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 18, 2020

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

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

1 Choosing Habitat Covariates

One of the crucial steps in building this carrying capacity model was choosing which habitat covariates to include. 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). Our goal was to select a group of covariates that captured as many different aspects of the stream habitat (e.g., substrate, flow, riparian condition, channel unit configuration, etc.) as possible, while still holding information about fish densities.

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. Habitat metrics were first grouped into broad categories that included channel unit configuration, complexity, fish cover, riparian areas, side channels, stream size, substrate, temperature, water quality, and woody debris. Habitat metrics measuring any large wood volume were scaled by the site length (in 100 m units). 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 the log of 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 the natural log of fish density and each habitat metric (Reshef et al. 2011). MIC is a measure of correlation that incorporates potential non-linear associations; for example, if there is a quadratic association the MIC value could be high, even when the standard correlation coefficient is low. We excluded categorical variables such as channel type (e.g., meandering, pool-riffle, plane-bed, etc.) because we assumed that other quantitative metrics would capture the differences between those qualitative categorical metrics.

Within each category, metrics were ranked according to their MIC value (Table S1 and Figure S1). The MIC value of each measured habitat characteristic and parr density was used to inform decisions on which habitat covariates to include in the QRF parr capacity model. We selected one or two variables amongst those with the highest MIC scores within each category, attempting to avoid covariates that were too highly correlated (Figure S3), while focusing on covariates we thought could influence fish behavior. For example, cumulative drainage area, mean annual flow and observed discharge are all highly correlated, but fish really only experience the observed discharge, so we chose to include that metric in our QRF model. We also tried to include covariates that can be directly influenced by rehabilitation actions or have been shown to impact salmonid juvenile density. Finally, we attempted to avoid metrics with too many missing values, or too many zero values, in the data set, as well as metrics that may have too much observer error (Rosgen et al. 2018).

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.

Rosgen, D., A. Taillacq, B. Rosgen, and D. Geenen. 2018. A technical review of the Columbia Habitat Monitoring Program's protocol, data quality.

3 Tables

Table S1: MIC statistic for top metrics within each habitat category, sorted by category and MIC value. The percent of records for which each habitat metric measurement was missing or zero is also shown. Metrics selected for the QRF model are in bold.

Category	Name	Abbrv	MIC	Percent Missing	Percent 0-value
ChannelUnit	**Channel Unit Frequency**	**CU_Freq**	**0.241**	**0.021**	**0.021**
ChannelUnit	Fast Turbulent Frequency	FstTurb_Freq	0.230	0.021	0.082
ChannelUnit	Fast NonTurbulent Frequency	FstNT_Freq	0.209	0.021	0.308
ChannelUnit	Slow Water Frequency	SlowWater_Freq	0.208	0.021	0.073
ChannelUnit	Fast Turbulent Percent	FstTurb_Pct	0.195	0.021	0.082
ChannelUnit	ChnlUnitTotal_Ct	ChnlUnitTotal_Ct	0.189	0.021	0.021
ChannelUnit	Channel Unit Count	CU_Ct	0.189	0.021	0.021
ChannelUnit	Fast Turbulent Count	FstTurb_Ct	0.178	0.021	0.082
ChannelUnit	Slow Water Percent	SlowWater_Pct	0.177	0.021	0.073
ChannelUnit	Fast NonTurbulent Percent	FstNT_Pct	0.169	0.021	0.308
Complexity	**Wetted Width To Depth Ratio Avg**	**WetWDRat_Avg**	**0.247**	**0.003**	**0.003**
Complexity	Bankfull Width To Depth Ratio Avg	BfWDRat_Avg	0.245	0.003	0.003
Complexity	**Wetted Depth SD**	**DpthWet_SD**	**0.232**	**0.003**	**0.003**
Complexity	**Wetted Channel Braidedness**	**WetBraid**	**0.212**	**0.003**	**0.003**
Complexity	Bankfull Channel Braidedness	BfBraid	0.211	0.003	0.003
Complexity	Wetted Channel Qualifying Island Count	Wet_QIsland_Ct	0.209	0.003	0.835
Complexity	Bankfull Width CV	BfWdth_CV	0.209	0.003	0.003
Complexity	Bankfull Width To Depth Ratio CV	BfWDRat_CV	0.202	0.003	0.003
Complexity	Detrended Elevation SD	DetrendElev_SD	0.196	0.003	0.003
Complexity	Bankfull Channel Qualifying Island Count	Bf_QIsland_Ct	0.193	0.003	0.780
Cover	**Fish Cover: Total**	**FishCovTotal**	**0.225**	**0.021**	**0.030**
Cover	Fish Cover: None	FishCovNone	0.224	0.021	0.021
Cover	Fish Cover: LW	FishCovLW	0.213	0.021	0.155
Cover	Fish Cover: Terrestrial Vegetation	FishCovTVeg	0.204	0.021	0.052
Cover	Percent Undercut by Length	UcutLgth_Pct	0.185	0.000	0.476
Cover	Percent Undercut by Area	UcutArea_Pct	0.184	0.000	0.476
Cover	Fish Cover: Aquatic Vegetation	FishCovAqVeg	0.166	0.296	0.631
Cover	Fish Cover: Artificial	FishCovArt	0.136	0.021	0.851
Riparian	Riparian Cover: Understory	RipCovUstory	0.206	0.000	0.000
Riparian	RipCovUstoryNone	RipCovUstoryNone	0.206	0.000	0.000
Riparian	Riparian Cover: No Canopy	RipCovCanNone	0.194	0.000	0.000
Riparian	**Riparian Cover: Some Canopy**	**RipCovCanSome**	**0.194**	**0.000**	**0.095**
Riparian	Riparian Cover: Big Tree	RipCovBigTree	0.184	0.000	0.183

