BUREAU OF RECLAMATION

Quantifying Fish Biomass X Distance from Environmental DNA Samples in a Hydrodynamically Complex Environment

Science and Technology Program
Research and Development Office
Final Report No. ST-2019-225-01

U.S. Department of the Interior

September 2021

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-09-2021	Research	10/22/2019 – 12/13/2020

4. TITLE AND SUBTITLE
Quantifying Fish Biomass X Distance from Environmental DNA Samples in a Hydrodynamically Complex Environment

5a. CONTRACT NUMBER
140R2019P0094, 140R2020P0099

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
1541 (S&T)

6. AUTHOR(S)
B. Mahardja (Biologist), A.A. Schultz, M. Johnston, S.M. Blankenship, M.B. Espe, G. Schumer, M.D. Bowen, C.A. Dean, K. Karpenko, M. Brennan and C. Hart

5d. PROJECT NUMBER
ST-2019-225

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Science Division, Bay Delta Office
U.S. Bureau of Reclamation – Interior Region 10
801 I Street, Suite 140, Sacramento, CA 95814

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Science and Technology Program
Research and Development Office
Bureau of Reclamation
U.S. Department of the Interior
Denver Federal Center
PO Box 25007, Denver, CO 80225-0007

10. SPONSOR/MONITOR'S ACRONYM(S)
Reclamation

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Final Report No. ST-2019-225-01

12. DISTRIBUTION/AVAILABILITY STATEMENT
Final Report may be downloaded from https://www.usbr.gov/research/projects/index.html

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A modeling and analysis tool, "artemis," was developed to analyze quantitative Polymerase Chain Reaction data from environmental DNA studies. Artemis models were fit to data from two field experiments in a tidal environment (Shag Slough, San Francisco Bay-Delta, CA): Experiment 1) 54 live Delta Smelt were placed in a live pen and Experiment 2) 25 Delta Smelt and 15 steelhead were placed in a live pen. Water samples were taken prior to live pen deployment in both experiments, providing control samples that represented zero target fish in Shag Slough (this was confirmed by no detections of Delta Smelt or steelhead DNA in the control samples). The ln[eDNA] declined with distance of the sampling point from the live pen. Furthermore, DNA concentration in samples increased with increasing Delta Smelt-biomass level. The tool developed in this project, artemis, is effective for designing an eDNA study and analyzing the results: once a site has been characterized, artemis can be used to select the number of samples to be taken and the distribution of those samples through time to achieve a desired probability of DNA detection.

15. SUBJECT TERMS
Delta Smelt, Steelhead, eDNA

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	THIS PAGE			Brian Mahardja
U	U	U			19b. TELEPHONE NUMBER *(Include area code)* (279) 234-1568

Mission Statements

The Department of the Interior (DOI) conserves and manages the Nation's natural resources and cultural heritage for the benefit and enjoyment of the American people, provides scientific and other information about natural resources and natural hazards to address societal challenges and create opportunities for the American people, and honors the Nation's trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities to help them prosper.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Disclaimer

Information in this report may not be used for advertising or promotional purposes. The data and findings should not be construed as an endorsement of any product or firm by the Bureau of Reclamation, Department of Interior, or Federal Government. The products evaluated in the report were evaluated for purposes specific to the Bureau of Reclamation mission. Reclamation gives no warranties or guarantees, expressed or implied, for the products evaluated in this report, including merchantability or fitness for a particular purpose.

Acknowledgements

The Science and Technology Program, Bureau of Reclamation, sponsored this research. We would also like to acknowledge René C. Reyes for his diligent permitting work. Without him, the use of live cultured Delta Smelt would not have been possible. We would also like to thank Brandon Wu for his expertise and data collection at the Tracy Fish Collection Facility. Special thanks to John Whitler and Sarah Anaya of Reclamation for program and contract assistance. Funding for this study was provided by the (California) State Water Contractors and a Reclamation - Science and Technology Program grant (Project ID: 225; contract no. 140R2020P0099 and 140R2019P0094).

Live animal use permits for Delta Smelt included 2017 - 2020 CDFW Scientific Collection Permit, SC-005544; 2020 - Present CDFW Scientific Collection Permit, Specific Use Permit ID: S-193170001-20084-001; CDFW MOU ID: 2081a-2019-0002-R3; 2019 USFWS §10(a)(1)(A) recovery permit: TE-027742-6.3; 2020 USFWS §10(a)(1)(A) recovery permit: TE-027742-6.5. Live animal use permit for steelhead trout was CDFW Scientific Collection Permit, specific use, S-190880001-20317-001. The views expressed are those of the authors and do not represent the official opinion of any employer, institution, or government agency.

Quantifying Fish Biomass X Distance from Environmental DNA Samples in a Hydrodynamically Complex Environment

Final Report No. ST-2019-225-01

prepared by

Bay-Delta Office – Science Division (California-Great Basin)
Brian Mahardja, Biologist (USBR – Interior Region 10, Bay Delta Office, Sacramento, CA)

Additional Authors

Andrew A. Schultz, USFWS – Interior Region 7, Vernal, UT
Myfanwy Johnston, Scott M. Blankenship, Gregg Schumer, Cheryl A. Dean, Katie Karpenko, Cramer Fish Sciences – Genidaqs, West Sacramento, CA
Matt B. Espe, Cramer Fish Sciences – Genidaqs, West Sacramento, CA and One Straw Consulting, Loveland, CO
Mark D. Bowen, Thomas Gast and Associates Environmental Consulting, Arcata, CA
Matt Brennan, Environmental Science Associates, San Francisco, CA
Christopher Hart, Fish Biologist (USBR – Interior Region 10, Tracy Field Office, Tracy, CA)

Peer Review

Bureau of Reclamation
Research and Development Office
Science and Technology Program

Final Report ST-2019-225-01

Report Title: Quantifying Fish Biomass X Distance from Environmental DNA Samples in a Hydrodynamically Complex Environment

BRIAN MAHARDJA
Digitally signed by BRIAN MAHARDJA
Date: 2021.09.22 10:05:03 -07'00'

Prepared by: Brian Mahardja
Biologist, U.S. Bureau of Reclamation

EVAN CARSON
Digitally signed by EVAN CARSON
Date: 2021.09.27 11:01:33 -07'00'

Peer Review by: Evan W. Carson
Fish & Wildlife Biologist, U.S. Fish and Wildlife Service

NICHOLAS BERTRAND
Digitally signed by NICHOLAS BERTRAND
Date: 2021.09.27 10:00:02 -07'00'

Peer Review by: Nicholas G. Bertrand
Fish Biologist, U.S. Bureau of Reclamation

Contents

Page

Executive Summary

This research was aimed at developing an environmental DNA (eDNA) methodology that can monitor for listed aquatic species upstream of water delivery facilities and improve water supply reliability. The eDNA methodology can detect listed species without capture, handling and "take," because it requires only water samples. The central research question was can eDNA monitoring in a hydrodynamically complex environment be used as a tool for monitoring listed fishes, e.g. Delta Smelt and steelhead. Our methods were two pronged: 1) develop a quantitative tool that can be used to design sampling and analyze eDNA data to provide high accuracy and less bias than commonly-used regression approaches and 2) conduct caged Delta Smelt and steelhead field experiments in a tidal environment to evaluate the effects of tidal action and distance from the DNA source on the quantity of eDNA in a water sample. The quantitative tool developed as part of this project, "artemis", was written in R and has been made publicly available. The artemis tool did produce less bias than commonly used regression approaches (see Chapter 1) and can be used to produce eDNA sampling designs that allow quantitative statements about the probability of detection. The caged fish experiments showed that the natural log of DNA quantity (ln[eDNA]) decreased with sampling distance from the cage just as expected. Furthermore, ln[eDNA] decreased through time after the removal of the caged fish from the water and artemis successfully estimated the range of values that described the strength of this negative relationship. An important conclusion was that Delta Smelt and steelhead DNA could be reliably detected in a tidally-influenced environment. However, a higher proportion of non-detection in replicates was observed in the tidal system than would have been expected in a unidirectional system (e.g. a canal) and drogue data from this project suggested this might be because the eDNA plume may not be homogenously mixed throughout the water column. So, increasing the number of samples in tidal-environment applications will be necessary to compensate for this phenomenon. Furthermore, under the range of conditions in which we worked, we found that Delta Smelt biomass was directly related to the probability of eDNA detection. In addition, when a positive DNA detection occurred, Delta Smelt biomass was directly related to the concentration of DNA in that sample. The next logical steps for the eDNA methodology and the validated quantitative tool artemis are field applications such as using eDNA sampling to guide target species capture or identification of listed species approaching a water intake.

1. Foundation

1.1 Preface

Water diversions are often reduced at certain times of the year when listed fish species are thought to be in the zone of influence of these operations. The studies described in this report lay the foundation for being able to reduce water diversions only when there is evidence that the listed fish are present using a non-invasive method that does not risk "take" of the listed species. This non-invasive approach, called environmental DNA, will also allow the U.S. Department of the Interior/Bureau of Reclamation (Reclamation) to contribute to listed species recovery by reducing water diversions when the listed fish's DNA is detected in the zone of influence.

When fish are present in a body of water, they routinely shed DNA. This DNA can be acquired through filtering a given amount of water from the water body in question. The DNA presence and concentration is then measured through quantitative polymerase chain reaction (qPCR) analysis. Thus, from a non-invasive water sample, DNA detection can provide indirect evidence of the presence of listed species. In this document, a tool is described that allows quantitative statements about the probability of detection when no positive DNA detections are made at a characterized sampling point. Thus, using this tool, the lack of detection can provide quantitative information regarding the absence of the listed species. In summary, in this report a quantitative tool is described, fit to data from field experiments, and used to evaluate the influence of distance of a sample point from fish-DNA source and biomass of the fish present at the source (a live pen that constrains the fish to a given location).

This report begins with Chapter 1 in which the quantitative modeling and analysis tool, "artemis," is described. Section 1.2 provides the complete manuscript sent to the journal "Environmental DNA" (https://onlinelibrary.wiley.com/journal/26374943). In Chapter 2, artemis models are fit to data from two field experiments. Sections 2.1 through 2.8 provide the complete manuscript that is in preparation for the journal "Public Library of Science – One" (https://journals.plos.org/plosone/). In Chapter 3, the influence of the target fish biomass in the live pen on the DNA concentration is described.

1.2 The artemis Package for Environmental DNA Analysis in R*

*: This chapter has been submitted to the journal Environmental DNA (https://onlinelibrary.wiley.com/journal/26374943)

1.2.1 Abstract

We introduce artemis, a new package for the R statistical programming environment designed to analyze quantitative Polymerase Chain Reaction (qPCR) data from environmental DNA studies. The artemis package directly addresses several challenges for analyzing eDNA survey data, including the censorship of data generated by qPCR analysis. The performance of artemis models is compared to commonly-used regression approaches using both simulated and experimental data. We demonstrate that the models in artemis have several favorable characteristics compared to standard regression approaches, yielding more accurate estimates, less bias, and better prediction performance while offering similar ease of use. Lastly, we discuss possible trade-offs and considerations for choosing the most appropriate analysis approach for eDNA survey data.

1.2.2 Introduction

Environmental DNA (eDNA) surveys provide an indirect yet noninvasive (Sigsgaard et al. 2015), unharmful, sensitive (Fernández et al. 2019; Pilliod et al. 2013), inexpensive (Akre et al. 2019), and rapid way to detect rare, cryptic, or invasive organisms in water (Fukumoto, Ushimaru, and Minamoto 2015; Goldberg et al. 2013). Given these strengths, sampling for and detecting eDNA using quantitative Polymerase Chain Reaction (qPCR) has gained popularity over the past 20 years (Dejean et al. 2012; Moyer et al. 2014) and is now broadly used to sample and indirectly infer presence of taxa in a variety of aquatic environments. However, widespread sampling and detection of eDNA by ecologists and conservation biologists is unstandardized, and the field is in need of standards for analysis and reporting (Fediajevaite et al. 2021). In addition, managers need assistance to understand the biological significance of a set of eDNA observations. This paper describes how probability statements about species presence can be developed using the artemis package and an eDNA sampling scheme.

Estimating [eDNA] via qPCR

In eDNA samples that undergo fluorescence-based quantitative real-time PCR, the amount of eDNA present in the sample is estimated from the number of quantification cycles of qPCR (hereafter the "Cq" value) completed before amplification takes place during qPCR. By this process, the concentration of eDNA is not directly measured. The relationship between eDNA concentration ([eDNA]) and Cq values is determined via a standard curve generated in the lab from the assay for the target species. The standard curve is specific to the lab reagents and techniques used. This standard curve formula typically takes the form:

$$Cq = \beta * ln([eDNA]) + a \quad (1)$$

Where α is the intercept and β is the slope for the standard curve equation and ln([eDNA]) is the natural logarithmic concentration of eDNA (ln[eDNA]). These coefficients are determined in the lab during the calibration process. A higher Cq value, i.e. more quantification cycles, corresponds to a

lower concentration of eDNA in a sample. Above a pre-determined threshold, additional quantification cycles are not attempted. Although this threshold would ideally correspond to when the [eDNA] concentration is zero, the threshold for Cq values typically corresponds to a non-zero eDNA concentration. Since "non-detection" is taken to be any sample which requires more than the threshold number of cycles to detect, a data censoring process occurs. Crucially, because the Cq values are dependent on the standard curve and hence the specifics of a particular lab, the censoring point is also lab dependent.

This censoring process can create several issues for analyzing qPCR data from eDNA samples. The most concerning issue is that failure to take the data censoring process into account may lead to biases in model estimates and invalid confidence or credible intervals. Additionally, when there is a large amount of data clustered at the censoring point, the estimated measurement error will be artificially low. This would in turn give rise to unrealistic expectations for the results of planned sampling efforts, and/or biased predictions.

To mitigate these issues in the analysis of our own eDNA data, we developed the artemis package for the R programming environment. In artemis, we implement a set of models to directly estimate the effect of predictors on the latent (unobserved) response variable, ln[eDNA]. This is accomplished by linking ln[eDNA]to the observed response variable (Cq) via the standard curve parameters. Our objects here are to introduce the censored latent variable models in the artemis R package, and to demonstrate how the artemis R package can be used in the analysis of qPCR data from eDNA samples. We compare the performance of artemis to several other commonly-used modeling approaches in eDNA research and discuss the benefits and trade-offs for each.

Common approaches to analyzing eDNA data

qPCR data from eDNA studies are often modeled via a binary response model, e.g. some form of binomial regression (Moyer et al. 2014; Song, Small, and Casman 2017; Hinlo et al. 2017) or occupancy models (Schmidt et al. 2013; Dorazio and Erickson 2018). In these, the response is a binary variable signifying the presence/absence of eDNA in the sample. In the case of occupancy models, presence/absence is a binary latent (unobserved) variable estimated from the observed binary detection of eDNA in the sample. In both, a binary variable indicates whether a sample had a Cq value below the censoring point, i.e., the detection threshold. Using standard and widely available statistical models and programmatic tools, these analysis methods allow for easy estimation of various covariates on the probability of the target species' presence.

Using a binary response for eDNA studies has the advantage of ease of analysis, as many statistical programs can estimate a binomial model. However, this ease of analysis within a study comes with a trade-off - it is difficult to comparing between studies. Binary response models are dependent on the threshold which defines a non-detection. This cutoff threshold is a function of 1) the standard curve, which defines the ln[eDNA] that corresponds to the threshold value, and 2) researcher decisions. For example, in response to the level of sensitivity of an assay, some researchers might use a maximum Cq threshold of 35 cycles (Huver et al. 2015), while others use 40 or even 45 cycles (Piggott 2016). Thus the ln[eDNA] which corresponds to the maximum Cq value for a particular set of extractions varies between studies, and therefore "presence" of a target species across studies can refer to different actual concentrations of eDNA in samples.

One solution to this quandary is to model either the Cq values themselves, or the concentration/copy number as a continuous response variable in a linear regression. Similar to a binary analysis, the effects of various covariates on the response can be estimated. In particular, using the ln[eDNA] or copy number avoids some of the issues outlined above. Similar to a binary response variable, modeling ln[eDNA] or copy number can be accomplished using common statistical software. However, as with binomial implementations, the continuous Cq, concentration, or copy number is still associated with the detection threshold: since the standard curve, which is lab-dependent, defines the concentration at which further qPCR cycles are not attempted, the standard curve defines a statistical censoring point for the response variable, regardless of whether it is modeled as binary or continuous.

