9-2024

Identifying Key Environmental Drivers of Reach-Scale Salmonid eDNA Recovery With Random Forest

Shawn A. Melendy

John R. Olson

Environmental DNA

WILEY

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

ORIGINAL ARTICLE OPEN ACCESS

Identifying Key Environmental Drivers of Reach-Scale Salmonid eDNA Recovery With Random Forest

Shawn A. Melendy | John R. Olson

Department of Applied Environmental Science, California State University, Monterey Bay, Seaside, California, USA

Correspondence: Shawn A. Melendy (smelendy1221@gmail.com) | John R. Olson (joolson@csumb.edu)

Received: 8 January 2024 | **Revised:** 1 August 2024 | **Accepted:** 5 August 2024

Funding: This work was supported by the NOAA Office of Education, #NA16SEC4810009.

Keywords: brook trout | discharge | eDNA fate | environmental DNA | random forest | rivers

ABSTRACT

Environmental DNA (eDNA) sampling from rivers has emerged as a promising new method for monitoring freshwater organisms of management concern. However, eDNA sampling cannot yet offer reliable estimates of a target species' abundance/biomass or confident determinations of a species' absence from a river segment. To unlock these abilities—and thereby greatly improve eDNA as a tool for management decision-making—the influence of local environmental factors on eDNA fate must be better understood. At nine river sites across the central California coast, we added a known quantity of novel eDNA (Brook Trout, *Salvelinus fontinalis*) and collected eDNA at sequential downstream distances for qPCR analysis. We then used random forest modeling to identify the most important environmental factors to reach-scale (≤ 200 m) sampling outcomes and characterize salmonid eDNA fate. Our final model identified six factors important to sampling outcomes, including five environmental factors (discharge, local catchment calcium oxide content, average depth of the sampling cross-section, presence of pools, and impervious cover of the watershed) and one factor regarding our experimental design (the number of qPCR technical replicates). Our results highlight the notable effects of cross-sectional area, turbulence, and catchment geology on eDNA fate, and we suggest the discharge and presence of pools as useful proxies for evaluating a site's favorability for eDNA recovery.

1 | Introduction

Monitoring the abundance and distribution of wild populations is essential to determine the success of conservation efforts. Such efforts are vital to protect rivers, which are among the most altered ecosystems globally (Dudgeon 2019; Reid et al. 2019; Vörösmarty et al. 2010). Salmonids are among the most important freshwater species to monitor, as they play critical ecological roles in nutrient cycling, habitat modification, and as a food source, but have suffered widespread impacts of habitat destruction, pollution, invasive species, overfishing, and climate change (Crozier et al. 2019; Mullan 1987; Nehlsen, Williams, and Lichatowich 1991). Imperiled (endangered, threatened, and vulnerable) salmonid species are in critical need of monitoring, particularly for the size and migrations of their anadromous populations, their spawning and juvenile activity, and expansions or contractions of their ranges. These aims have motivated numerous management efforts and studies in North America that rely on traditional surveillance methods (e.g., snorkel surveys, electrofishing, weirs) across a vast number of streams. These approaches rely on visual or hand counting of individuals and are inherently time-consuming and disruptive to animals. In recent years, the detection of organisms using environmental DNA has emerged as a complementary approach that could reduce or, in some cases, eliminate the need for such methods of direct observation.

Environmental DNA (eDNA) refers to DNA that is shed or excreted (e.g., tissue, mucous, saliva, urine, feces) into the environment by an organism (Taberlet et al. 2012). In aquatic

environments, eDNA can be analyzed from water samples, revealing the presence of target species without direct handling or observation. In this way, eDNA has already shown tremendous utility as a noninvasive biomonitoring tool in rivers. It has proven a viable, cost-effective method for targeted detection of species, including those that are rare, cryptic, or invasive (Bedwell and Goldberg 2020; Spence et al. 2021; Wilcox et al. 2013; Wittwer et al. 2018), and is increasingly used for assessing biodiversity (Civade et al. 2016; Deiner et al. 2016; Lodge et al. 2012; Pont et al. 2018; Valentini et al. 2016). However, further applications of eDNA in rivers are limited at present. Presence/absence monitoring via eDNA remains limited by the challenge of interpreting negative sampling outcomes: deciding if a negative sample indicates the absence of the target taxon, or simply indicates a failure to recover the target eDNA from the water column. Additionally, while recovered eDNA quantities have shown positive correlations with observed abundance/biomass of freshwater vertebrates, the relationship varies significantly across studies and species and is less clear at lower eDNA concentrations (Doi et al. 2017; Lacoursière-Roussel et al. 2016; Pilliod et al. 2013; Sepulveda et al. 2021; Wilcox et al. 2016). Thus, reliably estimating abundance/biomass from eDNA alone remains elusive.

These limitations result from the inability to predict how much eDNA will persist in the water column over time and distance, given a particular target organism and river environment. Accurate predictions of this kind could revolutionize aquatic biomonitoring, but they require a much-improved understanding of two complex phenomena: eDNA release profiles and eDNA fate. A release profile refers to the amount, rate, and particle size distribution of eDNA released by an organism and can vary by species, life stage, metabolic rate, and activities (Thalinger et al. 2021; Yates et al. 2021). A number of studies have investigated fish eDNA particle size distribution (Brandão-Dias et al. 2023; Jo et al. 2019; Sassoubre et al. 2016; Wilcox et al. 2015) and release rate (Klymus et al. 2015; Jo et al. 2019; Maruyama et al. 2014; Wilcox et al. 2016; and others reviewed in Rourke et al. 2021), but release profiles cannot yet be reliably predicted for wild individuals or populations. eDNA fate refers to the dispersion, decay (via chemical degradation, microbial consumption, and mechanical fragmentation), deposition (via benthic adhesion and gravitational settling), and/or resuspension of eDNA particles once they are released from an organism. Several studies on eDNA fate have been performed in controlled mesocosms and other environments, offering insight into processes of dispersion (Andruszkiewicz et al. 2019; Laporte et al. 2020; Wood et al. 2021), decay (Andruszkiewicz, Sassoubre, and Boehm 2017; Barnes et al. 2014; Sassoubre et al. 2016; Seymour et al. 2018; Shogren et al. 2017), deposition, and resuspension (Jerde et al. 2016; Shogren et al. 2017).

However, little progress has been made to understand the cumulative effect of these mechanisms, and thereby predict eDNA fate, given the complex set of environmental factors that influence them in rivers. Two prior studies have taken preliminary steps toward this understanding by searching for variables that can improve the relationship between target abundance/biomass and the amount of eDNA recovered downstream (Sepulveda et al. 2021; Tillotson et al. 2018). Using model selection by AIC approaches, Tillotson et al. (2018) found a minor effect of water temperature on their correlation between target abundance and eDNA concentration, and Sepulveda et al. (2021) found minimal support for including stream habitat attributes in their models (with one potentially significant correlation for percent pool). Importantly, both studies saw inclusion of random terms for site improve model performance, suggesting there are environmental factors consistently associated with eDNA concentrations, but they could not be identified.

We posited that a random forest (RF) modeling approach relating environmental factors to eDNA sampling outcomes is better suited for interpreting the effect of river environment on eDNA fate, given the multitude of potentially relevant variables and their high degree of interaction, as well as RF's ability to identify the most influential variables from high-dimensional data. Other machine learning algorithms have been used for related purposes: Ogburn et al. (2022) used boosted regression trees (BRTs) to explore the relationship between landscape-scale metrics (e.g., elevation, watershed area, land use) and presence/absence sampling outcomes for anadromous herring, inferring which habitat types within the watershed were preferred for spawning. Here, we took a similar approach but used RF modeling to explore the relationship between environmental factors (e.g., depth, percent riffle/pool, substrate cobble size) and the amount of target eDNA recovered downstream.

Specifically, we aimed to (1) identify the most important variables explaining the amount of eDNA recovered downstream using the *VSURF* R package (VSURF: variable selection using random forest) and (2) characterize the mechanisms of salmonid eDNA fate based on those variables and their modeled relationships with the response. Additionally, we sought to (3) evaluate which, if any, of said variables could be a useful proxy for the eDNA fate profile (i.e., how favorable local conditions are to eDNA recovery) of a given site.

2 | Methods

2.1 | Overview

We added novel salmonid eDNA (Brook Trout [*Salvelinus fontinalis*]) to river sites representing a range of environmental conditions, tracked the subsequent pulses of eDNA visually using fluorescein dye, and sampled from the pulses (leading edge to trailing edge) at sequential distances up to 200 m. We then developed a random forest (RF) model of the eDNA sampling outcomes at each distance as a function of river environmental factors. This experimental design allowed us to quantify the total amount of eDNA added to each river and likely minimized the effects of resuspension on sampling outcomes, since resuspension has been shown—at least in small experimental streams—to primarily occur multiple hours postdeposition (Shogren et al. 2017). Accordingly, this represents a novel approach to investigate how eDNA fate mechanisms—principally dispersion, decay, and deposition—are influenced by environmental factors at the reach-scale, offering valuable insight toward eDNA abundance/biomass estimates and presence/absence determinations for salmonids or other freshwater vertebrates.

2.2 | Site Selection and Data

We selected nine river sites within the central California coast, with the goal of including a range of environmental conditions and corresponding eDNA fate profiles (Figure 1). We emphasized stream order, discharge, substrate, level of anthropogenic disturbance, and accessibility as selection criteria. Table 1 shows a small selection of environmental factors to illustrate the range of conditions captured across all sites. The central California coast has a Mediterranean climate with dry summers and mild winters, in which flow regimes reflect patterns of precipitation. The sites we selected range from small headwater streams in hardwood woodlands to medium-sized rivers in conifer forests, urban landscapes, or agriculture-dominated plains.

2.3 | Site Characterization

We took a series of measurements to characterize each river site's 200 m experimental reach. Measurements were taken after the eDNA trials to avoid disruption of substrate. We measured discharge, total suspended sediment (TSS), and water chemistry metrics once at the top of our experimental reaches, 1 m upstream of the eDNA addition point. Discharge (m^3/s) was measured using a flow meter (SonTek, S19-02-1219). We sampled TSS (g/L) using a 1-L high-density polyethylene bottle secured inside a weighted-bottle sampler (Rickly Hydrological Co., US WBH-96). The 1-L bottle contents were dried and weighed for each study site. Water chemistry measurements (temperature [°C], specific conductivity [μS/cm], total dissolved solids [TDS] [g/L], salinity, dissolved O_2 [mg/L], and pH) were made with a handheld meter (YSI, 6050000). Additionally, we recorded macrohabitat type, channel slope, wetted width (m), depth (cm), substrate particle size (mm), substrate organic cover, substrate course particulate organic matter (CPOM) cover, and substrate biofilm thickness throughout the length of our experimental reaches. We measured the length in meters of each macrohabitat type (e.g., riffle, pool) by walking each reach and recording each contiguous segment. Channel slope was determined from elevation measurements taken at each riffle head, using a rotating laser (Topcon, RL-H5A) and measuring rod. We established cross-sectional transects every 10 m throughout each 200 m reach and measured wetted width (m). At 5 points along each transect (5%, 25%, 50%, 75%, and 95% across the wetted width), we measured depth (cm) and substrate particle size (mm) with a gravelometer field sieve (Wildco, 3-14-D40) and recorded the presence/absence of organic substrate and course particulate organic matter (CPOM). Biofilm thickness was measured by collecting one substrate particle from the midpoint of each cross-section, scraping a circular area (with a rubber stencil and a toothbrush) centered on the upper face of each particle, and measuring the combined dry mass of this material for each site following Steinman et al. (2017).

