Evaluation of potential habitat breaks (PHBs) for use in delineating the upstream extent of fish habitat in forested landscapes in Washington State

Study Design

Authors: Jason Walter, John Heimburg, Douglas Martin, Chris Mendoza, Cody Thomas, Don Nauer, Jenelle Black

Cooperative Monitoring
Evaluation & Research

CMER 2023.05.23

This Page intentionally left blank

Evaluation of potential habitat breaks (PHBs) for use in delineating the upstream extent of fish habitat in forested landscapes in Washington State

Prepared by
Jason Walter, John Heimburg, Douglas Martin, Chris Mendoza, Cody Thomas, Don Nauer, Jenelle Black

Project Manager
Anna Toledo

Washington State Forest Practices Adaptive Management Program

The Washington Forest Practices Board (FPB) has adopted an adaptive management program in concurrence with the Forests and Fish Report (FFR) and subsequent legislation. The purpose of this program is to:

> Provide science-based recommendations and technical information to assist the board in determining if and when it is necessary or advisable to adjust rules and guidance for aquatic resources to achieve resource goals and objectives. (Forest Practices Rules, WAC 222-12-045)

To provide the science needed to support adaptive management, the FPB made the Cooperative Monitoring, Evaluation and Research Committee (CMER) a participant in the program. The FPB empowered CMER to conduct research, effectiveness monitoring, and validation monitoring in accordance with guidelines recommended in the FFR.

Document Type and Disclaimer

This study design was prepared for the Cooperative Monitoring, Evaluation and Research Committee (CMER) to support design and implementation of Forest and Fish Adaptive Management research and monitoring studies. The project is part of the Stream Typing Program and was conducted under the oversight of the Instream Scientific Advisory Group (ISAG).

This study design was reviewed and approved by CMER and was reviewed and approved through the Adaptive Management Program's independent scientific peer review process (ISPR)[i]. As a CMER document, CMER is in consensus on the scientific merit, conclusions, and interpretations of the document. Recommendations contained within this document are those of the authors and may not reflect the views of any or all CMER members.

The Forest Practices Board, CMER, and all the participants in the Forest Practices Adaptive Management Program hereby expressly disclaim all warranties of accuracy or fitness for any use of this document other than for the Adaptive Management Program. Reliance on the contents of this document by any persons or entities outside of the Adaptive Management Program established by WAC 222-12-045 is solely at the risk of the user.

Proprietary Statement

This work was developed with public funding, as such it is within the public use domain. However, the concept of this work originated with the Washington State Forest Practices Adaptive Management Program and the authors. As a public resource document, this work should be given proper attribution and be properly cited.

---

[i] a process used to "determine if the scientific studies that address program issues are scientifically sound and technically reliable; and provide advice on the scientific basis or reliability of CMER's reports" (Board Manual Section 22.4.1).

Full Reference

Walter, Jason, John Heimburg, Douglas Martin, Chris Mendoza, Cody Thomas, Don Nauer, Jenelle Black. 2023. Evaluation of potential habitat breaks (PHBs) for use in delineating the upstream extent of fish habitat in forested landscapes in Washington State. Cooperative Monitoring Evaluation and Research. CMER 05-23-2023. Washington Department of Natural Resources, Olympia, WA.

Evaluation of potential habitat breaks (PHBs) for use in delineating the upstream extent of fish habitat in forested landscapes in Washington State

Study Design prepared for the Washington Forest Practices Board

(Revised from PHB Science Panel Draft 2019)

April 18, 2023

Submitted by:

Instream Science Advisory Group (ISAG) Project Team

Jason Walter (Weyerhaeuser), ISAG Chair
John Heimburg (WDFW)
Douglas Martin (WFPA)
Chris Mendoza (Conservation)
Cody Thomas (Eastside Tribes)
Don Nauer (WDFW)
Jenelle Black (CMER staff)
Anna Toledo (DNR Project Manager)

Preface

In 2018, the Potential Habitat Break (PHB) Science Panel convened by The Forest Practices Board (FPB or Board) developed a study design (PHB Science Panel 2019) to validate potential habitat breaks (PHBs). The study design (PHB Science Panel 2019) was reviewed and approved by Independent Scientific Peer Review (ISPR), however there were varying levels of comments and criticisms from all caucuses participating in the Forest Practices Adaptive Management Program (AMP) to particular aspects of the study design and the review process. In 2019, the Forest Practices Board remanded the project to the Department of Natural Resources' adaptive management science program, tasking the Cooperative Monitoring, Evaluation and Research (CMER) committee with revising the study design following CMER's protocols and standards (referenced in Forest Practices Board Manual Section 22). CMER assigned the study design revision to the Instream Science Advisory Group (ISAG). This revised study design was developed by a project team formed within ISAG. This document was adapted from the PHB Science Panel draft (2019) and includes substantial excerpts from this previous version.

Summary

The upstream extent of both fish distribution and fish habitat in forested watersheds is influenced by many factors including channel gradient, channel size, channel condition, nutrients, flow, barriers to migration, history of anthropogenic and natural disturbance, and/or fish abundance. Potential habitat breaks (PHBs) are defined as permanent, distinct, and measurable in-channel physical characteristics that limit the upstream extent of fish distributions. PHBs would be used in a Fish Habitat Assessment Methodology (FHAM), currently under development. The Washington Forest Practices Board has proposed three sets of criteria to be considered in determining PHBs between fish (Type F) and non-fish bearing (Type N) waters across the state. These criteria are based upon data that can be collected during a single Washington Department of Natural Resources (DNR) protocol electrofishing survey and include channel gradient, bankfull width, and both vertical and non-vertical non-deformable natural obstacles to upstream migration. Detailed information is needed on the uppermost fish location and associated habitat in small streams across Washington State to evaluate which physical criteria best define the end of fish (EOF) habitat (the uppermost stream segments that are actually or potentially could be inhabited by fish at any time of the year based on habitat accessibility and suitability). Some data on habitat conditions at uppermost detected fish

53   locations are available (e.g., from existing water type modification forms [WTMFs] submitted

54   to DNR), but these data were found to be insufficient to determine PHBs that defined

55   uppermost detected fish locations and associated habitat.

56   The purpose of this study is to develop criteria to characterize PHBs as accurately as possible

57   and to evaluate the utility and accuracy of PHB criteria selected by the Board for use in the Fish

58   Habitat Assessment methodology (FHAM) as part of a water typing rule. The study is designed

59   to assess combinations of gradient, channel width, barriers to migration, and other physical

60   habitat and geomorphic conditions associated with uppermost detected fish locations. Study

61   findings will 1) inform which Board-identified PHB criteria most accurately identify the

62   upstream extent of fish habitat in an objective and repeatable manner as applied in the FHAM;

63   2) evaluate whether an alternative set or combination of empirically derived criteria more

64   accurately achieves this goal; and 3) provide insight into how uppermost detected fish points

65   and associated stream characteristics may vary across geography, seasons, and years.

66   The study will be conducted across two sampling seasons (spring and fall/winter) in each of

67   three years at 350 sites statewide; 160 in Eastern and 190 in Western Washington. Uppermost

68   detected fish locations will be determined during each season at each site following modified

69   DNR protocols for electrofishing surveys. Once the uppermost fish is located during each

70   sampling event, the uppermost detected fish location will be flagged, GPS coordinates will be

71   recorded, and a longitudinal profile habitat survey will be conducted to characterize habitat

72   and geomorphic conditions 660 ft (200 meters) downstream and 660 ft upstream of the

73   uppermost detected fish location. To evaluate seasonal changes in the location of the

74   uppermost detected fish, the sites that can be accessed in the fall/winter season will be visited

75   with an augmented serially alternating panel design. One quarter of the sites will be assigned

76   to the fixed panel and will be surveyed every fall/winter, and the remainder will be allocated

77   to three alternating panels. One of the three alternating panels will be surveyed each year, and

78   the sample is augmented by the fixed panel of sites such that every accessible site will be

79   surveyed at least once during the fall/winter. If an uppermost detected fish location changes

80   during any subsequent survey, additional longitudinal profile survey data will be collected to

81   ensure that there are channel data 660 ft above and 660 ft below uppermost detected fish

82   locations for all seasons and years. Data will be analyzed using a suite of statistical methods

83   (e.g., random forest, classification, and regression) to determine the combinations of gradient,

84   channel width, and other geomorphic features associated with the uppermost detected fish

85   locations across all seasons and years at each site, which will define PHBs and EOF habitat, and

86   whether these vary across Eastern and Western Washington. Finally, a suite of PHB

87   performance analyses will be used to evaluate the effectiveness of Board-proposed or other

88   empirically derived PHB criteria resulting from this study in determining the regulatory break

89   between fish (Type F) and non-fish bearing (Type N) waters.

Table of Contents

List of Acronyms

AMP	Adaptive Management Program
BFW	Bankfull Width
CMER	Cooperative Monitoring, Evaluation & Research Committee
DNR	Washington State Department of Natural Resources
DPC	Default Physical Characteristics
eDNA	Environmental DNA
EOF	End of Fish (Last detected fish following a Protocol Survey)
EOFH	End of Fish Habitat
F/N Break	Regulatory break between fish and non-fish bearing waters
FHAM	Fish Habitat Assessment Method
FPB, or "Board"	Washington State Forest Practices Board
GIS	Geographic Information System
HCP	Habitat Conservation Plan
ISPR	Independent Scientific Peer Review
NVO	Non-vertical obstacle
PHB	Potential Habitat Break(s)
TFW	Timber, Fish & Wildlife
Type F	Fish Bearing Streams
Type N	Non-Fish Bearing Streams
WTM	Water Type Modification
WTMF	Water Type Modification Form

125

126 ## Introduction

127 In Washington State, forest practices are regulated by the Forest Practices Act (RCW 76.09)

128 established by the legislature, with rules established by the Washington Forest Practices Board

129 (Board). The goals of the rules include protecting public resources (water quality, fish, and

130 wildlife) and maintaining an economically viable timber industry. Rules pertaining to aquatic

131 and riparian habitats are specifically included in the Forest Practices Habitat Conservation Plan

132 (HCP), which provides coverage for approximately 9.3 million acres of forestland in Washington

133 (6.1 million acres west of the Cascade Crest and 3.2 million acres in eastern Washington).

134 Specific timber harvest and road prescriptions (rules) are applied to waters used by fish to

135 protect fish and their habitats.

136 The Board is responsible for rulemaking and overseeing the implementation of forest practice

137 rules. The evaluation of the effectiveness of these rules is administered by the Adaptive

138 Management Program of the Washington Department of Natural Resources. Water typing is an

139 important part of applying contemporary forest practice rules since prescriptions in riparian

140 areas are based in part on whether streams are or potentially could be used by fish. Streams

141 identified as having fish habitat are classified as Type F waters, defined in the water typing rule

142 (WAC 222-16-030), and have specific riparian buffer prescriptions and fish passage

143 requirements. Fish habitat is defined in WAC 222-16-010 as "…habitat, which is used by fish at

144 any life stage at any time of the year including potential habitat likely to be used by fish, which

145 could be recovered by restoration or management and includes off-channel habitat." Currently,

146 an interim rule allows for the delineation of Type F waters through the use of either default

147 physical characteristics (WAC 222-16-031) or a protocol electrofishing survey. DNR provides a

148 map showing stream segments of modeled fish habitat. The Forest Practice Rules require

149 forest landowners to verify, in the field, the type of any regulated waters identified within

150 proposed harvest areas prior to submitting a forest practices application/notification.

151 Landowners may use the default physical criteria or the results from protocol survey

152 electrofishing to identify the regulatory Type F/N break. Landowners are encouraged to submit

153 a Water Type Modification Form (WTMF) to the DNR to make permanent changes to the water

154   type maps. Thousands of WTMFs have been submitted to DNR to modify water types and

155   modify the location of the break between Type F and Type N waters.

156   The Board is currently in the process of establishing a permanent water typing rule. Ultimately,

157   the rule must be implementable, repeatable, and enforceable by practitioners and regulators

158   involved in the water typing system. An important part of the permanent rule will be guidance

159   on a specific protocol to determine the regulatory break between Type F and Type N waters.

160   The Board is considering the use of a fish habitat assessment method that incorporates known

161   fish use with PHBs to identify the upstream extent of fish habitat. The Board recommended

162   that PHBs be based on permanent physical channel characteristics such as gradient, stream

163   size, and/or the presence of non-deformable vertical and non-vertical natural obstacles as

164   potential barriers to upstream fish movement (WA Forest Practices Board 2017).

Study Purpose

166   The purpose of this study is to develop criteria for accurately identifying PHBs and to evaluate

167   the utility of PHB criteria for use in the Fish Habitat Assessment Methodology (FHAM) as part

168   of a water typing rule. The study is designed to assess which combinations of gradient, channel

169   width, barriers to migration, and other physical habitat and geomorphic conditions are

170   associated with uppermost detected fish locations. This will 1) inform which Board-identified

171   PHB criteria most accurately identify the upstream extent of fish habitat in an objective and

172   repeatable manner as applied in the FHAM and 2) evaluate whether an alternative set or

173   combination of empirically derived criteria more accurately achieves this goal (CMER 2020).

174   Additionally, this study is intended to provide insight into how uppermost detected fish points,

175   upstream extent of fish habitat based on FHAM, and PHBs proposed by the Washington Forest

176   Practice Board may vary across geography, seasons, and years. The Board is expected to use

177   the study findings to inform which PHB criteria to use in FHAM.

178   It is important to note that this study is not intended to evaluate the current water typing

179   system or the FHAM; nor is it intended to describe how the regulatory Type F/N break should

180   be determined. PHBs are defined in FHAM as permanent, distinct, and measurable changes to

181   in-channel physical characteristics. Other factors such as temperature, flow, water quality,

182   population dynamics, anthropogenic and natural disturbance, and biological interactions are

183   important covariates that might influence the distribution of fishes but do not affect PHBs.

184   Therefore, they are not being evaluated in this study.

185   ## Project Research Questions

186   The following project-specific research questions were developed to address key uncertainties

187   and provide information needed to evaluate the performance of the PHB criteria provided by

188   the Washington Forest Practices Board and empirically derived alternatives. They also address

189   certain aspects of the CMER Workplan Rule Group critical questions listed in Appendix A.

190   **UPSTREAM-MOST FISH LOCATIONS**

191   1. **How do the locations of the last (uppermost) detected fish vary interannually?**

192   2. **How do the locations of the last (uppermost) detected fish vary seasonally?**

193   3. **How do the locations of last (uppermost) detected fish vary geographically across the**
194   **state of Washington?**

195   **HABITAT ASSOCIATED WITH UPSTREAM-MOST FISH LOCATIONS**

196   4. **How do the physical channel and basin characteristics (e.g., bankfull width; average**
197   **gradient, basin size) associated with the identified end (upstream extent) of fish**
198   **habitat vary geographically across the state of Washington?**

199   5. **Where the location of the last (uppermost) detected fish changes (seasonally or**
200   **interannually), how does that influence which PHB would be associated with the F/N**
201   **break and how frequently does that occur?**

202   6. **How do the physical channel features at the locations initially identified as PHBs**
203   **change over the course of the study?**

204   7. **How often do similar features appear to limit upstream fish distributions in some**
205   **contexts but not others (e.g., further into the headwaters vs. downstream; different**
206   **flow levels)?**

207   **PHB PERFORMANCE ANALYSES**

208   8. **Which combinations of physical channel features and basin characteristics (for**
209   **example, gradient, channel width, barriers to migration) best identify the end of fish**
210   **habitat relative to the location of the last (uppermost) detected fish?**

211   9. **Can protocols used to describe PHBs be consistently applied among survey crews and**
212   **be expected to provide similar results in practice?**

213   10. **How well do the PHB criteria provided by the Washington Forest Practices Board**
214   **accurately identify the EOF habitat when applied in the Fish Habitat Assessment**
215   **Methodology (FHAM)?**

216

Approach

We will use data from electrofishing and physical habitat channel surveys in a spatially balanced sample of 350 streams across Eastern and Western Washington to address the project research questions above and to evaluate proposed criteria to be used as potential habitat breaks in the FHAM. We will conduct multiple surveys over a three-year period to document seasonal and interannual changes in fish distribution and to maximize the likelihood of identifying the upper extent of fish use in each stream. This will allow us to address questions about seasonal and interannual changes in uppermost fish location, and to evaluate proposed criteria to be used as potential habitat breaks in the FHAM. We will identify PHBs associated with the upper extent of fish habitat using a suite of physical channel attributes and basin characteristics. Three sets of PHB classification criteria proposed by the Board will be assessed and an independent set of criteria will be developed with statistical tools for classification.

Background (*adapted from PHB Science Panel 2019*)

Over the past 20 years, protocol electrofishing surveys have been conducted under WAC 222-16-031 with guidance provided by Board Manual Section 13 to determine the upper extent of Type F waters. These surveys often incorporate additional stream length upstream of the uppermost detected fish to include habitat "likely to be used by fish" (defined in WAC 222-16-010). Throughout Washington, the uppermost fish[1] detected during protocol electrofishing surveys is most often a salmonid, and in around 90% of cases the uppermost fish is a cutthroat trout (*Oncorhynchus clarki*) (D. Collins, Washington Department of Natural Resources, unpublished data; Fransen et al. 2006). Other salmonid species that have been documented at uppermost fish locations on water type modification forms across Washington include rainbow trout (*O. mykiss*), brook trout (*Salvelinus fontinalis* - an introduced non-native that has become established in many Washington streams), and (rarely) bull trout (*S. confluentus*). In headwater reaches that are accessible to anadromous fishes, coho salmon (*O. kisutch*) juveniles have been reported on occasion as the uppermost fish. Of the non-salmonid species documented at uppermost fish sites on WTMFs in western Washington, sculpins (*Cottus* spp.) were most

---

[1] WAC 222-16-010: "Fish" means for purposes of these rules, species of the vertebrate taxonomic groups of Cephalospidomorphi [lampreys] and Osteichthyes [bony fish].

