

# 1 Estimating Carrying Capacity for Juvenile Salmon using 2 Quantile Random Forest Models

3 Kevin E. See<sup>1,\*</sup>, Michael W. Ackerman<sup>1</sup>, Richard A. Carmichael<sup>1</sup>, Sarah L. Hoffmann<sup>1</sup>, and C

4 December 01, 2020

## 5 Abstract

6 Establishing robust methods and metrics to evaluate habitat quality is critical for the recov-  
7 ery of endangered Pacific salmonids. A variety of modeling approaches are used for status  
8 and trend monitoring of anadromous species throughout the Pacific Northwest, USA, but cur-  
9 rent methods may fail to capture the complex relationship between fish and habitat and are  
10 often limited in predictive power beyond specific watersheds. Further, the focus on species  
11 distribution and abundance is not easily manipulated to predict carrying capacity and tradi-  
12 tional stock-recruitment analyses are reliant on long-term data which are not always available.  
13 In this study, we developed a quantile random forest (QRF) model to provide estimates of  
14 habitat carrying capacity for Chinook salmon parr during the summer months, at both the site  
15 and watershed scale. QRF models allow for the consideration of noisy data, correlated vari-  
16 ables, and non-linear relationships: common features in fish-habitat datasets. We leveraged  
17 Columbia Habitat Monitoring Program (CHaMP) data to select habitat co-variates and predict  
18 capacity at those sites. We also identified a set of globally available attributes to extrapolate  
19 capacity estimate predictions throughout wadeable streams within the Columbia River basin.  
20 Total capacity estimates for watersheds closely matched estimates from alternative fish produc-  
21 tivity models. Carrying capacity estimates based on QRF, like those presented here, provide  
22 managers a framework to guide the identification, prioritization, and development of habitat

23 rehabilitation actions to recover salmon populations.

24 <sup>1</sup> Biomark, Inc. 705 South 8th St., Boise, Idaho, 83702, USA

25 \* Correspondence: Kevin E. See <Kevin.See@merck.com>

## 26 **Introduction**

27 The decline of anadromous Pacific salmonids (*Oncorhynchus spp.*) across the Pacific Northwest,  
28 USA has prompted numerous actions aimed at reversing that trend. These actions are often catego-  
29 rized into four H's: harvest modification, hatchery practices, hydrosystem operations, and habitat  
30 rehabilitation. Problematically, there is substantial uncertainty regarding the degree of change that  
31 can be exerted across and within these categories, and what combination of changes will most cost-  
32 effectively and sustainably reduce mortality. Freshwater habitat capacity deficits have recently  
33 been identified as a major factor directly impacting population abundance which has been largely  
34 overlooked in Columbia Basin salmonids (Bond et al. 2018, Hinrichsen and Paulsen 2020, NOAA  
35 Fisheries 2020). Specifically, restoring salmonid carrying capacity through tributary rehabilitation  
36 actions has been identified as a key component of recovery efforts for salmon and steelhead (*O.*  
37 *mykiss*) in the Pacific Northwest, USA (NOAA Fisheries 2016a, 2016b). Efforts have included in-  
38 creasing and improving existing habitat for both spawning adults and rearing juveniles. However,  
39 estimating habitat carrying capacity (both historic and contemporary) for various life-stages of Pa-  
40 cific salmon, as well as identifying important habitat characteristics that influence capacity, has  
41 been an ongoing but necessary challenge (Bond et al. 2018, Hinrichsen and Paulsen 2020, NOAA  
42 Fisheries 2020). Reliable methods to better understand fish-habitat relationships and estimate ca-  
43 pacity are necessary to identify those salmon and steelhead life-stages that are limited by habitat  
44 capacity to better direct tributary rehabilitation efforts.

45 When it comes to estimating carrying capacity, spawner-recruit models are the gold standard (Mous-  
46 salli and Hilborn 1986, Myers et al. 1999). However, such models require a long time-series of ac-

47 curate estimates of abundance for adults and juveniles, with variation in the number of adults. Such  
48 data are unavailable in most watersheds (Cramer and Ackerman 2009), and they do not necessarily  
49 allow one to link capacity to habitat characteristics, except perhaps at the watershed scale. Bioen-  
50 ergetics approaches, such as the net rate of energy intake (NREI) have been applied to salmonids  
51 to estimate capacity on the 200 - 600 m reach scale (Wall et al. 2016). However, there are some  
52 potential issues with how the food supply (i.e. invertebrate drift) is measured with these methods  
53 that could lead to biases in capacity estimates (Dodrill and Yackulic 2016) as well as difficulty  
54 in properly constraining drift depletion and inter-species competition, and computational and spa-  
55 tial limitation of this modelling approach (Wall et al. 2016, Carmichael et al. 2020). In addition,  
56 those authors did not take the step of scaling the capacity predictions at the reach scale to entire  
57 watersheds. In contrast, Sweka and Mackey (2010) estimated carrying capacity of Atlantic salmon  
58 (*Salmo salar*) parr at the watershed scale, using a quantile regression approach, but the only habitat  
59 covariate they included was cumulative drainage area. Estimates of salmonid carrying capacity  
60 that leverage fish-habitat relationships are lacking at the watershed scale in the Pacific Northwest.  
61 Most studies that have investigated fish-habitat relationships focus on predicting a species' distri-  
62 bution (presence / absence) or the average abundance or density: neither of which can be easily  
63 manipulated to predict carrying capacity. Further, many of these studies focus on only one or  
64 two measures of habitat. Fausch et al. (1988) conducted a thorough review of attempts to predict  
65 the abundance of fish from measurable habitat covariates from 1950 to 1985 and found that the  
66 vast majority of multiple linear regression models failed to detect a significant fish-habitat signal.  
67 Since that review, there has been progress in identifying some fish-habitat relationships for several  
68 salmonid species. Nickelson et al. (1992) demonstrated that juvenile coho salmon (*O. kisutch*)  
69 were found in higher densities within pool habitat on the Oregon coast. Similarly, pool and pond  
70 densities were good predictors of coho smolt densities in western Washington (Sharma and Hilborn  
71 2001). Bryant and Woodsmith (2009) found that juvenile coho abundance was positively related  
72 to large wood at the reach scale; however their results demonstrated a negative relationship be-  
73 tween abundance and the number of pools. Braun and Reynolds (2011) similarly found positive

74 associations between spawner densities of sockeye salmon (*O. nerka*) in the Fraser River and large  
75 wood, in addition to positive relationships to percent undercuts and percent pools. Densities of  
76 adult spawning coho were also higher in forested areas compared to urban or agricultural areas in  
77 the Snohomish River watershed (Pess et al. 2002). Mossop and Bradford (2006) examined juvenile  
78 Chinook salmon (*O. tshawytscha*) in the Yukon River and found positive correlations between the  
79 log of fish density and several metrics related to residual pool dimensions and large woody debris  
80 abundance as well as a negative correlation between fish density and gradient. These studies were  
81 focused on predicting observed fish densities, not necessarily capacity, and for most of them the  
82 predictive extent is limited to a particular watershed. In addition, they all assumed some form of  
83 linear fish-habitat relationship which often results in weak predictive power.

84 A number of studies have used other modeling approaches to elicit fish-habitat relationships. Dun-  
85 ham et al. (2002) used a quantile regression approach to show a negative relationship between  
86 cutthroat trout (*O. clarkii*) densities and the width:depth ratio of a stream for the upper quantiles of  
87 trout density. The same approach was also used to map the potential extent of sole (*Solea solea*)  
88 in the English Channel and southern North Sea (Eastwood et al. 2003). Machine learning models  
89 such as boosted regression trees and random forests have been used to examine species biomass,  
90 diversity, and distribution for a number of different species (Pittman et al. 2009, Knudby et al.  
91 2010, Compton et al. 2012). The results from these studies highlight the importance and effective-  
92 ness of using techniques that can accommodate non-linear fish-habitat relationships and provide  
93 motivation for furthering research in this realm.

94 For the purposes of this paper, we define carrying capacity as the maximum number of individu-  
95 als that can be supported given the quantity and quality of habitat available at a given life-stage.  
96 We assume that higher observed relative densities within a given life stage are a function of habitat  
97 quantity and quality. Furthermore, we assert that observed fish density is a poor proxy of habitat ca-  
98 pacity owing to both a paucity of individuals, especially for threatened or endangered species, and  
99 the existence of unmeasured variables that may serve to limit capacity. To address this, we have

100 developed a model to estimate juvenile rearing capacity for Pacific salmon in wadeable streams  
101 based on quantile random forest (QRF) (Meinshausen 2006) models using measurements of fish  
102 abundance and density and habitat characteristics. QRF models combine the theory and justifica-  
103 tion of quantile regression modeling (Koenker and Bassett Jr 1978, Cade and Noon 2003) with the  
104 flexibility and framework of random forest models (Breiman 2001). They account for unmeasured  
105 variables and can be used to describe the entire distribution of predicted fish densities for a given set  
106 of habitat conditions, not just the mean expected density. Random forest models have been shown  
107 to outperform more standard parametric models in predicting fish-habitat relationships in other con-  
108 texts (Knudby et al. 2010). Quantile random forests share many of the benefits of random forest  
109 models, such as the ability to capture non-linear relationships between independent and dependent  
110 variables, naturally incorporate interactions between covariates, and work with untransformed data  
111 while being robust to outliers (Prasad et al. 2006). In addition, quantile regression models have  
112 been used in a variety of ecological systems to estimate the effect of limiting factors (Terrell et al.  
113 1996, Cade and Noon 2003).