2.4 | eDNA Source

We selected Brook Trout (*Salvelinus fontinalis*) eDNA because we required a source organism that was not present in our selected rivers, to avoid background detections. eDNA was collected from a 120 × 4 m raceway at the California Department of Fish and Wildlife San Joaquin Hatchery. The raceway contained ~4900 trout (671 days old) on the date of our first eDNA collection (11/5/21), and ~500 trout (847 days old) on our final collection date (5/1/22). Following the approach of Snyder et al. (2021), water was scooped into pre-sterilized 5-gallon buckets. The buckets were closed with a lid and transported back to CSUMB campus for room temperature storage before being used at a river site 22–24 h after collection. Each experimental addition of eDNA occurred within this time window to ensure similar levels of degradation across trials. Just prior to adding the eDNA to each river, we collected 2-L

FIGURE 1 | Locations of the nine chosen river sites/experimental reaches, labeled with green pins.

TABLE 1 | Coordinates and select environmental factors for each river site/experimental reach.

Site	Coordinates	Environment	Subwatershed area (km^2)	Discharge (m^3/s)	Slope	Temp (°C)	Conductivity (μS/cm)	Dissolved O_2 (mg/L)	pH
Arroyo Seco	36°16′51″ N, 121°19′21″ W	Shrub/Agriculture (Hardwood Riparian Corridor)	63.1	0.787	0.11	13.3	353	8.3	6.1
Big Sur River	36°16′29″ N, 121°49′29″ W	Conifer Forest/Hardwood Woodland	151.3	1.660	0.75	8.4	406	20.1	4.7
Carmel River (Lower)	36°32′05″ N, 121°54′33″ W	Urban/Herbaceous (Hardwood Riparian Corridor)	80.1	0.776	0.13	9.7	284	11.9	5.5
Carmel River (Middle)	36°31′24″ N, 121°49′50″ W	Hardwood Woodland	80.1	0.360	0.99	10.8	269	10.0	7.1
Carmel River (Upper)	36°26′04″ N, 121°42′35″ W	Hardwood Woodland	90.3	0.094	1.23	13.9	499	17.6	5.5
Las Garzas Creek	36°26′14″ N, 121°50′42″ W	Hardwood Woodland	101.7	0.038	1.24	8.1	251	19.0	6.3
Pajaro River	36°53′39″ N, 121°38′39″ W	Agriculture (Hardwood Riparian Corridor)	135.2	0.564	0.17	14.7	921	8.3	5.6
San Clemente Creek	36°25′34″ N, 121°45′14″ W	Hardwood Woodland/Conifer Forest	90.3	0.028	6.44	5.8	222	13.5	7.6
San Jose Creek	36°29′13″ N, 121°51′50″ W	Hardwood Woodland/Conifer Forest	37.2	0.021	3.01	8.4	205	12.0	5.9

Note: Discharge and water chemistry metrics were recorded just after each eDNA experiment, 1 m upstream from the eDNA addition point. Environment is based on the predominant vegetation cover types at each site, according to the California Wildlife Habitat Relationship System (CDFW 2021).

preliminary samples from the bucket to determine the starting concentration.

2.5 | River Sampling

We added 5-gallon buckets of *S. fontinalis* eDNA to each river at a riffle head and at the thalweg, quickly and smoothly pouring out the bucket contents at the river surface. One bucket was added to each site, except for Arroyo Seco (two buckets) and the Pajaro River (four buckets) as fish counts at the hatchery declined. At the same moment we added the eDNA, we also added fluorescein dye (Thermo Scientific, 119240250), premixed with river water in a Nalgene bottle. We adjusted the amount of added fluorescein to account for variation in river discharge (2 g for Carmel River [Upper], Las Garzas Creek, San Clemente Creek, San Jose Creek; 3 g for Arroyo Seco, Carmel River [Lower], Carmel River [Middle], Pajaro River; 4 g for Big Sur River). This dye acted as a visual marker of where the eDNA pulse was located as it flowed downstream, thus informing crew members when to begin and end eDNA collection at each sampling distance.

Crew member pairs were positioned to collect eDNA samples from the thalweg (≤2 ft. below the surface) at 10, 50, 100, and 200 m downstream, with the exception of the Pajaro River site, where samples were collected at 10, 20, 40, and 60 m due to a debris flow blocking downstream access. If crew needed to position themselves in the water to access the thalweg, they walked and stood downstream of the sampling point. When the leading edge of the fluorescein dye plume first reached each pair of crew members, that pair began to pump their first sample. They pumped until 5 L of water had been filtered (or until the filter clogged, whichever came first), and immediately proceeded to take another sample, repeating the process until the entire plume of dye had passed. The sequential order of samples at a given distance was denoted as "A", "B", and "C" samples. Because the dye plume tended to elongate as it flowed downstream, this approach often required fewer samples at shorter distances (e.g., only "A" at 10 m), and more at longer distances (e.g., "A" through "C" at 200 m).

We collected samples following Carim, McKelvey et al. (2016), using peristaltic pumps (Geotech, 91350103) to direct river water through 1.6-μm glass microfiber filters (Whatman, 1820-047) and into outflow buckets. The filter holder/collection cup was lowered into the water pointing in the upstream direction. Filters were subsequently folded twice-over and transferred, using plastic forceps, to a 50-mL tube containing approximately 25 mL of silica-bead desiccant. We protected the tubes from sunlight and heat during their transport to the lab. Pump tubing, forceps, and other sampling equipment were sterilized with a 20% bleach solution and thoroughly rinsed with DI water prior to collection.

2.6 | eDNA Extraction

We performed eDNA extractions in a dedicated room, separate from where the eDNA buckets, pumps, tubing, and other field equipment were stored. We split each filter in half, with one side extracted and the other archived. Extractions used a combination of the DNeasy Blood and Tissue Kit (Qiagen, 69506) and Qiashredder Kit (Qiagen, 79654) in a modified protocol developed by Carim, Dysthe et al. (2016). We soaked each filter half in ATL buffer and proteinase K on a shaking incubator at 65°C overnight and used 80 μL AE buffer for the final elution step. Elutes were concentrated using a benchtop centrifugal vacuum concentrator (Labconco, 7810012) at 34°C–37°C for 30–90 min and resuspended in 20 μL of AE buffer in a shaking incubator at 65°C, 850–1000 rpm, for 10 min.

2.7 | qPCR

We set up qPCR reactions in a dedicated room, separate from where the eDNA buckets, pumps, tubing, and other field equipment were stored, and we loaded each plate in a laminar flow hood. We utilized a probe-based qPCR protocol, targeting the *S. fontinalis cytochrome c* oxidase I (COI) sequence (amplicon length: 103) identified by Hulley et al. (2019). F: CGGTACGGGGTGAACAGTTT, R: GGAAATGCCAGCTAAATGTAGGG, P: FAM-CTCGCCCA CGCAGGAGCTTC-QSY. Primer and probe concentrations followed Hulley et al. (2019). We ran three technical replicates for each eDNA extraction (from a half filter), with the exception of the 10-A, 100-A, and 200-B samples from Las Garzas Creek, for which only two technical replicates were run due to limited remaining sample volumes. Each 20 μL reaction contained: 10 μL TaqMan Environmental Master Mix 2.0 (Applied Biosystems, 4396838), 1 μL each primer, 1 μL TaqMan QSY probe (Applied Biosystems, 4482777), 2 μL internal positive control (IPC) mix (Applied Biosystems, 4308321), 0.4 μL IPC DNA (Applied Biosystems, 4308321), and 4.6 μL sample. We ran reactions in 96-well plates (VWR, 82006-664) with (1) a triplicate standard curve with 10^1–10^6 COI copies inserted to a linearized plasmid (Integrated DNA Technologies), (2) triplicate no template control with DEPC water, and (3) triplicate no amplification control for the IPC reaction. Plates were sealed with clear adhesive film for qPCR (VWR, 60941-078). A CFX96 Touch Real-Time PCR Detection System (Bio-Rad, 1855196) was set to the following thermocycling conditions: 10-min activation at 95°C, followed by 45 cycles of 95°C for 15 s and 60°C for 60 s. We considered samples with a single positive replicate as positive for *S. fontinalis* DNA. Positive replicates were confirmed by visually inspecting amplification curve morphology.

We modeled the assay's limit of detection and limit of quantification (LOD/LOQ) using the R script developed for eDNA applications by Merkes et al. (2019), and the approach outlined in Klymus et al. (2020). Based on 13 replicates of our standard curve, the script's curve-fitting method determined an LOD at 13.7 copies (95% detection probability, given three sample replicates) and an LOQ at 299.55 copies (CV = 0.35).

2.8 | Generalized Linear Modeling

To determine the effect of distance on the proportion of positive technical replicates (PPTR) (our proxy for the amount of recovered eDNA), we developed a generalized linear model

using the *glmmTMB* R package (Brooks et al. 2017). Since PPTR is proportion data, we chose to run a binomial regression with the logit link function. We also weighted the model by the number of qPCR technical replicates run for each sampling distance, to account for the systematic bias described in section 2.10.

2.9 | Occupancy Analysis

We used a Bayesian multiscale occupancy modeling approach to estimate the probability of occurrence of *S. fontinalis* eDNA in each water sample (θ), at each sampling distance, for each of our river sites. Using the *eDNAoccupancy* package in R (Dorazio and Erickson 2018), we fit a model with distance and sample order as covariates, using 11,000 Markov chain Monte Carlo iterations. Posterior means and 95% credible intervals for θ were estimated after a burn-in of 1000 iterations. We then calculated each sampling distance's aggregate probability of occurrence (θ_{Agg}) (the probability that *S. fontinalis* eDNA occurs in *any* of water samples "A", "B", or "C") with the following equation:

$$\theta_{\text{Agg}} = 1 - \left[\left(1-\theta_{\text{A}}\right) * \left(1-\theta_{B}\right) * \left(1-\theta_{C}\right)\right]$$

2.10 | Random Forest Modeling

We developed a random forest (RF) model (Breiman 2001) to interpret which river environmental factors have the greatest effect on downstream eDNA sampling outcomes. RF is a non-parametric regression and classification modeling approach that is well suited for numerous predictor variables with interacting effects. We used the proportion of positive technical qPCR replicates across all water samples at a given sampling distance (PPTR) as the response variable. For example, if one site's "A" water sample at 50 m had 2/3 positive technical replicates, and the "B" water sample at 50 m had 1/3 positive technical replicates, the PPTR for 50 m was 3/6 = 0.50. PPTR served as a rough proxy for amount of eDNA recovered, since the copy number values from our qPCR results were reliably below the assay's limit of quantification (LOQ = 299.55).

