Guidance document: QRF Extrapolation Process

Created: 4/7/2022

Updated: 4/7/2022 MR

Overview

The quantile random forest (QRF) extrapolation process is the final component for estimating reach, watershed, and Columbia River basin (CRB) carrying capacities for steelhead and Chinook Salmon at the juvenile and spawning life stages (See et al. 2021). The entire QRF process can be broken down to a “QRF model” and an “Extrapolation model” that, briefly:

**QRF Model**

1. Inputs habitat (CHaMP) and fish survey data
2. Estimates fish densities and imputes habitat data when necessary
3. Pairs fish observations with habitat data
4. Fits a QRF model to predict estimated densities based on habitat covariates
5. QRF model predicts fish distributions
6. Estimates carrying capacity by taking the 90th percentile of the estimated density
7. Predicts capacity for all paired fish/habitat sites

**Extrapolation Model**

1. Inputs globally available attributes (GAAs) and/or DASH data
2. Selects covariates for extrapolation
3. Fits an extrapolation model for the entire CRB (currently linear regression)
4. The extrapolation model uses QRF-based capacity predictions at all CHaMP sites to estimate the effect of the chosen covariates
5. Capacities are estimated at master points or 200 meter reaches depending on the script used, “extrap\_qrf\_mastPts.r” and “extrap\_qrf\_200rch.r”, respectively.
6. This outputs model fits for each combination of species and life stage (e.g., “extrap\_200rch\_juv\_summer”, “extrap\_mastPts\_juv\_summer”, “extrap\_mastPts\_redds”) as well as associated geopackage files.
7. Capacity at the desired scale (e.g., reach, watershed, etc.) is then obtained using model fits as well as an input shape file for each species and life stage desired
8. Shape files define the study area, habitat features, and species extents within the polygons
9. Capacity is then provided using the “calc\_watershed\_cap” function
10. Capacity outputs can then be summarized and mapped

Estimating capacity

This section assumes that QRF and extrapolation models are complete and covers the process to obtain capacity estimates for the desired reaches/streams/watershed (steps 7-10 above). For this process, you will need:

* Extrapolation model fit for the desired species, life stages, and data inputs. For example, summer juvenile capacity for 200 meter reaches with GAAs would require “extrap\_200rch\_juv\_summer”, while summer juvenile capacity with DASH data inputs would require the “extrap\_200rch\_juv\_summer\_dash” model fit.
* The “calc\_watershed\_cap” function
* Shape files prepped into .rda format. These contain watershed boundaries, spatial extents, habitat data, etc.

Preparing shape files

In general, there are four shapefiles that need to be prepared to estimate watershed/reach/site carrying capacity:

* HUC12 Watershed Boundaries
  + NAS/main/data/habitat/watershed\_boundaries/WBDHU12.shp
* The boundary for the watershed of interest
  + .shp file needs to be prepared manually in QGIS
* Morgan Bond's spatially continuous, 200m linear network layer
  + …/Git/QRFCapacity/data/rch\_200.rda
* QRF extrapolation output
  + main/data/qrf/extrapolations/Rch\_Cap\_RF\_juv\_summer\_dash.gpkg
  + main/data/qrf/extrapolations/Rch\_Cap\_RF\_juv\_winter.gpkg
  + main/data/qrf/extrapolations/Rch\_Cap\_RF\_redds.gpkg

These data are imported into a script (examples scripts are Git\NF\_assess\analysis\R\_scripts\prep\_NF\_data.R, or Git\UWW.plan\analysis\R\_scripts\prep\_UWW\_data) where they are cleaned and formatted into the shapefiles needed to estimate watershed/reach/site carrying capacity using the “calc\_watershed\_cap” function

* \*\*Still working on getting to describe this in more/better detail\*\*