# Technical Comments

**Subject:** Ongoing development and applications of quantile random forest (QRF) habitat capacity models

**Date:** *4/20/2022*

**Prepared for:** Bureau of Reclamation, Columbia/Snake Salmon Recovery Office (CSRO)

**Prepared by:** Mount Hood Environmental (MHE)

**Suggested Citation:** Ackerman, M.W. 2022. Technical Comments prepared for Bureau of Reclamation, Columbia/Snake Recovery Office regarding ongoing development and applications of quantile random forest (QRF) capacity models.

## Background

The purpose of this document is to provide a brief background of current and ongoing development of quantile random forest (QRF) habitat capacity models (See et al. 2021) by MHE personnel and collaborators. We also provide a summary of upcoming applications of QRF and related extrapolation models.

## Development

Many of the current developments are related to “Next Steps” provided in a presentation given to a Grande Ronde working group on November 27, 2021 by Kevin See and Mike Ackerman. Developments include:

### Improvements to QRF Covariates

Previous sets of habitat covariates used in the six QRF models (Chinook salmon and steelhead; summer parr, winter presmolts and redds) were chosen because of their high predictive power to estimate capacity across the Columbia River Basin. The summer parr models were validated using independent estimates of capacity from spawner-recruit data and performed very well. However, a subset of the covariates included in the models are not necessarily useful for restoration project monitoring, or to describe target conditions for restoration design due to the habitat covariate not being easily manipulated by project actions. Additionally, some of the CHaMP covariates used in previous models are difficult to replicate or measure using streamlined fish habitat protocols (DASH - Carmichael et al. 2019). To increase the utility of the QRF model for project monitoring/design and future data collection efforts, we are exploring alternative covariates to include in the model that: 1) maintain high predictive power, 2) are informative for restoration efforts and monitoring, 3) can be calculated by DASH, and 4) are not highly correlated with other covariates in the model. Using this criterion, we hope to develop a modified QRF model that is more informative for restoration design and monitoring, can easily be measured using newly developed stream habitat protocols, and maintains the same level of predictive power as the original QRF. These efforts will occur in April and May 2022.

### Improvements to Extrapolation Models

The QRF models are currently being fit using a relic ISEMP-CHaMP dataset. As such, these models can only predict capacity at locations where either CHaMP habitat data are available or CHaMP metrics can be replicated using a different protocol e.g., DASH. To make predictions of capacity across larger scales where CHaMP or DASH data are absent, we rely on an extrapolation model. Previously, extrapolation models were dependent on sample points and their associated attributes to estimate capacity (Larsen et al. 2016). Unfortunately, a sample point does not represent a stretch of river, rather a single location (lat/lon coordinates) that is meant to be representative of about a kilometer of stream. Recently, a group of NOAA researcher developed a dataset that classifies the entire Columbia River system into 200m reaches with various attributed assigned to each river reach. This new dataset covers the same area as the master sample point dataset but provides better interpretation and visualization. In 2021, the QRF model extrapolations began utilizing this spatially continuous, linear network; however, the current extrapolations will need to be updated after improved QRF covariates sets are chosen in spring 2022. We anticipate this to occur in summer 2022.

Map

Description automatically generated

**Figure 1.** *Plots of Chinook parr capacity in the Lemhi, using the master sample points method (A) and the 200m reach method (B).*

DASH Data Pipeline Development

A “data pipeline” to process stream habitat data available from the DASH survey protocol is under development and available at <https://github.com/mackerman44/DASH>. Significant progress was made in early (January – April) 2022 and a workflow was developed to efficiently import on-the-ground field data, QC the data, and append it to spatial centerlines generated using drone orthomosaics. Next steps include creating documentation (user manuals or vignettes) to describe the workflow and consider adding image analysis capabilities that populate habitat metrics to the dataset from drone-generated ortho-mosaics. We anticipate image analysis to begin in summer 2022.

Process DASH Data Available 2018-2021

In April 2022, we compiled all the on-the-ground DASH survey habitat data that has been collected to-date. The dataset spans four years (2018 – 2021) and includes 5,568 channel units and 710 habitat reaches (a group of consecutive channel units spanning approximately 200m in length). Soon, habitat metrics (covariates) will be calculated using the data so predictions of habitat capacity can be made using QRF models in late-spring/early-summer of 2022. The datasets, including QRF capacity predictions, will be used to support upcoming North Fork Salmon and Reference Reach assessments (described below) as well as ongoing efforts in the Upper Salmon Subbasin where DASH data are available.

Pre- and Post-Project Capacity Uplift Monitoring

One desired outcome of QRF and DASH is to be able to estimate capacity uplift at projects where pre- and post-project DASH data can be collected. As part of a pilot effort, we are collecting pre-project DASH data at ~ 5-mile reach near the confluence of Hayden Creek and Lemhi River in summer, 2022. Monitoring at this Lemhi/Hayden project will be used to understand expected capacity uplift from restoration projects and will help inform ongoing evaluations to understand the amount of restoration needed to achieve restoration objectives i.e., eliminate capacity deficits. A similar approach could be used to estimate capacity uplift at the Bird Track Springs – Longley complex, but would require including an adjacent “control” reach as pre-project DASH data are not available.

