

¹ **Environmental DNA provides quantitative estimates of abundance and distribution in support of fisheries management.**

³ Authors... in no particular order at present. Andrew Olaf Shelton^{1*}, Linda Park¹, Krista Nichols¹, Ana
⁴ Ramon-Laca, Abigail Wells², Sandra L. Parker-Stetter^{3†}, Rebecca Thomas³, Chu³, Julia³, Ryan P. Kelly⁴,
⁵ Jameal F. Samhouri¹, Blake E. Feist¹

⁶ ¹Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service,
⁷ National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112, U.S.A.

⁸ ² Lynker Associates, Under contract to Northwest Fisheries Science Center, National Marine Fisheries Service,
⁹ National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112, U.S.A.

¹⁰ ³Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine
¹¹ Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA
¹² 98112, U.S.A.

¹³ ⁴ University of Washington, School of Marine and Environmental Affairs, 3707 Brooklyn Ave NE, Seattle,
¹⁴ WA 98105, U.S.A.

¹⁵ [†] AFSC

¹⁶ * corresponding author: ole.shelton@noaa.gov

¹⁷ *Keywords:*

¹⁸ PNAS requirements:

¹⁹ Research reports describe the results of original research of exceptional importance. The preferred length of
²⁰ these articles is 6 pages, but PNAS allows articles up to a maximum of 12 pages. A standard 6-page article is
²¹ approximately 4,000 words, 50 references, and 4 medium-size graphical elements (i.e., figures and tables).

²² Primary research Article

²³ **Abstract**

²⁴ All creatures inevitably leave genetic traces in their environments, and the resulting environmental DNA
²⁵ (eDNA) therefore reflects the species present in a given habitat. It remains unclear, however, whether eDNA
²⁶ signals are sufficiently quantitative for use in regulatory or policy decisions on which human livelihoods or
²⁷ conservation successes may depend. Here, we report the results of the largest eDNA ocean survey to date
²⁸ (spanning 86,000 km² to depths of 500m) to understand Pacific hake (*Merluccius productus*), the target of the
²⁹ largest finfish fishery along the west coast of the United States. We sampled eDNA in parallel with traditional
³⁰ acoustic survey methods and show how eDNA provides a spatially smooth signature of hake relative to
³¹ the patterns seen in traditional acoustic survey methods. Despite local differences, when aggregated to
³² management relevant scales the two methods yield comparable information about the broad-scale spatial
³³ distribution and abundance of hake. This occurs despite eDNA arising from a limited number of discrete
³⁴ samples within the larger acoustic survey. The analysis also yields novel information about depth-specific
³⁵ spatial patterns of eDNA at a large spatial scale with strong depth-specific patterns in eDNA abundance and
³⁶ variability. We demonstrate the potential power and efficacy of eDNA sampling for applied problems and
³⁷ posit that eDNA methods have general quantitative applications that will prove especially valuable in data-
³⁸ or resource-limited contexts.

³⁹ (1)

40 **Introduction**

41 Environmental DNA, in which the DNA from target organisms is collected from an environmental medium
42 (e.g. soil or water), can detect species in a wide range of terrestrial, aquatic, and marine habitats (2) . As
43 a result, eDNA may efficiently survey species diversity and changes in community membership] REFs.
44 However, most applied natural resource questions depend upon estimates of abundance (e.g. fisheries or
45 managing species of conservation concern) and for these topics, eDNA must provide information about
46 abundance in order to be useful. While most studies find a positive relationship between eDNA concentrations
47 and other survey methods (3), uncertainty about the strength of the eDNA-abundance relationship due to
48 the complexity of eDNA generation, transport, degradation, and detection have limited the application of
49 eDNA in many quantitative applications (4). While the use of eDNA methods has skyrocketed, growing
50 XXX% over the past Y years (3), reflecting widespread adoption of eDNA technologies, basic questions about
51 the behavior of eDNA limit its practical application and slow its adoption in environmental management.

52 Rigorous, well designed surveys underlie the successful management and conservation of wild populations.
53 But field surveys are expensive – open oceans surveys involve ship time costing tens of thousands of dollars
54 per day, for example – and are typically tailored to a single or relatively narrow suite of species. eDNA
55 methods are appealing for open-ocean or other difficult-to-sample locations because sampling can be fast,
56 standardized, and non-lethal for many species simultaneously; sampling involves only the collection and
57 processing of environmental samples. Even modest improvements in sampling efficiency from current surveys
58 can reduce the duration of surveys potentially yielding substantial cost savings for focal species surveys and
59 freeing survey time to be reallocated to other understudied communities. However, much of this kind of
60 implementation depends upon providing eDNA-based estimates of abundance at management-relevant scales
61 (5–7).

62 Observations of eDNA differ in varying degrees from observations that use traditional methods (e.g. visual
63 (8, 9), capture (10, 11), or acoustic (7) surveys) and the degree of agreement between individual samples
64 of eDNA and traditional methods collected simultaneously often determines whether eDNA is viewed as
65 successful or a failure (7, 11). However, eDNA observations arise from fundamentally different processes
66 than observations from these traditional survey methods – most dramatically, by exponential amplification
67 of DNA molecules in an environmental sample, but also because the distribution of eDNA itself in the
68 environment differs significantly from the distribution of its source organisms. In the case of microbial eDNA,
69 this distributional distinction is negligible, but for animals of management relevance – such as pelagic fishes
70 – it is not. Conceptually, fish are discrete, while the genetic traces they leave in the water are continuous,
71 smoothing their environmental fingerprint over space and time.

