
Accounting for uncertainty when estimating drivers of imperfect detection: an integrated approach illustrated with snorkel surveys for riverine fishes

BENJAMIN A. STATON*

Fisheries Science Department, Columbia River Inter-Tribal Fish Commission, NE
Multnomah St., Ste. 1200, Portland, OR 97232

bstaton@critfc.org

CASEY JUSTICE

Fisheries Science Department, Columbia River Inter-Tribal Fish Commission, NE
Multnomah St., Ste. 1200, Portland, OR 97232

SETH WHITE

Fisheries Science Department, Columbia River Inter-Tribal Fish Commission, NE
Multnomah St., Ste. 1200, Portland, OR 97232

EDWIN R. SEDELL

East Region Fish Research, Oregon Department of Fish and Wildlife, One University
Blvd., La Grande, OR 97580

LAUREN A. BURNS

Fisheries Science Department, Columbia River Inter-Tribal Fish Commission, NE
Multnomah St., Ste. 1200, Portland, OR 97232

MATTHEW J. KAYLOR

Department of Fisheries and Wildlife, Oregon State University, 2820 SW Campus Way,
Corvallis, OR 97331

Keywords: abundance estimation, Bayesian analysis, detection probability, hierarchical models, snorkel surveys, variable selection

Manuscript accepted for publication in *Fisheries Research* on Dec. 20, 2021

DOI: [10.1016/j.fishres.2021.106209](https://doi.org/10.1016/j.fishres.2021.106209)

©2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

*Corresponding Author

1 Abstract

2 Imperfect detection is a common issue affecting the accuracy of surveys that quantify
3 animal abundance and distribution. To quantify detectability, counts are often calibrated
4 to independent measures of abundance (e.g., via mark-recapture) but stochastic sampling
5 variability in both data types is not typically accounted for. This practice may cause
6 detectability to be quantified inaccurately and lead to overly confident predictions for
7 out-of-sample applications. Our objective was to develop, apply, and simulation-test an
8 integrated approach for quantifying detectability that better-accommodates uncertainty in
9 the data. The method assumes mark-recapture and count surveys sample the same local
10 abundance with error, allowing the construction of a joint likelihood function for both
11 data sets. The model estimates coefficients that link detection probability to local
12 covariates through a logit-linear model, which enables correcting counts for imperfect
13 detection in locations where mark-recapture data are unavailable. We illustrate the
14 application of the model with an empirical data set of over 100 paired snorkel count and
15 mark-recapture electrofishing surveys for riverine salmonids in northeastern Oregon.
16 Covariates that best explained heterogeneity in detectability included species, visibility,
17 and channel unit type and depth, though substantial variability was attributed to
18 site-level random effects. Estimated detection probability ranged from 0.02 to 0.92 among
19 surveys and was higher for Chinook Salmon (*Oncorhynchus tshawytscha*) juveniles (mean:
20 0.38) than for steelhead/Rainbow Trout (*O. mykiss*; mean: 0.24). Simulation analyses
21 revealed that our integrated model performed better (relative to a method that treated
22 mark-recapture abundance estimates as known without error) with respect to (i) selection
23 of covariates, (ii) credible interval coverage, (iii) accuracy of estimated random variability
24 terms, and (iv) reduced sensitivity to violated mark-recapture assumptions surrounding
25 behavioral effects. This model represents an improvement over simpler calibration
26 methods, particularly for snorkel surveys, by applying more a rigorous statistical
27 treatment of sources of variability while explicitly describing the mechanistic link between

28 local conditions and detectability. The analytical methods we illustrate are general and
29 could be broadly applied to quantify detectability in other biological surveys with paired
30 abundance and count data.

31 **1 Introduction**

32 Spatial and temporal patterns in population abundance are key information sources for
33 understanding the ecology of fish populations and forming the knowledge base for
34 management and conservation. For example, quantifying abundance is critical for
35 assessing the efficacy of preserves (Mosquera et al. 2000), measuring population response
36 to restoration actions (Roni et al. 2008), and setting harvest quotas (Walters and Pearse
37 1996). Complete population censuses are often infeasible and therefore abundance is
38 typically sampled using counts, which may be visual or capture-based in nature. Such
39 counts are imperfect measures of population abundance because they result from
40 imperfect detection, i.e., a surveyor can either count or fail to count individuals in the
41 population, making the probability of successfully counting any one individual (termed
42 “detection probability”) less than one. Imperfect detectability (also known as catchability
43 for capture-based methods) introduces a directional bias that underestimates true
44 abundance (Kéry and Schmidt 2008; Kellner and Swihart 2014). Further, the use of count
45 data as an index of abundance (or other indices, such as catch-per-effort) assumes that
46 detectability is constant across individuals, sampling locations, and time periods.
47 Detectability may not be constant and the same factors that influence its spatiotemporal
48 variability may also affect local abundance (e.g., density of large wood for riverine fishes).
49 Thus, quantifying detectability and the factors that influence its spatiotemporal variability
50 are critical steps for accurately estimating population abundance from count data.

51 Methods commonly used to quantify detectability include those that assume some
52 attenuation with increasing distance from the observer (distance sampling, Buckland et al.
53 2001), use spatiotemporally replicated counts (N-mixture models, Royle 2004) or
54 presence/absence data (Royle and Nichols 2003), or use the frequency with which marked
55 individuals are subsequently recaptured (mark-recapture methods, Seber 1986). Another
56 method, sometimes referred to as “double sampling” (Williams et al. 2002, Section 12.6
57 therein) or “dual-gear validation” (e.g., Thurow et al. 2006), uses intensively monitored

58 sites to estimate total abundance which, when paired with independent counts, provide
59 an estimate of detection probability that may be applied to sites with count data only to
60 estimate abundance. This practice of pairing independent abundance and count data has
61 been widely used for quantifying detectability or otherwise obtaining count correction
62 factors for aquatic and terrestrial species. Examples include point counts paired with
63 census counts in nesting bird survey plots (Bart and Earnst 2002), spotlight counts of
64 crocodiles paired with mark-recapture (Hutton and Woolhouse 1989), spotlight counts of
65 deer paired with abundance from thermal imagery (Collier et al. 2007), aerial counts of
66 salmon paired with mark-recapture (Jones et al. 1998), and snorkel counts of fish paired
67 with abundance estimates from mark-recapture (Jonasson et al. 2016), electrofishing
68 (Thurow et al. 2006), or removal methods (Hankin and Reeves 1988; Hillman et al. 1992).

69 Given that detectability quantified using the double sampling design is generally
70 conducted on a small subset of sites and then applied to other locations, it is critical that
71 relationships with covariates and their out-of-sample uncertainty be accurately
72 characterized. To our knowledge, none of the examples of the double sampling design
73 above accounted for sampling variability in the abundance information used in the
74 analysis (i.e., stochasticity in sampling that creates uncertainty in resulting estimates). For
75 example, Thurow et al. (2006) used beta-binomial regression to quantify covariate effects
76 on snorkel survey detectability wherein abundance estimates derived from
77 efficiency-adjusted catch rates were treated as the binomial sample sizes and snorkel
78 counts were treated as successes. The abundance information that enables quantifying
79 detectability from a model like this has uncertainty, yet it was assumed to be a perfectly
80 known number of fish.

81 Ignoring sampling variability in the data that inform abundance estimates in double
82 sampling designs likely inserts additional noise into covariate relationships with
83 detectability, which could hinder our ability to identify optimal predictive models.
84 Further, unaccounted variability in abundance data is likely to inappropriately weight

85 observations during model fitting and result in overstated confidence in
86 estimated/predicted detection probability. These issues may be problematic given the
87 central role served by covariate relationships with detectability: they enable comparisons
88 of count data across space and time. Thus, it may be preferable to analyze data generated
89 from double sampling designs using a method that explicitly acknowledges abundance is
90 unknown.

91 Snorkel surveys are commonly used to assess fish abundance and distribution as
92 part of routine monitoring programs (e.g., Constable and Suring 2015; Fowler 2017) and
93 directed ecological studies (Som et al. 2018; Flitcroft et al. 2013). They are frequently used
94 in riverine habitats for salmonid fishes, both for juveniles (Thompson and Lee 2000;
95 Flitcroft et al. 2013; Som et al. 2018) and adults (Pinter et al. 2018; Korman et al. 2002;
96 Thurow et al. 2006) but have also been used for other fish taxa and in non-riverine
97 environments (e.g., Lawson et al. 2011; Weaver et al. 2014; Ulibarri et al. 2017). Counts
98 obtained from these surveys can inform fine-scale patterns of fish-habitat associations
99 unavailable from larger-scale monitoring methods alone (Som et al. 2018). Further,
100 because fish are never handled during snorkel surveys, these methods are more rapid and
101 less intrusive than other survey methods (e.g., multi-pass electrofishing), making them
102 well-suited for monitoring the abundance and distribution of threatened species at large
103 spatial scales (O'Neal 2007).

104 Given the difficult nature of counting small and mobile organisms in flowing water,
105 snorkel surveys are subject to imperfect detection and the variability in riverine habitat
106 types in which they are applied suggests that detectability is not spatially homogeneous.
107 Thus, snorkel counts should be corrected for heterogeneity in detectability before they can
108 be interpreted as reliable indicators of abundance. The double sampling approach has
109 been used widely to quantify detectability in snorkel surveys (Hankin and Reeves 1988;
110 Hillman et al. 1992; Thurow et al. 2006; Jonasson et al. 2016). N-mixture models have
111 recently been used to address this issue (Som et al. 2018; Doll et al. 2020), but have not yet

112 been widely applied for snorkel surveys and may not be ideal for all cases given the need
113 for within-site replication (which could otherwise be distributed spatially) and influential
114 assumptions that are easily violated (i.e., that abundance and detection probability are
115 constant among repeated visits to the same site, Barker et al. 2017). Thus, it would be
116 desirable to have a method for quantifying detectability that (a) is suitable for commonly
117 collected data sets (i.e., following a double sampling design), (b) accommodates
118 uncertainty in independent abundance estimates, (c) allows direct modeling and selection
119 of local covariates on detectability, and (d) propagates uncertainty in estimated
120 relationships to out-of-sample applications.

