environment studied and the novelty of the technique, there are fewer standards for eDNA detection compared to other methods. Here we examine each step of an eDNA assay, looking at common protocols and comparing their efficiencies in terms of time to process the samples, cost and how much DNA is recovered. We then analyze the data to provide a recommendation for best practices given different project constraints. For estuarine conditions, we suggest the use of glass fiber filtration, the use of paramagnetic beads for DNA extraction and the use of a secondary inhibitor removal step. We expect our findings to help managers decide their preferred approach before initiating a project, not only for estuarine conditions but for other aquatic habitats.

Introduction

Environmental management relies heavily on knowledge of the spatial distribution of species. In the past decade, environmental DNA (eDNA) monitoring has gained traction as one of the most sensitive and cost-effective monitoring methods[1], allowing researchers to better estimate species occupancy rates in each habitat. eDNA refers to genetic material present in an environmental sample, which is shed from all organisms in the form of fluids, skin cells/scales, decay and feces from the target species. Filtration or precipitation procedures can isolate eDNA or cells containing DNA from an environmental sample[2]. eDNA acts as a species-specific footprint, which then can be detected by researchers and used to infer the presence of a target species[3]. eDNA might also be used to monitor biodiversity of whole communities as all living beings possess and shed DNA. Although DNA is present in most environmental samples, chemical factors and physical forces can reduce with the DNA detection of an eDNA assay. Unfortunately, due to high variability in the physical and chemical characteristics of studied environments, there are no clear guidelines to assist investigators in choosing an optimal protocol for their eDNA monitoring studies.

One environment in which there has been comparably few eDNA studies are estuaries. The estuarine environment provides a challenge for eDNA biomonitoring as the elevated density of solid particles, measured by turbidity levels, can bind to the eDNA and clog the filter pores, limiting the volume of water that might be filtered. Also, estuarine water has been shown to contain elevated levels of PCR inhibitors[4,5]. While we focus on the estuarine habitat here, we assume that if our DNA amplification-based experiments work in these complex conditions, the same approach could also be applied to less turbid freshwater and marine conditions.

The typical protocol used to isolate eDNA from water samples can be described in four steps: filtration, DNA extraction, inhibitor removal and DNA amplification in order to estimate

the initial concentration of eDNA[6]. In the filtration step, the water samples, with preferred volumes of 1 liter[7–9], are pressure-pumped through a membrane filter which captures the free DNA as well as tissue and cells suspended in the water. One problem that arises is filter clogging, which leads to under-sampling and irregular sampling volumes. On the other hand, if the pore size or filter material does not capture enough particles, the DNA retention and therefore DNA yield will be reduced. These two diverging issues also have to be balanced with the cost of each filter, which can inflate the total cost of the project considering the number of sampled sites and the number of replicates per site. Currently, the best way to address these problems is to test a variety of filter materials with different pore sizes and identify the ones that have the best characteristics for the sampled system.

The next step is to extract DNA from the filter using conventional extraction methods, which were developed to isolate large nuclear DNA fragments from tissue. However, in the case of eDNA, it is preferable to target small fragments of mitochondrial genes as they have a higher copy number per cell compared to nuclear DNA and are more likely to be detected in an environmental sample. Currently, the standard DNA extraction protocols have been thoroughly optimized for DNA yield, although high cost and low throughput of these techniques are still problematic. Novel extraction protocols aim to address these problems yet have encountered difficulties in achieving high DNA yield of the standard DNA extraction protocols[10]. To increase DNA yield, our study adapted those novel protocols to the context of eDNA such that the sample volumes used are an order of magnitude larger than conventional DNA extractions for molecular biology assays.

Filters may also capture high concentrations of PCR inhibitors. PCR inhibitors are substances that can inhibit PCR amplification. Their inhibiting mechanism varies between

affecting the template DNA, the DNA polymerase, or other reagents necessary for the reaction. PCR inhibitors can be catalytic (e.g. proteases degrading proteins and phenol degrading DNA) or work through competitive binding (e.g. melanin forming a complex with the polymerase and humic acid interacting with the DNA template)[11]. Humic matter and proteases are typical PCR inhibitors present in high concentrations in turbid waters and other environmental samples[11,12]. It is often necessary to use a secondary inhibitor removal (SIR) step to further isolate the DNA from contaminants[7]. Since SIR is a column-based extraction, part of the eDNA present in the sample might be lost, as it stays bound to the filter after the last elution step. It is expected that the gain from stopping PCR inhibition outweighs the DNA loss. However, this step might be skipped in order to diminish costs whenever PCR inhibitors are not present in a sample.

Lastly, the isolated DNA is amplified using quantitative PCR (qPCR) with primers specific to the target species, and the initial amount of the target eDNA is determined based on the Cq value[13]. Although it is possible to use conventional PCR, this method significantly underperforms compared to qPCR in terms of sensitivity, and even when DNA is successfully amplified it provides less informative data[14]. Thus, the protocols in this study were only tested using qPCR.

In this study, we separate and optimize four important steps for eDNA biomonitoring of delta estuarine waters, which are characterized by elevated concentrations of solid suspended particles and fluctuating levels of salinity[15]. We targeted Chinook Salmon in our experiments for a variety of reasons. First, as a widespread species in the North American Pacific Northwest, it has invaluable importance for the stability of the marine ecosystem of the region[16] and at the same time, provides a critical source of income for historic fishing communities[17,18]. Little is yet known about the spatial-temporal distribution and estuarine habitat usage of pre-smolt juvenile Chinooks

during their annual out migration to the ocean. Developing a high precision, high throughput eDNA protocol optimized for estuarine waters will allow managers to have a better understanding of the habitats used by Chinook in their early life-stages. We comment on the specifics of each step for eDNA biomonitoring, providing a framework that will help investigators make more informed decisions about the best protocol for their study, considering various study constraints.

Methods

Ethics statement

Holding juvenile Chinook salmon in captivity to sample water for this study was approved by the University of California Davis Institutional Animal Care and Use Committee (USDA registration: 93-R-0433, PHS Animal Assurance A3433-01) under the protocol number #20608.

Accounting for cost and time in experimental design

To decide on the most practical estuarine eDNA protocol, we first need to determine what it means for a protocol to be efficient. In our case we listed our priorities in the following order: 1) The eDNA yield must be adequately sensitive in realistic scenarios; 2) The protocol must be fast and scalable, and 3) The protocol must be cost effective, considering that reagent cost is the main driver of cost per sample. This order of priorities is influenced by several factors that include species abundance and costs. If the target species is known to be present and potentially at high density, the DNA yield constraint can be loosened, allowing the use of faster and more cost-conscious protocols. If labor cost is inexpensive, choosing a more time intensive yet cheaper protocol will maximize the number of sampling points. On the other hand, in situations where labor accounts

for much of the costs, choosing less time intensive protocols will allow more sampling points for the project.

Experimental Design

The goal of this study was to identify the best combination of filter, extraction method, and inhibitor removal steps for studies with varying constraints. We tested three biological replicates for every combination of filter and extraction method, with and without inhibitor removal, and measure the amount of recovered eDNA using qPCR Cq values (Figure 1.1). A DNA standard curve was developed from a fin clip serial dilution on the same plate. We defined an equation that describes how the efficiency of each step influences the total amount of recovered eDNA:

$$Y \sim Y_f Y_e(f) - I_f I_e * \begin{cases} 0 \text{ , if secondary inhibitor removal is used} \\ 1 \text{ , otherwise} \end{cases} \quad (1.1)$$

where:

Y:ratio of input eDNA that was amplified by the qPCR

Y_f:ratio of input eDNA that binds to filter

$Y_e(F)$:ratio of eDNA bound to the filter that is isolated by the extraction method

I_f:filter inhibitor carryover

I_e:extraction method inhibitor carryover

$I_f I_e$:ratio of input eDNA not available to amplification due to inhibitors

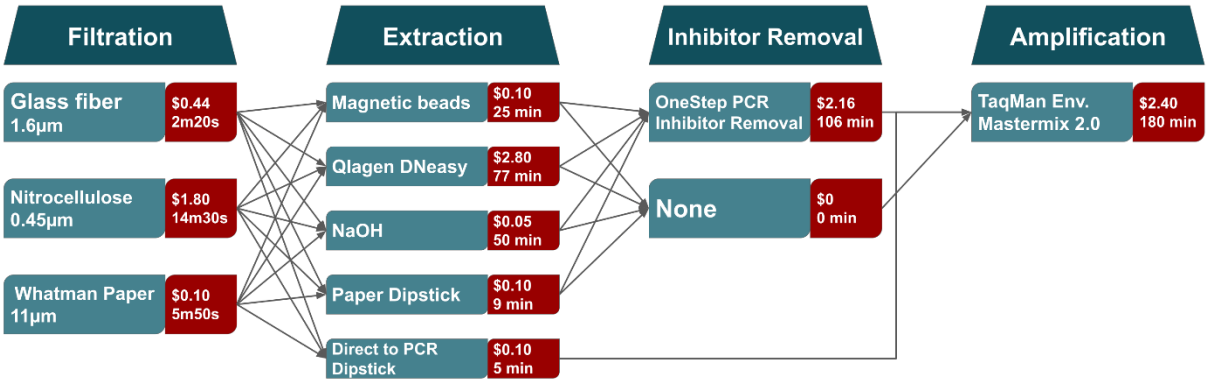


Figure 1.1: Scheme of steps for an eDNA protocol with tested methods for each step. Cost and processing times (in minutes) of each method are shown next to the method name. Number of samples for the measured times varies as the number of samples that can be run in parallel varies between steps. Costs are estimated per sample.

Then, based on equation 1.1 we used Automatic Differentiation Variational Inference (ADVI)[19] to estimate the distribution of the parameters that maximize the likelihood of the observed yields.

Sampling

To replicate realistic water conditions in terms of salinity, temperature and turbidity while also controlling the presence and amount of Chinook DNA, we combined water samples from a tank containing a high density of juvenile Chinook with an estuarine water sample from a representative location of pre-smolt Chinook habitat in the San Francisco Estuary. The estuarine water biological replicates consisted of 500 mL of surface water taken with a 1 L measuring cup (sterilized by rinsing in 20% bleach solution and then rinsing in DI water) from Deverton Slough, California (38°11'16.7"N 121°58'34.5"W) and collected in a 1 L Nalgene bottle. Next, using another sterile measuring cup, we added 500 mL of tank water known to hold Chinook salmon DNA to each

estuarine water biological replicate. The 680 L tank contained 906 Chinook salmon of approximately 11 cm in length. This mixture allowed us to both control Chinook density and observe similar PCR inhibitor levels as those observed in the estuary. In total, we produced 85 samples, which included a deionized water sample negative control, a tank water only positive control and a Deverton Slough water negative control.

Estimation of average input eDNA

To estimate the average input DNA from the tank water we spiked 10 samples of 1 L surface water from the Deverton Slough, California (38°11'16.7"N 121°58'34.5"W) with varying concentrations of isolated Chinook and green sturgeon (*Acipenser medirostris*) DNA totaling 3 samples with 1 ng/L, 3 samples with 0.1ng/L and 3 samples with 0.01 ng/L for the Chinook salmon samples and 3 samples with 10 ng/L, 3 samples with 1 ng/L and 3 samples with 0.1 ng/L for green sturgeon. We tested for green sturgeon concomitantly to validate the protocol for multiple species and verify that probe specificity and detection limit doesn't affect the DNA yield of the protocol. The tenth sample was not spiked and used as a negative control. Then we filtered the samples using a glass filter, extracted the DNA using the Qiagen DNeasy Blood & Tissue Kit (Cat No./ID: 69504) and removed PCR inhibitors using Zymo OneStep™ (Cat No./ID: D6030). Our serial dilution consisted of the same extracted DNA solution used to spike the samples. We estimated the average yield in percentage for this protocol by qPCR amplification. From the Qiagen protocol average yield we could estimate the average input DNA from the tank water. We also estimated that DNA yield percentage is mildly inverse correlated (p-value = 0.0023) to the initial DNA concentration (Figure 1.S1), while the probability of amplification is logistically correlated to the Log_{10}(initial DNA concentration) (Figure 1.S2).

Filtration

We filtered the samples one day after sampling to simulate real conditions, where it is not always possible to complete filtration on the same day as sampling. In each filtration run, 4 samples of 1 L were filtered in parallel at the speed of 310 rpm on a peristaltic pump and we timed each filtration event. Filters were folded in half 3 times and stored in a 2 mL microcentrifuge tube and stored at -20°C. Between runs, tubing and casing were sterilized using a bath of 20% bleach[20], rinsed twice using DI water to remove any remaining bleach and dried.

Extractions

For all DNA extraction protocols except the dipstick-based ones, we added 180 µL of ATL buffer and 20 µL of 5 U Proteinase K to the microcentrifuge tube and incubated at 56°C overnight using a rotisserie attachment for 2 mL microcentrifuge tubes. As the incubation time step doesn't require labor, we didn't add it to the total time of the protocol. Next, the filter was compressed inside of the microcentrifuge tube using a pipette tip and the supernatant was transferred to a clean 0.5 mL (NaOH extraction) or 1.5 mL microcentrifuge tube (magnetic beads and Qiagen).

NaOH-based extraction

For each 100 uL of supernatant we added 5.26 µL of 1M NaOH. In a benchtop thermocycler, we incubated the samples at 95°C for 20 min and ramped down the temperature at a pace of 0.7°C/min until reaching 4°C. Next, we added 10% of the total volume of 1M Tris-HCL. Samples were vortexed and centrifuged for 15 min at 4680 rpm. Without disturbing the pellet, 100 µL of the supernatant was extracted and transferred to a new 1.5 mL microcentrifuge tube and stored at -20°C.

Magnetic Beads

For each sample, 180 µL of Agencout AMPure XP (Beckman Coulter™; Cat No./ID: A63881) was added to the solution and incubated at room temperature for five minutes. Then the microcentrifuge tubes were placed onto the magnetic plate (DynaMag™-2; Cat No./ID: 123.21D) for 2 minutes. We removed the supernatant and washed the magnetic beads twice using 200 µL of a freshly made 70% ethanol solution with an incubation time of 30 s in the magnetic plate. We then air-dried the beads for 3 minutes. A total of 100 µL of TE solution was used to resuspend the particles and elute the DNA. The solution was incubated for 1 minute at room temperature before pulling down the magnetic beads with the plate for 2 minutes. Lastly, the supernatant was transferred to clean 1.5 mL microcentrifuge tubes.

Qiagen DNeasy cell and tissue

The Qiagen DNeasy extraction was performed following the manufacturer's recommendations. A total of 200 µL of AL buffer was added and the samples were incubated at 56°C for 10 minutes. We added 200 µL of ethanol to each sample and the solution was transferred to the column and centrifuged at 8000 rpm for 3 minutes. Then the column was washed using 500 µL of Wash Solution N°1 and centrifuged at 6000 rpm for 1 minute. Then the column was washed again with Wash Solution N°2 and centrifuged for 3 minutes at 1400 rpm. Next, 100 µL of AE solution was added and incubated for 20 minutes before centrifuging at 8000 rpm for one minute. The flowthrough was then stored at -20°C

Whatman paper dipstick

The Dipstick method was tested due to its cost efficiency, short processing time, and the fact that it allows for eDNA extraction in the field at the expense of DNA yield. Dipsticks were made following the protocol described in [10]. We used the qualitative Whatman filter n°1 to make our dipsticks and used an effective surface area of 8 mm^2 (2 mm width and 4 mm height). We added 200 µL of lysis buffer and ground the filter using a pipette tip until the filter was dissolved. Then we dipped the dipstick in the lysis buffer solution (20 mM Tris [pH 8.0], 25 mM NaCl, 2.5 mM EDTA, 0.05% SDS) 3 times, then dipped 3 times in 100 µL of wash solution (10mM Tris [pH 8.0], 0.1% Tween-20), and 3 times in a final solution of nuclease free water which then was stored at -20°C. In the case of "straight to qPCR" dipstick extraction, we directly dipped the dipstick after the wash step into the qPCR reaction. An advantage of the "straight to qPCR" dipstick is that, by skipping an elution step, we avoid dilution of the DNA bound to the dipstick. DNA captured on the dipstick is at a higher concentration than what would be eluted in nuclease free water. Though this method avoids a dilution step, PCR inhibitors will also be at a higher concentration and could lead to variable results from multiple qPCR reactions from the same sample.

Secondary Inhibitor removal

Zymo OneStep™ PCR Inhibitor Removal Kit (Zymo Research; Cat No./ID: D6030) was used following the manufacturer's protocol to remove any carryover PCR inhibitors from previous steps. We added 600 µL of Prep-solution to the column and centrifuged at 8000 g for 3 minutes, the flow-through was discarded, then 50 µL of DNA elute from previous steps were added to the column and centrifuged at 16000 g. The flowthrough was then stored at -20°C.

qPCR amplification oligos

Quantitative PCR detection for Chinook salmon was developed by adapting the protocol from[21]. Reaction solution totaling 20 µL was composed of 1× TaqMan™ Environmental Master Mix 2.0 (ThermoFisher Scientific; Cat No./ID: 4396838), 0.9 µM concentration of each primer, and 0.7 µM of the Taqman probe, and 6 µL isolated DNA extract from previous steps. The chosen primers, probes and Gblock were designed following[21] and are shown in Table 1.1. Thermocycling was performed on a Bio-Rad CFX96 real-time detector using the following profile: 10 min at 95°C, 40 cycles of 15s denaturation at 95°C and 1 min annealing–extension at 60°C.

Table 1.1: List of DNA oligonucleotides used in this study.

Oligonucleotide	Sequence
Probe sequence	FAM–5'–AGCACCCTCTAACATTTCAG–3'–ZEN/Iowa Black
Forward primer	5'–CCTAAAAATCGCTAATGACGCACTA–3'
Reverse primer	5'–GGAGTGAGCCAAAGTTTCATCAG–3'
Gblock sequence	5'–ACCATCGTTGTTATTCAACTACAAGAACCT AATGGCCAACCTCCGAAAAACCCATCCTCT CCTAAAAATCGCTAATGACGCACTAGTCGA CCTCCCAGCACCCTCTAACATTTCAGTCTG ATGAAACTTTGGCTCACTCCTAGGCCTATG TTTAGCCACCCAAATTCTTACCGGGCTCTT CTTAGCCATACACTATACCT–3'

Primers were used for DNA amplification. Probe was used for the qPCR step for DNA quantification. Gblock was used for creating a standard ladder for the qPCR reaction and made possible the conversion from initial DNA concentration to copy number.

Data analysis

Data analysis was performed in Python 3.7 and the analysis pipeline is available on https://github.com/sanchestm/eDNA-Protocol-Optimization. We measured interference between filter type and extraction method using two competing models, one that includes the interference effect and one that does not. Using ADVI inference we fitted the data to the models[19]. From the ADVI fitting for the best model we estimated the distribution of filter eDNA yield percentage, extraction eDNA yield percentage and PCR inhibitor carryover for filtration and extraction. To estimate which step of an eDNA experiment has the most variance between methods, and therefore can lead to the most significant gains when optimized, we trained a random forest regressor[22] with the collected data and estimated importance of each step of the experiment.

Results

Effect of filter type on DNA extraction yield

We first examined different methods for the initial two steps of an eDNA protocol, filtration and DNA extraction, and tested for interference between these steps. In general, when optimizing a protocol consisting of several steps, it is important to identify if previous steps interfere with the effectiveness of subsequent steps. In our study, the main possible interference is between the filter used and the extraction method. The yield percentage of a certain extraction method could change depending on which filter was used. Possible reasons for interference between filter and extraction method include different particles binding differentially to filters and extraction methods not isolating DNA from all types of particles at the same yield percentage. The models that we tested are the following:

Model with interference:

$$Y \sim Y_f\left(Y_e - interference(F)\right) - I_f I_e \begin{cases} 0 \text{ , if secondary inhibitor removal is used} \\ 1 \text{ , otherwise} \end{cases} \quad (2)$$

Model without interference:

$$Y \sim Y_f Y_e - I_f I_e * \begin{cases} 0 \text{ , if secondary inhibitor removal is used} \\ 1 \text{ , otherwise} \end{cases} \quad (3)$$

For both widely applicable information criterion (WAIC) and Leave-one-out cross validation (LOO) the model without interference was selected with a weight of 1 in both cases[10,23]. Other evidence for the absence of interactions is that the ranking order of filter yield (1st *Cellulose nitrate - 2nd Glass fiber - 3rd Filter paper N1)* does not change independently of which extraction method is chosen (Figure 1.2A). Similarly, the yield ranking for extraction methods is not affected by filter choice. (Figure 1.2B). Only the NaOH method breaks the independence rule for the nitrocellulose filter. In this case, target DNA could not be amplified from NaOH extractions without secondary inhibitor removal, resulting in an upwards skewed average of the DNA yield as the samples without secondary inhibitor removal were not taken into account.

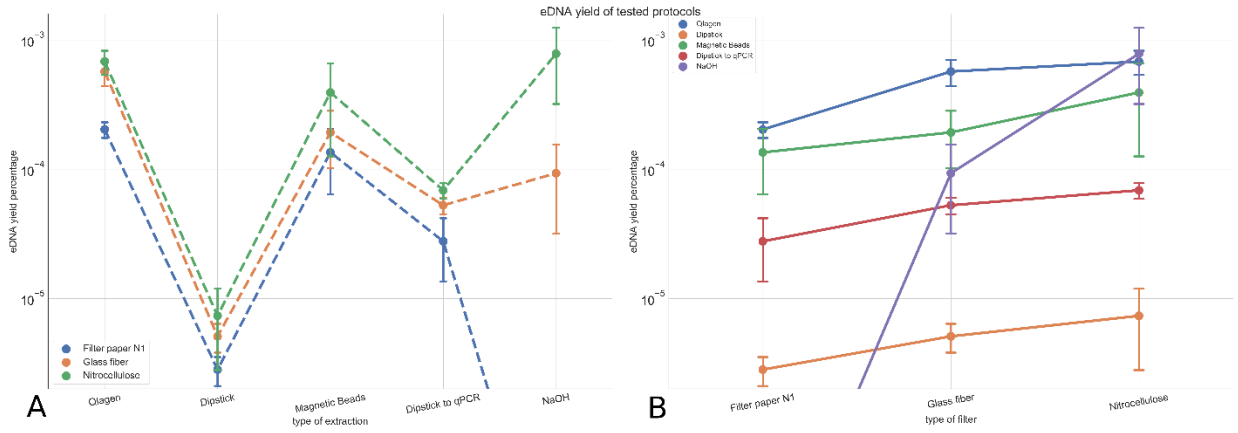


Figure 1.2: Relationship between total DNA yield and filtration and extraction protocols. (A) No crossing between lines indicates that on average, DNA yield ranking for the filters is independent of the extraction protocol. Error bars represent the 95% confidence intervals. The combination of Whatman filter and NaOH extraction wasn't able to amplify the target Chinook DNA. (B) No crossing between lines indicates that on average, DNA yield ranking for the extraction method is independent of the filter type. The NaOH extraction protocol is the only case where the ranking order is not maintained and can be explained by the added effects of carry-on inhibitors.

DNA retention and filtration time per filter type

Next, we compared DNA yields from three different filters. The nitrocellulose filter outperformed the glass fiber filter in terms of DNA yield by 1.6 times and the Whatman n°1 filter by 3.75 times on average (Figure 1.3). The percentage of captured eDNA copies were 0.00045 ± 0.00013, 0.00107 ± 0.00013, 0.00172 ± 0.00013 for paper filter N1, glass fiber, and nitrocellulose respectively. In other words, 1.6 L and 3.75 L of water would need to be filtered through a glass fiber filter or Whatman n°1 filter, respectively, to isolate the same amount of DNA as filtering 1 L of water through a nitrocellulose filter. However, the glass filter outperforms the nitrocellulose and Whatman filters in terms of filtration time, with the glass filter not only being drastically faster but also more consistent and resilient to variations in turbidity (Figure 1.4). Filtration times were 2.32 ± 0.08 min, 14.16 ± 1.86 min and 6.72 ± 1.99 min for glass fiber, nitrocellulose and paper filter N1, respectively.

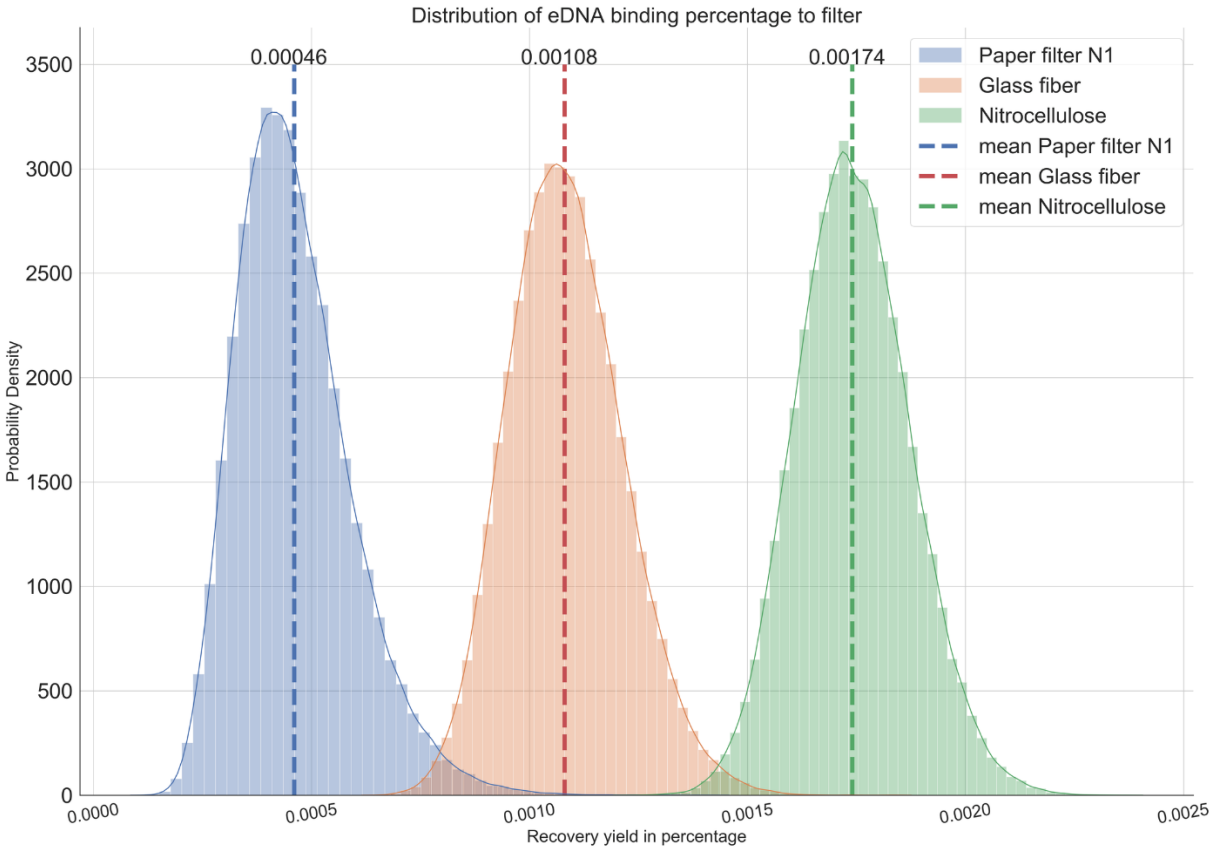


Figure 1.3: Distribution of DNA capture ratio for each filter type from the Automatic Differentiation Variational Inference model. The broadness of the curve shows the variability of the ratio of the input DNA that binds to the filter. The peak of each distribution is the mean yield ratio of DNA recovery for that filter type. The nitrocellulose filter yielded the highest recovery ratio with little efficiency overlap compared to glass and Whatman filters.