114 The fish abundance/density and habitat data used to fit the QRF model presented here were avail-  
115 able from seven watersheds within the interior Columbia River basin, Pacific Northwest, USA.  
116 Within the interior Columbia River basin two major runs of Chinook salmon occur, stream-type  
117 (i.e., spring/summer run) and ocean-type (i.e., fall run), each characterized by different life history  
118 characteristics. Stream-type Chinook salmon adults enter freshwater from the ocean earlier in the  
119 year, spawn in the upper reaches of a watershed, and the juveniles rear for up to 16 months in fresh-  
120 water before entering the ocean as smolts. Ocean-type Chinook salmon adults enter freshwater later  
121 (e.g., fall or winter) spawn lower in the watershed, and the juveniles may spend between several  
122 weeks and six months in freshwater before migrating to the ocean as subyearlings. Here, we focus  
123 on stream-type Chinook salmon, and in particular the juvenile summer rearing period during low  
124 flow, during which juveniles are often termed parr, referring to the camouflage markings that occur  
125 on their sides during this life-stage. Data presented here are from Chinook salmon populations in  
126 the Upper Columbia River spring-run and Snake River spring/summer-run Evolutionary Significant

127 Units (ESU). The Upper Columbia spring-run ESU is listed as endangered under the Endangered  
128 Species Act, the Snake River spring/summer-run is listed as threatened (NOAA Fisheries 2016a,  
129 2016b). Hereafter, we refer to both ESUs simply as Chinook salmon.

130 In this study, we developed a QRF model to:

- 131 • identify measured habitat characteristics that are most strongly associated with observed Chi-  
132 nook salmon parr abundance and density,
- 133 • elicit fish-habitat relationships for those habitat characteristics identified as important for  
134 determining fish abundance and density, using paired fish and habitat measurements,
- 135 • predict contemporary habitat carrying capacity at all sites where the important habitat char-  
136 acteristics are measured,
- 137 • extrapolate capacity predictions at measured habitat sites across a watershed using globally  
138 available attribute data to estimate the Chinook salmon parr capacity of that watershed, and
- 139 • validate estimates of carrying capacity from our approach across multiple watersheds using  
140 independent estimates of capacity (e.g., spawner-recruit relationships).

141 Our study incorporates multiple measures of stream habitat to estimate fish-habitat relationships  
142 that encompass the collinear nature of most stream habitat metrics and can be used to predict carry-  
143 ing capacity. Our approach moves across several spatial scales, inferring fish-habitat relationships  
144 from detailed, localized habitat data and extrapolating capacity predictions across wide swaths of  
145 unsampled locations. Additionally, this approach for estimating life-stage specific habitat-based  
146 carrying capacity can be used to quantitatively identify the magnitude of tributary habitat rehabil-  
147 itation necessary to support de-listing. Given the multitude of (often correlated) habitat metrics  
148 and the potentially non-linear fish-habitat relationships that define capacity as a function of habitat,  
149 we explore the application of QRF modeling to habitat capacity estimation, validated using data  
150 from Columbia River Chinook salmon. For perhaps the first time, the necessity of tributary habitat  
151 rehabilitation can be demonstrated, and the magnitude of required change can be placed in context  
152 with the other “H’s.”

## 153 **Methods**

### 154 **Study Site**

155 Habitat data used in our study were collected from eleven watersheds within the interior Columbia  
156 River basin, Pacific Northwest, USA (Figure 1). The Columbia River basin covers more than  
157 668,000 km<sup>2</sup> draining large portions of Idaho, Oregon, and Washington, and smaller portions of  
158 Montana, Nevada, Utah, and Wyoming, as well as the southeastern portion of British Columbia.  
159 The habitat data used to populate the QRF model were collected by the Columbia Habitat Monitor-  
160 ing Program (CHaMP) (Volk et al. 2017) and were downloaded from <https://www.champmonitorin>  
161 [g.org](https://www.champmonitoring.org). Data from the following eleven CHaMP watersheds were used in this study: Asotin, Entiat,  
162 John Day, Lemhi, Methow, Minam, South Fork Salmon, Tucannon, Upper Grande Ronde, We-  
163 natchee and Yankee Fork. Juvenile density and abundance data were collected in a subset of seven  
164 watersheds (see Table 1 and Figure 1), at CHaMP survey reaches and were graciously provided by  
165 a number of agencies and projects, including the Integrated Status and Effectiveness Monitoring  
166 Project (Volk et al. 2017).

### 167 **Data**

168 CHaMP sites are 200 m to 600 m reaches in wadeable streams across select watersheds within the  
169 interior Columbia River basin. The sites were selected based on a spatially balanced generalized  
170 random tessellation stratified sample selection algorithm (Stevens Jr and Olsen 1999, 2004). Habi-  
171 tat data within CHaMP sites were collected using the CHaMP protocol (CHaMP 2016) which calls  
172 for field data collection during the low-flow period, typically from June through October. CHaMP  
173 habitat data include, but are not limited to, measurements describing channel complexity, chan-  
174 nel units, disturbance, fish cover, large woody debris, riparian cover, stream size (depth, width,  
175 discharge), substrate, temperature, macroinvertebrate productivity, and water quality.

176 Juvenile fish surveys were conducted for Chinook salmon parr during the summer low-flow season

177 at many of the same sites surveyed using the CHaMP protocol. Survey methods included mark-  
178 recapture, three-pass removal sampling, two-pass removal sampling, single-pass electrofishing,  
179 and snorkeling. These data were used to estimate Chinook salmon parr abundance at all CHaMP  
180 sites where fish survey data were available. Three-pass removal estimates used the Carle-Strub  
181 estimator (Carle and Strub 1978), following advice from Hedger et al. (2013). Two-pass removal  
182 estimates used the estimator described by Seber (2002). Mark-recapture estimates used Chapman's  
183 modified Lincoln-Peterson estimator (Chapman 1951) and were deemed valid if they met the crite-  
184 ria described in Robson and Regier (1964). These estimates were made using the *removal* function  
185 from the *FSA* package (Ogle et al. 2020) or the *closedp.bc* function from the *Rcapture* package  
186 (Rivest and Baillargeon 2019) in R software (R Core Team 2019). Snorkel counts were trans-  
187 formed to abundance estimates using paired snorkel-electrofishing sites to calibrate snorkel counts.  
188 For sites with invalid estimates or that were sampled with a single electrofishing pass, we devel-  
189 oped an estimate of capture probability based on valid estimates, using a binomial generalized  
190 linear mixed effects model. Fixed effects were species, wetted width of the site, density of fish  
191 caught on the first pass and all possible two-way interactions. We included a random effect for fish  
192 crew/watershed. We used this model to predict abundances based on the number of fish caught on  
193 the first pass and any other covariates.

194 Abundance estimates at all sites were then translated into linear (parr/m) fish densities which were  
195 paired with the associated CHaMP habitat data. For sites that were sampled in multiple years,  
196 only the fish and habitat data from the year with the highest observed fish density was retained to  
197 avoid possible pseudo-replication, while remaining consistent with our goal of estimating carrying  
198 capacity. After removing duplicate sites, our initial dataset contained 327 unique sites with paired  
199 fish-habitat data (Table 1). We did explore using areal fish densities (parr/m<sup>2</sup>) as the response but  
200 found very similar results so in the interest of brevity we only present results based on linear fish  
201 densities.

## 202 **Habitat Covariate Selection**

203 A key step in developing a QRF model to predict fish capacities was selecting the habitat covari-  
204 ates to include in the model. The CHaMP program generated more than 100 habitat metrics at  
205 each site, many of which were correlated with each other to one degree or another, as is often the  
206 case with stream habitat variables. We sought to include a small set of covariates that were not  
207 overly redundant (i.e., not highly correlated), described many aspects of stream habitat (e.g., sub-  
208 strate, temperature, complexity, etc.) and were highly associated with the observed fish densities,  
209 presumably because they contained information about what types of habitat fish sought or avoided.  
210 Full details of how the twelve covariates used in the QRF model were selected can be found in  
211 Appendix S1.

## 212 **QRF Model Fit**

213 Using the selected habitat covariates (Table 2), we fit a QRF model to predict habitat rearing ca-  
214 pacity for Chinook salmon parr during summer months using the natural log of fish densities as  
215 the response. After constructing a random forest, predictions of the mean response can be made  
216 by averaging the predictions of all trees, similar to the expected value predictions from a statisti-  
217 cal regression model. The individual predictions from each tree, viewed collectively, describe the  
218 entire distribution of the predicted response; therefore, the random forest model can be used in the  
219 same way as other quantile regression methods to predict any quantile of the response. There were  
220 missing values for some habitat data; thus, any site visit with more than three missing covariates  
221 was removed from the dataset and the remaining missing habitat values were imputed using the  
222 *missForest* R package (Stekhoven and Bühlmann 2012, Stekhoven 2013). We fit the QRF mod-  
223 els using the *quantregForest* function from the *quantregForest* package (Meinshausen 2017) in R  
224 software (R Core Team 2019), incorporating data from 327 records (paired fish-habitat data) and  
225 twelve habitat covariates (27.2 data points per covariate) (Table 2). The 90th quantile of the pre-  
226 dicted distribution was used as a proxy for carrying capacity following the suggestion of Sweka  
227 and Mackey (2010), and to avoid higher quantiles that draw from the very upper tails of observed

228 fish density, where the variability of predictions may be influenced by small sample size issues.  
229 After model fitting, the QRF model was then used to predict capacity at sites with measurements  
230 of the habitat covariates that were used to fit the model. In our case, this includes all sites within  
231 CHaMP basins in the interior Columbia River basin. For CHaMP sites that were sampled in multi-  
232 ple years, we first calculated the mean for each habitat metric among years to make predictions. In  
233 total, we generated 589 predictions of Chinook salmon parr capacity, during summer months, for  
234 the following basins: Entiat, Grande Ronde (including Minam), John Day, Lemhi, Methow, South  
235 Fork Salmon, Tucannon, Wenatchee and Yankee Fork Salmon. CHaMP sampled between 1 and  
236 28% of the Chinook domain within these watersheds, with an average of 11%.