244   prevalent, followed by brook lamprey (*Lampetra* spp.), and less commonly dace (*Rhinichthys*

245   spp.), three-spine stickleback (*Gasterosteus aculeatus*), and Olympic mudminnow (*Novumbra*

246   *hubbsi*). The only non- salmonid uppermost fish species recorded in east-side Washington

247   streams were sculpins.

248   Many factors can limit the distribution of fishes including barriers to migration, stream gradient,

249   flow, and channel size. Understanding the current science on how these factors influence fish

250   distribution is important when discussing how they can be used to most accurately define the

251   upstream limits of fish habitat in forested streams of Washington State.

Obstacles to Migration

253   Natural stream habitat breaks that might obstruct or completely block upstream fish

254   movement to apparently suitable habitat include: vertical drops, cascades, bedrock sheets,
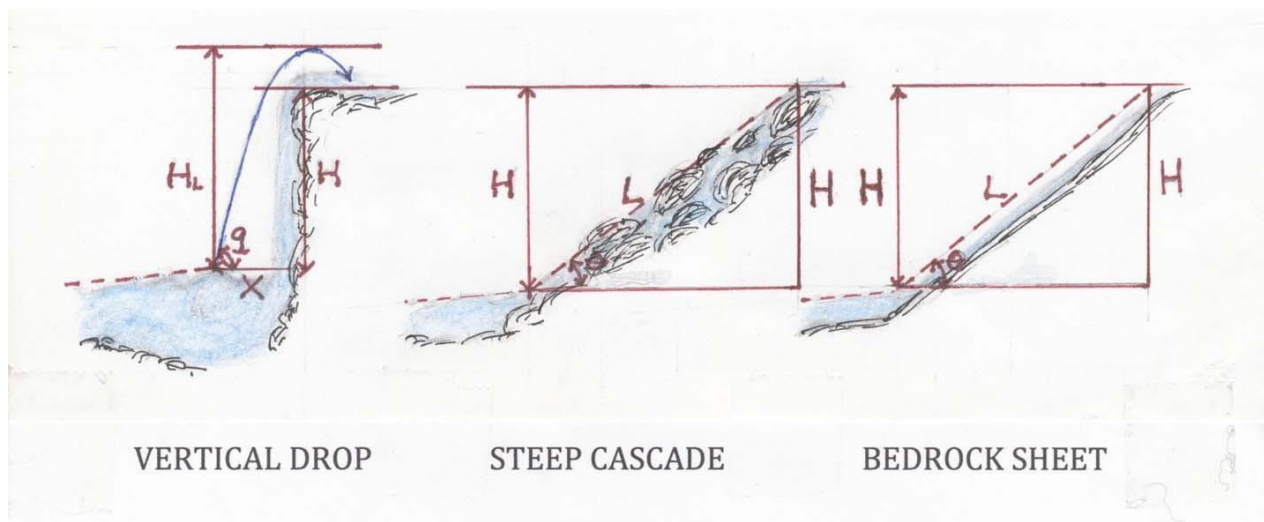
255   and/or chutes (Hawkins et al. 1993; Figure 1).

256

257   **Figure 1. Three types of features that could pose obstacles or barriers to upstream movement of**
258   **headwater fishes. (PHB Science Panel 2019)**

259

260   The ability of fishes to pass such obstacles is associated with the interactions between their

261   swimming and leaping abilities, environmental factors such as flow and temperature and the

262   dimensions of the obstacles. The swimming ability of fishes is typically described in terms of

263   cruising, prolonged, and burst speeds, which are measured in units of body lengths per second

264   (Watts 1974; Beamish 1978; Webb 1984; Bell 1991; Hammer 1995). Body form also affects
265   swimming ability, with more fusiform body shapes being advantageous for stronger burst
266   speeds in fishes such as cutthroat and rainbow trout (Bisson et al. 1988; Hawkins and Quinn
267   1996) in comparison to some other fishes, such as sculpin (*Cottus* spp.), commonly found at
268   EOF locations. Cruising speed is the speed a fish can sustain essentially indefinitely without
269   fatigue or stress, usually 2–4 body lengths per second. Cruising speed is used during normal
270   migration or movements through gentle currents or low gradient reaches. Prolonged speed
271   (also called sustained speed) is the speed a fish can maintain for a period of several minutes to
272   less than an hour before fatiguing, typically 4–7 body lengths per second. Prolonged swimming
273   speed is used when a fish is confronted with more robust currents or moderate gradients. Burst
274   speed is the speed a fish can maintain for only a few seconds without fatigue, typically 8–12
275   body lengths per second. Fish typically accelerate to burst speed when necessary to ascend
276   short, swift, steep sections of streams; to leap obstacles; and/or to avoid predators.

277   When leaping obstacles, fish come out of the water at burst velocity and move in a parabolic
278   trajectory (Powers and Orsborn 1985). Relationships for the height attained in the leap, and
279   the horizontal distance traversed to the point of maximum height are often used to assess
280   barriers. Depth at the point of takeoff is important for enabling fish to reach burst velocity.
281   Stuart (1962) found water depth of at least 1.25 times the height of an obstacle to be required
282   for successful upstream barrier passage. More recently, however, Kondratieff and Myrick
283   (2006) reported that small brook trout (size range 100-150 mm) could jump vertical waterfalls
284   as high as 4.7 times their body length from plunge pools only 0.78 times the obstacle height,
285   and larger brook trout (size ranges 150-200 mm and 200 mm+) could jump waterfalls with
286   heights 3 to 4 times their body length if the plunge pool depth was at least 0.54 times the
287   obstacle height.

288   To successfully ascend 4.7 body lengths in height, a back-calculation from the Powers and
289   Orsborn (1985) trajectory equation yields a burst speed of 22 body lengths per second (11.7
290   feet per second) for the 100-150 mm body-length brook trout reported by Kondratieff and
291   Myrick (2006). If it is assumed that other salmonids (e.g., cutthroat, rainbow trout or coho
292   salmon) could perform as well as brook trout in the size range typically found at uppermost fish

293   locations in Washington (Sedell et al. 1982; Fransen et al. 1998; Liquori 2000; Latterell et al.

294   2003; Peterson et al. 2013), then a burst speed of 22 body lengths per second (11.7 feet per

295   second) would allow the largest fishes in the size range typical of headwater-dwelling salmonids

296   (6.3 in, 160 mm) to leap a vertical obstacle 2.6 feet high, whereas a vertical obstacle of 3 feet

297   high would be impassable.

298   When leaping is not required, fishes may ascend steep cascades and other high-velocity habitat

299   units (Hawkins et al. 1993) by seeking pockets of slow water interspersed in areas with turbulent

300   flow (e.g., boundary layers near rocks or logs). For example, Bisson et al. (1988) reported the

301   average water velocity was only 24.8 ± 3.2 cm/s (0.8 ft/s) in shallow (10.0 ± 1.4 cm; 4 inches)

302   cascade habitat units of small western Washington streams. It is possible that fish may ascend

303   streams during periods of elevated flow by moving along the channel margins where water

304   velocities are reduced relative to mid-stream and small falls and boulder cascades are partially

305   or completely submerged.

306   Although studies examining fish migration through potential non-vertical obstacles are rare,

307   some studies have examined brook trout movement through steep cascades and reported fish

308   ascending cascades of more than 20% gradient (Moore et al. 1985; Adams et al. 2000; Björkelid

309   2005). For example, Adams et al. (2000) reported that adult brook trout ascended cascades

310   with slopes of 13% that extended for more than 67 m, and 22% for more than 14 m as well as

311   adult brook trout ascending a waterfall 1.2m high. Similarly, Björkelid (2005) reported invasive

312   brook trout colonizing 18 headwater streams in Sweden and found they ascended stream

313   segments with slopes of 22% (measured with a clinometer) and 31% (measured with GIS).

Gradient

315   In Washington streams, fish (not necessarily the uppermost fish) have been observed in

316   headwater segments with overall slopes as steep as 31% (S. Conroy, formerly Washington Trout

317   [now Wild Fish Conservancy], unpublished data), 35% (J. Silver, Hoh Indian Tribe, unpublished

318   data; D. Collins, Washington Department of Natural Resources, unpublished data), and in reach

319   gradients of 25% and steeper in Oregon streams (C. Andrus, Oregon Department of Forestry,

320   unpublished data; Connolly and Hall 1999). This range of channel steepness is consistent with

321   other observations in western North America (e.g., Leathe 1985; Fausch 1989; Ziller 1992;

322   Kruse et al. 1997; Watson and Hillman 1997; Dunham et al. 1999; Hastings et al. 2005; Bryant

323   et al. 2004, 2007) and Europe (Huet 1959). In the "trout zones" of European rivers

324   (headwaters), brown trout (*Salmo trutta*) predominate and reach gradients may be 10 to 25%

325   or steeper (Huet 1959; Watson 1993). In Washington, it is important to note that fish presence

326   in streams steeper than 15% accounted for only 10% of reported occurrences in forested

327   streams (Cole et al. 2006; J. T. Light, Plum Creek Timber, unpublished data). Kondolf et al. (1991)

328   reported that often the water surface slopes where fish occur in step-pool habitats have much

329   lower local gradients than the overall reach gradient and may range from only 0.4 to 4%, even

330   where overall reach gradients may be as high as 35% (Figure 2). These observations indicate

331   that in some cases fish habitat in headwater streams can extend into the types of steep step-

332   pool and cascade reaches described by Montgomery and Buffington (1993).
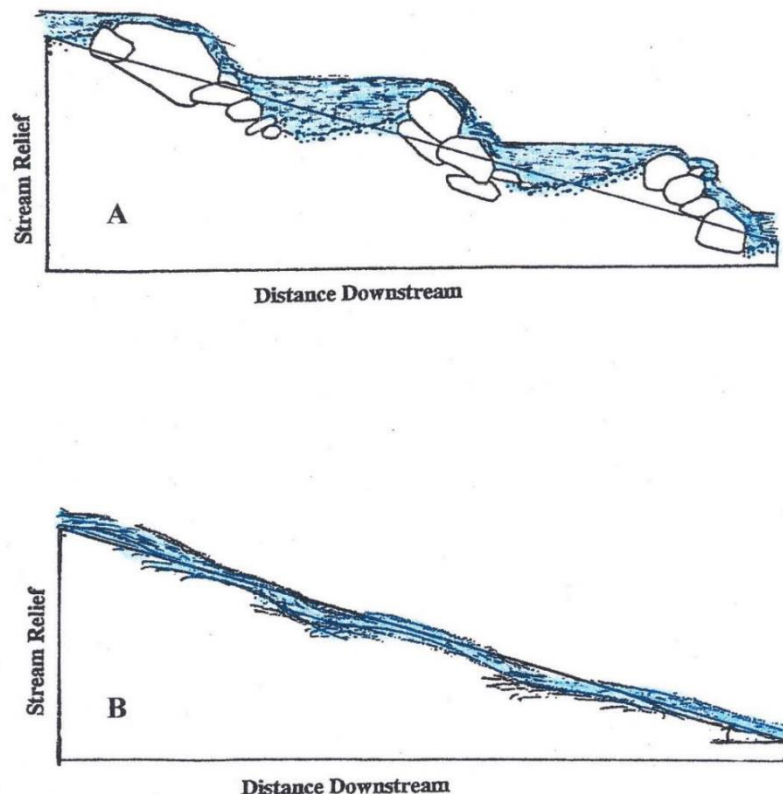

333

Figure 2. Two very different profiles of a headwater reach with the same overall reach gradient. Illustration (A) demonstrates how roughening elements create local gradients that are lower than the overall reach gradient, while reaches without such features (B) do not. (PHB Science Panel 2019)

337

Flow and Channel Size

Bankfull width (BFW) has been found to reflect the stage of discharge at which a stream does its habitat-building work (Andrews 1980; Leopold 1994; Rosgen 1996). Studies have shown that BFW is correlated with drainage area and varies with climate, geology, and topography of the basin (Castro and Jackson 2001). For example, Beechie and Imaki (2014) developed an equation for BFW for Columbia Basin streams based on annual precipitation and catchment (drainage) area. Although that equation was developed for larger streams, the PHB Science Panel (2019) tested it using empirical BFW data from multiple smaller streams across Washington State and found that it accurately predicted BFW in headwater streams. However, Castro and Jackson (2001) found that while BFW and drainage area relationships worked well in areas of similar lithology/geology and precipitation regimes to those for which they were developed, they were less useful in the Pacific coastal areas of western Washington where the geology and precipitation patterns are highly variable. Researchers continue to work on developing accurate and usable relationship models for highly variable headwater streams, which may become useful as more precise information and mapping of lithology, topography, and precipitation becomes available.

Because of the perceived relationship between channel width and discharge, BFW is often used as a surrogate for stream discharge (area, depth, and velocity), which is often important for determining the uppermost fish and upstream extent of fish habitat (Harvey 1993). Fransen et al. (1998) estimated mean annual flow rates at the upstream extent of fish distribution for 79 streams in the western Cascade foothills and Willapa Hills in Washington and found that 90% of these streams had mean annual flows of ~3.5 cfs or less at the upper boundary of fish presence; 80% had mean annual flows of ~2 cfs or less at the upper boundary; 65% had mean annual flows of ~1 cfs or less at the upper boundary; and approximately 25% of the sites had mean annual flows of 0.5 cfs or less at the upper boundary (Figure 3).
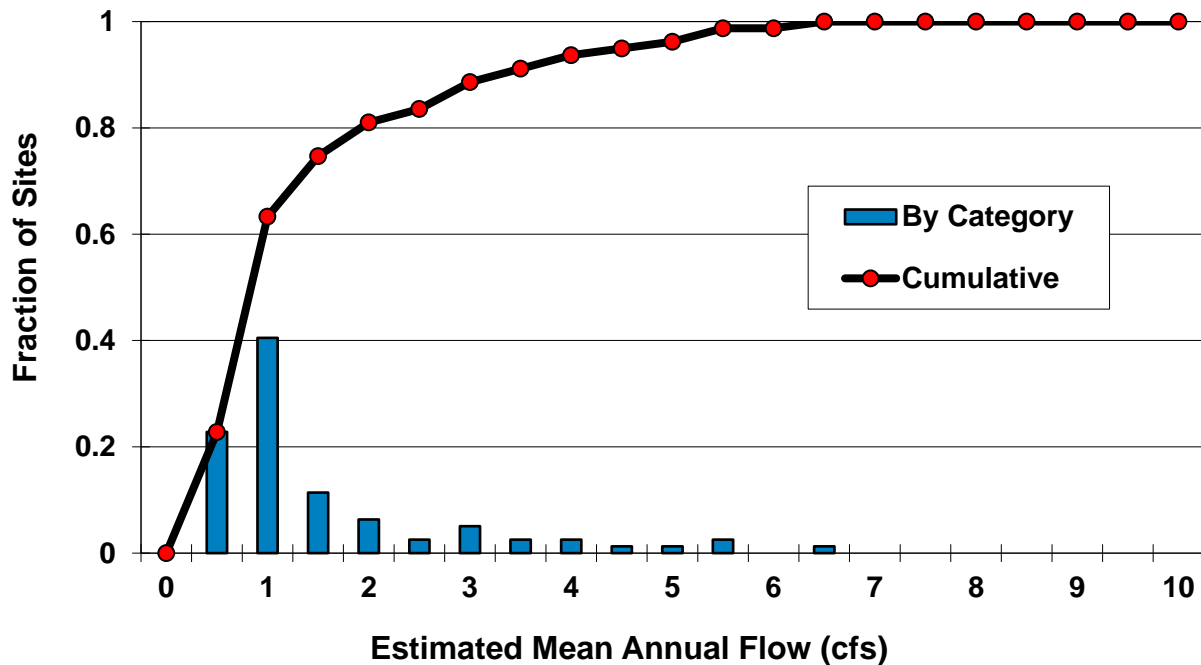
363

Figure 3. Estimated mean annual flows at uppermost fish locations in 79 streams in the Cascade foothills and Willapa Hills of western Washington (from Fransen et al. 1998)

366

Food Availability

368   Many studies, particularly in Pacific Northwest streams, have demonstrated strong food

369   limitations for fish inhabiting (using) small streams (Warren et al. 1964; Mason 1976; Naiman and

370   Sedell 1980; Bisson and Bilby 1998). Headwater segments are often characterized by closed

371   forest canopies, requiring primary energy sources from allochthonous inputs of coarse

372   particulate organic matter (CPOM). Shredder organisms occur in these reaches and feed on this

373   CPOM. These aquatic organisms, along with any terrestrial invertebrates that fall into the

374   stream, comprise the food base for trout and other predators (Vannote et al. 1980; Hawkins

375   and Sedell 1981; Triska et al. 1982; Wipfli 1997). The total production of macroinvertebrate

376   organisms is substantially lower in small headwater stream reaches than in the larger, lower-

377   gradient reaches further downstream (Northcote and Hartmann 1988; Haggerty et al. 2004).

378   As a result, resident fishes in small headwater stream reaches tend to be small bodied, which

379   limits their ability to negotiate obstacles to upstream movement and migration.

Fish Habitat Assessment Method (FHAM)

380

381   Water typing surveyors have used professional judgment to estimate "habitat likely to be used

382   by fish" when proposing regulatory fish bearing/non-fish bearing (F/N) water type breaks.