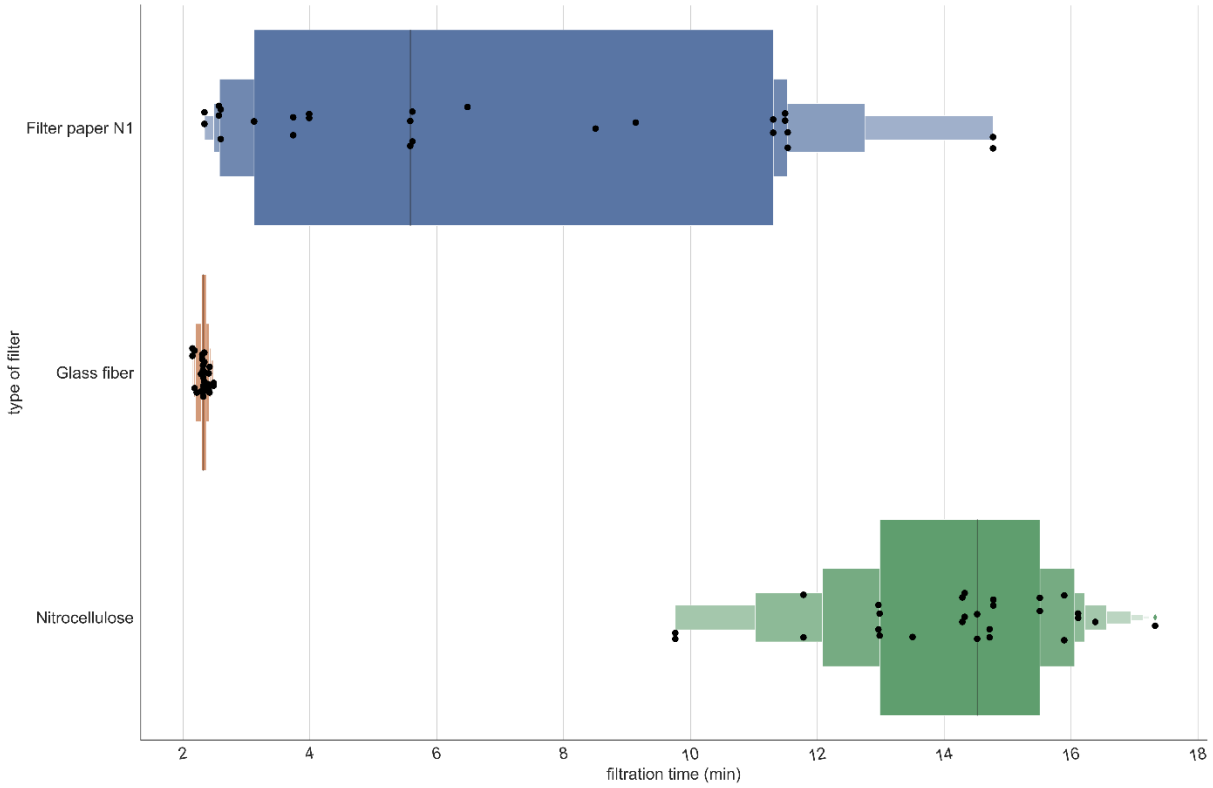


Figure 1.4: Percentiles and median filtration time in order to filter 1 L of estuarine water for each filtration method. The glass filter outperforms nitrocellulose and Whatman by a significant margin in terms of average filtering time and consistency in the filtering time. Dots are filtration events while the black line represents the median value filtering time. Boxes indicate 10% quantiles.

Comparison between DNA extraction protocols

All extraction methods could yield enough eDNA to be detectable by qPCR amplification. The Qiagen DNEasy kit had the highest DNA yield, outperforming NaOH by 1.7 times, magnetic beads by 2.26 times, direct to qPCR dipsticks by 9.71 times and regular dipsticks by 358 times (Figure 1.5). Observed DNA extraction yields in percentage were 0.475 ± 0.036, 0.047 ± 0.037, 0.287 ± 0.037, 0.206 ± 0.037, 0.132 ± 0.053 for QIagen, dipstick, NaOH, magnetic beads, direct dipstick, respectively. At the same time, the Qiagen kit is by a considerable margin the most time-

consuming method, requiring 77 minutes to process 18 samples. In contrast, the direct to qPCR dipstick approach was the fastest and most cost-efficient method by a wide margin.

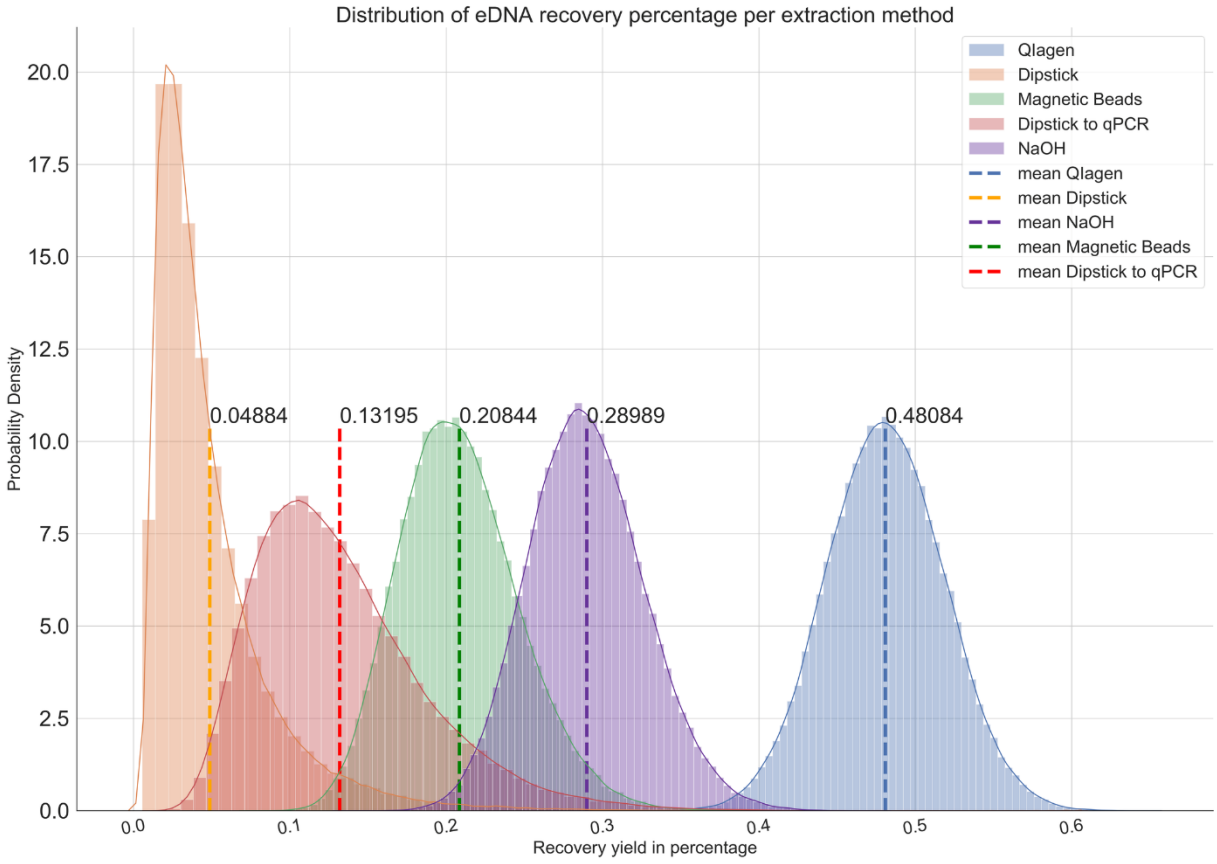


Figure 1.5: Modelled distribution of percentage yield for each extraction protocol. The width of each of the curves shows the variability in modelled yield. The peak of the distribution is the mean yield per extraction type. Qiagen DNeasy yields the best yield with little overlap with other methods. Meanwhile magnetic beads and NaOH have shown similar distributions with significant overlap, while both dipstick methods underperform the other methods. It is important to note that this plot does not take into account PCR inhibitor carryover, which might vary significantly between methods.

Importance of each protocol step on total DNA yield

The extraction method was shown to be the most influential factor for the eDNA yield from the random forest aggressor analysis (Figure 1.6). Therefore, further optimization experiments should focus on this step, experimenting with different protocols to extract the eDNA to maximize protocol eDNA yield. Meanwhile, in the context of our experiments, the removal of inhibitors was shown to have little impact to the total DNA yield estimated by qPCR, although published data[7] have shown that inhibitor removal highly influences the amplification probability of the qPCR reaction.

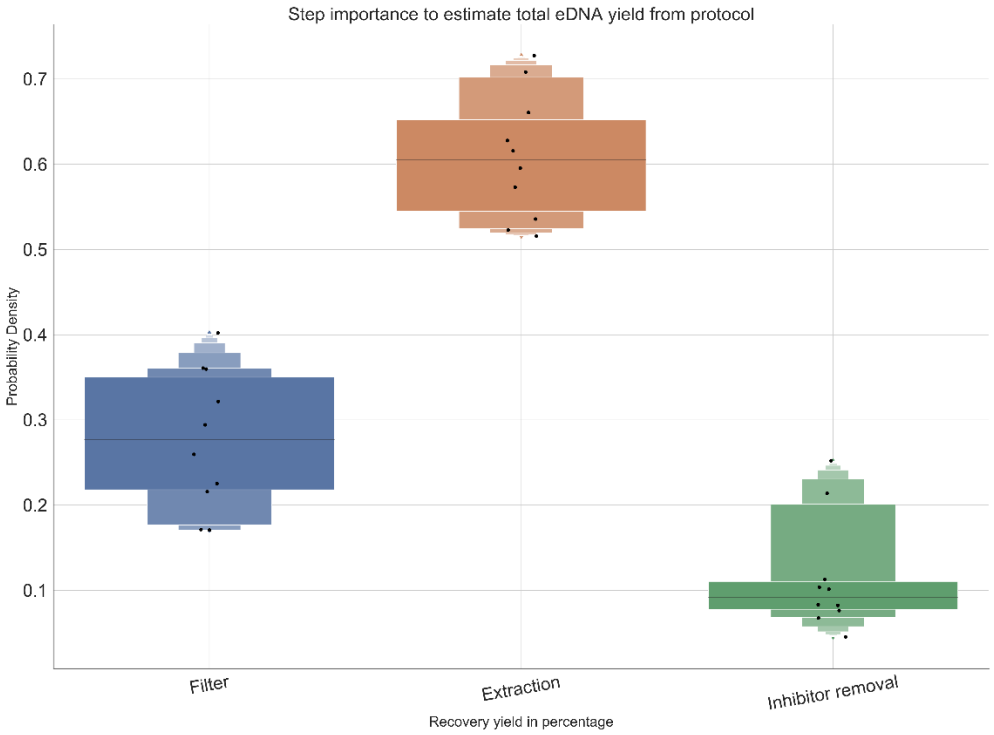


Figure 1.6: Influence of each eDNA protocol segment to total eDNA yield ratio estimate. The extraction method was the factor that had the highest influence on the total eDNA yield of the total protocol, while inhibitors didn't have a significant impact compared to the other components of the total protocol.

Conditions in which the use of SIR is necessary

We observed that the secondary inhibitor removal step always outperformed skipping this step. Regressions from Figs 1.7A and 1.7C were always positive and the distributions from Figs 1.7B and 1.7D were always greater than zero. The nitrocellulose filter and the NaOH extraction were the methods that carried the most PCR inhibitors, while the other methods for each step showed a high overlap of their carryover inhibitor distributions. Secondary inhibitor removal was essential to observe any amplification using the NaOH extraction method, which also suggested that this method is inefficient at removing PCR inhibitors.

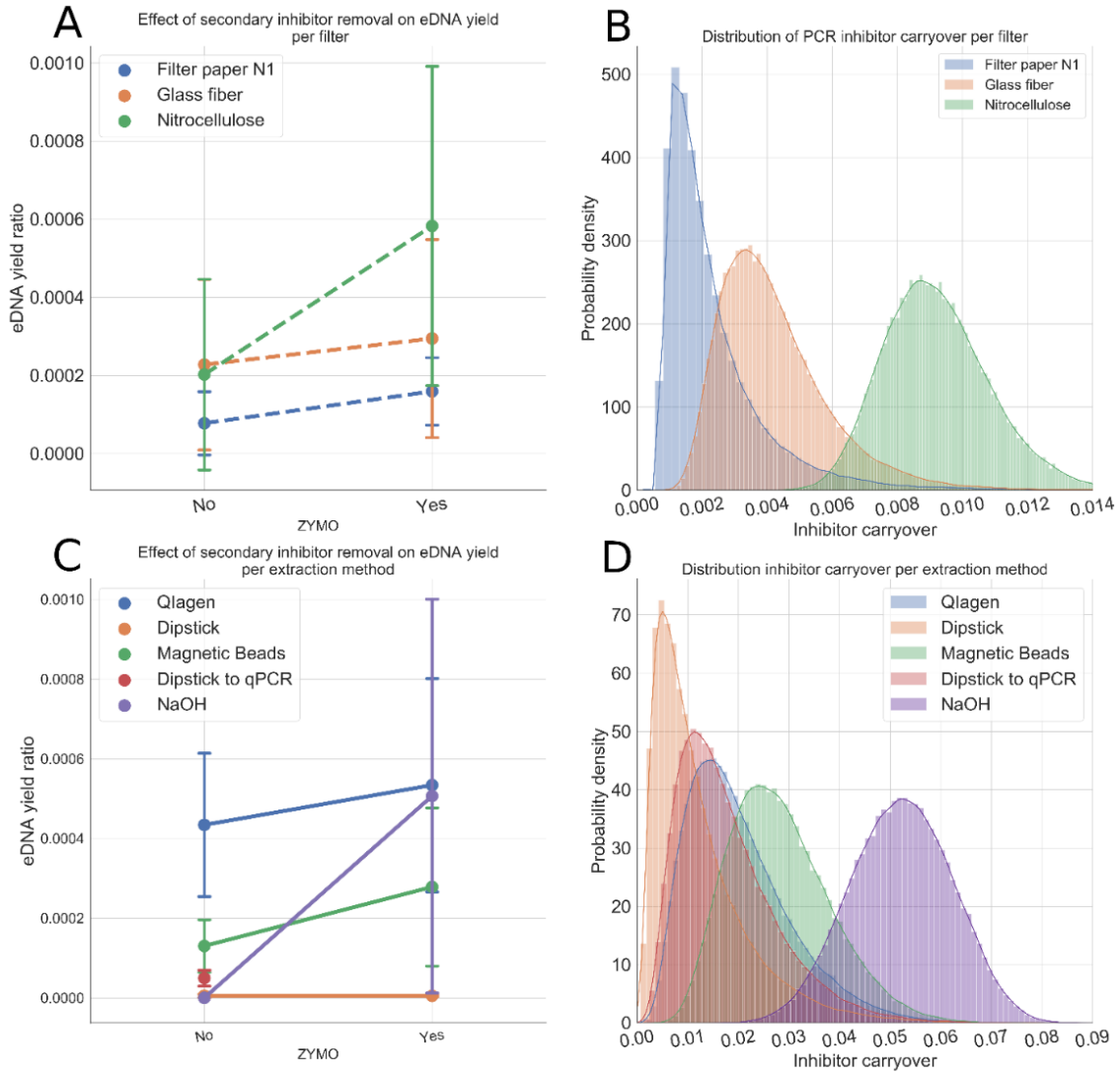


Figure 1.7: Effects of adding a secondary inhibitor removal step to the eDNA estimation protocol. (A-B) DNA yield variation from using a OneStep PCR Inhibitor Removal©. The nitrocellulose filter produced the highest inhibitor carryover levels at the same time it captured the highest percentage of free eDNA. This suggested that the nitrocellulose filter captured particulates indiscriminately with high efficiency (C-D) Estimated distributions for inhibitor carryover for filter and extraction method. Aside from NaOH extraction, other methods had similar distributions of carryover PCR inhibitors with high overlap. Therefore, NaOH extraction, even if it has an elevated eDNA yield, doesn't properly address the high levels of PCR inhibitors commonly encountered in environmental samples.

Discussion

As the eDNA field has progressed and we've learned more about how to detect DNA under different environmental conditions, the number of methods for each step of an eDNA assay test has exponentially increased. Due to the wide array of options, identifying the best protocol given the environment and the target species of the survey is complex. Several publications describe DNA yield comparisons between different methods[10,24–26]. We chose to test nitrocellulose, glass fiber and paper filter N1, as in the literature nitrocellulose consistently presents the highest DNA yield[27,28], while glass fiber is the most common filter material[2,29] and paper filter had shown to have high DNA binding[10] while it's cost is a fraction of other materials. The chosen pore size for glass fiber was chosen based on[27], while nitrocellulose pore size was chosen relative to the most common pore size[30], while paper filter pore size was chosen to maximize flow. We opted to test more extraction methods than filter material as we expected that most filter materials capture

particles indiscriminately in a similar fashion, while extraction methods do vary significantly on the mechanics of the protocol.

One of the novel observations from our experiment is that each step of an eDNA assay seems to work independently of the previous step. The lack of interference between steps shows that for future optimization tests, it may not be necessary to test all the possible combinations of filters and extraction methods at the same time. Instead, one might test each section of the protocol independently and still obtain an optimal eDNA assay protocol. This would allow more methods to be tested for each step and increase the number of replicates for each method in future optimization experiments. The observed low levels of interaction between steps also permits a better comparison to the literature. As an example, both this paper and [27] test approximately 1 µm glass fiber filtration. In [27], the 1 µm filter retained the most DNA, while the nitrocellulose filter had a better DNA retention than the 1.6 µm glass fiber filter in our experiment. Even though other steps of the assay might differ between the publications, we can still conclude that nitrocellulose outperforms glass fiber in any mesh size tested in [27].

Estuaries possess unique physical characteristics and biodiversity of major importance for both marine and riverine ecosystems. The Chinook salmon is a perfect example of a species in which the understanding of its annual out migration patterns in these estuaries are essential for their preservation. To understand the migration patterns of Chinook Salmon, we optimized an eDNA protocol for estuaries. For estuarine conditions, we concluded that glass fiber filters are the most efficient because estuaries possess high turbidity, which clogs the other tested filters. The high filtration time for the Whatman and nitrocellulose filters hinders the throughput of the sampling, possibly leading to under sampling in the total eDNA survey. For the DNA extraction, we considered magnetic beads to be the optimal method for estuarine waters, as dipstick methods

had subpar DNA yield while Qiagen DNeasy did not allow high throughput sampling due to cost and processing time. Even though Ampure XP has an elevated cost per sample, we are able to reduce this cost 100-fold by making a magnetic beads solution in-house[31]. Buying in bulk is also another alternative to reduce costs, though that might be limited to the initial funding of the project.

Our experiments have shown that the use of secondary inhibitor removal has little influence on total eDNA yield. This observation may be explained by several nonexclusive factors. First, inhibitors generally work by binding to DNA strands and do not act as catalysts. Therefore, if the ratio between eDNA to inhibitors is significantly elevated, which is expected in tank experiments, we would predict minimal effects of the inhibitors. Another possibility is that during the time we were sampling, there were fewer inhibitors than usually observed in estuaries. Furthermore, filtration and DNA extraction methods vary considerably in DNA yield, more than the observed effect from SIR. Lastly, PCR inhibitors might not affect the eDNA retrieval but only the probability of amplification. This last observation might also explain why the probability of amplification and DNA yield is not always fully correlated. Therefore, considering this experiment's results and previous findings[32,33], we advise the use of secondary inhibitor removal, if possible, as it improves the DNA yield and amplification probability in the context of estuarine samples.

Our study suggests that in most cases, using a glass fiber filter and magnetic beads would be the most practical method to generate the maximum amount of information obtained about fish distribution. We also concluded that DNA extraction from the filters is the most time-consuming step and most variable in terms of efficiency. Therefore, this is the step that should be decided with utmost care in order to maintain the high-throughput and useful detection limit of the desired methodology. For this reason, magnetic beads DNA extraction is a promising alternative to silica

column extraction, as this method strikes the balance between yield, amplification probability, carryover PCR inhibitors and time to process samples. Meanwhile, the cost of using magnetic beads can be mitigated by developing the necessary reagents in-house.

However, different protocols may yield better results under certain circumstances, such as when the target species is extremely rare or ubiquitous. In order to choose the best protocol, we constructed a simple decision tree for those scenarios (Figure 1.8). We also ranked the protocols, sorting them by DNA yield, which should be the main parameter for the protocol selection. Then, given an eDNA assay sensitivity cutoff, we can choose a protocol that is fast and cost-effective (Figure 1.9). To choose the optimal eDNA assay protocol, prior information about the target species' occupancy rates in the locations that will be sampled is beneficial. Even though detecting occupancy is also the goal of a species survey, having this prior information will help managers decide how sensitive the eDNA assay must be in order to have high confidence in negative results for a given site[34]. If there is no prior information about the species distribution in the region, we would assume a uniform distribution of the species. Otherwise, if prior information shows a species to be common, faster, and less sensitive tests can be used. Meanwhile, if the species is rare, one should avoid false negatives and therefore must use the most sensitive protocol. This is the most crucial question for the project design since the least desirable outcome is for the species to be completely undetected when there is a high false-negative rate. Another undesirable outcome is under sampling the study region. In those cases, even if there is high confidence in the species' presence at a given site, those sites may not be representative of the study region, therefore leading to under or overestimation of the species distribution. Sampling more sites or having higher confidence in a given site is the most complicated choice when designing an eDNA survey. Due to the impact of sampling size on the conclusions of a survey, we consider processing time to be

the second main factor when choosing an eDNA assay protocol. Lastly, we consider cost to be the third main factor. In cases where sensitivity is not a major factor, we can use cheaper assays to largely improve the sampling effort, providing robustness to the findings of the survey.

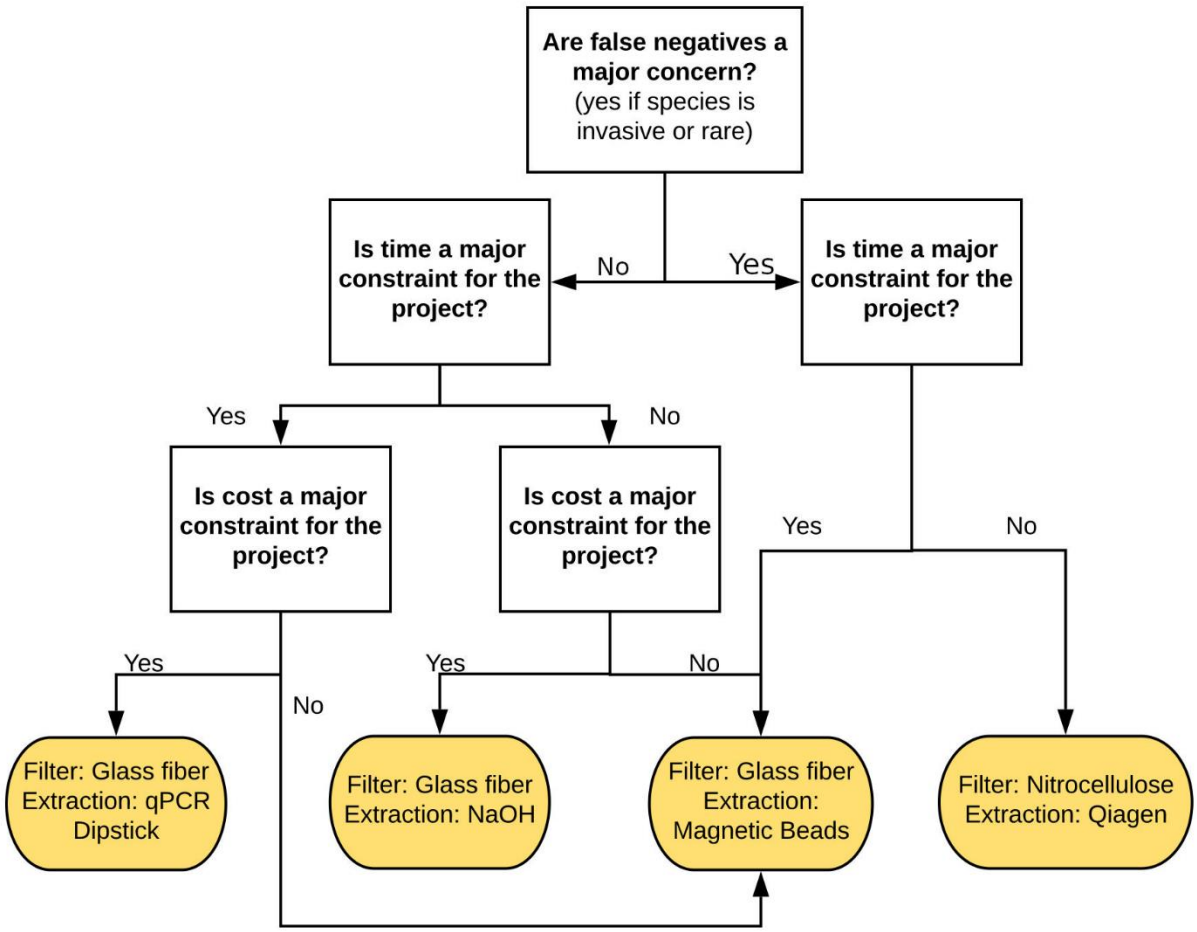


Figure 1.8: Decision tree for choosing the protocol which will yield the most information given research constraints. A glass filter is recommended in most cases, if the focus isn't maximizing DNA yield with no time or cost constraints. Magnetic beads also are advised in general for its balance between DNA yield and time to process the samples, while cost can be mitigated by producing magnetic beads solution in-house (~$0.55/mL) instead of buying Ampure XP ($15–$70/mL)[31].

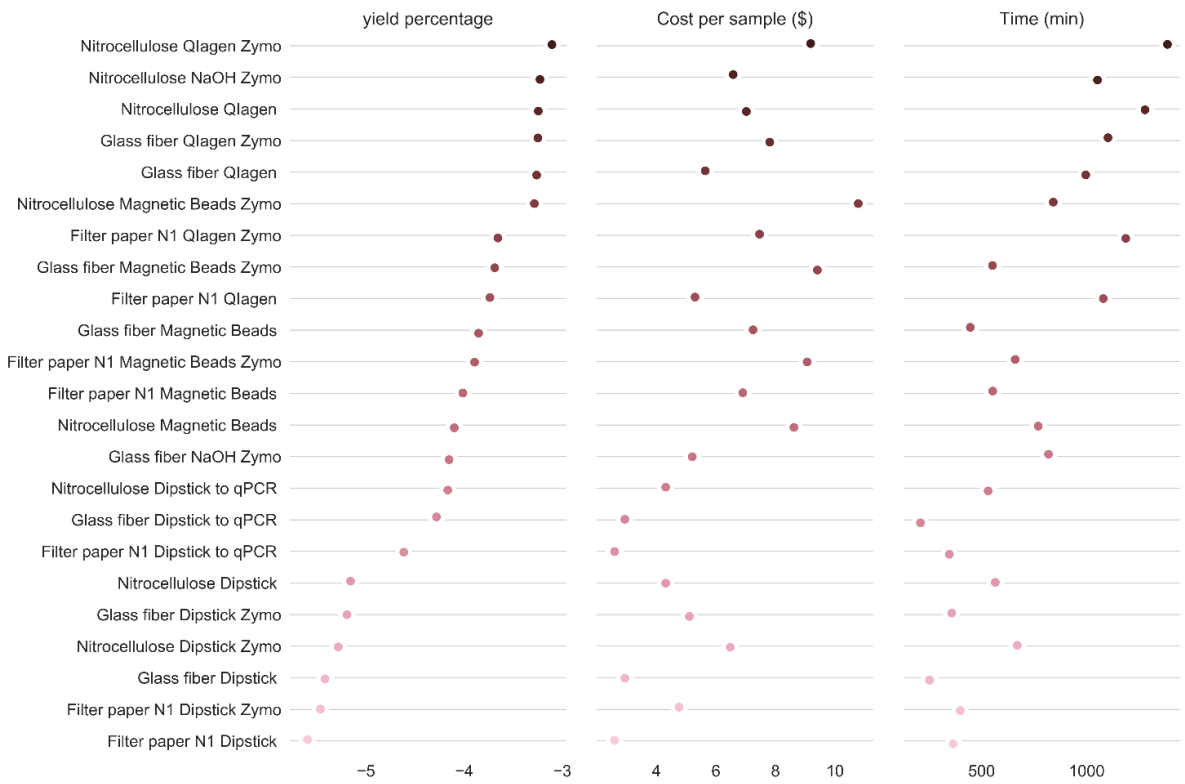


Figure 1.9: Comparison between eDNA protocols for DNA yield, cost and time to process 96 samples. Methods were sorted by yield and shown in $\log_{10}$ scale.