Using PPTR in this way introduced a potential systematic bias into our response. Since each fluorescein dye plume tended to stretch longitudinally while traveling downstream, more back-to-back water samples were required to continuously collect from the leading edge through the trailing edge. The same number of eDNA particles could then be diluted across more samples at the downstream distances (e.g., 100 and 200 m), potentially making positive qPCR technical replicates less likely. Thus, even in a hypothetical experiment with no lateral dispersion, degradation, or deposition, our resulting PPTR values could have falsely indicated a decline in the amount of thalweg eDNA over distance. We accounted for this potential bias in our model by including, as a predictor, the number of qPCR technical replicates run for each sampling distance (the total across "A", "B", and/or "C" water samples). Accordingly, this predictor's variable importance could be used as a measure of the bias strength, and its modeled relationship with PPTR (alongside our occupancy analysis) could indicate the eDNA plume dynamics in relation to the fluorescein.

We assessed the importance of 66 potential predictor variables to determine those with the greatest influence on PPTR (Table 2). We used the *VSURF* R package (VSURF: variable selection using random forest) for predictor and model selection, with default settings except that 200 forests were built for the thresholding step and 100 forests were built for the interpretation step (Genuer et al. 2015). The *VSURF* package first uses a thresholding step to eliminate all predictors below a threshold of variable importance (the percentage increase in mean squared error when a particular predictor's values are randomly permuted). For example, predictors with no variability would be eliminated at this thresholding step. Then, the interpretation step compares all nested models for the remaining predictors and selects one with the lowest out-of-bag error. This step aims to select all highly relevant predictors to the response, perhaps with some redundancy, whereas the final *VSURF* prediction step—which we did not use—further eliminates variables to find a model optimal for prediction (Genuer et al. 2015). In comparison with other random forest-based variable selection R packages, *VSURF* is more conservative in selecting truly relevant variables (Speiser et al. 2019). Thus, *VSURF* may miss some weakly relevant variables but is less prone to selecting false positives. For this reason, the *VSURF* interpretation step was ideal for our purpose of identifying only the top truly relevant variables to PPTR (i.e., predictors that play a large role in shaping eDNA fate and sampling outcomes across river environments).

Although correlated predictors do not influence the performance of RF models, they can bias variable importance estimates and lead to misleading conclusions about the influence of predictors (Strobl et al. 2008). To avoid this, we assessed correlations among each of the predictors selected by the *VUSRF* interpretation step, removing predictors to avoid any pairs with a Pearson's correlation of 0.7 or greater in our final model.

3 | Results

3.1 | eDNA Added

Our first collection of *S. fontinalis* eDNA yielded 2,800,000 total copies in the bucket (*after* removing the 2-L preliminary sample), which were added to the upper Carmel River site. Each of the eight following collections yielded much lower eDNA quantities, as fish counts declined from the hatchery raceway due to removal and mortalities. Of those eight collections, one low outlier was the Las Garzas Creek trial, with 300 total copies added from the bucket. All others ranged from 75,000 to 530,000 total copies added from the bucket(s). The upper Carmel River and Las Garzas Creek trials had the highest and lowest ratios of starting quantity to discharge, respectively, largely due to their outlier amounts of eDNA added (Figure 2).

3.2 | eDNA Sampling Outcomes

We successfully recovered *S. fontinalis* eDNA (i.e., at least 1 qPCR technical replicate amplified) downstream at each of our

TABLE 2 | List of predictor variables used in RF modeling, categorically grouped.

Group	Predictor variable	Unit
Channel morphology	Depth (average)	cm
	Depth (maximum)	cm
	Slope of channel	—
	Wetted width (average)	m
Geology	Al_2O_3 content*	%
	Bedrock depth*	cm
	CaO content*	%
	Fe_2O_3 content*	%
	K_2O content*	%
	MgO content*	%
	N content*	%
	Na_2O content*	%
	P_2O_5 content*	%
	S content*	%
	SiO_2 content*	%
Hydrology	Baseflow index*	—
	Discharge	m^3/s
	Impervious cover*	%
	Run-off (water-year average)*	mm
	Water table depth (seasonal average)*	cm
Macrohabitat	Percent pool	%
	Percent riffle	%
Sampling design	Depth (average) of sampling cross-section	cm
	Depth (maximum) of sampling cross-section	cm
	Distance from eDNA addition point	m
	Number of qPCR technical replicates	—
	Starting quantity/discharge	copy #/(m^3/s)
	Wetted width of sampling cross-section	m
Soils	Clay content*	%
	Erodibility factor (Kf)*	—
	Organic matter content*	%
	Permeability*	cm/h
	Sand content*	%
Substrate	Biofilm thickness	g
	Median particle diameter (D50)	mm
	Percent course particulate organic matter (CPOM) cover	%
	Percent fine (< 2 mm diameter) particles	%
	Percent organic cover	%

(Continues)

TABLE 2 | (Continued)

Group	Predictor variable	Unit
Water quality	Conductivity	mS/cm
	Dissolved oxygen	mg/L
	Percent dissolved oxygen	%
	pH	—
	Salinity	—
	Temperature	°C
	Total dissolved solids (TDS)	g/L
	Total suspended sediment (TSS)	mg/L

Note: Asterisks indicate predictors obtained from the EPA StreamCat data set (Hill et al. 2016), for which both local catchment- and watershed-scale metrics were included. Depth (average), depth (maximum), wetted width (average), percent pool, percent riffle, D50, percent CPOM cover, percent fine particles, and percent organic cover were calculated for the portions of every reach between the eDNA addition point (0 m) and each sampling location (e.g., the average depth between 0–10, 0–50, 0–100 and 0–200 m were predictors for the PPTR at each corresponding sampling distance). All other variables had one value to represent the entire reach, so the same predictor value was used for each PPTR within a reach.

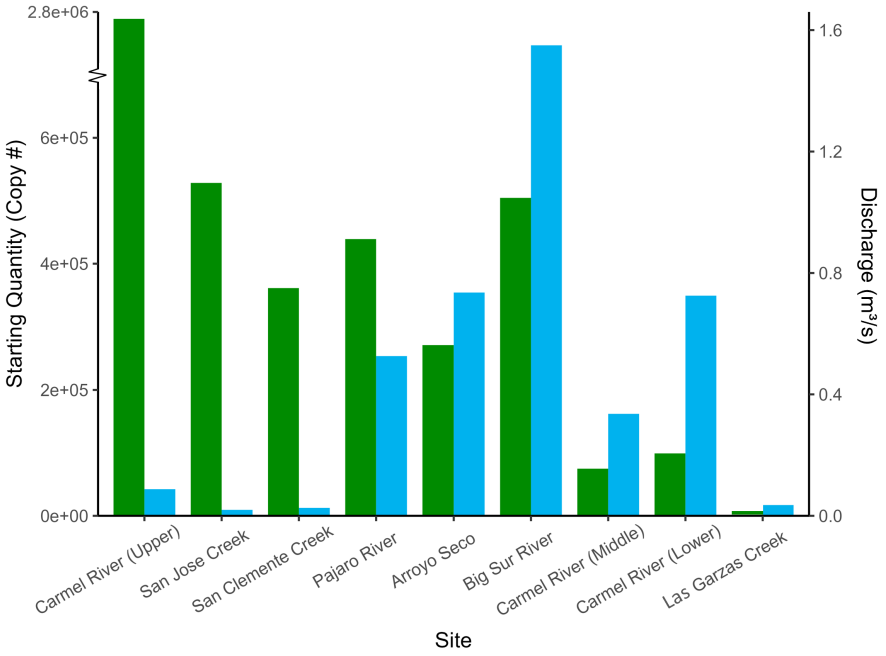

FIGURE 2 | Starting quantity of *Salvelinus fontinalis* eDNA added (copy #) and discharge (m^3/s) for each experimental reach. Green bars represent starting quantity, and blue bars represent discharge. Reaches are shown in decreasing order of starting quantity to discharge ratio, from left to right. The starting quantity axis (left) has a break to accommodate the upper Carmel River, which had a much greater starting quantity than any other site.

sites (Figure 3). Cycle threshold (C_t) values ranged from 33.71 to 43.11. The average proportion of positive technical replicates (PPTR) declined over distance from 0.71 at 10 m to 0.31 at 200 m, though highly varied PPTR values (from 0 to 1) occurred at each sampling distance (Figure 4). Aside from the upper Carmel River (which had a PPTR of 1 at each distance), seven of the eight remaining sites had their peak PPTR value at 10 or 50 m. Only 1 site (the Big Sur River) peaked at 100 m, and none peaked at 200 m (Figure 4). Our binomial generalized linear model ($n = 35$) revealed a significant decline in PPTR over distance ($\beta = -0.0061$, SE = 0.0024, $z = -2.475$, $p = 0.0133$) (Figure 4). The model's intercept was estimated at $\beta = 0.0643$ (SE = 0.2754, $z = 0.233$, $p = 0.8155$), and the model had a deviance of 143.0 and an AIC of 147.0, indicating a reasonable fit to the data.

3.3 | Occupancy Analysis

The estimated probability of occurrence of *S. fontinalis* eDNA in "A" water samples (collection beginning at the leading edge of the fluorescein dye) was moderately high (> 0.50) through 50 m and trended slightly down through 200 m, though credible intervals were wide (Figure 5). The occurrence probability in "B" water samples had posterior means consistently much lower than "A" but followed a very similar trend over distance with wide credible intervals. "C" water samples trended downward from 100 to 200 m in a similar fashion with wide credible intervals but had posterior means consistently less than "A" and greater than "B". The aggregate probability of occurrence of *S. fontinalis* eDNA (i.e., the probability of at least one positive