Statistical censoring is a well-studied phenomenon where data values above or below a certain threshold value are recorded as the threshold value. Conceptually, this represents a partially missing value - it is known that the value is beyond the threshold, but its exact value is unknown. A naive analysis of censored data which does not take this into account (such as the linear modeling of Cq, eDNA concentration, or copy number described above) will overestimate the certainty associated with values near or at the threshold. In eDNA studies, when all ln[eDNA] or copy number values are relatively high (i.e. far from the censoring point) the censoring point will have negligible impact on the analysis. However, when there are many values near the censoring point (i.e. near the limit of detection), estimates will be biased.

Therefore, there is a need to take the above issues into consideration in eDNA analyses, while also providing the ease of use of common statistical programs.

Modeling qPCR eDNA Data with artemis

We created the artemis R package to implement a set of Bayesian censored latent variable models, which mitigate the issues with commonly-used statistical analysis techniques on qPCR data. At its core, artemis is a specialized Generalized Linear Model (GLM), where the predictors are assumed to affect the latent response variable additively,

$$ln[eDNA]_i = X_i\beta \tag{2}$$

where β is a vector of effects on ln[eDNA]^i, and X^i is a vector of predictors. Since artemis directly models the effect of the predictors on the latent variable, ln[eDNA], it is unnecessary for the researcher to back-transform the data prior to modeling. Internally, artemis conducts this conversion using the user-supplied values for the standard curve formula,

$$\hat{Cq}_i = a_{std\_curve} + \beta_{std\_curve} * ln[eDNA]_i \tag{3}$$

Where α_{std_curve} and β_{std_curve} are fixed values from calibration in the lab prior to qPCR. Internally, the back-transformed ln[eDNA]i values are considered a sample with measurement error from the true ln[eDNA]^i value (ln[eDNA$\hat{}$]^i) in the extract, with values above the threshold censored to be equal to the threshold (i.e. a truncated normal distribution),

$$ln[eDNA]_i \sim Trunc.Normal(ln[\hat{eDNA}]_i, \sigma_{Cq}, U) \tag{4}$$

Where the observed ln[eDNA]^i values are censored at the predetermined concentration threshold, U. This threshold concentration value is internally calculated from the user-supplied threshold on Cq.

Importantly, the ln[eDNA$\hat{}$]^i values in the model are not censored, allowing the latent variable to reflect the "true" log-concentration of eDNA beyond the censorship point. To appropriately condition model estimates on the censoring process, the likelihood that a sampled ln[eDNA] value will exceed the threshold is a function of the measurement error and the estimated latent ln[eDNA$\hat{}$]^i value. We calculate this likelihood using the normal cumulative distribution function, $\Phi()$,

$$Pr(ln[eDNA]_i > U) = 1 - \Phi(ln[\hat{eDNA}]_i - \mu_i/\sigma) \quad (5)$$

Thus, the models in artemis account for the data censoring process by estimating the probability that the observed value will exceed the threshold.

This model formulation makes several assumptions, namely that 1) ln[eDNA] is uniform within a sample, 2) ln[eDNA] is sampled with normally-distributed errors with censorship at the detection threshold, and 3) there are no false detections, i.e. the measurement error cannot result in a positive detection when the target species' eDNA is not present in the sample.

Importantly, this formulation produces estimates of the effect sizes which are modeled directly on ln[eDNA] or copy number, rather than Cq, therefore are independent of the standard curve and can be compared between studies that use different standard curves. This model formulation also accounts for the data censoring at the upper limit of qPCR cycles, handling uncertainty and reducing bias in the estimates.

Overview of artemis Functionality

In addition to the modeling framework described above, the artemis package includes several utility and convenience functions associated with the planning and analysis of eDNA surveys and sampling. Taken all together, the functions in the artemis R package can be grouped into a few categories: modeling, simulation, post-hoc analyses, and utilities. Modeling and simulation are primarily introduced here, with detailed vignettes available for post-hoc analyses and utilities included in the package installation or via the package website (fishsciences.github.io/artemis).

Modeling

The modeling functions in artemis are intended to be drop-in replacements for lm() or glm() (R Core Team 2021) while utilizing the generative model as described in the previous section. An example call to the modeling function eDNA_lm() is,

```
eDNA_lm(Cq ~ Distance_m,
          data = eDNA_data,
          std_curve_alpha = 21.2,
          std_curve_beta = -1.5)
```

Note that the parameters for the conversion to ln[eDNA] are user-provided. Just as with other modeling functions in R, the user provides a formula for the model in the form response ~ predictors.

The data for the model formula is supplied as a data frame object passed to the data argument of eDNA_lm(). Although the model technically uses the latent variable ln[eDNA] "under the hood" as the response to the predictors, the formula in eDNA_lm() is expressed on Cq, since Cq is typically present in the raw output of qPCR analysis.

Internally, the conversion between Cq and ln[eDNA] is conducted using standard curve coefficients provided by the user. Importantly, these can be specified as a vector of a_{std_curve} and β_{std_curve} values corresponding to the rows of the user's input data. This allows the use of multiple standard curves within the same model. Thus, data from different studies or data which use multiple standard curves can easily be analyzed together.

For mixed- or random-effects models, the modeling function eDNA_lmer() can be used. The formula syntax follows the convention of lmer() (Bates et al. 2015) and specifies the random effects in the model with,

```
(parameter | grouping variable)
```

Both model types are fit using a Bayesian model fit via the Stan MCMC program (Stan Development Team 2021). In both artemis modeling functions, additional parameters can be passed to control the MCMC algorithm via the "..." arguments.

Simulation

The simulation functions sim_eDNA_lm() and sim_eDNA_lmer() allow researchers to see the implications of assumptions on the expected concentration of eDNA, e.g. how ln[eDNA] responds to hypothetical environmental effects. This can be important both to understand effects estimated by an artemis model fit to collected data, and/or as method to design a study prior to collecting data.

The simulation functions are based on the generative model outlined previously and are populated similarly. As with the artemis modeling functions, the relationships in the simulation are specified using a model formula. Then, the user provides a set of parameters (i.e. the "effects") for the linear model on ln[eDNA], the standard curve coefficients, and the measurement error on ln[eDNA]. Lastly, the user provides the covariate levels for which simulations are desired and the number of simulations to generate.

For example, a simulation call might be specified as,

```
sim_eDNA_lm(Cq ~ distance + volume,
                  variable_list = list(Cq = 1,
                                       distance = c(0, 10, 50),
                                       volume = c(50, 200)),
                                       betas = c(intercept = -10.6,
                                       distance = -0.05,
                                       volume = 0.01),
                    sigma_ln_eDNA = 1,
                    std_curve_alpha = 21.2,
```

```
                    std_curve_beta = -1.5)
```

Installation

The artemis package is open-source, and will be or is available via the Comprehensive R Archive Network (CRAN), and can be installed from within R via install.packages("artemis"). The latest development version and source code is also available via GitHub at https://github.com/fishsciences/artemis.

1.2.3 Methods

To demonstrate the strength and utility of the artemis package for modeling eDNA data, we compare the models in the artemis package to standard mixed-effects analysis. To ensure that the results were directly comparable, competing models were fit using the rstanarm R package (v2.21.1) (Goodrich et al. 2020), a Bayesian modeling package. The stanarm package was chosen for several reasons. First, the rstanarm and artemis packages both use the Stan probabilistic programming language as a back-end to estimate parameters (Stan Development Team 2021). Furthermore, many of the defaults in artemis functions mirror those in rstanarm, which in turn mirror those of the (g)lm (R Core Team 2021) and (g)lmer (Bates et al. 2015) functions. Both artemis and rstanarm models support similar model comparison metrics, allowing one-to-one comparisons between models.

Simulated data

We simulated 500 datasets using the generative process outlined previously in Section 1.3 with known parameter values. We then used two different eDNA modeling approaches to recover (estimate) the original parameters used to simulate the data: a linear mixed-effects model, and artemis's censored-data mixed-effects model.

For this task, only models which directly estimate effects on the latent or back-transformed ln[eDNA] values were compared. Hence, the simulations first simulated ln[eDNA] values then converted these to Cq values via a hypothetical standard curve. These Cq values were used directly by the artemis models (which internally convert to ln[eDNA]) but had to be back-transformed to ln[eDNA] prior to modeling with rstanarm. For the purposes of this demonstration, we simulated an experiment with two different filtered volumes, 50 and 100mL, and ten different distances from eDNA emission source spread equidistant from 0 to 1000 m. Additionally, we simulated three filters per measurement and three technical replicates per filter.

Note that while artemis contains similar functions to simulate data, we opted to replicate the data in base R, outside of artemis's functions, for transparency. All data and code is freely available at github.com/fishsciences/artemis_methods and included as supplemental information to this publication.

Experimental data

In addition to simulated data, we also employed a previously-collected experimental data set for the task of model comparison. The data used were the qPCR results from a Delta Smelt "live pen" experiment conducted in the primary intake channel of the Tracy Fish Collection Facility (TFCF).

The TFCF collects fish before they are permanently entrained into Reclamation's Central Valley Project (CVP). This water intake is located in the southwest portion of the CVP, in the Sacramento - San Joaquin River Delta, California, USA (Bowen et al. 2004). The experiment was completed on 2017-08-02, part of a series of live pen experiments at the CVP in coordination with the TFCF in August-September of 2017. All water filtering, eDNA extraction, and qPCR analysis procedures associated with the experiment followed those described in Schumer et al. (2019).

In this experiment, one-hundred live (cultured) Delta Smelt were placed in a rigid, meshed enclosure (the "live pen") that was suspended from the primary louvers in the intake channel. Two sets of three water samples each were then filtered at 10m intervals, beginning at 10 m and ending at 50 m from near to far in the downstream direction relative to the live pen. The first set of three samples pulled 50mL of water through each filter at each interval; the second set of three samples pulled 200mL per filter at each interval. This procedure was then replicated in the upstream direction (from far to near) relative to the live pen. Each filter was extracted and analyzed three times with qPCR (three technical replications). The qPCR data from these experiments is plotted in Figure 1.1. Approval for the experiment was via United States Fish and Wildlife Service (protocol 2017 §10(a)(1)(A) recovery permit TE-027742-5) and California Department of Fish and Wildlife (protocol 2017 MOU under Scientific Collecting Permit SC-005544).

To model the experimental data, we assume a fixed effect of distance (m) and volume sampled (mL). For mixed-effects models, we assume a random intercept term for each unique filter (FilterID).

Model comparison

Two different methods were used to compare performance between the standard mixed-effects linear model and the artemis model. First, the Pareto-Smoothed Leave-one-out Information Criteria was calculated for each model using the loo package in R (Yao et al. 2017). This metric assesses a model's performance by approximating predictive performance on out-of-sample data. This provides a measure of how well the model performs relative to its risk of overfitting to the data. Next, each model was used to predict the expected response values for a second experimental data set, collected in the same system and following the same procedures as described above. From these predictions, the Root Mean Square Error (RMSE) was calculated. This allowed for a realistic example of out-of-sample prediction error for each model.

To broaden the comparisons to other models commonly used to analyze eDNA data, we also compared the classification performance between binomial models. First, we fit a binomial mixed-effects model to the first experimental data set. The modeled response was "presence," a binary variable indicating whether the Cq of an observation was lower than the detection threshold, according to its standard curve. The model was fit similarly to the mixed-effects linear model, again using the rstanarm R package. We then generated a predicted value (presence or absence) for each observation in the in-sample data set (the data used to estimate model parameters) and the out-of-sample data set (data not used to estimate parameters). Since the Bayesian model produces many posterior predictions, we took a predicted "presence" to be when 50% or more of the posterior predictions were presences. To compare to the model fit with artemis, predictions were generated for each observation. The median predicted Cq value was taken as the predicted value, and each prediction was then classified as a presence or an absence, according to whether the predicted value was below or above the detection threshold.

Finally, to compare the classification performance for the in-sample and out-of-sample predictions, the precision (the proportion of positive predictions that agree with those found in the data) and recall (the proportion of positives in the data that were correctly predicted) were calculated. A precision of 1 indicates the model had no false positives. A recall of 1 indicates the model had no false negatives (Google 2020).

1.2.4 Results

Simulated data

Based on 500 simulated data sets, the standard linear mixed-effects models produced biased estimates. The estimates for the four generative model parameters (Intercept, βdistance, βvolume, and σ), were all skewed from the true values used to generate the data set (Figure 1.2). One effect of this bias was the 95% Credible Intervals produced from the standard model did not include the true values in 14% of cases, higher than the expected rate.

By contrast, the estimates produced by the censored-data mixed effects model in artemis were centered around the true values for all four parameters. Additionally, the 95% credible intervals included the true parameter values in all but 3.8% of cases, which is within the expected range.

Experimental data

When fit to the same experimental data, artemis models demonstrated favorable characteristics compared to alternative models. While the differences in parameter estimates were relatively small (Table 1.1), the predictive performance as measured by the Pareto-Smoothed Leave-One-Out Information Criteria suggested the artemis models fit the data better compared to widely-used alternatives (Table 1.2). Furthermore, when the parameter estimates for each model were used to generate predictions for a second set of experimental data, the artemis models had lower Root Mean Square Error (RMSE) on the predictions.

In a comparison between artemis models and a binomial mixed-effects model fit to the same data (Table 1.3), the binomial data had better precision and recall for the in-sample data relative to the artemis model (0.92 vs. 0.72 precision; 0.94 vs. 0.88 recall). However, when used to classify out-of-sample data, the binomial model's performance showed similar results to the artemis predictions in precision (0.57 vs 0.56 precision), but worse recall (0.69 vs 0.77 recall). This suggests that the binomial model was overfit to the original data. These metrics suggest the binomial model would produce more false negatives compared to the artemis classification predicting the same data.

1.2.5 Discussion

We were motivated to create the artemis package out of the need for better contextualization of both non-detections and the strength of positive detections in eDNA studies. To answer questions such as "Would we have detected the species of interest if it were within a certain distance from the sampling site?" we needed to understand the parameters of the sampling regime in that system. For example, how quickly did the concentration of eDNA decrease with distance? How much did the

filtered volume of water increase the concentration of eDNA in the filter? We first tried to address these questions using off-the-shelf statistical models, but the results were not robust.

Our analysis demonstrates the bias potentially introduced by data censorship, as well as how that bias can be mitigated with the use of models which explicitly account for censorship. In the case of censored data, like the data generated from a qPCR analysis, the assumptions of a standard linear model are not met, and we cannot reliably expect valid intervals or unbiased model estimates from it. A model that accounts for the data generation process is needed, but custom statistical models are often inaccessible to the average researcher. By mimicking the syntax of the lmer package and the lm function in base R, the artemis R package provides a simple replacement for standard linear models in R, and also contains several utilities and convenience functions for working with eDNA data and model estimates. More importantly, however, the artemis package implements statistical models which account for the data censorship inherent to qPCR data.

In situations where the primary interest is to understand how the concentration of eDNA responds to various sampling or environmental factors, the models in artemis are drop-in replacements for standard linear models. However, when the primary interest is to estimate presence or absence of a target species, artemis requires some additional steps. First, an artemis model needs to be fit to observed data from the system of interest, for example from a live pen or other controlled experiment. Then, using the estimates, we can produce predictions of the probability of positive detection for a certain set of conditions. In other words, after calibrating our estimates, we can predict how likely we would be to not detect a species if it were present, given some conditions on the sampling. While not the exact corollary to traditional occupancy or presence/absence modeling, this may be more informative in some situations.

Environmental DNA sampling studies are often of interest where other forms of sampling are difficult or impossible, or to supplement an existing sampling procedure (Adams et al. 2019). The species of interest may be quite rare. For these situations, a standard presence/absence analysis might lead one to conclude absence, with uncertainty. By contrast, while they do require the additional step of parameterizing for a given sampling procedure, environment, and design, the models implanted in artemis allow probability statements like the following to be made: "If the species of interest were within 50 m of our sampling site, we would have had 95% probability of detecting it."