References

1. Kelly, R. P. *et al.* Environmental monitoring. Harnessing DNA to improve environmental management. *Science* **344**, 1455–1456 (2014).
2. Deiner, K., Walser, J.-C., Mächler, E. & Altermatt, F. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation* vol. 183 53–63 (2015).
3. Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. *Environmental DNA: For Biodiversity Research and Monitoring*. (Oxford University Press, 2018).
4. Nielsen, K. M., Calamai, L. & Pietramellara, G. Stabilization of Extracellular DNA and Proteins by Transient Binding to Various Soil Components. in *Soil Biology* 141–157.
5. Matheson, C. D., Gurney, C., Esau, N. & Lehto, R. Assessing PCR Inhibition from Humic Substances. *Open Enzym. Inhib. J.* **3**, 46–52 (2014).
6. Hunter, M. E. *et al.* Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive burmese pythons. *PLoS One* **10**, e0121655 (2015).
7. Williams, K. E., Huyvaert, K. P. & Piaggio, A. J. Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLoS One* **12**, e0179282 (2017).
8. Thomas, A. C. *et al.* A system for rapid eDNA detection of aquatic invasive species. *Environmental DNA* (2019) doi:10.1002/edn3.25.
9. Lacoursière-Roussel, A., Côté, G., Leclerc, V. & Bernatchez, L. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *Journal of Applied Ecology* vol. 53 1148–1157 (2016).
10. Zou, Y. *et al.* Nucleic acid purification from plants, animals and microbes in under 30 seconds. *PLoS Biol.* **15**, e2003916 (2017).
11. Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* **113**, 1014–1026 (2012).
12. Gregory, J. B., Litaker, R. W. & Noble, R. T. Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of enteroviruses in environmental samples. *Appl. Environ. Microbiol.* **72**, 3960–3967 (2006).
13. Huggett, J., Nolan, T. & Bustin, S. A. MIQE: Guidelines for the Design and Publication of a Reliable Real-time PCR Assay. *Polymerase Chain Reaction: Theory and Technology* (2019) doi:10.21775/9781912530243.12.
14. Piggott, M. P. Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecol. Evol.* **6**, 2739–2750 (2016).
15. Schoellhamer, D. H. Influence of salinity, bottom topography, and tides on locations of estuarine turbidity maxima in northern San Francisco Bay. in *Proceedings in Marine Science* 343–357 (2000).
16. Vélez-Espino, L. A. *et al.* Relative importance of chinook salmon abundance on resident killer whale population growth and viability. *Aquat. Conserv.* **25**, 756–780 (2014).
17. Yoshiyama, R. M., Fisher, F. W. & Moyle, P. B. Historical Abundance and Decline of Chinook Salmon in the Central Valley Region of California. *N. Am. J. Fish. Manage.* **18**, 487–521 (1998).
18. Yoshiyama, R. M. A History of Salmon and People in the Central Valley Region of California. *Rev. Fish. Sci.* **7**, 197–239 (1999).
19. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. (2016) doi:10.7287/peerj.preprints.1686.
20. Wilcox, T. M. *et al.* Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. *Biol. Conserv.* **194**, 209–216 (2016).
21. Brandl, S. *et al.* Ten real-time PCR assays for detection of fish predation at the community level in the San Francisco Estuary-Delta. *Mol. Ecol. Resour.* **15**, 278–284 (2015).
22. Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **63**, 3–42 (2006).
23. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2016).

24. 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* **9**, e114329 (2014).
25. Boström, K. H., Simu, K., Hagström, Å. & Riemann, L. Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnology and Oceanography: Methods* vol. 2 365–373 (2004).
26. Laramie, M. B., Pilliod, D. S., Goldberg, C. S. & Strickler, K. M. Environmental DNA sampling protocol - filtering water to capture DNA from aquatic organisms. *Techniques and Methods* (2015) doi:10.3133/tm2a13.
27. Turner, C. R. *et al.* Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution* vol. 5 676–684 (2014).
28. Hinlo, R., Gleeson, D., Lintermans, M. & Furlan, E. Methods to maximise recovery of environmental DNA from water samples. *PLoS One* **12**, e0179251 (2017).
29. Jerde, C. L., Mahon, A. R., Lindsay Chadderton, W. & Lodge, D. M. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters* vol. 4 150–157 (2011).
30. Spens, J. *et al.* Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution* vol. 8 635–645 (2017).
31. DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743 (1995).
32. Tillotson, M. D. *et al.* Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biological Conservation* vol. 220 1–11 (2018).
33. McKelvey, K. S. *et al.* Sampling large geographic areas for rare species using environmental DNA: a study of bull trout Salvelinus confluentus occupancy in western Montana. *J. Fish Biol.* **88**, 1215–1222 (2016).
34. Trevethan, R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Frontiers in Public Health* vol. 5 (2017).

Supporting information

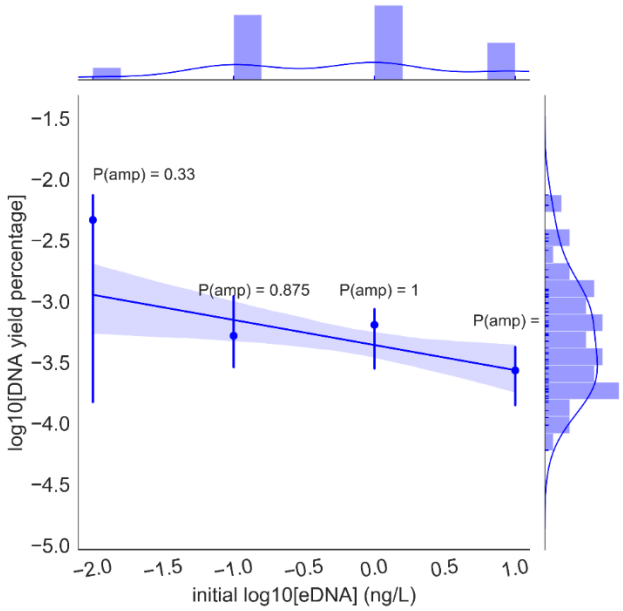


Figure 1.S1: Correlation between DNA yield and initial eDNA concentration. Blue dots - median value; vertical lines - 95% CI; horizontal line - linear regression between protocol DNA yield and input DNA, with both axes being represented in $\log_{10}$ scale. P(amp) - probability of amplification.

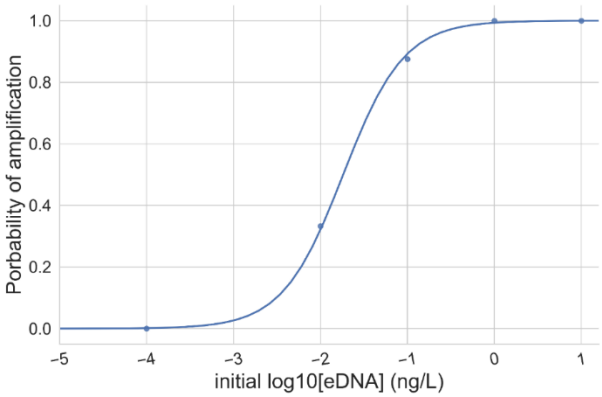


Figure 1.S2: Probability of amplification as a function of the input DNA concentration. Dots - probability of amplification from DNA spiking experiment; line - logistic fit to data points.

eDNA transport and novel sampling opportunities

Abstract

The use of environmental DNA (eDNA) to monitor species in aquatic environments has rapidly increased over the past decade. eDNA often outperforms other methods of detection, yet eDNA relies on an indirect measure to estimate the real distribution of a species. Therefore, understanding the environmental factors that disperse eDNA is of major importance. Here we used a one-dimensional reactive-diffusion-advection eDNA transport model to describe the influence of river advection on the distribution of eDNA produced by one or more eDNA sources (i.e., target species). We then used the model to explore how sampling along transects of different lengths affects the probability of detecting eDNA under different transport velocities. At similar sampling efforts, both higher advection and longer transect lengths increased detection probability by allowing water samples to integrate more potential eDNA sources. At the same time, the mixing of water from different locations effectively homogenized eDNA concentration across space, reducing the ability to precisely estimate the location of target species. Longer transects had a higher detection probability than shorter transects, yet even short transects (less than 250 m) can yield considerable benefits compared to discrete sampling. Our model suggests long transects should be used when detection of rare species is of paramount importance, while short transects are preferred when knowledge of the location of target species is important. In conclusion, we identify transect sampling as an exciting alternative method of eDNA sampling with benefits that may surpass the disadvantages of not being able to pinpoint the exact source location.

Introduction

A lingering concern about environmental DNA (eDNA) based surveys is that sampling for eDNA is an indirect sampling method, and the distribution of eDNA may not reflect the actual distribution of the target species. Multiple environmental processes influence eDNA degradation, interfere with eDNA detectability, dilute eDNA concentration (hereafter [eDNA]), or transport eDNA away from the location where it is produced. River velocity (m/s), turbulent dispersion, substratum and vegetation influence eDNA transport[35]. Microbial activity and UV irradiation impact the degradation of eDNA[36,37]. Suspended solid particles and PCR inhibitors in the environment reduce the amount of eDNA that can be extracted or detected from environmental samples[38]. Therefore, for proper design and evaluation of eDNA surveys in aquatic environments, it is important to understand how environmental processes distribute eDNA for a particular species' distribution.

Most eDNA surveys involve sampling water at a particular time and location. The results of such a sampling scheme reflect the presence of eDNA at that specific point in space and time, which is later used to generalize the results of the survey to the whole system. Like any survey method, eDNA detection has biases. Variations in how eDNA interacts with the environment add noise to the measurements of eDNA, leading to bias. During the survey design, the sampling strategy and sampling locations are chosen in a way to maximize the signal to noise ratio, where the signal is the true species distribution, and the noise is any factor that causes the eDNA measurement to deviate from that species distribution. One way to reduce measurement noise is to average multiple measurements, which can be done over space[39,40], time[41–43] or by adding more replicates to each measurement[44,45]. In net fishing, one method to reduce spatial noise is to use tow nets instead of cast nets (also called throw nets), since tow nets sample a higher volume of water and pass over more microenvironments. This type of sampling is known to have a higher catch on

average than cast net sampling[46]. Analogously, we can imagine that the same process could be applied to eDNA studies. Collecting water along a transect rather than at single points would increase the spatial scope of eDNA detection (defined by the detectable amplification of the target eDNA). Applying the concept of transect sampling to the eDNA survey methodology is a novel approach to improve detection rates, diminish sampling variability, and better identify global population trends, especially when the goal is to characterize large water bodies (Fig. 2.1). High volume water sampling is a form of transect sampling which can yield significantly higher eDNA detection[47], though, it is unclear whether higher detection rates result from higher sampled water volume or the transect sampling strategy. With the model developed in this study, we can estimate if transect sampling can yield better detection rates than discrete sampling with the same effort.

In this analysis, we investigated the effectiveness of transect sampling compared to discrete sampling for eDNA in the context of a river. First, we developed a one-dimensional eDNA transport model to determine how sampling strategy affected detection rates under scenarios of high and low abundance of target organisms in a river system. We then explored the optimal transect length that would provide the most information about the presence and spatial distribution of a target species. Last, we investigated the effects of each model parameter on detection probability to understand the magnitude of each parameter's effects. Our one-dimensional reactive dispersion eDNA model is publicly available for researchers to use to estimate the optimal sampling strategies for their specific target species and locale.

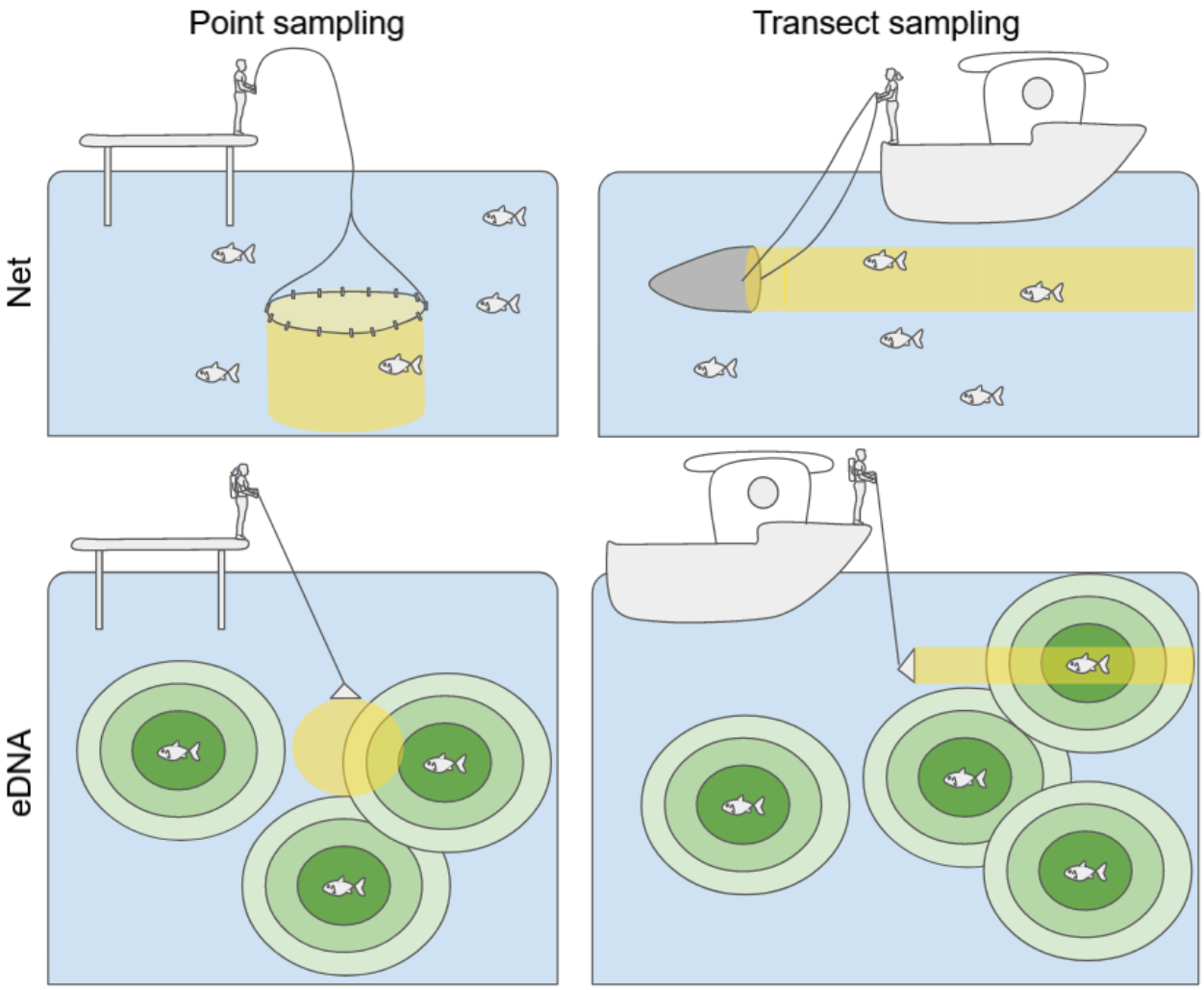


Figure 2.1: Sampling strategies for detecting aquatic species. eDNA sampling provides an indirect means to estimate the target species location while net sampling directly detects the species. We observe that transect strategies determine the average presence of a species over the transect while discrete sampling provides information for a single location.

Materials and Methods

Overview

To compare the detection rates of eDNA discrete sampling compared to eDNA transect sampling, we used a one-dimensional reactive diffusion advection model to simulate the steady state eDNA distributions along a channel under different conditions[48]. This type of model considers the rates in which eDNA is produced, the dispersion of the eDNA via turbulent diffusion and river flow and the degradation/sedimentation of eDNA, then finds how these forces interact to estimate the [eDNA] for every point within the area of interest. We simulated eDNA distributions in the environment for both uniformly distributed and non-uniformly distributed eDNA sources and for different eDNA source densities (i.e., we simulated how the eDNA is spread in the environment for rare and common species and for species that school and species that are randomly present in the surveyed location). Next, we estimated the total amount of eDNA captured by the filter for point and transect sampling for each simulated [eDNA] distribution. The copy number present on a filter can then be used to estimate the probability of detection. This estimation relies on a fitted sigmoid equation between eDNA copy number in a qPCR assay and the probability of detection of that assay[49]. Discrete sampling was modeled as a special case of transect sampling where the sampler velocity tends to 0 (10^{-5} was used in the model to avoid division by 0). Finally, we explored the optimal transect length to provide the most information regarding the presence and spatial distribution of a target species.

In the following sections, we present a series of mathematical equations that we used to model transect and discrete sampling. We do not explain the solutions of the equations in depth because our goal is not to model the eDNA particles themselves but to compare the two sampling approaches. Equations 2.3 and 2.5 are the main pillars of this paper, while equations 2.4, 2.6 and

2.7 are special cases that were chosen for this theoretical exercise. In case the full solution path of the equations is of interest, we refer readers to Wolfram Alpha step-by-step solutions[50] or the lecture slides of the 'Environmental Transport and Fate' course at the Thayer School of Engineering[51] which guided the development of our model.

Considering that sampled volume is commonly constrained by how much water can be filtered before filter clogs, all theoretical comparisons between discrete and transect sampling had the same sampled volume (e.g., discrete sampling was performed on an anchored boat for ten minutes totaling 1l of sample while transect sampling was performed on a moving boat for ten minutes totaling 1l of sample). We also compared the methods with the same sampling effort (e.g., ten discrete samples of 1l are equivalent of ten transect samples of 1l). Last, sampling location in our theoretical river occurred on the same locations for discrete and transect sampling (e.g., if we sampled at locations 10, 100 and 200 with discrete sampling, transect locations were [10-L/2, 10+L/2], [100-L/2, 100+L/2], [200-L/2, 200+L/2]).

Table 2.1: List of parameters, values and units used in the mathematical model.

Parameter	Symbol	Value	Units	Source
Diffusion Coefficient	D	0.01	$\frac{m^2}{s}$	[52]
eDNA degradation rate	λ	$7.33*10^{-4}$	$\frac{1}{s}$	[53]
eDNA production rate	μ	30937.5	$\frac{copies}{Kg * L * s}$	[20]
River cross section area	H	10	m^2	
Pump flow	P	$5*10^{-3}$	$\frac{L}{s}$	
Average River Velocity	V	[0-1.5]	$\frac{m}{s}$	

Transect Length	L		m
Space variable (position in the river)	x		m
Transect Time	T	600	s
Time variable	t		s
Weight of each eDNA source	B	[0.1-1]	Kg
Sampler Velocity	S	]0-0.33]	$\frac{m}{s}$
eDNA abundance	eDNA		copies
eDNA concentration	[eDNA]		$\frac{copies}{L}$
Average distance between eDNA sources	ASD	[5-100]	m
Location of eDNA source	d		m
Effective sampled distance	ESD		M
Zero-Inflation constant	a		$\frac{1}{copies}$

Definition of transect sampling and discrete sampling

In the context of this paper, we consider discrete sampling to be eDNA sampling that occurs at a single point in time and space. Two examples of discrete sampling are: 1) sampling with Nalgene bottles which are then filtered in the lab and 2) filtering on-site at a stationary location (from a dock, a pier, an anchored boat). Meanwhile, transect sampling is defined in our model as on-site filtration where the location that filtration starts is not the location where filtration stops (sampling from a moving boat or walking over the riverbanks) (Figure 2.1). In our model we assume that during transect sampling, the sampler will be moving at a constant speed (boat velocity) and the filtration pump will be turned off at the end of the transect. This type of sampling can be performed with commercial solutions such as the Smith-Root eDNA sampler[54].

eDNA production and degradation

We assumed that eDNA production was proportional to the target organism biomass multiplied by a species-specific shedding rate (*Lepomis macrochirus* in this case; Table 2.1)[40]. Meanwhile, eDNA degradation and settling can be modeled by an exponential decay[55,56]. The variation in eDNA concentration over time can be expressed as the amount of eDNA shed minus the amount of eDNA that has degraded or settled, producing equation 2.1.

$$\frac{\partial[\mathrm{eDNA}]}{\partial t} = B\mu_{(\mathrm{x})}H^{-1} - \lambda[\mathrm{eDNA}] \tag{2.1}$$

eDNA transport

To simulate eDNA transport we will deploy a one-dimensional advection–dispersion–reaction transport model[20,57–59] which dictates that eDNA can move upstream or downstream through turbulent dispersion with a coefficient D and through advection with a river velocity V (equation 2.2).

$$\frac{\partial[eDNA]}{\partial t} = D\frac{\partial^2[eDNA]}{\partial x^2} - V\frac{\partial[eDNA]}{\partial x} \tag{2.2}$$

[eDNA] spatial distribution

Our one-dimensional reactive transport model combines the right-side terms from equations 2.1 and 2.2, generating equation 2.3, which considers transport production and degradation of eDNA in a channel. This general equation can be then applied in specific cases that are solvable. We assume that the eDNA concentration is in a steady state ($\frac{\partial[eDNA]_{(x,t)}}{\partial t} = 0$) with a point source of

eDNA and that we are looking for a solution in the form of an exponential equation, meaning that eDNA concentration decays exponentially from the source[60,61]. Then, we solve for x using the negative exponential solution form $[eDNA]_{(x)} = Ae^{-bx}$, obtaining equation 2.4 for calculating [eDNA] upstream and downstream of the source, respectively. Although the negative exponential solution is cited most frequently in the literature, other equations can also be used to solve equation 2.3 such as the chi-square distribution.

$$\frac{\partial[eDNA]}{\partial t} = B\mu H^{-1} - \lambda[\text{eDNA}] + D\frac{\partial^2[eDNA]}{\partial x^2} - V\frac{\partial[eDNA]}{\partial x} \tag{2.3}$$

$$[eDNA]_{(x)} = \begin{cases} \frac{B\mu H^{-1}}{\sqrt{V^2+4D\lambda}} e^{\frac{V+\sqrt{V^2+4D\lambda}}{2D}x}, & if\ x < 0 \\ \frac{B\mu H^{-1}}{\sqrt{V^2+4D\lambda}} e^{\frac{V-\sqrt{V^2+4D\lambda}}{2D}x}, & if\ x \geq 0 \end{cases} \tag{2.4}$$

Sampled eDNA

The total amount of eDNA collected by the filter can be estimated as [eDNA] multiplied by the volume of water filtered. At every point of the transect we will spend $T*L^{-1}$ time, in which we will be able to collect $P*T*L^{-1}$ liters of water at $[eDNA]_{(x)}$ totaling $P*T*L^{-1}*[eDNA]_{(x)}$ copies of the target eDNA on the filter (Table 2.1). The total amount of eDNA collected will then be the sum of eDNA collected for all points within the transect, obtaining equation 2.5, assuming that the pump flow and sampler velocity are constant. Then, substituting $[eDNA]_{(x)}$ in equation 2.5 with the described eDNA distribution from equation 2.4, we generate equation 2.6.

$$eDNA = \frac{PT}{L}\int_{x_0-L/2}^{x_0+L/2}[eDNA]_{(x)}dx \tag{2.5}$$

$$eDNA = \frac{PT}{L}\int_{x_0-L/2}^{x_0+L/2}\frac{B\mu H^{-1}}{\sqrt{V^2+4D\lambda}}e^{\frac{V-\sqrt{V^2+4D\lambda}}{2D}x}dx \tag{2.6}$$

Integration over an exponential function is solvable, so we generate exact solutions for the three possible combinations of initial (x_i) and final positions (x_f) of the sampler (equation 2.7). If the transect begins and ends before encountering the eDNA source ($x_i < 0$ and $x_f < 0$) we obtain the first clause in equation 2.7. Else, if the transect begins and ends after encountering the eDNA source ($x_i > 0$ and $x_f > 0$) we obtain the second clause of equation 2.7. Last, if the transect crosses the eDNA source, the total amount of eDNA captured is equal to sampling from the source to x_i plus sampling from the source to x_f and we obtain the third clause of equation 2.7. Given that integration can be negative if $x_i > x_f$, we take the absolute value obtained from the integral.

$$eDNA = \begin{cases} \left|\frac{PTL^{-1}2B\mu DH^{-1}}{4D\lambda + V\sqrt{V^2+4D\lambda}+V^2}e^{\frac{V+\sqrt{V^2+4D\lambda}}{2D}x}\right|_{x_0-L/2}^{x_0+L/2}\right|, & if\ \ x_0-\frac{L}{2}\ and\ x_0+\frac{L}{2} \le 0 \\ \left|-\frac{PTL^{-1}2B\mu DH^{-1}}{4D\lambda - V\sqrt{V^2+4D\lambda}+V^2}e^{\frac{V-\sqrt{V^2+4D\lambda}}{2D}x}\right|_{x_0-L/2}^{x_0+L/2}\right|, & if\ \ x_0-\frac{L}{2}\ and\ x_0+\frac{L}{2} \ge 0 \\ \frac{PT2B\mu D}{HL}\left|\frac{1}{4D\lambda + V\sqrt{V^2+4D\lambda}+V^2}e^{\frac{V+\sqrt{V^2+4D\lambda}}{2D}x}\right|_{x_0-L/2}^{0} - \frac{1}{4D\lambda - V\sqrt{V^2+4D\lambda}+V^2}e^{\frac{V+\sqrt{V^2+4D\lambda}}{2D}x}\right|_{0}^{x_0+L/2}\right|, & if\ \ x_i * x_f < 0 \end{cases} \tag{2.7}$$

Knowing the solution for both discrete and transect sampling, we can compare the methods for a set of environmental parameters. For all comparisons we assumed a fixed amount of water filtered (3 L) and a fixed time of sampling (600 seconds). For all cases we will center the transect in its midpoint. For example, when measuring the sampled eDNA for a transect of 100 m at point 0, the transect starts at -50 m and ends at 50 m with its midpoint at 0 m.

Sampled eDNA from multiple sources

In cases where multiple eDNA sources are present, the total amount of eDNA sampled was estimated as the sum of the sampled eDNA collected from each of the sources within a range. The range was bounded by the farthest point upstream and downstream from the source that yields a probability of detection of 0.05. In other words, the range includes all the eDNA sources that can minimally add to the total amount of eDNA collected over the transect. With the sampler located at 0 m we obtain the set of source positions in relation to the sampler within the range, here denominated as S. The total amount of eDNA sampled will then be the sum of the collected eDNA for each source position in S as described in equation 2.8.

$$eDNA_{total} = \sum_{d \in S} eDNA(x_0 = -d|P, T, L, H, B, \mu, D, \lambda, V) \quad (2.8)$$

Model assumptions

eDNA is transported from its source in the shape of a tridimensional plume[62,63]. If advection is zero and therefore eDNA transport is limited to turbulent diffusion, the plume's shape will be a sphere. When advection is added, the shape of the plume will be conical, with its cone angle being defined by the ratio between diffusion and advection[64]. If advection is high, the cone angle will be small, meaning that by the time an eDNA particle moves laterally it will be well downstream of the source[64]. In these cases, lateral placement and the depth of the sampling site will greatly influence detection[35,39], especially when sampling close to the source, as the cross section of the eDNA plume will be small compared to the cross section of the river. This possible lower detection rate has been described in [59,61] and called breakout phase. The presence or absence of an observed

breakout phase in an eDNA study will be largely determined by the how sampling has been performed.

Considering that our model is one-dimensional, one of its assumptions is that eDNA is evenly distributed on the river cross section, which is not the case when advection is high. This assumption is broken when advection is significantly stronger than diffusion. Yet, the model would still be applicable if we were describing only the direct path downstream of the source instead of the whole cross section of the river. One way to verify the importance of accounting for lateral movement of eDNA particles is to measure the Péclet number of the river, denoted by $P = \frac{V * river_{width}}{D}$. If the Péclet number is large (>10), advection dominates, meaning that our model will be valid only for the water mass downstream of the source. If the Péclet number is small (<0.1), diffusion dominates, and our model will be valid for most of the river cross section[65].

Our one-dimensional eDNA transport model does not directly model the vertical transport of eDNA. Vertical transport is dependent on diffusion and the speed in which eDNA sinks over time which is primarily driven by gravitational settling[66]. The importance of eDNA settling in rivers will depend on the ratio between the river depth and flow[67]. In shallow rivers, settling importance will increase as it takes less time for the particle to settle. For our one-dimension model, the settling rate is encompassed by the degradation rate constant. Meaning that the 'degradation rate' in the model is compounded by both the settling of eDNA and the degradation of the eDNA. Thus, to maintain the validity of our model, we expect that the lotic system studied does not have variable morphology through its course, and that the degradation rate variable in our model will be a characteristic of the sampled river. Cage studies can be used to estimate degradation rates for each system.

eDNA resuspension is especially important for shallow rivers or rivers with low advection[35,57,60]. Resuspension can especially influence the persistence of eDNA in the environment[68,69], though for our model, since we assume stationary and persistent eDNA sources, the impact of resuspension can be lumped with the other factors accounted for by the degradation rate. The assumption of stationary and persistent eDNA sources also implies that the eDNA distribution reaches a steady state that is constant over time. This model might better fit non-motile species or territorial species. In case the target species does not follow this pattern, we suggest estimating the fish locations as a stable probability density function for species with known habitat preferences. eDNA concentration might also be stratified by depth in the cross section of a river[68,70,71]. Therefore, another requirement for our model is that all sampling must be performed at the same depth of the river to ensure consistency. Lastly, we assume that river advection is constant, and sources are shedding eDNA at a constant rate.

