Appendix S1

Estimating Carrying Capacity for Chinook Parr using Quantile Random Forest Models

Kevin E. See[1,*], Michael W. Ackerman[1], Richard A. Carmichael[1], Sarah L. Hoffmann[1], and Chris Beasle

June 09, 2020

[1] Biomark, Inc. 705 South 8th St., Boise, Idaho, 83702, USA

[*] Correspondence: Kevin E. See <Kevin.See@biomark.com>

1 Choosing Habitat Covariates

A key step in developing a QRF model to predict fish capacities was selecting the habitat covariates to include in the model. Random forest models naturally incorporate interactions between correlated covariates, which is essential since nearly all habitat variables are considered correlated to one degree or another, however, we aimed to avoid overly redundant variables (i.e., variables that measure similar aspects of the habitat). Further, including too many covariates can result in overfitting of the model (e.g., including as many covariates as data points).

To prevent overfitting, we pared down the more than 100 metrics generated by the CHaMP protocol describing the quantity and quality of fish habitat for each survey site. To assist in determining the habitat metrics to include in the QRF model, we used the Maximal Information-Based Nonparametric Exploration (MINE) class of statistics (Reshef et al. 2011) to determine those habitat characteristics (covariates) most highly associated with observed parr densities. We calculated the maximal information coefficient (MIC), using the R package *minerva* (Filosi et al. 2019), to measure the strength of the linear or non-linear association between two variables (Reshef et al. 2011). The MIC value between each of the measured habitat characteristics and parr density was used to inform decisions on which habitat covariates to include in the QRF parr capacity model.

Habitat metrics were first grouped into broad categories that included channel unit, complexity, cover, disturbance, riparian, size, substrate, temperature, water quality, and woody debris. Habitat metrics measuring volume or area were scaled to the wetted area of each site. Within each category, metrics were ranked according to their MIC value (**Figure ??**). Our strategy was to select one or two variables with the highest MIC score within each category so that covariates describe different aspects of rearing habitat (e.g., substrate, temperature, etc.). Additionally, we attempted to avoid covariates that were highly correlated (**Figure ??**) while including covariates that can be directly influenced by restoration actions or have been shown to impact salmonid juvenile density.

2 References

Filosi, M., R. Visintainer, and D. Albanese. 2019. Minerva: Maximal information-based nonparametric exploration for variable analysis.

Reshef, D. N., Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. 2011. Detecting novel associations in large data sets. Science 334:1518–1524.

3 Tables

Table S1: MIC statistic for top ten metrics in each habitat category.

Category	Name	MIC
ChannelUnit	Channel Unit Frequency	0.241
ChannelUnit	Fast Turbulent Frequency	0.230
ChannelUnit	Fast NonTurbulent Frequency	0.209
ChannelUnit	Slow Water Frequency	0.208
ChannelUnit	Fast Turbulent Percent	0.195
ChannelUnit	ChnlUnitTotal_Ct	0.189
ChannelUnit	Channel Unit Count	0.189
ChannelUnit	Fast Turbulent Count	0.178
ChannelUnit	Slow Water Percent	0.177
ChannelUnit	Fast NonTurbulent Percent	0.169
Complexity	Wetted Width To Depth Ratio Avg	0.247
Complexity	Bankfull Width To Depth Ratio Avg	0.245
Complexity	Wetted Depth SD	0.232
Complexity	Wetted Channel Braidedness	0.212
Complexity	Bankfull Channel Braidedness	0.211
Complexity	Wetted Channel Qualifying Island Count	0.209
Complexity	Bankfull Width CV	0.209
Complexity	Bankfull Width To Depth Ratio CV	0.202
Complexity	Detrended Elevation SD	0.196
Complexity	Bankfull Channel Qualifying Island Count	0.193
Cover	Fish Cover: Total	0.225
Cover	Fish Cover: None	0.224
Cover	Fish Cover: LW	0.213
Cover	Fish Cover: Terrestrial Vegetation	0.204
Cover	Percent Undercut by Length	0.185
Cover	Percent Undercut by Area	0.184
Cover	Fish Cover: Aquatic Vegetation	0.166
Cover	Fish Cover: Artificial	0.136
Land Classification	Natural Class PCA 2	0.271
Land Classification	Natural Class PCA 1	0.258
Land Classification	Disturbance Class PCA 1	0.242
Riparian	Riparian Cover: Understory	0.206
Riparian	RipCovUstoryNone	0.206
Riparian	Riparian Cover: No Canopy	0.194
Riparian	Riparian Cover: Some Canopy	0.194
Riparian	Riparian Cover: Big Tree	0.184
Riparian	Riparian Cover: Ground	0.182
Riparian	RipCovGrndNone	0.170
Riparian	Riparian Cover: Woody	0.168
Riparian	Riparian Cover: Non-Woody	0.166
Riparian	Riparian Cover: Coniferous	0.164
SideChannel	Bankfull Side Channel Width	0.223
SideChannel	Wetted Side Channel Width	0.213
SideChannel	Wetted Side Channel Percent By Area	0.209
SideChannel	SCSm_Freq	0.153
SideChannel	SCSm_Ct	0.153
SideChannel	SC_Area_Pct	0.153
Size	Mean Annual Flow	0.346
Size	Wetted Width Integrated	0.332

