

¹ **Environmental DNA provides quantitative estimates of abundance and distribution in the open ocean.**

³ Andrew Olaf Shelton^{1*}, Ana Ramón-Laca², Abigail Wells³, Julia Clemons⁴, Dezhang Chu⁴, Blake E. Feist¹,
⁴ Ryan P. Kelly⁵, Sandra L. Parker-Stetter^{4,6}, Rebecca Thomas⁴, Krista Nichols¹, Linda Park¹

⁵ ¹Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service,
⁶ National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112, U.S.A.

⁷ ²Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington at Northwest
⁸ Fisheries Science Center, National Marine Fisheries Service, Seattle, WA 98112, U.S.A.

⁹ ³Lynker Technologies, Under contract to Northwest Fisheries Science Center, National Marine Fisheries
¹⁰ Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112,
¹¹ U.S.A.

¹² ⁴Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine
¹³ Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA
¹⁴ 98112, U.S.A.

¹⁵ ⁵University of Washington, School of Marine and Environmental Affairs, 3707 Brooklyn Ave NE, Seattle, WA
¹⁶ 98105, U.S.A.

¹⁷ ⁶Resource Assessment & Conservation Engineering Division, Alaska Fisheries Science Center, National Marine
¹⁸ Fisheries Service, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA
¹⁹ 98115, WA

²⁰ *corresponding author: ole.shelton@noaa.gov

²¹ *Keywords:* environmental DNA, species distributions, fisheries, ocean surveys

²² **Abstract**

²³ All species inevitably leave genetic traces in their environments, and the resulting environmental DNA
²⁴ (eDNA) reflects the species present in a given habitat. It remains unclear whether eDNA signals can provide
²⁵ quantitative metrics of abundance on which human livelihoods or conservation successes depend. Here, we
²⁶ report the results of a large eDNA ocean survey (spanning 86,000 km² to depths of 500m) to understand the
²⁷ abundance and distribution of Pacific hake (*Merluccius productus*), the target of the largest finfish fishery
²⁸ along the west coast of the United States. We sampled eDNA in parallel with traditional survey methods
²⁹ and show how eDNA provides a spatially smooth signature of hake relative to traditional acoustic surveys.
³⁰ Despite local differences, at management-relevant scales the two methods yield comparable information
³¹ about the broad-scale spatial distribution and abundance of hake. Furthermore, we find depth and spatial
³² patterns of eDNA closely correspond to acoustic estimates for hake. We demonstrate the power and efficacy
³³ of eDNA sampling for estimation of abundance and distribution and move the analysis eDNA data beyond
³⁴ sample-to-sample comparisons to management relevant scales. We posit that eDNA methods are capable of
³⁵ providing general quantitative applications that will prove especially valuable in data- or resource-limited
³⁶ contexts.

37 Introduction

38 Environmental DNA, the DNA from target organisms collected from an environmental medium (e.g. soil
39 or water), can reflect species in a wide range of terrestrial, aquatic, and marine habitats [1]. eDNA has
40 the potential to revolutionize our understanding of natural communities by enabling rapid and accurate
41 surveys of many species simultaneously without capturing individuals [1]. At present, eDNA can efficiently
42 survey species diversity and changes in community membership [2–4]. However, many natural resource
43 questions depend upon quantitative estimates of abundance (e.g. fisheries or managing species of conservation
44 concern) and for these topics, eDNA must provide such information in order to be most useful [5]. While
45 most studies find a positive relationship between eDNA concentrations and other survey methods [reviewed
46 by 6], uncertainty about the strength of the eDNA-abundance relationship due to the complexity of eDNA
47 generation, transport, degradation, and detection have limited the application of eDNA in many quantitative
48 applications [5, 7, 8]. While the use of eDNA methods has grown exponentially from tens of publications in
49 2010 to over 600 in 2020 [6, 9], reflecting widespread adoption of eDNA technologies, basic questions about the
50 characteristics of eDNA limit its practical application and slow its adoption in environmental management.

51 Rigorous, well designed surveys underlie the successful management and conservation of wild populations.
52 But field surveys are expensive – for example, open-ocean surveys involve ship time costing tens of thousands
53 of dollars per day – and are typically tailored to one or a few species. eDNA methods are appealing for
54 open-ocean or other difficult-to-sample locations because sampling can be fast, standardized, non-lethal, and
55 detect many species simultaneously; sampling involves only the collection and processing of environmental
56 samples. Even modest improvements in sampling efficiency from current surveys can reduce the duration of
57 surveys, yield substantial cost savings for focal species surveys, and free survey time to be reallocated to other
58 understudied communities. However, such broad-scale implementation depends upon providing eDNA-based
59 estimates of abundance at management-relevant scales [10–12].

60 Observations of eDNA differ from observations derived from traditional methods (e.g. visual [13–14], capture
61 ([15–16], or acoustic [12] surveys) and the degree of agreement between individual samples of eDNA and
62 traditional methods collected simultaneously often determines whether eDNA-based methods are viewed as
63 successful or not [12,16]. However, eDNA observations arise from fundamentally different processes than
64 observations from these traditional survey methods – most dramatically, by exponential amplification of
65 DNA molecules in an environmental sample [18, 19], but also because the distribution of eDNA itself in the
66 environment is not identical to the distribution of its source organisms [5,7,8]. In the case of microbial eDNA,
67 this distributional distinction is negligible, but for larger animals – such as fishes or marine mammals – it is
68 not. Conceptually, fish are discrete, while the DNA traces they leave in the water are continuous, blurring

their environmental fingerprint over space and time [10]. For example, acoustic surveys of pelagic fishes reflect the patchy and skewed distribution of schooling fishes [20]. By comparison, we expect the associated eDNA to be distributed more evenly as a result of fish movement and the lag between shedding and decay processes [8, 11]. Understanding the ecology of eDNA [7] makes possible an honest assessment of the potential uses and limitations of eDNA for applied environmental problems, and allows each data stream to be used to its best advantage.

Here, we leverage the most spatially extensive eDNA survey of the oceans to date – spanning over 86,000 km², an area of ocean approximately equivalent to Portugal’s land area, and to depths of 500m – to document the empirical patterns of eDNA for a commercially important and abundant fish species, Pacific hake (*Merluccius productus*). Hake is a semi-pelagic schooling species and is among the most abundant fish species in the California Current Ecosystem [21, 22], supporting a large and important fishery along the Pacific coasts of US and Canada with coastwide catches in excess of 400,000mt annually and ex-vessel value in excess of \$60 million in recent years [22]. The rich datasets available for hake provide an opportunity to rigorously compare available information from traditional surveys with eDNA using parallel statistical models that relate observations from each data type to quantitative indices of abundance.