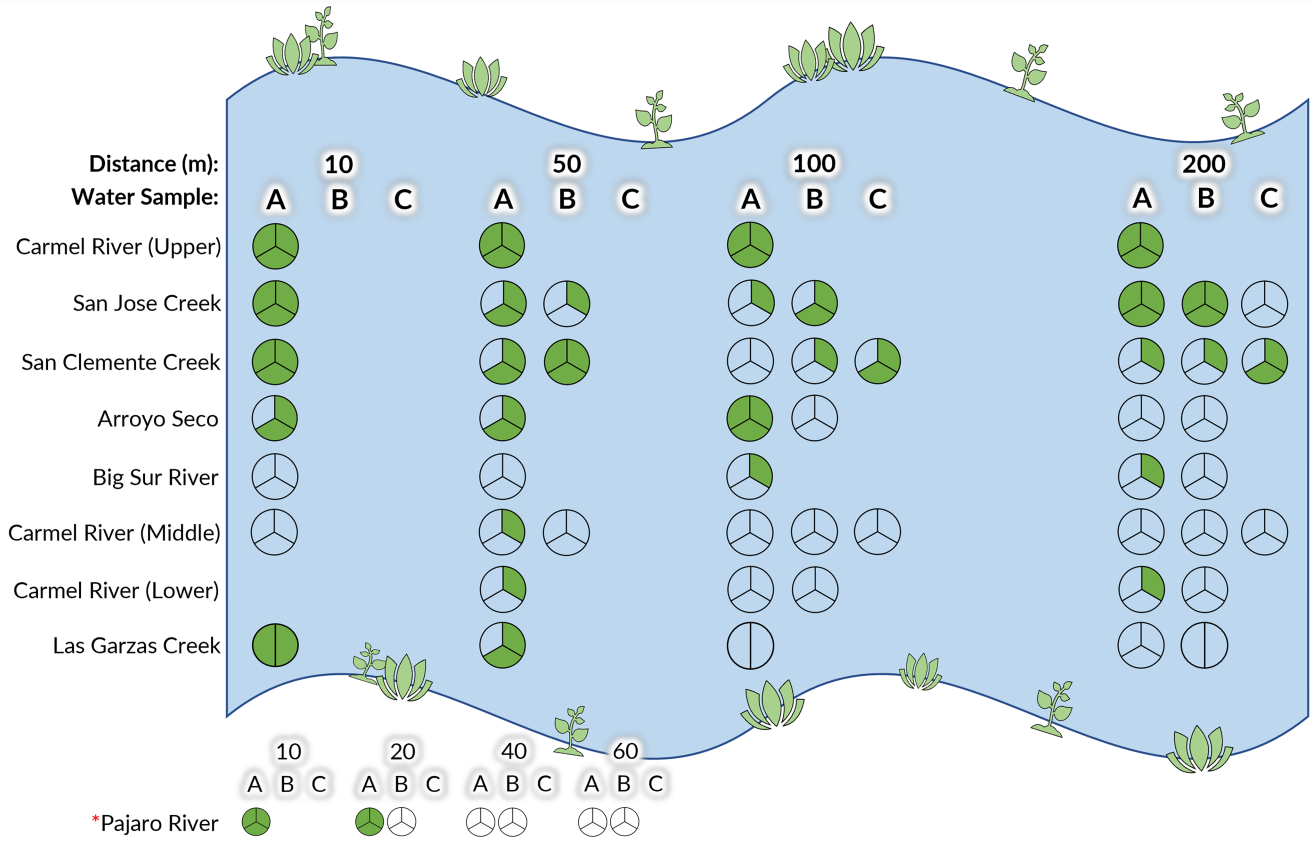

FIGURE 3 | *Salvelinus fontinalis* eDNA sampling outcomes for each water sample, at each distance, for every river site. Each circle represents a water sample. The number of divisions within circles represents the number of qPCR technical replicates run. Green shading indicates positive replicates. *Pajaro River, shown in the bottom left, had samples taken at different distances than every other reach.

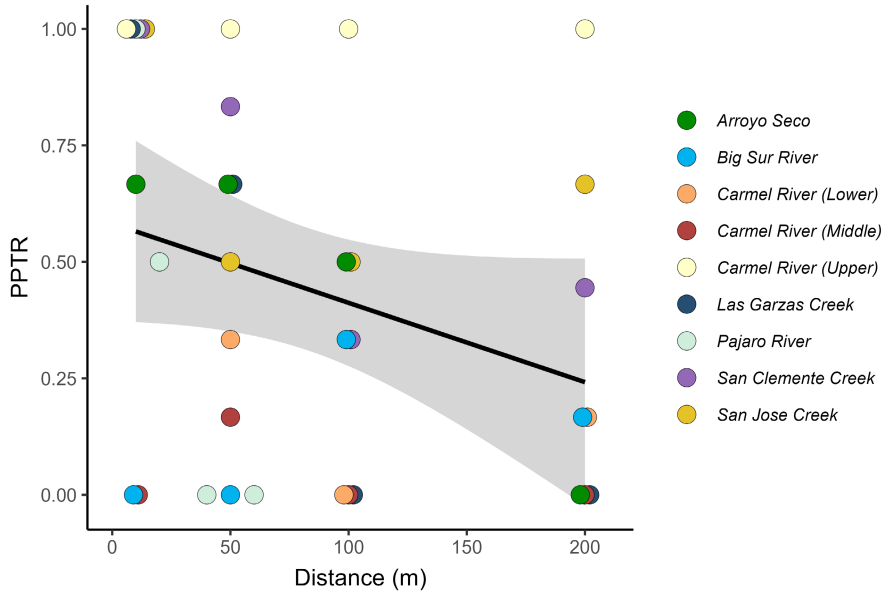

FIGURE 4 | Proportion of positive technical replicates (PPTR) for each sampling distance, for every river site. The black line represents our binomial regression model's predicted PPTR values over distance ($\beta = -0.0061$, $p = 0.0133$) and gray shading represents the 95% confidence interval.

qPCR technical replicate across all water samples at a given distance) had moderately high posterior means (> 0.50) regardless of distance. It did not trend downward, and perhaps trended slightly upward over distance, also with very wide credible intervals (Figure 5). This lack of a downward trend may indicate that the increased occurrence probability from additional water samples outweighs any decrease in occurrence probability associated with transport distance—at least through 200 m.

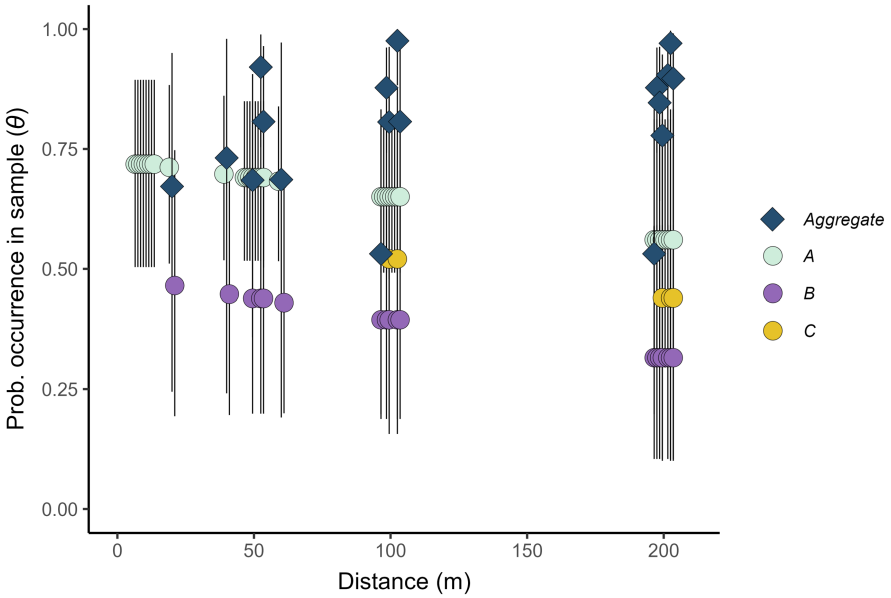

FIGURE 5 | Recovery probability of *Salvelinus fontinalis* eDNA over distance, for each sample at each river site. For distances at which multiple back-to-back water samples were required to capture the fluorescein plume, occurrence probability estimates from each water sample (θ) were aggregated into one recovery probability. Estimates are posterior means with 95% credible intervals. Distance values are jittered to aid in visualization. Actual distances include 10, 20, 40, 50, 60, 100, and 200 m.

3.4 | Random Forest Modeling

We modeled the proportion of positive technical replicates across all water samples at a sampling distance (PPTR), using environmental predictors from every experimental reach. Of 66 starting environmental predictors, five were eliminated in the first step (*VSURF* thresholding step) of our variable selection process (Appendix S1). Of the 61 remaining predictors, 54 were then eliminated in the second step (*VSURF* interpretation step), leaving the seven predictors most important to PPTR (Appendix S1). Of those seven, one predictor (average depth of the reach) was excluded due to high correlation with average depth of the sampling cross-section ($R^2 = 0.79$) and percent impervious cover of the watershed ($R^2 = 0.78$) (Appendix S2). This left six predictors for inclusion in our final model ($R^2 = 0.54$) (Figure 6). The starting quantity of eDNA normalized by discharge (variable importance, VI = 34.1) and the geologic CaO content of the catchment (VI = 32.4) had the highest importance, with strong positive and negative modeled relationships with PPTR, respectively (Figure 6). The next most important predictors were, in decreasing order, average depth at the sampling cross section (VI = 23.5), percent pool (VI = 17.8), percent impervious cover of the watershed (VI = 17.0), and the number of qPCR technical replicates (VI = 15.4), all with negative modeled relationships with PPTR (Figure 6).

4 | Discussion

This study uniquely contributes to the investigation of eDNA fate and sampling outcomes in lotic systems, given our novel experimental design in which the starting quantity of target eDNA released into each river was quantified, and the effect of a particular eDNA fate mechanism (resuspension) was likely minimized. Our results demonstrate a positive relationship between the quantity of eDNA we released and the proportion of positive qPCR technical replicates for our downstream samples (our proxy for the amount of recovered eDNA), supporting the sizeable body of evidence that a positive relationship exists between the biomass or density of target fish and the amount recovered downstream (Baldigo et al. 2016; Levi et al. 2019; Pochardt et al. 2020; and others, reviewed in Yao et al. 2022). While prior studies have done well to point out the strong dependence of lotic eDNA fate on local conditions (Goldberg et al. 2016; Spence et al. 2021; Wood et al. 2020), an understanding of fate that can be more broadly applied—at least across streams of similar order or within a geographic region—is an essential next step in the development of eDNA-based monitoring. With this step, useful predictions about a site's expected eDNA fate profile can be made, and eDNA sampling outcomes can be interpreted within their environmental context. By quantifying the strongest relationships between environmental factors and detections, our model allows us to characterize the mechanisms of reach-scale salmonid eDNA fate and identify potential proxy variables for the fate profile of a given sampling site.