72 If two methods are sampling different phenomena over different spatial or temporal scales, we expect individual
73 observations from those methods to differ. For example, acoustic trawls used in fisheries stock assessments
74 reflect the highly patchy density of schooling fishes in space and time. By comparison, we might expect
75 the associated eDNA to be distributed more evenly as a result of the lag between shedding and decay
76 processes. Understanding the ecology of eDNA (12) makes possible an honest assessment of the potential
77 uses and limitations of eDNA for applied environmental problems, and lets us use each data stream to its
78 best advantage.

79 Here, we leverage the most spatially extensive eDNA survey of the oceans to date – spanning over 86,000 km²
80 across 10 degrees of latitude, an area of ocean approximately equivalent to the land area Portugal, and to
81 depths of 500m – to document the empirical patterns of eDNA for a commercially important and abundant
82 species, Pacific hake *Merluccius productus*. Hake is a semi-pelagic schooling species and is among the most
83 abundant species in the California Current Ecosystem (13, 14), supporting a large and important fishery
84 along the Pacific coasts of US and Canada with coastwide catches in excess of 400,000t annually from 2017 to
85 2019 (14). The rich datasets available for hake provide an opportunity to compare available information from
86 traditional surveys with eDNA.

87 We demonstrate the presence of large-scale, depth-specific spatial patterns of hake DNA in the ocean ocean
88 using a quantitative PCR assay. We then show how eDNA can be aggregated to provide a depth-integrated
89 index of hake abundance comparable to acoustic-trawl survey results used for commercial stock assessments.
90 Consistent with the different expected distributions of eDNA and acoustic-trawl data, the two indices

⁹¹ are only modestly correlated at local scales (tens of km^2) but very strongly correlated when aggregated
⁹² to management-relevant scales (thousands of km^2). We derive metrics of the species' spatial distribution
⁹³ consistent with acoustics results, and find that eDNA provides nearly identical model precision as the acoustic
⁹⁴ data with [a fraction of the data | discrete bottle samples, etc].

95 EXTRA JUNK EXTRA JUNK EXTRA JUNK.

96 We and that eDNA has several appealing statistical characteristics of

97 (9) SALMON

98 (5) PNAS Using river distribution models to estimate abundance (SPATIAL STUFF)

99 people who have made rinky-dink joint models: (15)

100 (4) Review of eDNA processes (production, transport, etc.)

101 (10) used gillnets in combination with metabarcoding, lakes.

102 (11) Metabarcoding and trawls. New Jersey.

103 Here we show how statistical methods from the fields of ecology and fisheries to bear on eDNA data and

104 yield inference about the abundance and distribution of a commercially important fish, Pacific hake *Merc. . .*

105 We focus on hake because the rich datasets available for this species provided an opportunity to compare

106 and contrast available information with eDNA. We pair the largest eDNA survey of the coastal ocean to

107 date (spanning . . .) to traditional acoustic-trawl methods conducted in parallel with eDNA samples to

108 understand the

109 We show how eDNA can simultaneously differ substantially from acoustics estimates at local scales (kms)

110 and yet provide very similar patterns of abundance at large spatial scales.

111 In addition, we illustrate how the patterns of hake eDNA vary substantially spatially and by depth, revealing

112 new insights into the ecology of eDNA. Specifically, we show how small scale variability in eDNA appears to

113 decline with depth, and does not follow

114 Our work demonstrates that there is certainly a complex and interesting ecology of eDNA that needs additional

115 study. While this complexity obscures the relationships

116 eDNA has the potential to provide information simultaneously on a wide range of species including many

117 that are currently unstudied.

118 resulting in species of economic importance, exceptional charisma, or severe conservation concern, being

119 targeted for surveys while most species are ignored. In the context of a changing climate, this bias in surveys

120 leaves us blind to shifting distributions. . . other things.

121 A corollary to this fact is that we are treating eDNA samples as conceptually identical

122 More importantly, the question of what can be reasonably inferred from a given eDNA data set

123 Birds (Christmas counts), mammals (), fisheries In the oceans,

124 Integrated models are widely used in fisheries applications

125 By analogy deep understanding of the "ecology" of eDNA

126 While the promise of eDNA has been widely described and associated laboratory methods have been widely

127 reviewed, case greatly democratise which species can be surveyed as

128 sight unseen detection

129 • Behavior of dna in the world

130 • Depth stratification - unveiling a third dimension

131 • Comparison acoustics to dna

132 • Interpretation depends on spatial scale; box of interest

133 • Transboundary; why matters for rules

134 • Rarefaction - how little sample could we do and get same answer?

135 however, in most field applications that have relevance to management or conservation, the scale of interest

136 is not the scale of an individual sample but how multiple samples can be combined to infer the status and

137 trend of large units - a population, a meta-population, an evolutionary significant units, etc. -

- 138 Things that are interesting.
- 139 • 3D plot of hake by depth.
- 140 • methods of eDNA give and explicit estimate of detection thresholds.
- 141 • Comparison acoustics to dna
- 142 • Interpretation depends on spatial scale; box of interest
- 143 • Transboundary; why matters for rules
- 144 • Rarefaction - how little sample could we do and get same answer?