121 We present an analytical method for quantifying detectability from double sampling
122 designs in snorkel and other biological surveys intended to meet these criteria. Our
123 method jointly estimates abundance and covariate effects on detectability within a single
124 model rather than analyzing them sequentially. After describing the statistical structure
125 and its assumptions, we illustrate an application of the method using empirical snorkel
126 survey data collected in northeastern Oregon and use a simulation study to test its validity
127 in the face of violated assumptions and varying data quality and quantity.

128 **2 Methods**

129 **2.1 General data structure**

130 The method we describe for quantifying detectability requires paired abundance and
131 count data from a variety of locations and associated variables that may influence the ease
132 or difficulty of counting animals in each location. Our specific illustration assumes that
133 abundance data take the form of a two-sample mark-recapture design but other designs
134 could be accommodated by our approach. If possible, sampling locations should be
135 selected at the smallest spatial scale of interest to provide the highest resolution for
136 investigating the effects of local conditions on detectability. For example, riverine
137 sampling designs are often created hierarchically: channel units (i.e., individual pools,
138 riffles, or runs) are nested within a site, sites are nested within a tributary, and tributaries

139 are nested within a basin – in this case, channel units should be selected as the
140 observational unit for conducting double sampling.

141 **2.2 Modeling framework**

142 *2.2.1 Sub-model for detection probability*

143 Our goal was to develop a model to describe how detection probability (denoted by ψ_i) for
144 each observation (i ; a unique pair of abundance and count data) varies as a function of
145 covariates. We define ψ_i as the probability that each member of the population
146 (abundance denoted N_i) is counted during the count survey. Thus, given known N_i and
147 count (y_i), a standard way to estimate the effects of covariates $x_{1,i}, \dots, x_{n,i}$ on ψ_i is via
148 logistic regression methodology:

$$\begin{aligned} \text{logit}(\psi_i) &= \alpha + \beta_1 x_{1,i} + \dots + \beta_n x_{n,i} + \varepsilon_{j(i)}, \\ \varepsilon_j &\sim \text{N}(0, \sigma^2) \end{aligned} \tag{1}$$

149 where α is the intercept, β_1, \dots, β_n are coefficients quantifying covariate effects on the log
150 odds of detection, and ε_j is a random effect for level j of a blocking factor, and $j(i)$ denotes
151 the specific level that observation i belongs to (not all data sets will have a random
152 blocking factor). Logistic regression uses the assumption of binomial sampling:

$$y_i \sim \text{B}(\psi_i, N_i), \tag{2}$$

153 to build the likelihood function. However, this requires an abundance value N_i , and rarely
154 is this quantity known with certainty. We therefore constructed an integrated model that
155 uses this same logistic regression framework to model covariate effects on ψ_i , but
156 simultaneously estimates N_i from mark-recapture data and thus acknowledges its
157 uncertainty jointly with the uncertainty in count survey sampling. We chose to employ the
158 Bayesian inferential framework because it is useful for (a) hierarchical and integrated
159 models, (b) propagating uncertainty and parameter correlations to derived quantities, and
160 (c) estimation of posterior model probabilities for model-averaging purposes.

161 2.2.2 *Sub-model for abundance*

162 We designed the model to accommodate data collected by a two-sample mark-recapture
 163 design, in which the data for survey i are counts of the three observable capture histories;
 164 these include the number of individuals captured and marked in the first period and
 165 captured again in the second period ($z_{i,11}$), the number of individuals captured and
 166 marked in the first period that were not captured in the second period ($z_{i,10}$), and the
 167 number of individuals captured for the first time in the second period ($z_{i,01}$). For one
 168 observation, we organize these counts into the vector \mathbf{z}_i :

$$\mathbf{z}_i = [z_{i,11} \quad z_{i,10} \quad z_{i,01}]. \quad (3)$$

169 We used the conditional likelihood expression (Huggins 1989, 1991) that removes the
 170 unobservable capture history ($z_{i,00}$) from the calculation such that this three-element
 171 vector can be fitted assuming multinomial sampling conditional only on individuals
 172 captured in the study (more details in Williams et al. 2002, Section 14.1 therein). Consider
 173 the case where all individuals in the first and second periods have capture probabilities
 174 denoted by $p_{i,1}$ and $p_{i,2}$, respectively. The probability that a member of N_i was captured at
 175 all (denoted p_i^*) is then:

$$p_i^* = 1 - (1 - p_{i,1})(1 - p_{i,2}), \quad (4)$$

176 which can be used to create the vector ($\boldsymbol{\pi}_i$) of expected capture history probabilities:

$$\begin{aligned} \pi_{i,11} &= \frac{p_{i,1}p_{i,2}}{p_i^*} \\ \pi_{i,10} &= \frac{p_{i,1}(1 - p_{i,2})}{p_i^*} \\ \pi_{i,01} &= \frac{(1 - p_{i,1})p_{i,2}}{p_i^*}. \end{aligned} \quad (5)$$

177 This probability vector forms the expected proportions of observed capture histories, and
 178 is used in a multinomial likelihood by assuming:

$$\mathbf{z}_i \sim \text{M}(\boldsymbol{\pi}_i, r_i) \quad (6)$$

179 where $r_i = z_{i,11} + z_{i,10} + z_{i,01}$. Then, abundance can be derived as:

$$N_i = \frac{r_i}{p_i^*} \quad (7)$$

180 We embedded these calculations and their likelihood in the same model that includes the
 181 count survey data and logistic regression such that all unknowns could be simultaneously
 182 estimated (see Supplement A for integrated model pseudo-code written in the JAGS
 183 language).

184 Otis et al. (1978) introduced a family of mark-recapture models for closed population
 185 size estimation that vary in their assumptions about capture probabilities, and three of
 186 these models are relevant to the two-sample design we implemented. These models are
 187 commonly referred to as M_0 , M_t , and M_b (Table 1). Model M_0 assumes that $p_{i,1} = p_{i,2}$
 188 whereas model M_t assumes that $p_{i,1} \neq p_{i,2}$ and is equivalent to the well-known
 189 two-sample Lincoln-Peterson estimator. Model M_b assumes a behavioral response to
 190 capture and marking, such that capture probability in the second period depends on
 191 whether the individual was previously captured. Model M_b redefines $p_{i,2}$ to be the
 192 probability a fish is captured for the first time in the second period (and assumes
 193 $p_{i,1} = p_{i,2}$) and defines the recapture probability as $c_{i,2}$. For model M_b , the calculation to
 194 obtain π_i must be modified to accommodate these new definitions:

$$\begin{aligned} \pi_{i,11} &= \frac{p_{i,1}c_{i,2}}{p_i^*} \\ \pi_{i,10} &= \frac{p_{i,1}(1 - c_{i,2})}{p_i^*} \\ \pi_{i,01} &= \frac{(1 - p_{i,1})p_{i,2}}{p_i^*}, \end{aligned} \quad (8)$$

195 where the definition of p_i^* remains the same as in eq. 4. If behavioral effects are present,
 196 they can create large errors in abundance estimates when unaccounted for (e.g., Peterson
 197 et al. 2015). We thus designed our modeling approach to accommodate behavioral
 198 responses if evidence suggests they exist. We first assessed which model assumption (M_0 ,
 199 M_t , or M_b) was most supported by all observations with mark-recapture data only – we
 200 used the Watanabe-Akaike information criterion (WAIC) for this purpose (Watanabe 2013;

201 Hooten and Hobbs 2015). Although discerning between M_t or M_b is problematic for one
202 \mathbf{z}_i vector alone, by imposing the same assumption on all available \mathbf{z}_i vectors
203 simultaneously and considering their likelihoods jointly, we were able to select the
204 mark-recapture model that is most supported by the accumulation of multiple repeated
205 mark-recapture studies (see the results of simulation block H presented in Supplement A
206 section A2.2.8 for a verification of this). We then fitted the integrated model that includes
207 both count and mark-recapture data using the mark-recapture assumptions deemed most
208 appropriate by WAIC.

209 2.2.3 Variable selection and model-averaging

210 Identifying which covariates are useful for explaining variability in detection probability
211 (ψ_i) in eq. 1 is a key step in building a model that allows robust predictions for locations
212 with count data only. Our model accounts for model uncertainty in this regard with
213 additional parameters that enable estimation of the probability that each covariate should
214 be included in an optimal predictive model. These “indicator variables” (Hooten and
215 Hobbs 2015), denoted by $\omega_1, \dots, \omega_n$, were Bernoulli random variables that, when
216 multiplied by each of the β coefficients, had the purpose of toggling on ($\omega = 1$) or off
217 ($\omega = 0$) the effect of each covariate on the log odds of detection. This enabled evaluating
218 multiple models during the fitting process (Kuo and Mallick 1998; Hooten and Hobbs
219 2015) while providing the posterior model probabilities needed for Bayesian multi-model
220 inference. The posterior mean of each ω term represents the probability that the
221 corresponding covariate is a member of the optimal predictive model, with values greater
222 than 0.5 indicating that the covariate should be included (Barbieri and Berger 2004).
223 Model-averaged posteriors are calculated from $\omega\beta$ rather than β and posterior model
224 probabilities are calculated from the relative frequency with which the various ω terms are
225 jointly equal to 1 or 0.