4.1 | Effects of Dilution and Settling

Three of the most important predictors included in our final random forest (RF) model, eDNA starting quantity normalized by discharge (SQ/discharge), average depth of the sampling cross-section, and percent pool of the reach, implicate the effects of dilution and settling as key determinants of reach-scale sampling outcomes. To clarify terms, let us first stipulate that we are referring to dilution as the lowering of eDNA concentration *at the thalweg* (where we sampled) due to particles dispersing throughout the cross-sectional area and/or migrating toward the banks. In other contexts, dilution has been used to describe

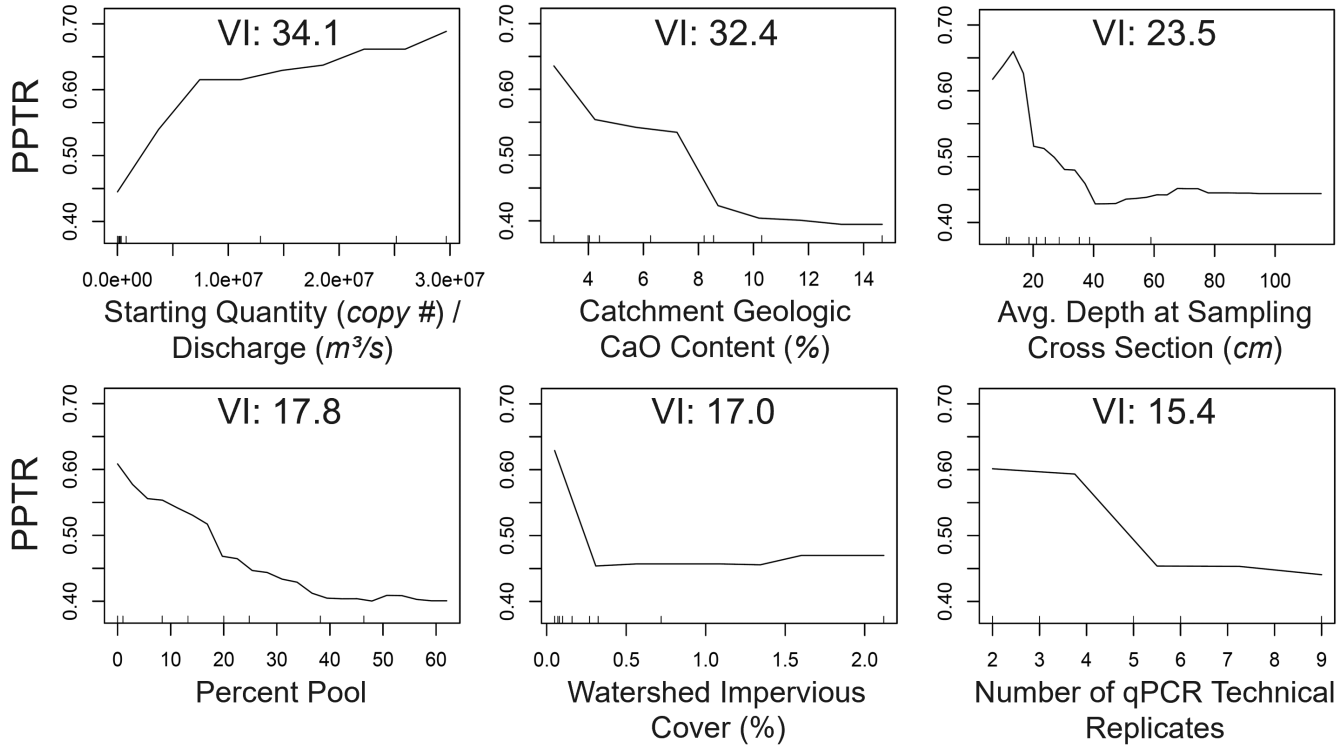

FIGURE 6 | Partial dependence plots (PDPs) of predictors selected for inclusion in our final random forest model ($R^2 = 0.54$), shown in decreasing order of variable importance (VI) (the percentage increase in mean squared error when a particular variable's values are randomly permuted). PDPs show how the proportion of positive qPCR technical replicates (PPTR), our proxy for the amount of eDNA recovered downstream, varies in response to individual predictors. The steeper the response curve, the more influential the variable is within that PPTR and variable range.

the lowering of total eDNA concentration (in the entire river) as the volume of water increases—either due to tributaries or increased groundwater contribution. Since we only sampled within 200 m of the source and selected reaches with no tributary confluences, we expected minimal dilution due to such increases in total water volume.

The most important predictor included in our final RF model was SQ/discharge, which had a strong positive modeled relationship with the proportion of positive qPCR technical replicates for our downstream samples (Figure 6). We normalized this predictor by discharge because we were unable to add a consistent amount of eDNA to each site since the amounts we collected from the hatchery were highly variable. In doing so we could account for this variation by converting the predictor from a term of absolute starting quantity to one approximating the starting *concentration*—a more relevant predictor for comparing sampling outcomes across rivers of different sizes. This predictor's top importance indicates that, unsurprisingly, the starting concentration of eDNA is a strong determinant of reach-scale sampling outcomes, even with minimized contribution of resuspended particles. Furthermore, its modeled relationship with the proportion of positive technical replicates (PPTR) implicates the previously well demonstrated effect of greater discharge reducing the amount of eDNA recovered downstream (Curtis et al. 2021; Jane et al. 2015; Levi et al. 2019; Thalinger et al. 2019; Van Driessche et al. 2023; Wood et al. 2021).

Dilution has been suggested to underlie this relationship between discharge and downstream sampling outcomes for two hypothesized reasons: because particles can ultimately become more dilute given a larger cross-sectional area, and because greater velocities drive particles to disperse throughout the entire cross-sectional area (or transition from midstream to the banks)—and thus become dilute more quickly (Jane et al. 2015; Pilliod et al. 2014; Pont et al. 2018; Wood et al. 2020; Van Driessche et al. 2023). However, as noted by Jane et al. (2015), it is important to consider that velocity is an imperfect proxy for the factors that truly drive dispersion (e.g., turbulence), and that it is possible to have a high velocity reach with relatively little dispersion because the flow is laminar.

eDNA particle fragmentation may be a third way that greater discharge tends to facilitate dilution. We suggest that the same turbulent action that hypothetically drives particles to disperse also subjects larger particles (e.g., clumps of cells) to mechanical fragmentation—creating numerous, smaller particles that themselves disperse more quickly and completely throughout the cross-sectional area. Thus, within a certain distance, the total amount of eDNA able to be recovered in a single water sample may tend to decrease with fragmentation, but the probability of recovering any eDNA—even a very small amount—may tend to increase. Van Driessche et al. (2023) made an observation consistent with this prediction, recovering lower concentrations of eDNA (from caged fish) at a high discharge site compared to a low discharge site, while both sites had similar estimated occurrence probabilities. Similarly, we observed a potential decoupling of PPTR and occurrence probability in our data, with an average

decline in PPTR over distance (Figure 4), but moderately high aggregate occurrence probabilities (regardless of distance) for all our sites (Figure 5). Though, our occurrence probabilities had wide credible intervals, and the average decline in PPTR over distance may, in part, be an artifact of our eDNA source and sampling methodology, as discussed below.

Additionally, if fragmentation is converting large particles into numerous small ones, such particles would be less likely to settle given their lighter weight (Choux, Druitt, and Thomas 2004; Snyder et al. 2023). Thus, fragmentation could help increase the distance over which eDNA persists in the water column. This aligns well with the fact—also pointed out by Van Driessche et al. (2023)—that higher discharge tends to be associated with greater eDNA detection distances (Jane et al. 2015; Thalinger et al. 2021; Van Driessche et al. 2023; Wilcox et al. 2016; Wood et al. 2021). It also suggests a tension between the potential diluting effect of fragmentation (driving down amounts recovered over shorter distances) and the potential preserving effect of fragmentation (allowing eDNA to remain suspended in the water column over longer distances).

The relationship between discharge and downstream sampling outcomes is further confounded by the expectation, also posited by Jane et al. (2015), that lower velocities tend to allow for increased settling of negatively buoyant eDNA particles. The fourth-most important predictor included in our final RF model, the percent pool of the reach, indicates (alongside findings by Sepulveda et al. 2021) this low-velocity-driven settling is an important phenomenon, since it demonstrates a clear negative relationship between pools (with their low velocities and role as transient storage zones) and PPTR (Figure 6).

The third-most-important predictor in our final RF model, average depth of the sampling cross-section, also had a negative modeled relationship with PPTR (Figure 6). We posit that this further speaks to the effect of dilution on downstream sampling outcomes, as the average depth of the sampling cross-section is likely serving as a proxy for the cross-sectional area—that is, there is simply more space for the particles to disperse into, so the amount one recovers from a thalweg water sample is liable to be more dilute. The fact of average depth of the reach being selected by the *VSURF* interpretation step, though it was ultimately excluded from the final RF model due to correlation with the average depth of the sampling cross-section ($R^2 = 0.79$) and percent impervious cover of the watershed ($R^2 = 0.78$) (Appendix S2), also likely speaks to this same effect.

To summarize, the set of hypotheses regarding eDNA fate we are outlining (based on our results in context with prior studies) is as follows:

- A greater cross-sectional area (for which greater discharge may serve as a useful proxy) allows for greater theoretical maximum dilution that can be achieved.
- Greater turbulence (for which greater velocity and/or discharge may serve as useful proxies) facilitates dispersion throughout the cross-sectional area and/or a transition of the plume from midstream-concentrated to bank-concentrated.
- Greater turbulence (for which greater velocity and/or discharge may serve as useful proxies) facilitates particle fragmentation, creating smaller particles that more readily disperse and are less likely to settle.
- Lower velocity (for which discharge may serve as a useful proxy) increases the amount of gravitational settling, with pools facilitating appreciable loss from the water column, accordingly.

4.2 | Effects of Sediment and Runoff

The second most important predictor included in our final RF model was the geologic calcium oxide (CaO) content of the local catchment, with a negative modeled relationship with the proportion of positive technical replicates (PPTR) (Figure 6). We initially hypothesized that CaO content is important to eDNA sampling outcomes because of its effects on water chemistry. Specifically, higher conductivity may drive chemical decay of eDNA and/or allow for greater microbial abundance and thus consumption, decreasing the recoverable amount downstream (Peixoto et al. 2023; Strickler, Fremier, and Goldberg 2015). However, we also expect CaO content may increase river pH, and eDNA is known to persist longer in neutral or slightly alkaline conditions, so these effects are difficult to parse (Lindahl 1993; Seymour et al. 2018; Strickler, Fremier, and Goldberg 2015). Moreover, we expect chemical decay and microbial consumption would have relatively minor impacts to sampling outcomes given the time and distance of our reach-scale experiment. Accordingly, the water chemistry implications of this predictor may, at best, only partially explain its importance.

A stronger hypothesis may be that CaO content, in addition to potentially driving chemical decay and microbial consumption, represents mechanical effects of higher overall sediment concentrations—especially given its correlation (> 0.70) with total suspended sediment (TSS) in our data set (Appendix S2). We posit that higher sediment concentrations would allow increased adsorption of eDNA particles, increasing the fraction of eDNA that is particle-bound versus free-floating. As Mauvisseau et al. (2022) points out, most extraction protocols from aquatic eDNA studies are unlikely to promote particle-bound eDNA to desorb. Thus, a higher fraction of particle-bound eDNA on a filter may lower the extraction efficiency, and thereby lower the qPCR copy number estimate. Furthermore, higher sediment concentrations caused some of our filters to clog while sampling, prompting us to swap in a new filter (e.g., replacing "A" with "B"), and potentially driving down the PPTR (as discussed below). Particle-bound eDNA may also be more likely to settle and become lost from the water column, given the weight of larger suspended particles.

