NRC Research Press

ARTICLE

N-mix for fish: estimating riverine salmonid habitat selection via N-mixture models

Nicholas A. Som, Russell W. Perry, Edward C. Jones, Kyle De Juilio, Paul Petros, William D. Pinnix, and Derek L. Rupert

Abstract: Models that formulate mathematical linkages between fish use and habitat characteristics are applied for many purposes. For riverine fish, these linkages are often cast as resource selection functions with variables including depth and velocity of water and distance to nearest cover. Ecologists are now recognizing the role that detection plays in observing organisms, and failure to account for imperfect detection can lead to spurious inference. Herein, we present a flexible N-mixture model to associate habitat characteristics with the abundance of riverine salmonids that simultaneously estimates detection probability. Our formulation has the added benefits of accounting for demographics variation and can generate probabilistic statements regarding intensity of habitat use. In addition to the conceptual benefits, model application to data from the Trinity River, California, yields interesting results. Detection was estimated to vary among surveyors, but there was little spatial or temporal variation. Additionally, a weaker effect of water depth on resource selection is estimated than that reported by previous studies not accounting for detection probability. N-mixture models show great promise for applications to riverine resource selection.

Résumé : Les modèles qui formulent des liens mathématiques entre l'utilisation d'habitats par les poissons et les caractéristiques des habitats sont utilisés à différentes fins. Pour les poissons de rivière, ces liens sont souvent présentés sous forme de fonctions de sélection de ressources avec des variables comme la profondeur et la vitesse de l'eau et la distance du couvert le plus proche. Les écologistes reconnaissent maintenant le rôle que joue la détection dans l'observation d'organismes, et le défaut de tenir compte d'une détection imparfaite peut mener à des inférences erronées. Nous présentons un modèle de mélange de N souple pour associer des caractéristiques de l'habitat à l'abondance de salmonidés de rivière qui estime simultanément la probabilité de détection. Notre formulation a aussi l'avantage de tenir compte des variations démographiques et peut générer des énoncés probabilistes concernant l'intensité de l'utilisation de l'habitat. En plus des avantages conceptuels, l'application du modèle à des données de la rivière Trinity (Californie, États-Unis) donne des résultats intéressants. S'il est estimé que la détection varie selon la personne qui réalise le relevé, il y a peu de variation spatiale ou temporelle. En outre, l'effet estimé de la profondeur de l'eau sur la sélection de ressources est plus faible que ce dont font état des études antérieures qui ne tiennent pas compte de la probabilité de détection. Les modèles de mélange de N sont très prometteurs pour des applications touchant à la sélection de ressources dans les cours d'eau. [Traduit par la Rédaction]

Introduction

Coupled physical and biological models often focus on interactions of target species with their environment for the purposes of exploring emergent properties (Harvey and Railsback 2009), evaluating management scenarios (Sandoval-Solis et al. 2013; Gard 2014), or predicting effects of future climate change (Holsinger et al. 2014). Key inputs to these models are information on the association of species with habitat characteristics, and quantifying these relationships allows dynamic models to simulate individual or population responses to physical habitat change. As such, an essential component of model construction is mathematically linking habitat quality to the model's spatial domain as a function of physical variables (Scheuerell et al. 2006).

Habitat models are commonly labeled as species distribution, resource selection, or habitat suitability models, and the last several decades have seen tremendous contributions regarding methods for modeling habitat quality and use as a function of physical variables. Initial methods relied on professional judgement (Bovee 1986), which has been modernized via improvements that include fuzzy inference methods and rule sets (Ahmadi-Nedushan et al. 2008; Conallin et al. 2010). Quantitative methods have become prevalent, with fisheries applications beginning with univariate frequency analysis (i.e., habitat suitability curves; HSC) (Hayes and Jowett 1994; Som et al. 2015). The suite of quantitative methods has continued to grow and includes applications of generalized linear models (Alldredge and Dasgupta 2003; Labonne et al. 2003), pattern recognition methods such as neural networks (Brosse et al. 1999), and more complex Bayesian hierarchical models (Boone et al. 2012). Methods have even been proposed for data containing presence-only observations (Phillips et al. 2006; Royle et al. 2012), which can be common in species distribution studies (Hefley and Hooten 2016), though some controversy remains re-

Received 20 January 2017. Accepted 31 August 2017.

N.A. Som. US Fish and Wildlife Service, Arcata FWO, Arcata, CA 95521, USA; Humboldt State University, Department of Fisheries Biology, Arcata, CA 95521, USA.
R.W. Perry and E.C. Jones. US Geological Survey, Western Fisheries Research Center, Cook, WA 98605, USA.
K. De Juilio. Yurok Tribal Fisheries Program, Weaverville, CA 96093, USA.
P. Petros. Hoopa Valley Tribal Fisheries, Hoopa, CA 95546, USA.
W.D. Pinnix and D.L. Rupert. US Fish and Wildlife Service, Arcata FWO, Arcata, CA 95521, USA.
Corresponding author: Nicholas A. Som (email: nicholas_som@fws.gov).

garding theoretical aspects of these models (Hastie and Fithian 2013). Given the breadth of methods available, the literature has hosted numerous methods comparisons (e.g., see Boyce et al. 2002; Baasch et al. 2010) and reviews (Ahmadi-Nedushan et al. 2006; Newcomb et al. 2007).

Ecologists and resource managers have recognized the role that detection plays in abundance estimation studies. It is now well understood that spurious ecological inference can result from a failure to account for imperfect detection (MacKenzie 2005; Dénes et al. 2015), particularly when factors driving abundance also affect detectability (Kéry and Royle 2010). Despite this recognition and various methods available to account for imperfect detection, many studies still fail to account for detection efficiency (Kellner and Swihart 2014). It appears that accounting for detection in species distribution studies has lagged behind other ecological pursuits (Lahoz-Monfort et al. 2014), though examples of species distribution studies incorporating detection are emerging (Hefley and Hooten 2016; Guillera-Arroita 2017; Koshkina et al. 2017).

Juvenile salmonids are no exception to the detection efficiency issue, as few sampling methods are 100% efficient (Sethi and Benolkin 2013). There are a variety of sampling and survey methods available to estimate detection efficiency, but many are not suitable for assessing habitat use at microhabitat spatial scales (patches on the order of a few square metres). Mark–recapture methods have a prolific legacy of fisheries applications to account for imperfect detection, but the assumptions and inferences associated with these methods are better suited for survival and abundance over larger spatial and temporal scales (e.g., Letcher and Horton 2008; Perry et al. 2010). Multipass and removal methods are frequently used to asses detectability while enumerating salmonids in lotic environments, but the assumptions of removal methods are difficult to fulfill (Bryant 2000; Rosenberger and Dunham 2005), leading to estimates that are not only frequently biased, but biased according to habitat and species characteristics (Peterson et al. 2004). This is particularly problematic when the inference goal is relating local abundances of fish to fine-scale habitat variables. Further, removal methods are commonly collected via electrofishing techniques that can result in musculoskeletal trauma of both target and bycatch species (Panek and Densmore 2013), and hence electrofishing might not be possible when sensitive species inhabit sampling areas. Telemetry has also been recently used to asses characteristics of riverine fish habitat use (Capra et al. 2017), but telemetry experiments are expensive to conduct (requiring specialized equipment), generally incorporate relatively small samples sizes, can negatively impact fish implanted with transmitters (Jepsen et al. 2015), and might be better suited to asses broader-scale demographics like movement and survival (Ebner and Thiem 2009).

N-mixture models (Royle 2004) are a class of models that hold promise for assessing juvenile salmonid habitat use at smaller scales. These models use site-specific replicated point counts to estimate both detection probability and abundance. Although the origins of N-mixture models lie in population abundance estimation, they have proven useful for quantifying how abundance varies with habitat covariates (Graves et al. 2011), while simultaneously accounting for imperfect detection (Kéry 2008). The benefits of N-mixture models are numerous, including flexibility regarding parametric distributions for the abundance and detection components of the model and the ability to include the same covariates for both abundance and detection (Kéry 2008). Additionally, N-mixture models obviate the need for marking or identifying individuals (Kéry 2008), a key advantage for their use with juvenile salmonids that are often too small for marking and too difficult to re-identify. Occupancy (presence–absence) models also account for imperfect detection (MacKenzie 2005) without the need to re-identify individuals and have proven useful in modeling the distribution of riverine salmonids (Rodtka et al. 2015) and other stream fishes (Ferreira et al. 2016). However, by incorporating counts instead of presence–absence observations, N-mixture models estimate the intensity of use by fish at survey sites rather than only the probability of presence (Sethi and Benolkin 2013).

N-mixture models assume that site-specific abundance follows a count (e.g., Poisson) distribution where the mean abundance can be expressed as a function of habitat covariates. However, site-specific abundance will vary through space and time owing to factors other than habitat. For example, interannual variation in juvenile salmonid abundance is driven by parental spawner abundance, whereas intra-annual variation is induced by mortality and migration (for anadromous salmonids). Consequently, when population abundance is low, suitable habitat may remain unoccupied, whereas at high population abundance, competition may push individuals into suboptimal habitat. Variation in abundance over and above that explained by habitat covariates can be incorporated into N-mixture models in a number of ways, including random effects (Latimer et al. 2006), zero-inflated Poisson models (Zuur et al. 2009), or by explicitly modeling the spatiotemporal evolution of abundance (Conn et al. 2015). Overdispersion is a commonly encountered issue in models of ecological count data (Martin et al. 2005) and occurs when the variance is greater than that specified by a model's mean and variance structure (Faraway 2006). For example, overdispersion in the detection component can be induced by lack of independence among individuals (Martin et al. 2011) or can occur in count and detection components when important covariates are not modeled (Kéry and Schaub 2012, p. 82). Although N-mixture models provide a powerful tool for understanding how abundance varies in response to habitat covariates, extensions allowing for different sources of variation are needed to adequately model variation in the data (Wu et al. 2015).