Table S1: MIC statistic for top metrics within each habitat category, sorted by category and MIC value. The percent of records for which each habitat metric measurement was missing or zero is also shown. Metrics selected for the QRF model are in bold. *(continued)*

Category	Name	Abbrv	MIC	Percent Missing	Percent 0-value
Riparian	Riparian Cover: Ground	RipCovGrnd	0.182	0.000	0.000
Riparian	RipCovGrndNone	RipCovGrndNone	0.170	0.000	0.003
Riparian	Riparian Cover: Woody	RipCovWood	0.168	0.000	0.000
Riparian	Riparian Cover: Non-Woody	RipCovNonWood	0.166	0.000	0.000
Riparian	Riparian Cover: Coniferous	RipCovConif	0.164	0.009	0.192
SideChannel	Bankfull Side Channel Width	BfSCWdth	0.223	0.796	0.796
SideChannel	Wetted Side Channel Width	WetSCWdth	0.213	0.832	0.832
SideChannel	Wetted Side Channel Percent By Area	WetSC_Pct	0.209	0.021	0.820
SideChannel	SCSm_Freq	SCSm_Freq	0.153	0.021	0.921
SideChannel	SCSm_Ct	SCSm_Ct	0.153	0.021	0.921
SideChannel	SC_Area_Pct	SC_Area_Pct	0.153	0.021	0.921
Size	Mean Annual Flow	MeanU	0.346	0.476	0.476
Size	**Wetted Width Integrated**	**WetWdth_Int**	**0.332**	**0.003**	**0.003**
Size	Bankfull Width Integrated	BfWdthInt	0.324	0.003	0.003
Size	Wetted Width Avg	WetWdth_Avg	0.324	0.003	0.003
Size	Drainage Area (Flowline)	CUMDRAINAG	0.302	0.341	0.341
Size	Bankfull Width Avg	BfWdth_Avg	0.298	0.003	0.003
Size	DpthThlwg_Avg	DpthThlwg_Avg	0.280	0.003	0.003
Size	**Discharge**	**Q**	**0.259**	**0.037**	**0.037**
Size	Bankfull Depth Avg	DpthBf_Avg	0.245	0.018	0.018
Size	Bankfull Depth Max	DpthBf_Max	0.240	0.018	0.018
Substrate	**Substrate < 6mm**	**SubLT6**	**0.237**	**0.049**	**0.055**
Substrate	Substrate < 2mm	SubLT2	0.227	0.049	0.082
Substrate	**Substrate: D16**	**SubD16**	**0.219**	**0.012**	**0.012**
Substrate	Substrate: Embeddedness Avg	SubEmbed_Avg	0.204	0.293	0.317
Substrate	Substrate: D50	SubD50	0.197	0.012	0.012
Substrate	Substrate Est: Sand and Fines	SubEstSandFines	0.190	0.021	0.030
Substrate	Substrate Est: Cobbles	SubEstCbl	0.185	0.021	0.027
Substrate	Substrate: D84	SubD84	0.185	0.012	0.012
Substrate	Substrate Est: Boulders	SubEstBldr	0.183	0.021	0.149
Substrate	Substrate: Embeddedness SD	SubEmbed_SD	0.181	0.302	0.320
Temperature	**Avg. August Temperature**	**avg_aug_temp**	**0.272**	**0.000**	**0.000**
Temperature	Elev_M	Elev_M	0.262	0.363	0.363
Temperature	August Temperature	aug_temp	0.188	0.155	0.155
Temperature	Solar Access: Summer Avg	SolarSummr_Avg	0.186	0.070	0.070
WaterQuality	Conductivity	Cond	0.254	0.024	0.027
WaterQuality	Alkalinity	Alk	0.225	0.009	0.027
WaterQuality	Drift Biomass	DriftBioMass	0.000	0.277	0.384
Wood	Large Wood Volume: Bankfull Slow Water	LWVol_BfSlow	0.213	0.003	0.232

Table S1: MIC statistic for top metrics within each habitat category, sorted by category and MIC value. The percent of records for which each habitat metric measurement was missing or zero is also shown. Metrics selected for the QRF model are in bold. *(continued)*

Category	Name	Abbrv	MIC	Percent Missing	Percent 0-value
Wood	**Large Wood Volume: Wetted Slow Water**	**LWVol_WetSlow**	**0.207**	**0.003**	**0.290**
Wood	Large Wood Frequency: Wetted	LWFreq_Wet	0.199	0.003	0.125
Wood	Large Wood Volume: Bankfull	LWVol_Bf	0.189	0.003	0.085
Wood	Large Wood Volume: Wetted Fast Turbulent	LWVol_WetFstTurb	0.187	0.003	0.274
Wood	Large Wood Frequency: Bankfull	LWFreq_Bf	0.178	0.003	0.085
Wood	Large Wood Volume: Bankfull Fast NonTurbulent	LWVol_BfFstNT	0.175	0.003	0.521
Wood	Large Wood Volume: Wetted	LWVol_Wet	0.166	0.003	0.125
Wood	Large Wood Volume: Wetted Fast NonTurbulent	LWVol_WetFstNT	0.159	0.003	0.595