The censored latent variable models in artemis performed better in multiple metrics compared to standard linear mixed effects models, against both simulated and observed data. While we demonstrate here that biased estimates are possible when fitting standard linear models to censored data, the degree of bias will depend on the exact characteristics of the data set. In general, the more observations which experience censorship, the more bias there can be in the estimates. As we observed in the model comparison task, this can invalidate computed Credible Intervals.

A mechanistic model of eDNA movement and diffusion from point source might be preferable to statistical models such as those estimated by artemis. In our experience conducting eDNA field surveys and experiments, however, parameterizing mechanistic models of particle diffusion and flow to contextualize eDNA data is difficult, cost-prohibitive, and time-consuming, and is often not fully transferable to new environments. The artemis package provides a modeling framework that allows for a simple method to generate probability statements from our qPCR data, which in turn allows us to quickly make decisions about future eDNA sampling designs.

Future work

Although artemis was designed to model ln[eDNA] values as the latent response, the functions are sufficiently general to allow other response variables as well. For example, copy number can easily be used, so long as the correct standard curve parameters are provided for the conversion from Cq to copy number. However, the functions have not been extensively tested on copy number, and there might be some aspects of copy number as a latent response variable which require a different model parameterization. More testing is needed.

It is generally known that qPCR results in more variable measurements as concentrations of eDNA decrease. During the Beta development stage, we implemented a basic parameterization to allow for this in the model. This initial attempt was based on normally-distributed errors which increase as concentrations decrease. However, the errors appear to be more likely Poisson-distributed. Initial testing reveals some issues with using normally-distributed errors. Further work is needed to evaluate and implement the most robust method to model the mechanisms involved.

1.2.6 Conclusion

The need for non-invasive, cost-effective sampling, especially for low-density, cryptic, or difficult-to-sample populations, is growing and will continue to drive development of molecular-based methods (Adams et al. 2019). This need also includes effective early-detection for invasive species (Milián-García et al. 2021; Sepulveda et al. 2019). As monitoring programs continue to adopt eDNA methods more broadly, there is an accompanying demand for robust, reliable, open-source analysis and modeling tools.

To address many of the issues that arise when analyzing data from qPCR analysis using standard statistical methods, we developed the artemis R package, which includes many utility and convenience functions in support of eDNA research and analysis. Importantly, the artemis package provides drop-in replacements for linear modeling functions in R. It does this with latent variable models that are customized to the data generating process inherent to qPCR data derived from eDNA samples. These models show less bias and more predictive accuracy for censored qPCR data when compared to standard linear models.

1.2.7 Figures

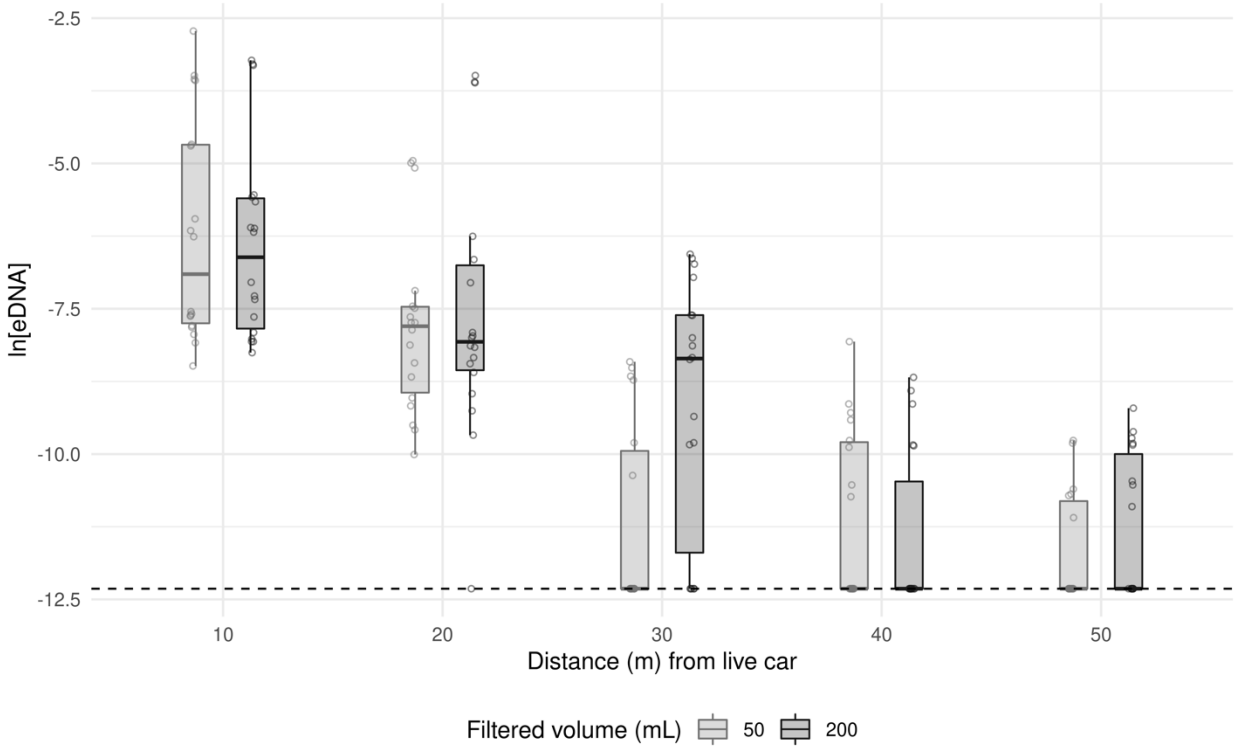


Figure 1.1. Plot of experimental live pen data from two experiments at the Tracy Fish Collection Facility (TFCF) in the Central Valley Project (CVP), California, USA. These data were used to test different analysis techniques for environmental DNA (eDNA) survey data. Each point represents one technical replicate. The Cq of each technical replicate has been converted to ln[eDNA] via the standard curve, where $ln[eDNA] = (Cq - \alpha_{StdCurve})/\beta_{StdCurve}$. The dotted line represents the threshold of detection for the assay.

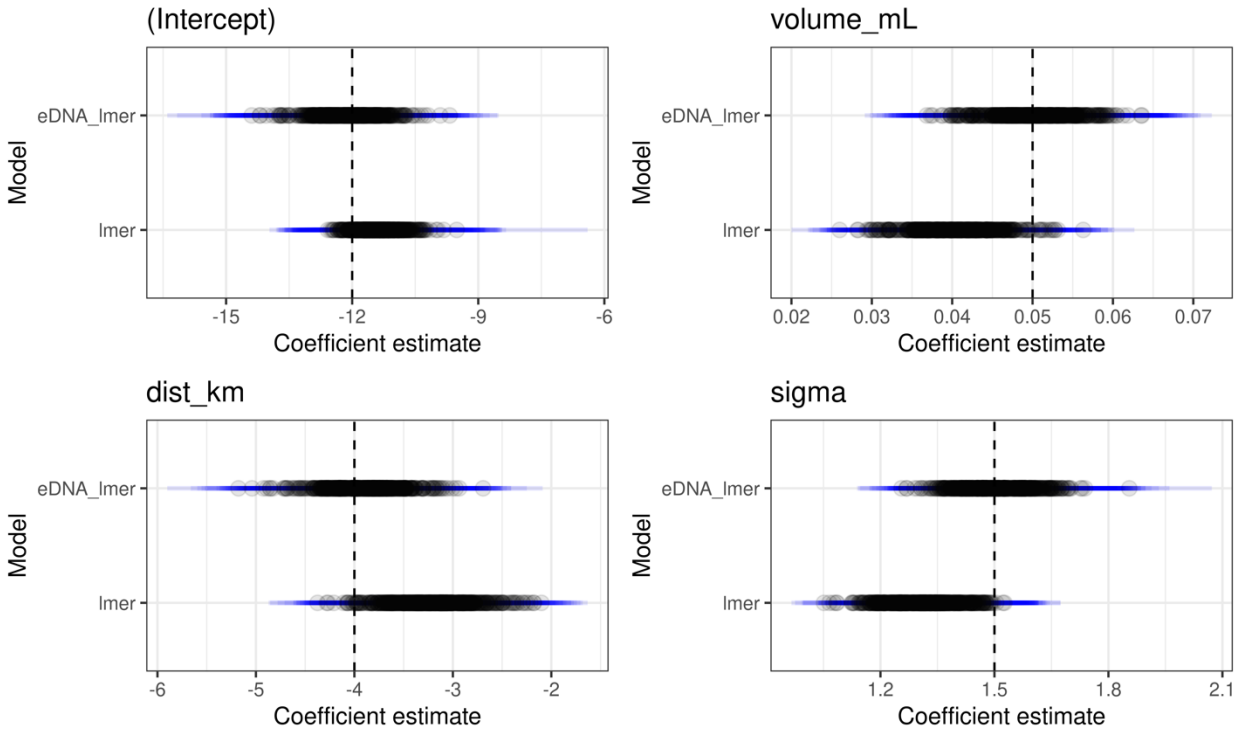


Figure 1.2: Comparison of lmer vs eDNA_lmer to recover parameters from 500 simulated datasets. Blue areas are the 95% Credible Intervals, while black areas are the median parameter estimates. The "true value" used in the simulation is marked as a dashed line.
Note: lmer = "Linear Mixed-Effects Models" package in base R (Bates et al., 2015); (https://www.rdocumentation.org/packages/lme4/versions/1.1-27.1/topics/lmer)).

1.2.8 Tables

Table 1.1: Model estimates and 90% credible intervals (CI) for the fixed effect parameters from the rstanarm and artemis packages fit to the same data

Parameter	rstanarm	90% CI	artemis	90% CI
(Intercept)	-5.6	-6.6 – -4.5	-5.4	-6.8 – -4.1
Distance_m	-0.14	-0.16 – -0.11	-0.17	-0.21 – -0.14
Volume_mL	0.0029	-0.002 – 0.0079	0.0049	-0.002 – 0.012
sigma	0.86	0.77 – 0.95	1.2	1 – 1.4

Table 1.2: Model comparison using Pareto-Smoothed Leave-One Out Information Criteria, ranked in descending order of predictive performance (top row is best). Lower values of Leave-One Out Information Criteria indicate better predictive performance.

Model	Leave-One Out Information Criteria	Effective Number of Parameters	Difference in Expected log-predictive density	Std. Error of the difference in expected log-predictive density
artemis: mixed-effects	497.9	49.4	0.0	0.0
rstanarm: mixed-effects	520.1	51.7	-11.1	17.1
artemis: fixed-effects	643.6	3.8	-72.9	12.8
rstanarm: fixed-effects	725.8	3.6	-114.0	16.5

Table 1.3: Model classification comparison of in-sample (data used to fit the original model) and out-of-sample (data not used to fit the original model) using precision and recall. Precision is the proportion of classifications which were in fact correct. A precision of 1.0 indicates no false positive classifications. Recall is the proportion of actual positives correct. A recall of 1.0 indicates no false negatives.

Model	Precision	Recall
binomial: in-sample	0.92	0.94
artemis: in-sample	0.72	0.88
binomial: out-of-sample	0.57	0.69
artemis: out-of-sample	0.56	0.77

2. Tidal influence on environmental DNA detections for Delta Smelt*

*: This chapter will be submitted to the journal PloS one (https://journals.plos.org/plosone/)

2.1 Abstract

Current ecosystem conditions within the San Francisco Estuary are considered inhospitable to many native estuarine species and has imperiled the endemic Delta Smelt (*Hypomesus transpacificus*). Programmatic monitoring regimes conducted by Government Agencies are insufficient for associating Delta Smelt occurrence with relevant habitat attributes, thus limiting inference about the relationships between putative habitat, restoration activities, and Delta Smelt population response. A further challenge to biological monitoring is the reluctance of authorities to permit directed surveys of listed fishes, e.g. Delta Smelt, due to the potential harm caused by physical handling. Indirect observation of macro-organisms via detection of DNA from environmental sampling has proved a compelling alternative monitoring approach, particularly for rare and/or protected species. Yet, the factors that influence detection of DNA in estuarine habitat are not well characterized, which has hindered refinement of eDNA sampling methods for detecting Delta Smelt occurrence. This study employed a fixed sampling array to explore how the ebb and flow of tidal phases affected Delta Smelt eDNA detection. Our primary objective was to estimate the effects of covariate metrics on ln[eDNA] observed in the tidal environment. Our secondary objectives were to quantify the difference between the effect of distance on ln[eDNA] in the tidal system with that of a unidirectional system and to estimate the effect of time elapsed since target species absence in tidal system on ln[eDNA] observed. Model predictors with a consistent effect on ln[eDNA] were: distance (from DNA source), eddy diffusivity, time (following removal of DNA source), side (tidal direction), and species (DNA source). The effect of distance from source on ln[eDNA] was consistently negative regardless of system, however, a higher proportion of non-detection was observed in the tidal system than would have been expected in a unidirectional system. DNA detections decreased over time, suggesting that positive DNA detections from estuarine water sampling are more likely to co-occur contemporaneously with the actual presence of individuals within habitat sampled. These findings improve the capabilities to design sampling strategies that will detect target species within an estimated probability, if individuals are present within a specified distance. Enhanced detections using eDNA sampling approaches would refine determinations of if and when Delta Smelt are present at a location, which in turn lessens known information gaps related to species occurrence and distribution.

2.2 Introduction

Dramatic alterations of the San Francisco Estuary have been ongoing since the Gold Rush (mid-19th century), with the current ecosystem conditions considered inhospitable to many native

estuarine watershed species (Cloern and Jassby 2012; MAST 2015). The Delta Smelt (*Hypomesus transpacificus*, McAllister) is a small osmerid fish (120-mm maximum length) endemic to the San Francisco Estuary that has experienced a significant decline in population abundance over the past several decades (Hobbs et al. 2017). The inadequacy of existing habitat to sustain Delta Smelt is exemplified by decreasing population abundance observed in long-term monitoring programs despite increasing effort, with an acute downward inflection occurring in the early 1980's (Stompe et al. 2020), resulting in the species being listed as 'threatened' in 1993 under the federal Endangered Species Act (ESA) (USFWS 1993). The most recent abundance index for Delta Smelt sub-adults has been zero 2018-2020 (White 2020).

A recent review of government agency Delta Smelt population-monitoring regimes utilizing various physical sampling gear (e.g., trawl, seine net) documented information gaps related to occurrence and distribution that created significant challenges to testing hypotheses about factors limiting population recovery (MAST 2015). For example, unknown distributions of smelt predators, Delta Smelt, and Delta Smelt prey negatively affects evaluations of both predation risk (top down) and food availability (bottom up) on Delta Smelt abundance and distribution. Additionally, life cycle modelling efforts lack inputs such as the habitat used by Delta Smelt for spawning and the habitat characteristics associated with reproductive success. The low encounter rates (high frequency of zero catches) result in imprecise survey estimates and inconclusive relationships between habitat attributes and population response (MAST 2015; Polansky et al. 2019). Additionally, permitted 'take' under the state and federal ESA generally curtail ancillary survey activities that directly target or that may encounter Delta Smelt. Therefore, sensitive and non-injurious monitoring approaches that do not require 'take' are needed to obtain biological information on Delta Smelt in their native habitat. One compelling approach to enhance detection of rare species applies molecular diagnostic methods that detect taxon-specific DNA in the environment.

Environmental DNA (eDNA) is simply DNA that is isolated from environmental samples, such as from water, rather than directly from an organism. Molecular techniques are used to amplify the target organism DNA once nucleic acids are extracted from the environmental samples. Fluorescence-based quantitative real-time Polymerase Chain Reaction (qPCR) – the benchmark for detection of nucleic acids – enables measurement of DNA with speed, specificity, and sensitivity (Wittwer et al. 1997; Bustin et al. 2009). As such, qPCR has been applied extensively in molecular diagnostics, medicine, and life sciences (Bustin 2000; Handelsman 2004; Kubista et al. 2006; Thomsen and Willerslev 2015; Cocolin and Ercolini 2015; Chiu and Miller 2019). Molecular species diagnostic tools have been particularly directed toward describing microbial diversity, as microbes are difficult to observe visually within samples taken from the environment (e.g., soil, water, air) (Morris et al. 2002; Venter et al. 2004; Martellini et al. 2005; Lee et al. 2020). The fields of biology and ecology have only recently embraced detection of macro-organism DNA from the environment to elucidate species presence (Ficetola et al. 2008; Thomsen et al. 2012; Rees et al. 2014; Roussel et al. 2015). Nevertheless, when one considers how to encounter and sample eDNA, it is instructive to consider it a particle (Turner et al. 2014) that degrades rapidly in aquatic systems (Balasingham et al. 2017), with general recommendations regarding approaches for sampling and subsequent analysis of eDNA available (Bustin et al. 2009; Goldberg et al. 2016).