For riverine fish, habitat quality is commonly modeled as a function of the depth and velocity of water and the distance to nearest cover (Persinger et al. 2011). These functions allow relative amounts of habitat to be summarized over spatial units for applications to population dynamics models that link habitat amounts to density-dependant demographics parameters. The size of spatial units can vary widely depending on the application and target species, with common examples in lotic applications ranging from tens (Gard 2014) to thousands (Heath et al. 2013) of metres in stream length. Habitat models in this context are often summarized as habitat suitability indices (HSI). HSI values range between 0 and 1, with values closer to 1 indicating desirable habitats likely to be selected by the focal species, and values closer to 0 suggesting poor-quality habitat that may be avoided.

Herein, we develop a field sampling design and associated statistical model for estimating abundance of juvenile salmon in relation to microhabitat variables and present an example application with data from the Trinity River, California, USA. The field sampling design was novel in that two-dimensional hydrodynamic models were used a priori to select sampling sites that yielded contrast in the habitat covariates of water depth, velocity, and distance to cover. A double observer snorkel survey was used to make independent counts of juvenile salmonids at selected sites. We then developed an N-mixture model with a binomial sampling protocol designed to estimate (*i*) spatial and temporal variance in abundance over that accounted for by habitat covariates, (*ii*) overdispersion in count data, and (*iii*) diver-specific detection efficiency. The model was implemented in a Bayesian framework that allowed the variance in abundance to be modeled via a zero-inflated Poisson distribution with spatial and temporal random effects and for overdispersion to be estimated in both the abundance and detection components of the model. We also highlight how our model can be naturally summarized for applications to population dynamics modeling.

Statistical model

N-mixture models

N-mixture models are a general class of models that assume observed data arise via a hierarchical structure working on a metapopulation (i.e., a collection of spatially distinct local populations; Royle and Dorazio 2008, p. 269). The first process is the unobserved local abundance "state process", and the second is an observation process leading to the observed counts. Metapopulation-level parameters drive local abundances and detection probabilities. The most standard implementation of the N-mixture model is the Poisson binomial (Kéry and Schaub 2012, p. 385), where local abundances (N_i) follow a Poisson distribution with mean λ_i, and replicated counts ($C_{i,k}$) follow a binomial distribution (with detection probability $p_{i,k}$) conditional on the local abundance:

$$\begin{aligned} [N_i|\lambda_i] &= \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \mathbf{x}_i'\alpha \\ [C_{i,k}|p_{i,k}, N_i] &= \text{binomial}(p_{i,k}, N_i) \\ \text{logit}(p_{i,k}) &= \mathbf{v}_{i,k}'\boldsymbol{\delta} \end{aligned} \quad (1)$$

where i references location, k references sampling occasion, $\mathbf{x}$ and $\mathbf{v}$ are design matrices containing covariate values, and $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ are vectors of regression parameter values. In this formula and throughout, square brackets ([]) designate a probability distribution, and vertical bars (|) indicate a conditional probability. As is common, the linear predictors for the Poisson mean and binomial detection probability are expressed as their canonical link functions, the log and logit ($\text{logit}(\boldsymbol{p}) = \log(\boldsymbol{p}(1-\boldsymbol{p})^{-1})$), respectively.

Mixed effects zero-inflated Poisson binomial mixture model

We next extend the model in eq. 1 to represent the abundance process as a zero-inflated Poisson distribution, akin to Wenger and Freeman (2008):

$$\begin{aligned} [N_i|\lambda_i, \omega_i] &= \text{Poisson}(\lambda_i\omega_i) \\ [\omega_i] &= \text{Bernoulli}(\pi_i) \\ \log(\lambda_i) &= \mathbf{x}_i'\alpha \\ \text{logit}(\pi_i) &= \mathbf{z}_i'\boldsymbol{\zeta} \\ [C_{i,k}|p_{i,k}, N_i] &= \text{binomial}(p_{i,k}, N_i) \\ \text{logit}(p_{i,k}) &= \mathbf{v}_{i,k}'\boldsymbol{\delta} \end{aligned} \quad (2)$$

where $\omega_i \in \{0, 1\}$, π_i is the probability of an observation having a strictly positive Poisson mean, $\mathbf{z}$ is a design matrix containing covariate values, $\boldsymbol{\zeta}$ is a vector of regression parameter values, the linear predictor for $\boldsymbol{\pi}$ is expressed as the canonical logit link function, and all other terms are as in eq. 1. In this formulation, zeros could arise from either the zero-inflation or Poisson mean component of the abundance process; the Bernoulli trial could result in a 0 ($\omega_i = 0$) bounding Poisson draws to a single outcome of 0, or the Poisson mean could be a small enough real number to result in stochastic outcomes of 0 (with $\omega_i = 1$):

$$\begin{aligned} [N_i = 0] &= (1 - \pi_i) + \pi_i \exp(-\lambda_i) \\ [N_i = n|n > 0] &= \pi_i \exp(-\lambda_i)(\lambda_i)^n(n!)^{-1} \end{aligned} \quad (3)$$

and we return to these probabilities later for population dynamics model applications.

Finally, we extend our model's linear predictors to account for our spatial and temporal random effects (Latimer et al. 2006), overdispersion (Wu et al. 2015), and an offset to account for variation in the size of areas sampled.

$$\begin{aligned} [N_\mathbf{g}|\lambda_\mathbf{g}, \omega_\mathbf{g}] &= \text{Poisson}(\lambda_\mathbf{g}\omega_\mathbf{g}) \\ [\omega_\mathbf{g}] &= \text{Bernoulli}(\pi_\mathbf{g}) \\ \log(\lambda_\mathbf{g}) &= \mathbf{x}_\mathbf{g}'\alpha + \tau_t + \xi_s + \phi_i + \log(A_i) \\ \text{logit}(\pi_\mathbf{g}) &= \mathbf{z}_\mathbf{g}'\boldsymbol{\zeta} + \varphi_t + \eta_s \\ [C_{\mathbf{g},k}|p_{\mathbf{g},k}, N_\mathbf{g}] &= \text{binomial}(p_{\mathbf{g},k}, N_\mathbf{g}) \\ \text{logit}(p_{\mathbf{g},k}) &= \mathbf{v}_{\mathbf{g},k}'\boldsymbol{\delta} + \gamma_t + \kappa_s + \epsilon_{i,k} \end{aligned} \quad (4)$$

$$\begin{aligned} &[\boldsymbol{\tau}] = \text{N}(0, \sigma_\tau^2) \quad [\boldsymbol{\xi}] = \text{N}(0, \sigma_\xi^2) \quad [\boldsymbol{\phi}] = \text{N}(0, \sigma_\phi^2) \\ &[\boldsymbol{\varphi}] = \text{N}(0, \sigma_\varphi^2) \quad [\boldsymbol{\eta}] = \text{N}(0, \sigma_\eta^2) \\ &[\boldsymbol{\gamma}] = \text{N}(0, \sigma_\gamma^2) \quad [\boldsymbol{\kappa}] = \text{N}(0, \sigma_\kappa^2) \quad [\boldsymbol{\epsilon}] = \text{N}(0, \sigma_\epsilon^2) \end{aligned}$$

where A represents the size of area sampled, $\mathbf{g}$ denotes the i,t,s index representing each sample's (i) spatial level (s) and temporal level (t). Random effects distributions are logit-normal on the response scale of the zero-inflation and detection probability linear predictors and log-normal for the abundance linear predictor.

Sampling

Study area

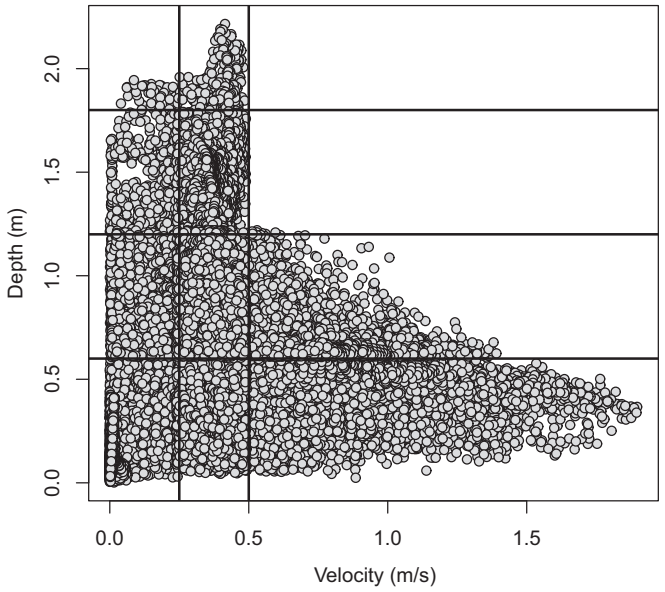
The Trinity River is located in northwestern California and flows 181 km from an anadromous barrier (Lewiston Dam) to its confluence with the Klamath River. The watershed drainage area is approximately 7700 km^2, and approximately 25% of the drainage area lies above Lewiston Dam. The most upstream 64 km of the mainstem (hereinafter: restoration reach) is the focus of a river restoration program. The Trinity River Restoration Program (TRRP; www.trrp.net) aims to restore anadromous native fish populations whose numbers declined after dam construction and water diversion, and legacy effects of mining and timber harvest. A fish population dynamics model is being constructed to inform habitat rehabilitation and water management decisions, and the dynamics model requires estimates of habitat availability.

Sampling design

We followed model-based sampling design principles in the generation of our survey design. An optimal regression sampling design is most efficient at estimating true regression parameter values when samples are allocated to the regions of the multivariate space of the explanatory variables with the greatest contrast (Kiefer 1958). To efficiently detect relationships that might be nonlinear (via quadratic regression terms) or to estimate interacting effects among the suite of explanatory variables, samples should also be collected across the mutually occurring combinations of those variables. Selecting sampling locations in this manner requires prior knowledge of depth, velocity, and distance to cover throughout the sampling domain. As these values are generally not known a priori, we relied on two-dimensional hydrodynamic models (2DHMs). Use of 2DHMs to inform optimal sampling designs has previously been demonstrated in Alexander et al. (2016). We used nine 2DHMs that were spatially distributed along the 64 km restoration reach. Each 2DHM predicts water depth and velocity over 400 m sections of the river, with computational cells varying between 0.25 and 0.56 m^2. More details about the construction and validation of these models can be found in Alvarez et al. (2015).