Category	Name	MIC
Size	Bankfull Width Integrated	0.324
Size	Wetted Width Avg	0.324
Size	Drainage Area (Flowline)	0.302
Size	Bankfull Width Avg	0.298
Size	DpthThlwg_Avg	0.280
Size	Discharge	0.259
Size	Bankfull Depth Avg	0.245
Size	Bankfull Depth Max	0.240
Substrate	Substrate < 6mm	0.237
Substrate	Substrate < 2mm	0.227
Substrate	Substrate: D16	0.219
Substrate	Substrate: Embeddedness Avg	0.204
Substrate	Substrate: D50	0.197
Substrate	Substrate Est: Sand and Fines	0.190
Substrate	Substrate Est: Cobbles	0.185
Substrate	Substrate: D84	0.185
Substrate	Substrate Est: Boulders	0.183
Substrate	Substrate: Embeddedness SD	0.181
Temperature	Avg. August Temperature	0.272
Temperature	Elev_M	0.262
Temperature	August Temperature	0.188
Temperature	Solar Access: Summer Avg	0.186
WaterQuality	Conductivity	0.254
WaterQuality	Alkalinity	0.225
WaterQuality	Drift Biomass	0.000
Wood	Large Wood Volume: Bankfull Slow Water	0.213
Wood	Large Wood Volume: Wetted Slow Water	0.207
Wood	Large Wood Frequency: Wetted	0.199
Wood	Large Wood Volume: Bankfull	0.189
Wood	Large Wood Volume: Wetted Fast Turbulent	0.187
Wood	Large Wood Frequency: Bankfull	0.178
Wood	Large Wood Volume: Bankfull Fast NonTurbulent	0.175
Wood	Large Wood Volume: Wetted	0.166
Wood	Large Wood Volume: Wetted Fast NonTurbulent	0.159