226 2.2.4 Model assumptions

227 Our model makes simplifying assumptions to enable estimation, and for our description
228 we divide these into three classes: those that apply to (a) random processes for
229 mark-recapture sampling, (b) random processes for count sampling, and (c) the joint
230 analysis of mark-recapture and snorkel data. All mark-recapture models we used assume
231 (i) the population is closed between the first and second periods, (ii) capture/non-capture
232 outcomes are the result of independent events across individuals, (iii) all individuals
233 within a period have homogeneous (re)capture probabilities, (iv) there is no mark loss, and
234 (v) marked and unmarked individuals are correctly identified as such. Mark-recapture
235 models M_0 , M_t , and M_b all make different assumptions about how capture probability
236 varies between periods and whether there is a behavioral effect of previous capture
237 (described above, also Table 1). Assumptions of the binomial model as the
238 count-generating process in equation 2 include (i) all members of N_i have homogeneous
239 ψ_i , (ii) detection/non-detection outcomes are the result of independent events across
240 individuals, and (iii) individuals are counted a maximum of one time and are correctly
241 identified (e.g., to species). Integrating the two data types into a joint analysis introduces
242 two additional assumptions. First, this requires the assumption that sampling variability
243 is independent between mark-recapture and count surveys, e.g., negative errors in
244 abundance estimation are not associated with higher- or lower-than-expected outcomes
245 from the count survey. Second, we must assume that the population abundance N_i is
246 unchanged in the time elapsed between mark-recapture and count surveys. This
247 assumption is less strict than complete closure as it does not assume no movement, but
248 rather that any movement of individuals in or out of the study area between surveys
249 results in net zero change on average across observations.

250 2.2.5 Computation

251 We fitted the model using Bayesian methods implemented with Markov chain Monte
252 Carlo (MCMC) sampling using program JAGS (Plummer 2003) invoked through program

253 R (R Core Team 2020) with the ‘jagsUI’ package (Kellner 2018). The R packages ‘postpack’
254 (Staton 2020b) and ‘posterior’ (Bürkner et al. 2020; Vehtari et al. 2021) were used for
255 posterior summarization and MCMC sampling diagnostics, respectively. We selected prior
256 distributions to be minimally informative while simultaneously discouraging implausible
257 values from being sampled by the MCMC algorithm. We discuss MCMC sampling,
258 diagnostics, and priors in greater detail in Supplement A and provide JAGS model
259 pseudocode therein as well. All code and data necessary to replicate the specific analyses
260 described below are provided in Staton (2020a).

261 **2.3 Empirical application**

262 *2.3.1 Study system*

263 Data were collected in the Grande Ronde River basin which originates in the Blue
264 Mountains of northeastern Oregon and flows northward to its confluence with the Snake
265 River (Fig. 1). Local salmon populations have declined precipitously from their historical
266 abundance (Nehlsen et al. 1991), prompting the listing of Snake River spring/summer-run
267 Chinook Salmon (*Oncorhynchus tshawytscha*) and Snake River steelhead (*O. mykiss*) as
268 threatened under the Endangered Species Act. Substantial resources have been invested in
269 restoring freshwater habitat conditions in this region, and snorkel surveys have been used
270 widely as a method for status and trends monitoring and evaluation of restoration
271 effectiveness (e.g., White et al. 2019).

272 *2.3.2 Field data collection*

273 The study design and details of data collection are fully described in Jonasson et al. (2016)
274 and we therefore provide only a brief overview here. Sampling occurred in the summers
275 of 2012 and 2015 and spanned a range of habitat types and stream sizes in the upper
276 Grande Ronde River, Catherine Creek, and their various tributaries (Fig. 1). Habitats were
277 classified into channel units (i.e., pools, riffles, and runs) based on differences in bedform
278 shape, gradient, and substrate size (CHaMP 2016). These channel unit delineations
279 provided the spatial scale at which snorkel surveys were conducted following White et al.

280 (2012), which is based on previously established protocols (Thurow 1994; O'Neal 2007).
281 Essentially, a snorkeler entered each channel unit at the downstream end, identifying fish
282 species and enumerating their abundance as they proceeded methodologically upstream.
283 Two snorkelers were used in wider stream channels, in which case snorkelers
284 communicated with each other during the survey to avoid duplicate counting. Data were
285 temporarily recorded on an underwater slate and after the entire channel unit was
286 surveyed, snorkelers permanently recorded the counts by species and the quality of
287 visibility underwater as poor, average, or good based on a number of factors including
288 water clarity, shade, and obstructions to line of sight. Target species were Chinook Salmon
289 and steelhead/Rainbow Trout.

290 Prior to mark-recapture sampling, block nets were installed spanning the upstream
291 and downstream ends of the channel unit to ensure population closure between capture
292 events (Peterson et al. 2005). Fish were captured with backpack electrofishing in both
293 events and marked using fin clips in the first event. We limited the data set to include only
294 observations with valid data including non-zero marked and recaptured fish and paired
295 snorkel and mark-recapture data that occurred within a two-day time frame (generally,
296 snorkel surveys were conducted prior to mark-recapture sampling). This left 136 total
297 observations (unique species by channel unit visit combinations) which included 100
298 unique channel units (45 and 59 channel units sampled in 2012 and 2015, respectively;
299 four units were sampled in both years) across 44 unique sites; 38 of the observations were
300 of Chinook Salmon and 98 were of *O. mykiss*.

301 We collected additional covariates intended to describe fish behavior and local
302 habitat conditions that we hypothesized could influence detectability of our snorkel
303 surveys. These covariates were measured using slight modifications to the CHaMP (2016)
304 protocol and included: channel unit type classification (pool versus non-pool), average
305 unit depth (m), density of large wood (greater than 15 cm in width and 3 m in length) in
306 the wetted stream channel (pieces \cdot m⁻²), snorkeler-determined quality of visibility (poor,

307 average, good), and species of observation.

308 2.3.3 Analysis

309 The model we applied to explain variability in snorkel survey detection probability had
310 the form:

$$\begin{aligned} \text{logit}(\psi_i) = & \alpha + \omega_1\beta_1\text{Chinook}_i + \omega_2\beta_2\text{Pool}_i \\ & + \omega_3\beta_3\text{LWD2}_i + \omega_4\beta_4\text{LWD3}_i \\ & + \omega_5\beta_5\text{VIS1}_i + \omega_6\beta_6\text{VIS3}_i \\ & + \omega_7\beta_7\text{Depth}_i + \omega_8\beta_8\text{Depth}_i \times \text{Pool}_i + \varepsilon_{j(i)}. \end{aligned} \quad (9)$$

311 All variables were binary except Depth_i , which was continuous and z-transformed prior
312 to analysis. We coded Chinook_i as 1 for Chinook Salmon and 0 for *O. mykiss*; likewise to
313 discern pool and non-pool units in the case of Pool_i . We categorized large wood density
314 because samples were not spread evenly along the observed continuum: we coded LWD2_i
315 as 1 if large wood density was non-zero but less than the median density of all non-zero
316 large wood density observations ($0.026 \text{ pieces} \cdot \text{m}^{-2}$) and 0 otherwise; likewise for LWD3_i
317 but for large wood densities greater than the median density. We coded VIS1_i as 1 for poor
318 visibility and 0 otherwise; likewise for VIS3_i but for good visibility. Zero wood density
319 (LWD1_i) and average visibility (VIS2_i) were the baseline categories and were quantified by
320 the intercept (α). Coding the covariates this way and assigning ω terms to each enabled us
321 to evaluate models where, for example, high wood density has an effect relative to no
322 wood but low wood density has no effect. We examined model residuals to determine
323 whether sampling year (2012 vs. 2015) needed to be included as a detectability covariate,
324 but found no heterogeneity in residuals across years (Supplement A, Fig. A3), and
325 excluded it from further consideration given it would not be useful for predictions in
326 other years. To account for potential spatial dependencies caused by factors other than
327 those explained by fixed effects, we used a random effect for site (j) such that observations
328 (i) made in channel units at the same site shared a common term (ε_j).

329 2.4 Simulation analysis

330 Stochastic simulation provides an opportunity to test the ability of the model to return
331 robust estimates because they can be compared to true and known parameters. We
332 designed the simulation analysis to address several questions:

333 (1) Can the integrated model return unbiased estimates of true detection probability
334 and abundance for both training data (locations with double sampling) and
335 prediction data (locations with count and covariate data only)?

336 (2) Are there benefits to accommodating the uncertainty in mark-recapture sampling, as
337 measured by performance metrics such as accuracy, precision, and identification of
338 ideal covariates?

339 (3) Do the inferences in questions 1 and 2 depend on varying degrees of sample size,
340 violated assumptions of the binomial count model and mark-recapture model,
341 quality of mark-recapture data, and unmeasured covariate effects?

342 (4) Does the variable selection approach we used perform well relative to a hypothetical
343 case in which the covariates in the true model are known?

344 To investigate these questions, we developed a series of scenarios intended to isolate the
345 effect of each factor on the reliability of the estimation model (scenarios and organization
346 into blocks summarized in Table 2). Scenarios encompassed various combinations of data
347 availability, data quality, violated assumptions, and effects of unmonitored covariates. We
348 fitted two models to each simulated data set: the integrated model described in Section 2.2
349 and one identical in all ways except that the maximum likelihood estimate of abundance
350 from mark-recapture data was used as known abundance with no uncertainty (referred to
351 as the “external” approach). For each simulation scenario, we obtained 100 replicates of
352 true states, data sets, and fitted estimates and calculated performance metrics reflecting
353 accuracy, precision, variable selection, and credible interval coverage. Complete details of
354 data simulation, model fitting, and calculation of performance metrics are provided in
355 Supplement A.

3 Results

3.1 Empirical application

Our analysis of mark-recapture data only strongly favored models M_t and M_b over the simplest model M_0 , and favored model M_t over model M_b by over 20 Δ WAIC units. Thus, we used the assumptions of model M_t for fitting the integrated model. Averaged across all observations, capture probabilities were nearly equal between capture periods: posterior median of 0.41 (0.38 – 0.43; 95% equal-tailed credible interval) for p_1 and 0.42 (0.40 – 0.44) for p_2 . Capture probability varied widely among observations, however, with a range of posterior medians of 0.02 – 0.89 and 0.03 – 0.83 for p_1 and p_2 , respectively. We calculated the Pearson correlation coefficient between p_1 and p_2 across observations for each MCMC iteration and found them to be moderately correlated: 0.56 (0.44 – 0.67), indicating channel units that had higher-than-average capture probability in the first period were also likely to have higher-than-average capture probability in the second period as well. The integrated model fitted the mark-recapture data well, and estimated abundance generally agreed well with maximum likelihood estimates of abundance (Supplement A; Figure A1). Across observations, Chinook Salmon abundance averaged 66 fish per channel unit (range of posterior medians: 2 – 624) and *O. mykiss* abundance averaged 53 fish per channel unit (range: 1 – 597).