Ahead of each sampling visit to each river location's corresponding 2DHM boundaries (hereinafter: site), we extracted 2DHM output associated with current river discharges and plotted the bivariate relationship between spatially indexed predicted depths and velocities. All observations were binned according to relative combinations of depth and velocity (Fig. 1). The majority of samples were selected from bins in or near the four most extreme corners (Fig. 1), with roughly seven samples per bin. At some sites at certain flow levels, few or no samples existed for some corner bins (e.g., the deepest and swiftest bins), or conditions were not

Fig. 1. Bivariate plot of spatially indexed depth and velocity values predicted by a two-dimensional hydrodynamics model at one site and discharge. Vertical and horizontal lines form bins from which spatially stratified individual sample locations were selected.



safe for sampling. Roughly three samples were selected from the remaining bins. We selected the sampling locations from each bin systematically with a simple spatial stratification procedure. First, all locations falling within each bin were sorted according to their riverine distance from Lewiston Dam. Second, the sorted list of locations was bracketed into as many groups as there were samples to be drawn. A randomly selected location was drawn from the most upstream bracket, and subsequent samples were selected systematically from the remaining brackets. This method ensured that samples were well distributed along the entire length of each site. To additionally design our data collection to capture variation associated with cover, the entire process was replicated separately for locations within 1 m of cover and locations considered further from cover (>1 m). This procedure resulted in sample sizes of 60–75 sampling locations at each site for each time that crews sampled a site.

The sample sizes at each site required a 2-week window for data to be collected at all nine sites. The 2-week windows represent time blocks, of which nearly eight full blocks were able to be completed between February and April 2013. Several sites were not surveyed in some blocks due to rare cases of storm-based dangerous flow conditions or excessively turbid water. For model fitting and parameter estimation, the time blocks were distilled to four time periods by combining adjacent blocks because of similarities in site-level observed fish counts as the survey season progressed.

Data collection

Data collection at each site, in each time block, occurred in two phases to avoid disturbance of juvenile fish that could bias surveys. In the first phase, crews snorkeled upstream using global positioning system (GPS) equipment to locate an individual sampling location. If hydraulic conditions did not match those specified at the target location, crews sampled the nearest location within 30 m of the original point matching the desired depth, velocity, and distance to cover. Once located, crews visually surveyed the area immediately around the point to delineate a zone of relatively homogeneous depth and velocity conditions (hereinafter: polygons). To maintain homogeneity, depth and velocity values were restricted to lie within 0.15 m and 0.15 $m \cdot s^{-1}$, respectively, of the target values. Polygons were limited to a maximum of 2 m on the shortest dimension to ensure visibility and proper species identification by the surveyors. Polygon corners were marked with painted rocks whose GPS coordinates were recorded. The depth and velocity were measured at the center of each polygon and recorded, and distance to cover was measured at each polygon's edge. This process led to polygons that ranged from 0.3 to 42.5 m^2, with a median value of 3.8 m^2.

The second phase of the sampling consisted of fish counting surveys that occurred at least 24 h after polygon delineation. Surveys were always conducted between 1100 and 1600 by two simultaneous divers. Using snorkel equipment, the pair would approach the polygon moving upstream and simultaneously count only fish within the polygon boundary. Fish counts were also separated by species and size class, with fry being defined as fish up to 50 mm in length and presmolts being defined as fish larger than 50 mm. These size class thresholds were chosen to be consistent with other TRRP monitoring programs. Though the two counts were obtained simultaneously, to maintain independence of the counts the snorkelers did not communicate with each other regarding spotting or counting of fish, and each count was provided to a data recorder beyond earshot of the other diver. Snorkeling pairs remained together for all surveys conducted within a single day, and pairs often remained together throughout the survey season. Each snorkeler's name was recorded with each of their counts for use as a fixed effects covariate. For the analysis herein, we focus only on counts of Chinook salmon (*Oncorhynchus tshawytscha*) fry.

Model reduction and fitting

In all, nearly 1800 paired counts were collected. We anticipated long run times for model fitting due to the large sample size and complexity of our model. We therefore proceeded with two phases of model building, akin to Miller et al. (2011). First, we used maximum likelihood techniques to evaluate the weight of evidence for including potential covariates. Second, inference proceeded from a Bayesian implementation of the model with selected covariates.

Covariate exploration

To more rapidly evaluate the merits of quadratic and interactive terms among the depth, velocity, and distance to cover covariates, we began by fitting a more constrained version of the model described above. To fit the model using maximum likelihood, both time and space were considered fixed effects, and a global overdisperion parameter was estimated. We commenced with estimation using functions suitably modified from the R package "unmarked" (Fiske and Chandler 2011). Acknowledging the potential for overdispersion, we began by estimating the parameters of a "full model" (Table 1) and estimated $\hat{c}$ via a deviance-based bootstrap procedure deriving $\hat{c}$ as the ratio of the observed full model deviance to the mean deviance among simulated model fits. Given $\hat{c}$, covariate evaluation proceeded in three stages.

The stages unfolded under the premise that the observed counts are conditioned on the detection process and then further on the zero-inflation component of the abundance process. Hence, the first stage consisted of holding parameters in the linear predictors for $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}$ fixed at their respective full model forms, and the quasi-Akaike information criterion (QAIC; Burnham and Anderson 2002, p. 70) was applied to evaluate a priori sets of reduced parameters for the p linear predictor. In the second stage, the full model elements of $\boldsymbol{\lambda}$ were again held fixed, the set of $\boldsymbol{p}$ linear predictor elements arising from the first stage were applied, and a priori reductions in the elements of the $\boldsymbol{\pi}$ linear predictor were evaluated via QAIC. Finally, the third stage consisted of evaluating a priori reductions in the elements of the $\boldsymbol{\lambda}$ linear predictor conditional on the $\boldsymbol{p}$ and $\boldsymbol{\pi}$ linear predictors arising from the first two stages. The resulting set of covariates we define as the "inference" model and consisted of reductions in the linear predictors of $\boldsymbol{p}$

Table 1. Covariates used for each parameter in the "full" model and retained in the "inference" model during the first phase of model evaluation.

Parameter	Covariates
Full model	
$\boldsymbol{\lambda}$	T_{cat}, S_{cat}, D, V, D2C, D^2, V^2, D2C^2, $D \times V$, $D \times$ D2C, $V \times$ D2C, $T_{cat} \times S_{cat}$
$\boldsymbol{\pi}$	T_{cat}, D, V, D2C, $T_{cat} \times D$, $T_{cat} \times V$, $T_{cat} \times$ D2C
$\boldsymbol{p}$	O_{cat}, Vis, D, T_{cat}
Inference model	
$\boldsymbol{\lambda}$	T_{cat}, S_{cat}, D, V, D2C, D^2, V^2, D2C^2, $D \times V$, $D \times$ D2C, $V \times$ D2C, $T_{cat} \times S_{cat}$
$\boldsymbol{\pi}$	T_{cat}, D, V, D2C
$\boldsymbol{p}$	O_{cat}, D, T_{cat}

Note: "Parameter" references which parameter's linear predictor in eq. 4 the listed "Covariates" contribute to. The subscript "cat" denotes covariates that were included as categorical variables, with all other variables included as continuous covariates. Covariates are as follows: T = time period, S = site-level spatial location, O = individual snorkelers, Vis = measured distance of visibility, D = water depth, V = water mean-column velocity, and D2C = the distance to nearest cover. A superscript of 2 denotes a quadratic term, and a "×" symbol denotes an interaction term. A log(Area) offset was also included in the linear predictor for $\boldsymbol{\lambda}$.

(removed water visibility term) and $\boldsymbol{\pi}$ (removal of all interaction terms), but no reductions in the linear predictor for $\boldsymbol{\lambda}$ (Table 1).

Parameter estimation and inference

The model complexity (inclusion of random effects in multiple linear predictors of our model) and the hierarchical structure induced by the latent and observation components of our model are well-suited for a Bayesian modeling framework. Using the now common expression of Berliner (1996), we can express the posterior distribution relative to the joint factorization of the data model (count generating process), process model (abundance generating process), and parameter model (priors; Cressie and Wikle 2011). Letting θ_Λ denote the set of parameters associated with the abundance linear predictors, θ_p denote the set of parameters associated with the detection probability linear predictor, and $\boldsymbol{\theta} = \{\theta_\Lambda, \theta_p\}$, we write the joint posterior of the data, the latent variables, and the parameters as

$$[\boldsymbol{\theta}, N|C] \propto [C|N, \boldsymbol{\theta}][N|\theta_\Lambda][\boldsymbol{\theta}] \quad (5)$$

where the proportionality is up to a normalizing constant, and C and N are as defined previously. We constructed the likelihood and specified prior distributions using BUGS language, and called JAGS (Plummer 2014) from R via the package "jagsUI" (Kellner 2014) to use Markov chain Monte Carlo (MCMC) simulation to draw samples from the joint posterior distribution of the parameters. Prior to drawing posterior samples, all continuous covariates were centered and scaled for numeric stability. For all regression coefficients, we specified mean-zero Gaussian priors with precision (variance^{-1}) values equaling 0.00001. For the variance components (each $\sigma_{\cdot}$ in eq. 4), we specified uniform (a, b) priors (Gelman 2006), with $a = 0$ and $b = 10$.

We ran three simultaneous MCMC chains and retained 3000 samples per chain after a burn-in period of 50 000 samples and a thinning rate of 250 (i.e., 9000 samples per parameter were summarized for inference). Convergence was assessed visually from the traceplots of each MCMC chain and quantitatively using the Rhat statistic (Gelman et al. 2014). Goodness of fit was assessed using a Bayesian p value with a χ^2 discrepancy measure (Kéry and Schaub 2012, p. 402).

Table 2. Summaries of parameter posterior densities, separated into the zero-inflation parameter linear predictor, the Poisson mean linear predictor, and the binomial detection probability linear predictor.