Additionally, Mauvisseau et al. (2022) noted that the presence of divalent cations (principally Ca^{2+} and Mg^{2+}) in solution may allow for increased adsorption of eDNA to negatively charged sorbents via "cation bridging" (Anastassopoulou 2003; Minasov, Tereshko, and Egli 1999; Serra et al. 2002; Sheng et al. 2019). Magnesium oxide (MgO) content of the local catchment was strongly correlated with CaO content (> 0.98) in our data set, so perhaps CaO content represents adsorption-favorable conditions

in our final RF model. To the extent that this is true, we would expect higher adsorption to both suspended/dissolved sediments and the river substrate. This could account for a significant loss of eDNA over the reach-scale (in addition to the proposed filtering and extraction challenges described above) and thereby explain the high importance of this predictor—though further research is needed to understand the complex effects of calcium oxide content on eDNA fate. This highlights the critical importance of understanding how filter recovery and extraction efficiencies vary according to eDNA state (e.g., intracellular, dissolved, particle-bound) and accounting for that variation when interpreting results.

The fifth most important predictor in our final RF model was percent impervious cover in the watershed, with a negative modeled relationship with PPTR (Figure 6). Impervious cover refers to surfaces that do not allow water to infiltrate the ground, including roads, parking lots, sidewalks, and rooftops (Schueler 1994). Impervious cover is well understood to negatively impact stream hydrology, channel stability, habitat, water quality, and biological diversity (reviewed in Schueler, Fraley-McNeal, and Cappiella 2009). However, to our knowledge, it has not been previously implicated in questions of eDNA fate. We posit that nutrient runoff associated with impervious cover could lead to increased microbial consumption of eDNA, resulting in decreased recovery downstream. Additionally, the effect of chemical runoff associated with impervious cover may play an important role in sampling outcomes. While the effects of most common chemical pollutants on eDNA remain unknown (Hussain et al. 2009), certain pesticides (e.g., hexachlorocyclohexane) and insecticides (e.g., diazinon) have been demonstrated to facilitate degradation of eDNA—even at very low concentrations (Eichmiller, Bajer, and Sorensen 2014; Joseph et al. 2022; Pourmoghadam et al. 2019). However, as previously mentioned, we expect microbial consumption and chemical decay would only have minor effects on reach-scale eDNA recovery, and further research is required to understand how impervious affects the many abiotic and biotic drivers of eDNA fate. Nonetheless, the fact of its inclusion in our final RF model raises the interesting possibility that eDNA persistence in the water column may generally correlate with the ecological condition of the system.

4.3 | eDNA Pulse Dynamics

The last predictor included in our final RF model was the number of qPCR technical replicates run for each sampling distance. We included this predictor in our data set to indicate whether systematic error in our experimental design was at all responsible for the average decline in the proportion of positive technical replicates (PPTR) (our proxy for the amount of recovered eDNA) over distance. We recognized the possibility of this systematic error during pilot experiments, while testing our ability to capture an artificial pulse of eDNA (from a bucket) using fluorescein dye as a visual marker. We realized, because our eDNA/fluorescein pulses tended to spread out longitudinally as they flowed downstream, that our further sampling distances (100 and 200 m) often required more back-to-back water samples to capture the entirety of the pulse. Thus, even in a hypothetical scenario with no eDNA loss from the thalweg through 200 m, the spreading out of eDNA particles would dilute them into a larger volume of water, likely decreasing the chances of amplification in each qPCR technical replicate. Accordingly, the PPTR could be driven down simply because eDNA particles are spread out across multiple filters, rather than concentrated onto one. We considered this effect a systematic error—rather than a consequence of dilution we should be interested in measuring—because it only exists given the "discrete pulse" nature of our eDNA source. When sampling from an eDNA plume emanating from a live fish (i.e., the natural phenomenon we are trying to make inferences about), there is no pulse that needs to be captured at a specific moment. Rather, a continuous plume is produced over time (though release rates tend to be stochastic), offering a replenishing stream of eDNA particles with no reference point to observe longitudinal dispersion (Figure 7).

Since our variable selection process did identify the number of qPCR technical replicates as an important predictor, we must presume that part of our measured decline in PPTR over distance can be attributed to this systematic error, and not the effects of environmental factors. Our occupancy analysis, which estimated the occurrence probability (θ) by order ("A", "B", or "C") of each water sample, hinted at the specific nature of how our eDNA pulses tended to disperse longitudinally. While credible intervals were wide, "A" samples had consistently higher posterior mean θ estimates than "B" and "C" samples, possibly indicating a "teardrop shape" tendency where eDNA particles were more concentrated toward the leading edge of the fluorescein and more dilute toward the trailing edge (Figure 5). Despite this bias, because we accounted for its effect in our model, we remain confident that the other five selected predictors represent real underlying effects on reach-scale eDNA sampling outcomes. Our choice to spike rivers with eDNA from a bucket offered a trade-off in our experiment, in which we incurred the downside of this systematic error but gained the ability to measure eDNA starting quantity—a benefit that is not possible in caged fish experiments. Recently, devices have been developed that drip-release an eDNA slurry at a set rate (Herman 2023). This approach solves the "continuous release" and "unknown starting quantity" problems, though it likely generates particle-size/state distributions that differ from a natural release. We hope these considerations may be useful for future research efforts that are interested in measuring the starting quantity of lotic eDNA.

4.4 | Useful Proxies for eDNA Fate Profiles

The top environmental factors in our final RF model are informative of salmonid eDNA fate dynamics, and potentially the fate dynamics of other species with similar eDNA release profiles. However, the question remains: are any of these environmental factors useful proxy variables for the eDNA fate profile (i.e., how favorable local conditions are to eDNA recovery) of a given site? Before addressing this question, let us clarify that an optimal fate profile for presence/absence determinations may differ from an optimal fate profile for abundance/biomass estimates—per our prior discussion of dilution, fragmentation, adsorption to sediments, etc. Here, we will discuss potentially useful proxies for environments favorable to quantitative sampling efforts.

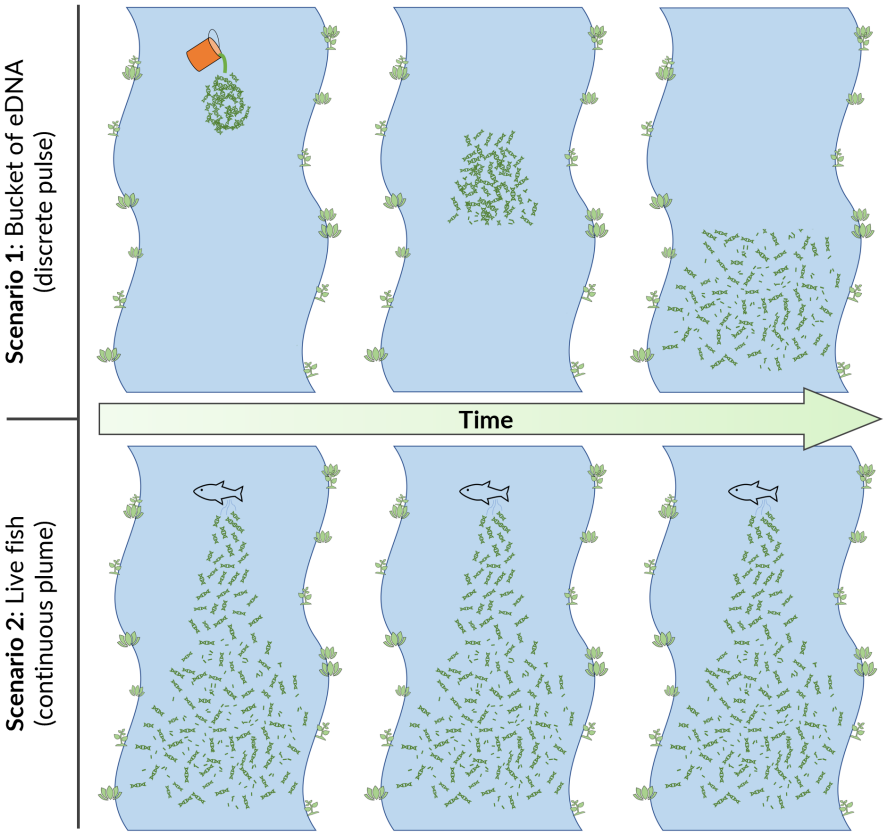

FIGURE 7 | Illustration comparing two eDNA transport scenarios over time. (1) The "discrete pulse" scenario, representing our experimental approach of adding a known quantity of eDNA particles from a bucket. (2) The "continuous plume" scenario, representing the natural process of eDNA release from a live fish.

We suggest discharge as a useful top-line metric for identifying optimal regions and seasons for increased eDNA recovery, particularly given its availability from stream-gaging networks. While higher discharge has been previously demonstrated as a useful proxy for eDNA detection distance (Thalinger et al. 2019, 2021; Van Driessche et al. 2023; Wilcox et al. 2016), our results add to growing evidence that lower discharge is a relevant proxy for relative quantity of eDNA recovered at the reach-scale (Van Driessche et al. 2023). Thus, all else being equal, researchers can expect more eDNA to be recovered in smaller streams, higher in watersheds. Additionally, wet seasons subject to higher discharge or more frequent hydrograph peaks are likely less optimal times for quantitative sampling.

For identifying optimal reaches within a catchment, our results suggest avoiding pools—especially deep, slow-moving pools—as they can be expected to facilitate settling and decrease recoverable amounts downstream. This is a useful proxy since it only requires visual identification of pools, and no time-intensive physical characterization of reaches. As an aside, if low-velocity-driven settling in pools facilitates significant loss from the water column, it follows that pool benthos may house a rich deposit of sedimentary eDNA that could provide a wealth of ecological information.

For selecting sampling sites within a reach, our results indicate shallower depths could represent optimal locations to collect a water sample. However, we have lower confidence this would be a reliable proxy since the potential negative effects of turbulence on recoverable eDNA quantity may often be counteracting (e.g., riffles). Furthermore, we expect depth was important to our sampling outcomes primarily because we collected water samples from the thalweg, near the top of the water column—an approach that ought to be improved upon in future studies. Calls for sampling eDNA across the wetted width of a river channel have emerged as the importance of dilution and dispersion to sampling outcomes has been revealed (Van Driessche et al. 2023; Wood et al. 2021). We echo this suggestion and add that sampling up and down the water column vertically—though it presents a logistical challenge—would likely have a positive impact on the amount of eDNA recovered as well. In addition, we suggest using dye (e.g., fluorescein) to visually identify the thalweg and degree of dispersion when deciding where to collect a water sample.

Regarding the other predictors from our final RF model, calcium oxide content of the catchment raises the interesting possibilities of low sediment levels and/or low cation concentrations being useful proxies for favorable fate profiles. However, RF models cannot extrapolate to novel environments, and while we selected a range of sites to make our model as general as possible, it remains unknown whether these factors are predictive of higher eDNA recovery across watersheds with different sediment and geology types (especially those outside the central California coast). Similarly, impervious cover of the watershed may implicate underappreciated

factors to eDNA fate, like runoff-associated degradation, but too little is known to expect impervious cover could reliably predict fate profiles.