145 **Results**

146 We provide the depth-specific distributions of hake eDNA over 10 degrees of coastal ocean. Patterns of mean
147 hake DNA concentration are distinct with highest concentration observed between 100 and 300m depths along
148 the continental shelf break and south of the Oregon-California border at 42°N. Concentrations at 500m were
149 generally low and showed limited spatial variation while the near surface layers (3m, 50m) showed generally
150 higher concentrations near the coast and in more northerly parts of the range (Figs. 1a, 1b). There was also
151 a large difference in the uncertainty around DNA concentration near the surface; among the predicted grid
152 cells, the median coefficient of variation (CV) was larger than 1 for both 3m and 50m but only about 0.2 for
153 depths 100m and deeper (Fig. 1g). Such uncertainty large observed differences DNA concentration between
154 replicate samples taken at the same sample location and substantial differences among proximate sampling
155 locations. Together, hake DNA concentration was far more predictable at depth than near the surface.

156 We combined DNA information between 50 and 500m to produce a depth-integrated estimate of hake DNA
157 concentration comparable to the acoustic-trawl survey results (Fig. 2). The eDNA index showed strong
158 spatial patterning with highest values along the continental shelf break. In comparison results acoustic-trawl
159 were more spatially variable, with some areas of very high hake density and others with very low density.
160 At the scale of individual 5km grid cells, eDNA and acoustic surveys were positively correlated ($\rho = 0.488$,
161 Pearson product-moment correlation) but there is considerable scatter in the relationship. High eDNA
162 values never occurred at locations which had very low acoustic biomass, but very high acoustic estimates
163 corresponded to moderate values of eDNA. Notably, across all 3,455 grid cells, acoustic biomass estimates
164 had a very right-skewed distribution – most values were near zero with very few high values – while eDNA
165 values were decidedly less skewed (Fig. 2).

166 When aggregated to larger spatial scales, the correlation between eDNA and acoustics increased substantially
167 ($\rho = 0.821$; Fig. 2) with acoustics and eDNA scaling approximately linearly. Such increased correlation is not
168 dependant upon the one degree breaks in Fig. 2 (see Supplement SX for alternate spatial groupings) and
169 shows that in terms of total biomass, eDNA and acoustic-trawls are providing nearly equivalent information
170 about relative spatial abundance. At a coast-wide scale, the CV of the acoustic-trawl estimate and eDNA
171 index were both 0.096 . This similarity in CV occurred despite the eDNA only being collected at 186 point
172 stations whereas the acoustic-trawl data includes 4,841 individual acoustic segments and 45 mid-water trawls
173 to determine age- and length-structure of the hake.

174 Within the projection bounds of this survey, we show very similar estimates of spatial distribution. The
175 methods produced nearly identical estimates of the center of gravity (median of the distributed biomass)
176 and very similar cumulative distributions (Fig. 4). The 90%CI overlapping for the two methos for the entire
177 latitudinal range sampled.

178 **Discussion**

- 179 • Building spatialstatistical model for eDNA is new, methods comparable to acoustics and to trawl
180 surveys
- 181 • Replication of DNA allows partitioning of uncertainty ways that are largely impossible in other methods.
- 182 • DNA is derived from many fewer observations, but provides comparable answers at management scales
- 183 • Comparable answers between do not occur at the local level.
- 184 • DNA is a wholly independent measure of abundance derived from the same amount of ship time.
- 185 • DNA is not in the units of biomass.... which hinders direct interpretation.
- 186 • Many metrics of abundance are like this... (e.g. CPUE)
- 187 • Acoustic and eDNA provide very similar CVs even though eDNA is
- 188 • Need to think about age-structure, other sources of biomass and eDNA.
- 189 Acoustic and eDNA are independently derived estimates. They are both related to the true abundance of
190 fish, but no information from the acoustics informs eDNA or vice versa. The application of eDNA, has, in
191 effect doubled your information about abundance without expanding the amount of ship-time at all.
- 192 There have been some rinky-dink attempts to make joint models for eDNA and other methods [(15);
193 fukaya2020estimating] but they make the mistake of assuming that the area sampled by each method is
194 equivalent (eDNA,individual
- 195 RANDO PARAGRAPH: Despite this debate, most papers that use eDNA implicitly assume that eDNA
196 accurately reflects abundance of the targeted community; all analyses that calculate common diversity or
197 community metrics (e.g. Shannon diversity, Simpson diveristy, Bray-Curtis divergence) from eDNA are
198 making the assumption that observed eDNA accurately reflects the true or at least the relative abundance of
199 species within the sampled community. It has been well documented that this will rarely be true when using
200 (Kelly et al. 2019, Amy's paper, several other.)

201 **Methods**

202 **Hake biology summary.**

203 cite Mike Malick, cite stock assessment, cite Martin Dorn,

204 **Field sampling and sample processing**

205 We collected eDNA samples during the 2019 Joint US-Canada Acoustic trawl survey (16). We collected water
206 from up to six depths (3, 50, 100, 150, 300, and 500m) at 186 CTD stations spread across 36 acoustic transects
207 (Fig. 2). We included 1769 individual 2.5L water samples collected at 892 depth-station combinations (a small
208 number of samples were contaminated or lost during processing). 710 depth-stations were collected at 50m
209 deep or deeper using Niskin bottles. For each Niskin sampled station, two 10L bottles were closed and after
210 CTD retrieval, 2.5L was collected from each bottle. Water samples from 3m were collected from the ship's
211 salt water intake line (MORE DETAILS). In addition we include 49 control samples collected ship board
212 (SEE THIS SECTION) and performed over 6,000 quantitative PCR reactions to produce an unprecedented
213 description of hake DNA in the coastal ocean. All CTD casts and therefore water collection for eDNA
214 occurred at night. Detailed water sampling and processing protocols can be found HERE (SUPPLEMENT?
215 OTHER PAPER?)