383   Stream segments that are accessible to fish and exhibit the same characteristics as those of

384   fish-bearing reaches are typically assumed to be fish habitat, whether or not fish are present

385   at the time of a survey. Surveyors have assessed barriers and measurable changes in stream size

386   and/or gradient to estimate the EOF habitat (Cupp 2002; Cole et al. 2006). Although research is

387   somewhat limited, the upstream extent of fish distribution in forest lands appears to be

388   strongly influenced by stream size, channel gradient, and access to suitable habitat (Fransen et

389   al. 2006; PHB Science Panel 2018). In response to these findings, the Board embraced the

390   concept of a Fish Habitat Assessment Methodology developed by a diverse group of AMP

391   technical stakeholders intended to be repeatable, implementable, and enforceable (WA Forest

392   Practices Board 2018; WA DNR 2019). The FHAM will utilize PHBs that reflect a measurable

393   change in the physical stream characteristics at or upstream from a detected fish point, above

394   which a protocol electrofishing survey would be undertaken (Figure 4). The first PHB located

395   at or upstream from the uppermost detected fish would serve as the end of fish habitat (F/N
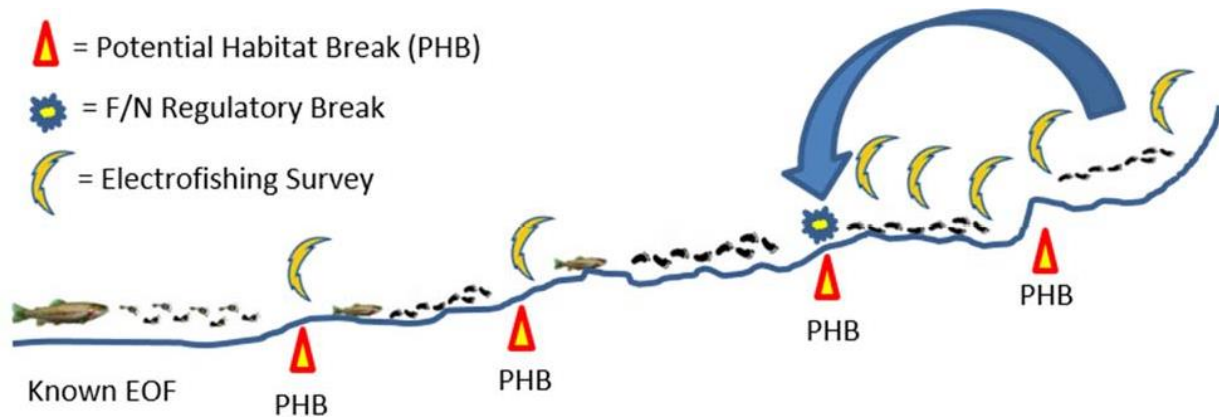
396   Break) when no fish are detected above this PHB.

397

398

399

400

Figure 4. Example of how the PHB criteria and Fish Habitat Assessment Methodology (FHAM) will be
applied in the field. The first step is to identify the uppermost detected fish location. Once the point
is identified, the survey team would begin to measure bankfull width, gradient, and barrier (obstacle)
criteria while moving upstream. Once a point in the stream meeting one of the PHB criterion
(gradient, barrier, change in channel width) is identified, the survey team would apply a fish survey
(e.g., electrofishing) upstream of the PHB to determine if fish are present upstream. If sampling yields
no fish ¼ mile upstream, then the F/N break would occur at the location where the survey
commenced (see arrow in the figure). If fish are encountered above any PHB, the process of
measuring and moving upstream would repeat until fish are not encountered. (PHB Science Panel
2019)

411

412   Per FHAM, PHBs are based on stream size, gradient, and access to fish habitat. The PHB Science

413   Panel reviewed the available science and data on PHBs and provided recommendations to the

414   Board for specific PHB criteria for eastern and western Washington (PHB Science Panel 2018).

415   The Panel considered a variety of potential PHB criteria, including the physical attributes of a

416   stream channel, water quality and quantity parameters, and other factors that might

417   contribute to measurable habitat breaks. These attributes were evaluated for the ability to

418   simply, objectively, accurately and repeatably measure them in the field, as well as the amount

419   and relevance of existing scientific literature pertaining to each. The Panel concluded that it

420   was possible to identify PHBs based on stream size, channel gradient, and natural non-

421   deformable obstacles. These three attributes satisfied the objectives of simplicity, objectivity,

422   accuracy, ease of measurement, and repeatability that can be consistently identified in the field

423   and can be incorporated into a practical survey protocol. The Board then selected three

424   combinations of stakeholder-proposed PHB criteria for these attributes at their 14 February

425   2018 meeting (WA FPB 2018) and instructed the PHB Science Panel to develop a field study to

426 evaluate the performance of these proposals (Table 1). It was important to the Board to
427 determine which of the proposed criteria most reliably identify PHBs in eastern and western
428 Washington. The Board also instructed the Science Panel to stratify sampling by ecoregion and
429 to examine crew variability in identifying PHBs, especially evaluating aspects of field
430 measurement practicality and repeatability (WA FPB August 2017). This study is designed to
431 evaluate which Board-identified PHB criteria most accurately identify the upstream extent of
432 fish habitat and to determine whether an alternative set or combination of empirically derived
433 criteria more accurately achieves this goal (CMER 2020).

434

435 **Table 1. Three combinations of barrier (obstacle), gradient, and width PHBs selected for evaluation**
436 **by the Washington Forest Practices Board during their February 2018 meeting. Descriptions are**
437 **abbreviated for readability from WA Forest Practices Board 2018. Criteria may be revised by the**
438 **Forest Practices Board before project is implemented.**

Type/	Description of Criteria
Criteria Set 1	
Width	2 ft BFW threshold (upstream BFW ≤2ft)
Gradient	Gradient increase of ≥10%
Vertical Obstacle	Obstacle height ≥3ft
Non-Vert Obstacle	Obstacle gradient ≥20%, AND elevation difference is ≥ 1x upstream BFW
Criteria Set 2	
Width	2 ft BFW threshold (upstream BFW ≤2ft)
Gradient	Gradient increase of ≥5%
Vertical Obstacle	Obstacle height ≥3ft AND ≥ 1x upstream BFW
Non-Vert Obstacle	Obstacle gradient ≥30%, AND elevation difference is > 2x upstream BFW
Criteria Set 3	
Width	20% BFW decrease (up- to downstream BFW ratio at tributary junctions ≤.8)
Gradient	Gradient increase of ≥5%
Vertical Obstacle	Obstacle height ≥3ft
Non-Vert Obstacle	Obstacle gradient ≥20%, AND elevation difference is ≥ upstream BFW

439   # Methods

440   ## Survey Design

441   **Sampling Frame and Study Sites**

442   Current F/N break points on the DNR Forest Practices water type map will serve as the sampling

443   frame for this study. The target population is defined as the set of all F/N break points on

444   streams on Forests and Fish (FFR) lands in Washington. A sampling frame that matches the

445   target population as closely as possible is needed for unbiased inference. Fish/non-fish stream

446   type break points extracted from the current DNR water type GIS map layer (DNR Forest

447   Practices   hydro,   watercourses   ("wchydro");   https://data-

448   wadnr.opendata.arcgis.com/datasets/wadnr::dnr-hydrography-watercourses-forest-

449   practices-regulation/about) represent an accessible source of possible study sites. Some of

450   these points are based on field surveys that were concurred (survey-based) through the WTM

451   review process while others are modeled points obtained from a logistic regression model that

452   predicts F/N points based on basin area, upstream and downstream gradients, elevation, and

453   precipitation (Conrad et al. 2003; Duke, 2005). The hybrid approach using both modeled and

454   concurred F/N break points as the sampling frame incorporates existing information while

455   allowing a broad scope of inference.

456   The study design will incorporate spatially balanced sampling. A spatially balanced sample

457   provides a sample that is geographically diverse, which generally means outcomes exhibit less

458   spatial correlation across units (Olsen et al. 2015). When outcomes are less correlated,

459   outcomes are more spatially independent of one another, thus increasing effective sample

460   sizes. Several types of spatially balanced sampling exist, including two-dimensional systematic

461   (or grid) samples, balanced acceptance sampling (BAS; Robertson et al. 2013), Halton iterative

462   partitioning (HIP; Robertson et al. 2018), and generalized random tessellation stratification

463   (GRTS; Stevens and Olsen 2003, 2004). Because the R package used to draw BAS & HIP samples

464   is currently not maintained on the CRAN server for R packages, the GRTS package maintained

465   by the EPA, spsurvey (Dumelle et al. 2022), will be used to draw the spatially balanced sample

466 to ensure best practices for security protocols and package functionality by using a currently-
467 maintained R package.

468 The spatially balanced sample of F/N points will be stratified by region (eastern or western
469 Washington)[2]. The western region of Washington consists of about one-third of the state's area
470 but has twice the stream density. Given the differences in stream distribution across the state
471 and the different sources of frame error in each region, east-west stratification will be applied
472 to ensure that spatial balance is maintained within each region.

473 Previous iterations of this study design incorporated ecoregion as a stratification variable.
474 Ecoregions reflect broad ecological patterns occurring on the landscape. In general, each
475 ecoregion has a distinctive composition and pattern of plant and animal species distribution.
476 Abiotic factors, such as climate, landform, soil, and hydrology are important in the
477 development of ecosystems and thus help define ecoregions. The Washington State Natural
478 Heritage Program modified ecoregions defined by the US EPA into Level III ecoregions specific
479 to   Washington,   each   of   which   is   described   at
480 http://www.landscope.org/washington/natural_geography/ecoregions (Figure 5). While it is
481 possible that there is something about ecoregions, particularly precipitation patterns, that
482 might cause differences in the barriers to fish movement, there is no strong reason to restrain
483 the analysis of results to that factor at the expense of our ability to investigate other, potentially
484 more important factors. We agree that there are likely to be differences among ecoregions in
485 where the fish and barriers to movement occur on the landscape but identifying those spatial
486 patterns of occurrence is not the purpose of the PHB study.

---

[2] We considered other finer scale stratification (e.g., geology, channel type, elevation, valley confinement), but
these were not logistically feasible and would greatly increase the sample size, cost and time needed to complete
the study. The Washington Forest Practices Board also instructed the PHB Science Panel to develop a study plan
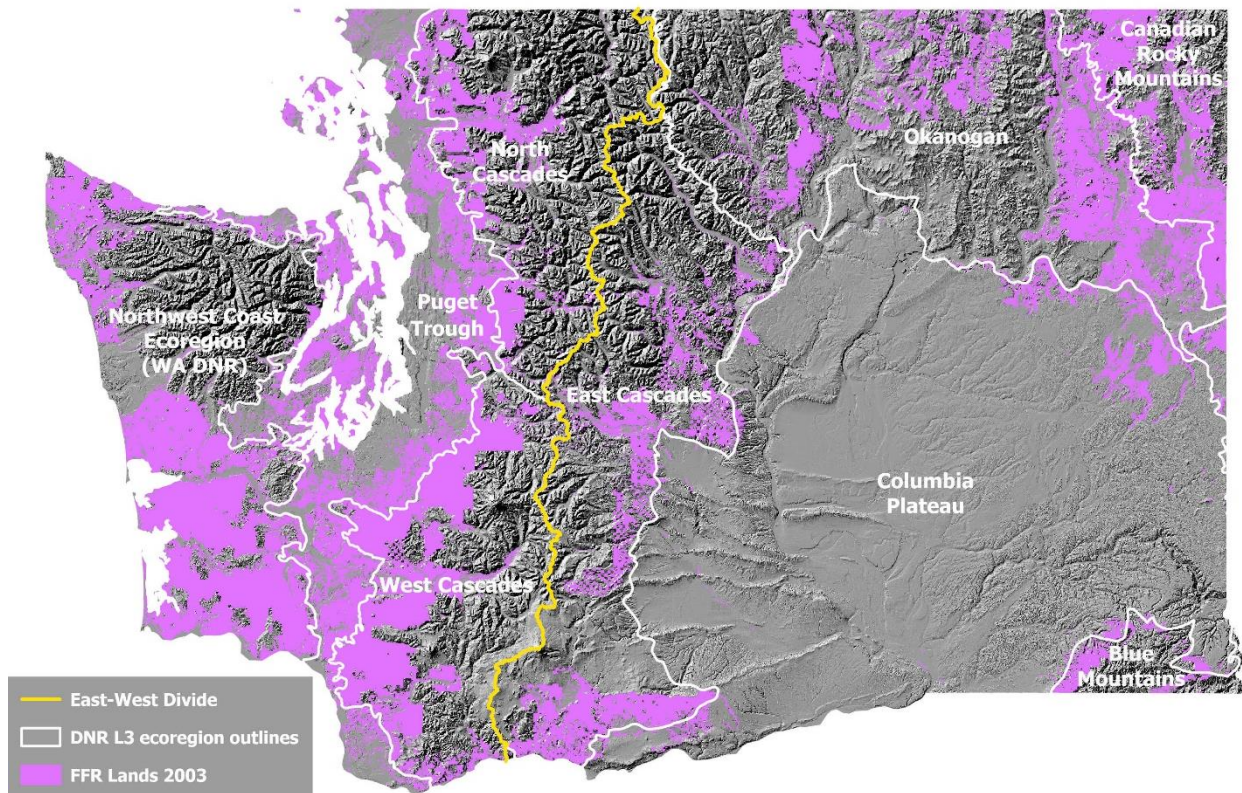that specifically included stratification by ecoregion.

488 **Figure 5. Washington Natural Heritage Program Level III ecoregions with Lands subject to the Forests**
489 **and Fish (FFR) forest practices rules designated in purple. Note the general absence of FFR lands in**
490 **the Columbia Plateau ecoregion. FFR lands mapped as of 2003. Ecoregion data downloaded from**
491 **https://data-wadnr.opendata.arcgis.com/datasets/wadnr::ecoregions-of-the-pacific-**
492 **northwest/explore?location=46.585091%2C-118.050200%2C6.03 in 2022.**

494 In this design, we do not propose the use of *a priori* stratification by ecoregion. A priori

495 stratification would be advisable for this study to model PHBs by ecoregion, to attain a desired

496 level of precision for each ecoregion, for administrative convenience, or to apply different

497 survey methodologies by ecoregion (Cochran 1977). However, none of these considerations

498 apply in this sampling design. We expect sampling effort to be allocated proportionally to the

499 relative area of ecoregions due to the implicit probability-proportional-to-size sampling

500 obtained from spatially balanced sampling. However, smaller ecoregions, such as the Blue

501 Mountains ecoregion, may receive fewer sampling points due to its smaller area and remote

502 location. "Islands" of sampling frame that are not contiguous can affect overall spatial balance

503   (Don Stevens, personal communication), in which case a priori stratification might be

504   necessary. When the sampling frame is available, the allocation of sites will be examined for

505   test sample draws to determine if adequate sample sizes within each ecoregion are obtainable.

506   Sampling effort will be apportioned among mapped terminal or lateral F/N break point types

507   (Figure 6) with post-hoc stratification. This approach is useful when the point types are not

508   known for each site before the survey, so no sampling frame is available to identify each

509   subpopulation for *a priori* stratification. Survey crews will record the point type at the time of

510   the survey and, when the desired sample size for a point type is satisfied, survey data from this

511   point type will not be collected at subsequent points of this type. Because the point type is not

512   known a priori so cannot be included as a survey design variable for stratification, employing

513   this technique will require adherence to the spatially balanced ordered list of sites to ensure

514   that the obtained sample of sites within each point type is also spatially balanced. The point

515   type should be recorded for each site so that inclusion probabilities for each site may be

516   calculated prior to analysis for any design-based summaries such as means and totals (Larsen

517   et al. 2008, section 2.4). This apportionment will only occur during the initial site surveys. If a

518   site changes from lateral to a terminal over the course of the study, we will not add any study

519   sites to accommodate that change.

520   Based on an analysis of observed variability in channel gradient and width upstream of

521   uppermost detected fish points from previous CMER studies and existing water type

522   modification forms (Appendix B), we propose to determine the location of uppermost

523   detectable fish at 160 sites in forested watersheds in eastern Washington and 190 sites in

524   forested watersheds in western Washington[3]. Habitat characteristics (gradient, channel width,

525   obstacles) will be measured using a longitudinal stream channel profile survey 660 ft (200 m)

526   above and 660 ft below the uppermost detected fish. The uppermost detected fish locations

527   will be determined during each sampling event via electrofishing surveys. The corresponding

528   habitat surveys surrounding the located uppermost fish point are expected to provide the data

529   necessary to evaluate differences among PHB criteria across the state and within the eastern

---

[3] The recommended sample size includes sites in addition to the minimum number calculated to meet the specified statistical requirements. This allows for site attrition over life of the project.

530   and western Washington regions. Data collected with consistent methods and crews might

531   have lower variability than the data we used to estimate sample size.