We investigate large-scale and depth-specific spatial patterns of hake DNA in the open ocean using a quantitative PCR assay targeting the 12S mitochondrial gene region [23]. We show how eDNA can be aggregated to provide a depth-integrated index of hake abundance comparable to acoustic-trawl survey results used for fisheries stock assessments [22]. The spatial-statistical model we use is a first for eDNA in the ocean and makes results for eDNA surveys comparable to other methods used in quantitative natural-resources management. We derive metrics of the species’ spatial and depth distribution and investigate the relative precision of the eDNA and acoustic surveys. Our results show that eDNA analyses can provide important information about abundance and distribution at management-relevant scales, provide relatively straightforward opportunities for supplementing existing surveys, and open the door for providing quantitative information for additional species that are currently un- or under-studied.

Materials and Methods

Study species

Pacific hake are a schooling, semi-pelagic fish that are a key component of the California Current ecosystem as both predator and prey, migrating to the surface at night and back to mid-water depths during the day [25]. They are targeted by major fisheries in both the United States and Canada [22] and adults perform

99 seasonal migrations between southern spawning areas and northern foraging areas [21, 24, 25].

100 **Field sampling and processing for eDNA**

101 We collected eDNA samples during the 2019 U.S.-Canada Integrated Ecosystem & Acoustic-Trawl Survey for
102 Pacific hake aboard the NOAA Ship *Bell M. Shimada* from July 2 to August 19 [26]. Detailed collection
103 protocols and laboratory analyses for all steps are provided in Ramón-Laca *et al.* [23]. We briefly summarize
104 those protocols here.

105 We collected seawater from up to six depths (3, 50, 100, 150, 300, and 500m) at 186 stations where a
106 Conductivity Temperature and Depth (CTD) rosette was deployed. These stations were spread across 36
107 acoustic transects (Fig. 2). We included 1,769 individual water samples collected at 892 depth-station
108 combinations (a small number of samples were contaminated or lost during processing). 710 depth-stations
109 were collected at 50m deep or deeper. Two replicates of 2.5L of seawater were collected at each depth and
110 station from independent Niskin bottles attached to a CTD rosette. Water samples from 3m were collected
111 from the ship's salt water intake line but processed identically to Niskin samples. Nearly all CTD casts and
112 therefore water collection for eDNA occurred at night while acoustic sampling (see below) took place during
113 daylight hours.

114 To account for possible contamination, negative sampling controls were collected routinely by filtering 2L of
115 distilled water from either the onboard evaporator or from distilled water brought from the laboratory for
116 this purpose (N=49 in total). The water was filtered immediately using 47mm diameter mixed cellulose ester
117 sterile filters with a $1\mu\text{m}$ pore size using a vacuum pump. The filters were stored at room temperature in
118 Longmire's buffer until DNA extraction [27].

119 The DNA was extracted using a modified phenol:chloroform method with a phase lock to increase the
120 throughput and yield. Quantification of Pacific hake was performed by qPCR using a specific TaqMan assay
121 on a QuanStudio 6 (Applied Biosystems) that included an internal positive control (IPC) of the reaction to
122 account for PCR inhibition. Any delay of more than 0.5 cycles from the IPC at the non-template controls of
123 the PCR was considered inhibition. For inhibited samples, we used a 1:5 dilution in subsequent analyses.

124 **Spatial eDNA Model**

125 We developed a Bayesian state-space framework for modeling DNA concentration in the coastal ocean.
126 State-space models separate the true biological process from the methods used to observe the process. We
127 use a relatively simple process model. Let D_{xyd} be the true, but unobserved concentration of hake DNA
128 (DNA copies L^{-1}) present at spatial coordinates $\{x, y\}$ (northing and eastings, respectively, in km) and

129 sample depth d (meters). We model the DNA concentration as a spatially smooth process at each depth
130 sampled ($d = 3, 50, 100, 150, 300$, or 500m) and linear on the \log_{10} scale,

$$\log_{10} D_{xyd} = \gamma_d + s(b) + t_d(x, y) \quad (1)$$

131 where γ_d is the spatial intercept for each depth, $s(b)$ indicates a smoothing spline as a function of bottom
132 depth in meters (b), and $t_d(x, y)$ is a tensor-product smooth that provides an independent spatial smooth for
133 each depth. We use cubic regression splines for both univariate and tensor-product smoothes. We investigated
134 a range of knot densities for smoothes in preliminary investigations (see *ESM*).

135 From the process model in eq. 1, we construct a multi-level observation model. First, we model the DNA
136 concentration in each Niskin bottle i , as a random deviation from the true DNA concentration at that depth
137 and location and include three offsets to account for variation in the processing of eDNA extracted from
138 Niskin bottles.

$$\log_{10} E_i = \log_{10} D_{xyd} + \delta_i + \log_{10} V_i + \log_{10} I_i + \mathbf{I}\omega \quad (2)$$

$$\delta_i \sim Normal(0, \tau_d) \quad (3)$$

139 where V_i is the proportion of 2.5 L filtered from Niskin i (in nearly all cases $V_i = 1$), I_i is the known dilution
140 used on sample i to eliminate PCR inhibition, and $\mathbf{I}\omega$ is an estimated offset for an ethanol wash error. Here,
141 \mathbf{I} is an indicator variable where $\mathbf{I} = 1$ for affected samples and $\mathbf{I} = 0$ otherwise (see also *ESM*).

142 When using qPCR, we do not directly observe eDNA concentration, we observe the PCR cycle at which each
143 sample can be detected (or if it was never detected). We use a hurdle model to account for the fact that there
144 is a probabilistic detection threshold (the PCR cycle of amplification is detected $G = 1$ or is not observed
145 $G = 0$). Conditional on being detected, we observe the PCR cycle (C) as a continuous variable that follows a
146 t distribution,

$$G_{ijr} \sim Bernoulli(\phi_{0j} + \phi_{1j} \log_{10} E_i) \quad (4)$$

$$C_{ijr} \sim T(\nu, \beta_{0j} + \beta_{1j} \log_{10} E_i, \eta) \quad if G_{ijr} = 1 \quad (5)$$

147 Here j indexes the PCR plate on which sample i and replicate r were run. We conducted 3 PCR reactions for
148 each E_i . We fix the degrees of freedom for the t-distribution ($\nu = 3$) to allow for heavy-tailed observations.
149 Note that there are different intercept (ϕ_{0j}, β_{0j}) and slope (ϕ_{1j}, β_{1j}) parameters for each PCR plate to allow
150 for among-plate variation in amplification. See the *ESM* for additional components of the statistical model.
151 We use diffuse prior distributions for all parameters (Table S1).