The regulatory structure of the United States as defined by the Federal Endangered Species Act necessitates the serial monitoring and management of individuals species, such as Delta Smelt. This structure is entirely suited to eDNA and species-specific qPCR. A point relevant to detection of Delta Smelt eDNA though is the factors that influence detection of DNA in estuarine habitat are

also not well characterized (Port et al. 2016; Schmelzle and Kinziger 2016; Shelton et al. 2019; Crane et al. 2021). Chapter 1 investigated the use of 'censored latent variable models' to study the response of qPCR-based eDNA particle detections to specific environmental and sampling variables, including volume of water sampled, number of qPCR technical replicates, and distance from eDNA source. Although early forms of the model examined how the log-concentration of eDNA (ln[eDNA]) in filtered samples responds to biomass of, and distance from, target organisms in unidirectional aquatic systems (e.g., streams and rivers), they did not evaluate how multidirectional, tidally- influenced systems affect detections (Jane et al. 2015; Schumer et al. 2019). The salient factor is the ebb and flow of tide could alter the detection of target particles (e.g., expected effect of distance or time) relative to a system with one-way (unidirectional) flow, which in turn could change predicted detections (i.e., survey design) or interpretation of a detection (i.e., context). This lack of quantitative data on the sensitivity of eDNA surveys to a quintessential characteristic of tidal environments limits further refinement of eDNA as a tool for detecting Delta Smelt occurrence and thus developing successful recovery actions for the species.

We implemented this study to explore how the ebb and flow of tidal phases affects our ability to detect Delta Smelt eDNA using qPCR and whether the sensitivity of qPCR permits development of eDNA surveys with a high probability of detecting Delta Smelt. The goal was to increase the sophistication of eDNA analysis interpretations to lessen the known information gaps for Delta Smelt population monitoring. Our first objective was to estimate the effects of covariate metrics on ln[eDNA] observed in the tidal environment, which would improve DNA detection predictions. Based on our previous observations in unidirectional systems (i.e., Chapter 1), we hypothesized that closer proximity to, and higher biomass of, the eDNA source positively affects sensitivity of qPCR by increasing the concentration of eDNA in the environment and subsequently within environmental samples. We further hypothesized that tidal phase would significantly affect the probability of detection in the San Francisco Estuary (estuarine habitat occupied by Delta Smelt), but in uncertain ways because the effects of tides on eDNA concentration are relatively unclear at this time in comparison to the effects observed in unidirectional systems. If our hypotheses are true, then we expected to observe consistently higher concentrations of eDNA in samples collected closer to eDNA sources and that tidal ebb and flow would affect detections in a significant and consistent way. Our subsequent objectives were to quantify the difference between the effect of distance on ln[eDNA] in the tidal system collections with samples taken in unidirectional systems and to estimate the effect of time elapsed since target species absence in tidal system on ln[eDNA] observed. We discuss how the quantification of effects on ln[eDNA] in filtered samples can be used to design powerful eDNA sampling strategies with consistent probabilities of detection to inform recovery planning with data related to occurrence and distribution for Delta Smelt.

2.3 Materials and Methods

2.3.1 Environmental DNA Sampling Array

The field study was designed to estimate the effects of tide and sampling effort on the sensitivity of qPCR to detect the eDNA of Delta Smelt in the San Francisco Estuary. Further, the experimental data were collected explicitly for analysis using the censored data models developed in Chapter 1. We implemented the study via two experiments (on October 22, 2019, and December 12, 2020, respectively) which coincided with suitable water temperatures to deploy live fish and were not associated with high freshwater inflows caused by storms. Additionally, all eDNA collection events

likely occurred prior to the Delta Smelt spawning season, which is estimated to occur from February to June (Moyle et al. 2016), to reduce the potential for detections of natural-origin Delta Smelt. We constructed a fixed distance eDNA collection array in Shag Slough (Figure 2.1), which allowed us to sample across a representative range of times and locations that might characterize tidal phases of Delta Smelt habitat. Shag Slough is located in Solano County, California in the northeastern portion of the San Francisco Estuary watershed where freshwater discharge enters into the tidally-influenced Delta (ESA 2014). In the first experiment (October 22, 2019), a known count and biomass of live (cultured) Delta Smelt (*Hypomesus transpacificus*) and a control species (i.e., Northern Anchovy, *Engraulis mordax)* were secured in floating cages at the array center (Table 2.1). In the second experiment (December 12, 2020), these same two species were used but in addition a third species was included in the floating cage: steelhead trout (*Oncorhynchus mykiss*) from Coleman National Fish Hatchery (Anderson, CA). Multiple control species were used as contingency in case the sampling or laboratory processes were compromised for a species. For both experiments, sampling stations were established in both along-channel directions at discrete distances to increase likelihood of collecting eDNA across the bi-directional ebb and flow of tidal phases. The floating cages and sampling stations were moored 15 meters from the water's edge using a laser rangefinder, buoys and either 18 kg (cages) or 5.5kg (stations) anchors. The distance between the array center (fish cages) and each sampling station was determined using a Garmin GPSMAP® 64. For crew safety, water filtration occurred onshore at each station by pumping water through three separate 20-30 m lengths of sterile Masterflex L/S® 15 peroxide-cured silicone tubing (Cole Parmer ®) that extended from cordless drill driven Masterflex L/S® Easy-Load® II peristaltic pump heads to the buoy-based water collection sites positioned 50-60 cm below water surface. All non-sterile array equipment was decontaminated with 20% bleach solution and rinsed with fresh water prior to deployment. Prior to live fish deployment, control samples were filtered from the array to confirm Delta Smelt and control species were not present.

2.3.2 Environmental DNA Collection

To reduce the possibility for among-station variation in eDNA detections associated sampling at different times, all stations within the array were sampled simultaneously at proscribed times for the duration of each experiment. Collections were synchronizing using two-way radios and verbal prompts. Control stations, located outside the array, were sampled immediately following each simultaneously sampled eDNA collection. Three filter replicates were collected per station, with water filtration procedures following (Bergman et al. 2016), but modified to use Millipore Sterivex™ 0.45 μm sterile filter units (EMD Millipore). Total water volume collected per filter was fixed at 100ml.

To limit the potential for contamination between sampling events, sterile gloves were worn, and all sampling consumables were pre-packaged in the laboratory, itemized by station and time point. Gloves were immediately disposed of after each use into a sealed trash bag. After water filtration, Sterivex™ filters were end-capped, labelled, and immediately placed into a sterile sealed secondary container stored on ice. All filters were transferred to a –20 C freezer for storage prior to DNA extraction. Sterile tubing was deployed at the start of each experiment, but installation was not altered during the sampling period. A coordinated equipment cleaning protocol for rinsing peroxide-cured silicone tubing was implemented prior to each sample event. Staff were first prompted to pump water through tubing without attaching a filter for approximately two minutes at maximum speed to replace and rinse water volume in each tubing line multiple times. Following

rinsing, filters were attached and eDNA collection occurred. After eDNA collections, pumps were reversed to evacuate water from all lines. Additionally, eDNA field controls (ultrapure water) were collected and analyzed to confirm equipment was clean of detectable target DNA.

Experimental parameters differed between events, such as array distance intervals, sampling times relative to tide, quantity of fish deployed in cages, and control species used. Sampling took place during two events: October of 2019 and December of 2020 (Table 2.1). On October 22-23, 2019, eDNA was collected every three hours over a full 25-hour tidal cycle. Floating cages containing Delta Smelt and Anchovy were deployed the first 13 hours of sampling, then removed, with sampling continuing an additional 12 hours. The three filter replicates per station were collected in parallel. The December 11-12, 2020, sampling event was a refinement of the 2019 effort. Sampling interval was reduced from 3 hours to every 45 minutes over the course of a single low to high tidal cycle, with live steelhead trout added as a control species. Filter replicates per station were collected in-series, with each filter taking 1-2 minutes to process. A single additional collection event was conducted on December 12, 2020, 24 hours after removal of the live pens present 11-December. After each experiment, all living individuals were euthanized according to professional society guidelines (Jenkins et al. 2014) and all specimens were maintained in cold storage: 1) Delta Smelt with the Reclamation's Tracy Fish Collection Facility and steelhead with Cramer Fish Sciences (West Sacramento, CA).

2.3.3 Environmental DNA Laboratory Analysis

DNA from all samples and controls were extracted using DNeasy PowerWater Sterivex™ Kit (Qiagen) following the manufacturer's recommendations. Sample and control technical replicates (Table 2.2) were analyzed for the presence of Delta Smelt and Northern Anchovy mitochondrial DNA using assays published in Baerwald et al. (2011) and Sassoubre et al. (2016), respectively. Samples collected on December 11-12, 2020 were also analyzed for presence of Rainbow Trout mitochondrial DNA (Brandl et al. 2015). The qPCR amplification mixture final volume of 10ul included 4ul template, 5ul 2x Applied Biosystems TaqMan Environmental PCR Master Mix and varying final primer and probe concentrations (Table 2.2). Thermocycling was performed using the Applied Biosystems QuantStudio™ 3 following: 10 min at 95° C, then 40 cycles of 15 sec at 95° C and 1 min at assay specific annealing temperature (Table 2.2). Each qPCR plate included three no template controls (ultra-pure water) and three positive controls (2 ng/µl Delta Smelt, Northern Anchovy or Rainbow Trout/Steelhead Trout genomic DNA). All PCR master mixes were made inside a UV sterilized PCR enclosed workstation. All PCR reactions were conducted on instruments located outside of the main lab in a separate portion of the building. Results of the qPCR reactions were analyzed using QuantStudio™ Design & Analysis Software with the magnitude of the qPCR signal reported as quantification cycle (Cq).

2.3.4 Hydrodynamic Modeling

A two-dimensional hydrodynamic model was developed and applied to characterize the movement of water that transported and mixed the eDNA material emitted by the fish samples in the floating cages. The modeling software, Delft3D-FLOW (Deltares 2021), was configured to represent the 4.5-km straight segment of Shag Slough that contained the floating cages (Figure 2.1). The model's horizontal grid cell size was four meters square and in the vertical direction a single layer was used for depth-averaged conditions. Bathymetry within the model domain was excerpted from a larger digital elevation data set (Fregoso et al. 2017). To generate currents within the domain, tidally-varying water levels were specified at the model's southern boundary. These water level boundary

conditions consisted of the observed water levels at the monitoring station at the southern end of Liberty Island, approximately 4.7 km south of the model's south boundary (California Data Exchange Center Station ID=LIB; 38.24210, -121.68490). In addition, observed wind speed and direction from the Hastings Tract East CIMIS station was applied as a surface stress across the model domain.

To predict the transport of eDNA by the tidal currents, a virtual tracer was introduced into the model domain in the grid cell containing the floating cages. The tracer was added to the cell at a constant rate for the same 13-hour period that the cages were deployed during the October 2019 sampling event. This tracer was advected by the currents predicted by the hydrodynamic model, yielding a continuous prediction of the concentration of the tracer throughout Shag Slough, relative to its initial concentration in the cell with the floating cages. The model only advects the tracer with the predicted average flow velocity at the scale of the grid cells. Smaller scale turbulence can also locally mix water and the constituents carried by the water (e.g., the virtual tracer and eDNA). To account for this sub-grid process, additional dispersal of the virtual tracer is predicted as a function of the horizontal eddy diffusivity (K). Without detailed velocity and/or dye-tracking measurements, K can only be estimated to within a likely range. To simulate range of likely tracer dispersal, the model was run three times, with identical conditions except for three levels of K: high diffusivity (K = 1.0 m^2/s), moderate diffusivity (K = 0.5 m^2/s), and low diffusivity (K = 0.1 m^2/s).

Flow-following Drogues

During the December 11, 2020 sampling event, a set of drogues were deployed to track the movement of water as a function of tidal currents and wind. The drogues were constructed from two thin plastic sheets, 15 cm by 45 cm, that were intersected along their shorter dimension to form a cross shape, thereby creating an underwater 'sail'. Tethered above the plastic sail was a waterproof box which contained a GlobalStat DG-200 GPS receiver and logger, with a nominal horizontal accuracy of 1-5 meters, depending on satellite connectivity. Weight was added just below the sail to align the sail vertically in the water column and to submerge nearly all the waterproof box, thereby minimizing the potential for wind acting on the box to transport the drogue. The loggers were configured to record position every 30 seconds.

Every half hour, from 9:30 am until 3:00 pm, one of the nine drogues was released just south of the floating cages. The drogues then freely drifted with the currents at the top of the water column, to track the likely trajectory of water which passed through the floating cages. If a drogue drifted close enough to shore to run aground, it was re-deployed a few meters from shore with a long pole.

2.3.5 Environmental DNA Statistical Analysis

The objectives of the statistical analysis were threefold: 1) Estimate the effects of covariate metrics on ln[eDNA] in the tidal experiments; 2) Quantify the difference between the effect of distance on ln[eDNA] in the tidal samples with samples taken in unidirectional systems; and 3) Estimate the effect of time elapsed since live pen removal on ln[eDNA] in the tidal experiments. We used the software package `artemis` (version 1.1.0, Espe et al. in review) in the R programming environment (version 4.0.5, R Core Team 2021) to fit all models in the analysis. Briefly, `artemis` implements a censored latent variable model (essentially a generalized linear model) to estimate the log-concentration of eDNA in environmental samples. The model assumes i) that observations are exchangeable (i.e., observations vary by measurement error but are otherwise uniform, given the model formula provided); ii) ln[eDNA] is sampled with normally-distributed errors, and iii) there are

no false detections, i.e. the measurement error cannot result in a positive detection when eDNA is not present within the sample.

To address analysis objectives 1 and 3, separate candidate models were fit to the data from each sampling event (October 2019 and December 2020). Nine candidate models were fit to the data from October 2019, incorporating different combinations of predictors, including the virtual tracer concentration values predicted at each distance interval by the three modeled eddy diffusivity (*K*) values. The other predictors considered in the October 2019 event candidate models were absolute distance in meters from the fish cage, target species (Delta Smelt or Anchovy), time elapsed before and/or after live pen removal, and array side (north or south of the cage). All continuous numeric variables were converted to z-scores prior to modeling to increase Markov Chain Monte-Carlo (MCMC) sampling efficiency. A unique ID for each filter was included in all candidate models as a random effect, to allow for the association of qPCR technical replicates conducted on each filter DNA extraction. A single model fit to the December 2020 data was similar in structure but differed by the tidal metrics included and did not include predictors for eDNA tidal dispersion. Tidal predictors for this experiment were water velocity in feet per second (ft/s) and river stage in feet (ft), both derived from publicly available data on the California Data Exchange Center for environmental station gauge (Station ID=SGG; 38.31833, -121.69306) located approximately 2,000 meters north of the deployed fish cage. The remaining model predictors were absolute distance from the cage (in meters, scaled to the October 2019 distances for ease of comparison and interpretation) and target species (Delta Smelt, Northern Anchovy, and steelhead trout). As with the October 2019 collection associated model, unique filter ID was included with random effects. Model comparison and selection was conducted within sampling events via Pareto-Smoothed Leave-one-out Cross Validation (LOO) model selection (Vehtari et al. 2017) using the 'loo' package (version 2.4.1, Vehtari et al. 2020). All modeled data and code are available at https://github.com/fishsciences/2020-Tidal-eDNA-Analysis.