Parameter	Mean	LCL	UCL
Zero-inflation (ζ)			
Intercept	1.14	−1.30	3.69
Velocity	−0.88	−1.08	−0.67
Depth	−0.73	−0.91	−0.56
D2C	−0.52	−0.68	−0.36
σ_φ	2.33	0.96	6.18
σ_η	0.48	0.22	0.96
Abundance (α)			
Intercept	−1.86	−2.60	−1.19
Velocity	−0.68	−0.85	−0.52
Depth	0.01	−0.14	0.16
D2C	−0.68	−0.91	−0.44
Depth^2	0.05	−0.02	0.13
Velocity^2	−0.01	−0.12	0.11
D2C^2	0.14	0.06	0.23
Depth × Velocity	0.20	0.04	0.37
Velocity × D2C	0.06	−0.10	0.23
Depth × D2C	−0.12	−0.31	0.08
σ_ϕ	1.38	1.30	1.48
σ_τ	0.67	0.25	1.76
σ_ξ	0.43	0.23	0.81
Detection (δ)			
Diver1	1.20	0.81	1.60
Diver2	0.75	0.29	1.21
Diver3	1.17	0.85	1.49
Diver4	1.22	0.82	1.61
Diver5	1.07	0.67	1.47
Diver6	1.04	0.70	1.37
Diver7	1.01	0.65	1.37
Diver8	0.46	0.02	0.92
Diver9	1.10	0.68	1.51
Diver10	1.14	0.65	1.63
Diver11	1.23	0.87	1.57
Diver12	0.29	−0.12	0.71
Depth	−0.26	−0.38	−0.13
σ_ϵ	1.22	1.09	1.37
σ_γ	0.12	0.01	0.46
σ_κ	0.08	0.00	0.26

Note: Point estimates (Mean) represent posterior distribution means, and the lower (LCL) and upper limits (UCL) of 95% credible intervals were created using the 0.025 and 0.975 quantiles of posterior distribution values. "D2C" = the distance to cover covariate. A superscript of 2 represents a quadratic coefficient, and a "×" symbol indicates a coefficient for the interaction of covariates. All reported values are on the scale of their respective link function.

Population model application

As noted above, a goal of this work was to create a function for generating HSI-type values aimed for inclusion in a population dynamics model. By incorporating the abundance component of our model, we can directly relate a probabilistic measure to intensity of use and avoid many concerns with habitat indices based on univariate preference curves (Railsback 2016). Further, by incorporating the posterior variation for each parameter, the generated HSI-type values reflect the parameter uncertainty of our model. A natural HSI-type measure in our context is $\psi = 1 -$ probability (abundance = 0) (Royle et al. 2005). Referencing eq. 3:

$$\begin{aligned} \psi_i &= 1 - [N_i = 0] \\ &= 1 - (1 - \pi_i) - \pi_i \exp(-\lambda_i) \end{aligned} \quad (6)$$

Fig. 2. Posterior densities of the highest (dashed line) and lowest (solid line) estimated probably of detection among survey divers, calculated at mean water depth and averaged over temporal and spatial effects. Posterior densities were created by applying the inverse-logit transformation to linear predictors representing the linear combination of logit-scale posterior distributions.

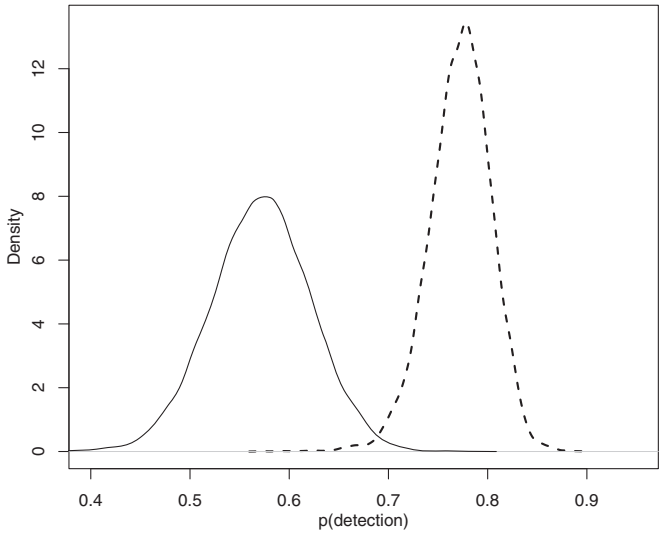
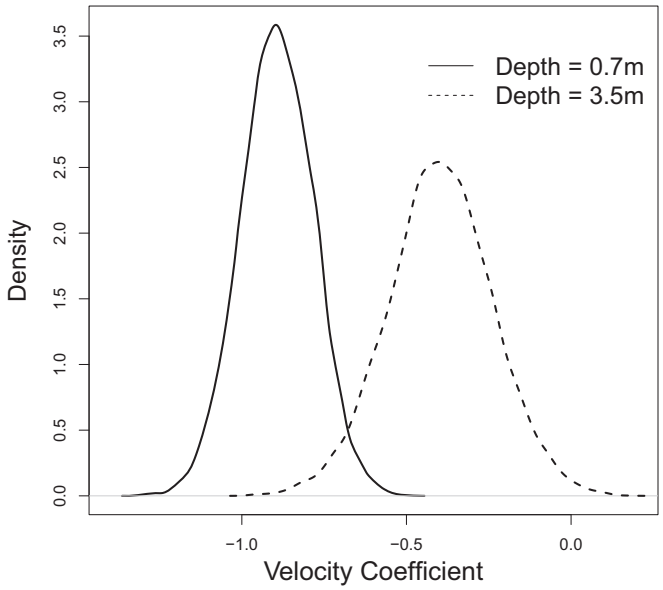


Fig. 3. Posterior densities of the estimated effects of velocity on abundance at average distances to cover, for both a relatively shallow location (0.7 m; solid line) and a relatively deep location (3.5 m; dashed line).



Results

The Rhat values for each parameter were $\ll$1.1, suggesting that all three MCMC chains converged to the same posterior space in all cases. Visually, the traceplots showed no indication that further burn-in, thinning, or additional posterior samples were necessary to proceed with inference. The Bayesian p value for these data and model was 0.16 and did not provide compelling evidence of lack of fit.

Based on posterior distribution estimates, we found evidence that detection probability varied with depth and among divers (Table 2: Detection). Detection probability estimates ranged from 0.57 to 0.77 (means of individual diver posterior distributions), at average water depths (Fig. 2). The 95% credible interval for the depth coefficient contained only negative values and suggested that detection probability declined with increasing depths (Table 2: Detection).

Posterior distribution estimates suggested that both the presence and abundance of fry were associated with the measured physical variables at the sampling locations. The probability of a site containing fry decreased with increasing water depth, velocity, and distance to cover (Table 2: Zero-Inflation). Given fry were present at a site, the posterior distributions suggested that abundance decreased with increasing distance from cover (Table 2: Abundance). Whereas there was evidence that increasing velocities lead to decreased abundances, the negative interaction term between velocity and depth revealed that the magnitude of velocity's effect on abundance declined with increasing depth (Table 2: Abundance and Fig. 3). In general, there was much less posterior evidence that other interactions and the quadratic terms were associated with variation in local abundances (Table 2: Abundance).

The posterior distributions for the standard deviations representing overdispersion for the Poisson mean (given abundance) and binomial detection linear predictors have no probability mass near zero, and suggest their inclusion captured extra Poisson and binomial variation, respectively (Table 2: Abundance, Detection). The posterior distributions for both the spatial and temporal random effects parameters reveal that detection probability showed little evidence of variation across either space or time, but that both the zero-inflation and Poisson mean abundance components showed greater variation over time than across space (Table 2). As expected, given the few number of temporal levels (4), the posterior distributions for the temporal random effects parameters are more diffuse than those for space (Table 2; Fig. 4).

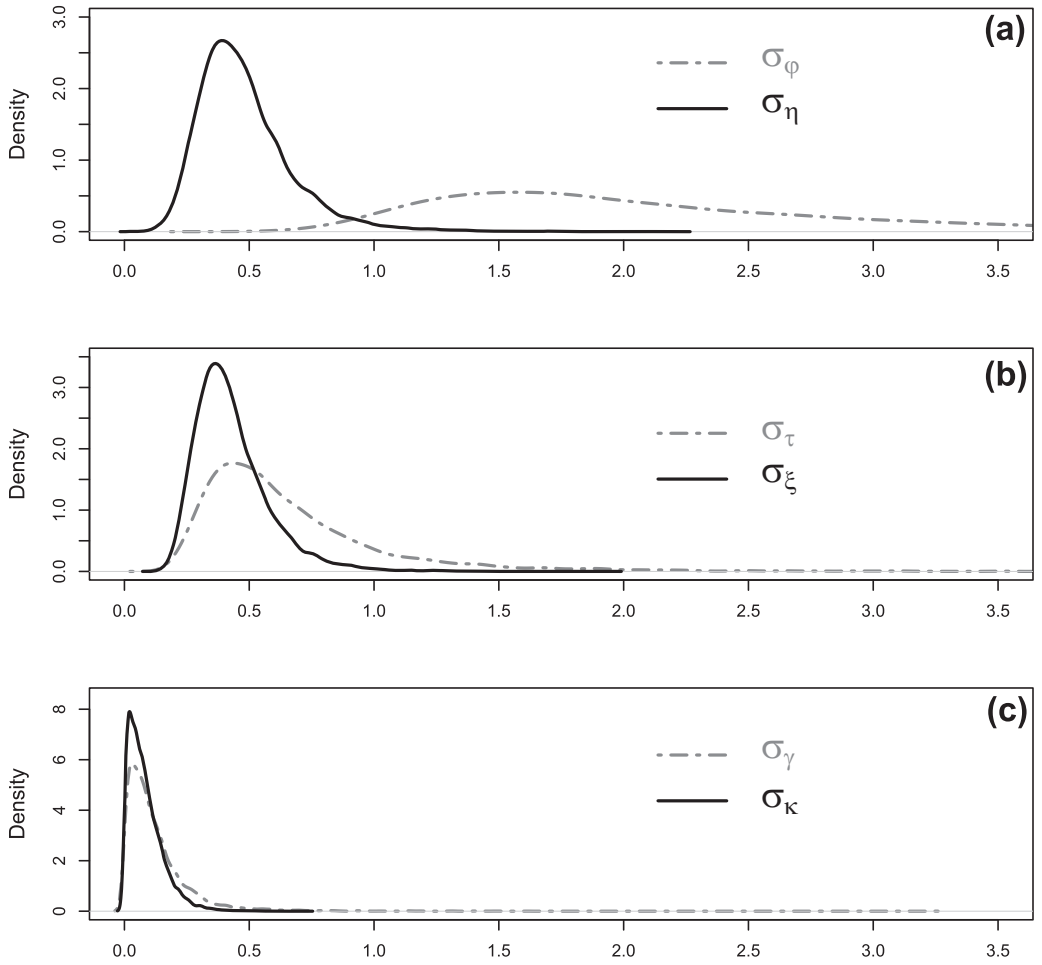
Given these results, the most highly suitable locations for Chinook salmon fry are those with mutually low values of the physical variables (i.e., shallower and slower water close to cover). In addition to statistical evidence for the relative strength of parameters, the utility of applying the model for population dynamics purposes could be assessed in its ability to predict discerning ψ values (i.e., not predict ψ values near 0.5 in all places). Our model and ψ generated according to eq. 6 do distinguish among habitat quality predictions according to varying levels of depth, velocity, and distance to cover (Fig. 5).