152 **Acoustic-trawl data**

153 In parallel with water collection for eDNA, we incorporated data on hake biomass derived from the contem-
154 poraneously collected data [26], consisting of 57 acoustic transects totaling 4,483 km in length. 45 midwater
155 trawls were deployed that provide information on the age, size and therefore signal strength of hake [22, 26].
156 Methods for converting raw acoustic and trawl data to biomass concentrations can be found in [21, 22] and
157 references therein. We used derived estimates of biomass concentration ($mt\ km^{-2}$) for hake ages 2 and older
158 that integrate the biomass in the water column between depths of 50 and 500m in all analyses. All acoustic
159 data and associated trawls were collected during daylight hours. Therefore there was a lag between collection
160 of acoustic and eDNA data, though for nearly all cases were separated by less than 24 hours. The temporal
161 separation of eDNA and acoustic sampling precluded direct comparisons at the single-sample level.

162 **Spatial acoustic-trawl model**

163 In parallel with the model for qPCR data, we estimated a spatial model for the hake biomass derived from
164 the acoustic-trawl survey. The biomass index created from the acoustic-trawl data for the entire survey area
165 (34.4°N to 54.7°N) is used in stock assessments that determine the allowable catch and allocation of hake
166 catch for fleets from the United States, Canada, and Tribal Nations [22]. As the eDNA samples only cover
167 a portion of this range (38.3°N to 48.6°N), we used the biomass observations within this latitudinal range
168 to generate spatially smooth estimates of biomass. Acoustic transects are divided into 0.926km (0.5nm)
169 segments and the biomass (age 2 and older) concentration within each segment is used as data [22].

170 Unlike the eDNA data, age-specific biomass estimates are available only as a biomass integrated across the
171 entire water column (from depths of 50 to 500m). We fit a Bayesian hurdle model using a form similar to the
172 eDNA, modeling biomass concentration (F_{xy} ; units: $mt\ km^{-2}$) using two separate spatial submodels: a) the
173 probability of occurrence and b) abundance conditional on the presence of hake. We model both components
174 as a function of bottom depth (smooth) and a spatial smooth,

$$H_{xy} \sim Bernoulli(\text{logit}^{-1}(\zeta_H + s_H(b) + t_H(x, y))) \quad (6)$$

$$F_{xy} \sim LogNormal(\zeta_F + s_F(b) + t_F(x, y) - 0.5\kappa^2, \kappa) \quad \text{if } H_{xy} = 1 \quad (7)$$

where H_{xy} is 1 if the observed biomass concentration is non-zero and 0 otherwise. In this formulation, ζ is the spatial intercept for each model component, $s(b)$ indicates a smoothing spline of as a function of bottom depth in meters (b), and $t(x, y)$ is a tensor-product smooth over latitude and longitude. κ is the standard deviation of the positive observations on the log scale. Table S2 provides the prior distributions for this model.

Model Estimation

We implemented both the eDNA and acoustic-trawl models using the Stan language as implemented in R (*Rstan*). All relevant code and data are provided in the online supplement. For the eDNA model, we ran 4 MCMC chains using 1,500 warm up and 9,000 sampling iterations. For the acoustic-trawl model, we ran 4 MCMC chains using 1,200 warm up and 3,000 sampling iterations.

We used traceplots and \hat{R} diagnostics to confirm convergence ($\hat{R} < 1.01$ for all parameters) – there were no divergent transitions in the sampling iterations. To generate design matrices necessary for estimating covariate effects we used the R package *brms* [28, 29]. We use diffuse prior distributions for all parameters (Table S1). Posterior summaries of parameters can be found in the *ESM*.

Coordinate systems, covariates, and spatial predictions

We generated 5km resolution gridded maps for both the acoustic-trawl and eDNA models to enable direct comparisons between models. This vector-based grid was developed and used by others [30] for interpolating various spatial models and uses a custom coordinate reference system that conserves area and distance reasonably well across the west coast of the United States (*ESM*) and was a suitable resolution for the purposes of our analyses.

To create spatial predictions for both eDNA and acoustic-trawl models, we took 4,000 draws from the joint posterior and generated predictions for the centroid of each grid cell. We calculated posterior means and uncertainty bounds among posterior draws. For the eDNA model we made projections for D_{xyb} ; we do not present results from including additional observation processes on top of the estimated DNA concentrations in the main text. We generated posterior predictive distributions for other model diagnostics checks (see *ESM*).

200 **Creating an eDNA index**

201 Our model provides direct predictions for hake DNA concentration at depths of 50, 100, 150, 300, and
202 500m. To produce an index spanning depths of 50 to 500m, we equally weighted depths between 50 and
203 500m using linear interpolation between the closest depths. We used posterior predictions at each depth
204 to provide predicted DNA densities at 200, 250, 350, 400, and 450m for each 5km grid cell. Because some
205 spatial locations have depths of less than 500m, we only include predicted DNA concentrations to a depth
206 appropriate for the bathymetry (e.g. a location with a depth of 180m only includes values from 50, 100 and
207 150m). We sum across all depths (between 50 and up to 500m) to generate a depth-integrated index of hake
208 DNA. This index will be proportional to the hake DNA found in the water column. However, as we are only
209 summing across discrete depths, not integrating values across the entire water column nor multiplying by the
210 total water volume within each grid cell, the absolute value of the index will depend upon the number of
211 discrete depths we use. We refer to this as an eDNA index to differentiate it from predictions to specific
212 depths.

213 **Results**

214 Hake were detected throughout the survey region (Fig. 1, 2), but hake DNA was far more commonly detected
215 than the acoustic signature of hake. As expected for a patchily distributed species, acoustic sampling identified
216 hake biomass in a minority of 0.983km long transect segments (1764 of 4841; 36%). By contrast, genetic
217 signatures of hake were nearly ubiquitous, detected in 94% of water samples (1670 of 1769 2.5L samples) and
218 98% of sampling stations (875 of 892 stations), reflecting considerable spatial smoothing of the eDNA signal
219 relative to the acoustic detections.