### 237 **Extrapolating to Other Sites**

238 To predict capacity at larger spatial scales, such as the watershed or population, we developed  
239 an extrapolation model based on globally available attributes (GAA) which were available for the  
240 entirety of tributary habitat utilized by a given population. The GAA data used here was taken from  
241 the list of generalized random tessellation stratified master sample sites that the CHaMP sites were  
242 originally selected from (Larsen et al. 2008, 2016). Possible covariates included temperature range,  
243 elevation, watershed, the first principal component of a natural feature classification and human  
244 disturbance classification (Whittier et al. 2011), the square root of cumulative drainage area, stream  
245 power, slope, channel type, bankfull width and average August temperature (Table 3). The natural  
246 log of the CHaMP site capacity predictions (parr/m) was used as the response variable in a multiple  
247 linear regression model that incorporated the design weights of the CHaMP sites using the *svyglm*  
248 function from the *survey* package (Lumley 2020) in R software (R Core Team 2019). The design  
249 weights are generated from how much of the watershed each site is meant to represent. Because the  
250 CHaMP sites were selected from strata that usually comprised unequal portions of that watershed,  
251 these weights must be accounted for to lead to unbiased model coefficients (Nahorniak et al. 2015).  
252 We fit two different extrapolation models, one that included watershed as a covariate for use in  
253 predicting capacity within CHaMP watersheds, and one that did not for predicting everywhere

254 else. We then made predictions of linear capacity at all master sample sites throughout the interior  
255 Columbia River basin, generally spaced about one kilometer apart. These points do not represent  
256 specific segments of streams, however, so we needed to do some spatial averaging of capacity  
257 predictions to generate larger scale capacity estimates.

258 To summarize capacity at larger scales, the mean linear capacity (e.g., parr/m) of the master sample  
259 points along a particular tributary is multiplied by the length of that tributary. We first restricted  
260 the upstream limit of master sample points and lengths of streams to those within the domain of  
261 spring/summer-run Chinook salmon, as defined by StreamNet (<http://www.streamnet.org>) or  
262 using expert opinion from local biologists, and the downstream limit by when streams were no  
263 longer wadeable (often determined by some combination of estimated bankfull width and cumula-  
264 tive drainage area). The capacities of various tributaries could then be summed to estimate capacity  
265 at almost any spatial scale. A conceptual diagram showing the data and modeling framework of  
266 the QRF and extrapolation models is shown in Figure 2.

## 267 **Model Validation**

268 Spawner-recruit data from several watersheds within the interior Columbia River basin were com-  
269 piled to validate the extrapolated QRF estimates of Chinook salmon parr capacity. Some water-  
270 sheds had direct estimates of parr, while some had estimates of pre-smolts and smolts (i.e., fall and  
271 spring emigrants) from rotary screw traps. For the latter, estimates of parr were calculated using  
272 estimates of over-winter survival to back-calculate parr from smolt estimates, and then adding that  
273 to pre-smolt estimates. A series of spawner-recruit functions were then fit to this data including  
274 Beverton-Holt, Ricker, and hockey stick (Froese 2008), using the *FSA* package (Ogle et al. 2020)  
275 in R. Estimates of capacity from each of these spawner-recruit curves were compared with QRF  
276 estimates of capacity for the same regions.

277 All code and data for the analyses presented here can be found in a GitHub repository  
278 (<https://github.com/KevinSee/QRFpaper>).

## 279 **Results**

### 280 **Habitat Covariate Selection**

281 We categorized 165 habitat measurements collected using the CHaMP habitat protocol (CHaMP  
282 2016) into eleven habitat categories, and for each habitat covariate the Maximal Information Crite-  
283 ria (MIC) value was calculated based on the strength of association between the habitat covariate  
284 and the response variable, parr density (parr/m) (See Appendix S1 for further details). We chose  
285 the following twelve CHaMP habitat covariates to fit the QRF model: wetted width, observed dis-  
286 charge, average August temperature, wetted width:depth ratio, percent fines less than 6 mm, total  
287 percent fish cover, channel unit frequency, standard deviation of the wetted depth, frequency of  
288 large wood in pools, percent riparian canopy cover, lower quantile of substrate size (D16) and  
289 braidedness (Table 2).

### 290 **QRF Model**

291 A QRF model was fit using those metrics and the *quantregForest* package (Meinshausen 2006) in  
292 R (R Core Team 2019) and the 90th quantile of the predicted distribution was used as a proxy for  
293 carrying capacity. After model fit, we examined the relative importance of each habitat covariate  
294 included in the model (Figure 3), quantified by the average decrease in residual sum of squares  
295 for splits on that variable amidst the trees in the random forest, implemented by the *importance*  
296 function from the *randomForest* package (Liaw and Wiener 2002). Moreover, QRF models allow  
297 one to visually examine the marginal effect of each habitat covariate on the quantile of interest  
298 using partial dependence plots. These plots show the marginal effect of changing a single habitat  
299 covariate while maintaining all other covariates at their mean values (Figure 4). However, given  
300 that many habitat metrics are somewhat correlated, these marginal effects are often not independent  
301 of one another and care should be taken when interpreting them. After model fitting, the QRF model  
302 was used to predict habitat capacity at all CHaMP sites within the interior Columbia River basin.

## 303 **Extrapolating to Other Sites**

304 We fit a linear regression extrapolation using QRF-based predictions of capacity at all CHaMP  
305 sites as the response, and various GAAs as the independent variables. The coefficients for the  
306 extrapolation model can be found in Table 3 and the summary of the model fit in Table 4. From  
307 this, we calculated estimates of capacity at every master sample point in the Columbia River basin,  
308 each representing roughly one kilometer of stream length.

## 309 **Model Validation**

310 Estimates of Chinook salmon parr capacity from the QRF and extrapolation models were compa-  
311 rable to independent estimates from spawner-recruit data (Table 5, Figure 5). QRF estimates had  
312 overlapping confidence intervals with one or more of the Beverton-Holt, Ricker, or hockey stick  
313 model estimates in each of the nine locations where comparisons were possible (Figure 5). Poten-  
314 tial additional uncertainty was not accounted for in estimates of spawners-per-redd or spawners-  
315 per-parr, which would increase the confidence intervals around spawner-recruit estimates and the  
316 overlap among estimates. Correlations between parr capacity estimates from the QRF model and  
317 spawner-recruit models ranged from 0.710 (Beverton-Holt) to 0.966 (Ricker). This favorable com-  
318 parison provides strong validation as the spawner-recruit estimates of Chinook salmon parr capacity  
319 were developed from completely independent datasets and using entirely different methods.

## 320 **Discussion**

### 321 **A Tool to Estimate Habitat Capacity**

322 In this study, we developed a novel approach to estimate the capacity of habitat to support Chinook  
323 salmon parr during summer months and in wadeable streams. Our model can be used to quantify  
324 juvenile rearing capacity in Chinook salmon watersheds or populations and, in turn, quantify the  
325 magnitude of tributary habitat rehabilitation that may be necessary to support Endangered Species

326 Act delisting. The QRF and extrapolation models presented here provide useful tools towards the  
327 prioritization, implementation, and evaluation of habitat rehabilitation actions to recover depleted  
328 salmon populations. Moreover, these models can be applied to multiple stages within the life cycle  
329 (e.g., parr, smolt, adult). Estimates of habitat carrying capacity for multiple life stages will allow  
330 biologists and managers to identify what life-stages and/or specific habitat patches may be limiting.  
331 As an example, QRF models and associated extrapolation models may demonstrate that habitat  
332 for a given population is sufficient to support adult spawning required to achieve delisting targets,  
333 but that juvenile rearing capacity may not be sufficient to support the target abundance. In such a  
334 case, habitat rehabilitation actions may be most cost-effectively and sustainably directed towards  
335 improving juvenile rearing habitat. Models to estimate habitat carrying capacity for multiple life  
336 stages will help to better direct habitat restoration actions and help guide not only the type of action,  
337 but also the location at which an action is performed.

338 The favorable comparison between QRF estimates of carrying capacity and the spawner-recruit  
339 based estimates in select watersheds helps support and validate this approach. Although built from  
340 completely different data, when these multiple lines of evidence converge it lends credence to the  
341 QRF capacity prediction results.

342 There are two aspects that make this approach “data hungry”, meaning a large dataset is needed to  
343 fit a QRF model like this. First, random forest models generally require more data than parametric  
344 models, due to the lack of parametric distribution assumptions and the lack of an assumed form of  
345 the relationship between dependent and independent variables. Second, it takes larger data sets to  
346 accurately predict the lower and higher quantiles in a quantile regression framework. For example,  
347 if a data set consisted of thousands, rather than hundreds, of data points, a researcher might feel  
348 comfortable using the 95th or the 98th quantile as a proxy for capacity, rather than the 90th. Our  
349 data set consisted of 327 sites, across a variety of habitats and years, providing contrast in all the  
350 habitat covariates and presumably satisfying the data hungriness of a QRF model, based on our  
351 validation with spawner-recruit data.