Discussion

Our analytical framework extends current approaches for estimating resource selection in riverine microhabitats by switching the focus of the response variable from presence–absence to abundance–fish density. The consequence is that the fundamental driver of microhabitat occupancy arises not from presence or absence per se, but from how fish density varies with key habitat variables. Thus, in our framework, occupancy arises intuitively as the probability that fish density is greater than zero, given the habitat covariates (eq. 6; Royle and Dorazio 2008).

Fish density is a more relevant response variable reflecting the outcome of population processes, which is often the focus of management actions. For example, historical uses of HSI were geared towards estimating the amount of useable habitat area in a presence–absence context (Railsback 2016), whereas contemporary applications are geared towards translating habitat area into more ecologically relevant metrics of abundance, such as habitat capacity (i.e., the maximum abundance that a given habitat area can support; Bartz et al. 2006; Ayllón et al. 2012*b*; Beechie et al. 2015). These approaches estimate capacity by combining estimates of habitat area with an estimate of maximum density. Our analytical framework retains favorable elements of HSI, but also extends inference to factors affecting fish density in microhabitats, which ultimately influences the capacity of streams to support juvenile salmonid populations.

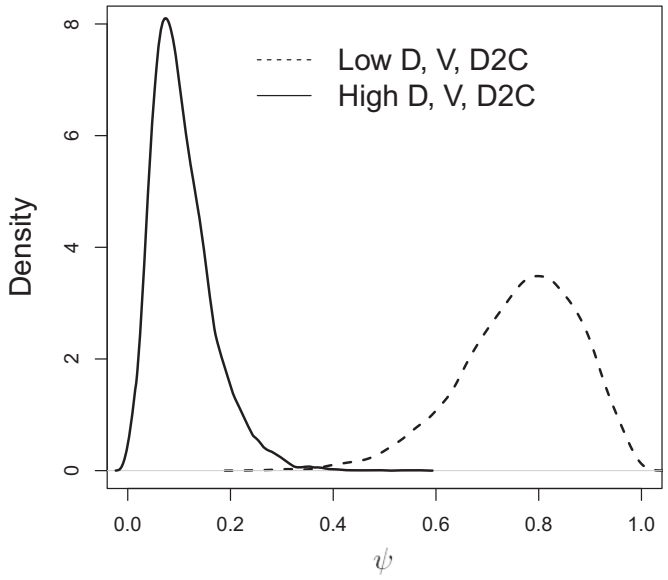
Fig. 4. Posterior density plots for the temporal (dashed line) and spatial (solid line) random effects in the (*a*) zero-inflation, (*b*) abundance, and (*c*) detection linear predictors of our model, each estimated from 9000 MCMC draws. The legend symbols in each plot correspond to their definitions in eq. 4 and summaries in Table 2.



Our model structure is also explicitly designed to fully quantify variation in fish density over that explained by habitat covariates. This aspect of our approach is critically important because it is well recognized that variation in habitat use arises from the interplay of a suite of abiotic and biotic factors, other than microhabitat variables, such as water temperature, food availability, growth potential, competition, and predation risk (Wall et al. 2016; Hayes et al. 2016; Rosenfeld et al. 2016). In our study, we expected to observe variation in local densities unrelated to habitat because (*i*) spawner densities varied spatially over the study area, and (*ii*) the study was conducted over the course of the juvenile residence period when fry initially emerge from spawning gravels, rear in nearshore habitats, grow, and then disperse downstream as they move towards the ocean. Variation in density owing to these population dynamics was captured, albeit at a course resolution, by inclusion of space and time random effects. However, even after accounting for these effects, our analysis measured non-negligible overdispersion in fish density, providing evidence that habitat covariates alone were insufficient to fully explain observed variation in fish density. Although our goal was to relate fish density to microhabitat variables, our modeling framework is amenable to inclusion of other covariates (e.g., invertebrate drift) that may help to explain variation in fish density that is otherwise relegated to unexplained variance as quantified by the overdispersion term.

We originally intended to model the co-dependency between the temporal and spatial effects in our model and data, but under simulations with similar levels of time and space effects (four levels for time, nine levels for space), we found the estimation power of their correlation too weak to alter prior distributions. A more refined and continuous treatment of the spatial and temporal random effects could lead to more precise parameter inference (Cressie and Wikle 2011) and potentially reduce or remove either of the overdispersion components of our model (Faraway 2006). On the temporal side, the coarse and small number of time periods was not nearly enough to consider continuous time-series models. Our spatial units (sites) were also relatively coarse, though we could have considered continuous spatial (i.e., geostatistical) models that modeled correlation as a continuous function of distance among samples within each site, albeit using a dimension reduction technique given large number of samples. Our sampling design did spatially stratify samples along the length of each site, and the point counts occurred in relatively small and discrete habitat areas. Given the distribution of samples within each of the sites, the site-level spatial random effects included, and the overdispersion parameters included at multiple portions of our model, we feel that conditional on the habitat covariate values, residual spatial autocorrelation is unlikely to drastically effect our posterior estimates. However, a continuous spatial model could be explored in the future, particularly in light of the estimated overdispersion. Indeed, this highlights a benefit of the model structure we have presented here, in that users have tremendous flexibility regarding components of the multiple linear predictors.

Fig. 5. Posterior density plots for $\psi = 1 - p(\text{abundance} = 0)$ for relatively low (10th percentile; dashed line) and high (90th percentile; solid line) values of water depth (*D*) and velocity (*V*) and distance to cover (D2C). The linear predictors include 1 standard deviation of each (spatial, temporal) random effects parameter for both the zero-inflation and abundance linear predictors. The density lines are smoothed over 9000 MCMC draws.



We acknowledge that the large amount of data presented in our example application is not attainable by all practitioners. The N-mixture model does require replicated samples, but this might be more accessible than capture–recapture methods because identification of individuals is not required, and replication is not required at all sampling locations (i.e., detection information can be borrowed across sites; Kéry and Schaub 2012, p. 385). It is also important to acknowledge that the N-mixture model is not solely responsible for the large amount of data collected for this study. First, the temporal and spatial breadth of our sampling would be required of any study aimed to capture the full suite of variation in local abundances attributed to time and space, even if methods to account for imperfect detection were not employed. This need will vary tremendously according to target species' life-history characteristics. Second, the data size was also influenced by our sampling design, which sought to capture the full breadth of covariate values. As this was the first Chinook salmon microhabitat study to our knowledge that accounted for detectability and spatiotemporal effects, we deemed it prudent to extensively sample the range of possible covariate values. Finally, we considered the need to model complex (e.g., interaction and quadratic terms) covariate relationships, which often result in relatively large data sets (Guillera-Arroita 2017). As traditional HSI approaches assume noninteractive covariate relationships, they naturally require relatively smaller samples sizes. The impacts to inference of reduced sampling according to any of the elements listed above could certainly be evaluated by the rich data set that was collected for this study.

Although not required for N-mixture models, our optimized sampling method relied on an extensive amount of auxiliary information, and we recognize that 2DHMs will not be available in all cases. Hydraulic models are likely to be available where flow–habitat relationships are desired, because 2DHMs are often applied to simulate hydraulic conditions across discharges in studied river sections (Rosenfeld et al. 2016). For other application settings, there are many other ways to tailor sampling designs to meet desired outcomes, and differing inference goals will lead to differences in optimal riverine sampling (Som et al. 2014). To optimize sampling around variables of potential habitat importance, one could utilize other information that might be more readily available than 2DHMs, like aerial photographs (Fitzgerald et al. 2006). Additionally, one could design sampling around differing habitat unit types (Pusey et al. 1998) or avoid reliance on preexisting habitat maps with systematic sampling (Hankin and Reeves 1988). There are many resources available to help fisheries scientists create efficient survey designs aligned with their inference goals, and these could be combined with recent research related to sampling for N-mixture models (Kowalewski et al. 2015) to inform future studies.