220 Distributions of hake DNA in the study area varied substantially with depth, with highest concentration
221 between 100m and 300m depth along the continental shelf break and south of the Oregon-California border
222 at 42°N (Fig. 1). Hake DNA concentration was far more homogeneous at depth than near the surface:
223 concentrations at 500m were generally low and showed limited spatial variation, while the near-surface layers
224 (3m, 50m) showed generally higher but more heterogeneous concentrations (Figs. 1, S8) among individual
225 sample bottles. The median coefficient of variation (CV) was larger than 1 for both 3m and 50m depths but
226 only about 0.3 for depths 100m and deeper (Fig. 1G). Such uncertainty reflects large observed differences in
227 DNA concentration between replicate samples taken at the same sample location and substantial differences
228 among proximate sampling locations (see *ESM*).

229 We combined DNA information between 50 and 500m to produce both a spatially smooth, depth-integrated

230 estimate of hake DNA concentration (Fig. 2B). Separately, we generated a spatially smooth estimate of age
231 2+ biomass from the acoustic-trawl survey (Fig. 2C). While spatial analyses for acoustic data are commonly
232 used in assessing pelagic species worldwide, this is a first application of these methods with eDNA data. The
233 eDNA biomass index showed strong spatial patterning with highest values along the continental shelf break
234 with notable peaks in central California and Oregon waters. In contrast, acoustic-trawl observations were highly
235 spatially variable – a common feature observed in acoustic surveys [31] – with high hake density and others
236 with very low density in close proximity (see also Fig. S17). At the scale of individual 25km^2 grid cells, eDNA
237 and acoustic surveys were modestly correlated ($\rho = 0.55[0.53, 0.57]$, Pearson product-moment correlation on
238 posterior mean prediction [90% CI]; Fig. 3) but there is considerable scatter in the relationship. Large eDNA
239 values never occurred at locations which had very low acoustic biomass, but very high acoustic estimates
240 corresponded to moderate values of eDNA. Notably, acoustic biomass estimates had a very right-skewed
241 distribution across the 3,455 25km^2 ocean cells considered – most values were near zero with very few high
242 values – while eDNA values were decidedly less skewed (Fig. 3). Taken together, these observations again
243 suggest a smoother distribution of eDNA information relative to the patchier acoustic detections.

244 When aggregated to one degree latitude bins, the correlation between eDNA and acoustics increased
245 substantially ($\rho = 0.88[0.65, 0.96]$; Fig. 3) with acoustics and eDNA scaling approximately linearly. Such
246 increased correlation is not dependant upon the spatial groupings in Fig. 3 (see Fig. S15 and S16 for results
247 from an alternate spatial grouping). At this scale, eDNA and acoustic-trawls provide nearly equivalent
248 information about relative biomass. At a coast-wide scale, the uncertainties (CVs) of the acoustic-trawl
249 estimate and eDNA index were nearly identical (both 0.09). This similarity occurred despite the eDNA only
250 being collected at 186 locations, whereas the acoustic-trawl data includes 4,841 acoustic transect segments
251 and 45 mid-water trawls to determine age- and length-structure of the hake.

252 Finally, the two methods produced nearly identical latitudinal distributional estimates as measured by center
253 of gravity (median value within the projection range) and cumulative distribution (90% CIs overlapping for the
254 entire latitudinal range; Fig. 4). Furthermore, averaged across space, hake DNA concentrations were highest
255 along the continental shelf break (bottom depths between 125 and 400m) and at water depths between 150m
256 and 300m (Fig. 4C). All of these observations are consistent with published descriptions of hake depth and
257 habitat preferences [21, 22, 32, 33].

258 Discussion

259 Ocean surveys are often used to generate large-scale, quantitative indices of species' abundances. At the
260 spatial scale relevant to management for hake along the U.S. west coast – our survey region encompasses

the majority of habitat for the Pacific hake stock – analysis of a limited number of discrete water samples for eDNA provides comparable indices of hake biomass to acoustic-trawl surveys despite far fewer eDNA observations. While other efforts have developed quantitative methods for eDNA within rivers [10], lakes [15, 33], estuaries [11], and nearshore marine habitats [12], we produce a large-scale study that can serve as a template for using eDNA to determine abundance and species distributions with clear practical applications to both conservation and fisheries. Importantly, we push beyond sample-to-sample comparisons of alternate sampling methods and make comparisons at the population-scale. The spatial scale investigated here (tens of thousands of km^2 or more) is roughly comparable to the scale at which most large ocean fisheries are managed both in the United States and internationally, suggesting eDNA approaches can begin to be broadly adopted for that purpose.

The kind of spatial-statistical model we report here brings eDNA analysis inline with the methods currently used in quantitative natural-resources management (e.g. [34, 35]. Despite the clear differences in biological processes producing eDNA signals versus acoustic trawl signals, these distinct data sets are both subject to rigorous analytical methods. We emphasize that eDNA data here are processed independently from acoustic-trawl data; no information from the acoustics informs eDNA or vice versa. Thus, the implementation of eDNA surveys provides a second survey of abundance for hake without requiring any additional days at sea, and should provide improved precision for estimated fish abundance when the two indices are incorporated into a stock assessment. Ultimately, eDNA holds unprecedented potential for improving the precision of abundance surveys, particularly when conducted in concert with existing surveys.

For determining an index of abundance over a very large area, we assert that eDNA works well because the concerns about the impact of DNA transport, degradation, and other processes [5, 7, 8] are negligible for our application (providing an index of abundance on large spatial scales). Hake DNA present within our survey boundaries was generated by hake present within the survey area; oceanographic processes like currents or upwelling are not of sufficient magnitude to transport meaningful amounts of water into or out of the survey domain on the time scale at which eDNA degrades [8]. Similarly, rates of DNA degradation are expected to be consistent across our sampling domain – cool, offshore, oceanic waters below 50m with relatively little among-sample variation in temperature, salinity, and other covariates identified as important for degradation [37, 38]. Such population closure and constant rate assumptions are reasonable [see also 12] and allow us to treat eDNA observations as analogous to other traditional sampling methods. We note that our modeling framework provides the flexibility to directly include relevant covariates into the observation model to account for relevant DNA processes if and when such information becomes available (see *Materials and Methods* and *ESM*). For hake, our eDNA results match available geospatial (Figs. 2, 4) and depth-specific patterns of hake

293 abundance [21, 26] (Fig. 4) from other methods, strongly suggesting our assumptions are reasonable and
294 justified. eDNA approaches may be less effective in applications focused on smaller temporal and spatial
295 scales such as detailed habitat-association studies where the precise locations of individuals are required [but
296 see 39, 40].