532   We will sample a small subset of sites across east/west regions concurrent with the site

533   selection year/process (during 'Year-0') in order to field test our methods without causing a
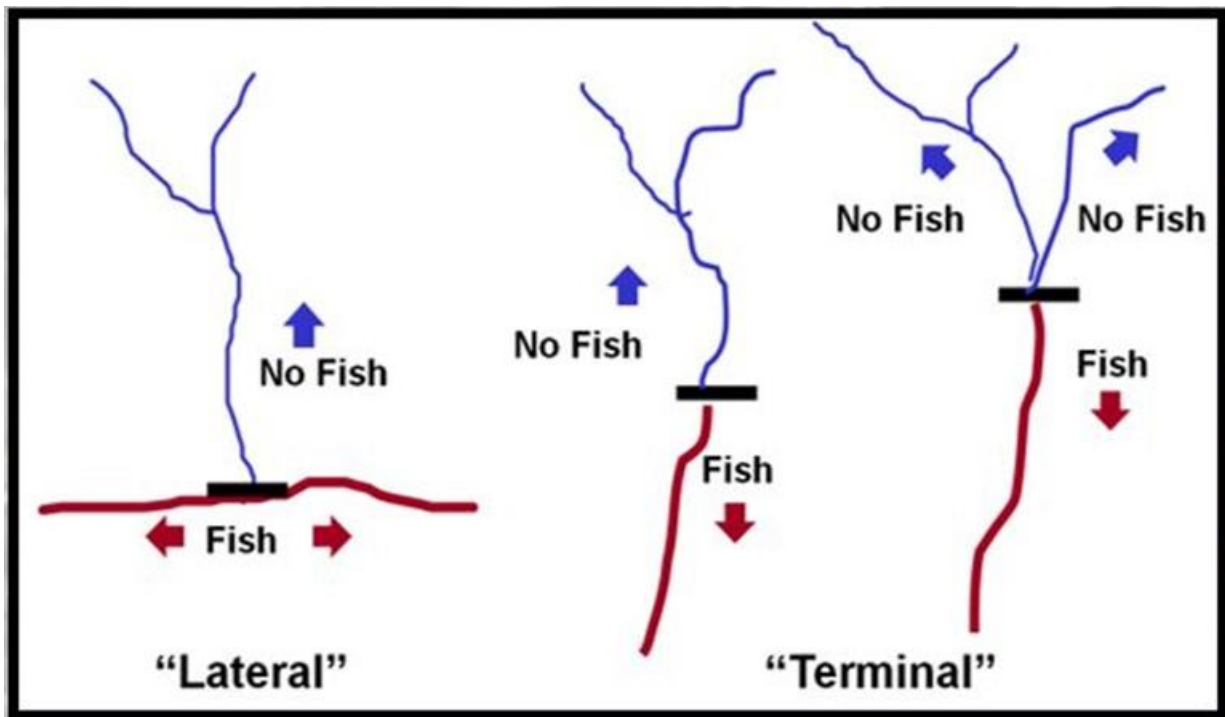
534   delay to project implementation.

535

536   **Figure 6. Schematic diagram of lateral versus terminal upstream limits of fish occurrence within**
537   **streams. The black bar(s) indicate the location of the uppermost fish (Fransen et al. 2006).**

538

539   **Site Identification**

540   The DNR Hydro Watercourses hydrography data layer contains stream channel locations across

541   the state. Stream lines are kept as segments with properties of each segment stored as

542   attributes. Segments are divided at intersections with other stream segments and any place

543   where their recorded properties change (e.g. - fish use/non-fish use). The points at which this

544   classification changes from fish (Type F) to non-fish (Type N) will be extracted from this hydro

545   layer. The properties of the fish use segment below the break will be retained with those data

546   points and stored in the new point layer. The attributes (properties) of interest for this study

547   include the criteria for fish use determination, such as whether it was a segment modeled as

548   likely fish habitat, a concurred point from a water type modification form, or a legacy

549   determination. Another attribute is whether that determination was based on biological

550   information (fish observation or electroshocking findings) or on physical habitat assessment.

551   Such information will be important for locating the optimum survey starting location but will

552   not be used for the purposes of selecting sample streams.

553   The F/N break points are intersected with the East/West Washington polygons to assign them

554   an East/West attribute. Points will also be intersected with the DNR Ecoregions polygon layer

555   to assign them an Ecoregion attribute. However, that attribute will be used as a covariate in

556   post-hoc analyses rather than as a stratification variable unless test sampling indicates

557   otherwise. The point layer will be subjected to the GRTS spatial randomization procedure,

558   which will assign a sequence number to each point. The points to be inspected for this study

559   will be selected from each side of the state in the sequence assigned. As points are discarded

560   according to our rejection criteria (below), the next sequential point will be added to the

561   sample population. In this way, spatial balance and random validity should be maintained.

562   In practice, batches of points will be selected and assessed for suitability, access permission,

563   and field crew accessibility to facilitate the sample set delineation prior to field surveys. These

564   batches will ensure that more points (streams) are ready to be sampled (and even perhaps

565   initially sampled) than are actually needed in case selected points are rejected during the first

566   study season. GRTS sample locations will be obtained from the sample draw in a GRTS design

567   file. Surveys that maintain the order of sites in the GRTS design file are spatially balanced

568   relative to the sampling frame from which the sample was drawn. Any sequential subset of

569   sites in the GRTS ordering is a spatially balanced subset of sites. Note that spatial balance does

570   not require that sites are *visited* in the order of the design file, but the sequential list of sites

571   should be fully enumerated by the end of the survey season with no skipped sites. This allows

572   field crews to visit the sites in an efficient manner while maintaining overall spatial balance of

573   the sample within any given year. For each site in the GRTS design file that is considered for

574   surveys, notes on any frame error or reasons for nonresponse will be recorded so that inclusion

575   probabilities for each site can be accurately calculated.

576   The F/N break point will identify the stream to be sampled, not necessarily the sample starting

577   point. The starting points will be the uppermost known fish location for that stream based on

578   any available information that can be obtained about that stream. The GIS layer contains some

579   information, such as the typing basis. Other information may be obtained from landowners,

580   tribal entities that monitor that stream area, and other local experts. In the case of tributary

581   streams that have no reliable fish observations, the electrofishing survey will start at the

582   confluence of the subject stream with the known fish-bearing mainstem stream. The initial

583   survey will determine lateral versus terminal status of the selected tributary for site allocation

584   purposes during site selection.

585   **Site Rejection Criteria**

586   Some potential study sites will be excluded from the sample population due to unforeseen

587   circumstances. During the site selection and field validation task, study sites may be dropped

588   as follows:

589   • Sites where the uppermost detected fish is associated with a man-made barrier;

590   • Streams showing evidence of recent (e.g., within five years) debris flows through the

591   subject stream;

592   • Sites where we cannot obtain landowner permission for the full survey length;

593   • Sites that are not safely accessible by field crews;

594   • Other reasons determined by project team.

595   In every case that a site is excluded from the sample, the reasons will be thoroughly

596   documented. Site rejection decisions will be approved by project managers and are not the

597   sole responsibility of field crews.

598   **Temporal Revisit Design**

599   Field surveys (electrofishing and habitat data collection) will be conducted during the

600   spring/early summer and the late fall/early winter sampling periods (seasons). These two

601   sample periods were chosen because they represent the most likely time periods for fish to be

602   found at their uppermost point in the stream network, and therefore should be adequate to

603   evaluate seasonal differences in the upper extent of fish use. While summer sampling may be

604   beneficial to compare seasons, due to the low flows typical of summer, it is unlikely that fish

605   would move higher into the system in that season (Cole and Lemke, 2006).

606   All sites will be surveyed every year during spring/early summer (current protocol electrofishing

607   survey window of March 1 to July 15) for three years to examine inter-annual changes in

608   uppermost detected fish locations. To evaluate seasonal changes in the location of the

609   uppermost detected fish, the sites that can be safely accessed in the fall/winter season will also

610   be visited with an augmented serially-alternating panel design. One quarter of the sites will be

611   assigned to the fixed panel and will be surveyed every fall/winter, and the remainder of sites

612   will be allocated to three alternating panels. One of three alternating panels will be surveyed

613   each year, with the sample augmented by the fixed panel to connect the sample across years

614   and seasons. The fixed panel will consist of the full count of sites from Table 2, while the

615   alternating panel counts will vary depending on site accessibility. The survey timing within both

616   sampling periods will be determined through consultation with regional experts to optimize

617   the timing based on local hydrology, fish life history, and potential for site access, and resurvey

618   timing will be consistent (within two weeks of the original survey date) across years.

619

620   **Table 2. Overall sampling schedule and number of sample sites by calendar year and season 2024 to**
621   **2026. All sites will be sampled in spring to early summer (March 1 to July 15) with the seasonal fixed**
622   **and alternating panel being resampled in fall to early winter high flow period (dates determined**
623   **through consultation with regional experts). A pilot study sampling 15 sites in eastern and 12 sites in**
624   **western Washington was completed in September of 2018 (Roni et al. 2018).**

Sampling Event	Pilot year (2018)	Year 1 (2024)	Year 2 (2025)	Year 3 (2026)
Spring to early summer		160 eastern Washington 190 western Washington	160 eastern Washington 190 western Washington	160 eastern Washington 190 western Washington
Late Fall/Winter Fixed Panel Sampled All Years (same sites)	27 to test methods	40 E WA 48 W WA	40 E WA 48 W WA	40 E WA 48 W WA
Late Fall/Winter Alternating panel, Sampled Only in Single Season		40 E WA 48 W WA	40 E WA 47 W WA	40 E WA 47 W WA
Reporting	Pilot study report	Annual report	Annual Report	Final Report

625   ## Data Collection

626   **Protocol Electrofishing and Habitat Surveys**

627   Electrofishing and habitat survey will provide a robust data set to inform the PHB and associated

628   analyses. Electrofishing surveys will be conducted to determine the location of the uppermost

629   fish at each survey event. Surveys at all study sites over three years will maximize the

630   probability of locating the upstream extent of fish habitat by incorporating both temporal and

631   spatial variability in fish movement due to physical (e.g., stream flow) and biological

632   (population dynamics) factors.

633   An intensive longitudinal thalweg and water surface profile survey (Roni et al. 2018) will be

634   conducted up- and downstream of the uppermost fish points following the electrofishing

635   surveys. The channel survey data will be used to partition the study reach into variable-length

636   stream segments that are scaled to lengths of homogeneous habitat attributes within the long-

637   channel profile. The length of segments will be based on changes in gradient and channel width

638   that are associated with inflection points and/or changes in habitat features (e.g., vertical and

639   non-vertical obstacle). Vertical and near-vertical obstacles will be captured as individual

640   segments, as such features will have some segment length associated with them.

641   Prior to sampling a site, the project team will review existing information from any available

642   sources on access, previous location of uppermost detected fish and habitat data, and obtain

643   landowner permission for access and sampling. In determining the upstream extent of fish

644   distribution, multiple upstream segments may be available for survey. When this situation

645   occurs, the selected surveyed segment will be the mainstem channel, defined as the stream

646   segment with the largest contributing basin area upstream from a tributary junction (should

647   have largest bankfull width, most flow, etc.). Where basin area upstream from a junction

648   appears approximately equal, rely on additional on-site metrics such as bankfull width and/or

649   flow to determine upstream direction of survey. Stream segments not included in the

650   hydrolayer may be encountered when moving upstream. These stream segments will be

651   included in the survey process in accordance with the above criteria.

652   Field crews will use modified DNR protocol electrofishing surveys with the intensity consistent

653   with methods being developed for FHAM to determine uppermost detected fish (Figure 7a) and

654   surveys will only be conducted when sampling conditions are suitable (avoiding periods of

655   extreme high/low flow or temperature, elevated turbidity, etc.). Water temperature (to the

656   nearest 0.1 °C), conductivity (microsiemens), and electrofishing setting (e.g., voltage,

657   frequency, pulse width) will be recorded at the beginning of each electrofishing survey. The

658   GPS coordinates of each uppermost detected fish location will be recorded, and the location

659   will be flagged and monumented with a marker including the survey date on an adjacent tree.

660   The fish species and approximate sizes will be recorded. Electrofishing surveys will continue

661   from the uppermost detected fish point upstream to at least the end of default physical fish

662   criteria (end DPC point). In the event the uppermost detected fish is found at the end of DPC,

663   electrofishing will continue 660 feet (upstream) to align with the extent of the detailed habitat

664   surveys. We will also record electrofishing survey time (shock seconds). In addition, coarse scale

665   habitat data will be collected on the full extent of the stream sampled during the e-fishing

666   survey.   These data will include channel gradient, bankfull width, wetted width and

667   confinement within unequal length segments of relatively uniform habitat character.

668   An intensive longitudinal thalweg and water surface profile survey (Roni et al. 2018) will be

669   used to assess key habitat attributes (i.e., gradient, bankfull and wetted width, water depth,

670   substrate size composition, and height of channel steps) below and above the uppermost

671   detected fish (Figure 7b). A previous study of variability on the upper limits of fish distribution

672   in headwater streams suggested that over 90% of the interannual variation in the uppermost

673   detected fish location occurred within 200 m (Cole et al. 2006). Therefore, we will use a

674   distance of 660 feet (200 m) below and 660 feet above the uppermost detected fish as our

675   intensive habitat survey reach.   The crew will measure 660 feet (horizontal distance)

676   downstream from the uppermost detected fish point to determine the beginning point for the

677   intensive stream habitat survey.

678   The intensive habitat survey involves surveying the streambed elevation along the deepest

679   portion of the stream (the thalweg), yielding a two-dimensional longitudinal profile of

680   streambed elevations.   This has been shown to be a reliable and consistent method for

681   measuring change in stream morphology and fish habitat independent of flow (Mossop and

682   Bradford 2006). We will also be recording water surface heights because surface levels are

683   what are important to fish with regard to obstacle heights. Survey measurements will be taken

684   every ten feet, and at any significant inflection points in topography or planform to be sure we

685   capture all changes in thalweg topography and gradient. A laser range finder mounted on a

686   monopod and a target on a second monopod will be used to collect distance and elevation

687   data. All data will be entered into a computer tablet in the field. Measurements and

688   observations at each point will include horizontal distance and slope between survey points,

689   water depths, wetted widths, bankfull width, dominant substrate (e.g., sand, gravel, cobble),

690   large wood, habitat feature type (e.g., pool, riffle, cascade), and general characterization of

691   flow and water conditions. Water surface elevation will be calculated after the survey from the

692   bed elevation plus the measured water depth. For steps and potential migration barriers, the

693   crew will record whether the step is formed by wood, bedrock, or another substrate. The

694   presence of wood is particularly important because wood-formed barriers and obstacles are

695   considered deformable and therefore are not PHBs. Crews will also note whether flow is

696   continuous or intermittent, the presence of beaver dams, groundwater inputs, and any other

697   unusual features (e.g., tunneled or sub-surface flow) that could influence fish distribution.

698   Because sites will generally be in small, constrained streams that are unlikely to change

699   significantly throughout the sampling year, it is likely that the habitat survey data for each

700   stream will only need to be collected once each year with the spring sampling effort. The survey

701   will be repeated annually to ensure we have a complete survey 660 feet above and 660 feet

702   below the uppermost detected fish found during each sampling event (Figure 7c). During each

703   survey, fixed elevation benchmarks will be placed at the bottom, middle (uppermost fish point)

704   and top of the intensive habitat survey reach to facilitate the coherence of repeat surveys. A

705   similar protocol based on Mossop and Bradford (2006) has been used to survey barrier removal

706   projects on small streams throughout the Columbia River Basin (Clark et al. 2019, 2020).

707   Evaluations of various regional stream habitat survey protocols have demonstrated that with

708   *well-trained* field crews, measurement error is small relative to naturally occurring variability

709   amongst sites (Kershner et al. 2002; Roper et al. 2002; Whitacre et al. 2007, Archer et al. 2004).

710   Therefore, all crews will participate in a three to five-day training course each year prior

711   to initiation of spring sampling to ensure consistency among crews in determining uppermost

712   detected fish locations, surveying habitat characteristics (long-profiles), and data collection.

713   Training should incorporate identifying potential sources of variation in measurement that can result

714   from dense vegetation, identification of features, and clarity of protocols (Roper et al. 2010). In

715   addition, mid-season check-in/corrections will be conducted with each crew to prevent

716   sampling drift (this process will be outlined in the Quality Assurance Plan). Moreover, to

717   quantify variability among crews in conducting longitudinal surveys, we propose that 10% of all

718   sites sampled each spring should be resampled during the same year and season by other crews

719   every year. Since variation in stream flow during subsequent surveys should not affect the

720   longitudinal bed profile, we don't expect flow changes to contribute to variability observed

721   among crews in these resurveys.

722
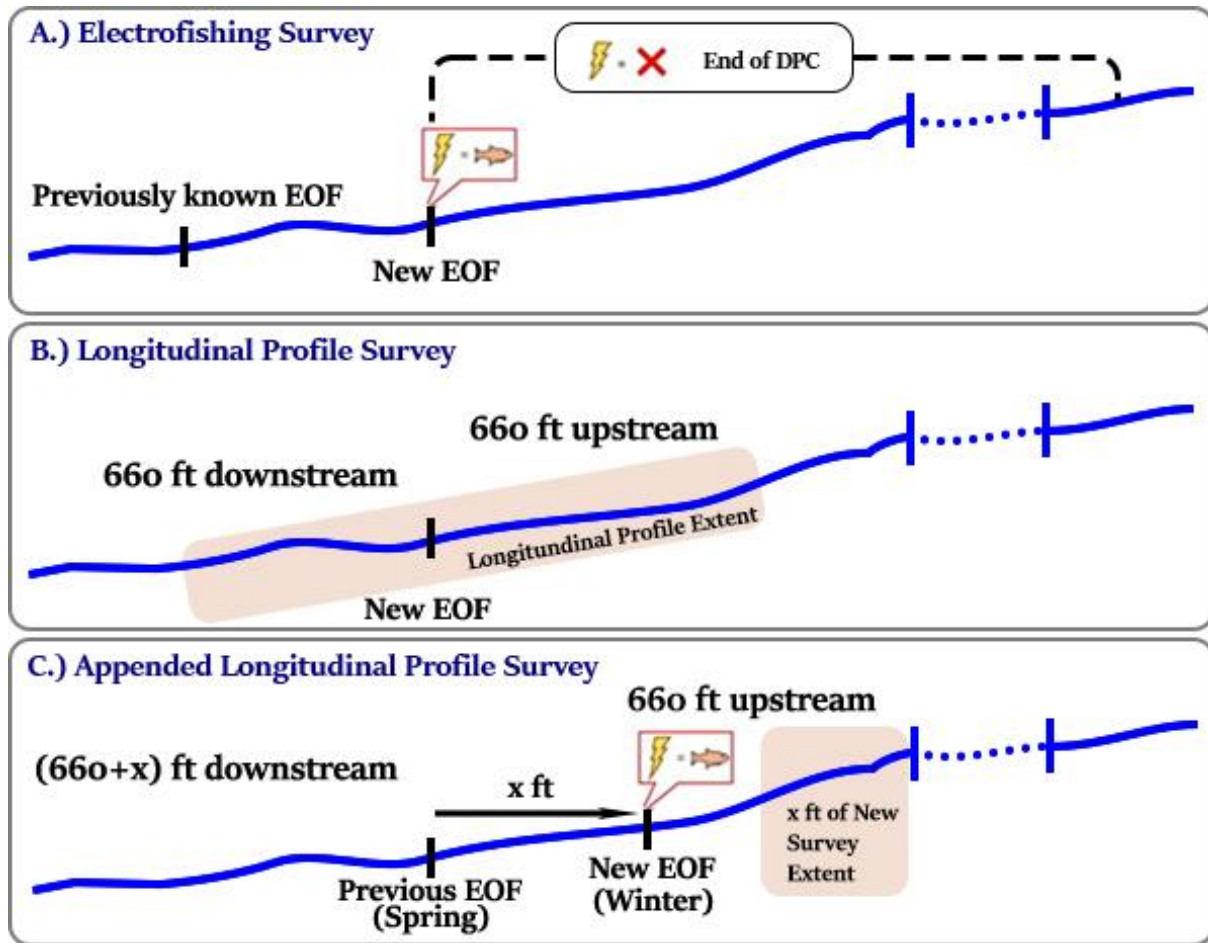
Figure 7. Components of field surveys demonstrating: (A) the extent of the protocol electrofishing survey to determine uppermost detected fish (EOF) point, (B) the range of the initial longitudinal profile habitat survey associated with the initial EOF point, and (C) an example of how the longitudinal profile survey would be appended if follow up protocol electrofishing surveys identify a new EOF point (adapted from PHB Science Panel 2019).