For statistical analysis objective 2, data from previously conducted live car studies allowed us to compare results from this tidally-influenced dataset to one collected from a unidirectional system (see Chapter 1). The unidirectional dataset was collected in the primary intake canal at Reclamation's Tracy Fish Collection Facility (near Byron, CA), and followed the same live pen deployment and eDNA sampling, extraction, and analysis procedures as the tidal sampling events. The unidirectional dataset was only used in the evaluation of statistical analysis objective 2, where all differences in covariate levels between experiments (number of Delta Smelt and volume filtered) could be accounted for with the incorporation of random effects. Statistical analysis objective 2 was then addressed with a single model combining data from the unidirectional experiment and both tidal experiments. By parameterizing the interaction between distance and system (tidal vs. unidirectional), the model was able to estimate the variance in the effect of distance between systems. The model structure is presented in Table 2.3.

2.4 Results

Data from 8,904 qPCR reactions were generated across the two tidal experiments analyzed, with technical replicates analyzed per filter, event, and species shown in Table 2.2. Overall, the proportion of positive detections (across target species) was 0.08. More specifically, 669 of 8904

total technical replicates had a Cq value lower than the assay DNA detection threshold of 40.0. The proportion of positive detections by tidal experiment, distance, and sampling times are plotted in Figures 2.2 and 2.3, overlaid with scaled metrics of tidal phase for the respective experiment.

2.4.1 Objective (1) Covariate effects on ln[eDNA]

An index of final models of (ln[eDNA]) and predictors is presented in Table 2.3. In the October tidal sampling event, all levels of eddy diffusivity (*K*) were positively associated with ln[eDNA]. Pareto-Smoothed Leave-one-out Cross Validation (LOO) model selection (Vehtari et al. 2017) of the models fit to the October 2019 data suggested all models had similar predictive performance. This implies that all levels of eddy diffusivity produced equivalent out-of-sample prediction error given the other covariates. For the October tidal sampling event, the north side of the array had a consistently positive effect on ln[eDNA]. Target species (live Delta Smelt) was strongly positively associated with ln[eDNA] (Table 2.3; Figure 2.4). In the models fit to the December 2020 sampling event, both distance and side (north) were found to have a consistent effect on ln[eDNA] (Table 2.4; Figure 2.4). The target species Rainbow Trout had a strongly positive effect on ln[eDNA] relative to Delta Smelt and Northern Anchovy. The 95% credible interval for the coefficient estimates of both tidal water velocity (ft/s) and river stage (ft) overlapped with zero (Figure 2.4), indicating no consistent effects on ln[eDNA] for this event.

2.4.2 Objective (2) Effect of distance in tidal vs. unidirectional experiments

Distance had a consistently negative effect on ln[eDNA] in both unidirectional and tidal systems (-0.763 – -0.081 95% CI). The effect of distance did not vary across systems consistently (-3.067 – 1.686). In other words, the effect of distance on ln[eDNA] was consistently negative (ln[eDNA] decreased with increasing distance from the live pen), regardless of system. While target eDNA were observed throughout the array's extent during experiments, a higher proportion of positive technical replicates were observed at collection stations closer to the fixed position of the live car and fewer positives at > 400 meters from the live pen (Figure 2.2-Figure 2.3).

2.4.3 Objective (3) Effect of time elapsed since live pen removal on ln[eDNA] in the tidal experiments

No detections were recorded in the samples taken 24 hours after removal of the floating cage used in the December 2020 event. In the October 2019 event, the proportion of positive technical replicates decreased with time following cage removal; the model estimated that time elapsed since live pen removal had a consistently negative effect on ln[eDNA] (-0.138 - -0.048 95% CI; Table 2.4; Figure 2.4).

2.5 Discussion

We conducted field trials to collect eDNA within tidally influenced Shag Slough, an area that lies within the historic range of Delta Smelt (Merz et al. 2011; MAST 2015). The Shag Slough field trial demonstrated that Delta Smelt, steelhead trout, and Northern Anchovy DNA could be reliably detected within the sampling array at concentrations (ln[eDNA]) and distances in alignment with sampling design expectations. Those expectations were that higher proportions of positive

detections would be observed closer to the source of eDNA modulated by covariate effect priors (e.g., 10^{-6} ng/μl DNA detection threshold (from standard curve), -2.476 ln[eDNA] per meter from DNA source), and that patterns of detection would be correlated with direction of tidal flow. However, overall, a higher proportion of *non-detection* in technical replicates were observed in the tidal system than would have been expected in a unidirectional system, particularly at distances closer to the floating cages.

Given the effect of distance on ln[eDNA] estimated from unidirectional system data (-2.476 ln[eDNA] per meter from DNA source; Table 2.4), we hypothesized that DNA detections would decrease with distance within the fixed distance array of tidally influenced Shag Slough. Yet, the proportion of positive detections at the closest absolute distances in the October 2019 sampling event (1m and 50m) was lower than expected, and the estimated effect of absolute distance from DNA source on ln[eDNA] collected was small: its 95% CI overlapped with zero (Table 2.4; Figure 2.4). We speculate that our monitored covariates failed to capture one or more effects operating on ln[eDNA] available for extraction from the filters. For example, during the October 2019 experiment, the plume of eDNA emitted from the floating cages may have been relatively narrow and patchy when close to the source. Variable mixing could have deflected this narrow plume away from the closest sampling stations, lowering the proportion of positive detections. At intermediate distances, as the plume spread out, it may have been more likely to be sampled.

The eDNA particle collection strategy performed here was more intensive than would typically occur, which allowed detection complexity to be uncovered that in practice would be unobservable during a more routine eDNA monitoring survey. A management application of this observation is that when designing a sampling protocol in a tidal environment a researcher may 1) increase the number of samples near the sampling point of interest to compensate for this phenomenon or 2) measure mixing rate. Detection strength of eDNA was statistically associated with the north tidal direction in both experimental events, particularly in the December 2020 event, suggesting tidal seasonality could also influence eDNA distribution. As the bulk of water movement in the December 2020 event was in the northward direction when the DNA source was present and eDNA collection interval was every 45 minutes, this suggests that the movement of eDNA particles newly introduced to a system are broadly reflective of water movement. More specifically, the "Side" covariate was estimated to have a positive effect size on concentration of eDNA (Figure 2.4). Additionally, an association was observed between proportion of positive technical replicates and distance/direction from DNA source during the October 2019 event (Figure 2.2). Intuitively, the distribution of particles (eDNA) present at time of sampling from an estuarine system would be influenced by the particle concentration, the underlying hydrological conditions, and the spatial relationship between DNA source and collection location (Port et al. 2016; O'Donnell et al. 2017; Shelton et al. 2019). In practice, the concentration of eDNA present will likely be low at time of sampling (i.e., in vicinity of level of detection) if the target macro-organism is rare (Jerde et al. 2016; Crane et al. 2021). As mentioned, an unexpected increase in non-detections were observed at certain times and stations within the fixed array. Mixing by wind, as indicated by the December 2020 drogue data, suggests that the eDNA plume may not be homogenously mixed by tidal currents alone on fine scales (10s of meters) and may have been deflected laterally away from the sampling array. The complex fine scale flow patterns observed here to influence positive detections of eDNA particles could complicate integration of eDNA detections into tools used by Government Agencies to manage hydrology more regionally (e.g., Department of Water Resources Delta Simulation Model II or CalSim 3). We suggest that management applications would benefit from emphasizing the design of sampling regimes where the concentration of eDNA is expected to be at detectable levels

given desired biological criteria (e.g., a specific biomass/distance range), rather than focus on the ln[eDNA] observed in a given sample. Further, while eDNA sampling designs should be robust to stochastic localized flow characteristics, randomization and aggregation of eDNA collection locations likely would be needed to align eDNA observations with other monitoring methods (O'Donnell et al. 2017; Shelton et al. 2019).

The relationship between the effects of distance from eDNA source and particle concentration (as predicted by eddy diffusivity, *K*) is complex. DNA particle density does decline with distance (Table 2.4); however, the strength of this decline varies by situation. For the observations from the October 2019 event, we note that *K* was predictive of ln[eDNA] when included in the same model as distance, but also that distance became predictive when the *K* parameter was omitted (supplementary materials). This suggests that the predictor *K* may be functioning in the model in much the same way as distance. Model selection only slightly favored the 'with-*K*' relative to ones that only included distance. Accounting for hydrological covariates in the tidal environment from which eDNA collections will occur could lead to better predictions regarding the distribution of DNA detections, although fine scale (<100 meters) differences in fish community structure have been observed even within high energy nearshore marine habitat (Port et al. 2016).

Not all hydrological metrics were predictive in these data. The tidal water velocity and river stage predictors in the December 2020 sampling event were not predictive of ln[eDNA], but distance was (Figure 2.4). Combined with observations above suggesting tidal direction influences the pattern of positive detections, and that there is a distance effect in both tidal and unidirectional systems, it is likely preferable to make eDNA collections during a consistent tidal phase for the intended application. For example, collecting during a slack tide may localize the particle distribution for fish close to the sampling station; and collecting during high tidal velocity would increase the probability of detecting eDNA emitted from a greater distance away (in relevant direction) from the collection location.

DNA detections decreased over time. It was unclear whether this was due to eDNA particle degradation or to particles being advected away from sampling array stations (e.g., transverse to shore) and decreasing in concentration below qPCR detection threshold (Jerde et al. 2016). Flow-following drogues, equipped with GPS loggers, during the December 2020 event suggested that when westerly cross-slough wind speeds exceeded 3 m/s, transverse flow could have been a factor (supplemental). For the first three hours after deployment of the fish cages, all the drogues remained close to the western bank of the slough (supplemental). Once the wind speed exceeded 3 m/s (supplemental), the drogues drifted eastward with the wind, and most of the drogues were located along the eastern bank for the remainder of the event. Although drogue movement is not a perfect surrogate for eDNA trajectory, the drogues' eastward translocation suggested that eDNA may have been moved away from the sampling array along the western bank.

Following the removal of the fish cage during the October 2019 experimental event, the model indicated with certainty that ln[eDNA] decreased over time, despite the "pulse" of eDNA that could have occurred in association with cage removal. In the December 2020 tidal experiment there were no positive detections the day after live pen removal. These observations in the two experiments suggested that for rare species such as Delta Smelt, positive DNA detections from estuarine water sampling would likely occur contemporaneously with the actual occurrence of individuals within habitat sampled. This observation is consistent with the effective sampling area being reported as

10s of meters to a few 100 meters in both marine (Port et al. 2016; O'Donnell et al. 2017; Shelton et al. 2019) and freshwater environment (Jane et al. 2015; Wilcox et al. 2016; Jerde et al. 2016).

As more experience is gained regarding factors that affect DNA detection in estuarine systems, surveys will be constructed with enhanced detection efficiency. For example, implementing different water collection strategies may add capabilities to eDNA collection. Fixed point sampling at large water diversions (e.g., pumping stations) is practical. Yet, given the scale of estuarine habitat within the range of Delta smelt (1,100 square miles; Suisun Bay to Sacramento on the Sacramento River and to Mossdale on the San Joaquin River) using fixed position sampling as the only sampling approach would likely be impractical for reliable detection of extremely rare eDNA (as is the case with Delta smelt). Rather for detection of rare DNA within a large geographic area, alternatives to fixed point sampling would likely be required, such as spatially integrated collections using transects to sample water. Transect sampling would then inform when and where to utilize point sampling at specific locations (e.g., restored habitat, water diversions). However, DNA detections of rare species occurred in this study in a tidal environment using point sampling methods. The paradigm that DNA is shed into the environment by macro-organisms and this DNA can be aggregated within environmental samples and detected using nucleic acid amplification seems without question. The challenges lie with improving specific means of encountering sufficient target DNA to achieve detection thresholds for rare organisms and the interpretations from observations generated.

2.6 Conclusion

Statistical model predictors with a consistent effect on ln[eDNA] were distance (from DNA source), eddy diffusivity, time (following removal of DNA source), side (tidal direction), and species (DNA source). The observed effect of distance on ln[eDNA] was reduced in tidal system relative to unidirectional system. DNA detections decreased over time, suggesting that positive DNA detections from estuarine water sampling would likely occur contemporaneously with the actual occurrence of individuals within habitat sampled. These findings, coupled with workflow customized for eDNA approaches (e.g. see artemis description in Chapter 1), enable sampling regimes to be constructed with more predictable detection capabilities targeting a specific biological question. Further, the general approach described here, using a fixed known DNA source to refine effect sizes for relevant sampling covariates within the desired study habitat offers a means to quickly characterize a location prior to designing a field effort for the target organism. Yet, efficiencies are needed for both deployment of control species and increased scale of eDNA collection. Specific to Delta Smelt, near-term needs would be to efficiently characterize any remaining habitat (by hydrological type) where sampling covariates are unconfirmed. To justify this activity, an agreement by regulatory authorities on characterization is needed. Longer-term objectives pertain to adoption of eDNA as a monitoring method (i.e., eDNA considered not experimental) and mechanisms for adopting new tools (eDNA) within the regulatory environment.

2.6 Figures

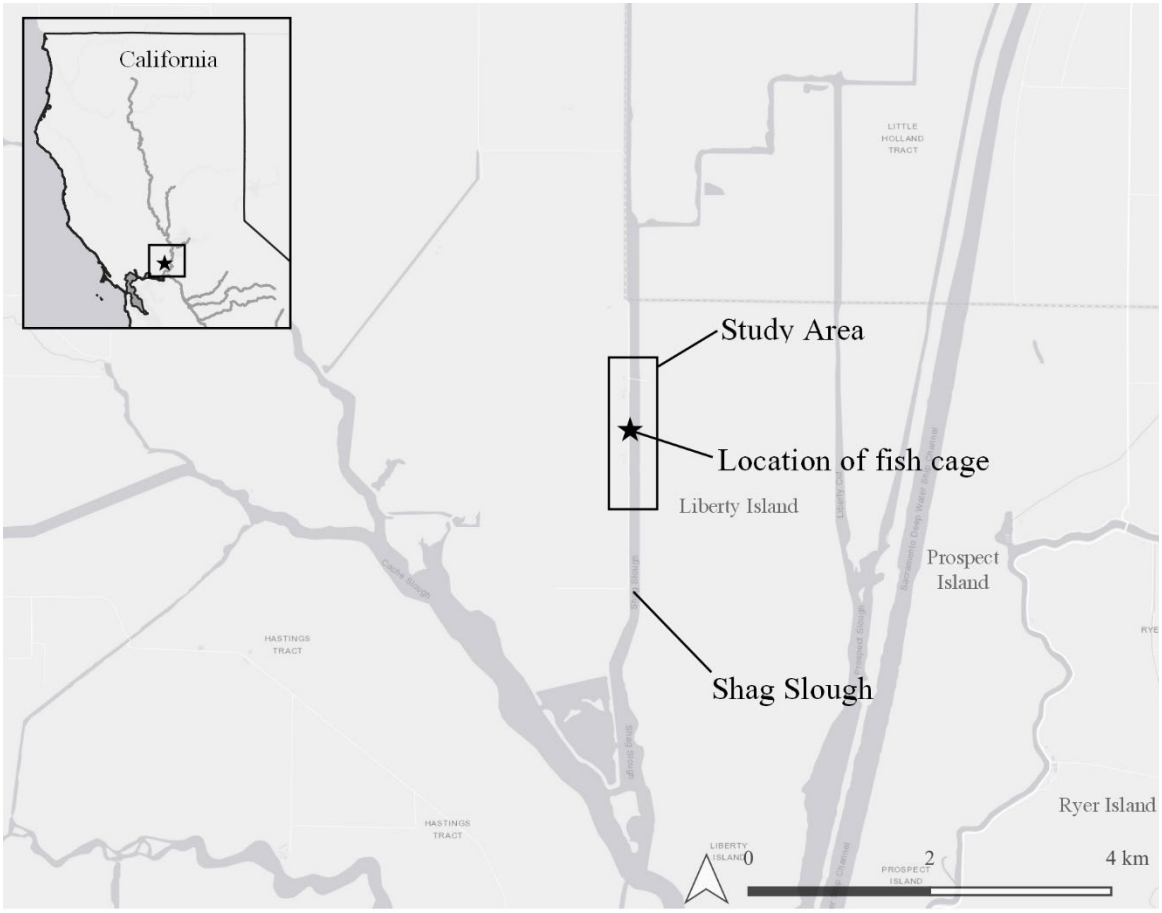


Figure 2.1. Location within Central Valley California, USA where eDNA collections occurred.