4 Figures

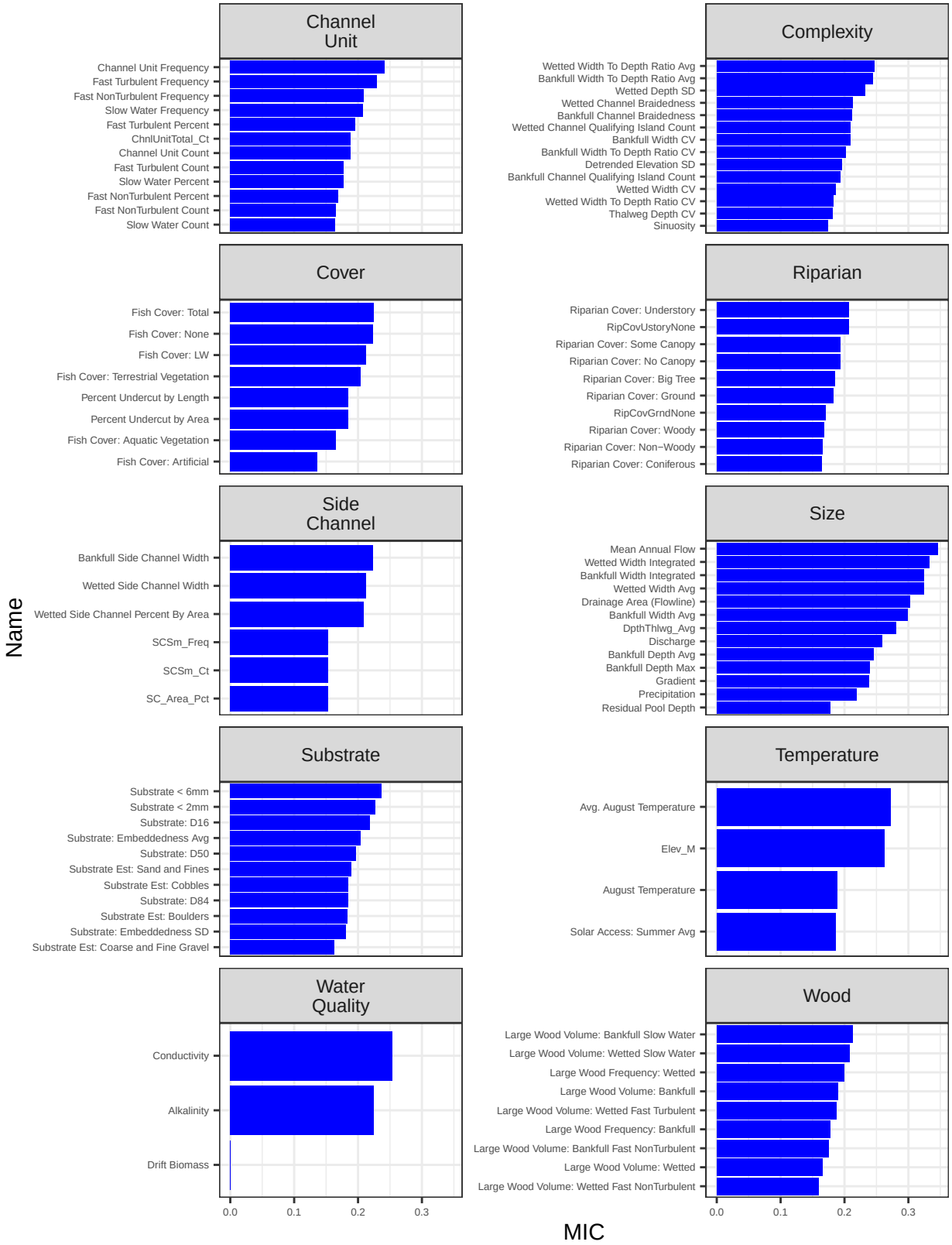


Figure S1: Barplots of MIC statistics, faceted by habitat category.