We found strong evidence that snorkel survey detectability varied by species, unit type, and depth, but that the depth effect was dependent on channel unit type (i.e., pool or non-pool; Fig. 2). Coefficients associated with these covariates all had parameter inclusion probabilities of 1, indicating their ω terms took on the value of 1 for each MCMC iteration and suggesting we should have complete confidence that these covariates should be accounted for when predicting detection probability at new locations within the basin. The species effect was strongly positive, indicating that Chinook Salmon juveniles were seen with greater probability than *O. mykiss* (Fig. 2). Visibility was only important if the observer determined it was good (VIS3 effect). Poor visibility (VIS1) had a low probability

383 of inclusion (0.15) and small effect size, indicating units with this attribute had similar
384 detection probabilities to those rated as average visibility. The covariates corresponding to
385 non-zero density of large wood (LWD2 and LWD3) were assigned low parameter
386 inclusion probabilities, indicating that channel units with these attributes had similar
387 detectability to units with no large wood at all. The single model with the highest
388 posterior probability (0.44) included all covariates except both large wood density
389 covariates and the poor visibility covariate (Supplement A, Table A1 therein).

390 Model-averaged detection probability response curves (Fig. 3) highlight the patterns
391 suggested by the coefficient estimates. Most observations for both species occurred in
392 channel units with no large wood present and average visibility; this limited our ability to
393 investigate interactions among some habitat characteristics or species. The interaction
394 between unit type and depth was strong, however, such that increasing depth had a
395 positive effect on detectability in non-pool units but a negative effect in pool units (Fig. 3).
396 The estimated standard deviation of site-level random effects was large (1.12; 0.88 – 1.48)
397 considering they are on the logit-scale. Posterior median detection probability for
398 observations ranged from 0.11 to 0.84 (mean across observations: 0.38) for Chinook
399 Salmon and 0.02 to 0.92 (mean: 0.24) for *O. mykiss*.

400 **3.2 Simulation analysis**

401 Due to the high dimensionality of the simulation results, only two blocks (B and D; Table
402 2) are displayed herein; complete block-specific results are presented in Supplement A and
403 displayed in a comparative layout in Supplement B.

404 In cases where model assumptions were met and data quality was not purposefully
405 reduced (blocks A, B, and F), both the integrated and external (mark-recapture-derived
406 abundance assumed to be the true value known without error) approaches returned
407 nearly unbiased fits to the data and predictions for out-of-sample data. Average median
408 percent error for abundance and detection probability across replicates was generally less
409 than 5% (positive or negative) for both modeling approaches and rarely did the median

410 percent error from any replicate exceed 10% positive relative bias. In terms of precision,
411 the integrated model tended to produce fitted values and out-of-sample predictions closer
412 to the true values, as evidenced by smaller median absolute percent error values, though
413 the difference was modest (Fig. 4b). Performance of the variable selection approach was
414 not perfect in all cases, but was at least as good or better when used in the integrated
415 model rather than the external method. In particular, the external method incorrectly
416 assigned high inclusion probabilities to truly unimportant covariates more frequently than
417 the integrated model (Fig. 4c2). Credible interval coverage was generally better for
418 detection probability than for abundance, and the integrated model had better coverage
419 than the external abundance approach in most cases (Fig. 4d). Furthermore, the external
420 abundance approach tended to over-estimate the standard deviation of site-level random
421 effects – the integrated model suffered from this issue to a lesser degree (Fig. 4e).

422 As an example of model sensitivity to violated assumptions, we have included
423 output summaries of simulation block D where it was possible to mistakenly count the
424 same individual twice during count surveys. The primary effect of this violated
425 assumption was to cause positive biases in the estimated and predicted detection
426 probabilities (Fig. 5a3,a4). Estimation and prediction of abundance was not affected nearly
427 to the same degree (Fig. 5a1,a2); this is because double-counting was possible in all count
428 surveys, including those with and without mark-recapture sampling (training and
429 prediction samples, respectively). So although we found the estimates of count survey
430 detection probability to be positively biased relative to the true value in the presence of
431 duplicate counts, largely unbiased predictions of abundance can still be obtained so long
432 as the prevalence of double-counts is the same in prediction samples as in the training
433 samples.

434 **4 Discussion**

435 We have presented a modeling approach for integrating multiple information sources that
436 are commonly collected when using a double sampling design to quantify detectability of

437 count surveys. When applied to an empirical data set, the model uncovered several useful
438 and intuitive patterns that demonstrated the effect of local conditions on the detectability
439 of snorkel surveys. Through simulation trials, we demonstrated the method can return
440 reasonably accurate and precise out-of-sample predictions. The simulation further
441 revealed that there are benefits, albeit sometimes modest in magnitude, to accounting for
442 uncertainty in mark-recapture sampling via the data integration technique, namely in
443 terms of better selection of important covariates, better credible interval coverage, and
444 greater accuracy in estimates of random variability. These patterns were consistent across
445 nearly all scenarios we considered but were most exaggerated for the cases where marked
446 fish actively avoided recapture (simulation blocks G and H). The only case in which we
447 found poorer performance of the integrated approach was with mark-recapture data
448 generated with very low capture/recapture probabilities (simulation block E, scenario 13,
449 see Supplement A section A2.2.5 for more details). These findings suggest that
450 investigators using the double sampling approach for quantifying the effects of covariates
451 on detectability may see improved statistical performance by applying the integrated
452 model we present (or similar) rather than treating abundance as a known quantity.

453 The magnitude and direction of estimated coefficients for the empirical data set were
454 intuitive and for the most part aligned with our expectations. For example, in our
455 experience with conducting snorkel surveys, *O. mykiss* are more evasive which makes
456 them more difficult to see than Chinook Salmon, an observation supported to a limited
457 extent by other studies (Hankin and Reeves 1988; Hillman et al. 1992). The depth by unit
458 type interaction is also intuitive, as shallow, fast water, non-pool units have more
459 turbulence and require more crawling (mask out of water) to progress through the survey;
460 shallow pools do not suffer from this issue but deeper pools have their own characteristics
461 that limit detectability. We had expected to identify a negative effect of high large wood
462 density, but we also had comparatively low replication at high large wood densities which
463 likely affected our ability to identify an effect. Moreover, another study also found

464 important effects of wood density, species, and visibility (Thurow et al. 2006) and their
465 estimates of detection probability for *O. mykiss* ranged between 0.15 and 0.2 (depending
466 on size class), with individual observations ranging from 0 to 0.6 – our analysis estimated
467 an average of 0.24 and a range of 0.02 to 0.92.

468 While covariates explained much of the variation in detectability in the empirical
469 data set, the estimated standard deviation of site-level random effects was large. This
470 indicates that either some important covariates were not included or the count process
471 was over-dispersed relative to the binomial process we assumed. Our simulation analyses
472 illustrated that both cases can cause positive bias in the random variability term but that
473 the accuracy of predictions was largely unaffected. Over-dispersion and failure to include
474 all important covariates are common in ecological analyses, and our simulation analyses
475 suggest that the integrated model showed better performance than the external approach
476 in these situations. Although the random term was needed to accommodate otherwise
477 unexplained variability in the empirical data, its magnitude indicates we should expect
478 low precision in out-of-sample predictions.

479 There are several important assumptions that, if violated, would affect the reliability
480 of our model. One especially worthy of note is the assumption of no net change in
481 abundance between snorkel and mark-recapture sampling. If fish counted during a
482 snorkel survey were inadvertently pushed out of the channel unit by observers before
483 block nets were inserted, then the abundance sampled by mark-recapture would be
484 smaller than that vulnerable to being counted by snorkel survey. This would have the
485 tendency to produce positive biases in our estimates of detection probability, but we
486 would not expect it to bias the estimates of covariate effects (which are log odds ratios).
487 We believe the assumption would need to be violated frequently and severely to render
488 our estimates unusable. Furthermore, this assumption is not specific to our integration of
489 abundance and count data into a single model, but rather to the use of the double
490 sampling design to quantify detectability. To the contrary, we illustrated via simulation

491 that in the face of some violated assumptions (e.g., ignoring behavioral effects in
492 mark-recapture sampling) the integrated model performed better than a non-integrated
493 method (Supplemental Fig. A10).

494 We believe this integrated approach is general enough for application to a wide
495 variety of systems that have similar data (paired count, independent abundance data, and
496 explanatory covariates). This is true so long as the count process can be reasonably
497 approximated by a binomial process, as illustrated by our simulation analysis. In our
498 simulation, the model generated snorkel count data using a binomial sampling process,
499 yet the counts could have instead been intended to represent counts from spotlight
500 surveys, aerial surveys, camera traps, or other methods and the conclusions would be
501 identical.

502 There are a variety of adaptations that could be made to the model for application to
503 new use cases. First, the count component is essentially a binomial generalized linear
504 model, meaning that its complexity can be tailored to available data and research needs.
505 With suitable survey design, there are many aspects that could be included such as (a)
506 variability due to year or observer, (b) higher order (e.g., quadratic) terms to accommodate
507 non-linearity, (c) more interaction terms, or (d) more complex structures for random effects
508 that better capture spatial autocorrelation. The indicator variable selection approach we
509 embedded within the model to accommodate model uncertainty is general and can be
510 altered to perform multi-model inference in more simplistic or complex cases (more
511 examples of this approach in ecological settings can be found in Coggins et al. 2014;
512 Dorazio et al. 2011; Gwinn et al. 2019a,b). Second, the mark-recapture likelihood could be
513 altered to accommodate independent abundance data from more complex designs or
514 covariates could be used to explain variability in capture probabilities ($p_{i,1}$ and $p_{i,2}$) which
515 could potentially increase precision of abundance estimates. In addition to our conditional
516 multinomial likelihood approach, data augmentation (e.g., Royle and Dorazio 2010; Kéry
517 and Schaub 2012, Ch. 6) is another method for expressing mark-recapture likelihoods that

518 may show promise in this context (M. Falcu, personal communication). This would be
519 particularly true if individual covariates (e.g., fish size) need to be accounted for, as data
520 augmentation methods model sampling processes on an individual basis. Further, we
521 believe non-mark-recapture designs for obtaining abundance estimates (e.g., depletion
522 sampling, catchability-adjusted catch rate data, etc.) could be accommodated by our
523 approach as well, so long as a sampling distribution can be assigned to the abundance
524 estimate (or raw data) and an associated measure of statistical (i.e., observational)
525 uncertainty is available. Finally, addressing particular research questions may necessitate
526 assigning counts or captured individuals to multiple groups (e.g., size, age, or sex); it is
527 plausible that the model could be extended to accommodate multinomial count outcomes.