## 352 **Biological Expectations from QRF Model**

353 The results of the QRF parr capacity model for Chinook salmon meet many biological expectations.  
354 Focusing on the partial dependence plots (Figure 4), the QRF model predicts capacity to increase  
355 when the wetted width, discharge and the width:depth ratio grow, when temperatures are cooler  
356 (Brett 1952, Raleigh et al. 1986, Bjornn and Reiser 1991), when there is less fine sediment (Hillman  
357 et al. 1987, Bjornn and Reiser 1991, Allen 2000), when there is more fish cover (Hillman et al. 1987,  
358 Bjornn and Reiser 1991, Holecek et al. 2009), when channel unit frequency increases and when  
359 the standard deviation of the wetted depth (a proxy for streambed complexity) increases. These are  
360 all patterns that emerged from the fish-habitat data, and where available, match those fish-habitat  
361 relationships identified qualitatively in other studies (Mossop and Bradford 2006).

362 The biggest driver of capacity identified in this study is stream size, whether measured by wetted  
363 width or discharge, which should be unsurprising since we are using fish per meter as our response.  
364 In many ways, these metrics define habitat quantity; however, other metrics used in our QRF model  
365 help define habitat quality, such as cooler temperatures in August, less pool-tail fine sediment, and  
366 higher channel unit frequencies (a measure of habitat complexity and surrogate for the number of  
367 pool-riffle sequences or potential sheer areas providing feeding zones), and fish cover. Metrics  
368 that describe habitat quantity set some bounds around possible capacity estimates, while metrics  
369 describing habitat quality refine those estimates to better match conditions at that site.

## 370 **Extrapolation Model**

371 Fish are mobile creatures and determining the appropriate spatial scale to estimate how their capac-  
372 ity may be determined by habitat characteristics is important. In the summer, for Chinook salmon  
373 parr, our fish data clearly shows movement between multiple channel units (e.g., pool, riffle, run),  
374 suggesting that fish are utilizing habitat at a larger scale than the channel unit. However, it is un-  
375 likely that they are moving up and down the entire watershed and we believe the 200 - 600 m  
376 reaches used in this study are an appropriate scale to capture the fish-habitat relationships that de-

377 fine carrying capacity. At the same time, we acknowledge that managers, life-cycle modelers, and  
378 others are often interested in capacity estimates at larger spatial scales. While our QRF model can  
379 provide site-specific estimates of carrying capacity derived from paired fish-habitat data, our ex-  
380 trapolation model allows for estimates at larger spatial extents, such as watershed and population  
381 levels. This is an efficient technique to leverage existing relationships for meaningful management  
382 decisions.

383 Our extrapolation model was focused on extrapolating to other master sample points, because that  
384 is the dataset available to us, but the methodology could be improved. Extrapolating to reaches  
385 on a stream network, as opposed to points on the landscape, could improve the interpretability of  
386 the results. This would require a stream network with relevant attributes attached to each reach,  
387 similar to the GAAs we used. Another approach could be to move towards sampling habitat in a  
388 more spatially continuous fashion, covering most or all of a watershed, and building a QRF model  
389 from that dataset. Even if the fish data were not collected continuously, estimates of capacity could  
390 be made directly from the QRF model across the entire stream network without the need for an  
391 extrapolation model.

392 One of the potential downsides to the extrapolation approach used here is that the GAAs generally  
393 do not change through time, and therefore may not reflect the dynamic nature of changing stream  
394 habitat. While the QRF model itself uses habitat characteristics that can be observed to change  
395 over the course of several years, most GAAs are static, generally derived remotely or from another  
396 model. This is the nature of extrapolating to such large spatial extents; it can be impossible to gather  
397 actual habitat data on such a scale, but with improvements in remote sensing, spatially continuous  
398 data (modeled or measured) may be on the horizon (Tonina et al. 2019).

### 399 **The Future: Improving Habitat Data**

400 Given the cost/extent of data necessary for QRF extrapolation in watersheds outside of the  
401 Columbia River basin, there is a pressing need to develop new tools for habitat analyses. Un-

402 manned Aerial Systems (UAS or drones, commonly) are gaining popularity in wildlife and  
403 ecosystem monitoring for their ease of use, safety, accessibility, and cost-efficiency (Jones  
404 IV et al. 2006, Chabot and Bird 2015). UAS produce high-resolution, permanent data at a  
405 fraction of the cost of on-the-ground habitat sampling. Advances in imaging techniques (e.g.,  
406 multispectral imaging) and post-processing (e.g., automation of data collection from imagery) are  
407 already demonstrating increase in the efficiency and accuracy of data collection (Whitehead and  
408 Hugenholtz 2014, LeCun et al. 2015, Weinstein 2018). Further, developments in Light Detection  
409 and Ranging (LiDAR) technology have allowed for the characterization of watershed scale  
410 geomorphologic and hydraulic variables not previously possible (McKean et al. 2008, Tonina et  
411 al. 2019).

412 Development of a standardized protocol to incorporate remotely sensed data (LiDAR, aerial im-  
413 agery) into the collection of habitat metrics would greatly increase the broadscale application of  
414 QRF. Rapid advances in drone technology further improve upon traditional habitat data collection  
415 by leveraging 1) sub-meter global navigation satellite system (GNSS) receivers; 2) cost-effective  
416 drone imagery collection, image stitching, and photogrammetry; and 3) semi-automated to auto-  
417 mated data post-processing. All data collection efforts would be georeferenced and topologically  
418 compatible to increase repeatability of methods and data collection locations; a primary criticism of  
419 previous CHaMP survey efforts. The implementation of such a protocol would circumvent the need  
420 to extrapolate by collecting data for individual channel units in a rapid manner using remote sens-  
421 ing technologies, thereby reducing labor, providing a cost-effective tool for habitat data collection  
422 supporting status and trend evaluation and model products to better inform habitat rehabilitation  
423 prioritization and planning.

## 424 **Conclusions and Next Steps**

425 If a species' carrying capacity is defined or constrained, at least in part, by the habitat in which  
426 it lives, then illuminating statistically how such habitat impacts carrying capacity can lead to un-  
427 derstanding how a species interacts with its environment. This understanding could be of crucial

428 importance in the realm of conservation when dealing with an endangered or threatened species,  
429 but species/habitat interactions are a core element of ecological studies more generally. We have  
430 demonstrated how a quantile regression approach, coupled with a random forest framework, can be  
431 used to estimate these relationships, and predict a habitat's capacity. As large ecological datasets  
432 become more accessible, and the ability to measure large swaths of habitat more feasible, we be-  
433 lieve this approach has many potential applications, from the North American breeding bird survey  
434 to groundfish trawl surveys. The framework could also be applied to depleted, non-migratory, and  
435 isolated populations (e.g., desert pupfish *Cyprinodon macularius*) to identify limiting factors in pop-  
436 ulations and/or determine whether a given habitat patch could support a viable population if limiting  
437 factors were addressed. Capacity estimates could also be used to evaluate potential translocation  
438 sites to determine if those sites could support an abundance considered viable before investing in  
439 translocation efforts.

## 440 **Acknowledgements**

441 Model development efforts have been funded by the Bonneville Power Administration through  
442 projects 2003-017-00 and 2011-006-00 and by the Bureau of Reclamation and Idaho Governor's  
443 Office of Species Conservation through contract BOR002 16. Fish sampling work in the Lemhi  
444 River was also funded through the Idaho Governor's Office of Species Conservation through the  
445 Pacific Coast Salmon Recovery Fund. Special thanks to staff from the following agencies for  
446 providing data to calculate Chinook salmon parr abundance and density estimates: Columbia River  
447 Inter-Tribal Fish Commission, Oregon Department of Fish and Wildlife, and U.S. Fish and Wildlife  
448 Service. The models in this study were improved by conversations with Eric Buhle.

## 449 Literature Cited

- 450 Allen, M. A. 2000. Seasonal microhabitat use by juvenile spring chinook salmon in the yakima  
451 river basin, washington. *Rivers* 7:314–332.
- 452 Bjornn, T., and D. Reiser. 1991. Habitat requirements of salmonids in streams. *American Fisheries*  
453 *Society Special Publication* 19:138.
- 454 Bond, M. H., T. G. Nodine, T. J. Beechie, and R. W. Zabel. 2018. Estimating the benefits of  
455 widespread floodplain reconnection for columbia river chinook salmon. *Canadian Journal of Fish-*  
456 *eries and Aquatic Sciences*:1–15.
- 457 Braun, D. C., and J. D. Reynolds. 2011. Relationships between habitat characteristics and breeding  
458 population densities in sockeye salmon (*oncorhynchus nerka*). *Canadian Journal of Fisheries and*  
459 *Aquatic Sciences* 68:758–767.
- 460 Breiman, L. 2001. Random forests. *Machine Learning* 45:5–32.
- 461 Brett, J. R. 1952. Temperature tolerance in young pacific salmon, genus *oncorhynchus*. *Journal of*  
462 *the Fisheries Board of Canada* 9:265–323.
- 463 Bryant, M., and R. Woodsmith. 2009. The response of salmon populations to geomorphic mea-  
464 surements at three scales. *North American Journal of Fisheries Management* 29:549–559.
- 465 Cade, B. S., and B. R. Noon. 2003. A gentle introduction to quantile regression for ecologists.  
466 *Frontiers in Ecology and the Environment* 1:412–420.
- 467 Carle, F. L., and M. R. Strub. 1978. A new method for estimating population size from removal  
468 data. *Biometrics*:621–630.
- 469 Carmichael, R. A., D. Tonina, E. R. Keeley, R. M. Benjankar, and K. E. See. 2020. Some like it  
470 slow: A bioenergetic evaluation of habitat quality for juvenile chinook salmon in the lemhi river,  
471 idaho. *Canadian Journal of Fisheries and Aquatic Sciences*.
- 472 Chabot, D., and D. M. Bird. 2015. Wildlife research and management methods in the 21st century:

473 Where do unmanned aircraft fit in? *Journal of Unmanned Vehicle Systems* 3:137–155.

474 CHaMP. 2016. Scientific protocol for salmonid habitat surveys within the columbia habitat moni-  
475 toring program. Bonneville Power Administration Portland, Oregon, USA.