297 Many challenges to implementing eDNA surveys remain. Surveys are primarily valuable because they inform
298 temporal trends; most surveys, particularly those of marine species, are not used as measures of absolute
299 abundance but as indices of abundance relative to previous years [22, 34]. It will accordingly require years to
300 accumulate the kinds of eDNA-based time series that parallel those used in current management. Furthermore,
301 there are additional data streams needed for management applications that are not currently possible from
302 eDNA. For example, physical specimens are needed to document age, size, sex, and condition, all of which
303 cannot be extracted from eDNA at present, though these are active areas of research [41, 42]. At present,
304 eDNA approaches should be regarded as supplementing existing surveys, not replacing them.

305 Despite these limitations, the characteristics of eDNA surveys have several advantages. First, the samples
306 collected and analyzed here for hake can be re-analyzed for other species. Analyses using species-specific
307 qPCR should provide similar quantitative data for additional species. DNA metabarcoding approaches
308 can detect many species simultaneously [1], but metabarcoding results are difficult to link to abundance or
309 biomass [18, 19]. Second, surveys of eDNA provide the potential for large-scale replication and high precision;
310 as many replicate samples as desired can be collected, enabling researchers to target and achieve a desired
311 level of precision. Such replication is often not possible for other sampling methods that involve capturing
312 individuals. For example, repeatedly trawling a particular location will deplete the fish present, and therefore
313 such repeated sampling is generally not helpful for estimating abundance. In theory there are few limits
314 on replication using eDNA and our results indicate that the amount of small-scale variation between water
315 samples declines with depth (Fig. 1G, Fig. S9), suggesting that the amount of statistical noise and therefore
316 the amount of sampling needed may vary concomitantly. It is wholly unknown if other marine species will
317 exhibit similar depth-specific patterns of variability to those observed in hake.

318 We developed and applied our eDNA approach to Pacific hake because of its broad geographic range, economic
319 importance, and decades of associated survey information. The ability of eDNA to provide similar indices of
320 abundance and distribution to existing surveys lend strong support for the applicability of eDNA methods to
321 the unstudied majority of species in ocean ecosystems.

322 **Data accessibility**

323 Additional methodology and analysis are provided in the electronic supplementary material and in the online
324 data and code repository [43]. <https://github.com/nwfsc-cb/eDNA-Hake-public/releases/tag/v1.0.2>

325 **Acknowledgements**

326 Funding for this project was provided by the NMFS Genomic Strategic Initiative. Special thanks to S. Allen,
327 B. Dewees, J. Witmer and J. Davis for assistance during sample collection and the captain and crew of the
328 NOAA ship *Bell M. Shimada* for overall support during the survey cruise. A. Billings, J. Pohl, and members
329 of the Fisheries Engineering and Acoustic Technologies team provided logistical and analytical support for
330 both the acoustic and eDNA components of the survey. E. Iwamoto and members of the molecular genetics
331 lab at the Northwest Fisheries Science Center supported laboratory work. A. Berger, M.J. Ford, and J.F.
332 Samhouri provided helpful comments on earlier versions of the manuscript.

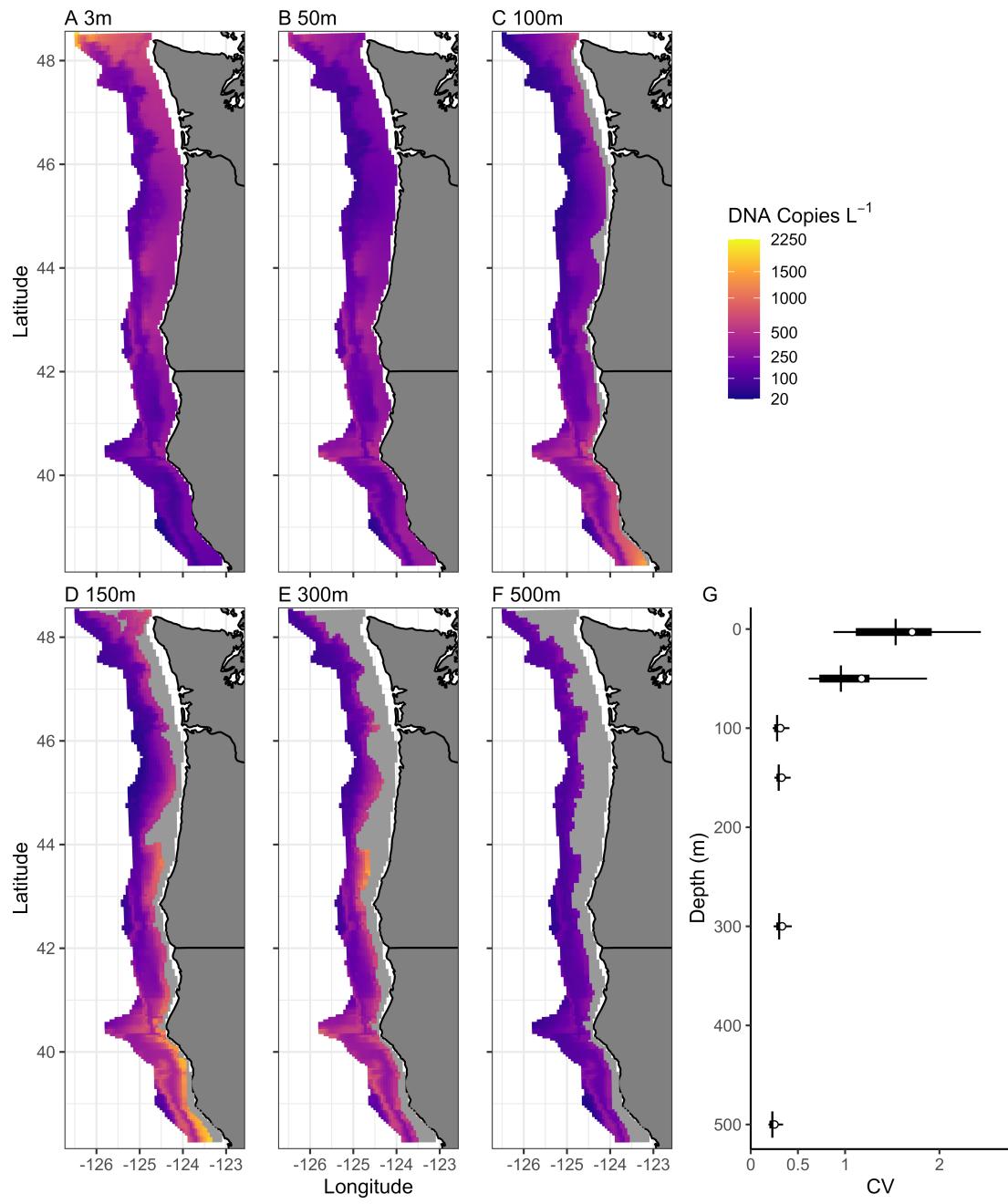


Figure 1: Predicted DNA concentration for six water depths shows clear spatial patterning in DNA concentration (A-F; posterior mean). G) Uncertainty around the posterior mean for each water depth as measure by the coefficient of variation. The distribution among all projected 25km^2 grid cells are shown (mean (circle), median (line), interquartile and 90% CI).