528 A common goal of many ecological research, monitoring, and evaluation programs
529 includes understanding and predicting animal abundance and distribution as influenced
530 by habitat quality, climate change, species interactions, trophic relations, and other factors.
531 However, the first step in any such program should include robust evaluation of the core
532 response variable – animal abundance – which often includes assessing heterogeneity in
533 count survey detectability. Our integrated method should be applicable to most double
534 sampling designs wherein the investigators wish to estimate covariate effects on
535 detectability while being explicit about uncertainty in abundance data or estimates. The
536 modeling approaches presented in this study point towards opportunities for future
537 research that will help guide studies targeted at quantifying detectability with the
538 intention of making more robust inferences from count data.

539 **Acknowledgements**

540 We thank the field personnel, who are too numerous to name individually, for collecting
541 the data that informed our empirical analysis. We are grateful for C. Horn's contributions
542 to the field study design. The analytical components of this research were supported by
543 the Bonneville Power Administration as part of the Columbia Basin Fish Accords
544 Agreement (Project #2009-004-00).

545 **Authors' Contributions**

546 **B.A. Staton:** conceptualization, methodology, software, verification, formal analysis,
547 writing – original draft, writing – review & editing; **C. Justice:** conceptualization,
548 methodology, investigation, data curation, writing – review & editing; **S. White:**
549 conceptualization, methodology, investigation, writing – review & editing, supervision,
550 project administration; **E.R. Sedell:** conceptualization, methodology, investigation,
551 writing – review & editing, supervision; **L.A. Burns:** conceptualization, methodology,
552 writing – review & editing; **M.J. Kaylor:** conceptualization, methodology, writing – review
553 & editing

554 **Data Availability**

555 All data and code used to run analyses and generate figures found herein are documented
556 on GitHub [DOI: [10.5281/zenodo.3928691](https://doi.org/10.5281/zenodo.3928691); Staton (2020a)].

References

- 557
558 Barbieri, M. M. and Berger, J. O. 2004. Optimal predictive model selection. *The Annals of Statistics*,
559 32(3):870–897. doi: [10.1214/009053604000000238](https://doi.org/10.1214/009053604000000238).
- 560 Barker, R. J., Schofield, M. R., Link, W. A., and Sauer, J. R. 2017. On the reliability of N-mixture
561 models for count data. *Biometrics*, 74(1):369–377. doi: [10.1111/biom.12734](https://doi.org/10.1111/biom.12734).
- 562 Bart, J. and Earnst, S. 2002. Double sampling to estimate density and population trends in birds.
563 *The Auk*, 119(1):36–45. doi: [10.1093/auk/119.1.36](https://doi.org/10.1093/auk/119.1.36).
- 564 Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. 2001.
565 *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University
566 Press, Oxford, UK.
- 567 Bürkner, P.-C., Gabry, J., Kay, M., and Vehtari, A. 2020. *posterior: Tools for Working with Posterior*
568 *Distributions*. R package version 0.1.3, URL: <https://mc-stan.org/posterior>.
- 569 CHaMP 2016. *Scientific protocol for salmonid habitat surveys within the Columbia Habitat Monitoring*
570 *Program*. Columbia Habitat Monitoring Program. URL:
571 <https://www.champmonitoring.org/Program/RetrieveProgramDocumentFile/Tab/1126>.
- 572 Coggins, L. G., Bacheler, N. M., and Gwinn, D. C. 2014. Occupancy models for monitoring marine
573 fish: A Bayesian hierarchical approach to model imperfect detection with a novel gear
574 combination. *PLoS ONE*, 9(9):e108302. doi: [10.1371/journal.pone.0108302](https://doi.org/10.1371/journal.pone.0108302).
- 575 Collier, B. A., Ditchkoff, S. S., Raglin, J. B., and Smith, J. M. 2007. Detection probability and sources
576 of variation in white-tailed deer spotlight surveys. *Journal of Wildlife Management*, 71(1):277–281.
577 doi: [10.2193/2005-728](https://doi.org/10.2193/2005-728).
- 578 Constable, Jr., J. and Suring, E. 2015. Juvenile salmonid monitoring in coastal Oregon and Lower
579 Columbia streams, 2014. Monitoring Program Report OPSW-ODFW-2014-1, Oregon
580 Department of Fish and Wildlife, Salem, OR. URL:
581 <https://nrimp.dfw.state.or.us/CRL/reports/WORP/WORP2014AnnualReport.pdf>.
- 582 Doll, J. C., Etchison, L., and Owensby, D. 2020. Population estimate of the state and federally
583 threatened Spotfin Chub using underwater observations. *North American Journal of Fisheries*
584 *Management*, 40(2):342–353. doi: [10.1002/nafm.10414](https://doi.org/10.1002/nafm.10414).
- 585 Dorazio, R. M., Gotelli, N. J., and Ellison, A. M. 2011. Modern methods of estimating biodiversity
586 loss from presence-absence surveys. In Grillo, O., editor, *Biodiversity Loss in a Changing Planet*,
587 pages 277 – 302. InTech. URL:
588 [https://www.intechopen.com/books/biodiversity-loss-in-a-changing-planet/
589 modern-methods-of-estimating-biodiversity-from-presence-absence-surveys](https://www.intechopen.com/books/biodiversity-loss-in-a-changing-planet/modern-methods-of-estimating-biodiversity-from-presence-absence-surveys).
- 590 Flitcroft, R., Burnett, K., Snyder, J., Reeves, G., and Ganio, L. 2013. Riverscape patterns among
591 years of juvenile coho salmon in midcoastal Oregon: Implications for conservation. *Transactions*
592 *of the American Fisheries Society*, 143(1):26–38. DOI: [10.1080/00028487.2013.824923](https://doi.org/10.1080/00028487.2013.824923).
- 593 Fowler, P. 2017. Southeast Alaska steelhead snorkel surveys of regional index streams, 2014 and
594 2015. Fishery Data Series 17-13, Alaska Department of Fish and Game, Anchorage, AK. URL:
595 <http://www.adfg.alaska.gov/FedAidPDFs/FDS17-13.pdf>.
- 596 Gwinn, D. C., Bacheler, N. M., and Shertzer, K. W. 2019a. Integrating underwater video into
597 traditional fisheries indices using a hierarchical formulation of a state-space model. *Fisheries*
598 *Research*, 219:105309. doi: [10.1016/j.fishres.2019.105309](https://doi.org/10.1016/j.fishres.2019.105309).
- 599 Gwinn, D. C., Todd, C. R., Brown, P., Hunt, T. L., Butler, G., Kitchingman, A., Koehn, J. D., and

- 600 Ingram, B. 2019b. Assessing a threatened fish species under budgetary constraints: Evaluating
 601 the use of existing monitoring data. *North American Journal of Fisheries Management*,
 602 39(2):315–327. doi: [10.1002/nafm.10271](https://doi.org/10.1002/nafm.10271).
- 603 Hankin, D. G. and Reeves, G. H. 1988. Estimating total fish abundance and total habitat area in
 604 small streams based on visual estimation methods. *Canadian Journal of Fisheries and Aquatic
 605 Sciences*, 45(5):834–844. doi: [10.1139/f88-101](https://doi.org/10.1139/f88-101).
- 606 Hillman, T. W., Mullan, J. W., and Griffith, J. S. 1992. Accuracy of underwater counts of juvenile
 607 Chinook salmon, coho salmon, and steelhead. *North American Journal of Fisheries Management*,
 608 12(3):598–603. doi: [10.1577/1548-8675\(1992\)012<0598:aoucoj>2.3.co;2](https://doi.org/10.1577/1548-8675(1992)012<0598:aoucoj>2.3.co;2).
- 609 Hooten, M. B. and Hobbs, N. T. 2015. A guide to Bayesian model selection for ecologists. *Ecological
 610 Monographs*, 85(1):3–28. doi: [10.1890/14-0661.1](https://doi.org/10.1890/14-0661.1).
- 611 Huggins, R. M. 1989. On the statistical analysis of capture experiments. *Biometrika*, 76(1):133–140.
 612 doi: [10.1093/biomet/76.1.133](https://doi.org/10.1093/biomet/76.1.133).
- 613 Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture
 614 experiments. *Biometrics*, 47(2):725–732. doi: [10.2307/2532158](https://doi.org/10.2307/2532158).
- 615 Hutton, J. M. and Woolhouse, M. E. J. 1989. Mark-recapture to assess factors affecting the
 616 proportion of a Nile crocodile population seen during spotlight counts at Ngezi, Zimbabwe,
 617 and the use of spotlight counts to monitor crocodile abundance. *The Journal of Applied Ecology*,
 618 26(2):381. doi: [10.2307/2404068](https://doi.org/10.2307/2404068).
- 619 Jonasson, B. C., Sedell, E. R., Tattam, S. K., Garner, A. B., Horn, C., Bliesner, K. L., Dowdy, J. W.,
 620 Favrot, S. D., Hay, J. M., McMichael, G. A., Power, B. C., Davis, O. C., and Ruzycki, J. R. 2016.
 621 Investigations into the life history of naturally produced spring Chinook salmon and summer
 622 steelhead in the Grande Ronde River subbasin. Annual Report BPA Project #1992-026-04,
 623 Oregon Department of Fish and Wildlife, La Grande, OR. URL:
 624 https://nrimp.dfw.state.or.us/web%20stores/data%20libraries/files/ODFW/ODFW_40892_2_1992-026-04_2015_FinalReport_BPA_Version.pdf.
 625
- 626 Jones, E. L., Quinn, T. J., and Alen, B. W. V. 1998. Observer accuracy and precision in aerial and foot
 627 survey counts of pink salmon in a southeast Alaska stream. *North American Journal of Fisheries
 628 Management*, 18(4):832–846. doi: [10.1577/1548-8675\(1998\)018<0832:oaapia>2.0.co;2](https://doi.org/10.1577/1548-8675(1998)018<0832:oaapia>2.0.co;2).
- 629 Kellner, K. 2018. *jagsUI: A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses*. R package version
 630 1.5.0.
- 631 Kellner, K. F. and Swihart, R. K. 2014. Accounting for imperfect detection in ecology: A
 632 quantitative review. *PLoS ONE*, 9(10):e111436. doi: [10.1371/journal.pone.0111436](https://doi.org/10.1371/journal.pone.0111436).
- 633 Kéry, M. and Schmidt, B. 2008. Imperfect detection and its consequences for monitoring for
 634 conservation. *Community Ecology*, 9(2):207–216. doi: [10.1556/comec.9.2008.2.10](https://doi.org/10.1556/comec.9.2008.2.10).
- 635 Kéry, M. and Schuab, M. 2012. *Bayesian population analysis using WinBUGS: a hierarchical perspective*.
 636 Academic Press, Waltham, MA.
- 637 Korman, J., Ahrens, R. N., Higgins, P. S., and Walters, C. J. 2002. Effects of observer efficiency,
 638 arrival timing, and survey life on estimates of escapement for steelhead trout (*Oncorhynchus
 639 mykiss*) derived from repeat mark-recapture experiments. *Canadian Journal of Fisheries and
 640 Aquatic Sciences*, 59(7):1116–1131. doi: [10.1139/f02-081](https://doi.org/10.1139/f02-081).
- 641 Kuo, L. and Mallick, B. 1998. Variable selection for regression models. *Sankhyā: The Indian Journal of
 642 Statistics, Series B*, 60(1):65–81. URL:

- 643 https://www.jstor.org/stable/25053023?seq=1#metadata_info_tab_contents.
- 644 Lawson, Z. J., Gaeta, J. W., and Carpenter, S. R. 2011. Coarse woody habitat, lakeshore residential
645 development, and largemouth bass nesting behavior. *North American Journal of Fisheries*
646 *Management*, 31(4):666–670. doi: [10.1080/02755947.2011.608990](https://doi.org/10.1080/02755947.2011.608990).
- 647 Mosquera, I., Cote, I. M., Jennings, S., and Reynolds, J. D. 2000. Conservation benefits of marine
648 reserves for fish populations. *Animal Conservation*, 3(4):321–332. doi:
649 [10.1111/j.1469-1795.2000.tb00117.x](https://doi.org/10.1111/j.1469-1795.2000.tb00117.x).
- 650 Nehlsen, W., Williams, J. E., and Lichatowich, J. A. 1991. Pacific salmon at the crossroads: Stocks at
651 risk from California, Oregon, Idaho, and Washington. *Fisheries*, 16(2):4–21. doi:
652 [10.1577/1548-8446\(1991\)016<0004:psatcs>2.0.co;2](https://doi.org/10.1577/1548-8446(1991)016<0004:psatcs>2.0.co;2).
- 653 O’Neal, J. S. 2007. Snorkel surveys. In Johnson, D. H., Shrier, B. M., O’Neal, J. S., Knutzen, J. A.,
654 Augerot, X., O’Neil, T. A., and Pearsons, T. N., editors, *Salmonid Field Protocols Handbook:*
655 *Techniques for Assessing Status and Trends in Salmon and Trout Populations*, pages 325–340.
656 American Fisheries Society, Bethesda, MD.
- 657 Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. 1978. Statistical inference from
658 capture data on closed animal populations. *Wildlife Monographs*, (62):3–135.
- 659 Peterson, J. T., Banish, N. P., and Thurow, R. F. 2005. Are block nets necessary?: Movement of
660 stream-dwelling salmonids in response to three common survey methods. *North American*
661 *Journal of Fisheries Management*, 25(2):732–743. doi: [10.1577/m04-051.1](https://doi.org/10.1577/m04-051.1).
- 662 Peterson, J. T., Scheerer, P. D., and Clements, S. 2015. An evaluation of the efficiency of minnow
663 traps for estimating the abundance of minnows in desert spring systems. *North American Journal*
664 *of Fisheries Management*, 35(3):491–502. doi: [10.1080/02755947.2015.1017125](https://doi.org/10.1080/02755947.2015.1017125).
- 665 Pinter, K., Lautsch, E., Unfer, G., and Hayes, D. S. 2018. Snorkeling-based fish stock assessment by
666 anglers—a valuable method for managing recreational fisheries. *North American Journal of*
667 *Fisheries Management*. doi: [10.1002/nafm.10246](https://doi.org/10.1002/nafm.10246).
- 668 Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs
669 sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna,*
670 *Austria*, 124. URL: <https://www.r-project.org/conferences/DSC-2003/Drafts/Plummer.pdf>.
- 671 R Core Team 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for
672 Statistical Computing, Vienna, Austria. Version 4.0.2.
- 673 Roni, P., Hanson, K., and Beechie, T. 2008. Global review of the physical and biological
674 effectiveness of stream habitat rehabilitation techniques. *North American Journal of Fisheries*
675 *Management*, 28(3):856–890. doi: [10.1577/m06-169.1](https://doi.org/10.1577/m06-169.1).
- 676 Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts.
677 *Biometrics*, 60(1):108–115. doi: [10.1111/j.0006-341x.2004.00142.x](https://doi.org/10.1111/j.0006-341x.2004.00142.x).
- 678 Royle, J. A. and Dorazio, R. M. 2010. Parameter-expanded data augmentation for Bayesian analysis
679 of capture–recapture models. *Journal of Ornithology*, 152(S2):521–537. doi:
680 [10.1007/s10336-010-0619-4](https://doi.org/10.1007/s10336-010-0619-4).
- 681 Royle, J. A. and Nichols, J. D. 2003. Estimating abundance from repeated presence–absence data or
682 point counts. *Ecology*, 84(3):777–790. doi: [10.1890/0012-9658\(2003\)084\[0777:EAFRPA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2).
- 683 Seber, G. A. F. 1986. A review of estimating animal abundance. *Biometrics*, 42(2):267. doi:
684 [10.2307/2531049](https://doi.org/10.2307/2531049).
- 685 Som, N. A., Perry, R. W., Jones, E. C., Juilio, K. D., Petros, P., Pinnix, W. D., and Rupert, D. L. 2018.

- 686 N-mix for fish: Estimating riverine salmonid habitat selection via N-mixture models. *Canadian*
687 *Journal of Fisheries and Aquatic Sciences*, 75(7):1048–1058. doi: [10.1139/cjfas-2017-0027](https://doi.org/10.1139/cjfas-2017-0027).
- 688 Staton, B. 2020a. *bstaton1/snk-eff-ms-analysis*: Final Release (v1.0) of repository storing code and
689 data to replicate analyses. doi: [10.5281/zenodo.3928691](https://doi.org/10.5281/zenodo.3928691).
- 690 Staton, B. 2020b. *postpack: Utilities for Processing Posterior Samples Stored in 'mcmc.lists'*. R package
691 version 0.5.2, URL: <https://bstaton1.github.io/postpack/>.
- 692 Thompson, W. L. and Lee, D. C. 2000. Modeling relationships between landscape-level attributes
693 and snorkel counts of Chinook salmon and steelhead parr in Idaho. *Canadian Journal of Fisheries*
694 *and Aquatic Sciences*, 57(9):1834–1842. doi: [10.1139/f00-135](https://doi.org/10.1139/f00-135).
- 695 Thurow, R. F. 1994. Underwater methods for study of salmonids in the intermountain West.
696 General Technical Report INT-GTR-307, Department of Agriculture, U.S. Forest Service,
697 Intermountain Research Station, Odgen, UT. doi: [10.2737/INT-GTR-307](https://doi.org/10.2737/INT-GTR-307).
- 698 Thurow, R. F., Peterson, J. T., and Guzevich, J. W. 2006. Utility and validation of day and night
699 snorkel counts for estimating bull trout abundance in first- to third-order streams. *North*
700 *American Journal of Fisheries Management*, 26(1):217–232. doi: [10.1577/m05-054.1](https://doi.org/10.1577/m05-054.1).
- 701 Ulibarri, R. M., Bonar, S. A., Rees, C., Amberg, J., Ladell, B., and Jackson, C. 2017. Comparing
702 efficiency of American Fisheries Society standard snorkeling techniques to environmental DNA
703 sampling techniques. *North American Journal of Fisheries Management*, 37(3):644–651. doi:
704 [10.1080/02755947.2017.1306005](https://doi.org/10.1080/02755947.2017.1306005).
- 705 Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. 2021. Rank-normalization,
706 folding, and localization: An improved Rhat for assessing convergence of MCMC. *Bayesian*
707 *Analysis*. doi: [10.1214/20-BA1221](https://doi.org/10.1214/20-BA1221).
- 708 Walters, C. and Pearse, P. H. 1996. Stock information requirements for quota management systems
709 in commercial fisheries. *Reviews in Fish Biology and Fisheries*, 6(1):21–42. doi: [10.1007/bf00058518](https://doi.org/10.1007/bf00058518).
- 710 Watanabe, S. 2013. A widely applicable Bayesian information criterion. *Journal of Machine Learning*
711 *Research*, 14:867–897.
- 712 Weaver, D. M., Kwak, T. J., and Pollock, K. H. 2014. Sampling characteristics and calibration of
713 snorkel counts to estimate stream fish populations. *North American Journal of Fisheries*
714 *Management*, 34(6):1159–1166. doi: [10.1080/02755947.2014.951808](https://doi.org/10.1080/02755947.2014.951808).
- 715 White, S., Justice, C., Burns, L., Graves, D., Kelsey, D., and Kaylor, M. 2019. Assessing the status
716 and trends of spring Chinook habitat in the upper Grande Ronde River and Catherine Creek:
717 Annual report 2018. Annual Report 19-04, Columbia River Inter-Tribal Fish Commission,
718 Portland, OR. URL: <https://www.critfc.org/wp-content/uploads/2019/05/19-04.pdf>.
- 719 White, S., Justice, C., and McCullough, D. 2012. *Protocol for snorkel surveys of fish densities*. Columbia
720 River Inter-Tribal Fish Commission. URL:
721 <https://www.monitoringmethods.org/Protocol/Details/499>.
- 722 Williams, B. K., Nichols, J. D., and Conroy, M. J. 2002. *Analysis and Management of Animal*
723 *Populations*. Academic Press, San Diego, CA.