216 In parallel with water collection, we incorporate data on hake from the Joint US-Canada acoustic-trawl survey
217 (16). We include data from 4483 km of acoustics transects representing 57 acoustic transects. 45 midwater
218 trawls were deployed that provide information on the age, size and therefore signal strength of hake (14,
219 16). We use derived estimates of biomass concentration ($mt\ km^{-2}$) that integrate the biomass in the water
220 column between depths of 50 and 500m in all analyses. All acoustic data and associated trawls were collected
221 during daylight hours. Therefore there is a lag between collection of acoustic and eDNA data, though for
222 nearly all cases collection were separated by less than 24 hours. Methods for converting raw acoustic and
223 trawl data to biomass concentrations can be found in (13, 14) and references therein.

224 **Water filtration, preservation, and DNA extraction** Overall, we follow the methods developed by
225 and described in Ramon-Laca et al In review. We briefly describe the sample processing steps but refer
226 readers to Ramon-Laca et al. (2021) for a detailed description of methods.

227 **Quantitative PCR methods** Include stuff on wash error here. How many samples, how we partitioned
228 some sampled sites to experimentally wash one sample and not the other.

229 **Controls and contamination**

230 We found very low levels of contamination in control samples. We present examination of the magnitude and
231 consequences of contamination in supplement S2.

232 **Spatial eDNA Model**

233 We developed a Bayesian state-space framework for modeling DNA concentration in the coastal ocean.
234 State-space models separate the true biological process from the methods used to observe the process (see...
235)

236 We use a relatively simple process model. Let D_{xyd} be the true, but unobserved concentration of hake DNA
237 ($DNA\ copies\ L^{-1}$) present at spatial coordinates $\{x, y\}$ (northing and easting, respectively, in km) and
238 sample depth d (meters). We model the DNA concentration as a spatially smooth process at each depth
239 sampled ($d = 3, 50, 100, 150, 300$, or 500m) and linear on the \log_{10} scale,

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

240 where γ_d is the spatial intercept for each depth, $s(b)$ indicates a smoothing spline of as a function of bottom
 241 depth in meters (b), and $t_d(x, y)$ is a tensor-product smooth that provides an independent spatial smooth for
 242 each depth. We use cubic regression splines for both univariate and tensor-product smoothes.

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

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

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

243 where V_i is the proportion of 2.5 L filtered from Niskin i (in nearly all cases this is 1), I_i is the known dilution
 244 used to on sample i to eliminate PCR inhibition, and $\mathbf{I}\omega$ is an estimated offset for a ethanol wash error (ω is
 245 the estimated effect of the wash error and \mathbf{I} is an indicator variable where $\mathbf{I} = 1$ for affected samples and
 246 $\mathbf{I} = 0$ otherwise; see PCR methods below for additional description of each offset).

247 When using qPCR, we do not directly observe eDNA concentration, we observe the PCR cycle at which each
 248 sample can be detected (or if it was never detected during the PCR). We use a hurdle model to account for
 249 the fact that there is a detection threshold (the PCR cycle of amplification is detected $G = 1$ or it is not
 250 $G = 0$). Conditional on being detected, we observe the PCR cycle (C) as a continuous variable that follows a
 251 t-distribution, ADD CONDITIONING BAR FOR THESE EQUATIONS.

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

$$C_{ijr} \sim T(\nu, \beta_{0j} + \beta_{1j} \log_{10} E_{ij}, \eta) \quad (5)$$

252 Here j indexes the PCR plate on which sample i and replicate r were run. We conducted 3 PCR reactions for
 253 each E_i . Note that there are different intercept (ϕ_{0j}, β_{0j}) and slope (ϕ_{1j}, β_{1j}) parameters for the j th PCR
 254 plate to allow for among-plate variation. The t-distribution allows for heavier tails than a normal distribution.
 255 We fix $\nu = 3$, and estimate the remaining parameters.

To calibrate the relationship between the number of DNA copies and PCR cycle, each PCR plate has replicate samples with a known number of DNA copies. These standards span six orders of magnitude (1 to 100,000 copies) and determine the relationship between copy number and PCR cycle of detection. Let K_j be the known copy number in PCR plate j , then, ADD CONDITIONING BAR FOR THESE EQUATIONS.

$$G_{jr} \sim Bernoulli(\phi_{0j} + \phi_{1j} \log_{10} K_j) \quad (6)$$

$$C_{jr} \sim Normal(\beta_{0j} + \beta_{1j} \log_{10} K_j, \sigma) \quad (7)$$

256 where for the standards we do not allow for heavy tails in the observed PCR counts and use a normal
 257 likelihood rather than a t-distributed likelihood.

258 Not that the use of standards provides an explicit model for the detection threshold of eDNA when using
 259 qPCR and this detection threshold can be directly incorporated to understand the ability to measure DNA
 260 concentrations in field samples.