5 | Conclusion

In conclusion, our novel experimental design and random forest modeling approach identified five river environmental factors that are important to reach-scale salmonid eDNA sampling outcomes in Mediterranean-climate streams (Figure 6). Those five factors and their modeled relationships with the proportion of positive qPCR technical replicates for our downstream samples (our proxy for amount of recovered eDNA) allowed us to characterize the mechanisms of salmonid eDNA fate and their relative impacts. Notably, we suggest that the dispersion, fragmentation, and corresponding dilution of eDNA particles may often be a stronger determinant of amount recovered at the reach-scale than any mechanism of loss from the water column—particularly when sampling from a single cross-sectional point (which is not recommended). We highlight cross-sectional area and turbulence as the direct drivers of this effect, and thus important factors to consider (despite typically being represented via proxy by discharge or velocity). We identify catchment geology (specifically, calcium oxide content) and its consequences for water chemistry and eDNA adsorption to sediments as a potentially underappreciated factor in eDNA fate. Plus, we suggest discharge and the presence of pools as useful proxies for evaluating a river site's favorability for quantitative eDNA sampling. While the collective understanding of fish eDNA fate dynamics supports increasingly reliable models of downstream detection distance (e.g., Pont 2024), abundance/biomass monitoring from eDNA alone remains impractical. We hope our findings will contribute towards quantitative sampling efforts, and call attention to important drivers of eDNA fate that may have been previously overlooked.

Author Contributions

S.A.M. and J.R.O. have both contributed to (1) the conception/design of the study, (2) the acquisition, analysis, and interpretation of the data, and (3) writing the manuscript.

Acknowledgments

Thank you to Dr. Susanna Theroux and Dr. Elizabeth Alter for your guidance and feedback on this project. Thank you to Dr. Nathaniel Jue, Dr. J.P. Dundore-Arias, and Dr. Doug Smith for allowing access to your labs and instruments. To our field crew members—thank you Ellery Charleton, Lexi Yokomizo, Kathy Ucello, Hannah Kim, Charley Abernathy, Cindy Matuch, Connor Quiroz, Alyssa Schaer, Kayla Fon, Lizbeth Amador, Natalie Hunter, Daniel Lopez, Rachel Aitchison, Isaac Nicholes, and Duncan Miller. Thank you to Dr. Brian Spence and Dave Rundio for your mentorship related to this project. Thank you to Dr. Steven Kim for your insight regarding data analysis. Thank you to William Branch with CDFW, the Santa Lucia Conservancy, and the Big Sur Land Trust for hatchery and river access. Finally, thank you to NOAA CCME for providing funding for this project.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The raw data and R script regarding the main results of this study are available on Zenodo at https://doi.org/10.5281/zenodo.13362717.

References

Anastassopoulou, J. 2003. "Metal–DNA Interactions." *Journal of Molecular Structure* 651–653: 19–26.

Andruszkiewicz, E. A., J. R. Koseff, O. B. Fringer, et al. 2019. "Modeling Environmental DNA Transport in the Coastal Ocean Using Lagrangian Particle Tracking." *Frontiers in Marine Science* 6: 477.

Andruszkiewicz, E. A., L. M. Sassoubre, and A. B. Boehm. 2017. "Persistence of Marine Fish Environmental DNA and the Influence of Sunlight." *PLoS One* 12, no. 9: e0185043.

Baldigo, B. P., L. A. Sporn, S. D. George, and J. A. Ball. 2016. "Efficacy of Environmental DNA to Detect and Quantify Brook Trout Populations in Headwater Streams of the Adirondack Mountains, New York." *Transactions of the American Fisheries Society* 146, no. 1: 99–111.

Barnes, M. A., C. R. Turner, C. L. Jerde, M. A. Renshaw, W. L. Chadderton, and D. M. Lodge. 2014. "Environmental Conditions Influence eDNA Persistence in Aquatic Systems." *Environmental Science and Technology* 48, no. 3: 1819–1827.

Bedwell, M. E., and C. S. Goldberg. 2020. "Spatial and Temporal Patterns of Environmental DNA Detection to Inform Sampling Protocols in Lentic and Lotic Systems." *Ecology and Evolution* 10, no. 3: 1602–1612.

Brandão-Dias, P. F. P., D. M. C. Hallack, E. D. Snyder, et al. 2023. "Particle Size Influences Decay Rates of Environmental DNA in Aquatic Systems." *Molecular Ecology Resources* 23: 756–770.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Brooks, M. E., K. Kristensen, K. J. van Benthem, et al. 2017. "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *R Journal* 9, no. 2: 378–400.

California Department of Fish and Wildlife [CDFW], Biogeographic Data Branch. 2021. *California Wildlife Habitat Relationship System, Version 10.x.x*. Sacramento, CA: CDFW.

Carim, K. J., J. C. S. Dysthe, M. K. Young, K. S. McKelvey, and M. K. Schwartz. 2016. "An Environmental DNA Assay for Detecting Arctic Grayling in the Upper Missouri River Basin, North America." *Conservation Genetics Resources* 8: 197–199.

Carim, K. J., K. S. McKelvey, M. K. Young, T. M. Wilcox, and M. K. Schwartz. 2016. "A Protocol for Collecting Environmental DNA Samples From Streams." General Technical Report RMRS-GTR-355 (United States Department of Agriculture [USDA], 1–18).

Choux, C., T. Druitt, and N. Thomas. 2004. "Stratification and Particle Segregation in Flowing Polydisperse Suspensions, With Applications to the Transport and Sedimentation of Pyroclastic Density Currents." *Journal of Volcanology and Geothermal Research* 138, no. 3–4: 223–241.

Civade, R., T. Dejean, A. Valentini, et al. 2016. "Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System." *PLoS One* 11, no. 6: e0157366.

Crozier, L. G., M. M. McClure, T. Beechie, et al. 2019. "Climate Vulnerability Assessment for Pacific Salmon and Steelhead in the California Current Large Marine Ecosystem." *PLoS One* 14, no. 7: e0217711.

Curtis, A. N., J. S. Tiemann, S. A. Douglass, M. A. Davis, and E. R. Larson. 2021. "High Stream Flows Dilute Environmental DNA (eDNA) Concentrations and Reduce Detectability." *Diversity and Distributions* 27, no. 10: 1918–1931.

Deiner, K., E. A. Fronhofer, E. Mächler, J. C. Walser, and F. Altermatt. 2016. "Environmental DNA Reveals That Rivers Are Conveyer Belts of Biodiversity Information." *Nature Communications* 7: 12544.

Doi, H., R. Inui, Y. Akamatsu, et al. 2017. "Environmental DNA Analysis for Estimating the Abundance and Biomass of Stream Fish." *Freshwater Biology* 62, no. 1: 30–39.

Dorazio, R. M., and R. A. Erickson. 2018. "*eDNAoccupancy*: An R Package for Multiscale Occupancy Modeling of Environmental DNA Data." *Molecular Ecology Resources* 18, no. 2: 368–380.

Dudgeon, D. 2019. "Multiple Threats Imperil Freshwater Biodiversity in the Anthropocene." *Current Biology* 29, no. 19: 960–967.

Eichmiller, J. J., P. G. Bajer, and P. W. Sorensen. 2014. "The Relationship Between the Distribution of Common Carp and Their Environmental DNA in a Small Lake." *PLoS One* 9, no. 11: e112611.

Genuer, R., J.-M. Poggi, C. Tuleau-Malot, and C. Tuleau. 2015. "VSURF: An R Package for Variable Selection Using Random Forests." *R Journal* 7, no. 2: 19–33.

Goldberg, C. S., C. R. Turner, K. Deiner, et al. 2016. "Critical Considerations for the Application of Environmental DNA Methods to Detect Aquatic Species." *Methods in Ecology and Evolution* 7: 1299–1307.

Herman, B. A. 2023. "Use of Foreign eDNA Tracers to Resolve Site- and Time-Specific eDNA Distributions in Natural Streams." *Cal Poly Humboldt Theses and Projects*. 629.

Hill, R. A., M. H. Weber, S. G. Leibowitz, A. R. Olsen, and D. J. Thornbrugh. 2016. "The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States." *Journal of the American Water Resources Association* 52, no. 1: 120–128.

Hulley, E. N., S. Tharmalingam, A. Zarnke, and D. R. Boreham. 2019. "Development and Validation of Probe-Based Multiplex Real-Time PCR Assays for the Rapid and Accurate Detection of Freshwater Fish Species." *PLoS One* 14, no. 1: e0210165.

Hussain, S., T. Siddique, M. Saleem, M. Arshad, and A. Khalid. 2009. "Chapter 5. Impact of Pesticides on Soil Microbial Diversity, Enzymes, and Biochemical Reactions." *Advances in Agronomy* 102: 159–200.

Jane, S. F., T. M. Wilcox, K. S. McKelvey, et al. 2015. "Distance, Flow and PCR Inhibition: eDNA Dynamics in Two Headwater Streams." *Molecular Ecology Resources* 15, no. 1: 216–227.

Jerde, C. L., B. P. Olds, A. J. Shogren, et al. 2016. "Influence of Stream Bottom Substrate on the Retention and Transport of Vertebrate Environmental DNA." *Environmental Science & Technology* 50: 8770–8779.

Jo, T., H. Murakami, S. Yamamoto, R. Masuda, and T. Minamoto. 2019. "Effect of Water Temperature and Fish Biomass on Environmental DNA Shedding, Degradation, and Size Distribution." *Ecology and Evolution* 9, no. 3: 1135–1146.

Joseph, C., M. E. Faiq, Z. Li, and G. Chen. 2022. "Persistence and Degradation Dynamics of eDNA Affected by Environmental Factors in Aquatic Ecosystems." *Hydrobiologia* 849, no. 19: 4119–4133.

Klymus, K. E., C. M. Merkes, M. J. Allison, et al. 2020. "Reporting the Limits of Detection and Quantification for Environmental DNA Assays." *Environmental DNA* 2, no. 3: 271–282.

Klymus, K. E., C. A. Richter, D. C. Chapman, and C. Paukert. 2015. "Quantification of eDNA Shedding Rates From Invasive Bighead Carp *Hypophthalmichthys nobilis* and Silver Carp *Hypophthalmichthys molitrix*." *Biological Conservation* 183: 77–84.

Lacoursière-Roussel, A., G. Côté, V. Leclerc, and L. Bernatchez. 2016. "Quantifying Relative Fish Abundance With eDNA: A Promising Tool for Fisheries Management." *Journal of Applied Ecology* 53, no. 4: 1148–1157.

Laporte, M., B. Bougas, G. Côté, et al. 2020. "Caged Fish Experiment and Hydrodynamic Bidimensional Modeling Highlight the Importance to Consider 2D Dispersion in Fluvial Environmental DNA Studies." *Environmental DNA* 2, no. 3: 362–372.

Levi, T., J. M. Allen, D. Bell, et al. 2019. "Environmental DNA for the Enumeration and Management of Pacific Salmon." *Molecular Ecology Resources* 19, no. 3: 597–608.

Lindahl, T. 1993. "Instability and Decay of the Primary Structure of DNA." *Nature* 362, no. 6422: 709–715.