TABLE 1. Three types of mark-recapture models for the two-sample case with closed populations. For models M_0 and M_t , $p_{i,1}$ and $p_{i,2}$ are capture probabilities for all individuals in each period. For model M_b , $p_{i,2}$ is the probability of first capture in the second period, and $c_{i,2}$ is the probability of recapture. The Δ WAIC column shows the difference in WAIC scores between each model and the lowest WAIC model for the Grande Ronde empirical analysis, which we used to inform which assumption should be made when fitting the integrated model.

Model	Assumptions	Description	Δ WAIC
M_0	$p_{i,1} = p_{i,2}$	Time-constant capture probability	101
M_t	$p_{i,1} \neq p_{i,2}$	Period-specific capture probability	0
M_b	$p_{i,1} = p_{i,2} \neq c_{i,2}$	Behavioral response affects recapture probability	22

TABLE 2. Breakdown of the scenarios used for the simulation analysis. See Supplement A for a description of how the scenarios were implemented.

Block	Description	Scenario	Training Samples	ω Known	Over-disp. Counts ^a	Double Count Prob.	p_1, p_2^b	c_2^c	Assumed Model ^d	Unobserved Covariates ^e	
A	Model assumptions met, true model known, vary training sample size	1	25	Yes	None	0	beta(10, 10)	p_2	M_t	None	
		2	50								
		3	100								
B	Model assumptions met, true model unknown, vary training sample size	4	25	No	None	0	beta(10, 10)	p_2	M_t	None	
		5	50								
		6	100								
C	Homogenous p_i assumption violated, vary the amount of over-dispersion	7	100	No	Little	0	beta(10, 10)	p_2	M_t	None	
		8									Some
		9									Lots
D	Single count only assumption violated, vary the probability a fish is counted twice	10	100	No	None	0.05	beta(10, 10)	p_2	M_t	None	
		11				0.1					
		12				0.2					
E	Model assumptions met, vary the quality of mark-recapture data	13	100	No	None	0	beta(20, 80)	p_2	M_t	None	
		14					beta(50, 50)				
		15					beta(80, 20)				
F	Model assumptions met, vary the contribution of unobserved covariates	16	100	No	None	0	beta(10, 10)	p_2	M_t	Little	
		17								Some	
		18								Lots	
G	Behavioral recapture response present, but not accounted for	19	100	No	None	0	beta(10, 10)	$p_2 \cdot 0.9$	M_t	None	
		20						$p_2 \cdot 0.7$			
		21						$p_2 \cdot 0.5$			
H	Behavioral recapture response present, use WAIC to determine if it needs accounting	22	100	No	None	0	beta(10, 10)	$p_2 \cdot 0.9$	Use WAIC	None	
		23				$p_2 \cdot 0.7$					
		24				$p_2 \cdot 0.5$					

^a We used a beta-Binomial count model to simulate over-dispersed counts of varying degrees by changing the sum of the two beta parameters; none = 1×10^{10} , little = 20, some = 10, lots = 5.

^b Capture probabilities for each period were sampled from the listed beta distribution for each observation.

^c Recapture probability for the second period is a function of the capture probability in that period.

^d For block H, WAIC was used to select the best mark-recapture assumptions to make prior to fitting the model with count data.

^e We used observation-level logit-normal deviates to simulate the effects of unobserved covariates. We varied their standard deviation: none = 0, little = 0.1, some = 0.3, lots = 0.5.

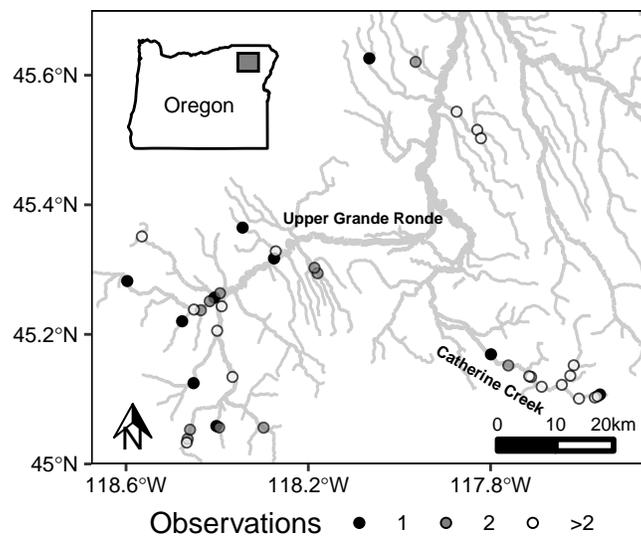


FIGURE 1. Map of the portion of the Grande Ronde River basin that served as the study area for the empirical application of the integrated model. Point colors denote the number of observations made at each site, i.e., the number of replicates for the site-level random effect.

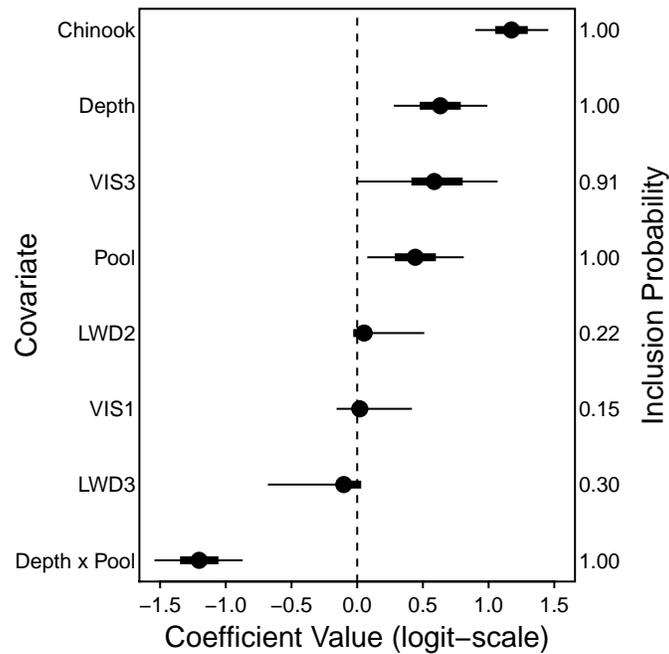


FIGURE 2. Model-averaged coefficient estimates ($\omega\beta$) from the empirical application of the integrated model. Positive coefficient values indicate that covariate increases snorkel detection probability; points represent posterior means, thick bars represent central 50% credible intervals and thin bars represent 95% credible intervals. Along the right axis are posterior probabilities that each covariate should be included in an optimal predictive model (posterior mean of each ω term).