261 Spatial models for acoustic-trawl data

262 In parallel with the model for qPCR data, we estimated a spatial model for the hake biomass derived from
 263 the acoustic-trawl survey. The biomass index created from the acoustic-trawl data for the entire survey area
 264 (34.4°N to 5XX°N;) is used in stock assessments that determine the allowable catch and allocation of hake
 265 catch for fleets from the United States, Canada, and Tribal Nations (14). As the eDNA samples only cover a
 266 portion of this range (38.3°N to 48.6°N), we used the biomass observations from the subset of transects on
 267 which we sampled eDNA to generate spatially smooth estimates of biomass. Acoustic transects are divided
 268 into 0.926km (0.5nm) segments and the biomass concentration within each segment is used as data.

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

$$H_{xy} \sim Bernoulli(\text{logit}^{-1}(\zeta_H + s_H(b) + t_H(x, y))) \quad (8)$$

$$F_{xy} \sim LogNormal(\zeta_F + s_F(b) + t_F(x, y)) - 0.5\kappa^2, \kappa \quad (9)$$

274 where H_{xy} is 1 if the observed biomass concentration is non-zero and 0 otherwise.

275 where ζ is the spatial intercept for each model component, $s(b)$ indicates a smoothing spline of as a function
 276 of bottom depth in meters (b), and $t(x, y)$ is a tensor-products smooth for each model component.

277 Model Estimation

278 We implemented both the eDNA and acoustic-trawl models using the Stan programing language as imple-
 279 mented in R (*Rstan*) language. All relevant code and data are provided in the online supplement. For
 280 the eDNA model, we ran 4 MCMC chains using 1200 warm up and 2000 sampling iterations. For the
 281 acoustic-trawl model, we ran 4 MCMC chains using 1200 warm up and 2000 sampling iterations.

282 We used traceplots and \hat{R} diagnostics to confirm convergence ($\hat{R} < 1.01$ for all parameters). There were no
 283 divergent transitions in the sampling iterations. To generate matrices necessary for estimating smoothes we
 284 used the R package *brms* (REF).

285 Coordinate systems, covariates, and spatial predictions

286 We project both the acoustic-trawl model and the eDNA model to a shared grid to enable direct comparisons
 287 between methods.

288 Reference Blake's 5km grid. We used bottom depth as a covariate. We derived bottom depth from HERE
 289 AND THERE. To ensure comparability between eDNA and acoustics, we project our results to a common
 290 grid (equal area... etc.) and between LAT and LAT. Bottom depth on this grid is depth-integrated across
 291 the entire grid cell.

292 To create a spatial prediction for eDNA we take 2,000 draws from the joint posterior and use eq. 1 to generate
 293 predictions for the centroid of each grid cell and calculate posterior means and uncertainty bounds among
 294 posterior draws. Importantly, as DNA concentration were the values of interest, this projection does not
 295 include projection information about the detailed observation process – replicated Niskin samples from a site
 296 for example.

297 Blake will write a paragraph about area-weighted mean depth used as a covariate.

298 Ref for the 5im grid:

299 (17)

300 Creating an eDNA index

301 Our model provides direct predictions for hake DNA concentration at depths of 50, 100, 150, 300, and 500m.
 302 To produce an index spanning depths of 50 to 500m, we need to equally weight depths between 50 and 500m.
 303 As we lack observations at other depths, we used linear interpolation between the closest depths using the
 304 posterior predictions at each depth to provide predicted DNA densities at 200, 250, 350, 400, and 450m for
 305 each 5km grid cell. For example, a prediction for a grid cell at 400m would be the mean of the predicted
 306 value for that grid cell from a single posterior draw for 300m and 500m. Because some spatial locations have
 307 depths of less than 500m, we only include predicted DNA concentrations to a depth appropriate for the

308 bathymetry (e.g. a location with a depth of 180m only includes values from 50, 100 and 150m). We sum
309 across all depths (between 50 and up to 500m) to generate a depth-integrated index of hake DNA. This index
310 will be proportional to the hake DNA found in the water column. However, as we are only summing across
311 discrete depths, not integrating values across the entire water column nor multiplying by the total water
312 volume within each grid cell, the absolute value of the index will depend upon the number of discrete depths
313 we use. As such, we refer to this as a DNA index to differentiate it from the predictions to a specific depths
314 which have units of copies per liter.