476 Chapman, D. G. 1951. Some properties of the hypergeometric distribution with applications to  
477 zoological sample censuses. University of California Press.

478 Compton, T. J., M. A. Morrison, J. R. Leathwick, and G. D. Carbines. 2012. Ontogenetic habitat  
479 associations of a demersal fish species, *pagrus auratus*, identified using boosted regression trees.  
480 *Marine Ecology Progress Series* 462:219–230.

481 Cramer, S. P., and N. K. Ackerman. 2009. Linking stream carrying capacity for salmonids to  
482 habitat features. Pages 225–254 *in* American fisheries society symposium.

483 Dodrill, M. J., and C. B. Yackulic. 2016. Nonlinear relationships can lead to bias in biomass calcu-  
484 lations and drift-foraging models when using summaries of invertebrate drift data. *Environmental*  
485 *biology of fishes* 99:659–670.

486 Dunham, J. B., B. S. Cade, and J. W. Terrell. 2002. Influences of spatial and temporal variation on  
487 fish-habitat relationships defined by regression quantiles. *Transactions of the American Fisheries*  
488 *Society* 131:86–98.

489 Eastwood, P. D., G. J. Meaden, A. Carpentier, and S. I. Rogers. 2003. Estimating limits to the  
490 spatial extent and suitability of sole (*solea solea*) nursery grounds in the dover strait. *Journal of Sea*  
491 *Research* 50:151–165.

492 Fausch, K. D., C. L. Hawkes, and M. G. Parsons. 1988. Models that predict standing crop of stream  
493 fish from habitat variables: 1950-85. Notes.

494 Froese, R. 2008. The continuous smooth hockey stick: A newly proposed spawner-recruitment  
495 model. *Journal of Applied Ichthyology* 24:703–704.

496 Hedger, R. D., E. De Eyto, M. Dillane, O. H. Diserud, K. Hindar, P. McGinnity, R. Poole, and G.

497 Rogan. 2013. Improving abundance estimates from electrofishing removal sampling. *Fisheries*  
498 *Research* 137:104–115.

499 Hillman, T. W., J. Griffith, and W. Platts. 1987. Summer and winter habitat selection by juve-  
500 nile chinook salmon in a highly sedimented idaho stream. *Transactions of the American Fisheries*  
501 *Society* 116:185–195.

502 Hinrichsen, R. A., and C. M. Paulsen. 2020. Low carrying capacity a risk for threatened Chinook  
503 salmon. *Ecological Modelling* 432:109223.

504 Holecek, D. E., K. J. Cromwell, and B. P. Kennedy. 2009. Juvenile chinook salmon summer  
505 microhabitat availability, use, and selection in a central idaho wilderness stream. *Transactions of*  
506 *the American Fisheries Society* 138:633–644.

507 Jones IV, G. P., L. G. Pearlstine, and H. F. Percival. 2006. An assessment of small unmanned aerial  
508 vehicles for wildlife research. *Wildlife society bulletin* 34:750–758.

509 Knudby, A., A. Brenning, and E. LeDrew. 2010. New approaches to modelling fish-habitat rela-  
510 tionships. *Ecological Modelling* 221:503–511.

511 Koenker, R., and G. Bassett Jr. 1978. Regression quantiles. *Econometrica: journal of the Econo-*  
512 *metric Society*:33–50.

513 Larsen, D. P., A. R. Olsen, and D. L. Stevens. 2008. Using a master sample to integrate stream  
514 monitoring programs. *Journal of Agricultural, Biological, and Environmental Statistics* 13:243.

515 Larsen, D. P., C. J. Volk, D. L. Stevens Jr, A. R. Olsen, and C. E. Jordan. 2016. An overview of  
516 the Columbia Habitat Monitoring Program’s (CHaMP) spatial-temporal design framework. *South*  
517 *Fork Research*.

518 LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. *Nature* 521:436–444.

519 Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. *R News* 2:18–22.

520 Lumley, T. 2020. *Survey: Analysis of complex survey samples*.

521 McKean, J. A., D. J. Isaak, and C. W. Wright. 2008. Geomorphic controls on salmon nesting  
522 patterns described by a new, narrow-beam terrestrial–aquatic lidar. *Frontiers in Ecology and the*  
523 *Environment* 6:125–130.

524 Meinshausen, N. 2017. QuantregForest: Quantile regression forests.

525 Meinshausen, N. 2006. Quantile regression forests. *Journal of Machine Learning Research* 7:983–  
526 999.

527 Mossop, B., and M. J. Bradford. 2006. Using thalweg profiling to assess and monitor juvenile  
528 salmon (*oncorhynchus* spp.) habitat in small streams. *Canadian Journal of Fisheries and Aquatic*  
529 *Sciences* 63:1515–1525.

530 Moussalli, E., and R. Hilborn. 1986. Optimal stock size and harvest rate in multistage life history  
531 models. *Canadian Journal of Fisheries and Aquatic Sciences* 43:135–141.

532 Myers, R. A., K. G. Bowen, and N. J. Barrowman. 1999. Maximum reproductive rate of fish at  
533 low population sizes. *Canadian Journal of Fisheries and Aquatic Sciences* 56:2404–2419.

534 Nahorniak, M., D. P. Larsen, C. Volk, and C. E. Jordan. 2015. Using inverse probability bootstrap  
535 sampling to eliminate sample induced bias in model based analysis of unequal probability samples.  
536 *Plos one* 10:e0131765.

537 Nickelson, T. E., J. D. Rodgers, S. L. Johnson, and M. F. Solazzi. 1992. Seasonal changes in habitat  
538 use by juvenile coho salmon (*oncorhynchus kisutch*) in oregon coastal streams. *Canadian Journal*  
539 *of Fisheries and Aquatic Sciences* 49:783–789.

540 NOAA Fisheries. 2016a. 2016 5-year review: Summary & evaluation of Upper Columbia River  
541 steelhead and Upper Columbia River spring-run Chinook salmon. NOAA.

542 NOAA Fisheries. 2016b. 2016 5-year review: Summary & evaluation of Snake River Sockeye,  
543 Snake River spring-summer Chinook, Snake River fall-run Chinook and Snake River steelhead.  
544 NOAA.

545 NOAA Fisheries. 2020. Endangered Species Act Section 7(a)(2) Biological opinion and  
546 Magnuson-Stevens fishery conservation and management act essential fish habitat response for  
547 the continued operation and maintenance of the Columbia River system. NMFS Consultation,  
548 NOAA National Marine Fisheries Service.

549 Ogle, D., P. Wheeler, and A. Dinno. 2020. FSA: Simple fisheries stock assessment methods.

550 Pess, G. R., D. R. Montgomery, E. A. Steel, R. E. Bilby, B. E. Feist, and H. M. Greenberg.  
551 2002. Landscape characteristics, land use, and coho salmon (*Oncorhynchus kisutch*) abundance,  
552 snohomish river, wash., usa. *Canadian Journal of Fisheries and Aquatic Sciences* 59:613–623.

553 Pittman, S. J., B. M. Costa, and T. A. Battista. 2009. Using lidar bathymetry and boosted regression  
554 trees to predict the diversity and abundance of fish and corals. *Journal of Coastal Research* 53:27–  
555 38.

556 Prasad, A., L. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques:  
557 Bagging and random forests for ecological prediction. *Ecosystems* 9:181–199.

558 Raleigh, R. F., W. J. Miller, and P. C. Nelson. 1986. Habitat suitability index models and instream  
559 flow suitability curves: Chinook salmon. National Ecology Center.

560 R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for  
561 Statistical Computing, Vienna, Austria.

562 Rivest, L.-P., and S. Baillargeon. 2019. Rcapture: Loglinear models for capture-recapture experi-  
563 ments.

564 Robson, D., and H. Regier. 1964. Sample size in petersen mark–recapture experiments. *Transac-  
565 tions of the American Fisheries Society* 93:215–226.

566 Seber, G. A. F. 2002. The estimation of animal abundance and related parameters. Blackburn Press  
567 Caldwell, New Jersey.

568 Sharma, R., and R. Hilborn. 2001. Empirical relationships between watershed characteristics and

569 coho salmon (*Oncorhynchus kisutch*) smolt abundance in 14 western Washington streams. *Canadian*  
570 *Journal of Fisheries and Aquatic Sciences* 58:1453–1463.

571 Stekhoven, D. J. 2013. MissForest: Nonparametric missing value imputation using random forest.

572 Stekhoven, D. J., and P. Bühlmann. 2012. MissForest—non-parametric missing value imputation  
573 for mixed-type data. *Bioinformatics* 28:112–118.

574 Stevens Jr, D. L., and A. R. Olsen. 1999. Spatially restricted surveys over time for aquatic resources.  
575 *Journal of Agricultural, Biological, and Environmental Statistics*:415–428.

576 Stevens Jr, D. L., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. *Journal*  
577 *of the American Statistical Association* 99:262–278.

578 Sweka, J. A., and G. Mackey. 2010. A functional relationship between watershed size and Atlantic  
579 salmon parr density. *Journal of Fish and Wildlife Management* 1:3–10.

580 Terrell, J. W., B. S. Cade, J. Carpenter, and J. M. Thompson. 1996. Modeling stream fish habitat  
581 limitations from wedge-shaped patterns of variation in standing stock. *Transactions of the Ameri-*  
582 *can Fisheries Society* 125:104–117.