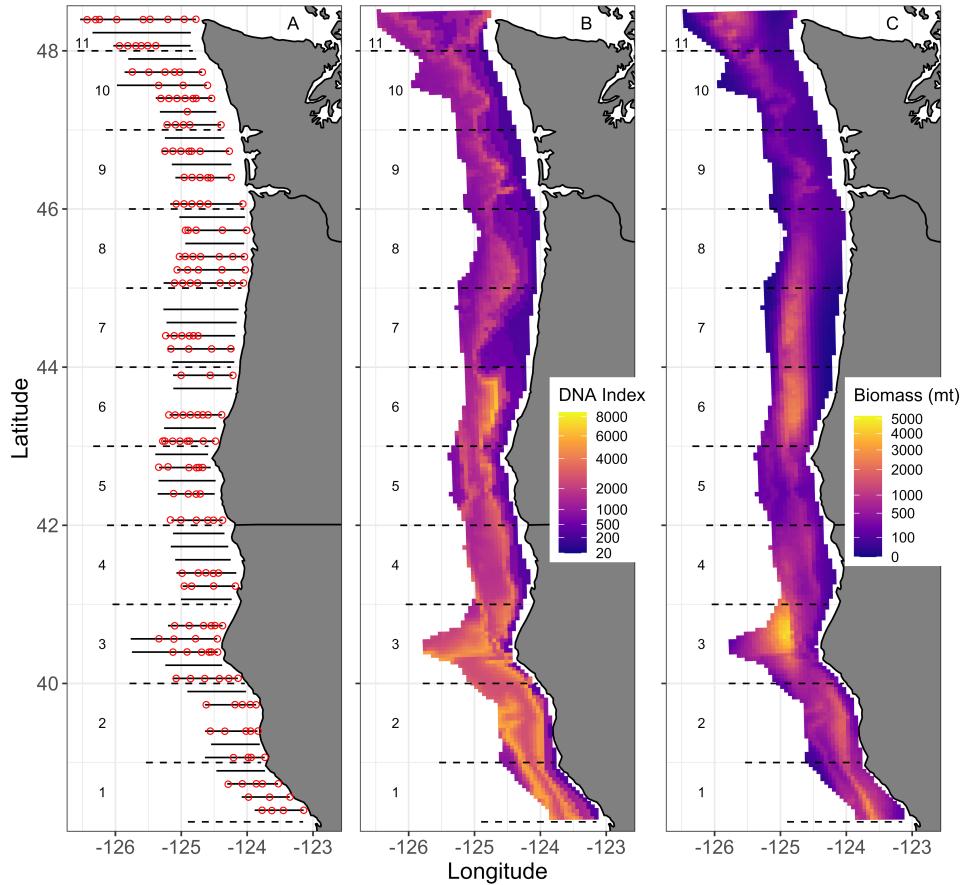


Figure 2: 2019 survey locations (A; red circles show eDNA sampling locations, lines show acoustic transects), depth-integrated index of hake DNA (B) and hake biomass from acoustic surveys (C). Both DNA and acoustic estimates are mean predicted values projected to a 5km grid and include information between 50 and 500m deep. All panels show one degree latitudinal bins used to aggregate abundance estimates over larger spatial scales (see Fig. 3).

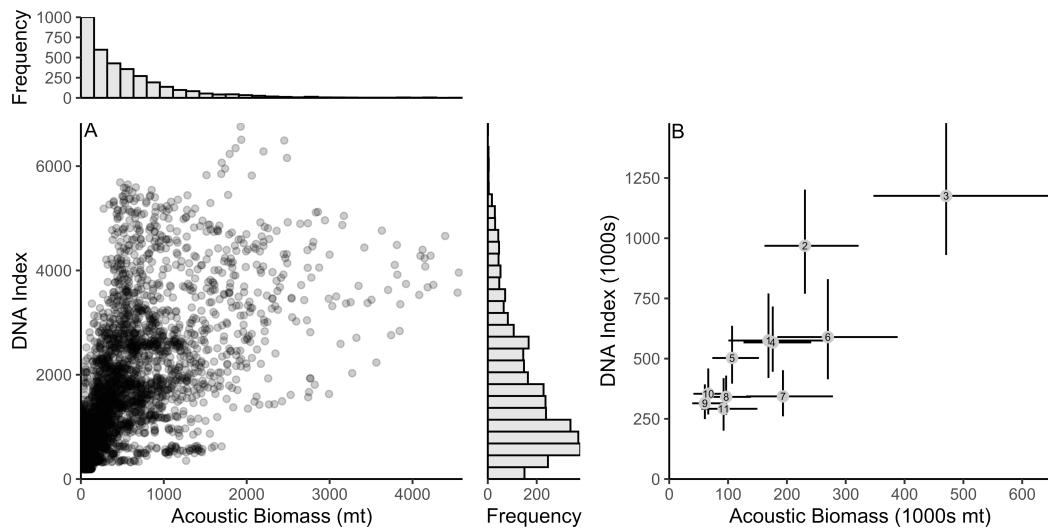


Figure 3: Pairwise comparison between DNA and acoustics-derived biomass. A) Posterior mean prediction from each method among the 3,455 25km^2 grid cells and includes the marginal histogram of posterior mean values for each method (correlation of posterior mean[90% CI]; $\rho = 0.55[0.53, 0.57]$). B) Correlation between methods among the 11, one degree latitude bins (posterior mean[90% CI] shown; $\rho = 0.88[0.65, 0.96]$). Numbers indicate regions identified in Fig. 2.