Traditional HSI approaches, like habitat suitability curves, are still the most commonly applied methods for habitat simulations in instream flow studies (Rosenfeld et al. 2016). This is likely due to their historical legacy and "ready-to-use" availability (Conallin et al. 2010). Despite widespread use, traditional HSI methods have received substantial critique. Railsback (2016) has recently summarized many of these complaints, which include selection of inappropriate spatial scales, assumptions of independence and equal effects among habitat variables, output indices without clear ecological meaning, and a general avoidance of modern modeling techniques. Another common complaint is the lack of measured uncertainty in habitat model outputs (Ayllón et al. 2012*a*; Zajac et al. 2015; Turner et al. 2016). Further, despite the fact that fisheries researchers have accounted for detectability as or more frequently than those studying other taxa (Kellner and Swihart 2014), this attention has not carried into HSI-type methods. Our approach and application of an N-mixture model have addressed all of these concerns. We opted to measure habitat variables and fish use at a microhabitat scale relevant to juvenile salmonids and to include variables (e.g., distance to cover) relevant to juvenile salmonid behavior. Our candidate N-mixture models included interactive terms among covariates, posed no restrictions on the weighting of importance for individuals effects, and naturally produced an output with clear ecological interpretation (probability of presence weighted directly by intensity of abundance). Finally, our estimation procedure resulted in posterior distribution samples that easily allow propagation of parameter uncertainties into habitat model outputs. These improvements were all conducted within a framework that explicitly accounted for imperfect detection. We believe these qualities make a compelling case for riverine habitat modeling via N-mixture models. As these models are relatively new in fisheries applications, we have focused on model details and an example of sampling design, analysis, and interpretation of the results. Akin to Rodtka et al. (2015), we will allow future work to directly compare the performance of output provided by our analysis with prior methods like HSI.

Estimating abundance in microhabitats while accounting for imperfect detection is fundamentally challenging because standard approaches that require geographic closure (e.g., multiple-pass removal sampling: Rosenberger and Dunham 2005; mark–recapture methods: Hillman et al. 1992; Thurow et al. 2006) cannot be applied in this setting. Our estimates of detection probability were relatively high (57% to 77% at average depths; Fig. 2) compared with detection probability estimates reported in other calibrated snorkel surveys. This is likely due to differences in the spatial scale of habitat units that we sampled (median area of 3.8 m^2), as detectability within microhabitats is generally not comparable to other studies that estimate detectability over mesohabitat size areas. For instance, Hillman et al. (1992) estimated snorkeling detection probabilities of 20%–50% in survey reaches that were 50–200 m in stream length, and Thurow et al. (2006) estimated snorkeling efficiencies of 12.5%–33.2% in reaches averaging 100 m in length. Some of the variation in these reported efficiencies can also be attributed to salmonid species diel behavior and other recorded attributes such as fish size and water temperature. Although ju-

venile Chinook salmon have been documented to be nocturnal and remain concealed in substrate during daytime (Bradford and Higgins 2001), our pilot surveys revealed little difference in fish counts between day and night samples in this system (Pinnix et al. 2016). It should also be noted that passive fish observation techniques will never detect fish that are completely hidden from the field of view over the duration of observation (no matter how many replicate surveys are conducted). Thus, given the availability to be observed, we attribute high detection probability to our focus on small microhabitat patches in which surveyors could focus on enumerating every potentially observable fish.

In some very recent work, Barker et al. (2017) detail concerns in N-mixture models. The authors highlight the critical loss of information when estimating detection without marked individuals and state that parameter identifiability in N-mixture models is afforded owing to the hierarchical structure of the model and strong assumptions on binomial detection, and requires high data quality. They note that under small values of detection probability, the variance-to-mean relationship for the detection component of the model can mimic that of a Poisson distribution (unrelated to the assumed distribution for abundances in the model), and this departure from the binomial assumption can lead to parameter identifiability problems, which can also arise when a data set does not account for variation in detection (i.e., their so-called "constant p" assumption). They also suggest a lack of methods for assessing specific assumptions in N-mixture models. While this may be the case in the maximum likelihood framework of their presentation, we opted for a Bayesian methodology, which granted us great flexibility to account for detection variation above that specified by the covariates and maintained a strict hierarchical model structure, and we applied a goodness-of-fit measure sensitive to distributional assumption deviations for both the abundance and detection components of the model. We have also noted the loss of detection information when individuals are not marked, and we accounted for this loss in the generation of our sampling design (and large sample size). Further, our results suggested estimates of detection probability much larger than those considered by Barker et al. (2017). Given these aspects of our design (and resulting high data quality), model, and results, coupled with extensive simulations of our model prior to applications with our data, and a rich literature of simulations for estimation in N-mixture models, we do not expect that concerns raised by Barker et al. (2017) have affected our results. The analysis by Barker et al. (2017) highlights the conditions under which inference from N-mixture models may be less reliable than in our study. Indeed, practitioners should recognize that much information is lost without marked individuals, and this information loss must be offset with thoughtful study design, simulation testing, and high levels for replication of repeated count data.

Our results suggest a less independent effect of water depth than has been previously reported in habitat selection calculations for Chinook salmon and other salmonid fry (Beakes et al. 2014; Gard 2014; Hardy et al. 2006; Hayes and Jowett 1994). The interaction estimates suggested that depth's role mitigated the effects of increasing velocity on reduced abundance, and this makes sense given that velocity is measured at 60% height from the riverbed to the water surface. As depths increase, the velocity at 60% water column height could potentially be much faster than near-bed velocities more suitable for drift-feeding fish (Railsback and Harvey 2011). Further, our parameter estimates suggested that depth impacted detection probability, and this highlights the potential for spurious inference when not modeling detection in scenarios where variables impact both abundance and detection, as suggested by Kéry and Royle (2010).

Although our sampling design and model could be fit in either a maximum likelihood or Bayesian framework, we found a number of advantages to the Bayesian framework. Although spatial and temporal blocks can be incorporated as fixed effects in a maximum likelihood framework for N-mixture models, fixed effects cannot be generalized to estimate expected variation in local abundance. In contrast, expressing spatial and temporal blocks as random effects allows for propagation of uncertainty in abundance owing to demographic processes driving spatiotemporal variation in abundance. In addition, the Bayesian framework allowed us to assess which process (i.e., abundance and (or) detection) generated overdispersion in the count data, as opposed to an omnibus estimate of overdispersion (e.g., $\hat{c}$; Burnham and Anderson 2002) applied to N-mixture models under a maximum likelihood framework (Mazerolle 2016). Finally, we were able to evaluate the strength of spatial and temporal random effects on abundance relative to detection. In our case, there was evidence that spatiotemporal variation was much larger for abundance than detection. This result is intuitive for a species that emerged and migrated over the course of our sampling period, and this relative relationship in variation has been observed in other organisms (Isaac et al. 2011; Dénes et al. 2015).

We were able to demonstrate the utility of a N-mixture model for assessing habitat selection of riverine juvenile salmonids. In addition to accounting for detectability, the model structure and Bayesian implementation proved quite flexible in incorporating structural and residual variance components. Further, we were able to formulate parameter estimates into a natural HSI congener that has the added benefit of interpretation as a probabilistic statement, as opposed to a simple index. N-mixture models show great promise for future applications to riverine habitat selection and components of population dynamics models.

Acknowledgements

The authors thank two anonymous reviewers whose insightful comments improved the quality of the manuscript. The authors also thank the field crew that worked very hard to collect the data discussed in this article and Aaron Martin and Sean Ledwin for thoughtful discussions regarding field survey techniques essential to the success of this project. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the US Fish and Wildlife Service. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Ahmadi-Nedushan, B., St-Hilaire, A., Bérubé, M., Robichaud, E., Thiémonge, N., and Bobée, B. 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Res. Appl. **22**: 503–523. doi:10.1002/rra.918.

Ahmadi-Nedushan, B., St-Hilaire, A., Bérubé, M., Ouarda, T., and Robichaud, E. 2008. Instream flow determination using a multiple input fuzzy-based rule system: a case study. River Res. Appl. **24**: 279–292. doi:10.1002/rra.1059.

Alexander, J., Bartholomew, J.L., Wright, K., Som, N.A., and Hetrick, N.J. 2016. Integrating models to predict distribution of the invertebrate host of myxosporean parasites. Freshw. Sci. **35**(4): 1263–1275. doi:10.1086/688342.

Alldredge, J.R., and Dasgupta, N. 2003. Multiple comparisons in resource selection using logistic regression. J. Agric. Biol. Environ. Stat. **8**(3): 356–366. doi:10.1198/1085711032219.

Alvarez, J., Goodman, D.H., Martin, A., Som, N.A., Wright, K.A., and Hardy, T.B. 2015. Development and validation of two-dimensional hydrodynamic models on the Trinity River, California. Technical Report Arcata Fish and Wildlife Office, Arcata Fisheries Technical Report Number TR 2015-24. US Fish and Wildlife Service, Arcata, Calif.

Ayllón, D., Almodóvar, A., Nicola, G.G., and Elvira, B. 2012*a*. The influence of variable habitat suitability criteria on PHABSIM habitat index results. River Res. Appl. **28**(8): 1179–1188. doi:10.1002/rra.1496.

Ayllón, D., Almodóvar, A., Nicola, G.G., Parra, I., and Elvira, B. 2012*b*. Modelling carrying capacity dynamics for the conservation and management of territorial salmonids. Fish. Res. **134**: 95–103. doi:10.1016/j.fishres.2012.08.004.

Baasch, D.M., Tyre, A.J., Millspaugh, J.J., Hygnstrom, S.E., and Vercauteren, K.C. 2010. An evaluation of three statistical methods used to model resource selection. Ecol. Modell. **221**(4): 565–574. doi:10.1016/j.ecolmodel.2009.10.033.

Barker, R.J., Schofield, M.R., Link, W.A., and Sauer, J.R. 2017. On the reliability of N-mixture models for count data. Biometrics. doi:10.1111/biom.12734.

Bartz, K.K., Lagueux, K.M., Scheuerell, M.D., Beechie, T., Haas, A.D., and Ruckelshaus, M.H. 2006. Translating restoration scenarios into habitat conditions: an initial step in evaluating recovery strategies for Chinook salmon

(*Oncorhynchus tshawytscha*). Can. J. Fish. Aquat. Sci. **63**(7): 1578–1595. doi:10.1139/f06-055.

Beakes, M.P., Moore, J.W., Retford, N., Brown, R., Merz, J.E., and Sogard, S.M. 2014. Evaluating statistical approaches to quantifying juvenile Chinook salmon habitat in a regulated California river. River Res. Appl. **30**: 180–191. doi:10.1002/rra.2632.