728

Reach- and Basin-Scale Explanatory Variables Derived from Office and Remote Sources

We will also collect data on several other factors that are thought to play a role in uppermost detected fish point and identification of PHBs from sources other than field data. These include:

elevation, aspect, drainage area, distance-from-divide[4], valley width, annual precipitation,

channel type[5], riparian stand condition[6], whether uppermost detected fish and PHB is at a mid-

---

[4] Palmquist (2005) found distance-from-divide to be less variable and more reliably calculated than basin area
[5] Montgomery & Buffington, 1993
[6] Watershed Analysis categories, WA DNR 1997

734 channel point (mainstem or terminal) or confluence (tributary or lateral tributary), dominant

735 drainage area geologic competence category[7], stream order, and whether a stream is

736 accessible to anadromous fish or only resident fish. Many of these variables will be derived

737 from existing GIS data layers. Drainage area, distance-from-divide, and valley width are

738 important because they, combined with annual precipitation, are related to flow and stream

739 size. The local geology around the stream determines whether stream substrate tends to

740 consist of hard, resistant, larger particles or friable, fine-grained substrates, which have been

741 shown to influence fish distribution (Gresswell et al. 2006; Torgersen et al. 2008).

Data Preparation

743 Physical attribute and fish presence data will be organized by site and variable-length segment

744 as laid out in Appendix F. To prepare data for analysis, the stream profile will be divided into

745 variable-length homogeneous segments, and each segment will be populated with a suite of

746 segment-scale physical attributes and fish presence or absence. Variable-length segments will

747 also be populated with associated basin-scale attributes that will be derived from GIS. Other

748 basin-scale characteristics will be included for each site. Measures such as gradient and

749 channel width can be used to form threshold variables and cumulative metrics (e.g., gradient

750 and width expressed over multiple segments) that can be assessed as predictors of PHBs. Data

751 sets will be developed for each sampling event to assess changes in distribution over time.

Data Analyses

Data Exploration, Summary Statistics, and Initial Tests

754 After data preparation is complete, initial data exploration will include graphical examination

755 of habitat metrics for segments within a site and segment means of physical characteristics for

756 each site (Figure 8). Distributions of physical attributes for variable-length segments at a site

757 can be compared for segments with and without fish by and across sites. The length of

758 segments will be based on changes in gradient and channel width that are associated with

759 inflection points and/or specific habitat features (e.g., vertical [falls] and non-vertical obstacles

---

[7] Competent/Incompetent, per McIntyre et al. 2009

760   [steep cascades]). Criteria for classifying variable-length segments and obstacles will be derived

761   during post-hoc data analysis using linear regression methods similar to those described by

762   Tompalski et al. (2017).   All statistical analysis described here presume the use of the R
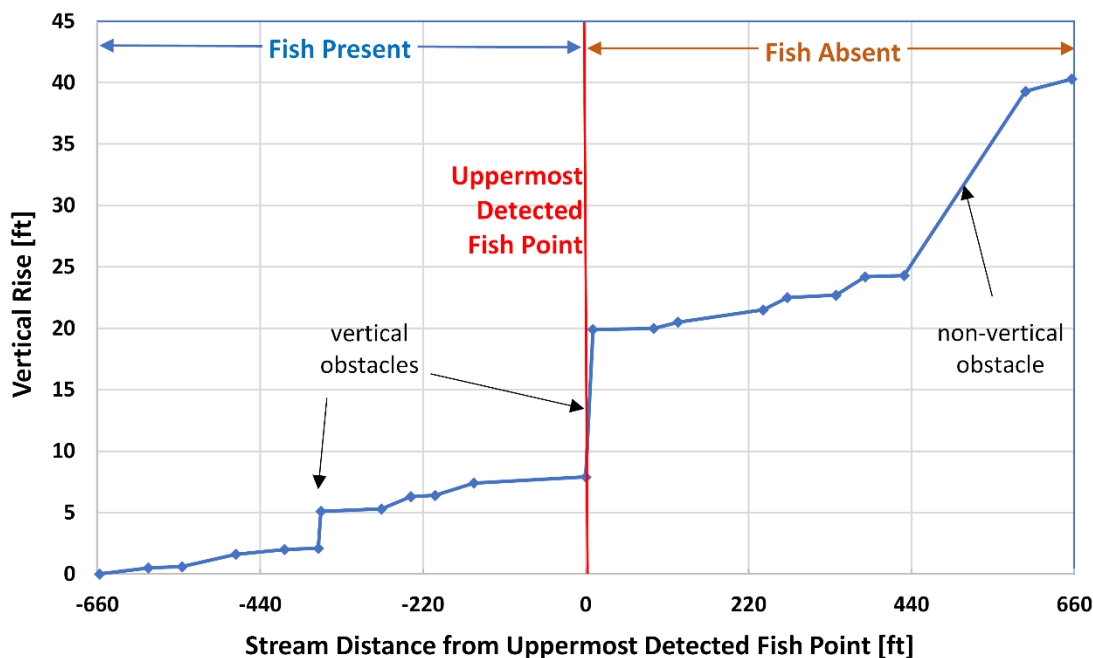
763   statistical programming language (R Core Team 2021).

764

765   **Figure 8. Schematic of channel long-profile survey showing variable-length segments (i.e., distance**
766   **between inflection points) and associated vertical and non-vertical obstacles.**

767

768   **Examining Uppermost Detected Fish Locations**

769   Research questions related to uppermost detected fish locations will address interannual

770   (Research Question #1), seasonal (Research Question #2), and spatial (Research Question #3)

771   dynamics. For sites in the fixed and alternating panels that are revisited over time, physical

772   attributes at each site may be summarized by year and by season (spring or fall/winter).

773   Stream profile plots (Figure 7) will be developed to compare uppermost detected fish points

774   across seasons and years.

775   To examine spatial patterns, physical attributes at each site will be summarized by region (east

776   or west), ecoregion, or other spatial classifications, and maps of attributes will be developed to

777   visually assess spatial patterns in distribution. Summaries may also be examined by point type

778   (lateral or terminal). For the subset of streams visited in the panel design, distances between

779   the lowest and highest uppermost detected fish locations will be computed for each stream

780   and mapped to examine spatial distributions of movements over time. Mapping the spatial

781   distribution of movements over time will contribute to adequate determination of PHBs based

782   on probability of observed fish movement.

783   **Examining Habitat Associated with Uppermost Detected Fish Locations**

784   Spatial patterns in physical channel and basin characteristics (e.g., bankfull width; average

785   gradient, basin size) associated with the identified upstream extent of fish habitat will be

786   examined to determine how these metrics vary geographically across the state of Washington

787   (Research Question #4). Maps and histograms of physical channel and basin characteristics will

788   be used to assess distributional patterns in attributes associated with the uppermost detected

789   fish. Summaries of physical channel and basin characteristics (mean, median, standard

790   deviation, range) will be calculated by spatial categories such as region (e.g., eastern versus

791   western Washington) and ecoregion. Generalized linear mixed models (GLMM ; McCullagh and

792   Nelder 2019, Bolker et al. 2009) of physical channel and basin characteristic metrics, as

793   response variables, will incorporate fixed effects for region, ecoregion, point type (terminal and

794   lateral), and other spatial factors. Random effects reflecting spatial structure (e.g., segments

795   within streams) will be incorporated to account for correlation. Surveys will identify the

796   uppermost detected fish point during each sample period at each study site, and the first PHB

797   encountered upstream from that point. Characteristics of these PHBs will be used to determine

798   how survey timing might influence which PHB would be associated with the proposed F/N

799   break and how frequently the PHB might be identified differently (Research Question #5).

800   Distributions of continuous habitat metrics (e.g., gradient, channel width) will be compared

801   with boxplots or violin plots for sites where fish have moved above PHBs compared to sites

802   where fish did not. These graphical summaries will be used to identify factors associated with

803   fish movement by year and season. The probability that the uppermost PHB at a site is

804   consistently selected during different survey occasions will be modeled as a function of season,

805   spatial factors, point type, and physical channel and basin characteristics to determine what

806   factors influence repeatability of identifying a PHB.

807   Physical changes in features originally identified as PHBs over time (Research Question #6) will

808   also be assessed. For each measured physical characteristic, a GLMM will be applied to examine

809   effects of time to estimate trends or changes over the course of the study. An examination of

810   how similar features appear to limit upstream fish distributions in some contexts but not others

811   (Research Question #7) will be conducted to examine any potential interactions among physical

812   characteristics (e.g., headwaters vs. downstream; different flow levels). These relationships will

813   be assessed in GLMMs with significance tests of the interaction effects.

814   **PHB Performance Analyses**

815   The primary goal of this project is to identify PHBs associated with EOF habitat using a suite of

816   physical channel attributes and basin characteristics (Research Questions #7 and #8). A subset

817   of physical channel attributes and basin characteristics will be identified as predictors to

818   develop PHB criteria using classification methods described below. The performance of these

819   developed PHB criteria and three sets of classification criteria proposed by the Board will be

820   evaluated. We first describe how random forests (Cutler et al. 2007, Trigal and Degerman 2015)

821   and interaction forest (Hornung and Boulesteix 2022) will be used to identify a subset of PHB

822   predictors that will be used in a classification and regression tree (CART; Breiman et al. 1984)

823   model to obtain thresholds for identifying PHBs. Then we describe the methods used to

824   compare the performance of each set of PHBs to inform the final selection of PHB criteria.

825   Random forest modeling will apply the *randomForest* package (Liaw and Wiener 2002),

826   interaction forest will utilize the *diversityForest* package (Hornung 2022) and generalized linear

827   mixed modeling will be conducted with the *glmmTMB* package (Brooks et al. 2017). CART

828   modeling and visualization will utilize the *rpart* package (Thernau and Atkinson 2022).

829   <u>PHB Classification Methods</u>

830   Given the complexity of identifying PHBs due to the variability in stream characteristics across

831   space and time and fish movement across obstacles, the classification of alternative PHBs will

832   incorporate: 1) Random forest modeling to determine variables important for separating fish

833 bearing segment from non-fish bearing; 2) Interaction forest models to identify characteristics

834 that in combination create PHBs; 3) an evaluation of variables related to probability of fish

835 movement using binomial GLMM; and 4) CART models to identify the thresholds for PHBs

836 based on the random forest and interaction forest outputs and the evaluation of probability

837 of fish movement.

838 Random forest (RF) methodology is a nonparametric approach used for classification and

839 prediction and can identify important predictor variables among a large suite of possible

840 covariates even when those covariates are highly correlated (Cutler et al. 2007, Kubosova et al.

841 2010). Random forest can also bin continuous data into discrete categories as part of the

842 analysis, as opposed to assigning arbitrary bins *a priori*. Cutler et al. (2007) found that random

843 forests had high classification accuracy compared to classification trees, generalized linear

844 models (logistic regression), and linear discriminant analysis. Random forest classification has

845 been used to classify salmonid habitat in Alaska (Romey and Martin 2021), fish assemblage

846 presence in stream segments in coastal Australia (Rose et al. 2016), and in macroinvertebrate

847 habitat in the Czech Republic (Kubosova et al. 2010). Random forest methods have been

848 extended to boosted random forests (Ko et al. 2015, Mishina et al. 2015) which features more

849 memory-efficient calculations. When classification covariates are impacted by spatial and/or

850 temporal correlation, binary mixed model forest (Speiser et al. 2019) or generalized mixed

851 effects random forest (Fontana et al. 2021, Seibold et al. 2019) can account for these sources

852 of correlation.

853 Random forest classification of fish use will be used to determine which segment-level,

854 cumulative (e.g., metrics such as gradient and width expressed over multiple segments), and

855 basin-scale characteristics are important variables for PHB establishment. Separate random

856 forest classification models may be applied to eastern and western sites and for lateral and

857 terminal points to identify influential variables independently in each system. The data will be

858 split into training and testing data sets to assess the performance of the random forest

859 classification. A random forest model will be developed from the training data set and then

860 applied to the test data set to assess classification. Classification performance metrics will

861 include sensitivity (proportion of presences correctly classified), specificity (proportion of

862   absences correctly classified), kappa (a measure of agreement computed across presences and

863   absences, Cohen 1960), and the area under the receiver operating characteristic curve (Fawcett

864   2006). The final model will be applied to the entire sample of uppermost detected fish points

865   at each site to obtain habitat variables related to the PHB associated with those points. Sites

866   with PHBs formed by vertical and non-vertical obstacles (e.g., waterfalls and cascades) can also

867   be analyzed separately from sites with width- and gradient-related PHBs so that random forest

868   models accurately reflect each type of PHB and more nuanced habitat relationships are not

869   missed. Vertical step height will be included as a segment-scale attribute. Alternatively, a single

870   model incorporating waterfall height (where height is zero if no waterfall is present) may

871   provide the basis for threshold definitions across all streams. Interaction random forest

872   modeling will be used to identify more complex relationships between habitat covariates

873   relative to PHBs. The covariates identified in the random forest and interaction forest models

874   will be used in the CART model to identify thresholds for PHB criteria. See the pilot data analysis

875   summary (Appendix C) for more information.

876   The probability of fish movement will be evaluated through a binomial GLMM based on

877   whether the uppermost detected fish location changed across surveys at a particular site. The

878   purpose is to identify weaker or stronger PHBs. After all data have been collected over the

879   three-year study, uppermost detected fish points identified during all surveys at all locations

880   will be categorized into two sets of PHBs: those that fish were and were not observed to move

881   beyond in an upstream direction over the course of the study. Physical channel and basin

882   characteristics will be calculated at the segment level and cumulatively across segments both

883   upstream and downstream of the uppermost detected fish point. A binomial GLMM will be

884   applied to the segment-level indicator that no fish was detected at or above the segment at a

885   particular survey occasion to model the probability that no movement occurs upstream of the

886   PHB, and a stream level random effect will be incorporated to account for the nesting of

887   segments within a stream. The model of the probability that fish do not move above a PHB may

888   contain classification and continuous covariates that describe physical habitat attributes (e.g.,

889   channel bankfull width, gradient) or explain seasonal movement, including the season, region

890   (east/west), ecoregion, and point type (lateral/terminal). Random effects for space and time

891   will ensure that standard errors for fixed effects estimates are not underestimated due to
892   correlation. Variance components may also incorporate habitat categories for which variance
893   heterogeneity in seasonal movement is observed (e.g., low vs high elevation). This model will
894   be used to assess the reliability of the PHBs identified by the CART model. A PHB that is
895   surpassed more often could be considered a weaker PHB, whereas a PHB that is surpassed less
896   frequently could be deemed a strong PHB.

897   CART models are a type of decision tree machine learning model that can identify variables of
898   importance, can accommodate unequal spatial sampling, and classify based on continuous and
899   categorical predictors (Morgan 2014, Loh 2011). We propose incorporating CART models
900   because, unlike random forest classification models, CART models return thresholds used at
901   splits in a decision tree. While, random forest models will likely have higher prediction accuracy,
902   they are not ideal for establishing thresholds. A random forest contains many individual
903   decision trees (a forest) to deal with the uncertainty that results from a single decision tree
904   (Maroco et al. 2011). CART models will be built for several combinations of variables (e.g.,
905   variables identified by random forest, interaction forest, or the FPB) to determine which
906   combination of variables produces the highest prediction accuracy and enables comparison of
907   model performance based on sensitivity, specificity, and Matthews Correlation Coefficient
908   (MCC). MCC is a statistical representation of all four confusion matrix categories (true positives,
909   true negatives, false positives, and false negatives) that is a reliable and holistic indicator of
910   model performance (Chicco and Jurman 2020). A visual decision tree will be presented for each
911   model to identify potential thresholds for variables in the model. We plan to compare the
912   existing Board criteria and alternatives by comparing the accuracy, sensitivity, specificity, and
913   MCC between models. These metrics will enable us to investigate trade-offs between model
914   accuracy and complexity for establishing putative thresholds. The model can also be tuned
915   based on the false negative cost to influence the model's emphasis for sensitivity or specificity.
916   Additionally, CART models may be built from data combined across years or may be developed
917   from data specific to a single year and then applied to a subsequent year to evaluate
918   classification accuracy.