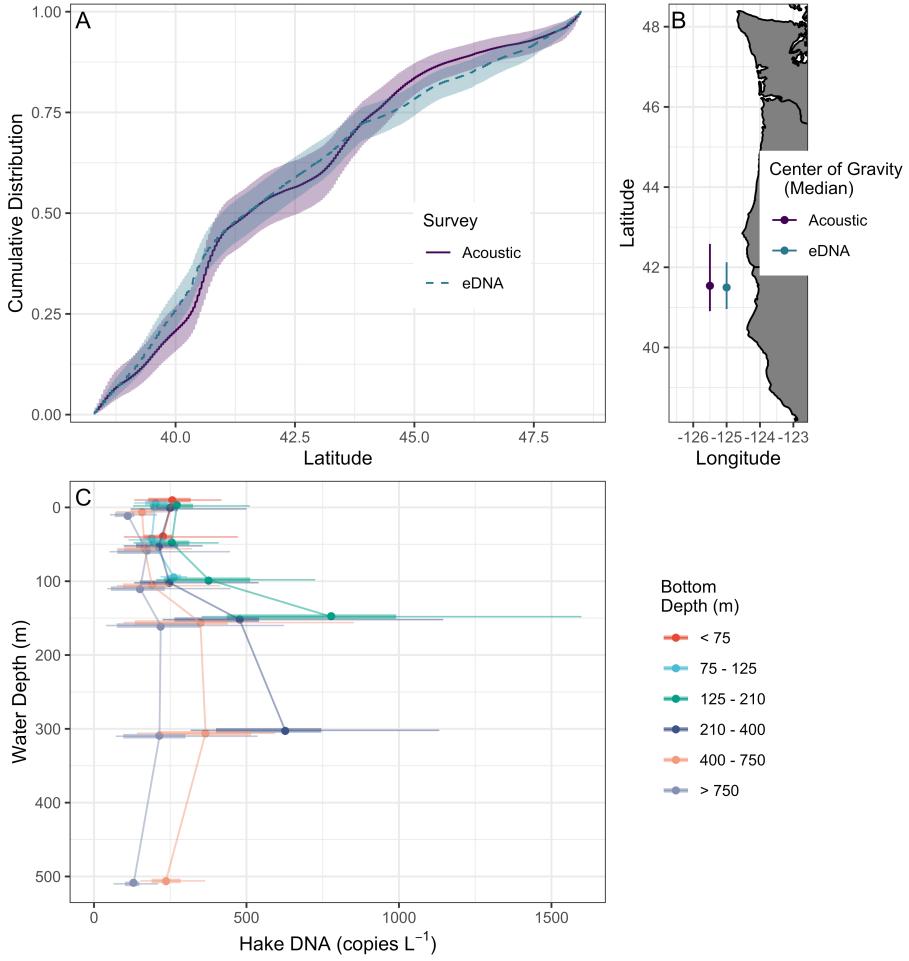


Figure 4: Estimates of distribution of Pacific hake. A) Cumulative distribution between 38.3 and 48.6°N (posterior means, 90% CI). B) Center of gravity (median of distribution) for each method (posterior means and 90% CI; only areas within the projection grid are included in this calculation ; see Figs. 1, 2). C) Posterior estimates of hake DNA concentration at each station-depth combination by the water depth sampled and categories of the depth of the bottom. The distribution of mean DNA concentration among station-depths (mean, interquartile range, and 90% CI among station-depths). Bottles at a sample location become increasingly similar at deeper sampling depths

333 **Citations**

- 334 1. Thomsen PF, Willerslev E. 2015 Environmental DNA—an emerging tool in conservation for monitoring
335 past and present biodiversity. *Biological Conservation* **183**, 4–18.
- 336 2. Boussarie G *et al.* 2018 Environmental DNA illuminates the dark diversity of sharks. *Science Advances* **4**,
337 eaap9661. (doi:10.1126/sciadv.aap9661)
- 338 3. Leempoel K, Hebert T, Hadly EA. 2020 A comparison of eDNA to camera trapping for assessment of
339 terrestrial mammal diversity. *Proceedings of the Royal Society B* **287**, 20192353.
- 340 4. Nguyen BN *et al.* 2020 Environmental DNA survey captures patterns of fish and invertebrate diversity
341 across a tropical seascape. *Scientific reports* **10**, 1–14.
- 342 5. Hansen BK, Bekkevold D, Clausen LW, Nielsen EE. 2018 The sceptical optimist: Challenges and
343 perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries* **19**, 751–768.
- 344 6. Rourke ML, Fowler AM, Hughes JM, Broadhurst MK, DiBattista JD, Fielder S, Wilkes Walburn J, Furlan
345 EM. 2021 Environmental dna (eDNA) as a tool for assessing fish biomass: A review of approaches and future
346 considerations for resource surveys. *Environmental DNA*
- 347 7. Barnes MA, Turner CR. 2016 The ecology of environmental DNA and implications for conservation
348 genetics. *Conservation genetics* **17**, 1–17.
- 349 8. Harrison JB, Sunday JM, Rogers SM. 2019 Predicting the fate of eDNA in the environment and implications
350 for studying biodiversity. *Proceedings of the Royal Society B* **286**, 20191409.
- 351 9. Rodríguez-Ezpeleta N *et al.* 2021 Biodiversity monitoring using environmental DNA. *Molecular Ecology*
352 *Resources* **21**, 1405–1409.
- 353 10. Carraro L, Hartikainen H, Jokela J, Bertuzzo E, Rinaldo A. 2018 Estimating species distribution and
354 abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences*
355 **115**, 11724–11729.
- 356 11. Shelton AO, Kelly RP, O'Donnell JL, Park L, Schwenke P, Greene C, Henderson RA, Beamer EM. 2019
357 Environmental DNA provides quantitative estimates of a threatened salmon species. *Biological Conservation*
358 **237**, 383–391.
- 359 12. Fukaya K *et al.* 2021 Estimating fish population abundance by integrating quantitative data on
360 environmental DNA and hydrodynamic modelling. *Molecular ecology* **30**, 3057–3067.
- 361 13. Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly

- 362 RP. 2016 Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular*
363 *Ecology* **25**, 527–541.
- 364 14. Tillotson MD, Kelly RP, Duda JJ, Hoy M, Kralj J, Quinn TP. 2018 Concentrations of environmental
365 DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biological Conservation*
366 **220**, 1–11.
- 367 15. Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ.
368 2016 Environmental DNA metabarcoding of lake fish communities reflects long-term data from established
369 survey methods. *Molecular ecology* **25**, 3101–3119.
- 370 16. Stoeckle MY, Adolf J, Charlop-Powers Z, Dunton KJ, Hinks G, VanMorter SM. 2021 Trawl and eDNA
371 assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey, USA. *ICES*
372 *Journal of Marine Science* **78**, 293–304.
- 373 17. Knudsen SW *et al.* 2019 Species-specific detection and quantification of environmental dna from marine
374 fishes in the baltic sea. *Journal of Experimental Marine Biology and Ecology* **510**, 31–45.
- 375 18. Kelly RP, Shelton AO, Gallego R. 2019 Understanding PCR processes to draw meaningful conclusions
376 from environmental DNA studies. *Scientific Reports* **9**, 1–14.
- 377 19. McLaren MR, Willis AD, Callahan BJ. 2019 Consistent and correctable bias in metagenomic sequencing
378 experiments. *Elife* **8**, e46923.
- 379 20. Zwolinski J, Fernandes PG, Marques V, Stratoudakis Y. 2009 Estimating fish abundance from acoustic
380 surveys: calculating variance due to acoustic backscatter and length distribution error. *Canadian Journal of*
381 *Fisheries and Aquatic Sciences* **66**, 2081–2095.
- 382 21. Malick MJ, Hunsicker ME, Haltuch MA, Parker-Stetter SL, Berger AM, Marshall KN. 2020 Relationships
383 between temperature and Pacific hake distribution vary across latitude and life-history stage. *Marine Ecology*
384 *Progress Series* **639**, 185–197.
- 385 22. Grandin CJ, Johnson KF, Edwards AW, Berger AM. 2020 Status of the Pacific hake (whiting) stock in
386 U.S. and Canadian waters in 2020., 273.
- 387 23. Ramón-Laca A, Wells A, Park L. 2021 A workflow for the relative quantification of multiple fish species
388 from oceanic water samples using environmental DNA (eDNA) to support large-scale 2021. *PLoS ONE* **16**,
389 e0257773.
- 390 24. Methot RD, Dorn MW. 1995 Biology and fisheries of north Pacific hake (*M. productus*). In *Hake*, pp.

- 391 389–414. Springer.
- 392 25. Ressler PH, Holmes JA, Fleischer GW, Thomas RE, Cooke KC. 2007 Pacific hake, *Merluccius productus*,
393 autecology: A timely review. *Marine Fisheries Review* **69**, 1–24.
- 394 26. de Blois S. 2020 The 2019 joint U.S.–Canada integrated ecosystem and Pacific hake acoustic-trawl survey:
395 Cruise report sh-19-06. *U.S. Department of Commerce, NOAA Processed Report NMFS-NWFSC-PR-2020-03*
- 396 27. Renshaw MA, Olds BP, Jerde CL, McVeigh MM, Lodge DM. 2015 The room temperature preservation
397 of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA
398 extraction. *Molecular Ecology Resources* **15**, 168–176.
- 399 28. Bürkner P-C. 2017 brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical
400 Software* **80**, 1–28. (doi:10.18637/jss.v080.i01)
- 401 29. Bürkner P-C. 2018 Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* **10**,
402 395–411. (doi:10.32614/RJ-2018-017)
- 403 30. Feist BE, Samhouri JF, Forney KA, Saez LE. 2021 Footprints of fixed-gear fisheries in relation to rising
404 whale entanglements on the US West Coast. *Fisheries Management and Ecology* **28**, 283–294.
- 405 31. Rivoirard J, Simmonds J, Foote K, Fernandes P, Bez N. 2008 *Geostatistics for estimating fish abundance*.
406 John Wiley & Sons.
- 407 32. Agostini VN, Francis RC, Hollowed AB, Pierce SD, Wilson C, Hendrix AN. 2006 The relationship between
408 Pacific hake (*Merluccius productus*) distribution and poleward subsurface flow in the California Current
409 System. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 2648–2659.
- 410 33. Buxton AS, Groombridge JJ, Zakaria NB, Griffiths RA. 2017 Seasonal variation in environmental DNA
411 in relation to population size and environmental factors. *Scientific reports* **7**, 1–9.
- 412 34. Maunder MN, Punt AE. 2004 Standardizing catch and effort data: A review of recent approaches.
413 *Fisheries research* **70**, 141–159.
- 414 35. Thorson JT, Shelton AO, Ward EJ, Skaug HJ. 2015 Geostatistical delta-generalized linear mixed models
415 improve precision for estimated abundance indices for west coast groundfishes. *ICES Journal of Marine
416 Science* **72**, 1297–1310.
- 417 36. Murakami H *et al.* 2019 Dispersion and degradation of environmental dna from caged fish in a marine
418 environment. *Fisheries science* **85**, 327–337.

- 419 37. Collins RA, Wangensteen OS, O'Gorman EJ, Mariani S, Sims DW, Genner MJ. 2018 Persistence of
420 environmental DNA in marine systems. *Communications Biology* **1**, 1–11.
- 421 38. Andruszkiewicz Allan E, Zhang WG, C Lavery A, F Govindarajan A. 2021 Environmental dna shedding
422 and decay rates from diverse animal forms and thermal regimes. *Environmental DNA* **3**, 492–514.
- 423 39. Pont D *et al.* 2018 Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers
424 despite its downstream transportation. *Scientific Reports* **8**, 10361.
- 425 40. Jeunen G-J, Knapp M, Spencer HG, Lamare MD, Taylor HR, Stat M, Bunce M, Gemmell NJ. 2019
426 Environmental dna (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats
427 connected by water movement. *Molecular Ecology Resources* **19**, 426–438.
- 428 41. Adams CI, Knapp M, Gemmell NJ, Jeunen G-J, Bunce M, Lamare MD, Taylor HR. 2019 Beyond
429 biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes* **10**, 192.
- 430 42. Mayne B, Korbie D, Kenchington L, Ezzy B, Berry O, Jarman S. 2020 A DNA methylation age predictor
431 for zebrafish. *Aging* **12**, 24817–24835. (doi:10.18632/aging.202400)
- 432 43. Shelton AO, Ramón-Laca A, Wells A, Clemons J, Chu D, Feist BE, Kelly RP, Parker-Stetter SL, Thomas
433 R, Nichols K, Park L. 2021 Environmental DNA provides quantitative estimates of abundance and distribution
434 in the open ocean. Available at: <https://github.com/nwfsc-cb/eDNA-Hake-public/releases/tag/v1.0.2>