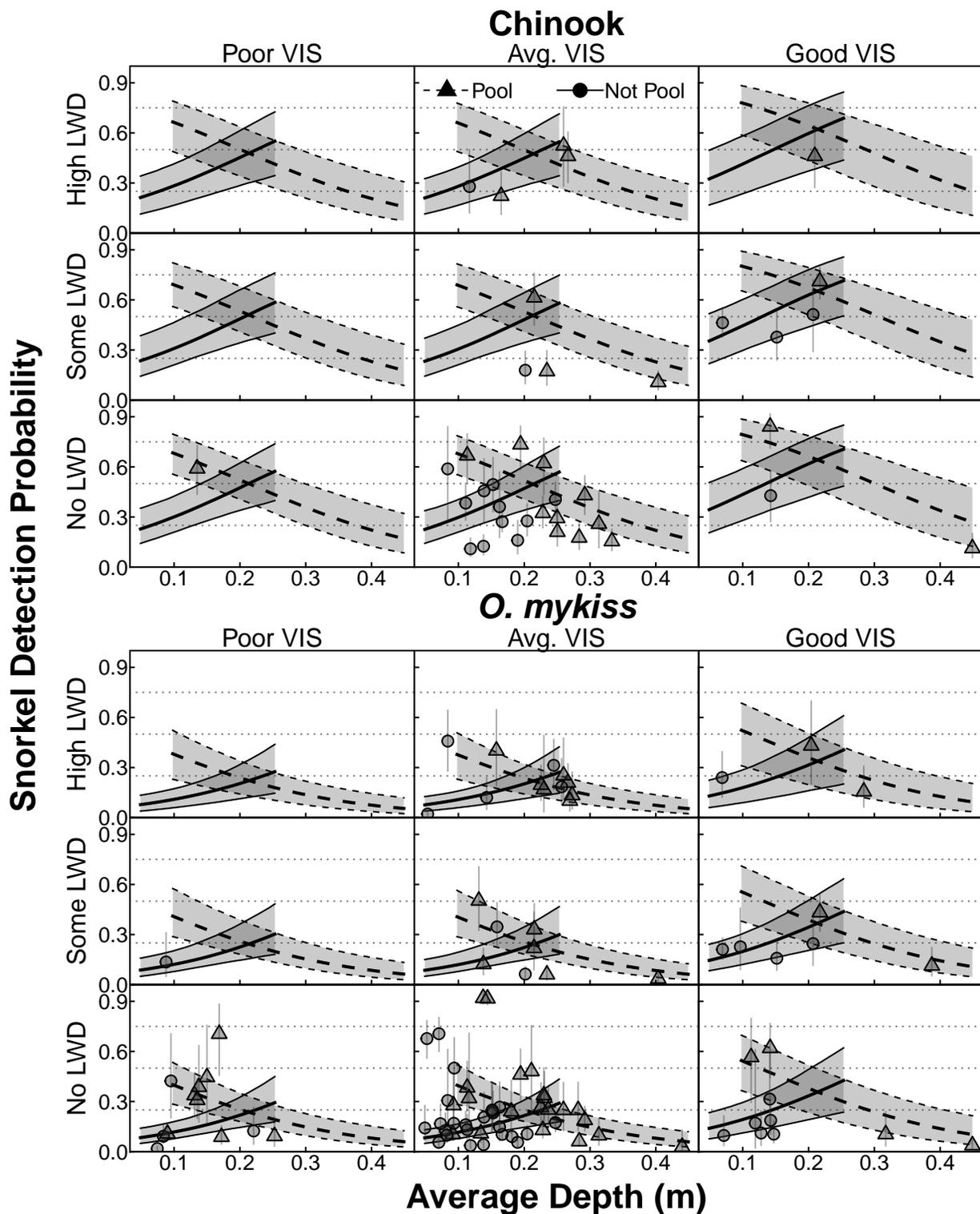


FIGURE 3. Response of snorkel survey detection probability to various covariates for two salmonid species in the empirical analysis. Each panel shows a unique combination of large wood density (LWD; rows) and snorkeler-determined visibility (VIS; columns) grouped by species (axes extent equal for all panels). Curves and points display the model-averaged posterior median fixed-effect relationship and observation-specific detection probability, respectively, for non-pool and pool units. Grey bands and error bars denote 95% equal-tailed credible intervals.

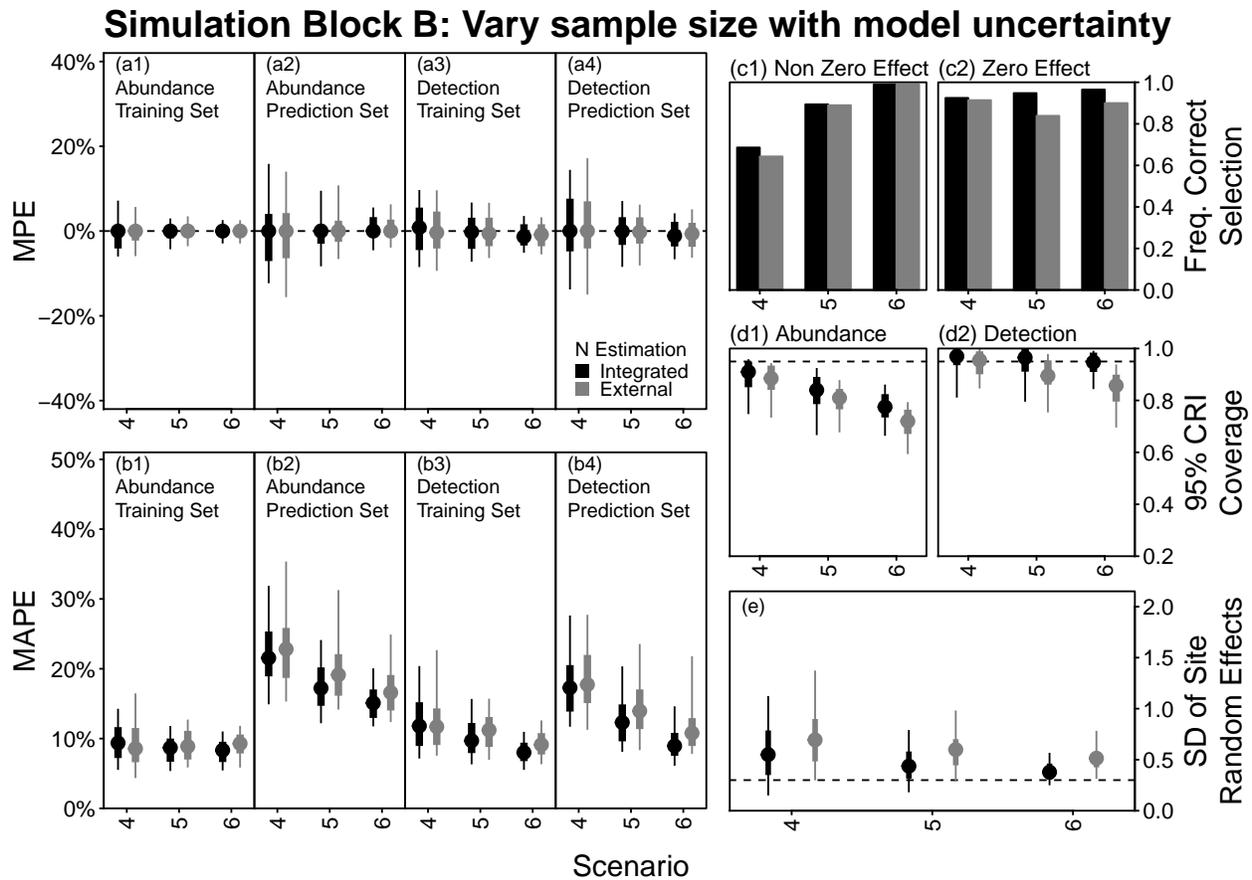


FIGURE 4. Summary of output from block B of the simulation trials, in which all model assumptions were met and the sample size was varied. Points represent medians across 100 replicate data sets, thick error bars represent the central 50% of outcomes, and thin error bars represent the central 95% of outcomes. (a) Distribution of median percent errors (MPE) across replicate data sets for abundance (a1 and a2) and detection probability (a3 and a4) – dashed line at 0 shows no error. (b) Same layout as (a), except for median absolute percent errors (MAPE). (c1) Proportion of the simulations in which truly non-zero effects were assigned probability of inclusion greater than 0.5. (c2) Proportion of the simulations in which truly zero-valued effects were assigned probability of inclusion less than 0.5. (d) Distribution of coverage statistics for abundance (d1) and detection (d2) – i.e., the fraction of 95% credible intervals that contained the true value (reference dashed line at 0.95 shows optimal coverage). (e) Distribution of estimated standard deviation of site-level random effects – dashed line shows the true value used to simulate the data. Matching figures for all simulation blocks are shown in Supplements A and B.

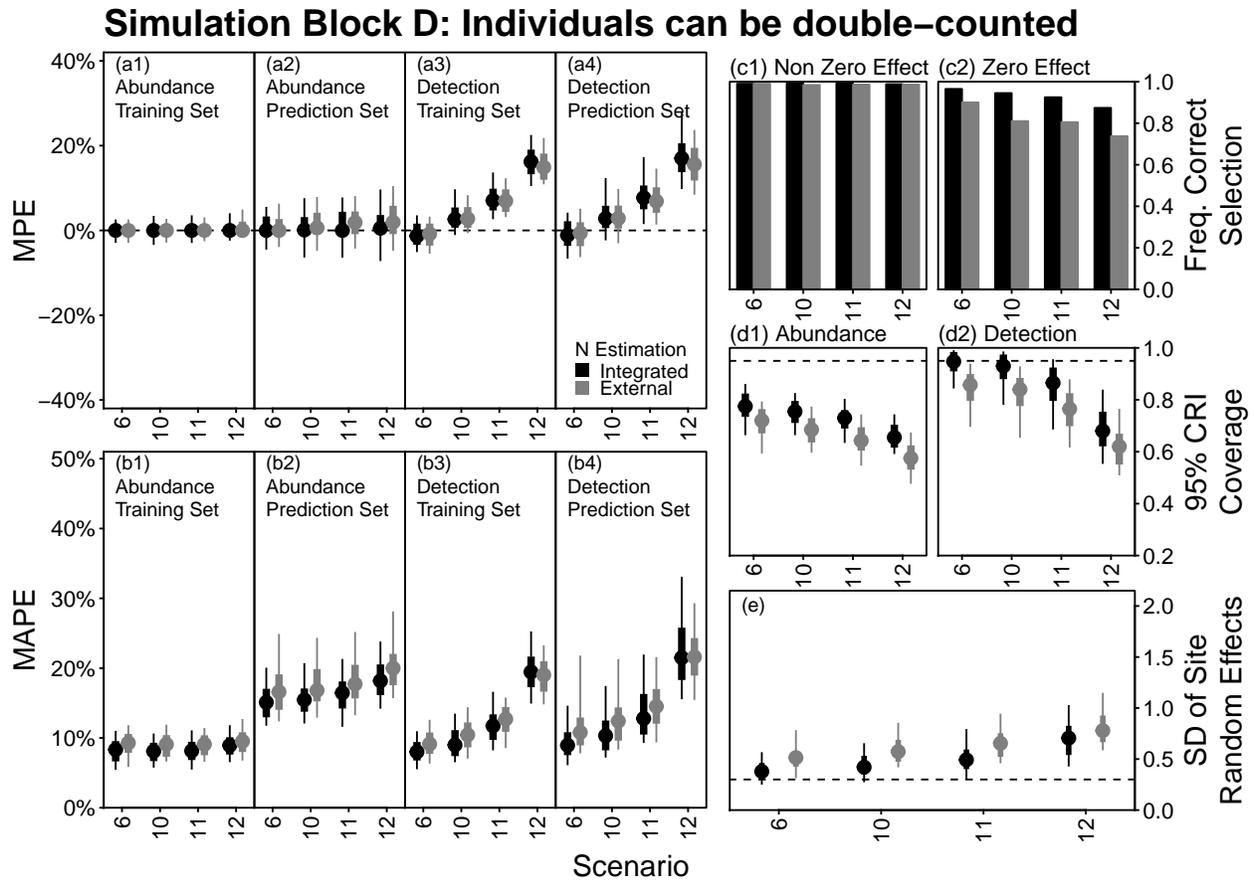


FIGURE 5. Summary of output from block D of the simulation trials in which it was possible to double-count individuals but the models did not account for it. Scenarios 6, 10, 11, and 12 had probabilities of double counting individuals of 0, 0.05, 0.1, and 0.2, respectively, and are identical in all other ways. Layout of this figure is identical to Figure 4, consult that figure caption for a description of the various panels.