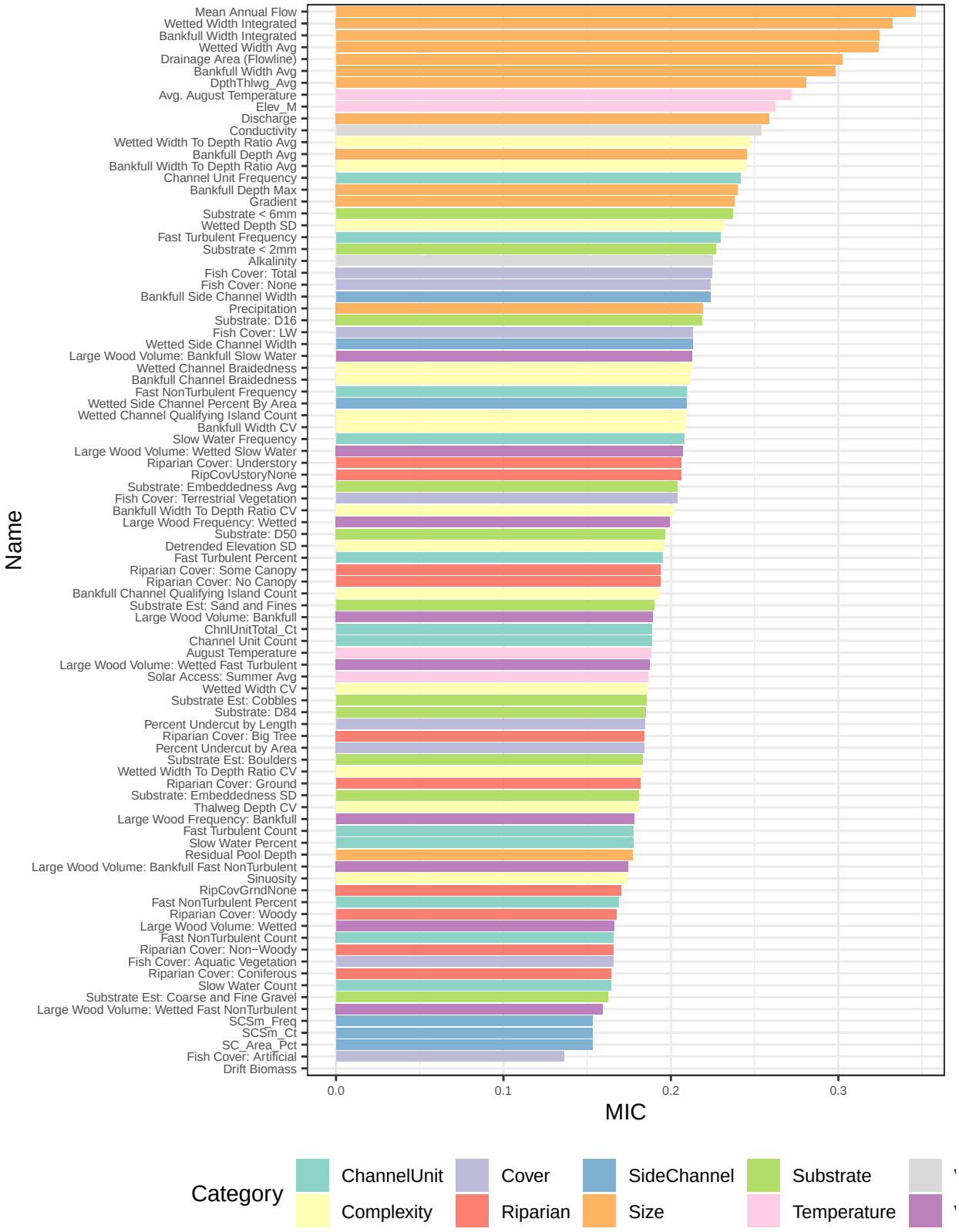
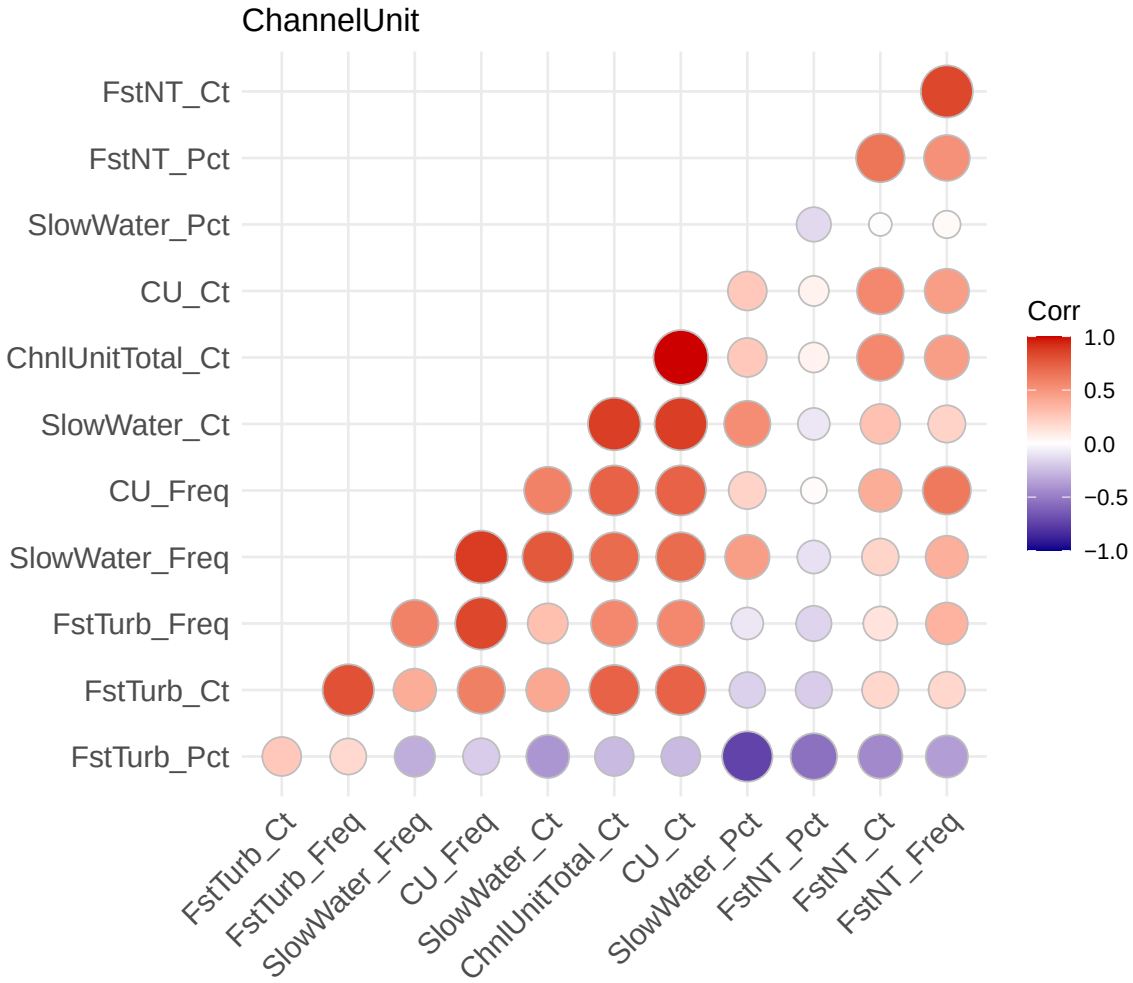
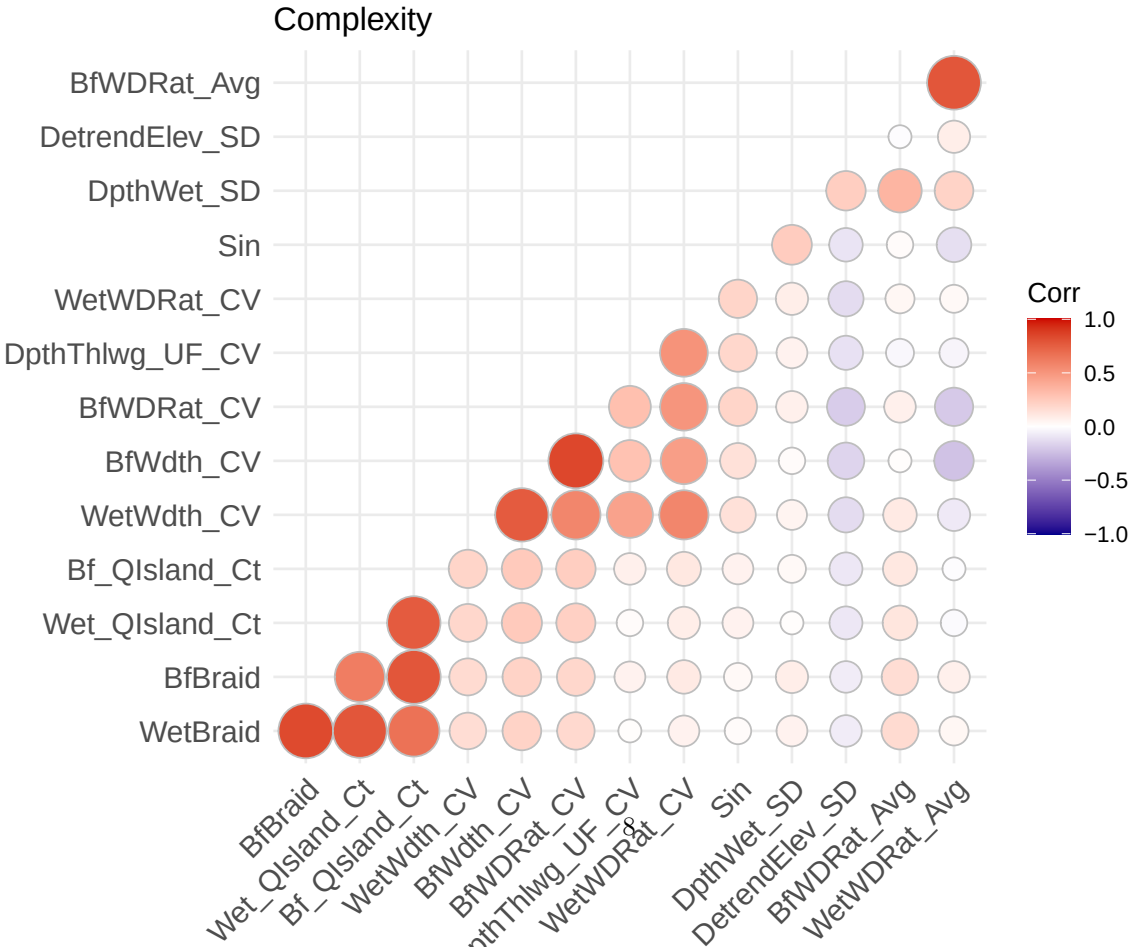