Cases of interest

For our modeling we focused on a few cases of interest. First, we compared sampling with a 500 m transect (T) and discrete sampling (D). Then we compared these two sampling strategies under the condition of little to no advection (LA) and high advection (HA). Last, we examined cases under four species distributions: a single individual (SI), a common species uniformly distributed (CU), a rare species randomly distributed (RNEA) and a common species randomly distributed (CNEA). A case of transect sampling with high advection and uniform species distribution will be denominated as T.HA.CU. If more than one condition is mentioned, we will use the x term instead. For example, when we compare transect sampling and discrete sampling in a high advection case with common uniform species distribution, it is represented as x.HA.CU (Fig. 2.2). Other

parameters are described in Table 2.1 and will be mentioned explicitly in the text if values are changed.

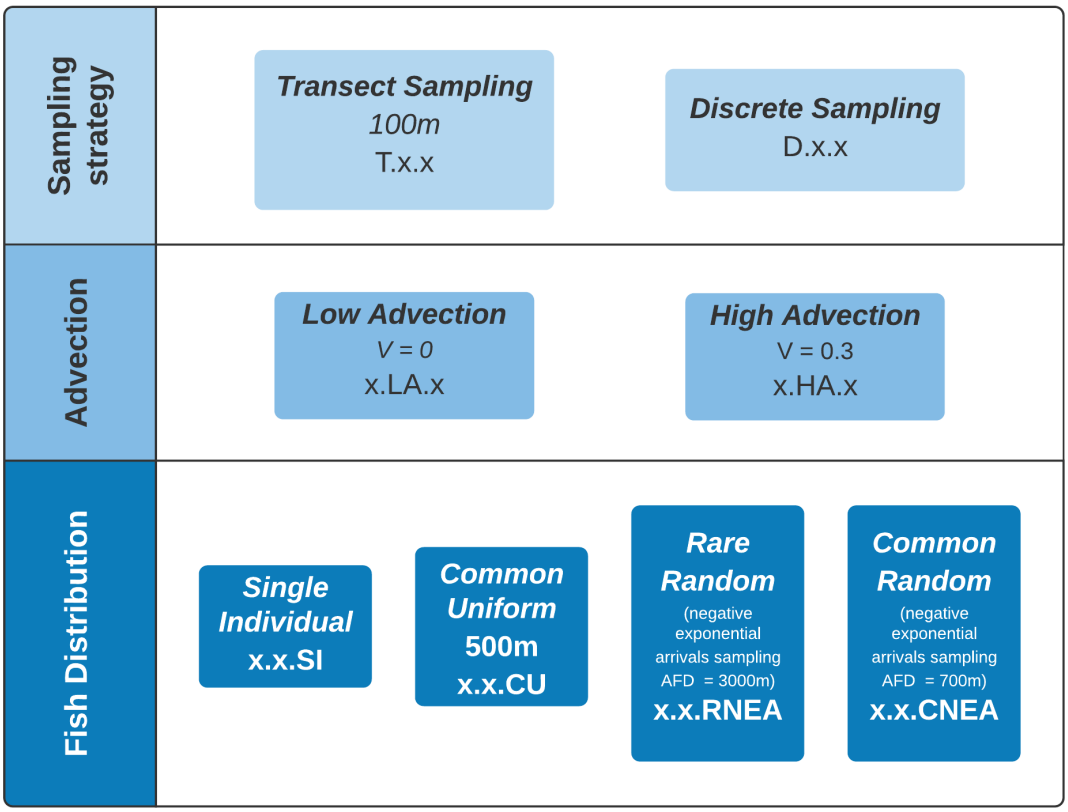


Figure 2.2: Combinations of cases of interest for the unidimensional reaction dispersion model for eDNA transport.

Generating random fish distributions

To simulate the distribution of random eDNA sources, we represented the distance between each eDNA source as a sample from a negative exponential distribution with an average sample distance of ASD. This type of random distribution assumes that each eDNA source does not interact with other eDNA sources, thus each source is completely independent. This random eDNA distribution would be appropriate for a species that does not school. More complex behaviors such as schooling

can be modeled by considering each school as a single unit or by using a different sampling function.

Average detection rate for randomly distributed eDNA sources

In cases where random distributions of eDNA sources are used to calculate average probabilities of detection, we used 1000 bootstrap replicates per average fish distance (l) to estimate the expected probability of detection, and calculate confidence intervals and standard deviations[72,73].

Estimation of effective sampled distance (ESD)

Knowing the maximum distance from the sampling site that an eDNA source can be detected is of major importance when trying to correlate a detection with physical location. Here, we will refer to this distance as effective sampled distance (ESD). ESD was measured as the sum of the maximum distance upstream and downstream which yields an eDNA detection rate above a defined threshold of 28.312%. We opted for the 28.312% detection rate as this value means that with 9 replicates per site, we will achieve a 95% detection rate ($(1 - 0.28312)^9 \sim 0.05$ chance of missing a detection). We opted to use the ESD terminology instead of detection radii as the latter refers to the maximum distance eDNA can be detected from its source (the source is at 0 m) while ESD refers to the maximum distance from the sampling site (the sampling site is at 0 m) where an eDNA source can be detected. Thus, ESD measures the length of the river that is tested for presence of the target species in each sample and is more appropriate when the fish locations are unknown.

Conversion from sample eDNA copy number to probability of detection

The probability of detection can be measured as a sigmoid function of the target eDNA copy number present in the filter. We utilized the fitted sigmoid from the serial dilution experiment from [49]. We opted to report our model in terms of probability of detection as a majority of eDNA studies report their results as presence or absence of detection instead of estimated copy number and the scale is simpler to comprehend and easier to compare between different scenarios.

Data analysis

Comparisons of transect sampling and discrete sampling using the above model were performed in Python 3.8 and the code has been made available on GitHub (https://github.com/sanchestm/eDNA-particle-modeling). The model is wrapped as a class to facilitate integration with other resources. The class contains multiple functions to analyze the expected results of an assay, to provide detection ranges, to provide expected probabilities of detection, and add flexibility to integrate this model onto other techniques. We also provide a separate interactive notebook to estimate the degradation rate and diffusion based on the detection rates or copy number results from a cage study experiment. Due to the similarity of our model to the one described in [60], we added a calculation for depositional velocity (v_{dep}) and uptake length (S_w) as a resource.

Results

Effects of each environmental parameter on the [eDNA] distribution

To identify the impacts and compare the magnitudes of each parameter on the [eDNA] distribution in the environment we compared extreme cases of D, V, λ, H^{-1}, B, μ from equation 2.4 (Figure

2.3). As constants H^{-1}, B, and μ are multiplied to each other (equation 2.4), we observe that increases of the same magnitude in each of these parameters are equivalent; thus, to understand the effects of those variations in the shape of the [eDNA] distribution we can consider $B\mu H^{-1}$ as a single constant while D, V and λ must be evaluated separately (Table 2.1). The constants $B\mu H^{-1}$ impact [eDNA] distribution by reducing the probability of amplification throughout the system (Fig. 2.3A and 2.3B). D and λ, on the other hand, mostly affect the distribution by increasing and decreasing its kurtosis, respectively. Last, the river's linear velocity is the parameter that most affects the [eDNA] distribution, with small increases sharply increasing kurtosis, reducing the probability of detection at its peak and removing the symmetry of the [eDNA] distribution over space.

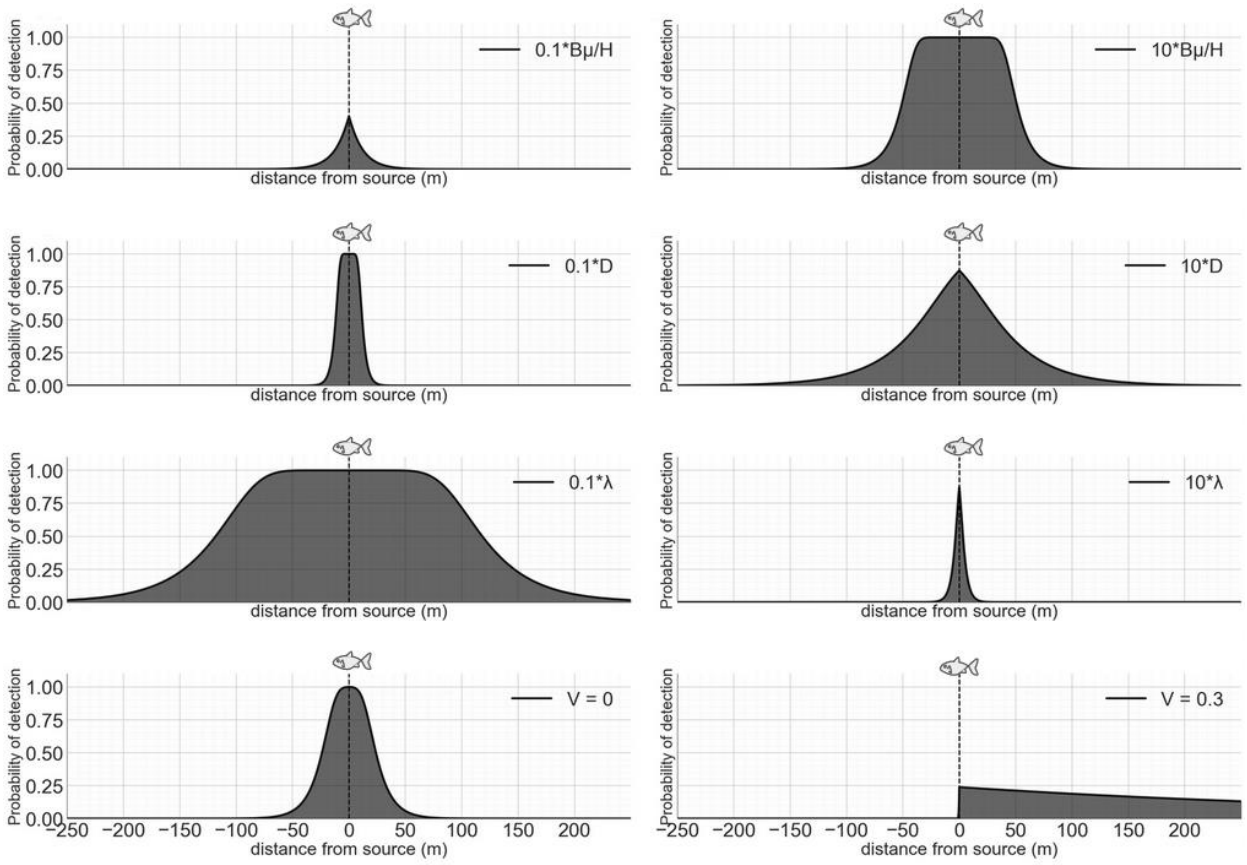


Figure 2.3: Effects of varying parameters on the distribution of [eDNA] in the environment generated by a single eDNA source. The magnitude of the parameter variation was set to 10 times lower or higher. We can observe that the constants H^{-1}, B, μ affect the quantity of eDNA present in the system, while D and λ affect the kurtosis of the [eDNA] distribution. Advection has shown to be the major parameter to determine the ESD and shape of [eDNA] distribution. B=1, V=0, u=30937.5, $\lambda=7.33*10^{-3}$, D=0.1, H=10, pf=0.005 (Table 2.1).

eDNA distribution reflects target species distribution in low advection systems

When flow was zero and eDNA particle movement was governed only by turbulent dispersion, eDNA distributions were narrow and concentrated around their sources with distinct zones of near zero [eDNA]. The eDNA distribution closely reflected the spatial distribution of the species. (Fig. 2.4; Table 2.1). The overall distribution of [eDNA] followed a negative exponential from the source with little interference between sources. When advection was introduced at 0.3 m/s, [eDNA] peaks became highly diminished, [eDNA] distributions overlapped between sources (Fig. 2.4B, 2.4C), and overall [eDNA] became more spatially homogeneous as advection velocity increased. Unlike the case with no advection, eDNA was found in most locations of the system, with small variations between [eDNA] peaks and background levels of [eDNA]. Under these conditions, eDNA at any given point could have originated from multiple distinct sources, with eDNA from up to 4 sources in Fig. 2.4B and 5 sources in Fig. 2.4C. Whenever [eDNA] from multiple sources overlapped, we observed an increase of the probability of detection, with the highest eDNA detection rate situated shortly downstream of the last source of the cluster. [eDNA]

spikes at the eDNA source locations still occurred albeit with decreased intensity as the eDNA admixes more efficiently in high advection situations.

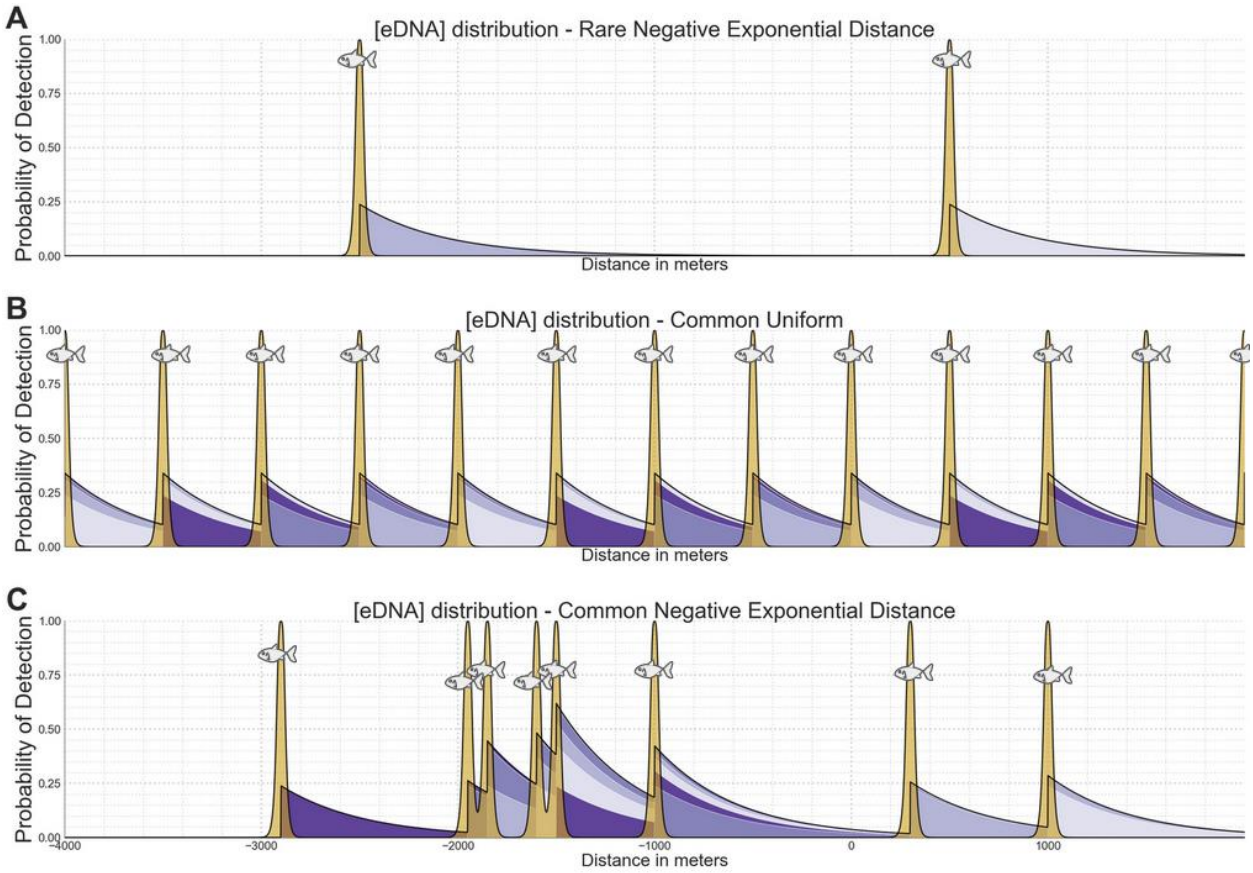


Figure 2.4: Expected distribution [eDNA] represented as probabilities of detection in the environment under the following conditions: A) rare sources distant from each other, B) multiple equidistant sources every 500 m, C) randomly distributed frequent sources. Yellow – with conditions of no advection. Purple scale – system with advection of 0.3 m/s, each layer of purple shading represents the gain in the probability of detection from each upstream source. B=1, V=[0,0.3], u=30937.5, $\lambda=7.33*10^{-3}$, D=0.1, H=10 (Table 2.1).

Transect sampling increases Effective Sampling Distance

In the D.LA.SI (Fig. 2.5; Table 2.1) scenario, detection can be obtained up to 50 m upstream or downstream of the source with a peak of 100% detection rate in this range. Discrete sampling farther away from the source yields a near zero probability of detection. Meanwhile in T.LA.SI we observed 2 differences: the probability of detection was lower (approximately 20%), but it plateaued for a long distance, approximately the entire length of the transect (500 m in this simulation) if the transect crossed the eDNA source at 0 (X_0-L/2 <0 and X_0+L/2> 0).

On the condition x.HA.SI (Fig. 2.5B) we observed differences from the condition of x.LA.SI (Fig. 2.5A). The first thing we noted was that neither transect sampling or discrete sampling reached 100% detection as the eDNA was pushed downstream. The second discrepancy was that we lost the symmetry of detection rate. In this case, detection of the source would only happen if we sampled water downstream of the eDNA source; thus, discrete sampling only detected eDNA if X_0 >0 while for transect sampling, detection was only achieved if the transect passed downstream of the source at any time of sampling. We did not observe the rapid drop in detection downstream of the eDNA source, meaning that detection could be obtained considerably farther downstream of the source for both transect and discrete sampling. In D.HA.SI, the peaks of detection probability are sharp, giving discrete sampling an appearance of a seesaw in D.HA.CNEA and D.HA.CU. On the other hand, the detection rate distribution in T.HA.x did not show clear peaks of detection, with the detection rate slowly varying as we moved downstream. Overall, transect sampling better characterized the distribution of the target species in conditions of high advection, as transect sampling smoothed the detection rate akin to a rolling window average, subduing sharp variations and facilitating the observation of [eDNA] trends. Last, when

comparing the left panel to the right panel of Fig. 2.5 we observe that advection strongly affects the probability of detection for discrete sampling while this effect is attenuated for transect sampling, as this method similarly mixed the waters from the transect range.

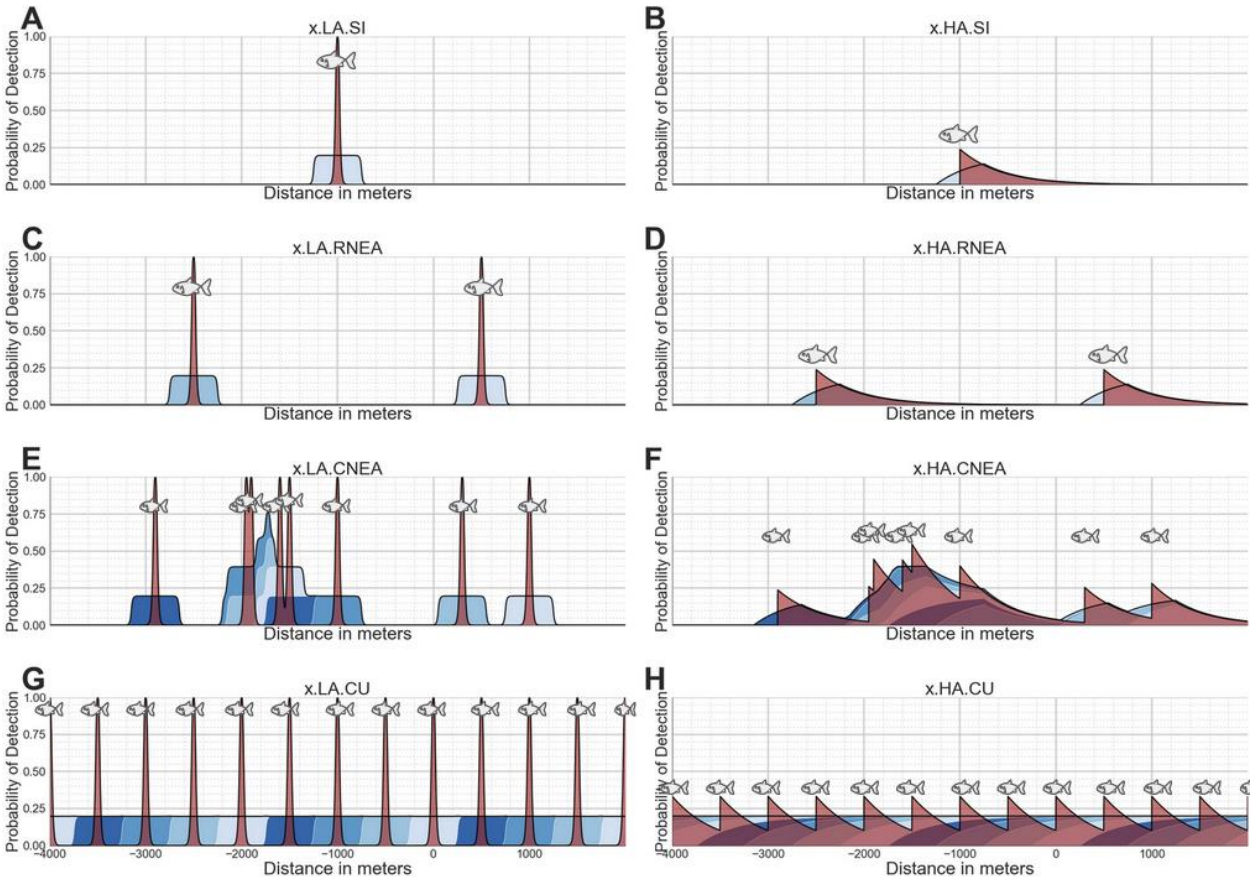


Figure 2.5: Probability of detection as a function of sampling method and location of sampling under varying conditions of advection and eDNA source distribution. A, C, E, G - no advection. B, D, F, H - advection set at 0.3 m/s. A, B - Single individual. C, D - Rare sources. E, F - common randomly distributed eDNA source. G, H common uniformly distributed eDNA sources. Blue stack plot - The probability of detection of each individual within the transect. Red - discrete sampling. B=1, V=[0,0.3], u=30937.5, $\lambda=7.33*10^{-3}$, D=0.1, H=10, pf=0.005, L=500, T = 600 (Table 2.1).

Advection and transect length interact to define the ESD of a single source

Both transect sampling and high advection increased the ESD (Table 2.1). To compare the effects of the transect length and advection onto the ESD we plotted the ESD at a detection rate of 28% (meaning that if we utilize 9 replicates per sampled site, we will obtain a 95% detection rate) as a function of the transect length and advection (Fig. 2.6).

For our eDNA model, increases in the transect length yielded continuous linear increases in the ESD (Fig. 2.6; arrow L). Increases in ESD were also observed when the advection of the system was increased (Fig. 2.6; arrow V), although in case of advection, ESD increased parabolically. As both advection and transect sampling mixed the sampled water in different ways, we observed that their effects both contributed to increasing the ESD. For example, point 3 in Fig. 2.6 has a higher ESD than points 1 or 2. Yet, the effects of transect sampling combined with the effects of advection on ESD were less than the sum of each of these factors alone. For example, the gain in ESD between points 0 and 3 was less than the sum of the ESD gain between points 0 and 1 and the gain in ESD between points 0 and 2.

Both transect length and advection increased the ESD up to a threshold (approximately 172 m for transect length only, 0.09 m/s for advection only), then dropped to a point where a single source could not be consistently detected anymore as [eDNA] was too diffused in the system (in the case of advection) or in the sample (in the case of transect length) and therefore existed in an undetectable concentration. The drop happened drastically for transect length while it was attenuated for advection.

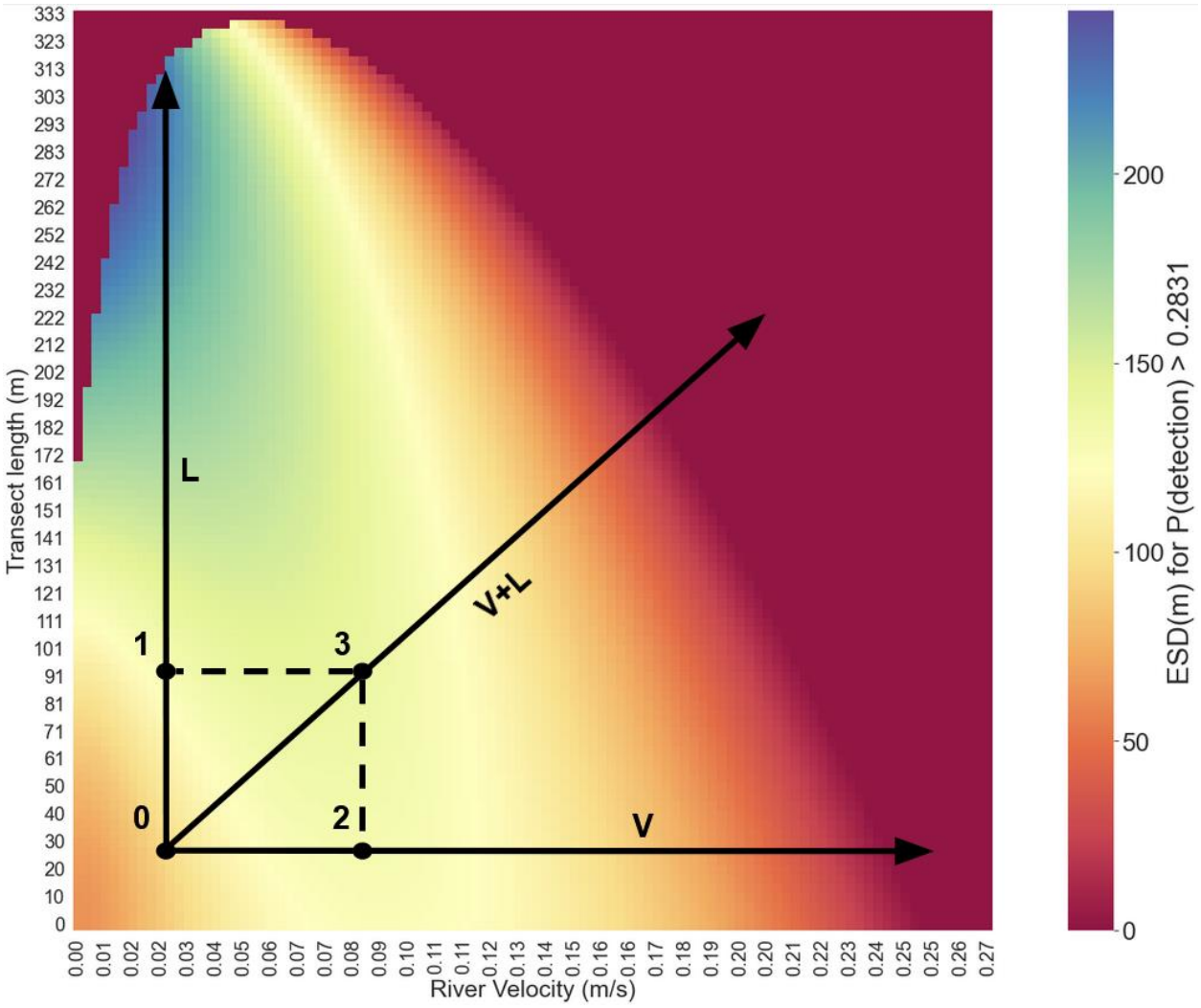


Figure 2.6: Effects of river advection and transect length onto the effective sampled distance (ESD) for a single source. L arrow - variation in transect length only; V arrow - variation in advection only; V+L arrow - variation in advection and transect length. Points 0 to 3 - example ESDs in different conditions of advection and transect length B=1, V=[0,0.3], u=30937.5, $\lambda=7.33*10^{-3}$, D=0.1, H=10, pf=0.005, T = 600 (Table 2.1).

Presence of multiple sources increases the advantage of transect sampling over discrete sampling

For both high advection and long transects in Fig. 2.6, it appeared that there was an optimal transect length and advection level to maximize the ESD. However, this measure as well as the ESD did not account for multiple sources of eDNA, as only a single eDNA source was measured for their estimation. In transect sampling and high advection cases, eDNA from multiple sources contributed to the total amount of sampled eDNA (equation 2.3). To measure the impact of multiple sources of eDNA in a system, we compared the efficiency of long and short transect samples for target species that were rare (high ASD) and common (low ASD) (Fig. 2.7).