Beechie, T.J., Pess, G.R., Imaki, H., Martin, A., Alvarez, J., and Goodman, D.H. 2015. Comparison of potential increases in juvenile salmonid rearing habitat capacity among alternative restoration scenarios, Trinity River, California. Restor. Ecol. **23**(1): 75–84. doi:10.1111/rec.12131.

Berliner, L.M. 1996. Hierarchical bayesian time series models. *In* Maximum Entropy and Bayesian Methods: Santa Fe, New Mexico, U.S.A. 1995 Proceedings of the Fifteenth International Workshop on Maximum Entropy and Bayesian Methods. *Edited by* K.M. Hanson and R.N. Silver. Springer, Dordrecht, the Netherlands. pp. 15–22. doi:10.1007/978-94-011-5430-7_3.

Boone, E.L., Stewart-Koster, B., and Kennard, M.J. 2012. A hierarchical zero-inflated Poisson regression model for stream fish distribution and abundance. Environmetrics, **23**: 207–218. doi:10.1002/env.1145.

Bovee, K.D. 1986. Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology. Technical Report Instream Flow Information Paper 21. US Fish and Wildlife Service.

Boyce, M.S., Vernier, P.R., Nielsen, S.E., and Schmiegelow, F.K. 2002. Evaluating resource selection functions. Ecol. Modell. **157**(2–3): 281–300. doi:10.1016/S0304-3800(02)00200-4.

Bradford, M.J., and Higgins, P.S. 2001. Habitat-, season-, and size-specific variation in diel activity patterns of juvenile chinook salmon (*Oncorhynchus tshawytscha*) and steelhead trout (*Oncorhynchus mykiss*). Can. J. Fish. Aquat. Sci. **58**(2): 365–374. doi:10.1139/f00-253.

Brosse, S., Guegan, J.F., Tourenq, J.N., and Lek, S. 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Modell. **120**(2–3): 299–311. doi:10.1016/S0304-3800(99)00110-6.

Bryant, M.D. 2000. Estimating fish populations by removal methods with minnow traps in southeast Alaska streams. N. Am. J. Fish. Manage. **20**: 923–930. doi:10.1577/1548-8675(2000)020<0923:EFPBRM>2.0.CO;2.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. Springer, New York.

Capra, H., Plichard, L., Bergé, J., Pella, H., Ovidio, M., McNeil, E., and Lamouroux, N. 2017. Fish habitat selection in a large hydropeaking river: strong individual and temporal variations revealed by telemetry. Sci. Total Environ. **578**: 109–120. doi:10.1016/j.scitotenv.2016.10.155. PMID:27839764.

Conallin, J., Boegh, E., and Jensen, J.K. 2010. Instream physical habitat modelling types: an analysis as stream hydromorphological modelling tools for EU water resource managers. Int. J. River Basin Manage. **8**(1): 93–107. doi:10.1080/15715121003715123.

Conn, P.B., Johnson, D.S., Hoef, J.M.V., Hooten, M.B., London, J.M., and Boveng, P.L. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecol. Monogr. **85**(2): 235–252. doi:10.1890/14-0959.1.

Cressie, N., and Wikle, C.K. 2011. Statistics for spatio-temporal data. Probability and statistics. Wiley, Hoboken, New Jersey.

Dénes, F.V., Silveira, L.F., and Beissinger, S.R. 2015. Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. Methods Ecol. Evol. **6**(5): 543–556. doi:10.1111/2041-210X.12333.

Ebner, B.C., and Thiem, J.D. 2009. Monitoring by telemetry reveals differences in movement and survival following hatchery or wild rearing of an endangered fish. Mar. Freshw. Res. **60**: 45–57. doi:10.1071/MF08027.

Faraway, J.J. 2006. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Chapman & Hall/CRC, Boca Raton, Fla.

Ferreira, M., Filipe, A., Bardos, D.C., Magalhães, M.F., and Beja, P. 2016. Modeling stream fish distributions using interval-censored detection times. Ecol. Evol. **6**(15): 5530–5541. doi:10.1002/ece3.2295. PMID:27551402.

Fiske, I., and Chandler, R. 2011. unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. J. Stat. Soft. **43**(10): 1–23. doi:10.18637/jss.v043.i10.

Fitzgerald, D.G., Zhu, B., Hoskins, S.B., Haddad, D.E., Green, K.N., Rudstam, L.G., and Mills, E.L. 2006. Quantifying submerged aquatic vegetation using aerial photograph interpretation: application in studies assessing fish habitat in freshwater ecosystems. Fish, **31**(2): 61–73. doi:10.1577/1548-8446(2006)31[61:FHQSAV]2.0.CO;2.

Gard, M. 2014. Modelling changes in salmon habitat associated with river channel restoration and flow-induced channel alterations. River Res. Appl. **30**: 40–44. doi:10.1002/rra.2642.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. **1**(3): 515–533. doi:10.1214/06-BA117A.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. 2014. Bayesian data analysis. 3rd ed. CRC Press, Boca Raton, Fla.

Graves, T.A., Kendall, K.C., Royle, J.A., Stetz, J.B., and Macleod, A.C. 2011. Linking landscape characteristics to local grizzly bear abundance using multiple detection methods in a hierarchical model. Anim. Conserv. **14**(6): 652–664. doi:10.1111/j.1469-1795.2011.00471.x.

Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography, **40**(2): 281–295. doi:10.1111/ecog.02445.

Hankin, D.G., and Reeves, G.H. 1988. Estimating total fish abundance and total habitat area in small streams based on visual estimation methods. Can. J. Fish. Aquat. Sci. **45**(5): 834–844. doi:10.1139/f88-101.

Hardy, T.B., Shaw, T., Addley, R.C., Smith, G.E., Rode, M., and Belchik, M. 2006. Validation of chinook fry behavior-based escape cover modeling in the lower Klamath River. Int. J. River Basin Manage. **4**(3): 169–178. doi:10.1080/15715124.2006.9635286.

Harvey, B.C., and Railsback, S.F. 2009. Exploring the persistence of stream-dwelling trout populations under alternative real-world turbidity regimes with an individual-based model. Trans. Am. Fish. Soc. **138**: 348–360. doi:10.1577/T08-068.1.

Hastie, T., and Fithian, W. 2013. Inference from presence-only data; the ongoing controversy. Ecography, **36**(8): 864–867. doi:10.1111/j.1600-0587.2013.00321.x. PMID:25492992.

Hayes, J.W., and Jowett, I.G. 1994. Microhabitat models of large drift-feeding brown trout in three New Zealand rivers. N. Am. J. Fish. Manage. **14**(4): 710–725. doi:10.1577/1548-8675(1994)014<0710:MMOLDF>2.3.CO;2.

Hayes, J.W., Goodwin, E., Shearer, K.A., Hay, J., and Kelly, L. 2016. Can weighted useable area predict flow requirements of drift-feeding salmonids? Comparison with a net rate of energy intake model incorporating drift-flow processes. Trans. Am. Fish. Soc. **145**(3): 589–609. doi:10.1080/00028487.2015.1121923.

Heath, M.W., Wood, S.A., Brasell, K.A., Young, R.G., and Ryan, K.G. 2013. Development of habitat suitability criteria and in-stream habitat assessment for the benthic cyanobacteria *Phormidium*. River Res. Appl. doi:10.1002/rra.2722.

Hefley, T.J., and Hooten, M.B. 2016. Hierarchical species distribution models. Curr. Landsc. Ecol. Rep. **1**(2): 87–97. doi:10.1007/s40823-016-0008-7.

Hillman, T.W., Mullan, J.W., and Griffith, J.S. 1992. Accuracy of underwater counts of juvenile chinook salmon, coho salmon, and steelhead. N. Am. J. Fish. Manage. **12**(3): 598–603. doi:10.1577/1548-8675(1992)012<0598:AOUCOJ>2.3.CO;2.

Holsinger, L., Keane, R.E., Isaak, D.J., Eby, L., and Young, M.K. 2014. Relative effects of climate change and wildfires on stream temperatures: a simulation modeling approach in a Rocky Mountain watershed. Clim. Change, **124**: 191–206. doi:10.1007/s10584-014-1092-5.

Isaac, N.J.B., Cruickshanks, K.L., Weddle, A.M., Marcus Rowcliffe, J., Brereton, T.M., Dennis, R.L.H., Shuker, D.M., and Thomas, C.D. 2011. Distance sampling and the challenge of monitoring butterfly populations. Methods Ecol. Evol. **2**(6): 585–594. doi:10.1111/j.2041-210X.2011.00109.x.

Jepsen, N., Thorstad, E.B., Havn, T., and Lucas, M.C. 2015. The use of external electronic tags on fish: an evaluation of tag retention and tagging effects. Anim. Biotelem. **3**(1): 49. doi:10.1186/s40317-015-0086-z.

Kellner, K. 2014. jagsUI: Run JAGS (specifically, libjags) from R; an alternative user interface for rjags. R package version 1.1.

Kellner, K.F., and Swihart, R.K. 2014. Accounting for imperfect detection in ecology: a quantitative review. PloS ONE, **9**(10): e111436. doi:10.1371/journal.pone.0111436. PMID:25356904.

Kéry, M. 2008. Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships. Auk, **125**(2): 336–345. doi:10.1525/auk.2008.06185.

Kéry, M., and Royle, J.A. 2010. Hierarchical modelling and estimation of abundance and population trends in metapopulation designs. J. Anim. Ecol. **79**: 453–461. doi:10.1111/j.1365-2656.2009.01632.x. PMID:19886893.

Kéry, M., and Schaub, M. 2012. Bayesian population analysis using WinBUGS. Academic Press, Oxford, UK.

Kiefer, J. 1958. On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. Ann. Math. Stat. **29**(3): 675–699. doi:10.1214/aoms/1177706530.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R.M., White, M., and Stone, L. 2017. Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection. Methods Ecol. Evol. **8**: 420–430. doi:10.1111/2041-210X.12738.

Kowalewski, L.K., Chizinski, C.J., Powell, L.A., Pope, K.L., and Pegg, M.A. 2015. Accuracy or precision: implications of sample design and methodology on abundance estimation. Ecol. Modell. **316**: 185–190. doi:10.1016/j.ecolmodel.2015.08.016.