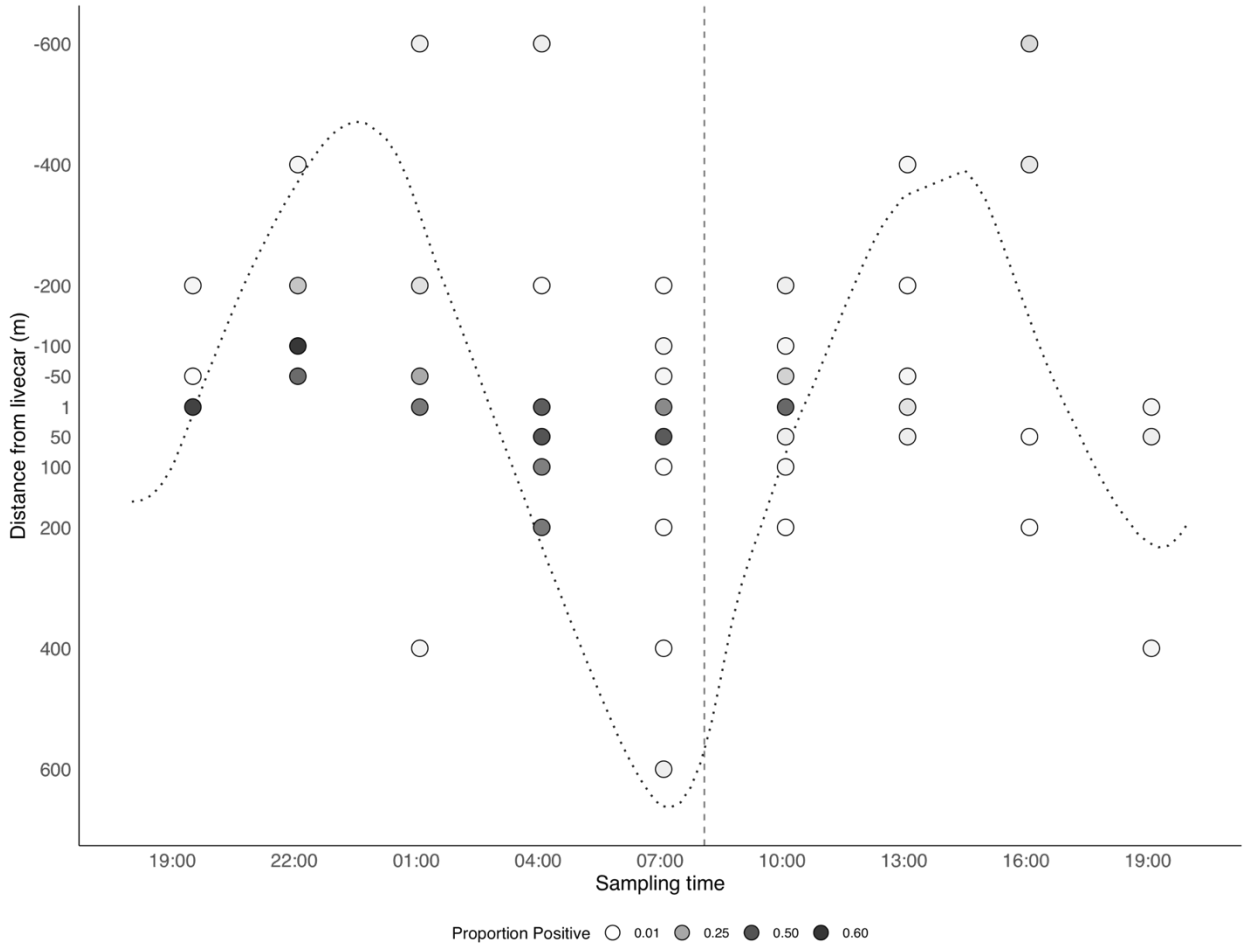


Figure 2.2. Proportion of positive detections (Anchovy and Delta Smelt combined) by sampling time (October 2019 event). The closest recorded river stage (ft) data has been matched to sampling time and overlaid on the plot by scaling its range (1.66 – 5.72 ft) to the primary y-axis. The dashed vertical line indicates live pen removal time. The distances closest to the live pen (-1 and 1) have been grouped at distance = "1", as they were for modeling.

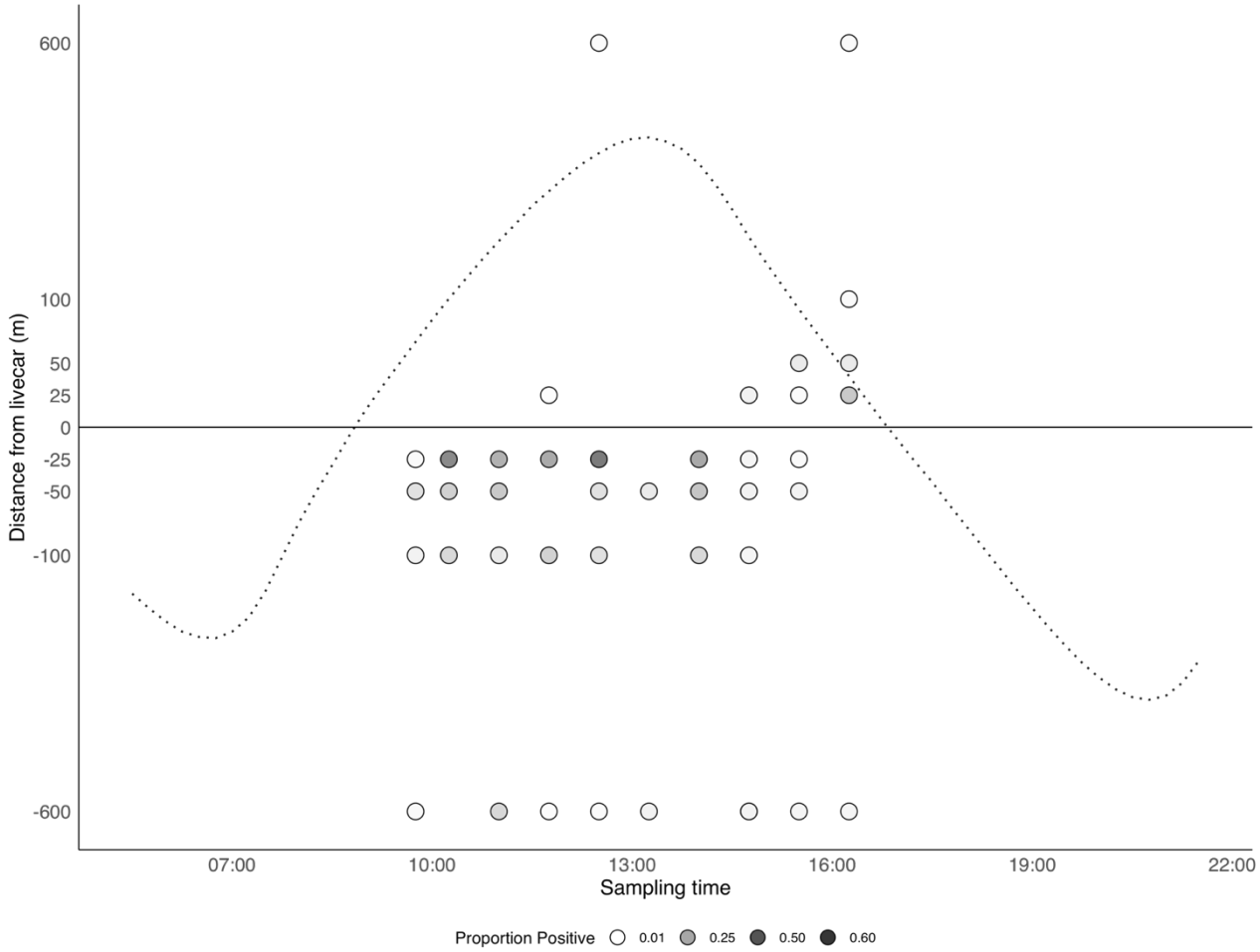


Figure 2.3. Proportion of positive detections (Anchovy, Steelhead, and Delta Smelt combined) by sampling time (December 2020 event). The closest recorded river stage (ft) data has been matched to sampling time and overlaid by scaling its range (1.66 – 5.72 ft) to the primary y-axis. For illustration of the delay in detection with river stage, more of the tidal cycle is shown than was captured by the experimental sampling intervals.

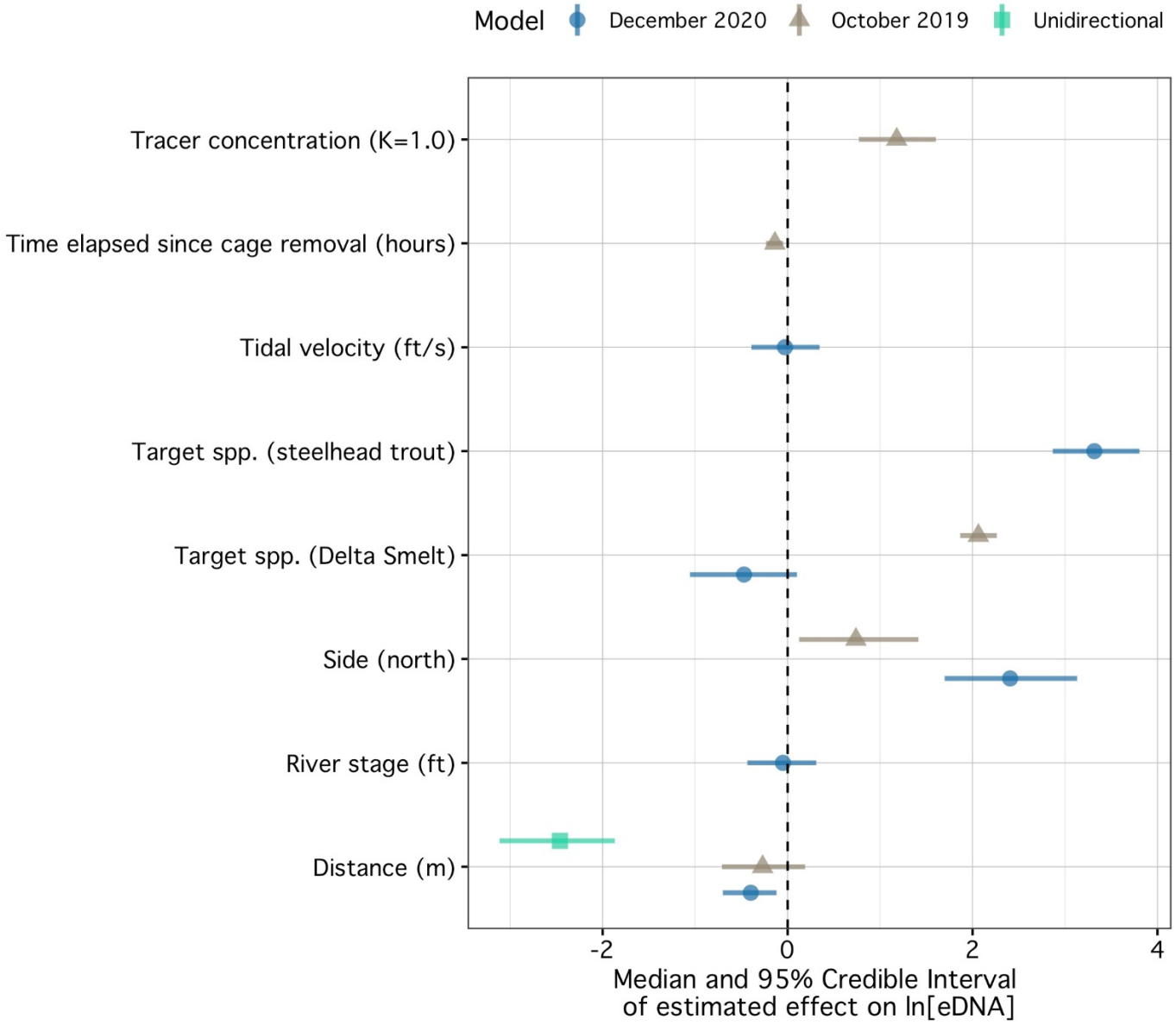


Figure 2.4. Estimated covariate effects by model. Dots are the median of the posterior probability; segments span 95% Credible Interval of posterior probability.

2.7 Tables

Table 2.1. Sampling event by live pen species deployed, time interval between water filter events, total hours of experiment, north (-) / south (+) distance intervals from live pen, +/- distance to control sample sites and filter technique (simultaneous filters collect 3 filters from same stream source; in-series filter collect each filter from the same stream source consecutively).

Event	Delta Smelt (count)	Anchovy (count)	Steelhead trout (count)	Sampling interval (min)	Total Hrs sampled	Array intervals (+/- (m))	Control interval (+/- (m))
2019-10-22	54	50	NA	180	25	1; 50; 100; 200; 400	600*
2020-12-11	25	25	15	45	6.5	25; 50; 100	600

Table 2.2. The qPCR amplification summary. Primary reference for qPCR primer and probe sequences given the target species, final primer and probe concentrations qPCR mixture volume of 10ul, thermocycling annealing temperatures used, and number of qPCR technical replicates performed per sample.

Target Species	**Primer and Probe**	**Primer and Probe**	**Annealing Temperature**	**Technical Replicates**	**Technical Replicates**
	Reference	Final Concentration		Oct 2019	Dec 2020
Delta Smelt (*Hypomesus transpacificus*)	Baerwald et al. 2011	900nM/60nM	63°C	10	5
Northern Anchovy, *Engraulis mordax*	Sassoubre et al. 2016	200nM/150nM	60°C	5	5
Steelhead Trout, *Oncorhynchus mykiss*	Brandl et al. 2014	900 nM/200nM	60°C	N/A	5

Table 2.3. Summary of mixed effects model formulas and their data sources. K refers to the particle concentration predicted by the level of eddy diffusivity, a value derived from ESA (2014) for the October 2019 tidal experiments.

Model Index	**Model Structure**	**Data Source**	**System type (sampling location)**
October 2019	Cq ~ K1.0 + Distance (m) + Species + Side + Time elapsed since cage removal (hrs) + FilterID	October 22-23, 2019 sampling event	Tidal (Shag Slough)
December 2020	Cq ~ Distance (m) + Species + Velocity (ft/s) + Side + River stage (ft) + FilterID	December 11, 2020	Tidal (Shag Slough)
Unidirectional	Cq ~ Distance (m) + FilterID	Delta smelt unidirectional live pen experiments, 2017-2018	Unidirectional (Central Valley Project)

Table 2.4. Estimated effect sizes for model estimates. Estimates are the effect on ln[eDNA]. Predictors with a consistent effect on ln[eDNA] are bolded. Intercepts represent the ln[eDNA] when all other predictors are held constant at 0 or their mean, depending on the predictor's scale. The model parameter "ln[eDNA] sigma" is a variance estimate corresponding to the standard deviation of the effects on Cq.

Model Index	Model Parameter	Mean	2.50%	97.50%
Unidirectional	Intercept	-10.048	-10.761	-9.408
Unidirectional	**Distance (m)**	**-2.476**	**-3.114**	**-1.868**
Unidirectional	ln(eDNA) sigma	1.188	1.008	1.409
October 2019	Intercept	-18.48	-19.224	-17.837
October 2019	**Particle concentration (K = 1.0)**	**1.18**	**0.77**	**1.603**
October 2019	Distance (m)	-0.266	-0.71	0.19
October 2019	**Time elapsed since removal (hrs)**	**-0.138**	**-0.233**	**-0.048**
October 2019	**Target Species (Delta Smelt)**	**2.065**	**1.866**	**2.263**
October 2019	**Side (north)**	**0.745**	**0.124**	**1.415**
October 2019	ln(eDNA) sigma	1.366	1.272	1.469
December 2020	Intercept	-21.194	-22.301	-20.157
December 2020	**Distance (m)**	**-0.4**	**-0.7**	**-0.12**
December 2020	Target spp (Delta Smelt)	-0.469	-1.056	0.101
December 2020	**Target spp (Steelhead)**	**3.322**	**2.867**	**3.805**
December 2020	Tidal water velocity (ft/s)	-0.029	-0.39	0.346
December 2020	River stage (ft)	-0.053	-0.434	0.311
December 2020	**Side (north)**	**2.413**	**1.699**	**3.131**
December 2020	ln(eDNA) sigma	2.298	2.024	2.62

3. Biomass Effects on Environmental DNA: Delta Smelt eDNA from Tidal Experiments

3.1 Introduction

The biomass of target fish species in the water is positively related to eDNA concentration in lentic (Doi et al. 2015) and lotic (Jane et al. 2015) habitats. In this component of the study, we studied the influence of fish biomass on eDNA in a tidal environment. We conducted two experiments. Each experiment began with a set of samples taken before Delta Smelt were introduced into the water at the experimental site in the San Francisco Bay-Delta (Delta). Each experiment then was executed with a different biomass of Delta Smelt (Table 3.1) inserted into a live pen and deployed in the waters of Shag Slough (2.1).