For x.LA.x, where ASD (Table 2.1) is very small (20 m), we observed a near 95% detection rate for discrete sampling and 100% detection rate for all transect sampling lengths (Fig. 2.7). The detection probability then dropped as the average distances between sources increased. However, the influence of ASD on detection probability was more pronounced for shorter transects, where detection probability for discrete sampling dropped rapidly to a plateau ranging from 5% to 20%. At the other extreme, detection probability for the longest transect we modelled (1000 m) maintained a higher than 95% probability of detection up to an ASD distance of 5 0 m and showed a more attenuated drop for longer ASDs. Intermediary transect lengths provided intermediary curves of detection probability. When accounting for multiple sources in the system, longer transects yielded higher probabilities of detection for all measured ASDs. The advantages of transect sampling were greatest when the ASD was higher than the ESD.

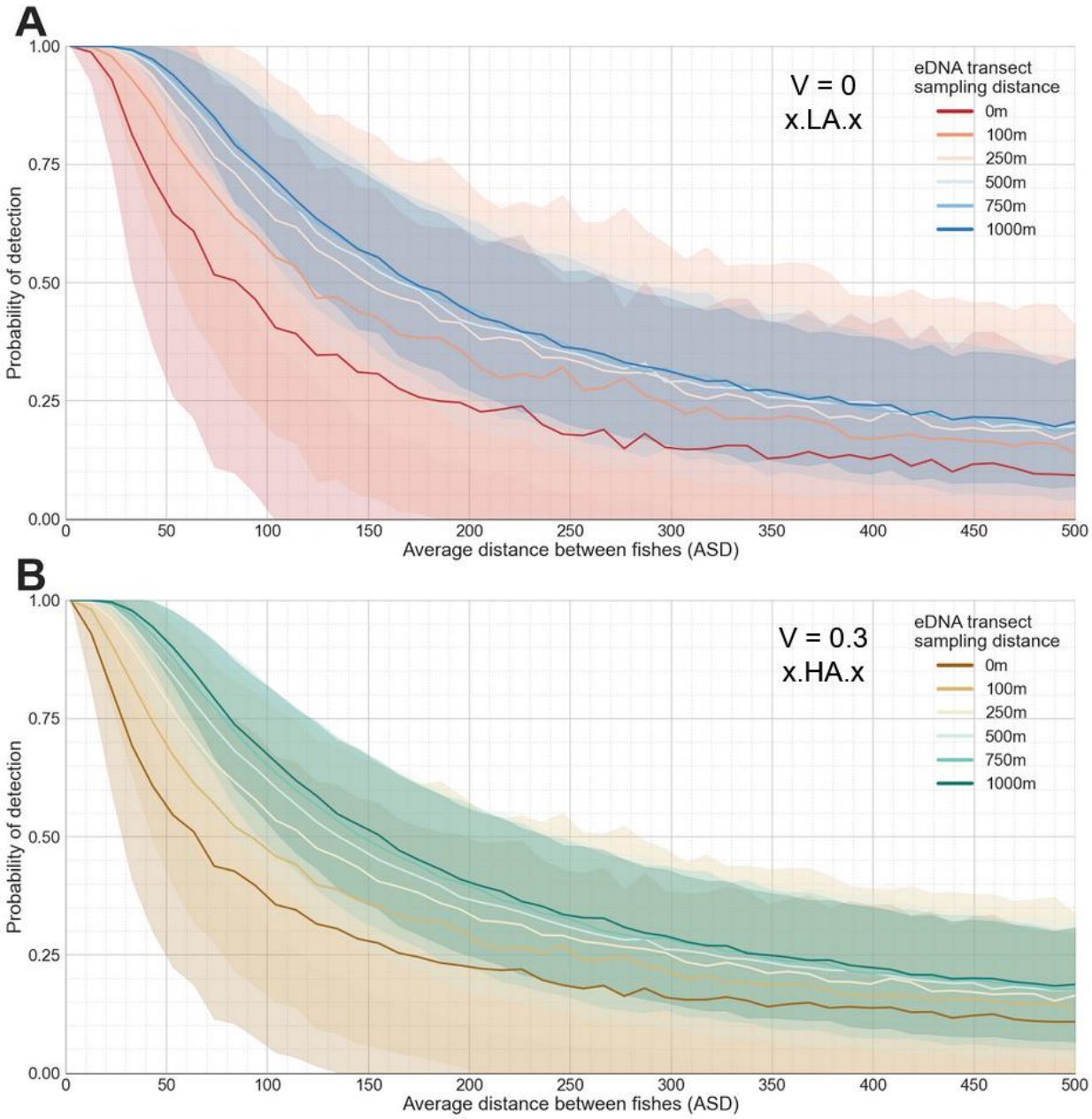


Figure 2.7: Effect of the average distance between eDNA sources (ASD) on the probability of detection in the condition of A) Low advection. Curves are shown for different transect lengths with shorter transects shown in warm colors and with longer transects shown in cool colors. Shaded areas indicate standard deviations. In this case, longer transects had a higher probability of detection, especially when sources were distributed between 50 to 300 m apart on average. B) High advection. Curves are shown for different transect lengths with shorter transects in warm

colors and longer transects in cool colors. In this case, transect length had a smaller effect on the probability of detection. B=1, V=[0,0.3], u=30937.5, $\lambda=7.33*10^{-3}$, D=0.1, H=10, pf=0.005, T = 600.

Differences between transect and discrete sampling are attenuated in high advection situations

The relative benefits of longer transects were less pronounced under conditions of higher advection (Fig. 2.7B). When advection was high, detection probabilities for all transect lengths were similar to that observed for the longest transect length under low advection conditions (Fig. 2.6). This result corroborated the interaction between advection and transect length observed in Fig. 2.6 where increased eDNA dispersion in the environment reduced the gain in ESD from transect sampling, as the eDNA is more homogeneously distributed in the environment and without [eDNA] peaks.

Longer transects have diminishing returns

Our model suggested that longer transects outperformed smaller ones in all cases. However, the gain in detection probability of longer transects $\frac{\partial(detection\ rate)}{dL}$ diminished. From Fig. 2.6, we observed that a transect of 500 m had a similar detection probability distribution as a transect of 1000 m.

Standard deviation of the detection probability is smaller for long transects

We observed that longer transects had a smaller standard deviation compared to shorter transects (Fig. 2.7A, B). To investigate the reasons for the drop in standard deviation due to longer transect

lengths, we made a histogram of the detection rate distribution from Fig. 2.5 for low and high advection scenarios (Fig. 2.2), checking the frequency of zones with high detection rates and zones with no detection (Fig. 2.8). In x.LA.SI and x.LA.RNEA we observed that for most of the discrete sampling cases, there is no detection at all, while when sampling close to the source we obtained an 100% detection rate. Thus, for discrete sampling there were zones of very high detection rate and zones with almost no detection. Meanwhile, for transect sampling, the difference between a zone of high detection and low detection was attenuated, as the detection rate peaks at 20%. With a smaller variation in detection ranges we observed lower standard deviations of detection, which represented higher repeatability from an eDNA survey. For example, we would expect different surveys of the same river to yield similar detection rates even if the sampled locations or eDNA source locations were not necessarily the same. x.LA.CU is the extreme version of this observation, where discrete sampling had regions of high and low detection rates while transect sampling always yielded a probability of detection of nearly 20%, as there was always the same amount of eDNA sources present in any transect. Therefore, transect sampling had a standard deviation of 0 and 100% repeatability. When we simulated a system with high advection (x.HA.x), the [eDNA] distribution in the environment was more homogeneous than in low advection scenarios (x.LA.x). In these cases, as there were no strong [eDNA] peaks and the background levels of [eDNA] were higher, we once again observed a lower standard deviation of the detection rate, similar to that observed for transect sampling. Overall, the lower standard deviation of the detection rate was accompanied by a higher overall average detection rate. This gain of average detection rate was further expanded if we assumed that a researcher might use several replicates to test a site, meaning that sites with lower detection rates for one test still can yield a high

positivity rate. For example, a site with 28.312% detection rate will have at least one positive result 95% of the time if we use 9 technical replicates.

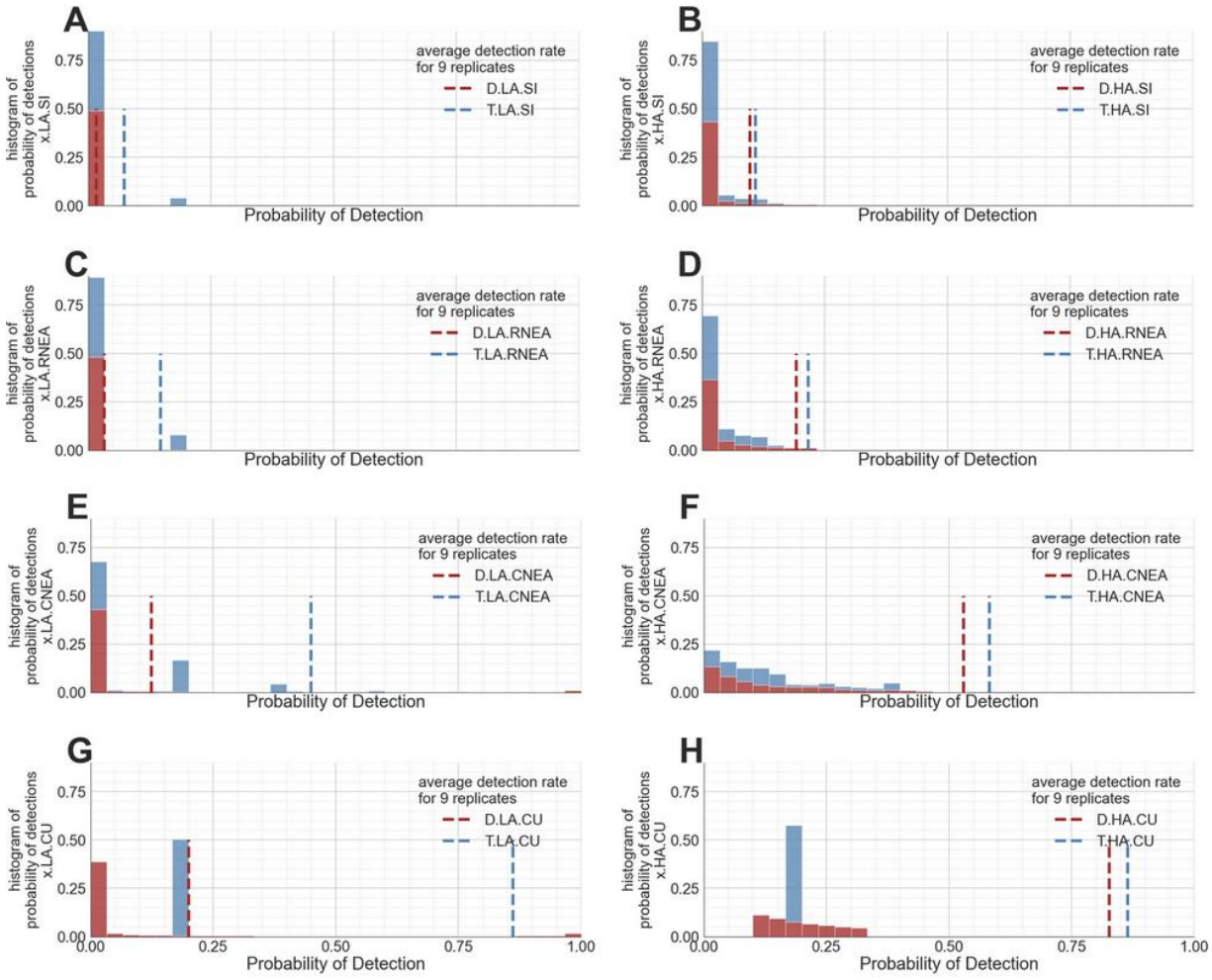


Figure 2.8: Probability of detection as a function of sampling method and location of sampling under varying conditions of advection and eDNA source distribution as shown in Figure 2.5. A, C, E, G - no advection. B, D, F, H - advection set at 0.3 m/s. A, B - Single individual. C, D - Rare sources. E, F - common randomly distributed eDNA source. G, H common uniformly distributed eDNA sources. Blue bars - transect sampling. Red - discrete sampling. Dashed lines - average positivity rate considering 9 replicates per sampled site. B=1, V=[0,0.3], u=30937.5, $\lambda=7.33*10^{-3}$, D=0.1, H=10, pf=0.005, L=500, T = 600.

Discussion

Due to the indirect nature of eDNA surveys, we need to understand factors that influence the transport, degradation, and shedding of eDNA in order to interpret the presence or absence of eDNA as the presence or absence of the target species[74]. Our model suggests that for linear systems with little to no advection, such as sloughs, eDNA does not travel far from its source before it degrades or settles in the substrate. In these situations, there is a high correlation between eDNA presence and target species presence[42,74], though this also means there is a high chance of a false negative if we do not sample close enough to the eDNA source. For example, a detection in our simulated low-advective conditions only occurred when an eDNA source was on the order of 50 to 100 meters away from the sampling site[75]. In contrast, in conditions of high advection, the eDNA may originate from one or more target individuals upstream of the sampled site (Fig. 2.2)[42,74,76]. In extreme cases, high advection values cause the eDNA to dilute to the point of no detection regardless of sampling strategy [63,77].

Transect sampling increases detection rate while attenuating the effects of microenvironments on the probability of detecting the target species. The increase in detection rate with transect sampling is maximal in systems with little to no advection, where eDNA transport is limited to turbulent dispersion, resulting in minimal eDNA particle dispersion from the source before degradation and settling[78]. Transect sampling is analogous to eDNA transport as we admix and homogenize the eDNA present over the transect. By sampling water throughout the transect, we increase odds of collecting water near a source, increasing the probability of detection. We believe that transect sampling is the optimal strategy for eDNA surveys if: there is low advection in the sampled system, the species is rare and current eDNA detection rates are low, the sampled system is large, the goal is to obtain proof of target species existence instead of fine-grained species

distribution, and if there are limitations to the sampling effort or to the volume of water that can be filtered.

Another benefit of transect sampling is increasing the effective sampled distance of a single source, reducing the probability of false negatives in which the target species occupies the region of interest but is located far from the sampling location (e.g., the target species is located on the other side of a river). The increase in ESD may impede the researcher in pinpointing the exact location of the eDNA source but allows for eDNA collection from multiple sources at distinct positions. In other words, longer transects better demonstrate the presence of a target species while discrete sampling pinpoints the individual locations of each eDNA source (Fig. 2.5).

When the intent is to detect rare species and reduce false negatives, particularly under conditions of low advection, longer transects appeared to yield great improvements in detection probability, but only to a point. In our simulations of low advection scenarios with fishes spread on average 100 m from each other, a transect of 200 m had a detection rate of 95% while a transect of 100 m had a 75% detection rate and discrete sampling had only a 55% detection rate. However, changing from a 200 m transect to a 1000 m transect only changes the detection rate from 95% to 97.5%, suggesting that the increase in transect length began to yield diminishing returns (Figure 2.7). Thus, very large transects may not be necessary to obtain clear improvements in detectability. For example, walking along the shore of a river at a reasonable pace (1.2 km/h) while sampling, and for a reasonable distance (200 m), should show large improvement in detection rate over discrete sampling or short transects[8,47,79]. For sampling on a boat, a speed of 1.2 km/h also allows the researcher to use a townet concomitantly with eDNA sampling, amplifying the amount of data gathered in the survey without an increase in sampling time. This also permits the easy addition of an eDNA survey in locations where continuous townet surveys are already used.

Our model demonstrates how abiotic factors can strongly influence the expected detection probabilities of an eDNA assay[80,81]. Given the strong influence of advection over the detection rate, we reiterate that reporting local advection speed at every sampled location is essential for interpreting the results of an eDNA survey [82,83]. Our model can also be used to report the expected range (instead of a single point) from where the eDNA originated. One of our model's greatest advantages is that the only information required is the river advection, turbulent diffusion, and degradation rates.

If the location where the eDNA survey is being conducted is well described, we suggest using Lagrange particle tracking[52,84] or fitted hydrological models[85,86] such as Delta Simulation Model II for the San Francisco Bay-Delta, which can better describe the dispersion of the eDNA particles. Lagrange Particle trackers also have the added benefit of simulating the eDNA particles in multiple states (eggs, skin, feces)[87] which can especially important as eDNA in different states has unique degradation rates, copy number and floatability[88,89], all of which can greatly impact the dispersion of eDNA in the system. However, these more complex methods require parametrization that is not commonly available and requires familiarity with hydrological modeling software. Therefore, our dispersion model is better suited for researchers that do not have the resources that are available in well-studied systems. In cases where precise modeling is available, we can apply equation 2.5 onto the simulated eDNA distribution, integrating over the dimensions simulated in the model. In these cases, we would generate a more precise comparison of transect versus discrete sampling and would allow the simulation of other transect sampling strategies or environments, such as a transect across the cross section of the river, using transect sampling in marine and lake environments, or even vertical transect sampling.

To adapt the model for conditions where the model assumptions are not met, mostly due to expectation of certain parameters being constant, we suggest testing the model for multiple values within the expected range of a variable or using average values. For example, in scenarios with varying river advection we suggest using an average river advection when those changes occur daily (e.g., tidal systems) while running the model at different advection parameters when changes are seasonal (e.g., wet season and dry season). Another example would be variability in eDNA production rates, especially for decaying eDNA sources or spawning events. In those cases, we suggest running the model at different eDNA production rates to observe if different forms of eDNA release would impact the conclusions of the study. Lastly, our model assumes that the target species remains at a single location for enough time that the eDNA distribution reaches a steady state and therefore the eDNA distribution is constant over time. In case the target species does not follow this pattern, we suggest estimating the eDNA source locations as a stable discrete probability density function in the case of species with known habitat preferences.

Our model shown a good fit with a published cage study data from [63] (Supplementary 2). One constraint of the model is that the use and interpretation of model results should be limited to lotic systems. Results from rivers with low lateral Péclet number[52] can be applied to the whole cross-section of the river. In systems with high Péclet values, the model can still be used, though conclusions will describe only a subsection of the river cross-section. To maximize detection and minimize the chance of observing an eDNA breakout phase[39,63,64] we suggest using water collection tools with a wide mouth kept parallel to the river cross section, sampling at the depth and water mass where the target species is expected to be located, or sampling across the cross-section of the river. If little is known about the target species, we advise sampling midchannel and closer to the bottom of the river to maximize eDNA detection[61,68,90]. Bankside sampling can be

complementary to midchannel sampling to better characterize the distribution of the target species across the cross-section of the river, especially in cases where the habitat is not homogeneous over the cross-section[63].

We reiterate that the appropriate sampling strategy depends on the environment, the target species distribution, and the goal of the study. If the goal of the study is to determine presence or absence of a species while minimizing false negatives, we suggest using of long transects with high total filtered volumes to maximize the probability of capturing sufficient genetic material of the target species to achieve threshold detection levels[47]. In this case, we would observe low standard deviation in detection probability resulting in higher repeatability and lower levels of false negatives throughout the study system. Given that filter clogging can be a constraint for turbid systems, longer transects can be a method to increase the detection of the target species without increasing the filtered volume or sampling effort [91,92]. Although not addressed in this theorical simulation, the difference between transect and discrete sampling may increase with more highly sensitive eDNA assays.

On the other hand, if the goal is to understand the environmental factors that dictate the presence of the rare target species, we suggest maximizing variation in the probability of detection for the complete survey by utilizing short transect samples contained within specifically defined habitat zones. This approach would allow researchers to detect most of the locations where the target species is present while at the same time having a low probability of detection where they are not present. Therefore, we would be able to differentiate areas in which the species is common from areas that the species is not frequently detected but still present at a baseline.

Since our model suggests that even short transects (< 250 m) can increase the detection rate of the target species, we recommend researchers attempt transect sampling when filtering on

site is possible. Commercial options for transect sampling equipment are described in [54]. We do not expect to see changes in cost or throughput when converting to transect sampling because there is no need to change mode of transportation. Sampling on foot can be performed while walking and boat sampling can occur with the vessel moving at a safe speed. eDNA transect sampling can also be done concomitantly to other surveys when they are available. If moving upstream, eDNA can be sampled from the front of a moving vessel to avoid contamination while at the same time a tow net is deployed from the back of a vessel.

We believe that this eDNA transport modelling exercise will allow researchers to adapt their standard operating procedures for eDNA collection to fit a diversity of aquatic systems and target species. Further cage studies comparing transect and discrete sampling in varying environmental conditions would help further validate our findings about the advantages of transect sampling. Comparing our one-dimensional model to a well-established hydrological model would provide important information on the limitations of the simplified eDNA transport models and define the conditions when they are appropriate. Lastly, developing a similar simplified model for lotic systems will permit the testing of transect sampling in lakes and large water bodies.

References

35. Jane, S. F. *et al.* Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol. Ecol. Resour.* **15**, 216–227 (2015).
36. Davy-Bowker, J., Hammett, M. J., Mauvisseau, Q. & Sweet, M. J. Rediscovery of the critically endangered "scarce yellow sally stonefly" Isogenus nubecula in United Kingdom after a 22 year period of absence. *Zootaxa* **4394**, 295–300 (2018).
37. Strickler, K. M., Fremier, A. K. & Goldberg, C. S. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation* vol. 183 85–92 (2015).
38. Lance, R. F. & Guan, X. Variation in inhibitor effects on qPCR assays and implications for eDNA surveys. *Canadian Journal of Fisheries and Aquatic Sciences* vol. 77 23–33 (2020).
39. Wood, Z. T., Erdman, B. F., York, G., Trial, J. G. & Kinnison, M. T. Experimental assessment of optimal lotic eDNA sampling and assay multiplexing for a critically endangered fish. *Environmental DNA* (2020) doi:10.1002/edn3.64.

40. Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M. & Minamoto, T. The release rate of environmental DNA from juvenile and adult fish. *PLoS One* **9**, e114639 (2014).
41. Djurhuus, A. *et al.* Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nat. Commun.* **11**, 254 (2020).
42. Eichmiller, J. J., Bajer, P. G. & Sorensen, P. W. The Relationship between the Distribution of Common Carp and Their Environmental DNA in a Small Lake. *PLoS ONE* vol. 9 e112611 (2014).
43. Itakura, H. *et al.* Environmental DNA analysis reveals the spatial distribution, abundance, and biomass of Japanese eels at the river-basin scale. *Aquat. Conserv.* **29**, 361–373 (2019).
44. Furlan, E. M., Gleeson, D., Hardy, C. M. & Duncan, R. P. A framework for estimating the sensitivity of eDNA surveys. *Mol. Ecol. Resour.* **16**, 641–654 (2016).
45. Furlan, E. M. & Gleeson, D. Improving reliability in environmental DNA detection surveys through enhanced quality control. *Mar. Freshw. Res.* **68**, 388 (2017).
46. Contente, R. F. & del Bianco Rossi-Wongtschowski, C. L. Complementarity among otter trawl, cast net, and encircling gill net improves the characterization of a subtropical subtidal fish assemblage. *Journal of Applied Ichthyology* vol. 35 1189–1196 (2019).
47. Schabacker, J. C. *et al.* Increased eDNA detection sensitivity using a novel high-volume water sampling method. *Environmental DNA* **2**, 244–251 (2020).
48. Heip, C. H. R., Middelburg, J. J. & Philippart, C. J. M. Functioning of ecosystems at the land–ocean interface. in *Treatise on Estuarine and Coastal Science* 1–3 (Elsevier, 2011).
49. Sanches, T. M. & Schreier, A. D. Optimizing an eDNA protocol for estuarine environments: Balancing sensitivity, cost and time. *PLoS One* **15**, e0233522 (2020).
50. Wolfram Mathematica. https://www.wolfram.com/mathematica.
51. Environmental Transport and Fate. https://cushman.host.dartmouth.edu/courses/engs43/slides.html.
52. Andruszkiewicz, E. A. *et al.* Modeling environmental DNA transport in the coastal ocean using Lagrangian particle tracking. *Front. Mar. Sci.* **6**, (2019).
53. Saito, T. & Doi, H. A model and simulation of the influence of temperature and amplicon length on environmental DNA degradation rates: A meta-analysis approach. *Front. Ecol. Evol.* **9**, (2021).
54. Thomas, A. C., Howard, J., Nguyen, P. L., Seimon, T. A. & Goldberg, C. S. eDNA Sampler: A fully integrated environmental DNA sampling system. *Methods Ecol. Evol.* **9**, 1379–1385 (2018).
55. Sansom, B. J. & Sassoubre, L. M. Environmental DNA (eDNA) Shedding and Decay Rates to Model Freshwater Mussel eDNA Transport in a River. *Environmental Science & Technology* vol. 51 14244–14253 (2017).
56. Wei, N., Nakajima, F. & Tobino, T. A Microcosm Study of Surface Sediment Environmental DNA: Decay Observation, Abundance Estimation, and Fragment Length Comparison. *Environ. Sci. Technol.* **52**, 12428–12435 (2018).
57. Pont, D. *et al.* Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Sci. Rep.* **8**, (2018).
58. Shogren, A. J. *et al.* Modelling the transport of environmental DNA through a porous substrate using continuous flow-through column experiments. *J. R. Soc. Interface* **13**, (2016).
59. Shogren, A. J. *et al.* Water flow and biofilm cover influence environmental DNA detection in recirculating streams. *Environ. Sci. Technol.* **52**, 8530–8537 (2018).
60. Shogren, A. J. *et al.* Controls on eDNA movement in streams: Transport, Retention, and Resuspension. *Sci. Rep.* **7**, 5065 (2017).
61. Jerde, C. L. *et al.* Influence of stream bottom substrate on retention and transport of vertebrate environmental DNA. *Environ. Sci. Technol.* **50**, 8770–8779 (2016).
62. Thalinger, B. *et al.* Lateral and longitudinal fish environmental DNA distribution in dynamic riverine habitats. *Environmental DNA* **3**, 305–318 (2021).
63. Wood, Z. T. *et al.* Spatial heterogeneity of eDNA transport improves stream assessment of threatened salmon presence, abundance, and location. *Front. Ecol. Evol.* **9**, (2021).

64. Laporte, M. *et al.* Caged fish experiment and hydrodynamic bidimensional modeling highlight the importance to consider 2D dispersion in fluvial environmental DNA studies. *Environmental DNA* **2**, 362–372 (2020).
65. Zhang, Y. *et al.* DNA tracer transport through porous media—the effect of DNA length and adsorption. *Water Resour. Res.* **57**, (2021).
66. Harrison, J. B., Sunday, J. M. & Rogers, S. M. Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proc. Biol. Sci.* **286**, 20191409 (2019).
67. Webster, J. R. *et al.* Experimental studies of physical factors affecting seston transport in streams1. *Limnol. Oceanogr.* **32**, 848–863 (1987).
68. Turner, C. R., Uy, K. L. & Everhart, R. C. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biol. Conserv.* **183**, 93–102 (2015).
69. Nevers, M. B. *et al.* Influence of sediment and stream transport on detecting a source of environmental DNA. *PLoS One* **15**, e0244086 (2020).
70. Jeunen, G.-J. *et al.* Water stratification in the marine biome restricts vertical environmental DNA (eDNA) signal dispersal. *Environmental DNA* **2**, 99–111 (2020).
71. Goldberg, C. S., Strickler, K. M. & Pilliod, D. S. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biol. Conserv.* **183**, 1–3 (2015).
72. Parsons, K. M., Everett, M., Dahlheim, M. & Park, L. Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. *R. Soc. Open Sci.* **5**, 180537 (2018).
73. Palacios Mejia, M. *et al.* The utility of environmental DNA from sediment and water samples for recovery of observed plant and animal species from four Mojave Desert springs. *Environmental DNA* **3**, 214–230 (2021).
74. Yamamoto, S. *et al.* Environmental DNA as a 'Snapshot' of Fish Distribution: A Case Study of Japanese Jack Mackerel in Maizuru Bay, Sea of Japan. *PLOS ONE* vol. 11 e0149786 (2016).
75. 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.* **14**, 109–116 (2014).
76. Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J.-C. & Altermatt, F. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nat. Commun.* **7**, 12544 (2016).
77. Curtis, A. N., Tiemann, J. S., Douglass, S. A., Davis, M. A. & Larson, E. R. High stream flows dilute environmental DNA (eDNA) concentrations and reduce detectability. *Divers. Distrib.* (2020) doi:10.1111/ddi.13196.
78. Gasparini, L., Crookes, S., Prosser, R. S. & Hanner, R. Detection of freshwater mussels (Unionidae) using environmental DNA in riverine systems. *Environmental DNA* **2**, 321–329 (2020).
79. Sepulveda, A. J. *et al.* Improved detection of rare, endangered and invasive trout in using a new large-volume sampling method for eDNA capture. *Environmental DNA* **1**, 227–237 (2019).
80. Spence, B. C., Rundio, D. E., Demetras, N. J. & Sedoryk, M. Efficacy of environmental DNA sampling to detect the occurrence of endangered coho salmon (Oncorhynchus kisutch) in Mediterranean-climate streams of California's central coast. *Environmental DNA* **3**, 727–744 (2021).
81. Fremier, A. K., Strickler, K. M., Parzych, J., Powers, S. & Goldberg, C. S. Stream transport and retention of environmental DNA pulse releases in relation to hydrogeomorphic scaling factors. *Environ. Sci. Technol.* **53**, 6640–6649 (2019).
82. Robinson, A. T. *et al.* Environmental DNA sampling of small-bodied minnows: Performance relative to location, species, and traditional sampling. *N. Am. J. Fish. Manag.* **39**, 1073–1085 (2019).
83. Thalinger, B., Wolf, E., Traugott, M. & Wanzenböck, J. Monitoring spawning migrations of potamodromous fish species via eDNA. *Sci. Rep.* **9**, 15388 (2019).
84. Delandmeter, P. & van Sebille, E. The Parcels v2.0 Lagrangian framework: new field interpolation schemes. *Geosci. Model Dev.* **12**, 3571–3584 (2019).
85. Matte, P., Secretan, Y. & Morin, J. Hydrodynamic modeling of the st. Lawrence fluvial estuary. I: Model setup, calibration, and validation. *J. Waterway Port Coast. Ocean Eng.* **143**, 04017010 (2017).
86. Risley, J. C. Using the precipitation-runoff modeling system to predict seasonal water availability in the upper Klamath River basin, Oregon and California. *Scientific Investigations Report* (2019) doi:10.3133/sir20195044.