919   Crew-variability testing conducted within this study will provide insight into our ability to

920   identify the same PHBs using data collected by different survey crews when implementing

921   FHAM in the field in the future (Research Question #9). Data from the subset of streams

922   surveyed multiple times by different survey crews will be used to assess crew variability in

923   measuring the physical stream characteristics that would be used to identify PHBs. Physical

924   characteristics measured at the same streams by different survey crews will be modeled to

925   identify attributes that are more susceptible to survey crew variability. Distances between PHBs

926   identified at the same stream based on data collected by different crews will be modeled as a

927   function of spatial characteristics such as region and ecoregion to determine if spatial factors

928   influence crew variability.

929   <u>Performance Evaluation of Board-Accepted PHB Criteria</u>

930   The three sets of classification criteria proposed by the Washington Forest Practices Board

931   (Research Question #10) will be assessed in three different ways. The first method will be to

932   compare frequencies that the various criteria occur above and below the uppermost detected

933   fish. The performance of each type of PHB variable (i.e. – gradient, obstacle characteristic,

934   channel width) and criterion within the three proposed criteria sets will be assessed individually

935   and then in combination with the others. The second will be to create a confusion matrix and

936   MCC for the Board criteria, as compared to the alternative PHBs determined by the CART

937   models. The third method will use CART analysis including only the physical habitat variables

938   utilized in the Board criteria. The resulting critical values, or thresholds, identified by the CART

939   model will be compared to the values in each criteria set established by the Board.

940   For each set of Board criteria, distance between the PHB and the uppermost detected fish will

941   be examined as a measure of PHB prediction performance. The mean, median, standard

942   deviation, and range of the set of distances for each set of Board criteria will be calculated and

943   compared to the distances obtained with PHB criteria from the CART analysis. The distances

944   between a PHB and the uppermost detected fish will be modeled with GLMMs as a function of

945   covariates, and the associated covariates identified in the model will be used in the random

946   forest and CART models to identify new PHBs. The distribution of distances between a PHB and

947   the uppermost detected fish will also be compared for the alternative PHB criteria from the

948   CART models to the Board Criteria. The proposed analysis methods are summarized by research

949   question in Table 3.

950   Following the first year of data collection we will perform a demonstration analysis to verify

951   desired outputs and analytical approaches described here within.

952   <u>Analysis of Pilot Study Data</u>

953   Data from a 2018 pilot PHB study (Roni et al. 2018) that used similar habitat data collection

954   methods as those proposed in this current design were analyzed to demonstrate available

955   analysis tools to identify habitat attributes associated with the uppermost detected fish

956   (Appendix C). Random forest models, interaction forest models, and CART models were

957   applied to habitat covariates obtained from the pilot data to identify important habitat

958   covariates associated with the uppermost detected fish. Additionally, random forest

959   methodology was used to assess the Forest Practices Board-proposed PHB criteria. Covariates

960   identified by random forest and interaction forest models were used in CART models to

961   identify PHB criteria. Accuracy, sensitivity, specificity, and Matthews correlation coefficient

962   (MCC; Chicco and Jurman 2020) were used to assess performance. The pilot study data set

963   does not include temporal replication and therefore could not inform inference on seasonal

964   and/or annual fish movement.

965

966   **Table 3. Proposed data analysis methods by Research Question.**

Research Question	Question	Proposed Methods	Data Sets
1	How do the locations of the last (uppermost) detected fish vary interannually?	Stream profile plots, summaries of physical channel and basin characteristics by year, summaries/models of distances between lowest and highest uppermost detected fish points across seasons by year	All data excluding crew variability data and error distance surveys
2	How do the locations of the last (uppermost) detected fish vary seasonally?	Stream profile plots, summaries of physical channel and basin characteristics by season, summaries/models of distances between lowest and highest uppermost detected fish points between seasons within years	Yearly data excluding crew variability data and error distance surveys

Research Question	Question	Proposed Methods	Data Sets
3	How do the locations of last (uppermost) detected fish vary geographically across the state of Washington?	Stream profile plots, maps of distances between lowest and highest uppermost detected fish points within streams among all survey occasions.	Stream and PHB attributes associated with uppermost detected fish points for each site
4	How do the physical channel and basin characteristics (e.g., bankfull width; average gradient, basin size) associated with the identified end (upstream extent) of fish habitat vary geographically across the state of Washington?	Maps of physical channel and basin characteristics associated with the identified end (upstream extent) of fish habitat, summaries of physical channel and basin characteristics associated with the identified end (upstream extent) of fish habitat for spatial categories such as region and ecoregion, models of physical channel and basin characteristics metrics with fixed effects for region, ecoregion, and other spatial factors.	Stream and PHB attributes associated with uppermost detected fish points for each site
5	Where the location of the last (uppermost) detected fish changes (seasonally or interannually), how does that influence which PHB would be associated with the F/N break and how frequently does that occur?	For each visit to a stream, determine the PHB corresponding to the uppermost detected fish for that visit then model the indicator of whether or not a fish was observed upstream of each PHB as a function of physical channel and basin characteristics to assess the probability that a PHB remains the "PHB of rule".	All data excluding crew variability data and error distance surveys
6	How do the physical channel features at the locations initially identified as PHBs change over the course of the study?	For the subset of PHBs visited at least twice, model changes each physical characteristic as linear trends, seasonal effects, and/or nonlinear effects. Include site random effects to examine spatial patterns in physical channel feature variation. Note that changes in physical characteristics can be related to crew effects.	The subset of PHBs visited at least twice

Research Question	Question	Proposed Methods	Data Sets
7	How often do similar features appear to limit upstream fish distributions in some contexts but not others (e.g., further into the headwaters vs. downstream; different flow levels)?	Assess interactions between physical characteristics in GLMM of distances between uppermost detected fish locations and PHB	Stream and PHB attributes associated with uppermost detected fish points for each site
8	Which combinations of physical channel features and basin characteristics (for example, gradient, channel width, barriers to migration) best identify the end (upstream extent) of fish habitat relative to the location of the last (uppermost) detected fish?	CART models informed by random forest, interaction forest, Board criteria, and covariates from a GLMM of distances between the uppermost detected fish and PHB defined from Board criteria. Assess segment-level performance of CART model thresholds with confusion matrices; measures of sensitivity, specificity, MCC, and classification accuracy. Assess stream-level performance of CART model thresholds by comparing the mean, median, range, and SD of distances between the uppermost detected fish and PHB across all streams and select PHB criteria that minimize those metrics.	Stream and PHB attributes associated with uppermost detected fish points for each site will be used to develop a potential alternative to the FPB-selected criteria sets, but all uppermost detected fish points would be used for the probability of movement test of PHB strength
9	Can protocols used to describe PHBs be consistently applied among survey crews and be expected to provide similar results in practice?	Physical characteristics measured in repeated surveys by different crews at the same sites will be used to identify PHBs. Models of PHB consistency relative to the uppermost PHB will be used to estimate the probability that crews identify the same PHB. Physical characteristics will be modeled to identify attributes that are more susceptible to measurement error among survey crews.	Crew variability data

Research Question	Question	Proposed Methods	Data Sets
10	How well do the PHB criteria provided by the Washington Forest Practices Board accurately identify the EOF (upstream extent of fish) habitat when applied in the Fish Habitat Assessment Methodology (FHAM)?	Assess segment-level performance with confusion matrices; measures of sensitivity, specificity, MCC, and classification accuracy. Assess stream-level performance by comparing the mean, median, range, and SD of distances between the uppermost detected fish and PHB.	Stream and PHB attributes associated with uppermost detected fish points for each site

967

Potential Challenges

969 Although the methods we propose have been widely used to quantify habitat conditions and

970 identify the location of uppermost detected fish, there are some potential challenges. These

971 include location of sites that meet selection criteria, access to initially identified sites, and access

972 to these sites throughout the two seasons and three years. It is possible that we may not have

973 access to selected sample sites due to issues with land ownership, landowner willingness to

974 permit access, or problems with the road networks. Thus, if a site is not suitable due to access

975 or for other reasons a different site (the next consecutive site number from the initial random

976 selection) would be used to replace the non-suitable site, and the reasons the site is excluded

977 will be documented.   This study is targeted at identifying the features and channel

978 characteristics that limit the upstream extent of fish distribution, which should not be strongly

979 dependent on particular land uses or ownership types. Therefore, results should have broad

980 applicability despite any site selection biases that may occur. A more challenging scenario

981 would be if accessibility changes between or among seasons and years. For example, forest

982 fires, heavy early or late snow, or road failures could affect repeat surveys at a site. In such

983 cases, we would continue to sample sites during other seasons and years when possible. The

984 recommended sample size includes sites in addition to the minimum number calculated to

985 meet the specified statistical requirements. This allows for some site attrition over life of the

986 project.

987   An additional challenge with study implementation will be largely financial and could result
988   from underestimating or overestimating the amount of time and cost needed to adequately
989   sample sites initially and repeatedly. Similarly, we need to ensure that the data collected will
990   allow us to answer the PHB study questions. To proactively assess these critical uncertainties,
991   a pilot (feasibility) study was conducted in August of 2018 to test and refine protocols, and
992   estimate the time needed to conduct a survey and collect data at a site (Roni et al. 2018). The
993   pilot study included conducting longitudinal thalweg profile surveys upstream and downstream
994   of known uppermost detected fish points at 27 sites on private, state, and federal forestlands
995   in western and eastern Washington. The analysis of longitudinal survey data from the pilot
996   study demonstrated that PHBs based on gradient, BFW, and obstacles being examined by the
997   Board could be easily determined from the survey data. The field surveys helped identify
998   several modifications to the initial proposed protocol that are needed to assure the proposed
999   and other potential PHBs can be easily identified (e.g., spacing of the survey points, habitat
1000   types, minimum habitat length, and substrate categories). It also provided important
1001   information on time needed to conduct surveys, which we have incorporated into the study plan
1002   and estimated cost to conduct the full validation study.

1003   This study does not address long-term changes in small streams that may render them
1004   unsuitable for fish occupancy, or conversely, may render previously unsuitable streams
1005   habitable for fish. At any point in time, some headwater streams are not used by fish during
1006   any season of the year due to a blockage, to invasion, or to unfavorable physical conditions (e.g.,
1007   gradient) in the channel itself. Factors that determine whether small streams can be used by
1008   fish are typically related to disturbances such as exceptionally high discharge, landslides, debris
1009   flows, and windstorms. Such episodic disturbances are erratic and can be widely spaced in time
1010   (decades to centuries), but their overall effect in drainage systems is to create a mosaic of
1011   streams suitable for fish occupancy that changes over long intervals (often hundreds of years)
1012   in response to local disturbance regimes (Kershner et al. 2018; Penaluna et al. 2018). An
1013   important implication of the notion that the potential use of small tributaries by fish can change
1014   over time is that while some stream segments are not now occupied by fish, there is no
1015   guarantee that they may not become suitable in the future, or that those which are currently

1016 habitable will always remain so. This study, however, does not address the expansion and
1017 contraction of fish habitat over long time intervals, because the sample time is limited to three
1018 years and the methods cannot predict with certainty where and in what form large
1019 disturbances capable of transforming a stream segment's ability to support fish will occur.

Expected Results and Additional Studies

1020

1021 Highly precise measurements of stream channel conditions both upstream and downstream of
1022 uppermost detected fish locations will provide a nearly continuous dataset of physical stream
1023 characteristics within the surveyed area. Thus, we will be able to objectively identify the
1024 physical stream characteristics most closely associated with uppermost detected fish. These
1025 data will be used to test the different PHB criteria under consideration by the Board in 2018,
1026 and also to identify alternative physical stream characteristics that may function as PHBs. We
1027 expect that the study will assess the performance of proposed and/or identify alternative PHB
1028 criteria for gradient, channel width, and obstacles that are most frequently associated with the
1029 furthest upstream of all uppermost detected fish points found at each stream across the time
1030 period of the study. Seasonal and inter-annual sampling will allow us to examine the variation
1031 of uppermost detected fish locations across years and seasons, which will help identify PHBs
1032 that are consistently associated with the upstream extent of fish habitat across years, seasons,
1033 and flow conditions regardless of where fish are found on any given day. Because we will be
1034 using some sites for which a WTMF already exists and the location of the uppermost detected fish
1035 was potentially identified, examining longer-term inter-annual variation in the uppermost
1036 detected fish may be possible for a subset of sites where uppermost detected fish has been
1037 previously identified and monumented. In addition, study sites could be revisited in the future
1038 to look at longer-term changes in uppermost detected fish locations, if desired.

1039 Ultimately, the analysis will provide the distances (upstream and downstream) from uppermost
1040 detected fish to the different proposed PHB criteria, if and how that differs among years and
1041 seasons, whether one set of criteria performs better in terms of consistently identifying EOF
1042 habitat across seasons and years, and whether different PHB criteria should be applied for
1043 different regions or should be stratified by other factors. While the focus of the study is to test
1044 the three different sets of PHB criteria being considered for adoption by the board, we expect

1045   that the analyses will help identify other criteria that might more consistently be associated with

1046   the uppermost detected fish and therefore better indicate upstream extent of fish habitat

1047   when integrated with FHAM.

1048   The results should also help inform the protocols for measuring gradient, bankfull width, and

1049   obstacles in the field to minimize variability among field crews and assure consistent

1050   identification of PHBs. Focus should be placed on specific protocols used to consistently and

1051   accurately identify and measure physical stream characteristics, including gradient, bankfull

1052   width, obstacles, and any other criteria that may be used to identify PHBs in this study.

1053   We will also examine seasonal and inter-annual changes in uppermost detected fish locations

1054   in headwater streams across the state. While this would potentially lay the groundwork for

1055   continued monitoring of long-term variability in the upstream extent of fish distribution, it is

1056   not designed as a long-term study on such variability. Depending on results, we may

1057   recommend that sites continue to be periodically revisited in the future to examine this longer-

1058   term variability. It is possible that a 3-year study period may not capture a sufficiently broad

1059   range of hydrological conditions associated with shifts in climatic cycles (e.g., El-Nino/La-Nina)

1060   to allow for the estimation of the best "average" upon which a PHB boundary can be

1061   determined. This can only be assessed once the 3-years of sampling have been completed.

DPC Study Integration

1063   The electrofishing and habitat surveys for each PHB study stream will extend up to or beyond

1064   the end of current DPCs. Therefore, the PHBs study will yield a data set that can be analyzed

1065   regarding the frequency with which fish are found up to the limits of current DPCs, including

1066   how this varies between seasons, years, and geography. The coarse-scale data collected during

1067   the electrofishing survey will also provide channel profiles and other data for the reaches

1068   between EOF/H and end of current DPC that can be analyzed for possible explanations as to

1069   what habitat attributes and/or features are limiting fish distributions for those sites where fish

1070   use does not extend to end of current DPCs. These data will include channel gradient, bankfull

1071   width, wetted width and confinement within unequal length segments of relatively uniform

1072   habitat character. The results might suggest appropriate metrics for vertical and non-vertical

1073   obstacles that could be used in conjunction with width and gradient to add an element of

1074   accessibility to the DPCs, thereby improving their accuracy and utility. In particular, this would

1075   reduce the degree to which the current DPCs, when used on their own in the absence of a

1076   protocol survey, predict fish use where there are no fish, and are not likely to ever be.

1077

References

1078

1079 Adams, S. A., C. A. Frissell, and B. E. Rieman. 2000. Movements of non-native brook trout
1080   in relation to stream channel slope. Transactions of the American Fisheries Society
1081   129:623-638.

1082 Andrews, E. D. 1980. Effective and bankfull discharges of streams in the Yampa River Basin,
1083   Colorado and Wyoming. Journal of Hydrology 46:311-330.

1084 Archer, E.K., B.B. Roper, R.C. Henderson, N. Bouwes, S.C. Mellison, J.L. Kershner. Testing
1085   Common Stream Sampling Methods for Broad-Scale, Long-Term Monitoring. General
1086   Technical Report RMRS-GTR-122.

1087 Beamish, F. H. 1978. Swimming capacity. Pages 101-187 *in* W. S. Hoar and D. J. Randall,
1088   editors. Fish physiology, Vol. 7. Academic Press, New York.

1089 Beechie, T., and H. Imaki. 2014. Predicting natural channel patterns based on landscape
1090   and geomorphic controls in the Columbia River basin, USA. Water Resources
1091   Research 50 39-57.

1092 Bell, M. C. 1991. Fisheries handbook of engineering requirements and biological criteria.
1093   U.S. Army Corps of Engineers Office of Chief Engineer, Fish Passage Development and
1094   Evaluation Program, Portland, Oregon.

1095 Bisson, P. A. and R. E. Bilby. 1998. Organic matter and trophic dynamics. Pages 373-398 *in*
1096   R. J. Naiman and R. E. Bilby. River ecology and management: lessons from the Pacific
1097   coastal ecoregion. Springer-Verlag, New York.

1098 Bisson, P. A., K. Sullivan, and J. L. Nielson. 1988. Channel hydraulics, habitat use, and body
1099   form of juvenile coho salmon, steelhead, and cutthroat trout in streams. Transactions
1100   of the American Fisheries Society 117:262-273.

1101 Björkelid, L. 2005. Invasiveness of brook charr (*Salvelinus fontinalis*) in small boreal
1102   headwater streams. Master's Thesis. University of Gothenberg, Umeå, Sweden.