₃₁₅ **Figures**

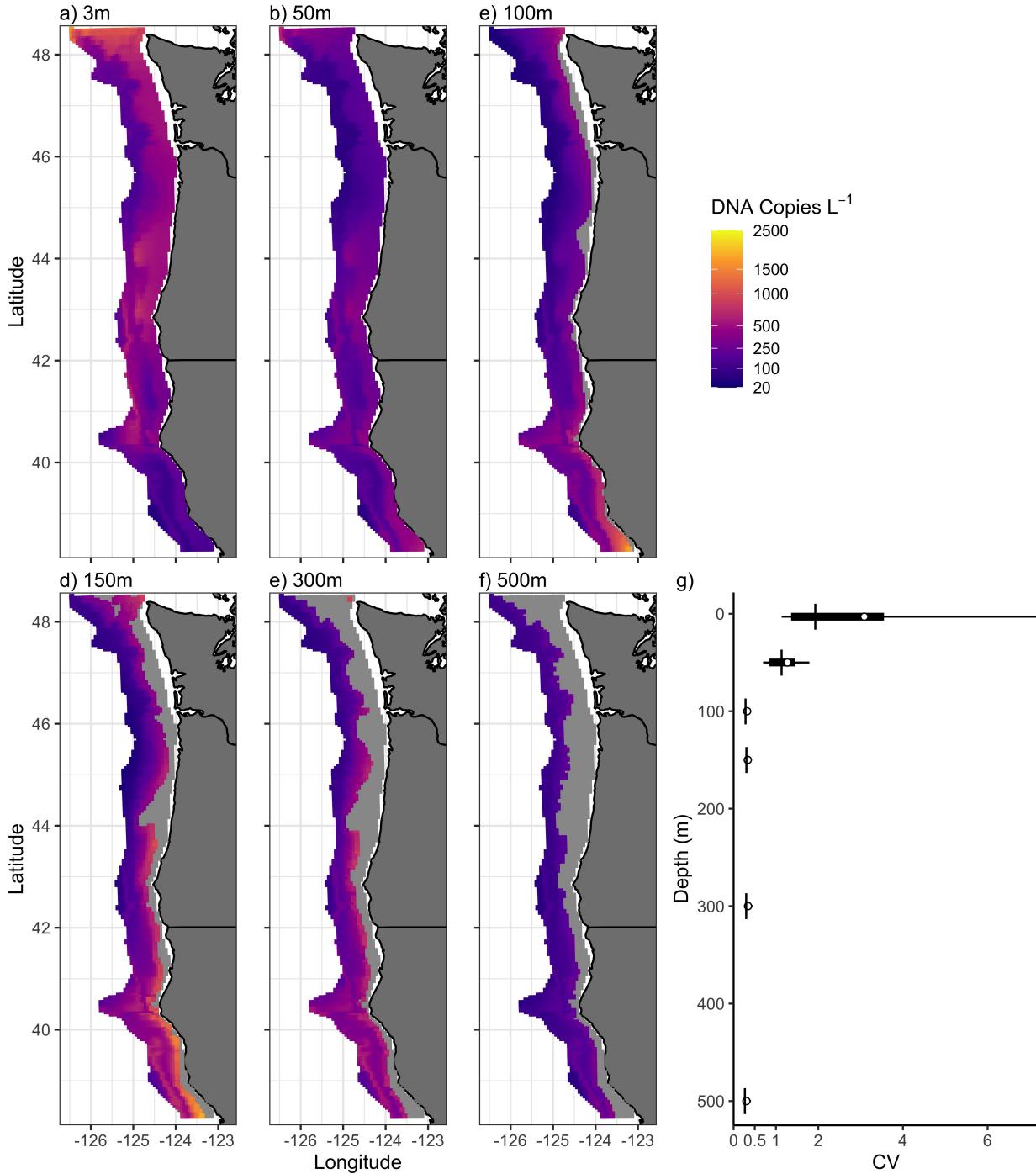


Figure 1: Predicted DNA concentration for six water depths.