Labonne, J., Allouche, S., and Gaudin, P. 2003. Use of a generalised linear model to test habitat preferences: the example of *Zingel asper*, an endemic endangered percid of the River Rhone. Freshw. Biol. **48**(4): 687–697. doi:10.1046/j.1365-2427.2003.01040.x.

Lahoz-Monfort, J.J., Guillera-Arroita, G., and Wintle, B.A. 2014. Imperfect detection impacts the performance of species distribution models. Glob. Ecol. Biogeogr. **23**: 504–515. doi:10.1111/geb.12138.

Latimer, A.M., Wu, S., Gelfand, A.E., and Silander, J.A., Jr. 2006. Building statistical models to analyze species distributions. Ecol. Appl. **16**(1): 33–50. doi:10.1890/04-0609. PMID:16705959.

Letcher, B.H., and Horton, G.E. 2008. Seasonal variation in size-dependent survival of juvenile Atlantic salmon (*Salmo salar*): performance of multistate

capture-mark-recapture models. Can. J. Fish. Aquat. Sci. **65**(8): 1649–1666. doi:10.1139/F08-083.

MacKenzie, D.L. 2005. What are the issues with presence–absence data for wildlife managers? J. Wildl. Manage. **69**: 849–860. doi:10.2193/0022-541X(2005)069[0849:WATIWP]2.0.CO;2.

Martin, J., Royle, J.A., Mackenzie, D.I., Edwards, H.H., Kry, M., and Gardner, B. 2011. Accounting for non-independent detection when estimating abundance of organisms with a Bayesian approach. Methods Ecol. Evol. **2**(6): 595–601. doi:10.1111/j.2041-210X.2011.00113.x.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., and Possingham, H.P. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. **8**(11): 1235–1246. doi:10.1111/j.1461-0248.2005.00826.x. PMID:21352447.

Mazerolle, M.J. 2016. AICcmodavg: model selection and multimodel inference based on (Q)AIC(c). R package version 2.1-0.

Miller, M.W., Pearlstine, E.V., Dorazio, R.M., and Mazzotti, F.J. 2011. Occupancy and abundance of wintering birds in a dynamic agricultural landscape. J. Wildl. Manage. **75**(4): 836–847. doi:10.1002/jwmg.98.

Newcomb, T.J., Orth, D.J., and Stauffer, D.F. 2007. Habitat evaluation. *In* Analysis and interpretation of freshwater fisheries data. *Edited by* M.L. Brown and C.S. Guy. American Fisheries Society, Bethesda, Md. pp. 843–886.

Panek, F.M., and Densmore, C.L. 2013. Frequency and severity of trauma in fishes subjected to multiple-pass depletion electrofishing. N. Am. J. Fish. Manage. **33**(1): 178–185. doi:10.1080/02755947.2012.754803.

Perry, R.W., Skalski, J.R., Brandes, P.L., Sandstrom, P.T., Klimley, A.P., Ammann, A., and MacFarlane, B. 2010. Estimating survival and migration route probabilities of juvenile chinook salmon in the Sacramento–San Joaquin River Delta. N. Am. J. Fish. Manage. **30**(1): 142–156. doi:10.1577/M08-200.1.

Persinger, J.W., Orth, D.J., and Averett, A.W. 2011. Using habitat guilds to develop habitat suitability criteria for a warmwater stream fish assemblage. River Res. Appl. **27**(8): 956–966. doi:10.1002/rra.1400.

Peterson, J.T., Thurow, R.F., and Guzevich, J.W. 2004. An evaluation of multipass electrofishing for estimating the abundance of stream-dwelling salmonids. Trans. Am. Fish. Soc. **133**(2): 462–475. doi:10.1577/03-044.

Phillips, S.J., Anderson, R.P., and Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. **190**: 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

Pinnix, W.D., De Juilio, K., Petros, P., and Som, N.A. 2016. Feasibility of snorkel surveys for determining relative abundance and habitat associations of juvenile chinook salmon on the mainstem Trinity River, California. Technical Report Arcata Fish and Wildlife Office, Arcata Fisheries Technical Report Number TR 2016-24. US Fish and Wildlife Service, Arcata, Calif.

Plummer, M. 2014. Jags: a program for analysis of Bayesian graphical models using Gibbs sampling [online]. *In* Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), 20–22 March 2003, Vienna, Austria. ISSN 1609-395X. *Edited by* Kurt Hornik, Friedrich Leisch, and Achim Zeilies. Available from https://www.r-project.org/conferences/DSC-2003/Proceedings/.

Pusey, B.J., Kennard, M.J., Arthur, J.M., and Arthington, A.H. 1998. Quantitative sampling of stream fish assemblages: single- vs multiple-pass electrofishing. Aust. J. Ecol. **23**: 365–374. doi:10.1111/j.1442-9993.1998.tb00741.x.

Railsback, S.F. 2016. Why it is time to put PHABSIM out to pasture. Fisheries, **41**(12): 720–725. doi:10.1080/03632415.2016.1245991.

Railsback, S.F., and Harvey, B.C. 2011. Importance of fish behaviour in modelling conservation problems: food limitation as an example. J. Fish Biol. **79**: 1648–1662. doi:10.1111/j.1095-8649.2011.03050.x. PMID:22136244.

Rodtka, M.C., Judd, C.S., Aku, P.K.M., and Fitzsimmons, K.M. 2015. Estimating occupancy and detection probability of juvenile bull trout using backpack electrofishing gear in a west-central Alberta watershed. Can. J. Fish. Aquat. Sci. **72**(5): 742–750. doi:10.1139/cjfas-2014-0175.

Rosenberger, A.E., and Dunham, J.B. 2005. Validation of abundance estimates from mark–recapture and removal techniques for rainbow trout captured by electrofishing in small streams. N. Am. J. Fish. Manage. **25**: 1395–1410. doi:10.1577/M04-081.1.

Rosenfeld, J.R., Beecher, H., and Ptolemy, R. 2016. Developing bioenergetic-based habitat suitability curves for instream flow models. N. Am. J. Fish. Manage. **36**(5): 1205–1219. doi:10.1080/02755947.2016.1198285.

Royle, J.A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics, **60**: 108–115. doi:10.1111/j.0006-341X.2004.00142.x. PMID:15032780.

Royle, J.A., and Dorazio, R.M. 2008. Hierarchical modeling and inference in ecology. Academic Press, London, UK.

Royle, J.A., Nichols, J.D., and Kéry, M. 2005. Modelling occurrence and abundance of species when detection is imperfect. Oikos, **110**(2): 353–359. doi:10.1111/j.0030-1299.2005.13534.x.

Royle, J.A., Chandler, R.B., Yackulic, C., and Nichols, J.D. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods Ecol. Evol. **3**(3): 545–554. doi:10.1111/j.2041-210X.2011.00182.x.

Sandoval-Solis, S., Teasley, R.L., McKinney, D.C., Thomas, G.A., and Patiño Gomez, C. 2013. Collaborative modeling to evaluate water management scenarios in the Rio Grande basin. J. Am. Water Resour. Assoc. **49**(3): 639–653. doi:10.1111/jawr.12070.

Scheuerell, M.D., Hilborn, R., Ruckelshaus, M.H., Bartz, K.K., Lagueux, K.M., Haas, A.D., and Rawson, K. 2006. The Shiraz model: a tool for incorporating anthropogenic effects and fish-habitat relationships in conservation planning. Can. J. Fish. Aquat. Sci. **63**(7): 1596–1607. doi:10.1139/f06-056.

Sethi, S.A., and Benolkin, E. 2013. Detection efficiency and habitat use to inform inventory and monitoring efforts: juvenile coho salmon in the Knik River basin, Alaska. Ecol. Freshw. Fish, **22**(3): 398–411. doi:10.1111/eff.12034.

Som, N.A., Monestiez, P., Ver Hoef, J.M., Zimmerman, D.L., and Peterson, E.E. 2014. Spatial sampling on streams: principles for inference on aquatic networks. Environmetrics, **25**(5): 306–323. doi:10.1002/env.2284.

Som, N.A., Goodman, D.H., Perry, R.W., and Hardy, T.B. 2015. Habitat suitability criteria via parametric distributions: estimation, model selection and uncertainty. River Res. Appl. **32**(5): 1128–1137. doi:10.1002/rra.2900.

Thurow, R.F., Peterson, J.T., and Guzevich, J.W. 2006. Utility and validation of day and night snorkel counts for estimating bull trout abundance in first- to third-order streams. N. Am. J. Fish. Manage. **26**: 217–232. doi:10.1577/M05-054.1.

Turner, D., Bradford, M.J., Venditti, J.G., and Peterman, R.M. 2016. Evaluating uncertainty in physical habitat modelling in a high-gradient mountain stream. River Res. Appl. **32**(5): 1106–1115. doi:10.1002/rra.2915.

Wall, C.E., Bouwes, N., Wheaton, J.M., Saunders, W.C., and Bennett, S.N. 2016. Net rate of energy intake predicts reach-level steelhead (*Oncorhynchus mykiss*) densities in diverse basins from a large monitoring program. Can. J. Fish. Aquat. Sci. **73**(7): 1081–1091. doi:10.1139/cjfas-2015-0290.

Wenger, S.J., and Freeman, M.C. 2008. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. Ecology, **89**(10): 2953–2959. doi:10.1890/07-1127.1. PMID:18959332.

Wu, G., Holan, S.H., Nilon, C.H., and Wikle, C.K. 2015. Bayesian binomial mixture models for estimating abundance in ecological monitoring studies. Ann. Appl. Stat. **9**(1): 1–26. doi:10.1214/14-AOAS801.

Zajac, Z., Stith, B., Bowling, A.C., Langtimm, C.A., and Swain, E.D. 2015. Evaluation of habitat suitability index models by global sensitivity and uncertainty analyses: a case study for submerged aquatic vegetation. Ecol. Evol. **5**(13): 2503–2517. doi:10.1002/ece3.1520. PMID:26257866.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. 2009. Mixed effects models and extensions in ecology with R. Springer, New York.