87. Barnes, M. A. & Turner, C. R. The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.* **17**, 1–17 (2016).
88. Garcia, T., Murphy, E. A., Jackson, P. R. & Garcia, M. H. Application of the FluEgg model to predict transport of Asian carp eggs in the Saint Joseph River (Great Lakes tributary). *J. Great Lakes Res.* **41**, 374–386 (2015).
89. Schilling, H. T. *et al.* Multiple spawning events promote increased larval dispersal of a predatory fish in a western boundary current. *Fish. Oceanogr.* **29**, 309–323 (2020).
90. 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.* **146**, 99–111 (2017).
91. Hunter, M. E., Ferrante, J. A., Meigs-Friend, G. & Ulmer, A. Improving eDNA yield and inhibitor reduction through increased water volumes and multi-filter isolation techniques. *Sci. Rep.* **9**, 5259 (2019).
92. Takahashi, S., Sakata, M. K., Minamoto, T. & Masuda, R. Comparing the efficiency of open and enclosed filtration systems in environmental DNA quantification for fish and jellyfish. *PLoS One* **15**, e0231718 (2020).
93. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55 (2016).
94. Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A. & Boehm, A. B. Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. *Environ. Sci. Technol.* **50**, 10456–10464 (2016).

Supplementary 2: Fitting the eDNA dispersion model to published data

Published data was fitted to the one-dimensional eDNA transport model to ensure that the model fits real conditions. We used a published cage study design provided in the Supplementary Table 1 from [63] and the data was fitted using Automatic Differentiation Variational Inference (ADVI)[93]. The concept of this modeling is to estimate unknown variables in available cage study data provided in [63] and verify that the estimated unknown variables are within acceptable boundaries from other published data (Table 2.1). The cage study from[63] provided the biomass (Kg), distance from source (m), advection (m/s) and [eDNA] of sample while river cross-section area was assumed to be 10 m^2. These parameters were used to estimate the eDNA production rate, turbulent diffusion constant and eDNA degradation rate. The data was assumed to be a zero-inflated normal distribution, with a probability P of the eDNA failing to amplify (equation 2.9). This probability P was derived as a logistic function of the eDNA concentration, meaning that in samples with low [eDNA], there is a high probability of no detection while samples with high [eDNA] will most likely amplify. This method also allowed us to estimate the probability of detection from the eDNA concentration of the sample.

$$eDNA_{sample} \sim ZeroInflatedNormal(eDNA_{Model}, P, \sigma^2)$$
$$eDNA_{Model} = \frac{PT}{L}\int_{x_0-L/2}^{x_0+L/2} \frac{B\mu H^{-1}}{\sqrt{V^2+4D\lambda}} e^{\frac{V-\sqrt{V^2+4D\lambda}}{2D}x} dx \quad (2.9)$$
$$P = (1 + e^{a*eDNA_{Model}-5})^{-1}$$

eDNA degradation rate was estimated at $5.56*10^{-3}$ s^{-1} ± $0.63*10^{-3}$. Turbulent diffusion was estimated at 1.0936 m^2/s ± 0.035. Last, eDNA production rate was estimated at 220 $\frac{pg}{Kg*L*s}$ ± 0.15. It's important to note that in this cage study [eDNA] was measured in pg instead of copy number. Our estimation for salmon eDNA production rate was 220 $\frac{pg}{Kg*L*s}$, or 792 $\frac{pg}{g*L*h}$ which was similar

to the estimated eDNA production rate 737 $\frac{pg}{g*h}$ of Pacific Chub Mackerel described in [94]. The estimated turbulent diffusion and degradation rates from the cage study data were both approximately 10 times higher than the parameters used in Table 2.1. Yet, since the diffusion increases the eDNA dispersion and degradation rates reduces the eDNA dispersion (Figure 2.2), the eDNA distribution is similar to the one observed using the parameters from Table 2.1. We can see the quality of the model fitting by plotting the observed and simulated eDNA concentrations as a function of the distance from the cage (Figure 2.S1). Detections above 1pg are estimated to be in the order of 100s of meters away from the eDNA source similarly to the values observed in Figure 2.2.

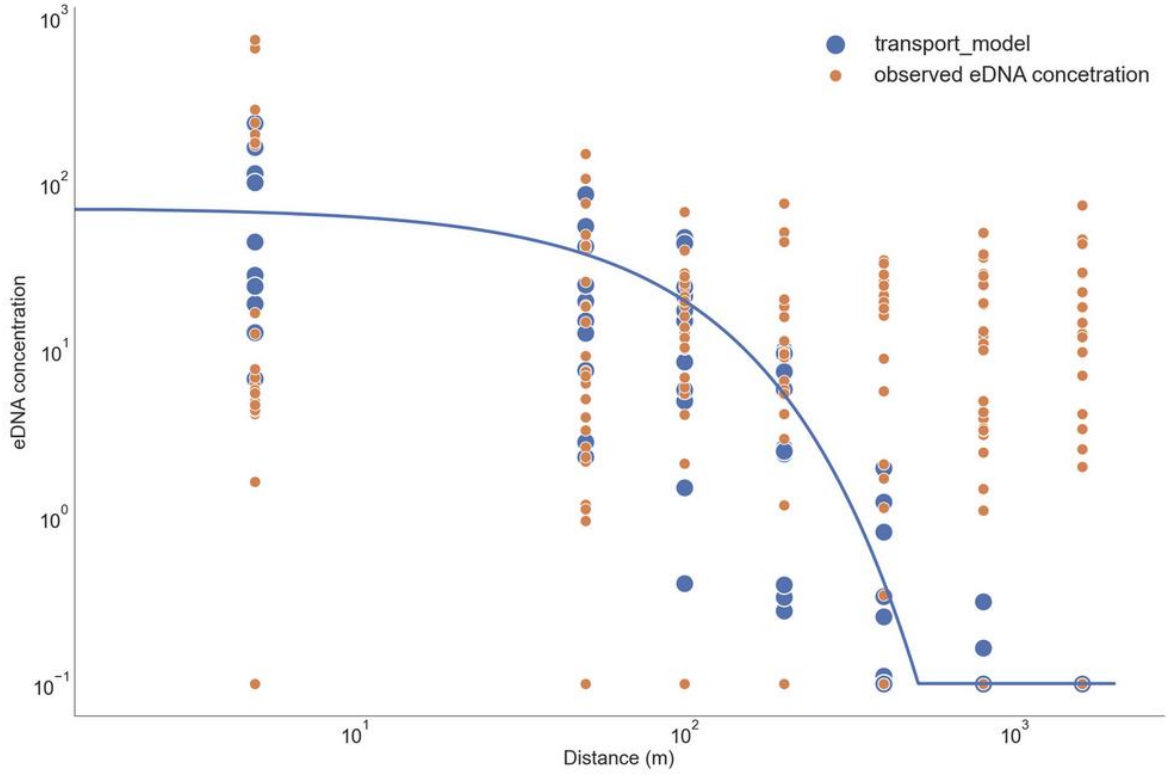


Figure 2.S1: Fitting of the one dimensional eDNA transport model onto the observed values from [63]. Expected values are described in blue, with a trendline of expected values for the average.

Comparative analysis of environmental DNA and trawling surveys for juvenile Chinook salmon in the San Francisco Bay-Delta

Abstract

Chinook Salmon populations in the San Francisco Bay-Delta (SFBD) are vulnerable to anthropogenic pressures. Efficient and precise survey of the species is necessary to understand their biology and assess the health of these populations. Between the years of 2018 and 2021, we performed pairwise surveys of eDNA and trawling for juvenile Chinook Salmon and compared detection rates to assess potential and identify biases attributable to the ecologically diverse environmental conditions of the SFBD, thus allowing better interpretation and integration of these survey methods. The eDNA survey was more sensitive than trawling for detecting Chinook Salmon, with Chinook Salmon eDNA detected in 45% of the sites while trawling captured juvenile Chinook in 13% of the sites. While eDNA outperformed trawl for detecting juvenile Chinook Salmon under all environmental conditions considered in our analysis, eDNA especially out-performed trawl in drier years, in terminal channels, and along undercut marsh banks. We observed that eDNA detection was present in broad sets of environmental parameters while trawling detection was present only in a narrow range of dissolved oxygen and salinity of the sampled sites, which can be partially explained by the limited trawling detection in drier years (when SFBD salinity levels were higher) and in upper marsh terminal channels. The Bayesian logistic regression model indicated that the eDNA survey was not more spatiotemporally autocorrelated than trawling, indicating that the eDNA detections were not repeat detections of Chinook salmon. While

eDNA was more effective for detecting juvenile Chinook Salmon, trawling can provide information about the juveniles captured that eDNA cannot provide, such as life stage, fork-length, health, and distinguishing hatchery origin from naturally produced juveniles. For these reasons we conclude that these two methods were complementary in the information that was obtained. We recommend that eDNA surveys be used to determine the most likely locations to detect juvenile Chinook salmon, while trawling be used to corroborate these detections and provide information that cannot be obtained from eDNA alone.

Introduction

Chinook Salmon in California's Central Valley have experienced a long-term and continuing decline due to anthropogenic pressures in the region [95]. A major component of the regional strategy to recover the Chinook Salmon is the restoration of rearing habitat, much of it in the tidal estuary[96]. However, actual use of Central Valley estuarine habitat for rearing is poorly understood and has been identified as a critical data gap for developing more effective management actions[97–99]. One reason for this poor understanding is the difficulty of detecting salmon in estuarine rearing habitat using currently applied net-based sampling techniques. eDNA is a rapidly advancing survey method that can address the technical complexities of conventional sampling[100,101]. To better understand the ability of eDNA compared to a trawling net to detect salmon in estuarine habitat, we compared detection rates from a four-year eDNA survey conducted in parallel with a survey using a surface trawling net specially modified and fished to catch pre-smolt juvenile salmon.

Climate change is predicted to cause longer and prolonged dry periods in the Western United States coupled with warmer and more extreme high temperatures. These landscape scale drivers are already producing observable detrimental effects on juvenile salmon rearing habitat

and oceanward migration success, including narrowing windows with suitably low water temperatures, increased pathogen outbreaks, and increased predation during migration[102,103]. A potentially mitigating factor for these trends is the large extent of estuarine rearing habitat that is completed or planned for restoration, some of which remains cooler than river habitat due to the marine influence. Currently, there is poor understanding regarding the actual use of this habitat by juvenile salmon. This knowledge gap has in turn limited the ability to design and assess effectiveness of restoration actions, and the ability to guide valley-wide salmon management in general[97–99]. A primary cause of this knowledge gap is the rare detection of pre-smolt salmon in conventional monitoring of estuarine habitat.

The rarity of salmon detection in the estuary is partly due to large scale drivers that limit salmon presence in the estuary in dry years with low river outflow[104]. However, the geomorphological complexity of the habitat also made salmon detection difficult using conventional net sampling. While small rearing salmon are generally expected to prefer shallow complex edge habitat providing refuge from both velocity and predators, regular monitoring of salmon in the estuary currently relies on surveys that use either surface Kodiak trawls or bottom oriented otter trawls conducted in open water or in the center of marsh channels[105,106]. For this reason, we instituted a four-year trawl survey using both a net and a trawling technique specially designed to sample shallow-water edge habitat in marsh sloughs. In our first pilot season, low detections prompted us to add eDNA sampling to our protocol part way through the season, allowing us to compare the relative effectiveness of these two sampling techniques for detecting juvenile salmon.

Environmental DNA monitoring is a promising strategy that is not subject to many of the methodological challenges associated with conventional field sampling for fishes such as size-

specific gear limitations and behavioral avoidance of nets[107]. Other advantages of eDNA are that sample collection does not disturb habitat like conventional net sampling gear, which often rakes the channel bottom or shoreline. Similarly, eDNA does not disturb target or not-target species like conventional net sampling gear, and therefore does not have permit constraints on sampling intensity [108,109]. which can limit the ability to detect and map distributions of species, particularly when highly sensitive or rare species are present[110,111]. In most cases, the simplicity of eDNA sampling approach may promote a highly standardized approach throughout a surveyed area and allow sampling in locations where conventional gear cannot be deployed safely or effectively[112].

While there are many benefits to surveying species using eDNA, there are considerations specific to interpretation of eDNA detections that must be considered. The probability of detecting eDNA in an aqueous environment is largely dependent on the concentration of eDNA in the water from which the sample is taken. Several factors can influence site-specific eDNA concentration including the distance from an eDNA source, a source's production rate of eDNA, hydrological transport and dispersion, and eDNA degradation, which is arguably less important under conditions of moderate dispersion and low eDNA production (i.e., low organism biomass). In addition to these field-based factors, changes in water quality from sample to sample can alter PCR inhibition rates in the lab[113]. The influence of all these factors can be difficult to estimate and account for when modeling species distributions using eDNA detections. However, conventional sampling methods are also subject to site-specific biases that are difficult to estimate and account for, such as differences in gear efficiencies in different environments[114]. Ultimately, the potential influence of these factors must be understood and considered in the interpretation of data from either eDNA or trawl surveys.

To understand the environmental biases affecting eDNA and trawling surveys of Chinook Salmon, we compared the results of a three year pairwise survey conducted in the San Francisco Estuary's Suisun Bay and Suisun Marsh. While we address salmon habitat use in another manuscript, for this analysis we focused on site-specific factors that influenced detections for the trawl and eDNA surveys. We expected both relatively constant characteristics (water body type and geomorphology), and transient characteristics (water depth and water quality characteristics) at each sampling site to influence the relative efficiency of each technique during each sampling event. Water quality characteristics were expected to have a higher effect on eDNA detection by influencing the eDNA transport and degradation, while channel geomorphological characteristics that may interfere with trawling, particularly channel edge type, would have a stronger influence on trawling detection. Site-specific conditions we measured and hypotheses regarding the influence of these conditions on our sampling methods are described in Table 3.1.

We expected this information would provide us the advantages and disadvantages of each method and improve interpretation of these methods for juvenile Chinook salmon distributions under different environmental conditions. This knowledge also let us think of a new framework to improve combined use of eDNA and trawl surveys.

Table 3.1: Expected impact of environmental conditions on the relative detection of juvenile Chinook salmon by trawling versus eDNA.

Feature	Predictions
Salinity	Lower salinity levels can be correlated with an increase of eDNA degradation and can affect eDNA extraction[115]. We predicted that higher values of salinity will yield higher than expected detection rates for eDNA while not affecting the trawl surveys[116]
Dissolved oxygen	Lower dissolved oxygen can be a proxy for higher biological activity in the region, which would be associated with higher eDNA degradation rates. Very low dissolved oxygen could diminish juvenile salmon swimming performance and their ability to evade trawling nets, yet we did not see dissolved oxygen concentrations at these levels. We predicted that higher values of dissolved oxygen would yield higher than expected detection rates for eDNA relative to trawl surveys.
Depth	Channel depth increases turbulent dispersion. While deeper channels could also allow greater "dispersal" of salmon across channel depths, salmon prefer to migrate nearer the surface where we fished our trawls. At the same time reticulate channel sides and uneven channel bottoms in shallower sites could reduce trawl efficiency due to possible interference with nets and the provision of "hiding" locations for juvenile salmon, but these mechanisms would not affect eDNA. Therefore, we predicted water depth would increase trawl detection relative to eDNA detection.
Temperature	Higher temperatures are correlated with higher degradation of eDNA, while juvenile salmon swimming performance declines rapidly when temperatures are at or above 19°C, likely affecting their ability to evade trawls. We predicted that higher temperatures would yield lower than expected detection rates for eDNA relative to trawl surveys.
Turbidity	Turbidity can clog the filters used for eDNA collection and inhibit eDNA amplification with PCR inhibitors. At the same time, turbidity is associated with higher efficiency for trawl sampling because impaired visibility impairs a fish's ability to detect and evade the nets. We predicted that turbid sites would have a negative bias for eDNA detection relative to trawling.
pH	Lower pH levels are correlated with higher degradation of eDNA, with no discernible effect on trawl efficiency. We predicted that lower values of pH

	would yield lower than expected detection rates for eDNA relative to trawl surveys.
Chlorophyll A	Can be a proxy for microbial activity. Higher microbial activity is correlated with a higher eDNA degradation rate. We predict that high values of CHL-A will yield lower than expected detection rates for eDNA while not affecting the trawl surveys.
Hydrological class	While sample locations were limited to shallow water, locations were distributed evenly among open bay shoals, distributary marsh channels, and terminal marsh channels. Terminals are expected to have the highest water residence time followed by distributaries and open bays. Open bays are also expected to have a complex set of currents. These two factors led us to predict that eDNA detectability will be highest in terminals due to reduced eDNA dispersion while not affecting trawl detectability.
Geomorphology	We expected the geomorphology of the sampled site to influence trawl ability to detect salmon when present, while having less appreciable influence on eDNA detectability. We expected trawls to be most efficient fishing shoals and earthen levees, followed by riprap, then natural edge, emergent vegetation and cutbanks.
Time autocorrelation	eDNA can linger in the environment after the source organism has left an area, leading to a longer period of possible eDNA detection than the period of organism presence, and possible "bridging" of eDNA presence between discrete periods of organism presence. We predicted that eDNA would have a higher time autocorrelation than trawling.
Spatial autocorrelation	eDNA disperses as an expanding plume from a source over time creating a larger area where detection is possible relative to trawl, which must fish the precise point location where the fish is present. Therefore, the possibility of nearby detections would be higher for eDNA relative to trawl. We predicted that eDNA would have a higher spatial autocorrelation than trawling.

Methods

Study Area

The study was conducted in the western Sacramento–San Joaquin River Delta, upper San Francisco Estuary, and San Pablo Bay, encompassing connected waters from Sherman Lake to the constriction point between Point San Pablo and Point San Pedro. Hydrology in the entire study area is tidal. Therefore, water depth and water quality at sample sites (e.g., salinity, turbidity, water temperature) vary hourly, while larger scale spatial patterns in water quality along the river to ocean corridor and temporal patterns over the migration season and year to year vary mainly as a function of river outflow. In general, higher outflows mute spatial and temporal variation by pushing the influence of ocean water and tides downstream. In addition, the shallow, muddy and densely vegetated habitat in the upper SFBD that is used by these juvenile Chinook brings numerous difficulties to conventional field sampling methods[117].

It is known that yearly outflow not only affects the overall number of salmon in the estuary, but also heavily influences the water conditions in the San Francisco Bay-Delta (citation on WQ conditions)[104]. In high outflows such as occurred in 2018 and 2019, fresh water pushes marine water towards San Pablo Bay, leading to a large drop in salinity in the Suisun Bay main corridor and the upper marshlands alike. The freshwaters from wet years also increased the turbidity and reduced the pH of all sites by bringing suspended solids from the Sierra Nevada mountains. Within a year we observe seasonal variations of mostly temperature, followed by variations in pH and salinity. Marsh habitats present in the highest latitudes of the Suisun Bay have lower dissolved oxygen, higher chlorophyll and lower pHs compared to the main channel.

Sample site selection

Each juvenile migration season, biweekly sampling typically commenced in December after the first detections of juvenile salmon were reported in lower Sacramento River monitoring conducted by the California Department of Fish and Wildlife and U.S. Fish and Wildlife Service, indicating salmon may be entering the estuary. Sampling commenced by the second week of January regardless of salmon detections in monitoring. Yearly descriptive statistics followed the California water year, which extends from October 1st to September 30th (e.g., both December 2019 and January 2020 sampling are referred to as "2020 migration season"). Once commenced, sampling continued until the end of the juvenile migration season in June. Spatially balanced sampling locations were selected randomly within the study area using the Generalized Random Tessellation Stratified (GRTS) program[118], which reduces spatial correlation and provides more information per sampling unit. The GRTS program discretizes an ArcGIS map layer of targeted habitat into grids and randomly selects locations from among those grids, allowing weighting of grid selection probability according to criteria defined by the user. While we cannot confirm the absence of adult Chinook salmon in the sampled area, we expected their occurrence to be minimal during the sampling period, and that most if not all the eDNA detected to be originated from juvenile individuals. We began with a fixed habitat layer defined by waters of the study area shallower than two meters below mean low tide because pre-smolt salmon in the San Francisco Estuary are rarely found in deeper tidal habitat except when entrained during extreme outflow events. These areas of shallow water habitat were then segregated into broad regions. Habitat within each region was then weighted based on three habitat types: open bay shoal, terminal slough (channels in a terminal slough network), and distributary slough (channels not in a terminal slough network). In 2020 and 2021, we combined the Upper Suisun with East Marsh region, and the Lower Suisun with West

Marsh region, to create a better balance of habitat types within each region because we found that a paucity of terminal sloughs in several of the initial regions caused sampling of this habitat type to be concentrated in a few small areas [119].

Each biweekly sampling period, sampling was restricted to only regions with a probability equal to or greater than 0.2 that juvenile salmon would be present using the modeled probabilities based on day of year and current Sacramento River outflow from [120]. However, if salmon were detected by trawl or eDNA in the most oceanward region, the next most oceanward region was also sampled, continuing in this manner until the most oceanward region in the study area was sampled (San Pablo Bay).

There were large differences in climatic conditions across study years that strongly influenced river outflow. Both 2020 and 2021 had very low outflow during sampling periods, while 2019, and to a lesser extent 2018, had high outflow (Figure 3.1). The number of samples taken in 2020 and 2021 was smaller compared to 2019 due to Covid pandemic related restrictions on fieldwork. There were also fewer samples taken in 2018, which was a pilot year in which sampling did not commence until April 16th, 2018.

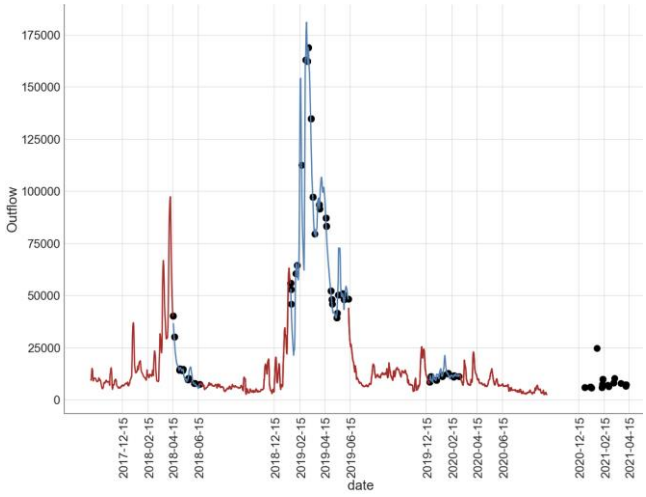


Figure 3.1: Daily outflows for the four years of pairwise sampling. Points - positive sites, Blue line - sampling season, Red line - outside of sampling season.

Trawling and water quality sampling

Trawl survey methods are described in detail in (Hassrick et al. in prep). Briefly, trawls were conducted with a specially designed Mamou net fished at surface along a transect beginning at each sampling site. In open bay shoals trawling was performed parallel to the nearest shoreline, while continuing along the shoreline as close to the edge as the gear would allow if the site was a channel. All fish species sampled were counted, and all juvenile salmon were sacrificed for diet and tissue stable isotope studies (Sturrock et al., in prep; Johnson et al. in prep).

At the start point of each trawl, a YSI Exo sonde was used to measure water quality, including temperature, salinity, turbidity, chlorophyll A, dissolved oxygen, and pH. Each sampling location was also characterized according to the following geomorphological classes: shoals (77), earthen levee (49), cutbank (with undercut - 38), riprap (2), natural edge (14) and emergent vegetation (12).

eDNA sampling

In the years of 2018 and 2019, three discrete 1 L grab samples of surface water were collected at each trawling site[121] using a sterilized wide mouth Nalgene bottle. Water collection occurred before trawling at each location to minimize the risk of contamination from the vessel or the trawling gear. At each site an extra bottle was filled with DI water, capped, and tested for Chinook eDNA, serving as negative control to detect possible contamination during the bottle handling, filtration or extraction processes. Samples were then kept in ice chests on wet ice until filtration,

which was performed within 24 h of sampling. Samples were filtered in parallel, four samples at a time, using a peristaltic pump (Longerpump © BT300-2J) at 300 rpm. The filter type used was a 52 mm glass fiber with 1.6 micron mesh to maximize filtered volume and eDNA capture[27]. When filters were clogged with suspended sediment, the pressure buildup was released at the pump head and then the pressure was added again. After every use, bottles, tubing and filter casings were rinsed with 20% bleach, followed by two DI water rinses and drying. Filter holders were placed inline between bottles and the pump heads to minimize the amount of tubing that the sample water contacted prior to filtration.

In the years of 2020 and 2021, to collect eDNA samples in a manner consistent with trawl sampling, a Smith-Root eDNA sampler[54] with inline filtration was used to pump sample water along the continuous transect during each trawl. In this new sampling strategy, four 25 mm diameter 1.6 micron mesh glass fiber syringe filters were each attached on the tip of a metal manifold that was immersed in the sampled water body. Then the manifold was connected to a self-priming diaphragm pump to filter the water. This approach reduced possible contamination from exposure of filters to handling during field collections and lab filtration, reduced the possibility of contamination from filtering tubing and minimized eDNA degradation due to delayed filtration and preservation[122]. eDNA samples were collected by a dedicated individual from the bow of the boat during trawls to eliminate cross contamination from trawl personnel, and to prevent sampling of eDNA disturbed by the passage of the boat and net gear. Pump rate was adjusted to target filtration volumes of 3 L per site. In cases where filters clogged in the middle of a transect, the filter was replaced mid-sample as quickly as possible. Filters were stored in individual sealed plastic bags and immediately placed on wet ice for transport to the lab at the end

of each sampling day. In the lab, filter casings were cut using a 1-1¼ inch ratchet-type PVC pipe cutter to remove filters and stored in 1.5 mL microcentrifuge tubes at -20 ˚C until DNA extraction.

eDNA extraction and amplification

In 2018, Qiagen DNeasy extraction was performed following the manufacturer's recommendations. For subsequent years, magnetic beads were used for DNA extraction following the protocol described in [49]. Once extracted, there was enough DNA elute to run two PCR assays on each sample (i.e. including blank, eight PCR assays were run for each location). In all years, secondary inhibitor removal was performed using Zymo OneStep™ PCR Inhibitor Removal Kit (Zymo Research; Cat No./ID: D6030), following the manufacturer's protocol. Quantitative PCR for Chinook Salmon detection was performed following the protocols described in [49] and [21]. The primers used for the amplification targeted the mitochondrial cytochrome-b gene, (forward primer 5'-CCTAAAAATCGCTAATGACGCACTA, reverse primer 5'-GGAGTGAGCCAAAGTTTCATCAG, probe 5'-AGCACCCTCTAACATTTCAG). Primer concentrations were 0.9 μM each, 0.7 μM for the Taqman probe, and 6 μL of eDNA extract. Thermocycling was performed on a Bio-Rad CFX96 and the thermal profile was 10 min at 95°C, 40 cycles of 15 s denaturation at 95°C and 1 min annealing–extension at 60°C. A site was considered positive if at least one field replicate was amplified. If a field blank amplified, the location was excluded from further analysis for both eDNA and trawling.