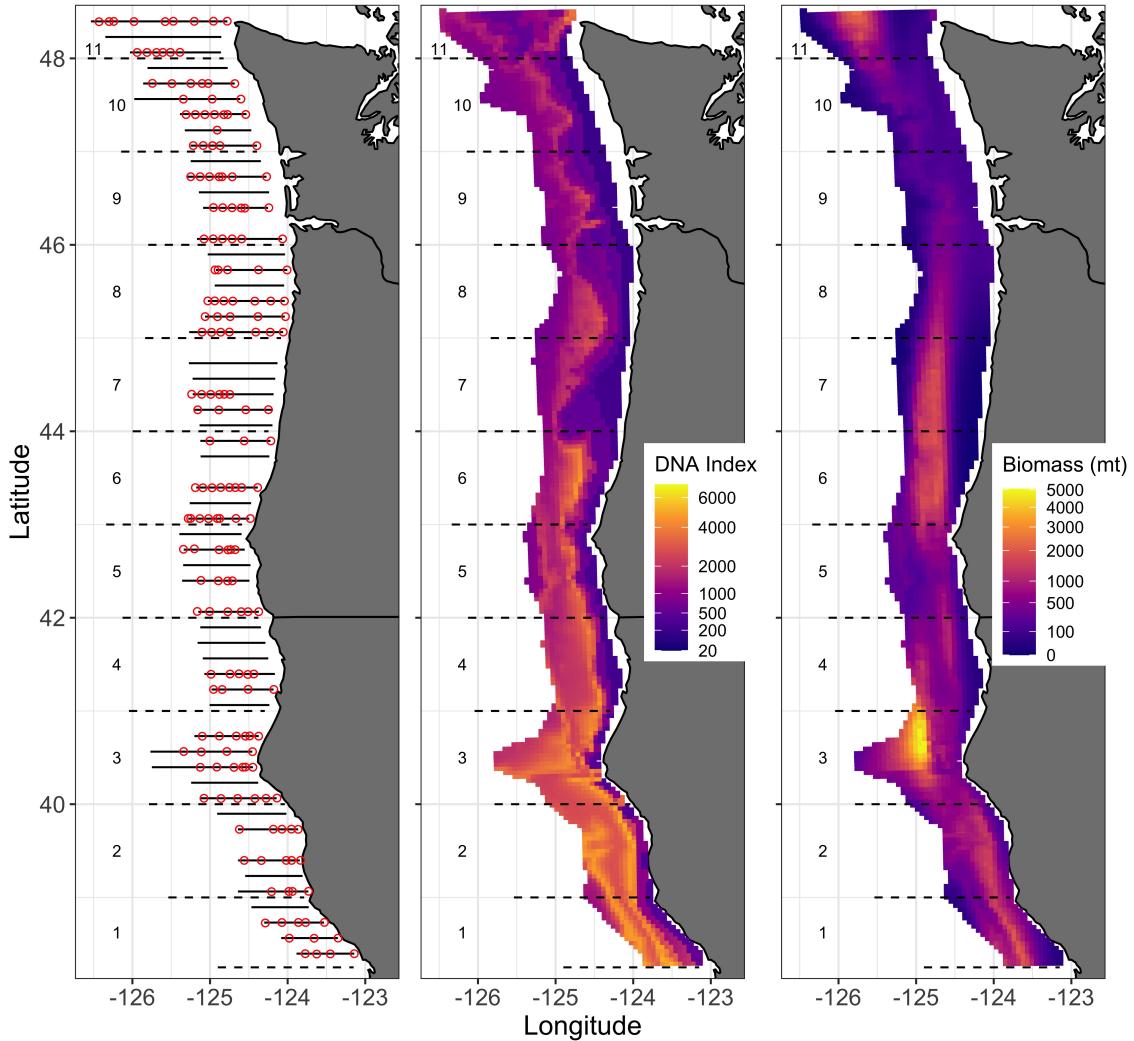


Figure 2: 2019 survey locations (left; red circles show eDNA sampling locations, lines show acoustic transects), depth-integrated index of hake DNA (middle) and hake biomass from acoustic surveys (right). Both DNA and acoustic estimates are mean predicted values projected to a 5km grid and include information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

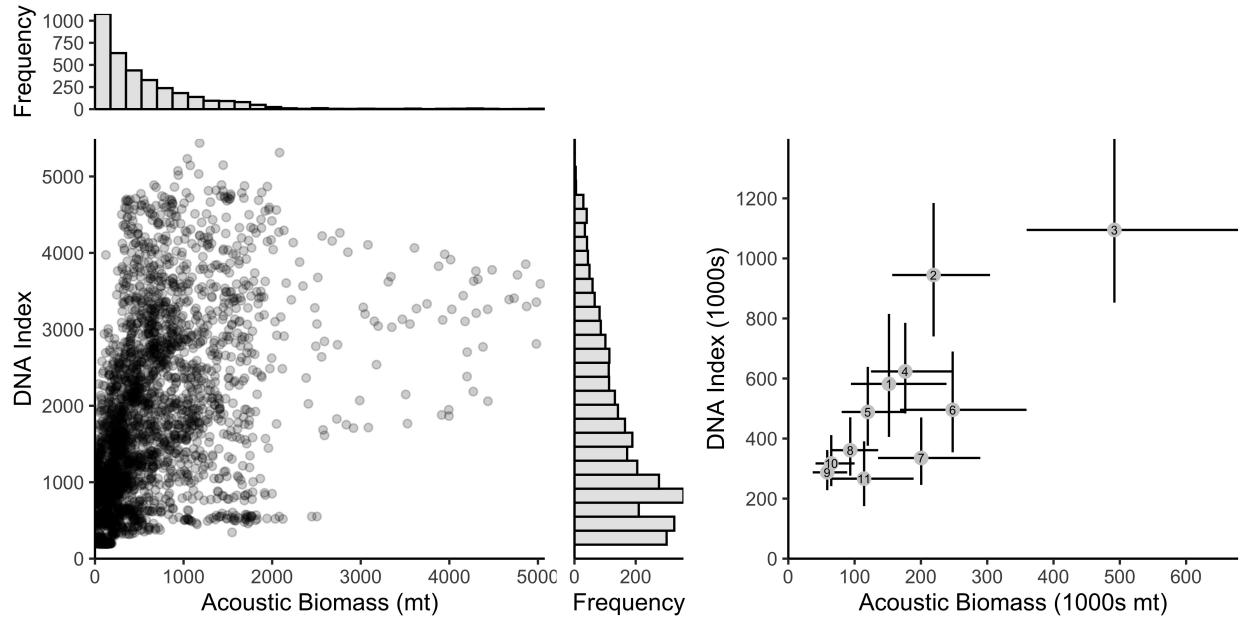


Figure 3: Pairwise comparison between DNA and acoustics-derived biomass. Left panel show the posterior mean prediction from each method among the 3,455 25km² grid cells and includes the marginal histogram of posterior mean values for each method ($RHO = 0.488$). Right panel shows correlation between methods among the 11, one degree bins (posterior mean, 90 CI; $RHO = 0.8208$). Numbers indicate associated region identified in Fig. 2.