Figure S2: Barplot of MIC statistics, colored by habitat category.


ChannelUnit
FstNT_Ct
FstNT_Pct
SlowWater_Pct
CU_Ct
ChnlUnitTotal_Ct
SlowWater_Ct
CU_Freq
SlowWater_Freq
FstTurb_Freq
FstTurb_Ct
FstTurb_Pct
FstTurb_Ct
FstTurb_Freq
SlowWater_Freq
CU_Freq
SlowWater_Ct
ChnlUnitTotal_Ct
CU_Ct
SlowWater_Pct
FstNT_Pct
FstNT_Ct
FstNT_Freq
Corr
1.0
0.5
0.0
−0.5
−1.0


Complexity
BfWDRat_Avg
DetrendElev_SD
DpthWet_SD
Sin
WetWDRat_CV
DpthThlwg_UF_CV
BfWDRat_CV
BfWdth_CV
WetWdth_CV
Bf_QIsland_Ct
Wet_QIsland_Ct
BfBraid
WetBraid
BfBraid
Wet_QIsland_Ct
Bf_QIsland_Ct
WetWdth_CV
BfWdth_CV
BfWDRat_CV
DpthThlwg_UF_CV
WetWDRat_CV
Sin
DpthWet_SD
DetrendElev_SD
BfWDRat_Avg
WetWDRat_Avg
Corr
1.0
0.5
0.0
−0.5
−1.0