1103 Bryant, M. D., T. Gomi and J. Piccolo. 2007. Structures linking physical and biological
1104   processes in headwater streams of the Maybeso watershed, southeast Alaska. Forest
1105   Science 53:371-383.

1106 Bryant, M. D., N. D. Zymonas, and B. E. Wright. 2004. Salmonids on the fringe: abundance,
1107   species composition, and habitat use of salmonids in high-gradient headwater
1108   streams, southeast Alaska. Transactions of the American Fisheries Society 133:1529-
1109   1538.

1110 Castro, J.M. and P.L. Jackson. 2001. Bankfull discharge recurrence intervals and regional
1111   hydraulic geometry relationships: Patterns in the Pacific Northwest, USA. Journal of

1112 the American Water Resources Association Vol. 37, No. 5, p.1249-1262. October
1113 2001.

1114 Chicco, D., and Jurman, G. 2020. The advantages of the Matthews correlation coefficient
1115 (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics.
1116 21(6).

1117 Clark, C., P. Roni, J. Keeton, and G. Pess. 2020. Evaluation of the removal of impassable
1118 barriers on anadromous salmon and steelhead in the Columbia River Basin. Fisheries
1119 Management and Ecology 27(1):102-110.

1120 Clark, C., P. Roni, and S. Burgess. 2019. Response of juvenile salmonids to large wood
1121 placement in Columbia River tributaries. Hydrobiologia 842(1):173-190.

1122 Cochran, W.G. 1977. Sampling Techniques. John Wiley & Sons, 3rd edition. 413 pp.

1123 Cole, M. B., D. M. Price, and B. R. Fransen. 2006. Change in the upper extent of fish
1124 distribution in eastern Washington streams between 2001 and 2002. Transactions of
1125 the American Fisheries Society 135:634-642.

1126 Cole, M. B. and J. L. Lemke. 2006. Annual and seasonal variability in the upper limit of fish
1127 distribution in Eastern Washington streams – Final Report. Prepared by ABR, Inc. –
1128 Environmental Research and Services under PSC 05-145 for Washington State
1129 Department of Natural Resources, Olympia Washington.

1130 Connolly, P. J., and J. D. Hall. 1999. Biomass of Cutthroat Trout in unlogged and previously
1131 clear-cut basins in the central coast range of Oregon. Transactions of the American
1132 Fisheries Society 128:890-899.

1133 Conrad, R.H., B. Fransen, S. Duke, M. Liermann, and S. Needham. 2003. The development
1134 and assessment of the preliminary model for identifying fish habitat in western
1135 Washington. CMER Report #03-313. Washington State Department of Natural
1136 Resources, Olympia, WA.

1137 CMER. 2020. CMER Strategy for completing Water Typing Study Designs. Memo to the
1138 Forest Practices Board, May 1, 2020. Delivered to and approved by the Washington
1139 Forest Practices Board May 13, 2020. File "bc_fpb_wtstrategy_20200513.pdf" in
1140 Board meeting materials for May 13, 2020 meeting.

1141 Cupp, C. E. 2002. Data collection for development of Eastern Washington water typing
1142 model. Cooperative Monitoring Evaluation and Research Report CMER 01-100.
1143 Washington Department of Natural Resources, Olympia.

1144 Duke, S. 2005. Eastside Model Development. PowerPoint presentation dated March 9,
1145 2005.

1146 Dumelle, Michael., Kincaid, T. M., Olsen, A. R., and Weber, M. H. 2022. spsurvey: Spatial
1147   Sampling Design and Analysis. R package version 5.3.0.

1148 Dunham, J. B., M. M. Peacock, B. E. Rieman, R. E. Schroeter, and G. L. Vinyard. 1999. Local
1149   and geographic variability in the distribution of stream-living Lahontan Cutthroat
1150   Trout. Transactions of the American Fisheries Society 128:875-889.

1151 Fausch, K. D. 1989. Do gradient and temperature affect distributions of, and interactions
1152   between, Brook Charr (*Salvelinus fontinalis*) and other resident salmonids in streams?
1153   Physiological Ecology (Japan) Special Vol. 1:303-322.

1154 FFR (Forest and Fish Report). 1999. Forests and Fish Report. WA DNR. Available at:
1155   http://file.dnr.wa.gov/publications/fp_rules_forestsandfish.pdf. (August 2018).

1156 Franklin, J. F. and J. W. Matthews, editors. Ecological research in National Parks of the
1157   Pacific Northwest. Proceedings of the 2nd Conference on Scientific Research in the
1158   National Parks. Oregon State University, Forest Research Laboratory, Corvallis.

1159 Fransen, B. R., R. E. Bilby, S. Needham, and J. K. Walter. 1998. Delineating fish habitat based
1160   on physical characteristics associated with the upper extent of fish distributions.
1161   Paper presented at the 1998 Annual General Meeting, North Pacific International
1162   Chapter American Fisheries Society, March 18-20, 1998, Union, Washington.

1163 Fransen, B. R., S. D. Duke, L. D. McWethy, J. K. Walter, and R. E. Bilby. 2006. A logistic
1164   regression model for predicting the upstream extent of fish occurrence based on
1165   geographical information systems data. North American Journal of Fisheries
1166   Management 26:960-975.

1167 Goldberg, C. S., K. M. Strickler, and D. S. Pilliod. 2015. Moving environmental DNA methods
1168   from concept to practice for monitoring aquatic macroorganisms. Biological
1169   Conservation 183:1-3.

1170 Gresswell, R. E., C. E. Torgerson, D. S. Bateman, T. J. Guy, S. R. Hendricks, J. E. B. Wofford.
1171   2006. A spatially explicit approach for evalauating relationships among coastal
1172   cutthroat trout, habitat, and disturbance in small Oregon streams. Pages 457-471 *in*
1173   R. M. Hughes, L. Wang, and P. W. Seelbach, editors. Landscape influences on stream
1174   habitats and biological assemblages (Volume 48). American Fisheries Society,
1175   Bethesda, Maryland.

1176 Haggerty, S. M., D. P. Batzer, and C. R. Jackson. 2004. Macroinvertebrate response to
1177   logging in coastal headwater streams of Washington, U.S.A. Canadian Journal of
1178   Fisheries and Aquatic Sciences 61:529–537.

1179 Hammer, C. 1995. Fatigue and exercise tests with fish. Comparative Biochemistry and
1180   Physiology 112A:1-20.

1181 Harrelson, C. C., C. L. Rawlins, J. P. Potyondy. 1994. Stream channel reference sites: An
1182   illustrated guide to field technique. General Technical Report RM-GTR-245. U.S.
1183   Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins,
1184   Colorado.

1185 Harvey, B. C. 1993. Benthic assemblages in Utah headwater streams with and without
1186   trout. Canadian Journal of Zoology 71:896-900.

1187 Hastings, K., C. A. Frissell, and F. W. Allendorf. 2005. Naturally isolated coastal Cutthroat
1188   Trout populations provide empirical support for the 50-500 rule. Pages 121-124 in P.
1189   J. Connolly, T.

1190 Hawkins, C. P., J. L. Kershner, P. A. Bisson, M. D. Bryant, L. M. Decker, S. V. Gregory, D. A.
1191   McCullough, C. K. Overton, G. H. Reeves, R. J. Steedman, and M. K. Young. 1993. A
1192   hierarchical approach to classifying stream habitat features. Fisheries 18(6):3-12.

1193 Hawkins, C. P., and J. R. Sedell. 1981. Longitudinal and seasonal changes in functional
1194   organization of macroinvertebrate communities in four Oregon streams. Ecology
1195   62:387-397.

1196 Hawkins, D. K., and T. P. Quinn. 1996. Critical swimming velocity and associated
1197   morphology of juvenile Coastal Cutthroat Trout (*Oncorhynchus clarki clarki*),
1198   Steelhead Trout (*Oncorhynchus mykiss*), and their hybrids. Canadian Journal of
1199   Fisheries and Aquatic Sciences 53:1487-1496.

1200 Heede, B. H. 1972. Influence of a forest on the hydraulic geometry of two mountain
1201   streams. Water Resources Bulletin 8:523-530.

1202 Hornung, R. 2022. Diversity Forests: Using Split Sampling to Enable Innovative Complex
1203   Split Procedures in Random Forests.SN Computer Science 3:1. doi: 10.1007/s42979-
1204   021-00920-1.

1205 Hornung, R. and L. A. Boulesteix. 2022. Interaction Forests: Identifying and Exploiting
1206   Interpretable Quantitative and Qualitative Interaction Effects. Computational
1207   Statistics and Data Analysis 171: 107460. doi: 10.1016/j.csda.2022.107460.

1208 Huet, M. 1959. Profiles in biology of western European streams as related to fish
1209   management. Transactions of the American Fisheries Society 88:155-163.

1210 Jane, S. F., T. M. Wilcox, K. S. McKelvey, M. K. Young, M. K. Schwartz, W. H. Lowe, B. H.
1211   Letcher, and A. R. Whiteley. 2015. Distance, flow and PCR inhibition: eDNA dynamics
1212   in two headwater streams. Molecular Ecology Resources 15(1):216-27.

1213 Kershner, J. L., E. Archer, R. C. Henderson, and N. Bouwes. 2002. An evaluation of physical
1214   habitat attributes used to monitor streams. Journal of the American Water Resources
1215   Association38:1-10.

1216 Kershner, J. L., P. A. Bisson, P. C. Trotter, P. Roni, R. K. Timm, J. Maroncy, K. Ross, H. Berge.
1217   2018. A review of the role of small tributary streams as salmonid habitat. Report
1218   prepared for the Washington Forest Practices Board. Olympia, Washington.

1219 Kondolf, G. M., G. F. Cada, M. J. Sale, and T. Felando. 1991. Distribution and stability of
1220   potential salmonid spawning gravels in steep boulder-bed streams of the eastern
1221   Sierra Nevada. Transactions of the American Fisheries Society 120:177-186.

1222 Kondratieff, M. C. and C. A. Myrick. 2006. How high can Brook Trout jump? A laboratory
1223   evaluation of Brook Trout jumping performance. Transactions of the American
1224   Fisheries Society 135:361- 370.

1225 Kruse, C. G., W. A. Hubert, and F. J. Rahel. 1997. Geomorphic influences on the distribution
1226   of Yellowstone Cutthroat Trout in the Absaroka Mountains, Wyoming. Transactions
1227   of the American Fisheries Society 126:418-427.

1228 Larsen, D.P., Olsen, A.R. and Stevens Jr, D.L. 2008. Using a master sample to integrate
1229   stream monitoring programs. Journal of Agricultural, Biological, and Environmental
1230   Statistics, pp.243-254.

1231 Latterell, J. J., R. J. Naiman, B. R. Fransen, and P. A. Bisson. 2003. Physical constraints on
1232   trout (*Oncorhynchus* spp.) distribution in the Cascade Mountains: a comparison of
1233   logged and unlogged streams. Canadian Journal of Fisheries and Aquatic Sciences
1234   60:1007-1017.

1235 Leathe, S. A. 1985. Cumulative effects of micro-hydro development on the fisheries of the
1236   Swan River drainage, Montana. III. Fish and habitat inventory of tributary streams,
1237   final report. US Dept. of Energy Bonneville Power Administration, Division of Fish and
1238   Wildlife Contract No. DE-A179-82BP36717, Project 82-19

1239 Leopold, L. B. 1994. A view of the river. Harvard University Press, Cambridge,
1240   Massachusetts.

1241 Liquori, M. 2000. A preliminary examination of the controls on small headwater channel
1242   morphology and habitat influence in managed forests. Poster presented at 10th
1243   Annual Review, Center for Streamside Studies, University of Washington, Seattle.

1244 Loh, W. Y. (2011). Classification and regression trees. Wiley interdisciplinary reviews: data
1245   mining and knowledge discovery, 1(1), 14-23.

1246 Maroco, J., Silva, D., Rodrigues, A. et al. 2011. Data mining methods in the prediction of
1247   Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear
1248   discriminant analysis, logistic regression, neural networks, support vector machines,
1249   classification trees and random forests. BMC Res Notes 4, 299.

1250   Mason, J. C. 1976. Response of underyearling Coho Salmon to supplemental feeding in a
1251   natural stream. Journal of Wildlife Management 40:775-788.

1252   McIntyre, A. P., M. P. Hayes, and T. Quinn. 2009. Type N Feasibility Study. A report submitted
1253   to the Landscape and Wildlife Advisory Group, Amphibian Research Consortium, and the
1254   Cooperative Monitoring, Evaluation, and Research Committee. Washington Department
1255   of Natural Resources, Olympia, Washington.

1256   Montgomery, D. R. 1999. Process domains and the river continuum. Journal of the
1257   American Water Resources Association 35:397-410.

1258   Montgomery, D. R., and J. M. Buffington. 1993. Channel classification, prediction of channel
1259   response and assessment of channel condition. Washington Department of Natural
1260   Resources Report TFW-SH10-93002, Olympia, Washington.

1261   Montgomery, D. R., and J. M. Buffington. 1997. Channel-reach morphology in mountain
1262   drainage basins. Geological Society of America Bulletin 109:596-611.

1263   Moore, S. E., G. L. Larson, and B. Ridley. 1985. Dispersal of brook trout in rehabilitated
1264   streams in Great Smoky Mountains National Park. Journal of the Tennessee Academy
1265   of Science 60:1–4.

1266   Morgan, J. 2014. Classification and regression tree analysis. Boston: *Boston University, 298*.

1267   Mossop, B., and M. J. Bradford. 2006. Using thalweg profiling to assess and monitor juvenile
1268   salmon (*Oncorhynchus* spp.) habitat in small streams. Canadian Journal of Fisheries
1269   and Aquatic Sciences 63:1515–1525.

1270   Naiman, R. J., and J. R. Sedell. 1980. Relationships between metabolic parameters and
1271   stream order in Oregon. Canadian Journal of Fisheries Aquatic Sciences 37:834-847.

1272   Northcote, T. G., and G. F. Hartman. 1988. The biology and significance of stream trout
1273   populations (*Salmo* spp.) living above and below waterfalls. Polish Archives of
1274   Hydrobiology 35(3-4):409- 442.

1275   Omernik, J. M. 1987. Ecoregions of the conterminous United States. Annals of the
1276   Association of American Geographers 77(1):118-125.

1277   PHB Science Panel. 2018. Review and recommendations for potential fish habitat breaks to
1278   begin protocol surveys to determine end of fish habitat on state and private forest
1279   lands in Washington State. Report to the Washington Forest Practices Board, January
1280   16, 2018. Washington Department of Natural Resources, Olympia.

1281   PHB Science Panel. 2019. Evaluation of physical features that define fish habitat in forested
1282   landscapes across Washington State, - [DRAFT] Study plan prepared for the

1283   Washington Forest Practices Board. March 20, 2019. Washington Department of
1284   Natural Resources, Olympia, WA.

1285   Palmquist, R. 2005. Type N stream demarcation study Phase I: Pilot results. Cooperative
1286   Monitoring, Evaluation and Research Report (not numbered), Washington Department
1287   of Natural Resources. Olympia, Washington.

1288   Penaluna, B. E., G. H. Reeves, C. Z. Barnett, P. A. Bisson, J. M. Buffington, C. A. Dolloff, R. L.
1289   Flitcroft, C. H. Luce, K. H. Nislow, J. D. Rothlisberger, M. L. Warren. 2018. Using natural
1290   disturbance and portfolio concepts to guide aquatic–riparian ecosystem
1291   management. Fisheries 43(9):406-422.

1292   Peterson, N. P. 1982. Immigration of juvenile Coho Salmon (*Oncorhynchus kisutch*) into
1293   riverine ponds. Canadian Journal of Fisheries and Aquatic Sciences 39(9):1308-1310.

1294   Peterson, N. P., R. K. Simmons, T. Cardoso, and J. T. Light. 2013. A probabilistic model for
1295   assessing passage performance of coastal cutthroat trout through corrugated metal
1296   culverts. North American Journal of Fisheries Management 33(1):192-199.

1297   Pleus, A., D. Schuett-Hames, and L. Bullchild. 1999. Method manual for the habitat unit
1298   survey.

1299   Timber, Fish, and Wildlife Monitoring Program. Publication No. TFW-AM9-99-004.,
1300   Olympia, Washington.

1301   Powers, P. D., and J. F. Orsborn. 1985. Analysis of barriers to upstream fish migration.
1302   Report submitted to Bonneville Power Administration, Contract DE-A179-
1303   82BP36523, Project 82-14. Washington State University Department of Civil and
1304   Environmental Engineering, Pullman.

1305   Rees, H. C., B. C. Maddison, D. J. Middleditch, J. R. M. Patmore, K. C. Gough, and E. Crispo.
1306   2014.

1307   REVIEW: The detection of aquatic animal species using environmental DNA - a review of
1308   eDNA as a survey tool in ecology. Journal of Applied Ecology 51(5):1450-1459.

1309   Roni, P., K. Ross, H. Berge, P. Bisson, J. Kershner, J. Maroney, P. Trotter, J. Walter. 2018.
1310   Potential habitat break (PHB) pilot study final report. Prepared for Washington
1311   Department of Natural Resources Olympia, Washington. 30 pages.

1312   Roper, B., J. L. Kershner, E. Archer, R. C. Henderson, and N. Bouwes. 2002. An evaluation of
1313   physical habitat attributes used to monitor streams. Journal of the American Water
1314   Resources Association 38:1-10.

1315   Roper, B., J.M. Buffington, S. Bennet, S.H. Lanigan, E. Archer, S.T. Downie, J. Faustini, T.W.
1316   Hillman, S. Hubler, K. Jones, C. Jordan, P.R. Kaufmann, G. Merrit, C. Moyer, A. Pleus.