Image & Orthomosaic Analysis

Most habitat metrics being generated from DASH surveys are reliant on on-the-ground field data. However, it is our desire to be able to calculate some of these metrics using image analysis of drone-generated ortho-mosaics (e.g. volume large woody debris) to make field protocol more efficient. Additionally, we are also interested in the possibility of calculating new metrics using the drone-generated ortho-mosaics. Meetings will be scheduled in late spring or early summer with MHE, CRITFC, and Biomark to discuss options and ideas further.

Pair Fish Abundance/Density Surveys with DASH

The QRF models are currently fit using >5 year old ISEMP/CHaMP fish-habitat data. Habitat conditions at these ISEMP/CHaMP sites and throughout the CRB are dynamic, and the fish-habitat relationships developed using this relic data may, or may not, remain static. Our desire is to begin pairing fish surveys at locations where DASH habitat data are being collected to continue to improve the fish-habitat relationships that to be used for future applications of QRF. Currently, no funding is in place for fish data collection. However, we are looking for opportunities to pair DASH with channel-unit scale snorkel surveys. Snorkel protocols available from CRITFC (and associated snorkel calibration models to estimate juvenile fish abundance) could be adopted to pair with DASH.

Develop QRF Viewer Application

Once we complete QRF covariate modifications and extrapolation models in spring and early-summer 2022, we intend to estimate carrying capacity for all the watersheds in the Upper Salmon Subbasin, as well as capacity estimates across the spatially continuous, 200m linear network developed by NOAA. Once that is complete, we will develop a user-friendly application (e.g., R Shiny App) where QRF extrapolations can be shared online to stakeholders, collaborators, and practitioners in the Upper Salmon. Development of the web application is anticipated to begin in late 2022 and into 2023.

Applications

QRF and extrapolation models will be applied in three upcoming projects:

Upper Walla Walla Watershed Assessment

The QRF framework and extrapolation outputs are currently being used in support of an assessment in the Upper Walla Walla watershed. QRF extrapolation results, along with a Generalized Capacity Model (Appendix C in Idaho OSC Team 2019) are being used to estimate required and available capacity for Chinook salmon and steelhead in the watershed; capacity deficits are then being estimated as available minus required capacity to identify limitations. A first draft and results have been completed and a final assessment is anticipated in spring/summer 2023.

North Fork Salmon River Assessment

A North Fork Salmon River (NFSR) Assessment will occur in 2022 to help identify, prioritize, and develop high-quality rehabilitation projects in the watershed that will improve instream habitat for ESA listed salmonids. Like the Upper Walla Walla assessment, this project will identify habitat limitations and guide biologically sound and geomorphically appropriate restoration strategies to improve spawning and rearing of Chinook salmon and steelhead along ~85km of NFSR mainstem and tributary habitat. The NFSR Assessment will incorporate capacity estimates across the watershed from both QRF extrapolations, as well as predictions from DASH surveys completed in the NFSR in 2019 and 2021.

Reference Reach Assessment

A “reference reach” assessment will be done in 2022 to inform target conditions for project design in the Upper Salmon Subbasin. Reference reaches were sampled during summer 2021 and included Elk Creek, Bear Valley Creek, Marsh Creek, and Knapp Creek in the upper Middle Fork Salmon River plus Wildhorse Creek, Star Hope Creek, and East Fork Big Lost River in the upper Big Lost watershed.

Due to the nature of the relic fish-habitat datasets (i.e., ISEMP-CHaMP) and the spatially random study design that was implemented, much of that habitat data and paired fish information does not possess high-quality habitat characteristics. Although the relic ISEMP-CHaMP dataset and study design to provide the groundwork and distribution of characteristics required to develop robust fish-habitat models, it does not contain the necessary distribution of high-quality habitat and resulting reference information that is necessary to describe target conditions throughout the basin. DASH habitat data from the reference reaches will be compared with the CHaMP dataset as well as DASH habitat collected elsewhere outside of the reference reaches to 1) indicate how much “high-quality” habitat CHaMP might have sampled (or not) and 2) to provide a comparison of the relative quality of habitat where DASH habitat data are previously available (Lemhi River, Pashimeroi River, Upper Salmon River [Sawtooth Valley]) to reference reaches. Reference reach DASH data will further be used to describe target habitat conditions for restoration design in the Upper Salmon Subbasin. The reference reach information and dataset will be used as parallel supporting information to previous analyses and literature reviews. Assessing a dataset of reference habitat will ensure that target conditions are most appropriately described and analyzed within the current framework of the reach assessments.

We plan to leverage reference reach data to provide upper limits for quality habitat characteristics for a multitude of models, but also utilize the reference information for comparisons across the Upper Salmon Subbasin. Properly characterizing the highest quality accessible habitat will allow for further understanding and constraint of models as well as provide benchmarks for habitat characteristics and capacity goals that can reasonable be achieved post-restoration.

References

Carmichael, R.A., M.W. Ackerman, K. See, B. Lott, T. Mackey, and C. Beasley. 2019. Drone Assisted Stream Habitat (DASH) Protocol, DRAFT. May 2019.

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

See, K.E., M.W. Ackerman, R.A. Carmichael, S.L. Hoffmann and C. Beasley. 2021. Estimating Carrying Capacity for Juvenile Salmon Using Quantile Random Forest Models. Ecosphere 12(3):e03404.