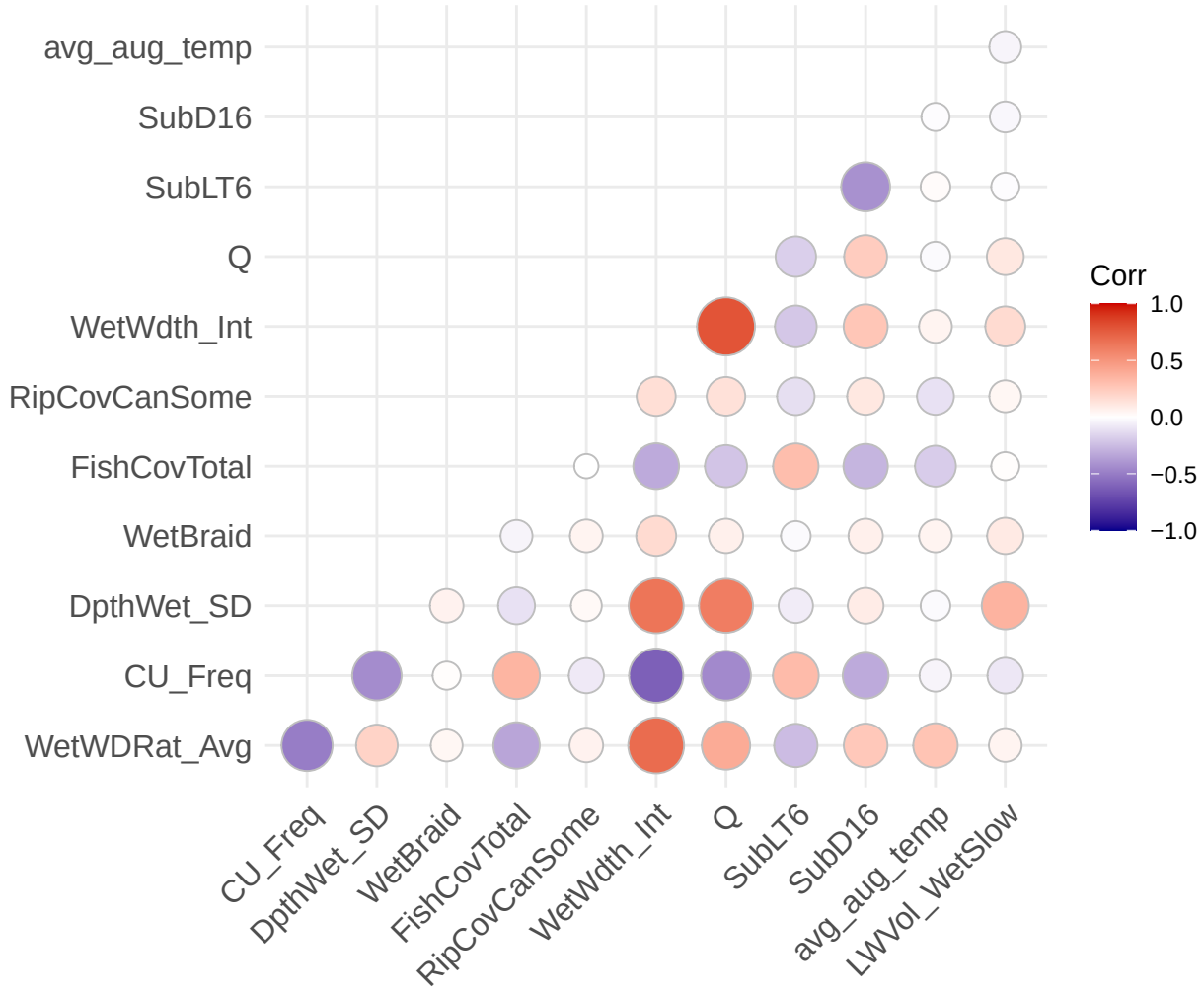


Figure S4: Correlation plot of habitat metrics used in the QRF model.