4 Figures

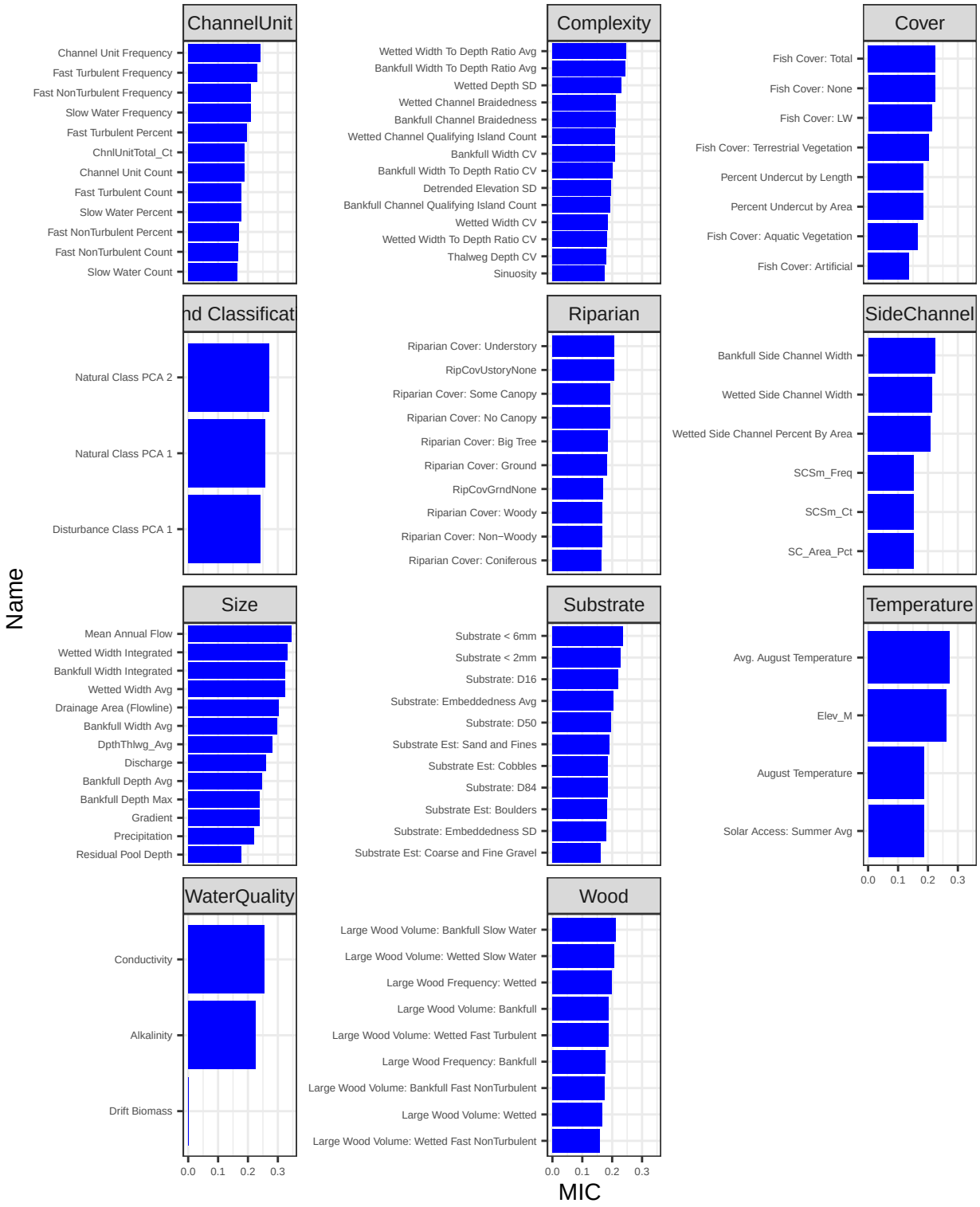

Figure S1: Barplots of MIC statistics, faceted by habitat category.

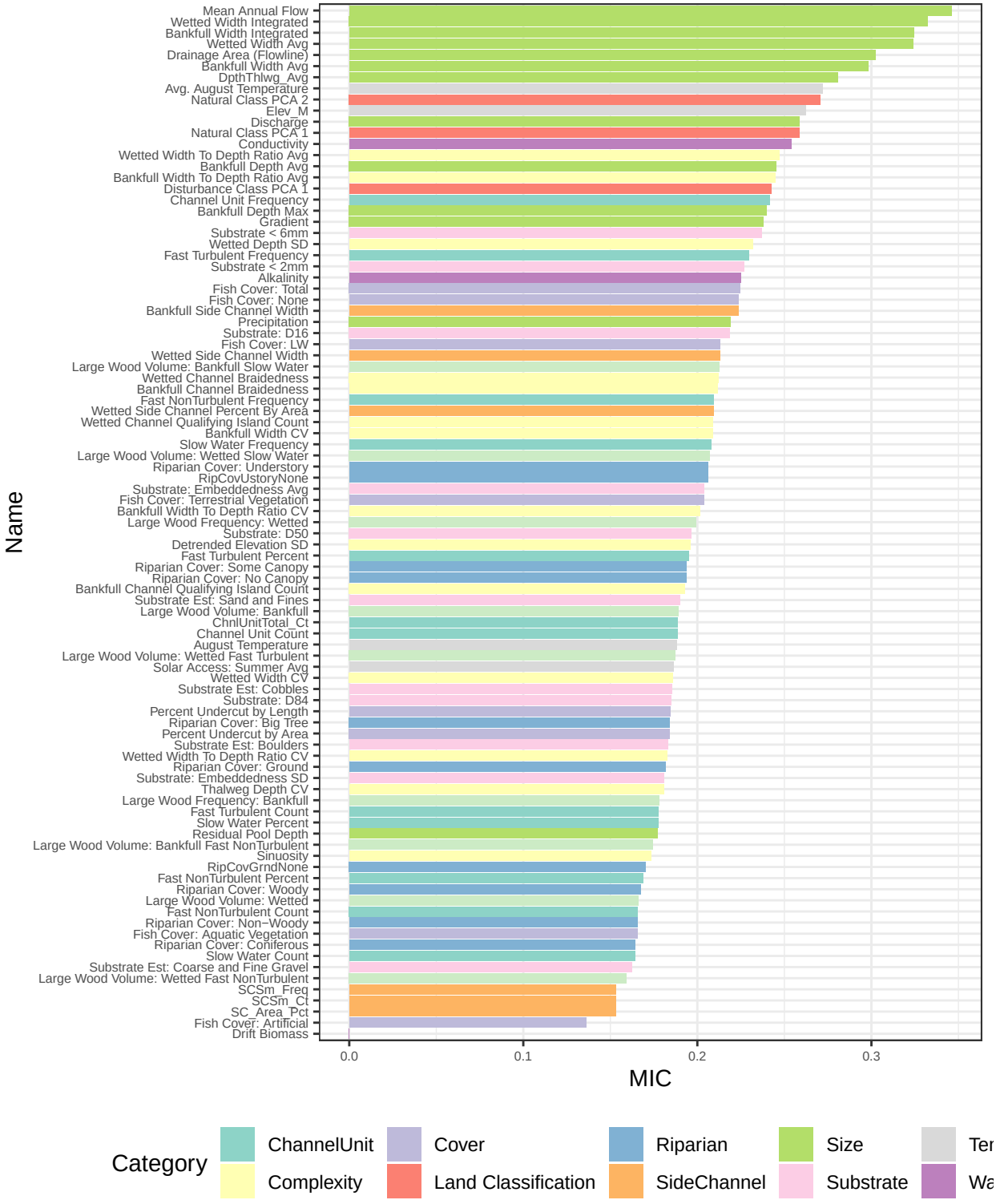

Figure S2: Barplot of MIC statistics, colored by habitat category.