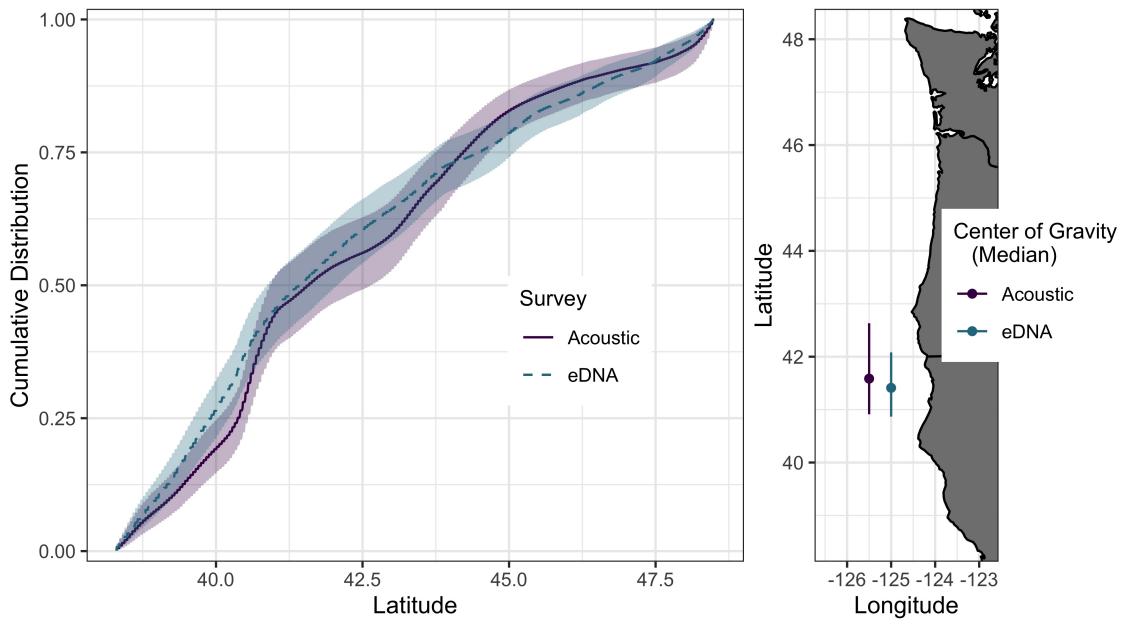


Figure 4: Estimates of distribution from acoustic and eDNA methods. left: Cumulative distribution between 38.3 and 48.6N (posterior means and 90 CI). Right: Center of gravity (median of distribution) for each method (posterior means and 90 CI). Only areas within the projection grid are included in this calculation (Figs. 1, 2) are included.

316 **Citations**

- 317 1. H. Murakamiet al., Dispersion and degradation of environmental dna from caged fish in a marine
318 environment. *Fisheries science* **85**, 327–337 (2019).
- 319 2. P. F. Thomsen, E. Willerslev, Environmental DNA—an emerging tool in conservation for monitoring past
320 and present biodiversity. *Biological Conservation* **183**, 4–18 (2015).
- 321 3. M. L. Rourkeet al., Environmental dna (eDNA) as a tool for assessing fish biomass: A review of approaches
322 and future considerations for resource surveys. *Environmental DNA* (2021).
- 323 4. J. B. Harrison, J. M. Sunday, S. M. Rogers, Predicting the fate of eDNA in the environment and implications
324 for studying biodiversity. *Proceedings of the Royal Society B* **286**, 20191409 (2019).
- 325 5. L. Carraro, H. Hartikainen, J. Jokela, E. Bertuzzo, A. Rinaldo, Estimating species distribution and
326 abundance in river networks using environmental dna. *Proceedings of the National Academy of Sciences* **115**,
327 11724–11729 (2018).
- 328 6. A. O. Sheltonet al., Environmental dna provides quantitative estimates of a threatened salmon species.
329 *Biological Conservation* **237**, 383–391 (2019).
- 330 7. K. Fukayaet al., Estimating fish population abundance by integrating quantitative data on environmental
331 dna and hydrodynamic modelling. *Molecular ecology* (2020).
- 332 8. J. A. Portet al., Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA.
333 *Molecular Ecology* **25**, 527–541 (2016).
- 334 9. M. D. Tillotsonet al., Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance
335 at fine spatial and temporal scales. *Biological Conservation* **220**, 1–11 (2018).
- 336 10. B. Hänflinget al., Environmental dna metabarcoding of lake fish communities reflects long-term data
337 from established survey methods. *Molecular ecology* **25**, 3101–3119 (2016).
- 338 11. M. Y. Stoeckleet al., Trawl and eDNA assessment of marine fish diversity, seasonality, and relative
339 abundance in coastal new jersey, usa. *ICES Journal of Marine Science* **78**, 293–304 (2021).
- 340 12. M. A. Barnes, C. R. Turner, The ecology of environmental dna and implications for conservation genetics.
341 *Conservation genetics* **17**, 1–17 (2016).
- 342 13. M. J. Malicket al., Relationships between temperature and pacific hake distribution vary across latitude
343 and life-history stage. *Marine Ecology Progress Series* **639**, 185–197 (2020).
- 344 14. C. J. Grandin, K. F. Johnson, A. W. Edwards, A. M. Berger, “Status of the pacific hake (whiting) stock
345 in u.s. And canadian waters in 2020.” (Prepared by the Joint Technical Committee of the U.S.; Canada
346 Pacific Hake/Whiting Agreement, National Marine Fisheries Service; Fisheries; Oceans Canada, 2020).
- 347 15. T. Chambert, D. S. Pilliod, C. S. Goldberg, H. Doi, T. Takahara, An analytical framework for estimating
348 aquatic species density from environmental dna. *Ecology and evolution* **8**, 3468–3477 (2018).
- 349 16. S. de Blois, The 2019 joint u.s.–Canada integrated ecosystem and pacific hake acoustic-trawl survey:
350 Cruise report sh-19-06. *U.S. Department of Commerce, NOAA Processed Report NMFS-NWFSC-PR-2020-03*
351 (2020).
- 352 17. B. E. Feist, J. F. Samhouri, K. A. Forney, L. E. Saez, Footprints of fixed-gear fisheries in relation to
353 rising whale entanglements on the us west coast. *Fisheries Management and Ecology* **28**, 283–294 (2021).