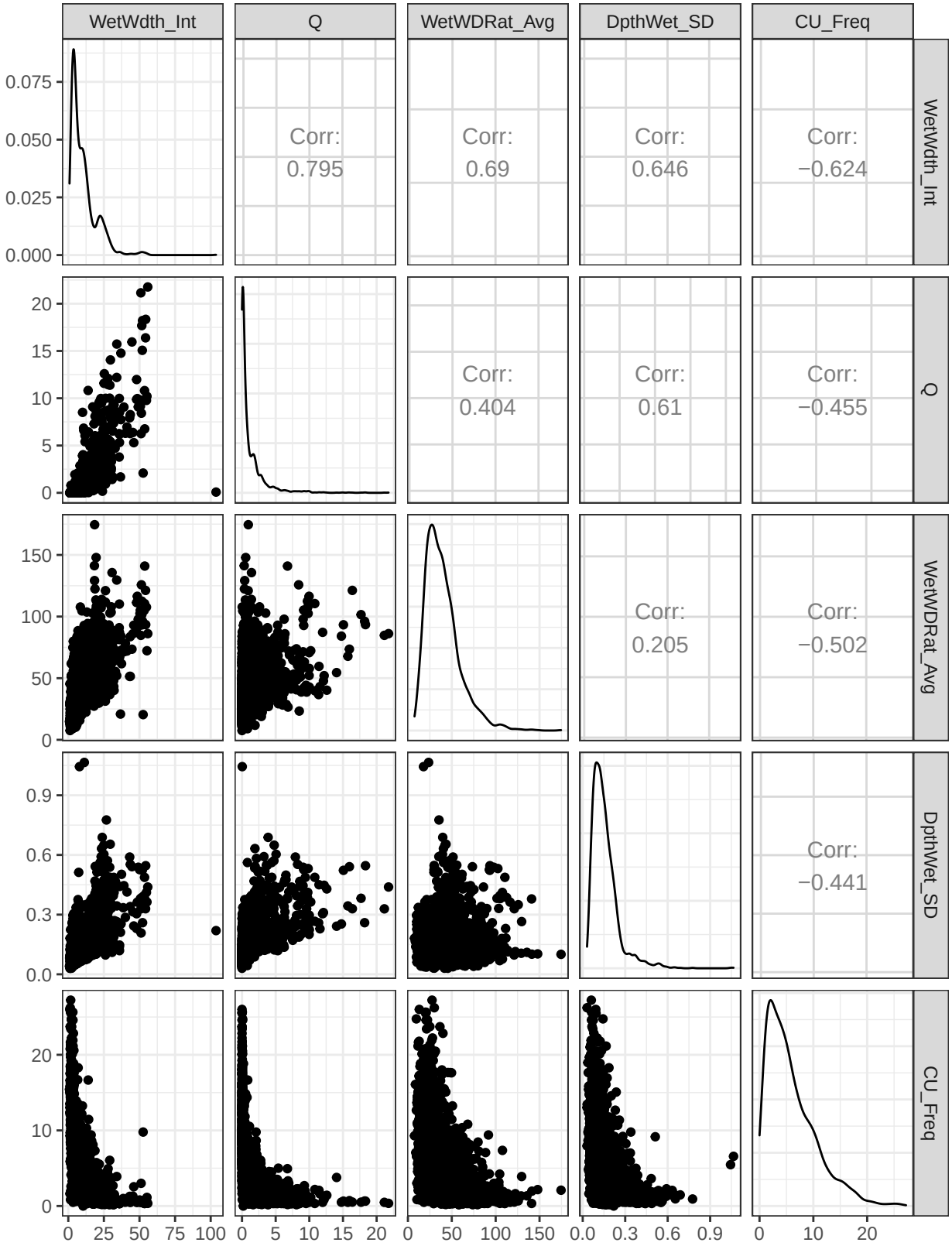


Figure S5: Pairs plot of habitat metrics used in QRF model with a correlation coefficient greater than 0.5.