583 Tonina, D., J. A. McKean, R. M. Benjankar, C. W. Wright, J. R. Goode, Q. Chen, W. J. Reeder, R.  
584 A. Carmichael, and M. R. Edmondson. 2019. Mapping river bathymetries: Evaluating topobathy-  
585 metric lidar survey. *Earth Surface Processes and Landforms* 44:507–520.

586 Volk, C., N. Bouwes, C. Jordan, J. Wheaton, P. Nelle, C. Beasley, S. Walker, B. Bouwes, M. Na-  
587 horniak, A. Hill, J. Heitke, K. Whitehead, S. Bangen, and E. Ward. 2017. Integrated status and  
588 effectiveness monitoring program (ISEMP) and Columbia habitat monitoring program (CHaMP)  
589 2016 annual combined technical report. Bonneville Power Administration.

590 Wall, C. E., N. Bouwes, J. M. Wheaton, W. C. Saunders, and S. N. Bennett. 2016. Net rate of energy  
591 intake predicts reach-level steelhead (*Oncorhynchus mykiss*) densities in diverse basins from a large  
592 monitoring program. *Canadian Journal of Fisheries and Aquatic Sciences* 73:1081–1091.

593 Weinstein, B. G. 2018. A computer vision for animal ecology. *Journal of Animal Ecology* 87:533–  
594 545.

595 Whitehead, K., and C. H. Hugenholtz. 2014. Remote sensing of the environment with small un-  
596 manned aircraft systems (uass), part 1: A review of progress and challenges. *Journal of Unmanned*  
597 *Vehicle Systems* 2:69–85.

598 Whittier, T. R., A. T. Herlihy, C. E. Jordan, and C. Volk. 2011. Landscape classification of pacific  
599 northwest hydrologic units based on natural features and human disturbance to support salmonid  
600 research. NOAA, National Marine Fisheries Service.

601 **Tables**

Table 1: The number of unique sites, by watershed, with paired fish-habitat data used to populate the spring/summer-run Chinook salmon parr capacity QRF model

Watershed	n Sites	Percent
Entiat	61	18.7%
John Day	75	22.9%
Lemhi	33	10.1%
Minam	20	6.1%
South Fork Salmon	30	9.2%
Upper Grande Ronde	86	26.3%
Wenatchee	22	6.7%
Total	327	100.0%

Table 2: Habitat metrics and descriptions of metrics included in the QRF model to predict spring/summer-run Chinook salmon parr capacity. Metrics are ranked in order of relative importance.

Rank	Metric	Metric Category	Description
1	Wetted Width	Size	Average width of the wetted polygon for a site.
2	Discharge	Size	The sum of station discharge across all stations. Station discharge is calculated as depth x velocity x station increment for all stations except first and last. Station discharge for first and last station is 0.5 x station width x depth x velocity.
3	Avg. August Temp.	Temperature	Average predicted daily August temperature from NorWest, averaged across the years 2002-2011.
4	Width:Depth Ratio	Complexity	Average width to depth ratio of the wetted channel measured from cross-sections. Depths represent an average of depths along each cross-section.
5	Fines	Substrate	Average percentage of pool tail substrates comprised of sediment <6 mm.
6	Fish Cover	Cover	Percent of wetted area with the following types of cover: aquatic vegetation, artificial, woody debris, and terrestrial vegetation.
7	Channel Unit Frequency	ChannelUnit	Number of channel units per 100 meters.
8	Depth Complexity	Complexity	Standard Deviation of water depths within the wetted channel.
9	Large Wood Freq. in Pools	Wood	Total volume of large wood pieces within the wetted channel and Slow Water/Pool channel units, scaled by site length.
10	Riparian Canopy	Riparian	Percent of riparian canopy with some vegetation.
11	Substrate: D16	Substrate	Diameter of the 16th percentile particle derived from pebble counts.
12	Braidedness	Complexity	Ratio of the total length of the wetted mainstem channel plus side channels and the length of the mainstem channel.

Table 3: Globally available attribute (GAA) habitat covariates used to extrapolate quantile random forest (QRF) model predictions of spring/summer-run Chinook parr capacity to a larger scale (e.g., watershed, population), with their coefficients and standard errors.

Covariate	Units	Spatial Scale	Estimate	Std. Error
Temperature Range	C	Reach-2km	-0.044	0.081
Elevation	m	Site-300m	-0.243	0.155
CHaMP Watershed	-	Region	-	-
Natural Class PCA 1	-	Watershed-HUC12	-0.092	0.070
Disturbance Class PCA 1	-	Watershed-HUC12	-0.064	0.064
Drainage Area (sqrt)	km <sup>2</sup> (square root)	Reach-2km	-0.141	0.077
Stream Power	-	Reach-2km	0.049	0.033
Slope	m/m	Reach-2km	-0.513	0.100
Channel Type	-	Site-300m	-	-
Bankfull Width - modeled	m	Site-300m	0.216	0.099
NorWeST Aug. Temperature	C	Reach-2km	-0.149	0.119

Table 4: Summary of extrapolation model fits, split by whether the extrapolation model used CHaMP watershed as a covariate or not.

Model	Response	$r^2$	Adjusted $r^2$
CHaMP	fish/m	0.481	0.454
non-CHaMP	fish/m	0.360	0.339

Table 5: Estimates of parr capacity from both spawner-recruit data (Beverton-Holt, Ricker, hockey stick) and from extrapolated estimates of parr capacity from the quantile random forest (QRF) model. Numbers in parentheses are coefficients of variation.

Population	n Yrs	Adult Data	Parr Data	Beverton Holt	Ricker	Hockey Stick	QRF
Catherine Creek	20	Spawners	RST	135,387 (0.269)	103,021 (0.141)	99,921 (0.21)	190,857 (0.162)
Chiwawa River	20	Spawners	Parr Surveys	248,586 (0.24)	166,139 (0.148)	174,216 (0.184)	216,451 (0.363)
Hayden Creek	7	Spawners	RST	58,394 (0.244)	65,958 (0.195)	48,351 (0.174)	121,676 (0.202)
Lostine River	17	Redds	RST	196,259 (0.24)	146,982 (0.159)	144,415 (0.201)	152,493 (0.316)
Minam River	14	Spawners	RST	1,309,223 (2.18)	484,810 (1.444)	662,802 (1.726)	365,338 (0.261)
South Fork Salmon River	17	Redds	RST	87,260 (0.407)	62,456 (0.265)	64,654 (0.317)	221,362 (0.142)
Tucannon River	27	Redds	RST	4,791,131 (13.016)	11,234,653 (8.566)	31,922,692 (10.082)	2529,223 (0.196)
Upper Grande Ronde River	8	Spawners	RST	171,607 (0.388)	168,137 (0.298)	127,052 (0.317)	200,228 (0.23)
Upper Lemhi	7	Spawners	RST	333,229 (0.322)	229,635 (0.212)	242,637 (0.252)	269,626 (0.217)

602 **List of Figures**

603 1 Watersheds with CHaMP habitat data. Watersheds in black also contain paired fish  
604 data. Watershed names are: 1 - Entiat, 2 - John Day, 3 - Lemhi, 4 - Methow, 5 -  
605 Minam, 6 - Secesh, 7 - South Fork Salmon, 8 - Tucannon, 9 - Upper Grande Ronde,  
606 10 - Wenatchee, 11 - Yankee Fork. . . . . 32

607 2 Conceptual diagram showing input data sources, modeling decisions, model out-  
608 puts etc. for the QRF and extrapolation models. . . . . 33

609 3 Relative importance of each habitat covariate included in the quantile random  
610 forest (QRF) model to predict habitat capacity, during summer months, for  
611 spring/summer-run Chinook salmon parr . . . . . 34

612 4 Partial dependence plots for the spring/summer-run Chinook salmon parr capacity  
613 quantile random forest (QRF) model, depicting how parr capacity shifts as each  
614 habitat metric changes, assuming all other habitat metrics remain at their mean  
615 values. Tick marks along the X-axis depict observed values. . . . . 35

616 5 Spawner-recruit data from nine watersheds. Solid lines are the spawner-recruit  
617 curve, dashed lines are the estimated capacity, and shaded polygons depict the 95%  
618 confidence intervals of capacity. Red corresponds to Beverton-Holt models, purple  
619 to Ricker models, blue to hockey stick models, and green to QRF estimates. The  
620 QRF solid curve is a Beverton-Holt model with the capacity parameter fixed to  
621 the QRF estimate of capacity. A few curves with high capacity estimates were not  
622 plotted to improve readability. . . . . 36