Data analysis

To compare relative detection capability between the methods under different environmental conditions independent of the influence of those conditions on presence/absence, only sites where

at least one of the methods was positive were used in data analyses. For all data analyses, continuous water quality values were normalized and centered. Extreme values of water quality variables (> 3 SD) were suspected to be erroneous and were set to the value of the third standard deviation to preserve the detection data point while minimizing the extreme values influence.

To determine whether the values of continuous water quality variables clustered according to specific water body or edge types, we performed uniform manifold approximation and projection (UMAP) analysis which is a non-parametric machine learning tool which projects data values into a reduced dimensional space similar to principal components analysis and non-metric multidimensional scaling. For the UMAP we used the default settings for number of neighbors equal to 15, minimum distance of 0.1, Euclidean metric, and a learning rate of 1. We compared the distribution of the normalized and centered water quality variables associated with eDNA and trawl detection using t-tests. We then applied a logistic regression model, which assumes a linear relationship between the detections (y) and the independent variables (x) described in Table 3.1 (equation 3.1). The logistic regression was chosen based on its clear connection with probabilities, and high explicability compared to other algorithms.

$$\hat{y} = logistic(X \cdot \beta_{method} + intercept_{method}) \quad (3.1)$$

In this framework each alpha represents the base detection rate for each sampling method (eDNA and trawl) relative to all positive detections, and beta represents the bias for each sampling method associated with each independent variable (i.e., water quality and physical characteristics at sampling locations). For the logistic regression model, the open bay and natural edge geomorphology were not used due to its high collinearity with shoals and distributaries

respectively. An ADVI optimizer was used to fit the model and estimate parameter values. To assess if the biases are independent of each other, we appended an interaction term to the model. Although our site selection was designed to avoid sampling locations near to each other on any given day, we tested for conditional autocorrelation to ensure this was the case. We used WAIC to compare the model with and without the interaction and autocorrelation terms. Bayesian modeling and model selection was performed using pymc3[123] in Python 3.8, and the analysis pipeline can be found at https://github.com/sanchestm/eDNA-salmon-migration.

Results

eDNA showed higher positivity rates for all sampled years and the difference in positivity is accentuated in dry years.

Between the years 2019, 2020, and 2021, we observed the highest overall positivity rate in 2019 (Table 3.2). This was very close to the annual positivity rates for eDNA alone because most of the detections were by eDNA. Note that the positivity rate was highest in 2018 because sampling mainly occurred during the period of heavy hatchery fish migration through the study area. Although both eDNA and trawl had lower positivity rates in the dry years of 2020 and 2021, the dry year effect on positivity was more pronounced for the trawl survey.

While there were frequent eDNA detections with no corresponding captures with trawling, there were few trawl detections without corresponding eDNA detection, and these few trawl-only detections occurred almost exclusively in the year 2019 (Figures 3.2).

Table 3.2: Positivity rates (Proportion of positive samples relative to total number of samples)

Oligonucleotide	2018	2019	2020	2021	Total
eDNA positivity	54%	45%	35%	22%	38%
Trawl positivity	13%	17%	2%	3%	10%
Overall positivity	57%	48%	36%	23%	42%
Total samples per year	66	186	95	110	457
Total detections per year	38	94	34	26	192

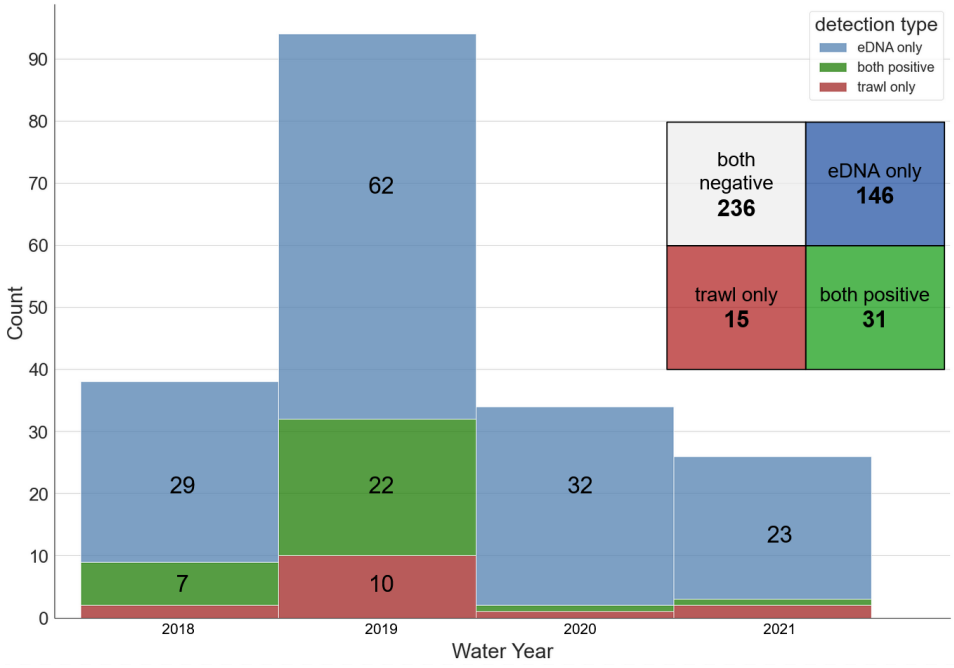


Figure 3.2: Number and overlap in positive detections for eDNA and trawl by year. Most positive detections occurred in the wet year of 2019 and for all year's most positive detections were from the eDNA survey

Spatial distribution of positive sites for eDNA and trawl

In 2019, both eDNA and trawl detections were widely distributed across the study area (Figure 3.1), although detections were concentrated along the main migration corridor. In 2020 and 2021, eDNA detections were again widely distributed, while the few trawl detections occurred only along the migration corridor.

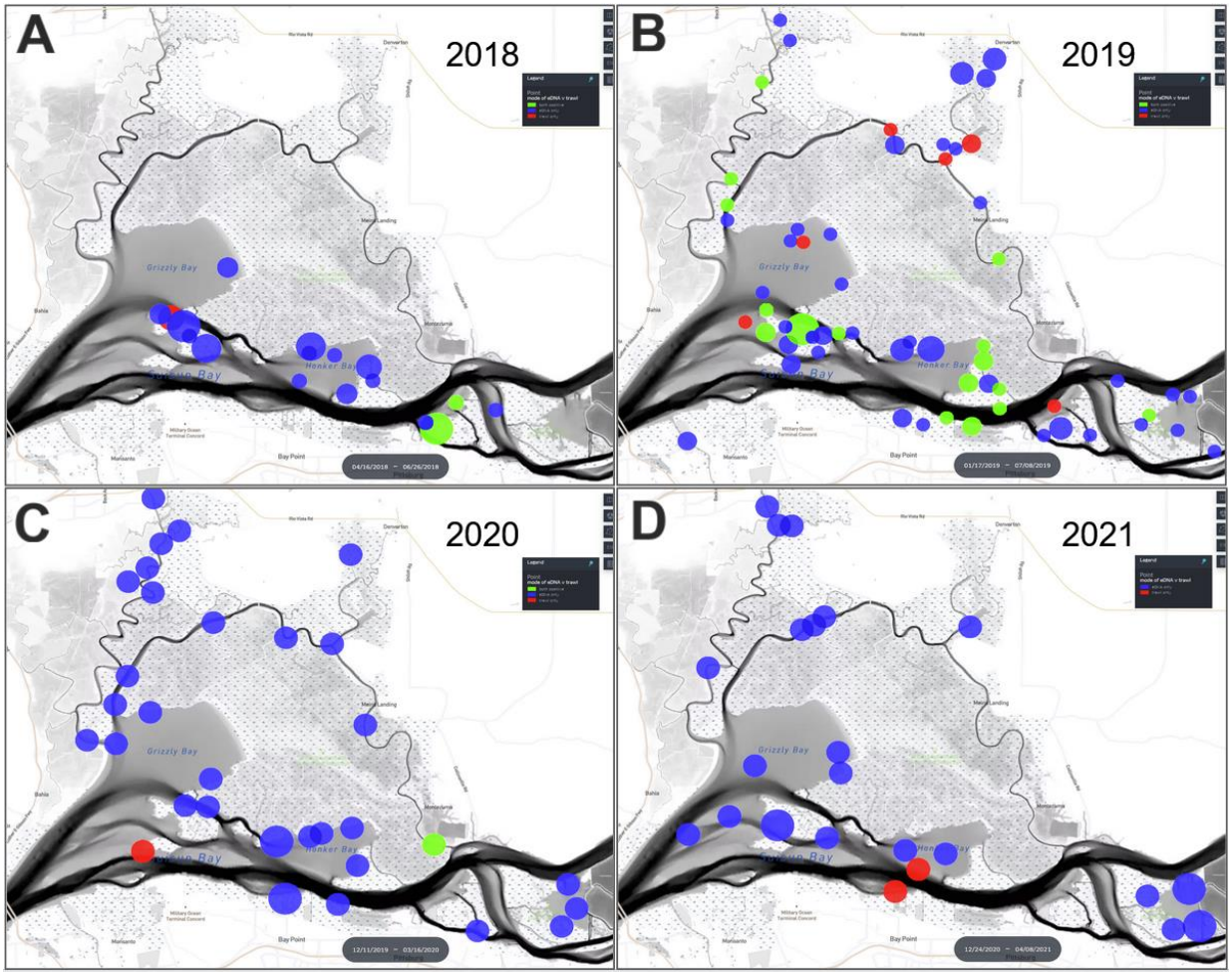


Figure 3.3: Spatial distribution of trawling and eDNA positive sites throughout the Suisun Bay and Suisun Marsh between the years of 2018 and 2021. Raster coloring represents the depth at the sampling site, with black coloration indicating depths below 10m.

eDNA has higher detectability in all measured morphological and hydrological conditions and the detection gap is most pronounced in cut banks and terminals

Positive detections were well distributed across geomorphological and hydrological classes. Apart from shoals, which occur mostly and reciprocally in open bay sites, most geomorphologies are well represented in the hydrological sites, indicating low collinearity between the hydrology and geomorphology (Figure 3.4A). When we compared the proportion of eDNA and trawling detections for each of the hydrological and geomorphological site classes, we found that eDNA detections were predominant in every case (Figure 3.4B,C). However, the proportion of eDNA over trawl detections was highest in terminal channels and cut banks.

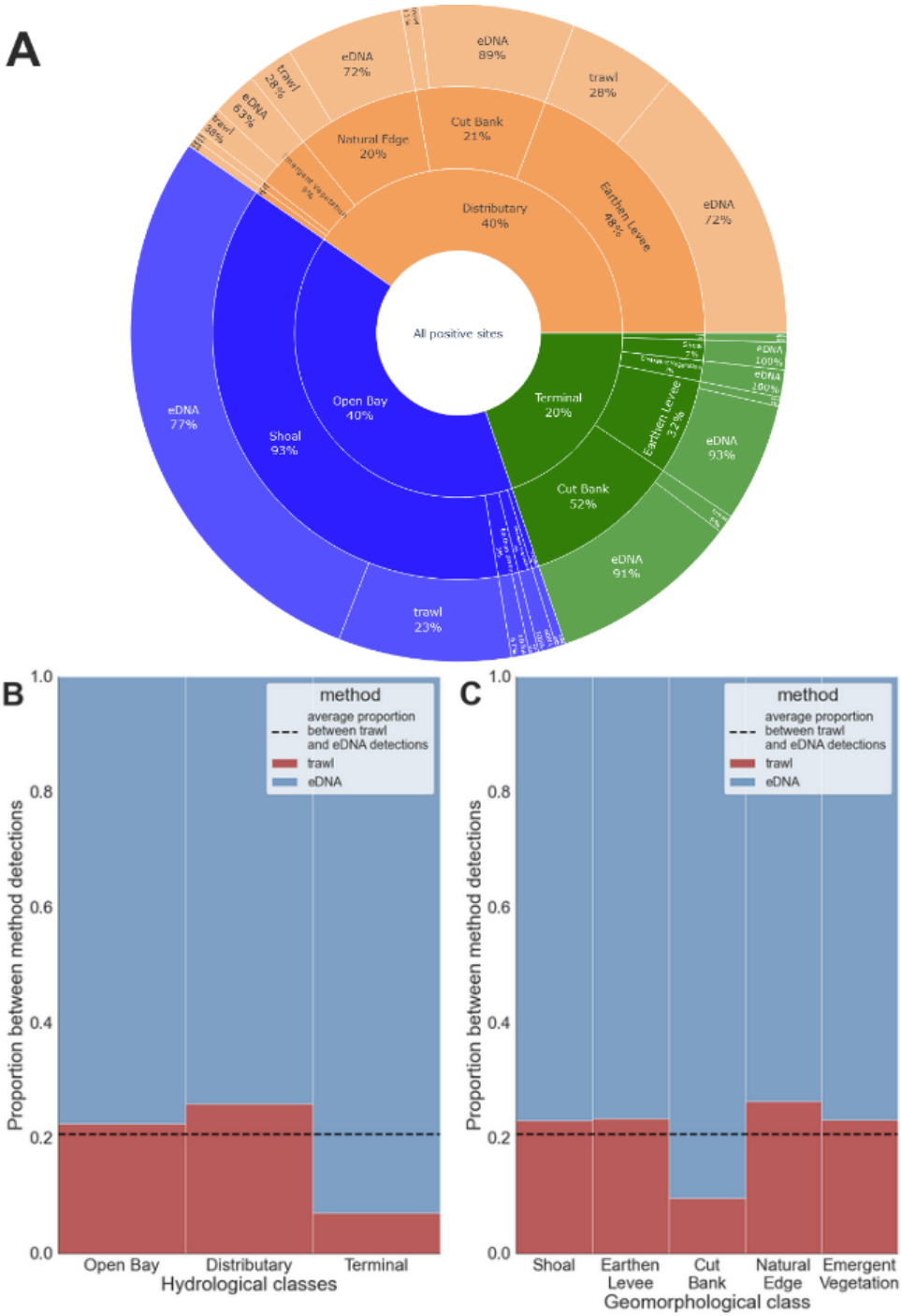


Figure 3.4: A) Sunburst diagram highlighting the proportions of eDNA and trawl detections for each combination of hydrological and geomorphological classes. B-C) Proportions of eDNA detections and trawl detections for hydrological classes and geomorphological classes separately.

eDNA detections were more frequent for all classes. The higher number of eDNA detections was more prominent for terminals and cut banks.

Salinity is the most statistically significant water quality parameter that differentiates eDNA and trawling detections

There did not appear to be a strong relationship between any of the continuous water quality variables warranting exclusion of any variable from further analyses (Figure 3.5A). There also did not appear to be strong clustering of water quality according to any of the fixed hydrological or geomorphological site classes, suggesting any influence of water quality variables would not simply reflect differences observed across site classes (Figure 3.5B,C).

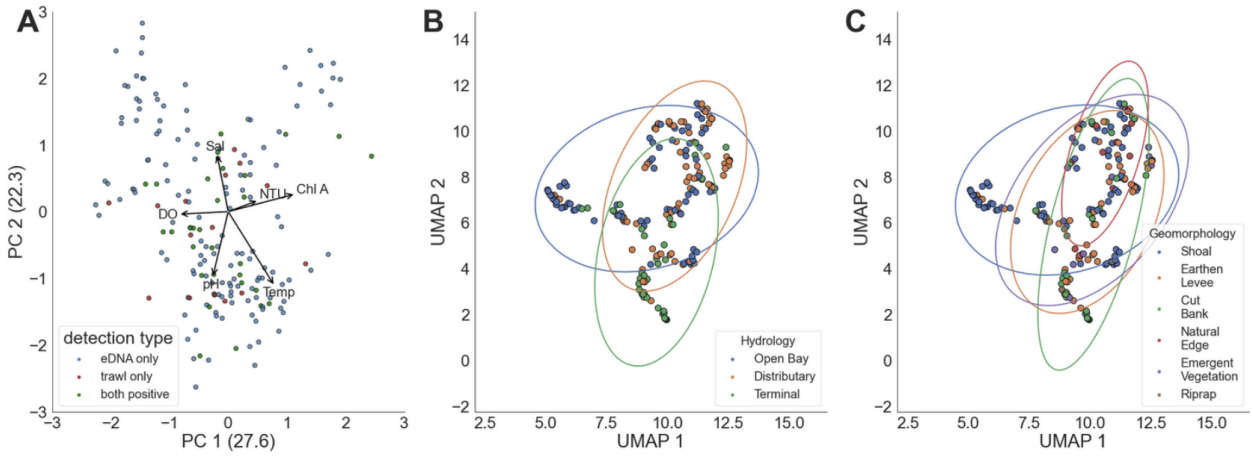


Figure 3.5: A) Principal component decomposition of water quality parameters for positive sites. Points - PCA representation of positive sites, Vectors - direction of measured water quality parameters. B-C) UMAP representation of water quality parameters for positive sites, ellipses represent the 95% confidence interval for hydrological and geomorphological classes.

The t-tests comparing water quality variables between eDNA, and trawl detection indicated that salinity had the lowest p-value (p-value: 0.0106), where eDNA detections occurred at higher salinities compared to trawling (Figure 3.6). Trawling detection was mostly limited to lower salinity values with 90% of the trawling detections occurring in salinities under 1.92 ppt, which were concentrated in the wet year of 2019, while 90% eDNA detections occurred in salinities under 6.09 ppt and spread across the sampled years. Depth was the only other water quality variable with a p-value < 0.05 (p-value: 0.0374). Although the distributions of temperature associated with eDNA and trawl detections appeared quite similar, the median temperature for eDNA detections was considerably lower than for trawl detections. In general, across water quality variables, the range of values at which eDNA was detected were broader than for trawl detections.

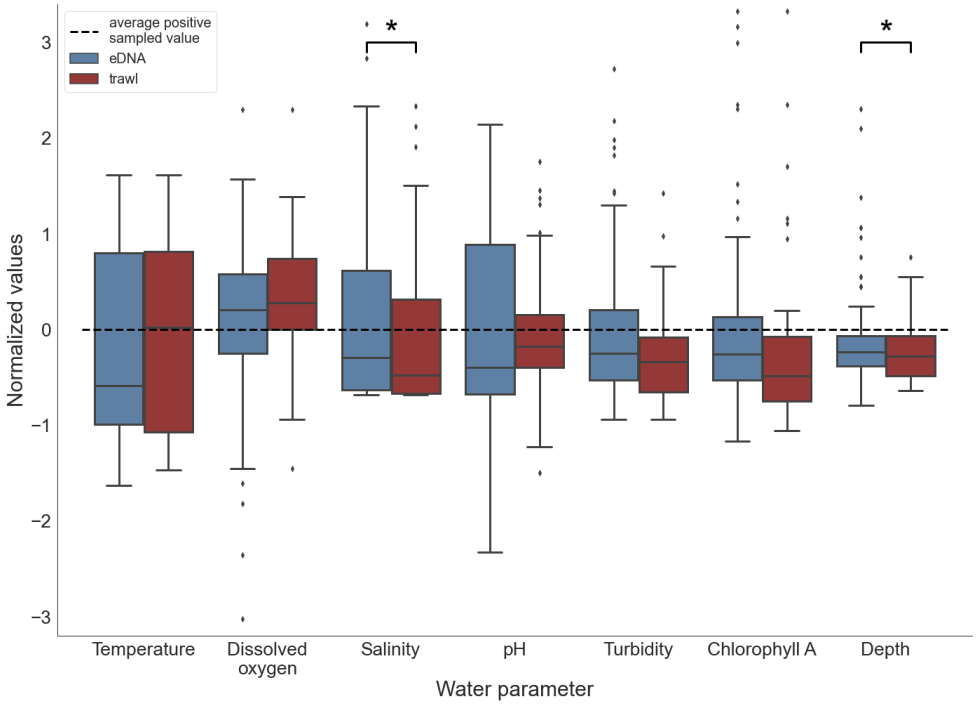


Figure 3.6: Boxplot of normalized environmental parameter values separated by detection method. Salinity and depth were significantly different between eDNA and trawl detections. * p-value < 0.05.

Logistic regression model

The logistical regression model allowed us to quantify and compare the effects of the water quality conditions and the fixed site classifications on the relative detection rates of trawling and eDNA. The WAIC ranking gave preference for the independent logistic regression model (Table 3.S3), although the Logistic Conditional Autoregressive model and feature interaction models had higher predictive power (Figure 3.S3, 3.S4). The effects of adding the conditional autoregressive factor to the logistic regression model did not modify substantially the impact of the measured water parameters, and the impact of the spatiotemporal autocorrelation was relatively small with values of 0.17 for eDNA and 0.23 for trawling (Figure 3.S3). The selected model had an overall accuracy of 0.87 and 0.72 for eDNA and trawling positive detections, respectively. Yet the precision for negative eDNA detection and positive trawl detection was 0.22 and 0.42, reflecting the imbalance between eDNA and trawl positive detections.

The estimated parameter values for each variable represent the influence of that variable on detection rates, in other words the bias on detection probability caused by that variable (Figure 3.7). The importance of each variable is reflected both by the magnitude of its displacement from zero, and the direction of its displacement (positive or negative) indicates how the variable influenced detection probability. Separation of eDNA and trawl bias estimates was an indication that the variable had a distinctly different influence on each survey's detections. Narrow distributions indicate higher confidence in the model's bias estimate. Broad bias estimates for fixed sites partly reflected fewer sample sites for each site class. Since there were few trawl detections without a corresponding eDNA detection, each of these instances was highly influential to the model bias predictions reflecting negative influence on eDNA detection probability and positive

trawl detection probability and are likely less representative of the true distribution of bias estimates for these variables.

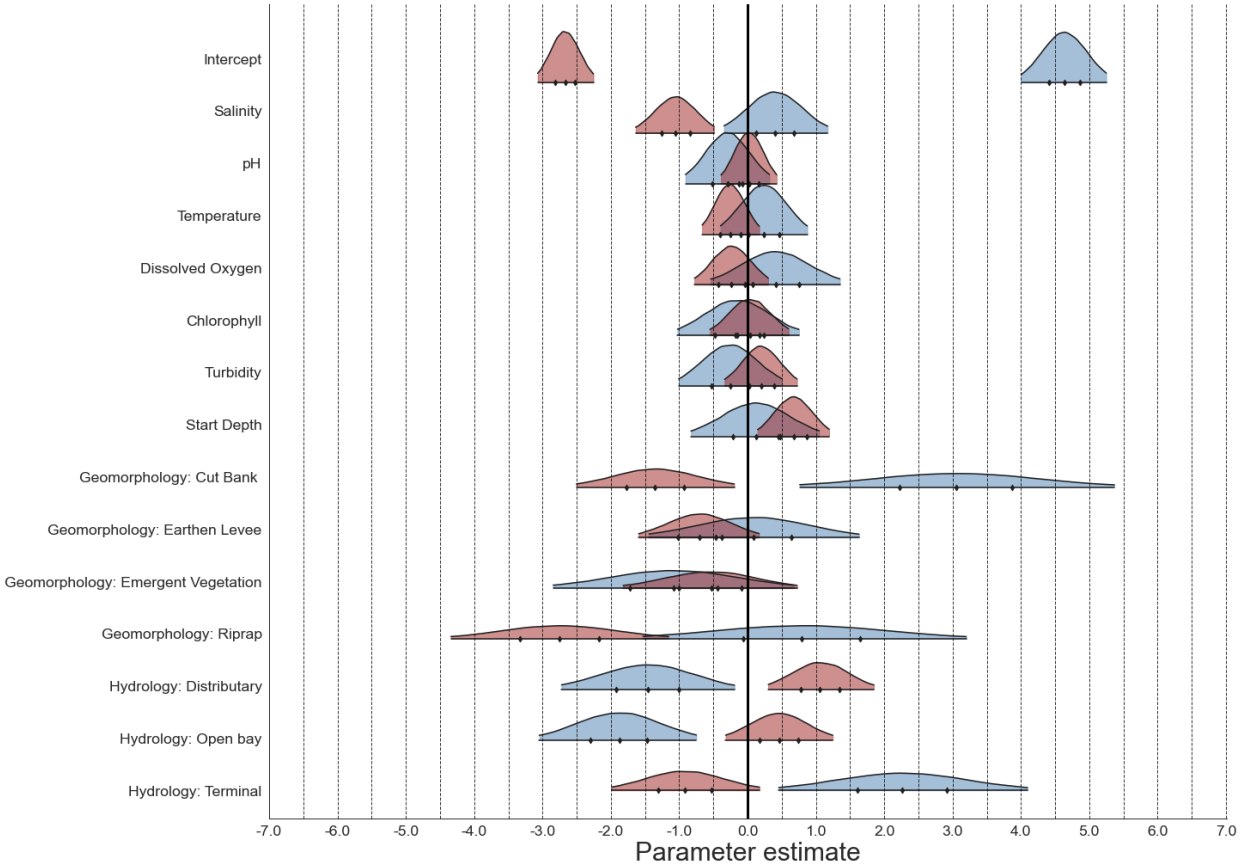


Figure 3.7: Estimated base detection rate (intercept) and detection biases caused by measured environmental conditions. Red - Trawl biases; Blue - eDNA biases.

The large estimate for the eDNA and trawl intercepts is representative of the difference in base detection rates between eDNA and trawl, reflecting the higher number of eDNA detection rates previously described (Figure 3.7). The model predicted that among sites where there is some indication of salmon presence (i.e., a positive detection) we would expect 99% of these sites to have a positive eDNA detection, while trawl would detect salmon at only 13% of these sites. This indicates that eDNA is expected to outperform trawl regardless of the environmental conditions

for the surveyed area as the difference in base detection rates are by far the most important parameter for the model. In contrast the model predicted that most of the site's environmental characteristics would have relatively small impacts on the detectability of either eDNA or trawl.

Although some of the bias estimates for the categorical site classes appeared highly influential, as indicated by the large parameter estimate means, the spread of the distributions showed a comparably low confidence in these estimates compared to the water quality variables. Among the hydrological conditions, the model predicted terminal channels would have increased eDNA detectability and reduced trawl detectability, while the opposite was observed in distributaries and open bay. Among the geomorphological conditions, the model predicted cut banks would increase eDNA detectability and reduce the detection rate for trawl. The other geomorphological classes are not predicted to greatly impact the detectability of either trawl or eDNA, except riprap, for which there were only two samples.

Salinity was the water quality parameter predicted to have the greatest impact on relative eDNA versus trawl detection rates. Water depth was the only other water quality variable with a bias estimate that did not substantially overlap zero, where higher depths are predicted to increase trawling detection.

Discussion

We estimated the ability of eDNA to detect Chinook Salmon relative to a trawl survey, while also examining how particular environmental conditions can bias the efficiency of the surveys. Understanding these biases improves our ability to interpret the complementary information and allows a comprehensive understanding of Chinook salmon distributions and habitat use.

eDNA has higher detection rate than trawling

Overall, eDNA had a nearly four times higher detection rate for Chinook salmon in the SFBD compared to trawling. This corroborates previous research comparing eDNA and net-based surveys for other fishes in an estuary and a river, which also found higher detection rates for eDNA [20,124]. The striking difference in detection rate between the methods demonstrates the challenge of detecting juvenile Chinook salmon using conventional field sampling methods in the shallow and complex edge habitats of the SFBD where juvenile salmon tend to rear. While we cannot completely rule out the possibility that some eDNA detections were caused by adult salmon migrating upstream through the study area, we expect such detections were rare or did not occur. Adult salmon tend to migrate upstream through Suisun Bay along the channel bottom and near the channel center along the main channel flow paths, well away from the shallow water shoals we targeted for sampling. Adult salmon are even less likely to occur near our channel and slough sampling locations within the marshes; in 33 years of monthly sampling in Suisun Marsh, adults have only been detected on 11 occasions, all occurring between mid-August and mid-December outside the sampling period for our study (UC Davis Suisun Marsh Survey unpublished data).

Other characteristics of the SFBD that may have contributed to the low detection by trawling relative to eDNA was the sheer scale of the waterbody to be sampled coupled with the low abundance of juvenile Chinook in most years. While a juvenile salmon must be contained in the volume of water sampled by a trawl net, an eDNA sampling transect need only pass through the eDNA plume emitted by a salmon, a probability that may be much higher than the odds of the salmon being contained in the net (i.e. eDNA may cast a wider net than a trawl). Additionally, eDNA does not actively evade the sampling equipment like a juvenile salmon. The superior sensitivity of eDNA for detecting salmon at lower densities was exemplified during the dry year

conditions of 2020 and 2021 when rearing salmon were known to be rare in the study area. During these years detection rates were reduced for eDNA but became nearly non-existent for the trawl survey (Table 3.2). This suggests that trawl efficiency may decline faster than eDNA as juvenile salmon become less abundant.