4.0.1 Colophon

This report was generated on 2020-06-18 16:46:37 using the following computational environment and dependencies:

```
#> - Session info ---------------------------------------------------------------
#>  setting  value
#>  version  R version 3.6.1 (2019-07-05)
#>  os       macOS Mojave 10.14.6
#>  system   x86_64, darwin15.6.0
#>  ui       X11
#>  language (EN)
#>  collate  en_US.UTF-8
#>  ctype    en_US.UTF-8
#>  tz       America/Los_Angeles
#>  date     2020-06-18
#>
#> - Packages -------------------------------------------------------------------
#>  package        * version    date       lib source
#>  assertthat       0.2.1      2019-03-21 [2] CRAN (R 3.6.0)
#>  backports        1.1.6      2020-04-05 [1] CRAN (R 3.6.2)
#>  bookdown         0.17       2020-01-11 [1] CRAN (R 3.6.0)
#>  broom            0.5.3      2019-12-14 [1] CRAN (R 3.6.0)
#>  callr            3.4.3      2020-03-28 [1] CRAN (R 3.6.2)
#>  cellranger       1.1.0      2016-07-27 [2] CRAN (R 3.6.0)
#>  cli              2.0.2      2020-02-28 [1] CRAN (R 3.6.0)
#>  colorspace       1.4-1      2019-03-18 [2] CRAN (R 3.6.0)
#>  corrr          * 0.4.0      2019-07-12 [1] CRAN (R 3.6.0)
#>  crayon           1.3.4      2017-09-16 [2] CRAN (R 3.6.0)
#>  DBI              1.0.0      2018-05-02 [2] CRAN (R 3.6.0)
#>  dbplyr           1.4.2      2019-06-17 [2] CRAN (R 3.6.0)
#>  desc             1.2.0      2018-05-01 [1] CRAN (R 3.6.0)
#>  devtools         2.2.1      2019-09-24 [1] CRAN (R 3.6.0)
#>  digest           0.6.25     2020-02-23 [1] CRAN (R 3.6.0)
#>  dplyr          * 0.8.5      2020-03-07 [1] CRAN (R 3.6.0)
#>  ellipsis         0.3.0      2019-09-20 [2] CRAN (R 3.6.0)
#>  evaluate         0.14       2019-05-28 [2] CRAN (R 3.6.0)
#>  fansi            0.4.1      2020-01-08 [1] CRAN (R 3.6.0)
#>  farver           2.0.3      2020-01-16 [1] CRAN (R 3.6.0)
#>  forcats        * 0.5.0      2020-03-01 [1] CRAN (R 3.6.0)
#>  fs               1.3.1      2019-05-06 [2] CRAN (R 3.6.0)
#>  generics         0.0.2      2018-11-29 [2] CRAN (R 3.6.0)
#>  GGally           1.4.0      2018-05-17 [1] CRAN (R 3.6.0)
#>  ggcorrplot       0.1.3      2019-05-19 [1] CRAN (R 3.6.0)
#>  ggplot2        * 3.3.0      2020-03-05 [1] CRAN (R 3.6.0)
#>  glue             1.4.0      2020-04-03 [1] CRAN (R 3.6.2)
#>  gtable           0.3.0      2019-03-25 [2] CRAN (R 3.6.0)
#>  haven            2.2.0      2019-11-08 [2] CRAN (R 3.6.0)
#>  highr            0.8        2019-03-20 [2] CRAN (R 3.6.0)
#>  hms              0.5.3      2020-01-08 [1] CRAN (R 3.6.0)
#>  htmltools        0.4.0      2019-10-04 [2] CRAN (R 3.6.0)
#>  httr             1.4.1      2019-08-05 [2] CRAN (R 3.6.0)
#>  jsonlite         1.6        2018-12-07 [2] CRAN (R 3.6.0)
#>  kableExtra     * 1.1.0      2019-03-16 [1] CRAN (R 3.6.0)
#>  knitr            1.27       2020-01-16 [1] CRAN (R 3.6.0)
```

```
#>  labeling         0.3        2014-08-23 [2] CRAN (R 3.6.0)
#>  lattice          0.20-38    2018-11-04 [2] CRAN (R 3.6.1)
#>  lifecycle        0.2.0      2020-03-06 [1] CRAN (R 3.6.0)
#>  lubridate        1.7.8      2020-04-06 [1] CRAN (R 3.6.2)
#>  magrittr         1.5        2014-11-22 [1] CRAN (R 3.6.0)
#>  memoise          1.1.0      2017-04-21 [1] CRAN (R 3.6.0)
#>  minerva        * 1.5.8      2019-05-27 [1] CRAN (R 3.6.0)
#>  modelr           0.1.5      2019-08-08 [2] CRAN (R 3.6.0)
#>  munsell          0.5.0      2018-06-12 [2] CRAN (R 3.6.0)
#>  nlme             3.1-142    2019-11-07 [1] CRAN (R 3.6.0)
#>  pillar           1.4.3      2019-12-20 [1] CRAN (R 3.6.0)
#>  pkgbuild         1.0.8      2020-05-07 [1] CRAN (R 3.6.2)
#>  pkgconfig        2.0.3      2019-09-22 [2] CRAN (R 3.6.0)
#>  pkgload          1.0.2      2018-10-29 [1] CRAN (R 3.6.0)
#>  plyr             1.8.5      2019-12-10 [1] CRAN (R 3.6.0)
#>  prettyunits      1.1.1      2020-01-24 [1] CRAN (R 3.6.0)
#>  processx         3.4.2      2020-02-09 [1] CRAN (R 3.6.0)
#>  ps               1.3.3      2020-05-08 [1] CRAN (R 3.6.2)
#>  purrr          * 0.3.4      2020-04-17 [1] CRAN (R 3.6.2)
#>  QRFpaper       * 0.0.0.9000 2020-06-08 [1] local
#>  quantregForest   1.3-7      2017-12-19 [2] CRAN (R 3.6.0)
#>  R6               2.4.1      2019-11-12 [2] CRAN (R 3.6.0)
#>  randomForest     4.6-14     2018-03-25 [2] CRAN (R 3.6.0)
#>  RColorBrewer     1.1-2      2014-12-07 [2] CRAN (R 3.6.0)
#>  Rcpp             1.0.4.6    2020-04-09 [1] CRAN (R 3.6.1)
#>  readr          * 1.3.1      2018-12-21 [2] CRAN (R 3.6.0)
#>  readxl           1.3.1      2019-03-13 [1] CRAN (R 3.6.0)
#>  remotes          2.1.0      2019-06-24 [1] CRAN (R 3.6.0)
#>  reprex           0.3.0      2019-05-16 [2] CRAN (R 3.6.0)
#>  reshape          0.8.8      2018-10-23 [1] CRAN (R 3.6.0)
#>  reshape2         1.4.3      2017-12-11 [2] CRAN (R 3.6.0)
#>  rlang            0.4.6      2020-05-02 [1] CRAN (R 3.6.2)
#>  rmarkdown        2.1        2020-01-20 [1] CRAN (R 3.6.1)
#>  rprojroot        1.3-2      2018-01-03 [1] CRAN (R 3.6.0)
#>  rstudioapi       0.11       2020-02-07 [1] CRAN (R 3.6.0)
#>  rvest            0.3.5      2019-11-08 [2] CRAN (R 3.6.0)
#>  scales           1.1.0      2019-11-18 [2] CRAN (R 3.6.0)
#>  sessioninfo      1.1.1      2018-11-05 [1] CRAN (R 3.6.0)
#>  stringi          1.4.6      2020-02-17 [1] CRAN (R 3.6.0)
#>  stringr        * 1.4.0      2019-02-10 [2] CRAN (R 3.6.0)
#>  testthat         2.3.2      2020-03-02 [1] CRAN (R 3.6.0)
#>  tibble         * 3.0.1      2020-04-20 [1] CRAN (R 3.6.2)
#>  tidyr          * 1.0.3      2020-05-07 [1] CRAN (R 3.6.2)
#>  tidyselect       1.0.0      2020-01-27 [1] CRAN (R 3.6.0)
#>  tidyverse      * 1.3.0      2019-11-21 [1] CRAN (R 3.6.0)
#>  usethis          1.5.1      2019-07-04 [1] CRAN (R 3.6.0)
#>  vctrs            0.2.4      2020-03-10 [1] CRAN (R 3.6.0)
#>  viridisLite      0.3.0      2018-02-01 [2] CRAN (R 3.6.0)
#>  webshot          0.5.2      2019-11-22 [1] CRAN (R 3.6.0)
#>  withr            2.2.0      2020-04-20 [1] CRAN (R 3.6.2)
#>  xfun             0.12       2020-01-13 [1] CRAN (R 3.6.0)
#>  xml2             1.2.2      2019-08-09 [2] CRAN (R 3.6.0)
#>  yaml             2.2.0      2018-07-25 [2] CRAN (R 3.6.0)
#>
```

```
#> [1] /Users/seek/Library/R/3.6/library
#> [2] /Library/Frameworks/R.framework/Versions/3.6/Resources/library
```

The current Git commit details are:

```
#> Local:    master /Users/seek/Documents/GitProjects/MyProjects/QRFpaper
#> Remote:   master @ origin (git@github.com:KevinSee/QRFpaper.git)
#> Head:     [ae5812d] 2020-06-18: moved appendix to it's own folder. added a _bookdown.yml file to mak
```