623 **Figures**

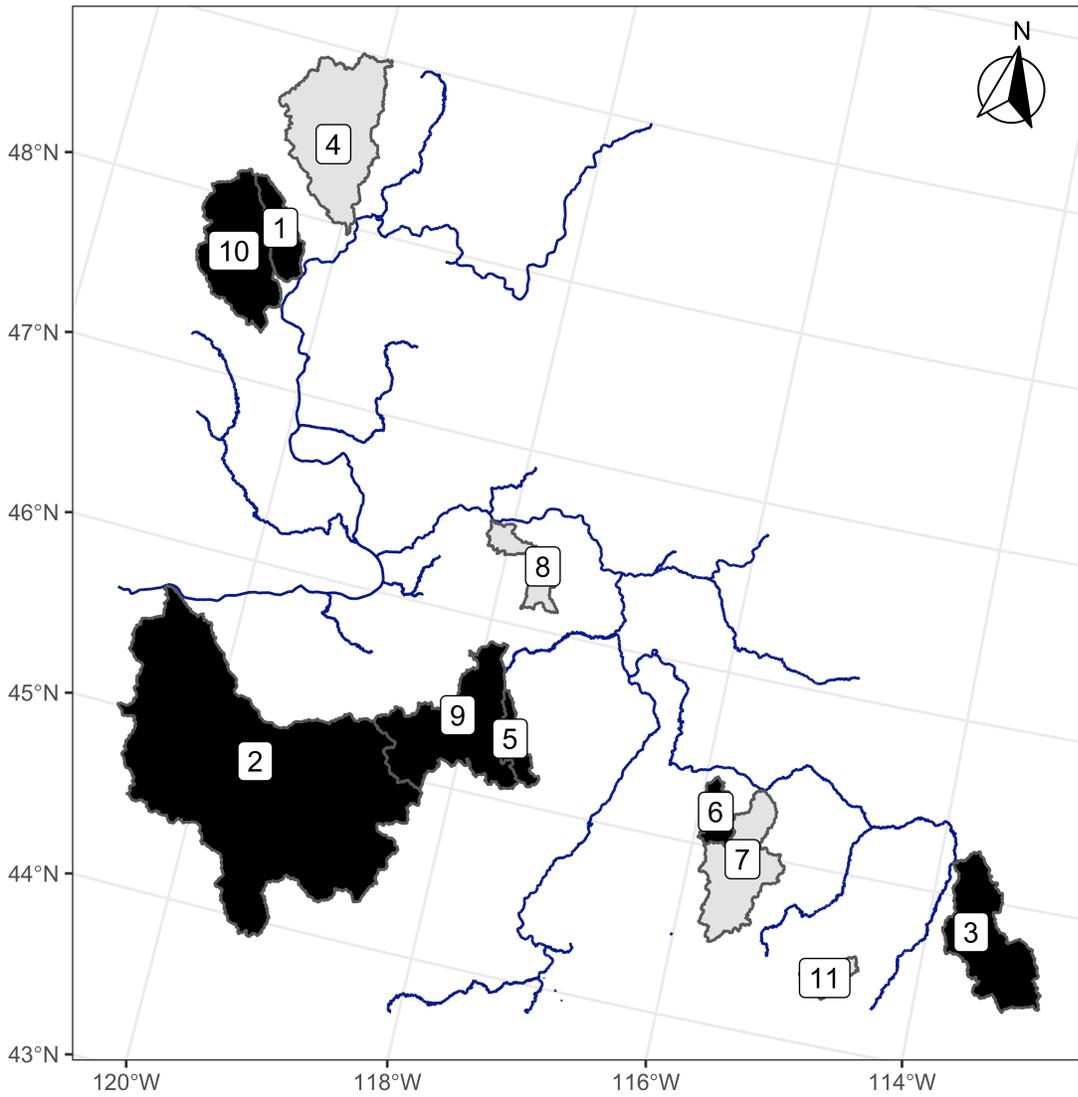


Figure 1: Watersheds with CHaMP habitat data. Watersheds in black also contain paired fish data. Watershed names are: 1 - Entiat, 2 - John Day, 3 - Lemhi, 4 - Methow, 5 - Minam, 6 - Secesh, 7 - South Fork Salmon, 8 - Tucannon, 9 - Upper Grande Ronde, 10 - Wenatchee, 11 - Yankee Fork.

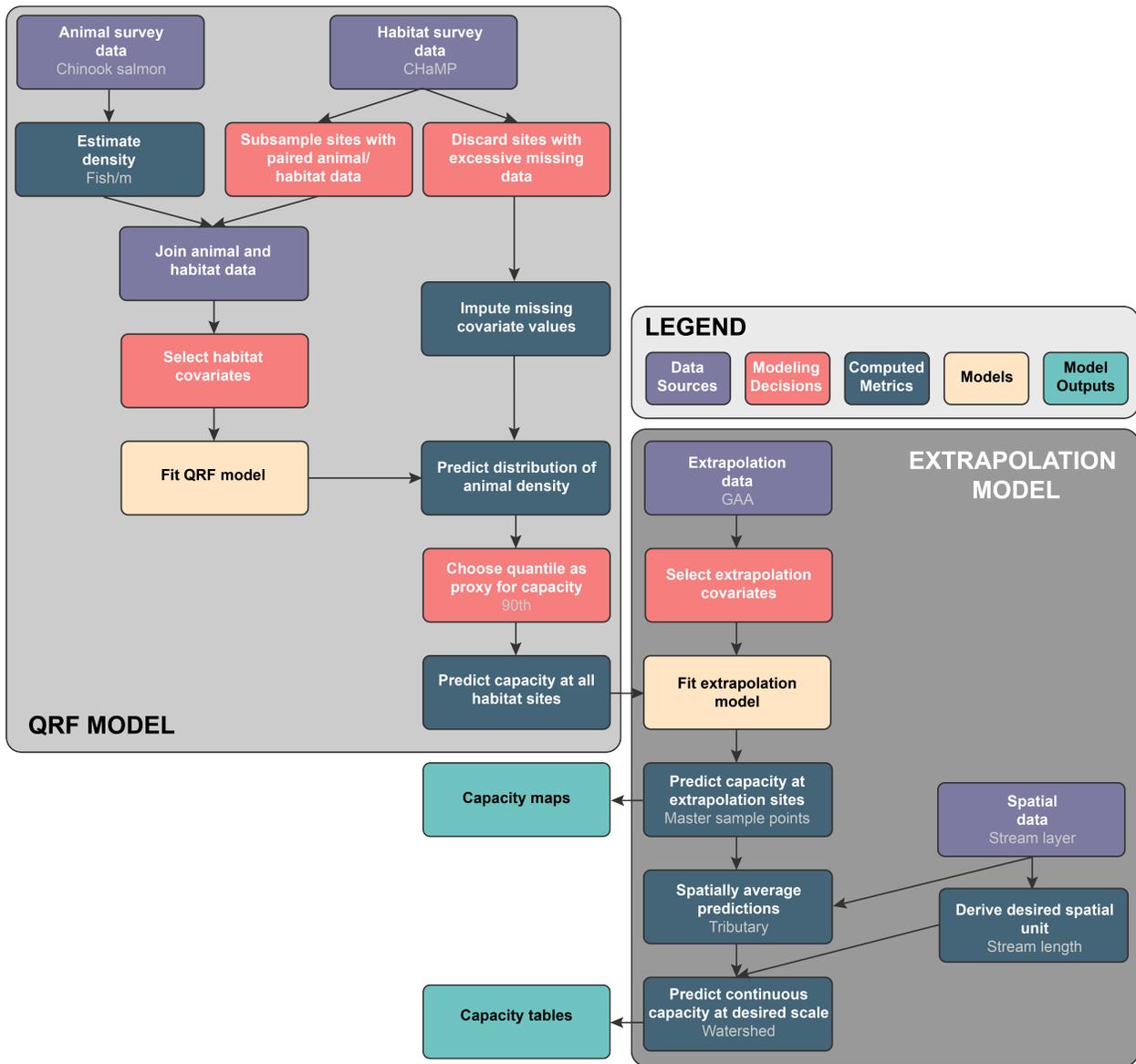


Figure 2: Conceptual diagram showing input data sources, modeling decisions, model outputs etc. for the QRF and extrapolation models.

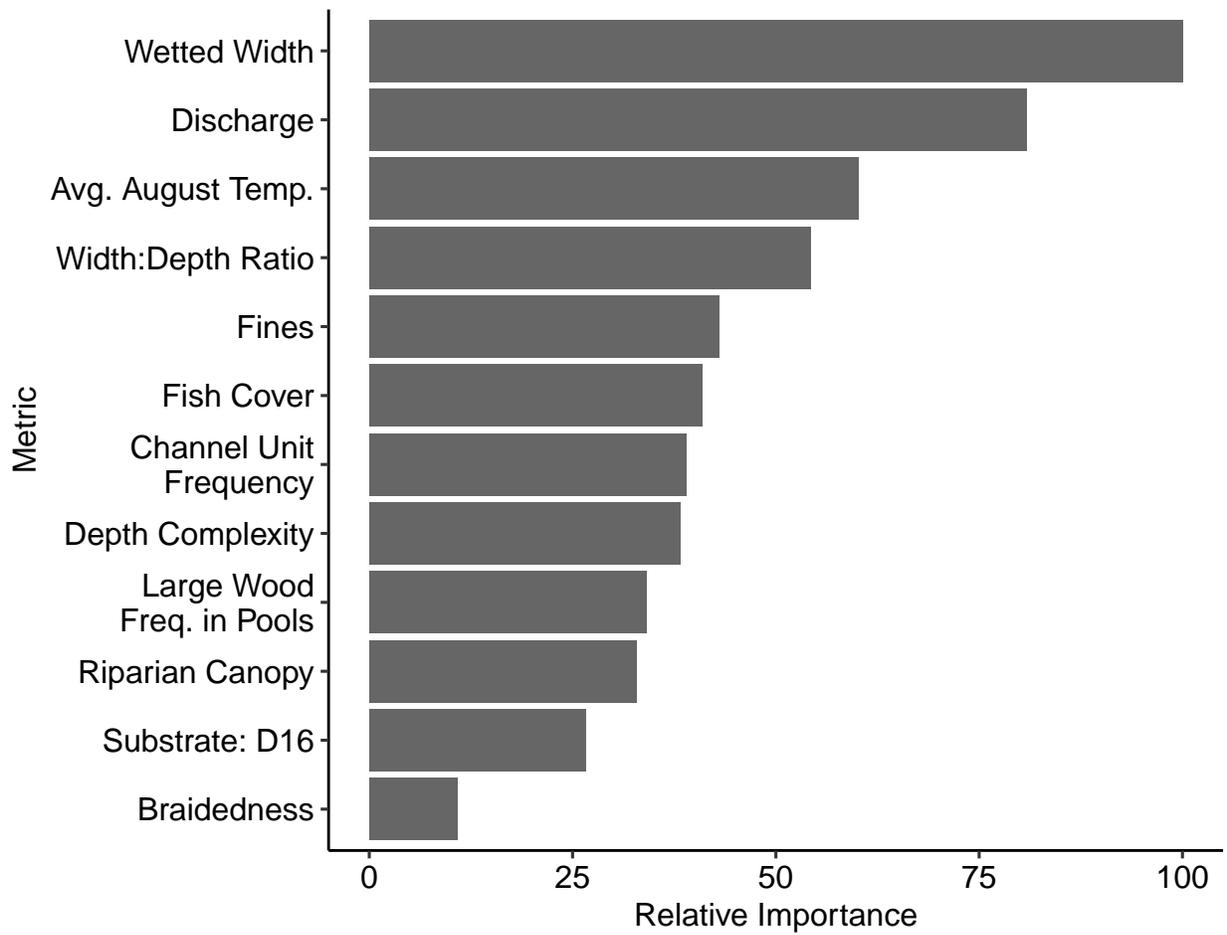


Figure 3: Relative importance of each habitat covariate included in the quantile random forest (QRF) model to predict habitat capacity, during summer months, for spring/summer-run Chinook salmon parr