Lodge, D. M., C. R. Turner, C. L. Jerde, et al. 2012. "Conservation in a Cup of Water: Estimating Biodiversity and Population Abundance From Environmental DNA." *Molecular Ecology* 21, no. 11: 2555–2558.

Maruyama, A., K. Nakamura, H. Yamanaka, M. Kondoh, and T. Minamoto. 2014. "The Release Rate of Environmental DNA From Juvenile and Adult Fish." *PLoS One* 9, no. 12: e114639.

Mauvisseau, Q., L. R. Harper, M. Sander, R. H. Hanner, H. Kleyer, and K. Deiner. 2022. "The Multiple States of Environmental DNA and What Is Known About Their Persistence in Aquatic Environments." *Environmental Science & Technology* 56, no. 9: 5322–5333.

Merkes, C., K. Klymus, M. Allison, et al. 2019. "Reporting the Limits of Detection (LOD) and Quantification (LOQ) for Environmental DNA Assays." U.S. Geological Survey Data Release.

Minasov, G., V. Tereshko, and M. Egli. 1999. "Atomic-Resolution Crystal Structures of B-DNA Reveal Specific Influences of Divalent Metal Ions on Conformation and Packing." *Journal of Molecular Biology* 291, no. 1: 83–99.

Mullan, J. W. 1987. *Status and Propagation of Chinook Salmon in the Mid-Columbia River Through 1985*. Washington, DC: U.S. Dept. of the Interior Fish and Wildlife Service Research and Development.

Nehlsen, W., J. E. Williams, and J. A. Lichatowich. 1991. "Pacific Salmon at the Crossroads: Stocks at Risk From California, Oregon, Idaho, and Washington." *Fisheries* 16, no. 2: 4–19.

Ogburn, M. B., L. V. Plough, C. W. Bangley, et al. 2022. "Environmental DNA Reveals Anadromous River Herring Habitat Use and Recolonization After Restoration of Aquatic Connectivity." *Environmental DNA* 5: 25–37.

Peixoto, S., M. Mota-Ferreira, C. Chaves, G. Velo-Antón, P. Beja, and B. Egeter. 2023. "Multi-Species Occupancy Modeling Reveals Methodological and Environmental Effects on eDNA Detection of Amphibians in Temporary Ponds." *Environmental DNA* 5: 796–811.

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, no. 8: 1123–1130.

Pilliod, D. S., C. S. Goldberg, R. S. Arkle, and L. P. Waits. 2014. "Factors Influencing Detection of eDNA From a Stream-Dwelling Amphibian." *Molecular Ecology Resources* 14, no. 1: 109–116.

Pochardt, M., J. M. Allen, T. Hart, S. D. L. Miller, D. W. Yu, and T. Levi. 2020. "Environmental DNA Facilitates Accurate, Inexpensive, and Multiyear Population Estimates of Millions of Anadromous Fish." *Molecular Ecology Resources* 20, no. 2: 457–467.

Pont, D. 2024. "Predicting Downstream Transport Distance of Fish eDNA in Lotic Environments." *Molecular Ecology Resources* 24: e13934.

Pont, D., M. Rocle, A. Valentini, et al. 2018. "Environmental DNA Reveals Quantitative Patterns of Fish Biodiversity in Large Rivers Despite Its Downstream Transportation." *Scientific Reports* 8, no. 1: 10361.

Pourmoghadam, M. N., H. Poorbagher, J. M. de Oliveira Fernandes, and O. Jafari. 2019. "Diazinon Negatively Affects the Integrity of Environmental DNA Stability: A Case Study With Common Carp (*Cyprinus carpio*)." *Environmental Monitoring and Assessment* 191: 672.

Reid, A. J., A. K. Carlson, I. F. Creed, et al. 2019. "Emerging Threats and Persistent Conservation Challenges for Freshwater Biodiversity." *Biological Reviews of the Cambridge Philosophical Society* 94, no. 3: 849–873.

Rourke, M. L., A. M. Fowler, J. M. Hughes, et al. 2021. "Environmental DNA (eDNA) as a Tool for Assessing Fish Biomass: A Review of Approaches and Future Considerations for Resource Surveys." *Environmental DNA* 4: 9–33.

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 and Technology* 50, no. 19: 10456–10464.

Schueler, T. 1994. "The Importance of Imperviousness." *Watershed Protection Techniques* 1, no. 3: 100–111.

Schueler, T. R., L. Fraley-McNeal, and K. Cappiella. 2009. "Is Impervious Cover Still Important? Review of Recent Research." *Journal of Hydrologic Engineering* 14, no. 4: 309–315.

Sepulveda, A. J., R. Al-Chokhachy, M. B. Laramie, et al. 2021. "It's Complicated…Environmental DNA as a Predictor of Trout and Char Abundance in Streams." *Canadian Journal of Fisheries and Aquatic Sciences* 78, no. 4: 422–432.

Serra, M. J., J. D. Baird, T. Dale, B. L. Fey, K. Retatagos, and E. Westhof. 2002. "Effects of Magnesium Ions on the Stabilization of RNA Oligomers of Defined Structures." *RNA* 8, no. 3: 307–323.

Seymour, M., I. Durance, B. J. Cosby, et al. 2018. "Acidity Promotes Degradation of Multi-Species Environmental DNA in Lotic Mesocosms." *Communications Biology* 1, no. 1: 4.

Sheng, X., C. Qin, B. Yang, et al. 2019. "Metal Cation Saturation on Montmorillonites Facilitates the Adsorption of DNA via Cation Bridging." *Chemosphere* 235: 670–678.

Shogren, A. J., J. L. Tank, E. Andruszkiewicz, et al. 2017. "Controls on eDNA Movement in Streams: Transport, Retention, and Resuspension." *Scientific Reports* 7, no. 1: 5065.

Snyder, E. D., J. L. Tank, K. Bibby, et al. 2021. "Exploring the Role of Biofilm Colonization on the Transport and Fate of Environmental DNA (eDNA) [Video]." YouTube. https://www.youtube.com/watch?v=P1JPltlGNFg.

Snyder, E. D., J. L. Tank, P. F. Brandão-Dias, et al. 2023. "Environmental DNA (eDNA) Removal Rates in Streams Differ by Particle Size Under Varying Substrate and Light Conditions." *Science of the Total Environment* 903: 166469.

Speiser, J. L., M. E. Miller, J. Tooze, and E. Ip. 2019. "A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling." *Expert Systems With Applications* 134: 93–101.

Spence, B. C., D. E. Rundio, N. J. Demetras, and M. Sedoryk. 2021. "Efficacy of Environmental DNA Sampling to Detect the Occurrence of Endangered Coho Salmon (*Oncorhynchus kisutch*) in Mediterranean-Climate Streams of California's Central Coast." *Environmental DNA* 3, no. 4: 727–744.

Steinman, A. D., G. A. Lamberti, P. R. Leavitt, and D. G. Uzarski. 2017. "Biomass and Pigments of Benthic Algae." In *Methods in Stream Ecology*, 3rd ed., edited by F. R. Hauer and G. A. Lamberti, 223–241. Cambridge, MA: Academic Press Elsevier.

Strickler, K. M., A. K. Fremier, and C. S. Goldberg. 2015. "Quantifying Effects of UV-B, Temperature, and pH on eDNA Degradation in Aquatic Microcosms." *Biological Conservation* 183: 85–92.

Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. "Conditional Variable Importance for Random Forests." *BMC Bioinformatics* 9: 307.

Taberlet, P., E. Coissac, M. Hajibabaei, and L. H. Rieseberg. 2012. "Environmental DNA." *Molecular Ecology* 21, no. 8: 1789–1793.

Thalinger, B., K. Deiner, L. R. Harper, et al. 2021. "A Validation Scale to Determine the Readiness of Environmental DNA Assays for Routine Species Monitoring." *Environmental DNA* 3, no. 4: 823–836.

Thalinger, B., E. Wolf, M. Traugott, and J. Wanzenböck. 2019. "Monitoring Spawning Migrations of Potamodromous Fish Species via eDNA." *Scientific Reports* 9, no. 1: 15388.

Tillotson, M. D., R. P. Kelly, J. J. Duda, M. Hoy, J. Kralj, and T. P. Quinn. 2018. "Concentrations of Environmental DNA (eDNA) Reflect Spawning Salmon Abundance at Fine Spatial and Temporal Scales." *Biological Conservation* 220: 1–11.

Valentini, A., P. Taberlet, C. Miaud, et al. 2016. "Next-Generation Monitoring of Aquatic Biodiversity Using Environmental DNA Metabarcoding." *Molecular Ecology* 25, no. 4: 929–942.

Van Driessche, C., T. Everts, S. Neyrinck, and R. Brys. 2023. "Experimental Assessment of Downstream Environmental DNA Patterns Under Variable Fish Biomass and River Discharge Rates." *Environmental DNA* 5: 102–116.

Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, et al. 2010. "Global Threats to Human Water Security and River Biodiversity." *Nature* 467, no. 7315: 555–561.

Wilcox, T. M., K. S. McKelvey, M. K. Young, et al. 2013. "Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity." *PLoS One* 8, no. 3: e59520.

Wilcox, T. M., K. S. McKelvey, M. K. Young, W. H. Lowe, and M. K. Schwartz. 2015. "Environmental DNA Particle Size Distribution From Brook Trout (*Salvelinus fontinalis*)." *Conservation Genetics Resources* 7, no. 3: 639–641.

Wilcox, T. M., K. S. McKelvey, M. K. Young, et al. 2016. "Understanding Environmental DNA Detection Probabilities: A Case Study Using a Stream-Dwelling Char *Salvelinus fontinalis*." *Biological Conservation* 194: 209–216.

Wittwer, C., S. Stoll, D. Strand, T. Vrålstad, C. Nowak, and M. Thines. 2018. "eDNA-Based Crayfish Plague Monitoring Is Superior to Conventional Trap-Based Assessments in Year-Round Detection Probability." *Hydrobiologia* 807, no. 1: 87–97.

Wood, Z. T., B. F. Erdman, G. York, J. G. Trial, and M. T. Kinnison. 2020. "Experimental Assessment of Optimal Lotic eDNA Sampling and Assay Multiplexing for a Critically Endangered Fish." *Environmental DNA* 2, no. 4: 407–417.

Wood, Z. T., A. Lacoursière-Roussel, F. LeBlanc, et al. 2021. "Spatial Heterogeneity of eDNA Transport Improves Stream Assessment of Threatened Salmon Presence, Abundance, and Location." *Frontiers in Ecology and Evolution* 9: 168.

Yao, M., S. Zhang, Q. Lu, et al. 2022. "Fishing for Fish Environmental DNA: Ecological Applications, Methodological Considerations, Surveying Designs, and Ways Forward." *Molecular Ecology* 31, no. 20: 5132–5164.

Yates, M. C., D. M. Glaser, J. R. Post, M. E. Cristescu, D. J. Fraser, and A. M. Derry. 2021. "The Relationship Between eDNA Particle Concentration and Organism Abundance in Nature Is Strengthened by Allometric Scaling." *Molecular Ecology* 30, no. 13: 3068–3082.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.