3.2 Methods

We initiated the two experiments on October 22, 2019 and December 12, 2020, respectively. We constructed a fixed distance eDNA collection array in Shag Slough, located in Solano County, California in the northeastern portion of the San Francisco Estuary watershed where freshwater discharge enters into the tidally-influenced Delta. In each experiment, a known count and biomass of live (cultured) Delta Smelt (*Hypomesus transpacificus*) were deployed in floating cages, i.e. live pens (Table 3.1). In addition, a control species (dead Northern Anchovy, *Engraulis mordax*) was secured in floating cages at the array center, with sampling stations established in both along-channel directions at discrete distances. Prior to live fish deployment, control water samples were filtered from sampling points ranging from 600 meters north (-600 in Figure 3.2) to 600 meters south (+600 in Figure 3.2) of the live pen location to confirm Delta Smelt and control species DNA were not present. To evaluate the success of eDNA detections across the various combinations of distance and biomass, we used Delta Smelt qPCR assays described in previous sections of this report and report the summary statistics of the Quantification Cycles (Cq) resulting from the qPCR runs.

3.3 Results

The biomass reported in Table 3.2 is the total mass of the Delta Smelt individuals reported in Column 2. The number of filters represents the number of water samples taken at each biomass

level. Descriptive statistics (mean, sd, and quantiles) report the findings for Quantification Cycles (Cq) which are inversely related to the concentration of DNA in each sample: Lower Cq values signify higher concentrations of environmental DNA in the sample. Cq values equal to 40 signify a non-detection of Delta Smelt DNA.

The number of positive DNA detections and the DNA concentration in samples increased with biomass (Figure 3.1). We initiated each experiment with "control" samples and those were graphed as biomass of zero (0) g (Figure 3.2, Top): All samples taken when Delta Smelt biomass was zero (0) g were negative (i.e., no detection of Delta Smelt eDNA). Then, the live pen with Delta Smelt was deployed at distance "0." We plotted data with respect to biomass and distance from DNA source in Figure 3.2. On inspection, we found that:

- The number of positive DNA detections decreased with distance from the source, and
- DNA concentration as measured by Quantification Cycles (Cq) increased with higher biomass.

3.4 Discussion

Under the range of conditions in which we worked, we found that Delta Smelt biomass was directly related to the probability of eDNA detection (Table 3.3). This made sense because 1) we studied three biomass levels, 0, 132, and 480 g (Table 3.2), 2) these biomass levels were composed of 0, 25, and 54 individual Delta Smelt, respectively, and 3) each individual fish generates intra- and extracellular DNA through release of mucus, feces, gametes and skin cells (Rourke et al. 2021). We observed some of the Delta Smelt adults we deployed expressed gametes but mucus, feces and skin cells were also all probable sources of DNA released into the water of Shag Slough during these experiments. Put simply, in the same period of time, more individuals tend to shed more DNA into the water than fewer individual fish.

The fewer quantification cycles (Cq) needed to detect the target DNA, the higher the concentration of DNA in the sample (see Chapter One "Estimating [eDNA] via qPCR" above; Ruiz-Villalba et al. 2021). So, it was evident from Figure 3.1, that there were more samples with smaller Cq values in the highest biomass level (480 g) compared to the intermediate level (132 g). Similarly, the intermediate biomass level exhibited more samples with smaller Cq values than did the minimum biomass level (zero (0) g). We concluded that when there was a positive detection, Delta Smelt biomass and distance from the DNA source were both related to the DNA concentration in a sample (Figure 3.2). This result was consistent with those of Rourke et al. (2021) who found that 57 of 63 reviewed studies found a positive relationship between eDNA and the abundance or biomass of the target fish species.

3.5 Figures

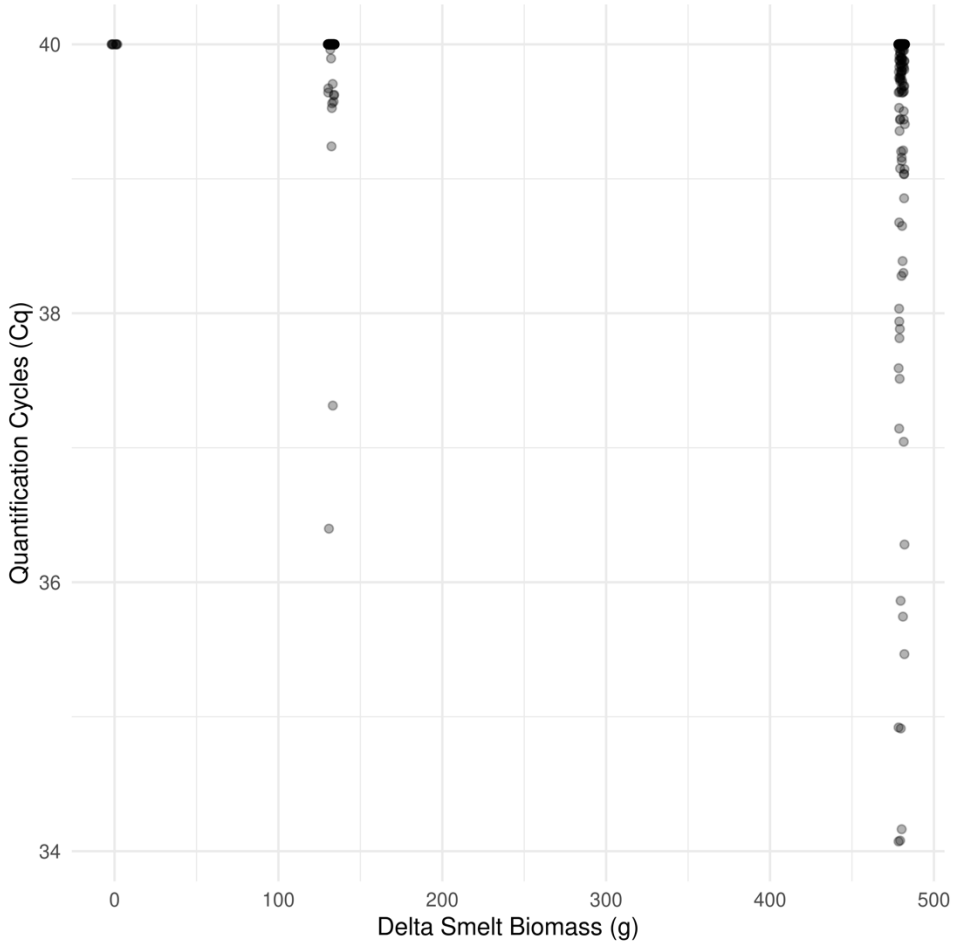
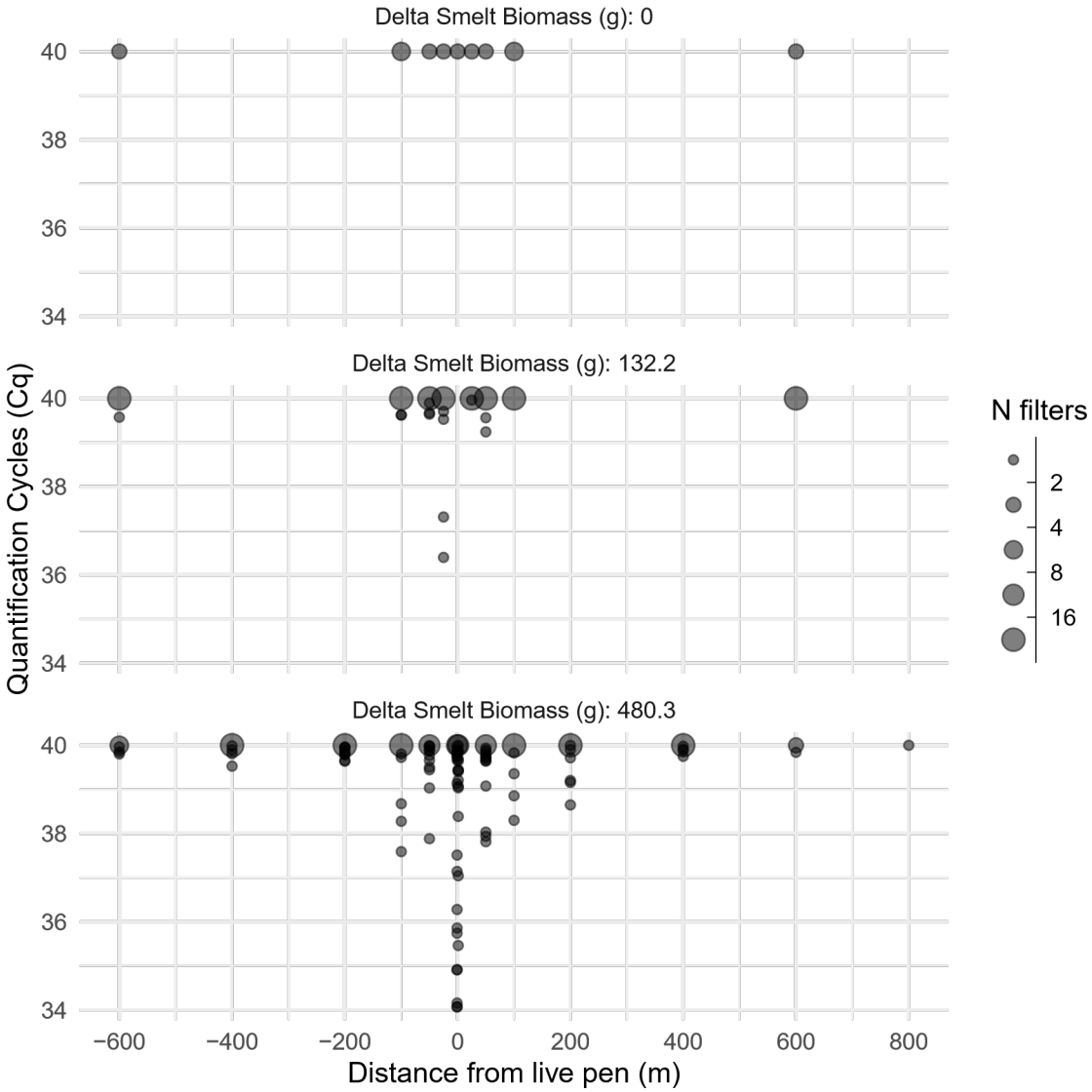


Figure 3.1: Observed qPCR Quantification Cycle (Cq) values for each biomass level. Lower Cq values signify higher concentrations of environmental DNA in the sample. Cq values equal to 40 signify a non-detection.


Delta Smelt Biomass (g): 0
Delta Smelt Biomass (g): 132.2
Delta Smelt Biomass (g): 480.3
Quantification Cycles (Cq)
40
38
36
34
−600
−400
−200
0
200
400
600
800
Distance from live pen (m)
N filters
2
4
8
16

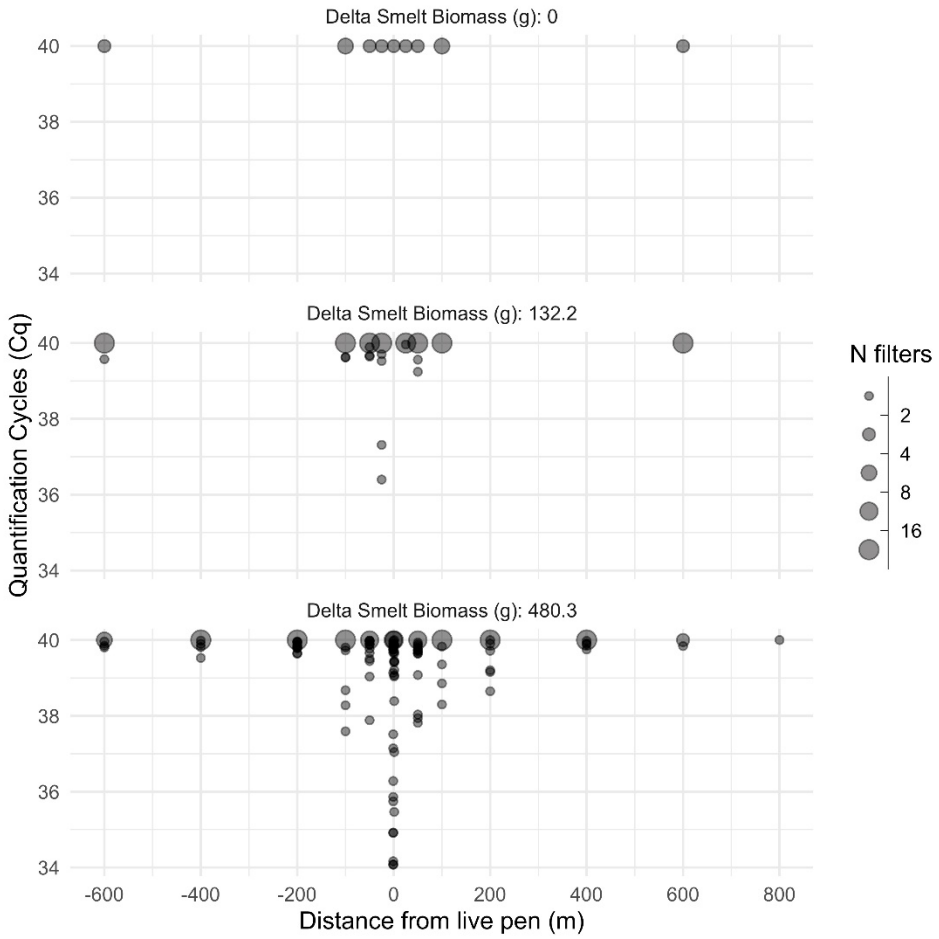


Figure 3.2: Observed qPCR Quantification Cycle (Cq) values for each biomass level by distance sampled from live pen. Negative distances are upstream of the live pen on an outgoing tide, while positive are downstream. Lower Cq values signify higher concentrations of environmental DNA in the sample. Cq values equal to 40 signify a non-detection.

3.6 Tables

Table 3.1: Summary of tidal environmental DNA experiments with two different biomass values of Delta Smelt placed in a live pen.

Experiment Name	Location	Environment type	Date	Delta Smelt count	Delta Smelt biomass (g)
Tidal-01	Shag Slough	Tidal	Oct. 2019	54	480.3
Tidal-03	Shag Slough	Tidal	Dec. 2020	25	132.2

Table 3.2: Summary of qPCR Quantification Cycles (Cq) values by biomass for two eDNA experiments using Delta Smelt.

Biomass (g)	N Delta Smelt	N filters	Mean	Standard deviation	Quantile (2.5%)	Quantile (97.5%)
0.0	0	33	40.00	0.00	40.00	40
132.2	25	234	39.96	0.30	39.61	40
480.3	54	283	39.67	0.97	35.88	40

Table 3.3: Summary of detections, a positive Delta Smelt DNA detection was defined as a quantification cycle (Cq) result below the cutoff threshold of 40.

Biomass (g)	N filters	N positive detection	% positive detections
0.0	33	0	0.0
132.2	234	13	5.6
480.3	283	91	32.2

References

Adams, C.I.M., M. Knapp, N.J. Gemmell, G. Jeunen, M. Bunce, M.D. Lamare, and H.R. Taylor. 2019. Beyond biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? Genes 10 (3): 192. https://doi.org/10.3390/genes10030192. Accessed 09/15/2021.