Terminal channels and cut banks strongly impacted eDNA versus trawl detectability

While eDNA performed better than trawl at detecting salmon in all types of water bodies and geomorphology, this was especially pronounced when sampling terminal channels and along cut banks. This was reflected both in the relative rates of detection for these habitat types (Figures 3.3,3.4), and in modeled biases (Figure 3.7). Consistent with our predictions (Table 3.1), we propose several direct mechanisms that may explain these observations. Terminal channels have higher water residence times than distributary channels and open bays[125], which may allow eDNA to achieve higher concentrations, increasing the probability of eDNA detection when salmon are present without any appreciable effect on the probability of trawl detection. The increased relative detection rates of eDNA when sampling along cut banks likely reflects a decreased efficiency of trawl. Undercut banks provide refuge for juvenile salmon where they may avoid the trawl net more effectively compared to other habitats, while at the same time having little influence on eDNA concentrations and therefore detectability. The higher eDNA assay performance in the upper marshes, where adult salmon are least likely to be present, corroborates our assumption that adults Chinook would not measurably increase the eDNA detection rate and that eDNA detections were primarily associated with juvenile individuals.

Salinity had the strongest bias in the eDNA and trawl detectability

Among the water quality variables associated with eDNA and trawl detections, salinity had the greatest difference between methods (Figure 3.6). This was also reflected in bias estimates of the logistic regression model, for which higher salinity was associated with a slight positive bias in eDNA detection, and much sharper negative bias in trawl detection (Figure 3.7). Salinity in our study area is generally higher during low outflow years. We believe the relationship observed between salinity and superior eDNA detection rates reflected eDNA's superior detection rates during the low outflow years when salmon were less abundant, rather than due to any specific mechanistic link between salinity and the efficiency of either survey method.

While modeling suggested water depth had a slight positive bias on trawl detection rates (Figure 3.7), observed eDNA detections occurred in slightly deeper water on average relative to trawl detections (Figure 3.6). The higher average depth of eDNA detections was driven mainly by a small number of eDNA detections at extreme water depths. These are not necessarily contradictory results. In general, trawl detections happened more frequently in the shoals along the main migratory corridor of the bay, with much sparser detections in the shallower upper marshes. Although detections were concentrated in the main corridor for both trawl and eDNA surveys, the eDNA survey was better at detecting salmon at shallower sites. This indicated that depth had a higher impact on trawl detectability than on eDNA detectability (Figure 3.7).

Variation in temperature, dissolved oxygen, pH, and chlorophyll had no discernable relationship to trawl and eDNA detectability. We hypothesized these variables would be linked to detection rates mainly through eDNA degradation processes. The lack of any relationship suggests any localized effect of degradation on eDNA concentration was overwhelmed by the influence of hydrological dispersion and dilution of eDNA[66,68]. While we experienced filter clogging during

field-based filtration at sites with higher turbidity, this did not seem to measurably affect eDNA detection as we hypothesized it would. The limited impact of most variables on the relative rates of eDNA versus trawl detections indicates that eDNA is reliable across a broad range of conditions compared to trawl surveys. This conclusion was supported by the underlying variation in modeled detection probabilities for eDNA, which was much lower than for trawl (Figure 3.7).

eDNA shown little to no cross-contamination between sites

One worry about eDNA surveys where the goal is to generate a fine-scale species distribution is that dispersion of eDNA from a site containing the target species can 'contaminate' another sampling site. Therefore, when designing a survey it is essential to consider the spatial scale of eDNA dispersion[39,126] to minimize spatiotemporal autocorrelation. Essentially this is done by minimizing the chance that two samples will detect eDNA from a plume emitted by a single organism. Autocorrelated detections from sampling locations that are too close together are less informative since they cannot be considered independent samples and are better interpreted as replicates from the same location. While eDNA dispersion dynamics can vary significantly between or within systems, and are difficult to predict, we compared results of the CAR model to the independent logistic regression model as an indicator of whether autocorrelation among eDNA detections was partly responsible for the apparent superiority of the approach. We found that the space-time autocorrelation was similar for eDNA and trawling, which supports the assumption that the eDNA detections represented samples from independent eDNA plumes, meaning eDNA detections represented the spatiotemporal locations of independent juvenile Chinook salmon. This was not surprising given the sampling design maximized the distance between samples, and most sample locations were separated by several kilometers or several weeks. The result of the model

comparison corroborates findings of other eDNA studies in estuarine or near-shore marine environments, where eDNA has been found to disperse rapidly with distance to low detectability levels, on the order of 10s to 100s of meters from the source organism[32,61].

The addition of the conditional autocorrelation did increase the accuracy of our logistic regression model, indicating that in a selected few cases detectability at one site was related to the detectability at a nearby site for both trawling and eDNA. This effect could be caused by the underlying out-migration pattern of juvenile Chinook salmon that tended to occur more frequently along the main migration corridor.

Effects of eDNA - trawl detection imbalance on model predictions

Since only sites where Chinook salmon were detected by trawling or eDNA were included in our modeling, our modeling approach measured relative biases, but did not estimate the true biases of detection probability, which could only have been estimated if true presence of juvenile Chinook salmon at sampling locations were known. In our comparative survey most of the detections were eDNA detections, many of them with no corresponding trawl detection. This provided a relatively rich dataset to examine conditions in which trawl was successful. In contrast, cases in which trawl detection was observed without a corresponding eDNA detection were rare (Figure 3.3), causing predictions of negative eDNA bias estimates and positive trawl bias estimates to depend on a small set of highly influential sampling events. This may explain why the only environmental conditions showing modeled eDNA and trawl biases clearly distinct from zero or each other were those conditions producing positive eDNA detection bias with corresponding negative trawl detection bias. The exceptions were detections at distributary channels and open bay sites, which we

interpreted as reflecting the clear superiority of eDNA detection rate observed in terminal channels (Figure 3.4B).

Integrating eDNA and conventional sampling

Given eDNA was better at detecting salmon under all conditions in our study, when deciding on an approach to maximize detection probability, does it matter that eDNA performed better than trawling under some conditions? Why not always use eDNA? For many studies, capture of salmon may be necessary, requiring the use of nets. In these cases, our study provides a better understanding of environmental conditions that may increase or lessen the chance of false negatives for shallow-water trawl surveys such as we conducted. For this reason, further comparative studies may be useful to verify our bias estimates.

It would also be useful to compare eDNA with a range of other net-based survey methods to understand how biases differ with these methods, which would in turn allow better informed selection and integration of survey methods for estimating Chinook salmon distributions and habitat use. In addition to paired surveys, interpretation of eDNA detections and detection biases would be greatly improved by models that account for the influence of hydrology on eDNA transport and detection probability. This would in turn further improve interpretation of pairwise eDNA and conventional survey techniques.

The higher potential of eDNA to detect Chinook salmon at low abundance makes it especially useful for defining the edges of salmon distributions, such as the start and end of juvenile presence at a given location during migration, and for understanding salmon habitat use when abundances are low. However, eDNA surveys have several characteristics that continue to limit its usefulness for some applications, although some of these limitations may be remedied by future

technological development. Perhaps foremost is the uncertainty of whether an eDNA detection is related to a large but far away eDNA source, or a close but small source – which we have termed the "density-distance conundrum". This uncertainty can be somewhat remedied by spatially intensive sampling[63], or external verification (as with trawl catch). Improvement in eDNA transport modeling is also a promising remedy to the density-distance conundrum.

eDNA is also currently unable to distinguish between adult and juvenile life stages[52,63], or to distinguish between salmon runs or populations, which is of particular interest in the Central Valley because only two of the four extant salmon runs are protected by federal and state endangered species acts. Different life stages and runs are sympatric in many places during many times of the year, limiting the ability of eDNA to inform life-stage or run-specific management. To meet these challenges, run/population-specific eDNA assays, and life-stage specific eRNA assays have been proposed or are in development, but until they are available it will be necessary to catch individuals to tabulate these demographic metrics.

Catching of individuals will always be necessary to track physiological characteristics like size and health, to obtain tissue and diet samples, or to determine whether salmon are naturally produced or of hatchery origin. One strategy that could be used in cases where catch of salmon is necessary is hierarchical sampling. eDNA sampling could be used first to determine the areas where juvenile Chinook salmon are present or at higher densities, followed by targeted trawling. This targeted sampling strategy would increase the trawling efficiency and reduce the impact of trawling on the environment, while continuing to improve our understanding and interpretation of eDNA data.

Future of eDNA surveys in the San Francisco Bay-Delta

As the use of eDNA expands, the technology and its applications are rapidly advancing. Relatively recent development of sequencing technology and metabarcoding pipelines allow eDNA surveys to target multiple species simultaneously, combining monitoring of endangered and invasive species in the same effort[127]. CRISPR-based assays such as SHERLOCK[128] and DETECTR[129] can increase sensitivity and soon will be able to differentiate different runs of Chinook salmon (Baerwald et al., in prep). In addition to potentially distinguishing between juvenile and adult life stages, eRNA assays are also promising as a potential solution to the density-distance conundrum, because eRNA degrades more quickly than eDNA[69]. This rapid degradation would limit transport distances and the distance between an eRNA source and detection location[130].

Strategies such as new forms of high volume eDNA sampling[47] and novel materials with high eDNA affinity such as hydroxyapatite[131,132] are promising tools to expand the use of eDNA surveys for turbid environments, circumventing problems that arise due to filter clogging. As discussed above, eDNA study precision can be increased by better integration with hydrological models[52,64], particularly in locations such as the SFBD where advanced hydrological models already exist (e.g. Delta Simulation Model II or DSM2)[133]. Last, automated eDNA samplers have been developed[134,135], allowing large-scale automated sampling in remote locations. Passive eDNA samplers that do not require water filtration are also an area of interest for their low-cost and ease of use[136,137]. These alternative eDNA sampling strategies can increase the numbers of samples taken, standardize timing between samples and allow data collection when in-person sampling is not feasible, further enhancing the toolkit for monitoring Chinook salmon populations in the SFBD.

References

95. Fisheries, N. Chinook Salmon (Protected). *NOAA* https://www.fisheries.noaa.gov/species/chinook-salmon-protected.
96. State of California. California EcoRestore. https://water.ca.gov/Programs/All-Programs/EcoRestore (2021).
97. Welcome to Science Action Agenda. https://scienceactionagenda.deltacouncil.ca.gov/.
98. Windell, S. Scientific framework for assessing factors influencing endangered Sacramento River winter-run Chinook salmon (Oncorhynchus tshawytscha) across the life cycle. (2017) doi:10.7289/V5/TM-SWFSC-586.
99. Johnson, R. C. *et al.* Science advancements key to increasing management value of life stage monitoring networks for endangered Sacramento river Winter-Run Chinook salmon in California. *San Franc. Estuary Watershed Sci.* **15**, (2017).
100. Shelton, A. O. *et al.* Environmental DNA provides quantitative estimates of a threatened salmon species. *Biological Conservation* vol. 237 383–391 (2019).
101. Bowers, H. *et al.* Towards the optimization of eDNA/eRNA sampling technologies for marine biosecurity surveillance. *Water* **13**, 1113 (2021).
102. Lehman, B. *et al.* Disease in Central Valley Salmon: Status and Lessons from Other Systems. *San Francisco Estuary and Watershed Science* vol. 18 (2020).
103. Nobriga, M. L., Michel, C. J., Johnson, R. C. & Wikert, J. D. Coldwater fish in a warm water world: Implications for predation of salmon smolts during estuary transit. *Ecol. Evol.* **11**, 10381–10395 (2021).
104. Munsch, S. H. *et al.* Science for integrative management of a diadromous fish stock: interdependencies of fisheries, flow, and habitat restoration. *Can. J. Fish. Aquat. Sci.* **77**, 1487–1504 (2020).
105. Mahardja, B. *et al.* Leveraging Delta Smelt monitoring for detecting juvenile Chinook Salmon in the San Francisco Estuary. *San Franc. Estuary Watershed Sci.* **19**, (2021).
106. Aguilar-Medrano, R., Durand, J. R., Cruz-Escalona, V. H. & Moyle, P. B. Fish functional groups in the San Francisco Estuary: Understanding new fish assemblages in a highly altered estuarine ecosystem. *Estuar. Coast. Shelf Sci.* **227**, 106331 (2019).
107. Zaiko, A., Pochon, X., Garcia-Vazquez, E., Olenin, S. & Wood, S. A. Advantages and Limitations of Environmental DNA/RNA Tools for Marine Biosecurity: Management and Surveillance of Non-indigenous Species. *Front. Mar. Sci.* **0**, (2018).
108. Maunder, M. N. & Punt, A. E. A review of integrated analysis in fisheries stock assessment. *Fish. Res.* **142**, 61–74 (2013).
109. Pimm, S. L. *et al.* Emerging Technologies to Conserve Biodiversity. *Trends Ecol. Evol.* **30**, 685–696 (2015).
110. Smart, A. S., Tingley, R., Weeks, A. R., van Rooyen, A. R. & McCarthy, M. A. I Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. *Ecol. Appl.* **25**, 1944–1952 (2015).
111. Mizumoto, H., Mitsuzuka, T. & Araki, H. An environmental DNA survey on distribution of an endangered Salmonid species, parahucho perryi, in Hokkaido, japan. *Front. Ecol. Evol.* **8**, (2020).
112. Schadewell, Y. & Adams, C. I. M. Forensics meets ecology – environmental DNA offers new capabilities for marine ecosystem and fisheries research. *Front. Mar. Sci.* **8**, (2021).
113. Carraro, L., Hartikainen, H., Jokela, J., Bertuzzo, E. & Rinaldo, A. Estimating species distribution and abundance in river networks using environmental DNA. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 11724–11729 (2018).
114. Standard Methods for Sampling North American Freshwater Fishes. *Standard Methods for Sampling North American Freshwater Fishes* (2009) doi:10.47886/9781934874103.fmatter.
115. Saito, T. & Doi, H. Degradation modeling of water environmental DNA: Experiments on multiple DNA sources in pond and seawater. *Environmental DNA* **3**, 850–860 (2021).
116. Collins, R. A. *et al.* Persistence of environmental DNA in marine systems. *Commun Biol* **1**, 185 (2018).
117. Grimaldo, L. *et al.* Forage Fish Larvae Distribution and Habitat Use During Contrasting Years of Low and High Freshwater Flow in the San Francisco Estuary. *San Francisco Estuary and Watershed Science* vol. 18 (2020).

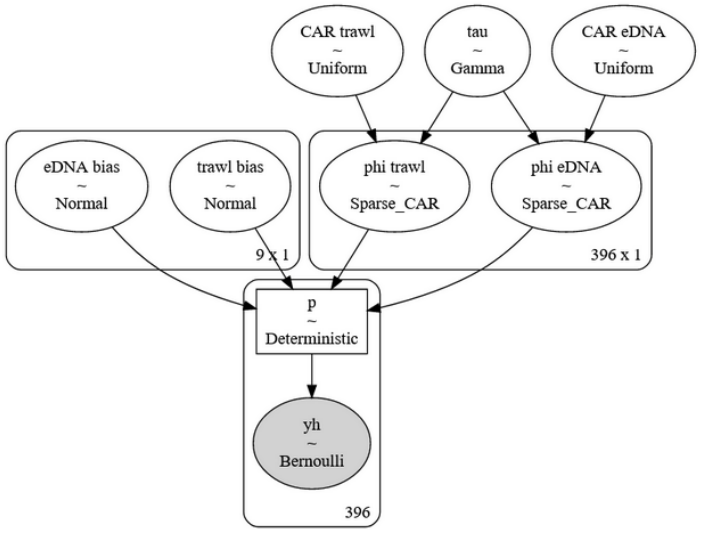
118. Starcevich, L. A. H., Irvine, K. M. & Heard, A. M. Impacts of temporal revisit designs on the power to detect trend with a linear mixed model: An application to long-term monitoring of Sierra Nevada lakes. *Ecol. Indic.* **93**, 847–855 (2018).
119. Stevens, D. L., Jr & Olsen, A. R. Spatially balanced sampling of natural resources. *J. Am. Stat. Assoc.* **99**, 262–278 (2004).
120. Munsch, S. H. *et al.* Warm, dry winters truncate timing and size distribution of seaward-migrating salmon across a large, regulated watershed. *Ecol. Appl.* **29**, e01880 (2019).
121. Kamoroff, C. & Goldberg, C. S. An issue of life or death: using eDNA to detect viable individuals in wilderness restoration. *Freshw. Sci.* 000–000 (2018).
122. Sales, N. G., Wangensteen, O. S., Carvalho, D. C. & Mariani, S. Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. *Environmental DNA* **1**, 119–130 (2019).
123. Martin, O. *Bayesian Analysis with Python: Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ, 2nd Edition*. (Packt Publishing Ltd, 2018).
124. Coulter, D. P. *et al.* Nonlinear relationship between Silver Carp density and their eDNA concentration in a large river. *PLoS One* **14**, e0218823 (2019).
125. Walters, R. A., Cheng, R. T. & Conomos, T. J. Time scales of circulation and mixing processes of San Francisco Bay waters. *Hydrobiologia* **129**, 13–36 (1985).
126. Carraro, L., Stauffer, J. B. & Altermatt, F. How to design optimal eDNA sampling strategies for biomonitoring in river networks. *Environmental DNA* **3**, 157–172 (2021).
127. Baloğlu, B. *et al.* A workflow for accurate metabarcoding using nanopore MinION sequencing. *Methods Ecol. Evol.* (2021) doi:10.1111/2041-210x.13561.
128. Baerwald, M. R. *et al.* Rapid and accurate species identification for ecological studies and monitoring using CRISPR-based SHERLOCK. *Mol. Ecol. Resour.* **20**, 961–970 (2020).
129. Williams, M.-A. *et al.* The application of CRISPR-Cas for single species identification from environmental DNA. *Mol. Ecol. Resour.* **19**, 1106–1114 (2019).
130. Marshall, N. T., Vanderploeg, H. A. & Chaganti, S. R. Environmental (e)RNA advances the reliability of eDNA by predicting its age. *Sci. Rep.* **11**, 2769 (2021).
131. Verdier, H., Konecny, L., Marquette, C. & Lefebure, T. Towards a simple way to collect eDNA using a 3D-printed passive sampler. *ARPHA Conference Abstracts* **4**, (2021).
132. Kumar, A., Kargozar, S., Baino, F. & Han, S. S. Additive manufacturing methods for producing hydroxyapatite and hydroxyapatite-based composite scaffolds: A review. *Front. Mater. Sci.* **6**, (2019).
133. State of California. DSM2: Delta Simulation Model II. https://water.ca.gov/Library/Modeling-and-Analysis/Bay-Delta-Region-models-and-tools/Delta-Simulation-Model-II.
134. Hansen, B. K. *et al.* Remote, autonomous real-time monitoring of environmental DNA from commercial fish. *Sci. Rep.* **10**, 13272 (2020).
135. Yamahara, K. M. *et al.* In situ autonomous acquisition and preservation of marine environmental DNA using an autonomous underwater vehicle. *Front. Mar. Sci.* **6**, (2019).
136. Simmons, M., Tucker, A., Lindsay Chadderton, W., Jerde, C. L. & Mahon, A. R. Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences* vol. 73 76–83 (2016).
137. Bessey, C. *et al.* Passive eDNA collection enhances aquatic biodiversity analysis. *Commun Biol* **4**, 236 (2021).
138. Lee, D. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat. Spatiotemporal Epidemiol.* **2**, 79–89 (2011).

Supplementary 3: alternative models

Logistic CAR model

For the CAR model[138] our interest is twofold. First, we are interested in measuring the space-time autocorrelation, and second, we want to know if this autocorrelation changes the effects of the water parameters on the biases of the methodology. Space-time autocorrelation converged to a value of 0.24 for trawling and 0.16 for eDNA, suggesting that the space-time autocorrelation exists for both trawling and eDNA, albeit with a small effect compared to the water parameters (Figure 3.S3). The observed conditional autocorrelation was similar for eDNA and trawling, meaning that the effects of space-time autocorrelation originated from the distribution of juvenile Chinook salmon and not due to eDNA transport over long distances. We also tested a CAR model with space and time as separated factors, which had a worse performance than the space-time model, also indicating that eDNA does not linger over long periods of time in the sampled sites. This effect might be explained by the surface sampling for eDNA, where eDNA cannot accumulate over time as it is transported downstream or settles in the substrate. The addition of the space-time CAR factor in the model did not greatly affect the detectability (intercept) of eDNA or trawl compared to the simpler logistic regression model (Figure 3.S3). Neither the biases measures for water condition nor hydromorphological characteristics changed considerably.

A



$$\hat{y} = logistic(X \cdot \beta_{method} + intercept_{method} + N(0, [\tau_{spacetime}(I - \alpha_{method} W_s)]^{-1}))$$

B

method	feature	mean	sd	hdi_3%	hdi_97%	feature impact SHAP
trawl	intercept	-4.19	0.31	-4.76	-3.61	2.91% base detection rate
	Salinity	-2.49	0.41	-3.25	-1.70	-5.15%
	Habitat: Terminal	-2.13	0.78	-3.58	-0.69	-2.1%
	Habitat: Distributary	1.69	0.58	0.61	2.77	2.79%
	Edge_Type: Cut Bank	-1.68	0.81	-3.22	-0.19	-1.32%
	Start_Depth	1.02	0.37	0.33	1.72	2.19%
	Edge_Type: Earthen Levee	-0.86	0.67	-2.11	0.42	-0.91%
	Edge_Type: Riprap	-0.72	1.08	-2.81	1.26	-0.02%
	Habitat: Shoal	0.58	0.60	-0.53	1.72	0.82%
	Turbidity	0.45	0.39	-0.30	1.16	0.83%
	pH	-0.11	0.31	-0.72	0.46	-0.24%
	Edge_Type: Emergent Vegetation	-0.09	0.91	-1.81	1.57	-0.03%
	Dissolved_Oxygen	-0.04	0.39	-0.79	0.68	-0.12%
	Chlorophyll	-0.02	0.44	-0.87	0.80	-0.03%
	Temperature	0.02	0.33	-0.61	0.62	0.05%
	CAR	0.24	0.16	0.01	0.54	NA
eDNA	intercept	6.67	0.44	5.85	7.53	99.9% base detection rate
	Habitat: Terminal	2.33	1.18	0.04	4.43	0.08%
	Edge_Type: Cut Bank	2.21	1.36	-0.27	4.88	0.06%
	Edge_Type: Emergent Vegetation	-1.96	1.19	-4.11	0.38	-0.04%
	Habitat: Shoal	-1.61	0.80	-3.21	-0.17	-0.09%
	Habitat: Distributary	-1.20	0.87	-2.80	0.45	-0.06%
	Edge_Type: Riprap	-1.19	1.52	-3.91	1.78	-0.0%
	Salinity	1.19	0.55	0.15	2.23	0.07%
	Turbidity	-0.59	0.53	-1.59	0.41	-0.04%
	Temperature	-0.44	0.47	-1.30	0.44	-0.04%
	Start_Depth	0.34	0.69	-0.98	1.61	0.02%
	Edge_Type: Earthen Levee	0.31	1.03	-1.66	2.17	0.02%
	pH	-0.11	0.45	-1.01	0.70	-0.01%
	Chlorophyll	-0.08	0.66	-1.28	1.21	-0.0%
	Dissolved_Oxygen	0.07	0.67	-1.21	1.32	0.0%
	CAR	0.16	0.14	0.00	0.44	NA

C

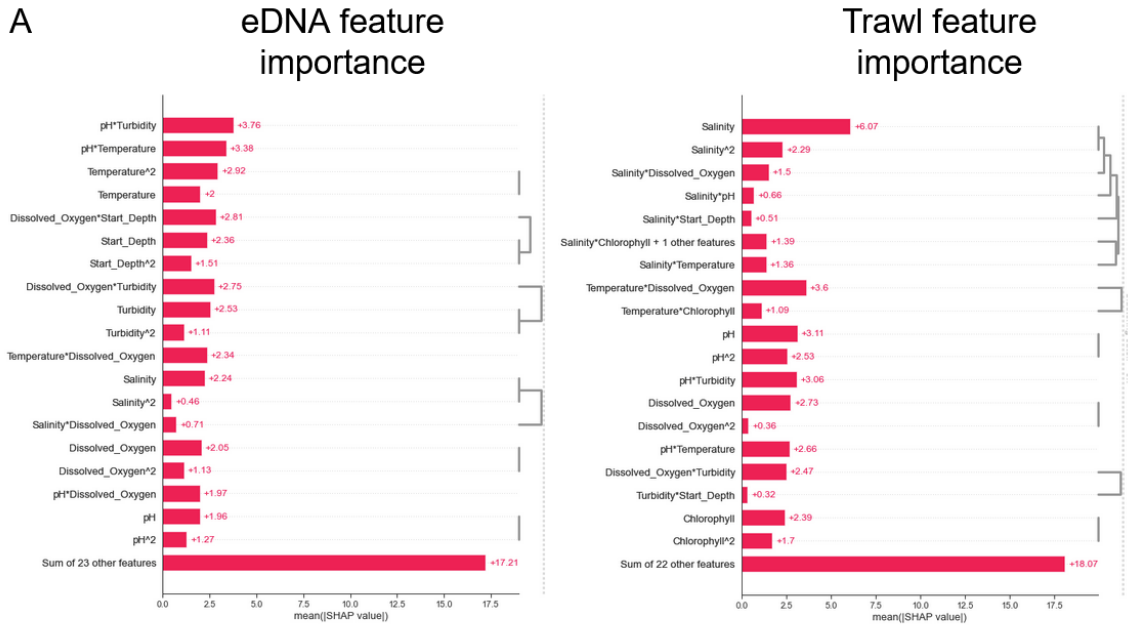
	eDNA			trawl		
	Negative detection eDNA	Positive detection eDNA	accuracy eDNA	Negative detection trawl	Positive detection trawl	accuracy trawl
precision	0.492222	0.950838	0.916094	0.888596	0.645725	0.832172
recall	0.465954	0.958227	0.916094	0.892999	0.642170	0.832172
f1-score	0.468537	0.954351	0.916094	0.890307	0.640144	0.832172
support	16.380000	181.620000	0.916094	151.363333	46.636667	0.832172

Figure 3.S3: A) Model representation of conditional autoregressive logistic regression model. B) Bias effect and confidence interval of each of the environmental parameters. For trawling, salinity had the highest influence on the detection rate. Meanwhile, for eDNA, terminal channels had the highest effects. Directionality of the effects was similar between the CAR model and the independent logistic regression model. C) Cross validation scores for both eDNA and trawling. We can observe that the CAR model had a good precision for all classes, especially considering the imbalance in frequency of positive and negative classes.

Feature interaction model

To understand if the biases identified for environmental conditions measured in this study have an interactive factor, we fitted the independent logistic model to the environmental conditions but in this case, we added the pairwise combination of each of the water parameters, producing in total 42 features that we tested for possible biases. Comparing the feature interaction model to the CAR and logistic models, we observe that the base detection rate (intercept) of eDNA and trawl were similar to the CAR model, with trawling base detection rate nearing 3% and eDNA detection rate nearing 99% (Figure 3.S4). In this case, trawling detection would only be seen if environmental conditions were favorable, with those conditions being related to low salinity.

A



B

	eDNA			trawl		
	Negative detection eDNA	Positive detection eDNA	accuracy eDNA	Negative detection trawl	Positive detection trawl	accuracy trawl
precision	0.355333	0.942109	0.896267	0.858151	0.553406	0.785139
recall	0.379740	0.945424	0.896267	0.860418	0.559909	0.785139
f1-score	0.345592	0.943330	0.896267	0.858146	0.549343	0.785139
support	15.576667	176.423333	0.896267	145.833333	46.166667	0.785139

Figure 3.S4: A) Bias effect of the most influential environmental parameters. For trawling, salinity and the product between temperature and dissolved oxygen had the highest influence on the

detection rate. Meanwhile, for eDNA, the interaction between pH and turbidity had the highest effect, followed by the interaction between temperature and other features. Directionality of the effects was similar between the CAR model and the independent logistic regression model. B) Posterior predictive check for both eDNA and trawling.

Supplementary 4: Model Comparison

When comparing models, we found that samples were mostly independent from each other for both eDNA and trawling. This modeling again points to the higher eDNA detection in the upper marshlands while trawling detection was mostly present in the wet year of 2020 and in open water locations. This observation implies that the more complex models were subject to overparameterization which was penalized by the WAIC. On the other hand, the significant increase of the classification scores from the more complex models point out that accounting for space-time autocorrelation and feature interaction is advised when developing models with elevated predictive power.

Table 3.S4: WAIC rank and weights for tested models

Model	Rank	WAIC
Logistic regression	0	-177.46
CAR Logistic regression	1	-220.08
Logistic regression with feature interaction	2	-255.79
CAR Logistic regression with feature interaction	3	-331.49