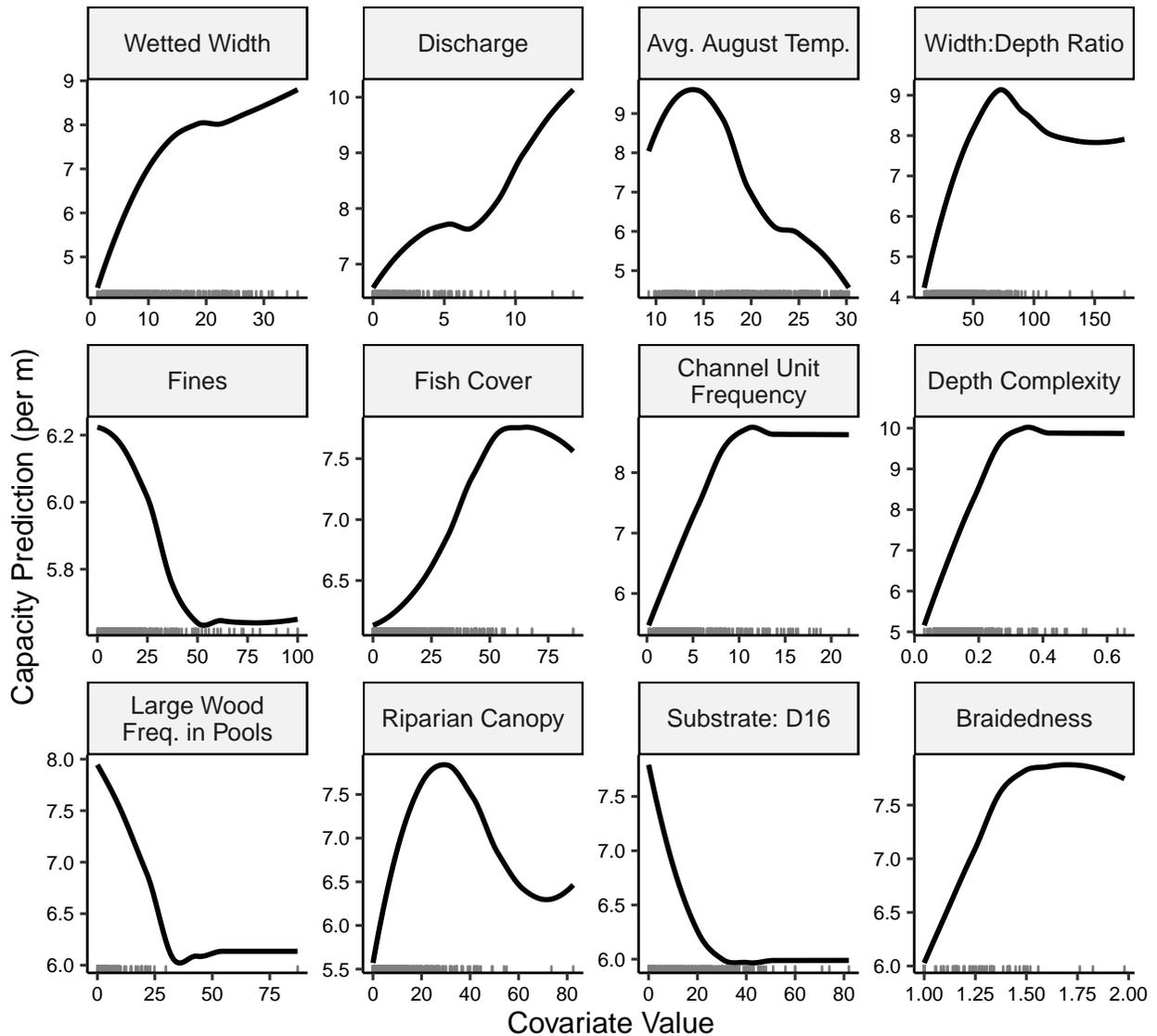


Figure 4: Partial dependence plots for the spring/summer-run Chinook salmon parr capacity quantile random forest (QRF) model, depicting how parr capacity shifts as each habitat metric changes, assuming all other habitat metrics remain at their mean values. Tick marks along the X-axis depict observed values.

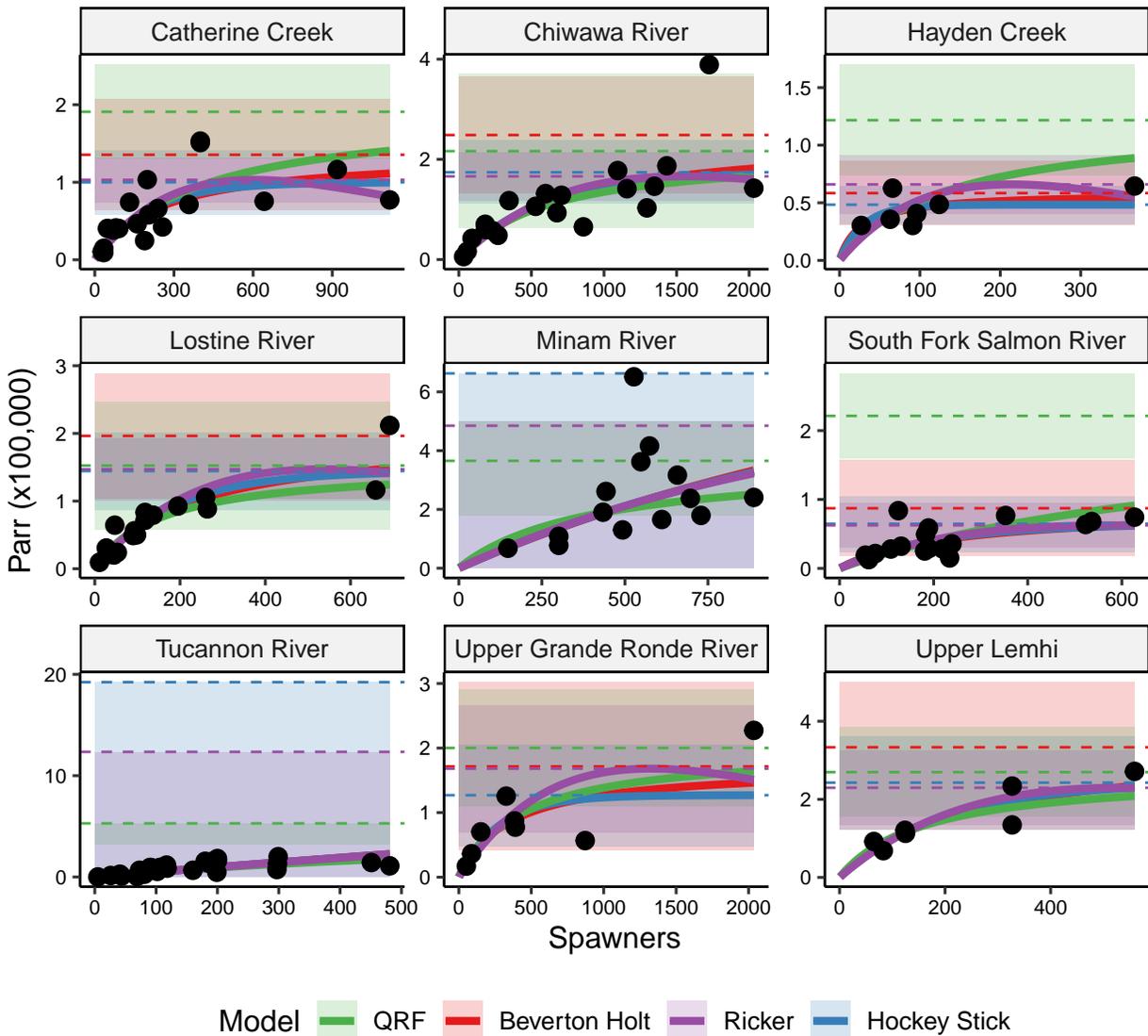


Figure 5: Spawner-recruit data from nine watersheds. Solid lines are the spawner-recruit curve, dashed lines are the estimated capacity, and shaded polygons depict the 95% confidence intervals of capacity. Red corresponds to Beverton-Holt models, purple to Ricker models, blue to hockey stick models, and green to QRF estimates. The QRF solid curve is a Beverton-Holt model with the capacity parameter fixed to the QRF estimate of capacity. A few curves with high capacity estimates were not plotted to improve readability.