Akre, T.S., L.D. Parker, E. Ruther, J.E. Maldonado, L. Lemmon, and N.R. McInerney. 2019. Concurrent visual encounter sampling validates eDNA selectivity and sensitivity for the endangered wood turtle (*Glyptemys insculpta*). PLoS ONE 14(4): e0215586. https://doi.org/10.1371/journal.pone.0215586.

Baerwald, M.R., G. Schumer, B.M. Schreier, and B.P. May. 2011. TaqMan assays for the genetic identification of delta smelt (*Hypomesus transpacificus*) and wakasagi smelt (*Hypomesus nipponensis*). Molecular Ecology Resources 11(5):784–785.

Balasingham, K. D., R. P. Walter, and D. D. Heath. 2017. Residual eDNA detection sensitivity assessed by quantitative real-time PCR in a river ecosystem. Molecular Ecology Resources 17(3):523–532.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bergman, P.S., G. Schumer, S. Blankenship, and E. Campbell. 2016. Detection of adult Green Sturgeon using environmental DNA analysis. PLOS ONE 11(4):e0153500.

Bowen, M. D., B. B. Baskerville-Bridges, K. W. Frizell, L. Hess, C. A. Karp, S. M. Siegfried, and S. L. Wynn. 2004. Empirical and Experimental Analyses of Secondary Louver Efficiency at the Tracy Fish Collection Facility: March 1996 to November 1997. Vol. 11. U.S. Bureau of Reclamation, Technical Service Center, Denver, Colorado and Mid-Pacific Region, Sacramento, California.

https://www.usbr.gov/mp/TFFIP/docs/tracy-reports/tracy-rpt-vol-11-empirical-experimental-analyses.pdf. Accessed 09/15/2021.

Brandl, S., G. Schumer, B.M. Schreier, J.L. Conrad, B.May, and M.R. Baerwald. 2015. Ten real-time PCR assays for detection of fish predation at the community level in the San Francisco Estuary–Delta. Molecular Ecology Resources 15(2):278–284.

Bustin, S.A. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. Journal of Molecular Endocrinology 25(2):169–193.

Bustin, S.A., V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G.L. Shipley, J. Vandesompele, and C.T. Wittwer. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55(4):611–622.

Chiu, C.Y., and S.A. Miller. 2019. Clinical metagenomics. Nature Reviews Genetics 20(6):341–355.

Cloern, J. E., and A. D. Jassby. 2012. Drivers of change in estuarine-coastal ecosystems: Discoveries from four decades of study in San Francisco Bay. Reviews in Geophysics 50(RG4001). https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2012RG000397. Accessed 09/15/2021.

Cocolin, L., and D. Ercolini. 2015. Zooming into food-associated microbial consortia: a 'cultural' evolution. Current Opinion in Food Science 2:43–50.

Crane, L.C., J.S. Goldstein, D.W. Thomas, K.S. Rexroth, and A.W. Watts. 2021. Effects of life stage on eDNA detection of the invasive European green crab (*Carcinus maenas*) in estuarine systems. Ecological Indicators 124:107412.

Dejean, T., A. Valentini, C. Miquel, P. Taberlet, E. Bellemain, and C. Miaud. 2012. Improved Detection of an Alien Invasive Species Through Environmental DNA Barcoding: The Example of the American Bullfrog *Lithobates catesbeianus*: Journal of Applied Ecology 49 (4): 953–59. https://doi.org/10.1111/j.1365-2664.2012.02171.x. Accessed 09/15/2021.

Deltares. 2021. Delft3D-FLOW - Simulation of multi-dimensional hydrodynamic flows and transport phenomena, including sediments. User Manual. https://content.oss.deltares.nl/delft3d/manuals/Delft3D-FLOW_User_Manual.pdf . Accessed 09/15/2021.

Doi, H., K. Uchii, T. Takahara, S. Matsuhashi, H. Yamanaka, and T. Minamoto. 2015. Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. PLoS ONE 10(3): e0122763. doi:10.1371/journal.pone.0122763.

Dorazio, R.M., and R.A. Erickson. 2018. "ednaoccupancy : An R package for multiscale occupancy modelling of environmental DNA data." Molecular Ecology Resources 18 (2): 368–80. https://doi.org/10.1111/1755-0998.12735.

Environmental Science Associates (ESA). 2014. Hydrodynamic, Wind-wave, and Suspended Sediment Transport Modeling at a Breached Delta Wetland. Report prepared for the University of Washington, U.S. Fish and Wildlife Service, and California Department of Fish and Wildlife. ESA, San Francisco, CA.

Espe, M.B., M. Johnston, S.M. Blankenship, C. A. Dean, M.D. Bowen, A.A. Schultz, and G. Schumer. in review. The artemis package for eDNA analysis in R: An R package for the design and analysis of environmental DNA studies. Environmental DNA.

Fediajevaite, J., V. Priestley, R. Arnold, and V. Savolainen. 2021. "Meta-analysis shows that environmental DNA outperforms traditional surveys, but warrants better reporting standards." Ecology and Evolution, March, ece3.7382. https://doi.org/10.1002/ece3.7382.

Fernández, S., E. Arboleya, E. Dopico, A. Ardura, and E. Garcia-Vazquez. 2019. Nonindigenous fish in protected spaces: Trends in species distribution mediated by illegal tocking. Aquatic Conservation: Marine and Freshwater Ecosystems 29 (12): 2240–52. https://doi.org/10.1002/aqc.3238.

Ficetola, G.F., C. Miaud, F. Pompanon, and P. Taberlet. 2008. Species detection using environmental DNA from water samples. Biology Letters 4(4):423–425.

Fregoso, T.A., R.-F. Wang, E. Ateljevich, and B.E. Jaffe. 2017. A new seamless, high-resolution digital elevation model of the San Francisco Bay-Delta Estuary, California. US Geological Survey. https://pubs.usgs.gov/of/2017/1067/ofr20171067.pdf. Accessed 09/15/2021.

Fukumoto, S., A. Ushimaru, and T. Minamoto. 2015. A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: A case study of Giant Salamanders in Japan. Journal of Applied Ecology 52 (2): 358–65. https://doi.org/10.1111/1365-2664.12392.

Goldberg, C.S., C.R. Turner, K. Deiner, K.E. Klymus, P.F. Thomsen, M.A. Murphy, S.F. Spear, A. McKee, S.J. Oyler-McCance, R.S. Cornman, M.B. Laramie, A.R. Mahon, R.F. Lance, D.S. Pilliod, K. M. Strickler, L.P. Waits, A.K. Fremier, T. Takahara, J.E. Herder, and P. Taberlet. 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution 7(11):1299–1307.

Goldberg, C.S., A. Sepulveda, A. Ray, J. Baumgardt, and L.P. Waits. 2013. Environmental DNA as a new method for early detection of New Zealand Mudsnails (*Potamopyrgus antipodarum*). Freshwater Science 32 (3): 792–800. https://doi.org/10.1899/13-046.1.

Goodrich, B., J. Gabry, I. Ali, and S. Brilleman. 2020. Rstanarm: Bayesian applied regression modeling via Stan. https://mc-stan.org/rstanarm.

Google. 2020. Classification: Precision and Recall. https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall. Accessed 09/15/2021.

Handelsman, J. 2004. Metagenomics: Application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews 68(4):669–685.

Hinlo, R., E. Furlan, L. Suitor, and D. Gleeson. 2017. Environmental DNA monitoring and management of invasive fish: Comparison of eDNA and fyke netting. Management of Biological Invasions 8 (1): 89–100. https://doi.org/10.3391/mbi.2017.8.1.09.

Hobbs, J., P.B. Moyle, N. Fangue, and R.E. Connon. 2017. Is extinction inevitable for Delta Smelt and Longfin Smelt? An opinion and recommendations for recovery. San Francisco Estuary and Watershed Science 15(2). https://escholarship.org/content/qt2k06n13x/qt2k06n13x.pdf. Accessed 09/15/2021.

Huver, J.R., J. Koprivnikar, P.T.J. Johnson, and S. Whyard. 2015. Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic ecosystems. Ecological Applications 25 (4): 991–1002. https://doi.org/10.1890/14-1530.1.

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

Jenkins, J.A. H.L. Bart, Jr., J.D. Bowker, P.R. Bowser, J.R. MacMillan, J.G. Nickum, J.D. Rose, P.W. Sorensen, and G.W. Whitledge, J.W. Rachlin, and B.E. Warkentine. 2014. Guidelines for the use of fishes in research. American Fisheries Society, Bethesda, Maryland. https://fisheries.org/docs/wp/Guidelines-for-Use-of-Fishes.pdf

Kubista, M., J.M. Andrade, M. Bengtsson, A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A. Ståhlberg, and N. Zoric. 2006. The real-time polymerase chain reaction. Molecular Aspects of Medicine 27(2):95–125.

Lee, J., J. Choi, M. Fatka, E. Swanner, K. Ikuma, X. Liang, T. Leung, and A. Howe. 2020. Improved detection of mcyA genes and their phylogenetic origins in harmful algal blooms. Water Research 176:115730.

Martellini, A., P. Payment, and R. Villemur. 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. Water Research 39(4):541–548.

Management, Analysis, and Synthesis Team (MAST). 2015. An updated conceptual model of Delta Smelt biology: our evolving understanding of an estuarine fish. Interagency Ecological Program, Technical Report 90. https://pubs.er.usgs.gov/publication/70141018. Accessed 09/15/2021.

Merz, J.E., S. Hamilton, P.S. Bergman, and B. Cavallo. 2011. Spatial perspective for delta smelt: a summary of contemporary survey data. California Fish and Game 97(4):164–189.

Milián-García, Y., R. Young, M. Madden, E. Bullas-Appleton, and R.H. Hanner. 2021. Optimization and validation of a cost-effective protocol for biosurveillance of invasive alien species. Ecology and Evolution 11 (5): 1999–2014. https://doi.org/10.1002/ece3.7139.

Miya, M., Y. Sato, T. Fukunaga, T. Sado, J.Y. Poulsen, K. Sato, T. Minamoto, S. Yamamoto, H. Yamanaka, H. Araki, M. Kondoh, and W. Iwasaki. 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society Open Science 2(7):150088.

Morris, C.E., M. Bardin, O. Berge, P. Frey-Klett, N. Fromin, H. Girardin, M.-H. Guinebretière, P. Lebaron, J.M. Thiéry, and M. Troussellier. 2002. Microbial biodiversity: Approaches to experimental design and hypothesis testing in primary scientific literature from 1975 to 1999. Microbiology and Molecular Biology Reviews 66(4):592–616.

Moyer, G.R., E. Díaz-Ferguson, J.E. Hill, and C. Shea. 2014. Assessing environmental DNA detection in controlled lentic systems. PLoS ONE 9 (7): e103767. https://doi.org/10.1371/journal.pone.0103767.

Moyle, P.B., L.R. Brown, J.R. Durand, and J.A. Hobbs. 2016. Delta Smelt: Life history and decline of a once-abundant species in the San Francisco Estuary. San Francisco Estuary and Watershed Science 14(2). https://escholarship.org/uc/item/09k9f76s. Accessed 09/15/2021.

Piggott, M.P. 2016. Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. Ecology and Evolution 6 (9): 2739–50. https://doi.org/10.1002/ece3.2083.

Pilliod, D.S., C.S. Goldberg, R.S. Arkle, and L.P. Waits. 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. Canadian Journal of Fisheries and Aquatic Sciences 70 (8): 1123–30. https://doi.org/10.1139/cjfas-2013-0047.

Polansky, L., L. Mitchell, and K.B. Newman. 2019. Using multistage design-based methods to construct abundance indices and uncertainty measures for Delta Smelt. Transactions of the American Fisheries Society 148(4):710–724.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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

Rourke, M.L., A.M. Fowler, J.M. Hughes, M.K. Broadhurst, J.D. DiBattista, S. Fielder, J.W. Walburn, and E.M. Furlan. 2021. Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. Environmental DNA. https://doi.org/10.1002/edn3.185.

Roussel, J.-M., J.-M. Paillisson, A. Tréguier, and E. Petit. 2015. The downside of eDNA as a survey tool in water bodies. Journal of Applied Ecology 52(4):823–826.

Ruiz-Villalba, A.; J.M. Ruijter, and M.J.B. van den Hoff. 2021. Use and misuse of Cq in qPCR data analysis and reporting. Life 2021, 11, 496. https://doi.org/10.3390/life11060496.

Sassoubre, L.M., K.M. Yamahara, L.D. Gardner, B.A. Block, and A.B. Boehm. 2016. Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. Environmental Science & Technology 50(19):10456–10464.

Schmelzle, M.C., and A.P. Kinziger. 2016. Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. Molecular Ecology Resources 16(4):895–908.

Schmidt, B.R., M. Kéry, S. Ursenbacher, O.J. Hyman, and J.P. Collins. 2013. Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. Methods in Ecology and Evolution 4 (7): 646–53. https://doi.org/10.1111/2041-210X.12052.

Schumer, G., K. Crowley, E. Maltz, M. Johnston, P. Anders, and S. Blankenship. 2019. Utilizing environmental DNA for fish eradication effectiveness monitoring in streams. Biological Invasions 21 (11): 3415–26. https://doi.org/10.1007/s10530-019-02056-z.

Shelton, A.O., R.P. Kelly, J.L. O'Donnell, L. Park, P. Schwenke, C. Greene, R.A. Henderson, and E.M. Beamer. 2019. Environmental DNA provides quantitative estimates of a threatened salmon species. Biological Conservation 237:383–391.

Sepulveda, A.J., C. Schmidt, J. Amberg, P. Hutchins, C. Stratton, C. Mebane, M.B. Laramie, and D.S. Pilliod. 2019. Adding invasive species biosurveillance to the U.S. Geological Survey Streamgage Network. Ecosphere 10 (8). https://doi.org/10.1002/ecs2.2843.

Sigsgaard, E., H.C. Egelyng, P.R. Møller, and P.F. Thomsen. 2015. Monitoring the near-extinct European Weather Loach in Denmark based on environmental DNA from water samples. Biological Conservation 183 (March): 46–52. https://doi.org/10.1016/j.biocon.2014.11.023.

Song, J.W., M.J. Small, and E.A. Casman. 2017. Making sense of the noise: The effect of hydrology on Silver Carp eDNA detection in the Chicago Area Waterway System. Science of The Total Environment 605-606 (December): 713–20. https://doi.org/10.1016/j.scitotenv.2017.06.255.

Stan Development Team. 2021. Stan Modeling Language Users Guide and Reference Manual. https://mcstan.org.

Stompe, D.K., P.B. Moyle, A. Kruger, and J.R. Durand. 2020. Comparing and integrating fish surveys in the San Francisco Estuary: Why diverse long-term monitoring programs are important. San Francisco Estuary and Watershed Science 18(2). https://escholarship.org/content/qt0b91527k/qt0b91527k.pdf.

Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21(8):2045–2050.

Thomsen, P.F., J. Kielgast, L.L. Iversen, C. Wiuf, M. Rasmussen, M.T.P. Gilbert, L. Orlando, and E. Willerslev. 2012. Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology 21(11):2565–2573.

Thomsen, P.F., and E. Willerslev. 2015. Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation 183:4–18.

Turner, C.R., M.A. Barnes, C.C. Xu, S.E. Jones, C.L. Jerde, and D.M. Lodge. 2014. Particle size distribution and optimal capture of aqueous macrobial eDNA. Methods in Ecology and Evolution 5(7):676–684.

United States Fish and Wildlife Service. (USFWS). 1993. Endangered and Threatened Wildlife and Plants; Determination of Threatened Status for the Delta Smelt. USFWS editor, 50 CFR Part 17.

Vehtari, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing 27(5):1413–1432.

Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D. Wu, I. Paulsen, K.E. Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H. Rogers, and H.O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304(5667):66–74.

Wittwer, C.T., M.G. Herrmann, A.A. Moss, and R.P. Rasmussen. 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. BioTechniques 22(1):130–138.

Yao, Y., A. Vehtari, D. Simpson, and A. Gelman. 2017. Using stacking to average Bayesian predictive distributions. Bayesian Analysis. https://doi.org/10.1214/17-BA1091.