1317   2010. A comparison of the performance and compatibility of protocols used by seven
1318   monitoring groups to measure stream habitat in the Pacific Northwest. North
1319   American Journal of Fisheries Management. 30:565-587.

1320 Rose, P.M., Kennard, M.J., Moffatt, D.B., Sheldon, F. and Butler, G.L., 2016. Testing three
1321   species distribution modelling strategies to define fish assemblage reference
1322   conditions for stream bioassessment and related applications. PLoS One, 11(1),
1323   p.e0146728.

1324 Rosgen, D. L. 1994. A classification of natural rivers. Catena 22:169-199.

1325 Rosgen, D. L. 1996. Applied river morphology. Wildland Hydrology, Pagosa Springs,
1326   Colorado.

1327 Sedell, J. R., P. A. Bisson, J. A. June, and R. W. Speaker. 1982. Ecology and habitat
1328   requirements of fish populations in South Fork Hoh River, Olympic National Park.
1329   Pages 35-42 *in* E. E. Starkey,

1330 Skeesick, D. G. 1970. The fall Immigration of juvenile Coho Salmon (*Oncorhynchus kisutch*)
1331   into a small tributary. Research Reports of the Fish Commission of Oregon 2:90-95.

1332 Stevens, D. L., and A. R. Olsen. 2003. Variance estimation for spatially balanced samples of
1333   environmental resources. Environmetrics 14:594-610.

1334 Stevens, D. L., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources.
1335   Journal of the American Statistical Association 99(465):262-278.

1336 Stuart, T. A. 1962. The leaping behavior of salmon and trout at falls and obstructions.
1337   Department of Agriculture and Fisheries for Scotland, Freshwater and Salmon
1338   Fisheries Research Report 28. Edinburgh, U.K.

1339 Therneau T, Atkinson B (2022). rpart: Recursive Partitioning and Regression Trees_. R
1340   package version 4.1.19, https://CRAN.R-project.org/package=rpart.

1341 Tompalski, P., N. C. Coops, J. C. White, M. A. Wulder, and A. Yuill. 2017. Characterizing
1342   streams and riparian areas with airborne laser scanning data. Remote Sensing of
1343   Environment 192:73-86.

1344 Torgerson, C. E., R. E. Gresswell, D. S. Bateman, and D. P. Hockman-Wert. 2008. Effects of
1345   landscape pattern on the distribution of coastal cutthroat trout in headwater
1346   catchments in Western Oregon. Coastal Cutthroat Trout Symposium: Status,
1347   Management, Biology, and Conservation. Oregon Chapter, American Fisheries
1348   Society, 2008.

1349   Triska, F. J., J. R. Sedell, and S. V. Gregory. 1982. Coniferous forest streams. Pages 292-332
1350   in R. L. Edmonds, editor. Analysis of coniferous forest ecosystems in the western
1351   United States. Hutchinson Ross, Stroudsburg, Pennsylvania.

1352   Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The
1353   river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130-
1354   137.

1355   Walter, J., et al. (in prep) Temporal variability in the upper extent of fish distributions in
1356   southwest Washington.

1357   Warren, C. E., J. H. Wales, G. E. Davis, and P. Doudoroff. 1964. Trout production in an
1358   experimental stream enriched with sucrose. Journal of Wildlife Management 28:617-
1359   660.

1360   WA DNR. 1997. Forest Practices Board Manual, Standard methodology for conducting
1361   watershed analysis, v. 4.0. Washington Department of Natural Resources, Olympia.

1362   WA DNR. 2002. Forest Practices Board Manual, Section 13, Guidelines for determining fish
1363   use of the purpose of typing waters. Washington Department of Natural Resources,
1364   Olympia.

1365   WA DNR. 2004. Forest Practices Board Manual, Section 2. Standard methods for identifying
1366   bankfull channel features and channel migration zones. Washington Department of
1367   Natural Resources, Olympia.

1368   WA DNR. 2019. Memo from Marc Engel (DNR), to Forest Practices Board, dated April 17,
1369   2019 regarding Water Typing System Rule and Board Manual Development Update.

1370   WA DNR. 2019b. Draft Rule Proposal for a Permanent Water Typing System, dated April
1371   22, 2019. In Forest Practices Board meeting materials for Board meeting of May 8,
1372   2019. https://www.dnr.wa.gov/publications/bc_fpb_mtgmaterial_20190508.pdf

1373   Washington State Forest Practices Board 2017. Minutes from August 2017 Forest Practices
1374   Board   meeting.
1375   https://www.dnr.wa.gov/publications/bc_fpb_mtg_packet_20180213.pdf?p3kh5ut

1376   Washington State Forest Practices Board 2018.   Minutes and meeting materials from
1377   February   2018   Forest   Practices   Board   meeting.
1378   https://www.dnr.wa.gov/publications/bc_fb_mtgminutes_20180213_14.pdf,
1379   https://www.dnr.wa.gov/publications/bc_fpb_mtg_packet_20180213.pdf?p3kh5ut
1380   FPB meeting minutes Feb 2018: "Berge reiterated that a PHB is not necessarily the
1381   Type F/N break, but rather the first point of potential unfavorable habitat and the
1382   starting point for a protocol survey."

1383   Watson, R. 1993. The trout: a fisherman's natural history. Swan Hill Press, Shrewsbury, U.K.

1384   Washington State Forest Practices Board 2019a. Minutes from August 2017 Forest
1385   Practices Board meeting.
1386   https://www.dnr.wa.gov/publications/bc_fpb_mtg_packet_20180213.pdf?p3kh5ut

1387   Watson, G., and T. W. Hillman. 1997. Factor affecting the distribution and abundance of
1388   Bull Trout: and investigation at hierarchical scales. North American Journal of
1389   Fisheries Management 17:237-252.

1390   Watts, F. J. 1974. Design of culvert fishways. University of Idaho Water Resources Research
1391   Institute, Project A-027-IDA, Moscow, Idaho. [Not seen, cited in Bjornn, T. C., and D.
1392   W. Reiser. 1991. Habitat requirements of salmonids in streams. Pages 83-138 *in* W.
1393   R. Meehan, editor. Influence of forest and rangeland management on salmonid fishes
1394   and their habitats. American Fisheries Society Special Publication 19, Bethesda,
1395   Maryland.

1396   Webb, P. W. 1984. Body form, locomotion and foraging in aquatic vertebrates. American
1397   Zoologist 24:107-120.

1398   Whitacre, H. W., B. B. Roper, and J. L. Kershner. 2007. A comparison of protocols and
1399   observer precision for measuring stream attributes. Journal of the American Water
1400   Resources Association 43:923-937.

1401   Wigington Jr., P. J., J. L. Ebersole, M. E. Colvin, S. G. Leibowitz, B. Miller, B. Hansen, H. R.
1402   Lavigne, D. White, J. P. Baker, M. R. Church, J. R. Brooks, M. A. Cairns, and J. E.
1403   Compton. 2006. Coho Salmon dependence on intermittent streams. Frontiers in
1404   Ecology and the Environment 4(10):513-518.

1405   Williams, H, and R. E. Gresswell, editors. 2005. The 2005 Coastal Cutthroat Trout
1406   Symposium: Status, Management, Biology, and Conservation. Port Townsend,
1407   Washington. Oregon Chapter of the American Fisheries Society, Portland.

1408   Wipfli, M. S. 1997. Terrestrial invertebrates as salmonid prey and nitrogen sources in
1409   streams: contrasting old-growth and young-growth riparian forests in southeastern
1410   Alaska, U.S.A. Canadian Journal of Fisheries Aquatic Science 54:1259-1269.

1411   Ziller, J. S. 1992. Distribution and relative abundance of Bull Trout in the Sprague River
1412   subbasin, Oregon. Pages 18-29 *in* P. J. Howell and D. V. Buchanan, editors.
1413   Proceedings of the Gearhart Mountain Bull Trout workshop. Oregon Chapter,
1414   American Fisheries Society, Corvallis.

1415   # Appendix A. CMER Workplan and prior science panel study questions

1416   **CMER Workplan Water Typing Rule Group Critical Questions**

1417   The following are the critical questions of the water typing rule group program this study will

1418   address:

1419   **CQ 1.**   How can the line demarcating fish- and non-fish habitat waters be accurately
1420   identified?

1421   **CQ 2.**   To what extent does the current water typing survey window capture seasonal and
1422   annual variability in fish distribution considering potential geographic differences?

1423   **CQ 3.**   How do different fish species use seasonal habitats (timing, frequency, duration)?

1424   **CQ 4.**   How does the upstream extent of fish use at individual sites vary seasonally and
1425   annually?

1426   **CQ 5.**   How does the delineation of the upstream extent of fish habitat change seasonally?

1427

1428   **Science Panel Document Study Questions**

1429   • Do the PHB criteria provided by the Washington Forest Practices Board accurately capture
1430   the EOF habitat when applied in the Fish Habitat Assessment Methodology (FHAM)?

1431   • Based on data collected, what is the most accurate combination of metrics for
1432   determining PHB by region or ecoregion?

1433   • Are there differences in PHB criteria by Environmental Protection Agency (EPA) Level III
1434   ecoregion, eastern vs western Washington, or some other geographic or landscape
1435   strata?

1436   • Are there additional variables (e.g., geology, drainage area, valley width, land use, channel
1437   type, and stand age) that could improve the accuracy of existing criteria?

1438   • What is the influence of season/timing of survey on PHB identification?

1439   •   What is the typical inter-annual variability in last detected fish and PHBs?

1440   •   Can protocols used to describe PHB be consistently applied among survey crews and be

1441   expected to provide similar results in practice?

1442   •   Answering these questions requires identifying the last detected fish and surveying

1443   habitat above and below these points in a random representative sample of streams

1444   across the state.

1445

1446

1447   Appendix B. Sample Size Estimation Memo of Jan 4, 2022

1448

ENVIRONMENTAL & STATISTICAL CONSULTANTS

2725 NW Walnut Blvd., Corvallis, OR 97330
Phone: 541 738 6198 ⬦ www.west-inc.com

1449

1450

1451   # MEMO

1452

1453   To: Instream Science Advisory Group
1454   From: Leigh Ann Starcevich (WEST, Inc.)
1455   Date: January 4, 2022
1456   Re: Sample size approximation from Eastern WA and Western WA data

1457

1458   The Instream Science Advisory Group (ISAG) is developing a sampling design for surveys of potential
1459   habitat breaks (PHB) for fish use. A sample size approximation is needed to ensure that the data collected
1460   to assess criteria defined by the Washington Forest Practices Board (Board) for the Fish Habitat
1461   Assessment methodology (FHAM) yield useful covariates for PHB modeling. Cooperative Monitoring,
1462   Evaluation, and Research (CMER) data from eastern Washington surveys conducted in 2001, 2002, and
1463   2005 were provided by Chris Mendoza. Stream habitat data associated with uppermost detected fish
1464   points from concurred water type modification forms for surveys conducted in western Washington
1465   between 2016 and 2020 were provided by Weyerhaeuser. These data were used to approximate sample
1466   sizes needed to estimate means of PHB model covariates with desired levels of precision and accuracy.

1467   **Eastern Washington Data**

1468   The eastern Washington data were collected in 2001 by Terrapin Environmental (Cupp 2002) and in 2002
1469   and 2005 by ABR, Inc. Environmental Research & Services (Cole and Lemke 2003, 2006). Channel
1470   characteristic metrics included mean channel widths and means gradients for reaches extending up to
1471   100m above and 100m below the last fish point obtained in the 2001 survey. Data for barriers were
1472   collected but inconsistencies in how barriers were classified and recorded prevented sample size
1473   evaluation specific to barriers. For surveys conducted after 2001, the last fish distance relative to the 2001
1474   last fish was provided. A metric for the maximum change in distance from the 2001 last fish point was
1475   calculated for each site. Using the 2001 point as baseline, the range of distances where the last fish was
1476   observed during subsequent surveys was calculated and used to inform the sample size approximation.

1477   Data screening was used to limit the data set to a subset of locations with natural habitat breaks.
1478   Unscreened data sets included sites where large woody debris jams were found, no surface flow occurred
1479   for at least 100m, and surveys were conducted past July 15. The screened data sets eliminated many of
1480   these sites. Sites where fish passage was limited by culverts were removed from all data sets. About 46%
1481   of the unscreened points were classified as lateral points.

1482   **Western Washington Data**

1483   Water type modification form data from western Washington were collected between 2016 and 2021 and
1484   included gradient and bankfull width metrics for stream segments upstream and downstream of the last

1485 fish point. For many lateral points, only the upstream measurements were provided because the point was
1486 located on a river mainstem. At these points, data on gradient and bankfull width metrics downstream of
1487 the confluence were not always collected, so these points are omitted for sample size calculations based
1488 on the downstream metrics. About 70% of the points were classified as lateral points.

1489 **Sample Size Approximation**

1490 Estimated means of channel characteristic metrics and change in last fish locations among years were
1491 used as the basis for the sample size approximation. Let z reflect the quantile of a standard normal random
1492 variable for a given Type I error rate (α). For $\alpha = 0.10$ we have that $z = 1.645$. Let d be the maximum
1493 absolute error (i.e. confidence interval half-width), let r be the relative precision of the estimate, and let γ
1494 be the coefficient of variation (CV). The coefficient of variation is a standardized measure of precision
1495 calculated as the standard deviation (SD) of the outcome divided by the mean of the outcome (Thompson
1496 2002). The sample size approximation formula below is applied with the mean and standard deviation for
1497 each outcome of interest. The sample size needed to obtain an estimate that is within $100*r$% of the true
1498 mean with probability $1 - \alpha$ was calculated. In other words, the confidence interval half-width of the mean
1499 should be $100*r$% of the true mean. The sample size to accomplish this goal is based on a normal
1500 approximation and calculated as:

1501
$$n = \frac{z^2 \gamma^2}{r^2}.$$

1502 For each outcome of interest from the eastern Washington data sets, the coefficient of variation was
1503 computed from the mean and standard deviation of the screened (Tables 1 through 3) and unscreened
1504 (Tables 4 through 6) data, and sample sizes were approximated for relative precision values of 0.10, 0.15,
1505 0.20, and 0.30. Variation was slightly higher in the unscreened data set, resulting in slightly larger
1506 sample sizes. For the eastern data, the coefficients of variation were higher for terminal points than for
1507 lateral points for the upstream reach gradient, reach gradient difference, and maximum change in distance
1508 (Tables 2 and 3, Tables 5 and 6). The coefficients of variation were higher for lateral points than for
1509 terminal points for downstream reach gradient and downstream bankfull width.

1510 Similar results were observed for the western Washington data. For estimation of mean channel metrics
1511 across point types, coefficients of variation ranged from 0.69 to 0.79 for reach gradient metrics and for the
1512 bankfull width above the point. However, bankfull width measured below the last fish point was less
1513 precise than in the eastern Washington data set with a CV of 1.28 (Table 7). The precision for the gradient
1514 difference was similar to that observed for the eastern Washington data with coefficients of variation near
1515 or above one. For the western data, the coefficients of variation were higher for terminal points than for
1516 lateral points for the reach gradient difference (Tables 8 and 9). The coefficients of variation were higher
1517 for lateral points than for terminal points for reach gradient metrics and the downstream bankfull width.
1518 The higher variability in these metrics suggest larger sample sizes are needed for precise estimation of
1519 means. While mean estimation of channel characteristics is not the ultimate inferential goal, we assume
1520 that samples large enough to provide information on the range of values for each of the potential PHB
1521 modeling covariates will yield a useful data set for modeling.

1522

1523 The maximum change in distance from the eastern data was highly variable and generated large sample
1524 sizes for levels of desired precision. The difference in reach gradient exhibited high variability across both
1525 the eastern and western data sets, and sample sizes needed for precise mean estimation are large. To
1526 obtain relative precision of 0.15, the required sample size is nearly double that calculated for relative
1527 precision of 0.20. Note that the sum of the sample sizes calculated for lateral and terminal points
1528 generally exceeds the sample size calculated from data pooled across point types. This indicates that
1529 overall sample sizes may need to be larger than indicated by the pooled analysis to achieve the same level
1530 of precision for means of channel characteristics for lateral and terminal points.

1531 Table 1: Estimates of means, standard deviations, and coefficients of variation from *screened eastern WA*
1532 *data pooled across point types* with sample size approximations for four levels of relative precision.

Outcome	n	Est. Mean	SD	CV	*r* = 0.10	*r* = 0.15	*r* = 0.20	*r* = 0.30
Reach gradient (%) above LF point	193	21.56	13.98	0.65	114	50	28	13
Reach gradient (%) below LF point	161	10.31	6.73	0.65	115	51	29	13
Reach gradient difference (%)	161	9.96	11.19	1.12	341	152	85	38
Bankfull width (m) above LF point	197	2.14	1.41	0.66	117	52	29	13
Bankfull width (m) below LF point	174	1.84	1.35	0.74	146	65	37	16
Maximum change in distance (m)	121	73.26	186.34	2.54	1751	778	438	195

1533
1534
1535 Table 2: Estimates of means, standard deviations, and coefficients of variation from *screened eastern WA*
1536 *data at lateral point types* with sample size approximations for four levels of relative precision.

Outcome	n	Est. Mean	SD	CV	*r* = 0.10	*r* = 0.15	*r* = 0.20	*r* = 0.30
Reach gradient (%) above LF point	67	24.03	12.36	0.52	72	32	18	8
Reach gradient (%) below LF point	53	8.30	9.25	1.11	336	149	84	37
Reach gradient difference (%)	53	18.30	10.77	0.59	94	42	23	10
Bankfull width (m) above LF point	74	1.42	0.79	0.55	83	37	21	9
Bankfull width (m) below LF point	64	0.83	0.74	0.89	214	95	53	24
Maximum change in distance (m)	13	72.12	72.49	1.01	273	121	68	30

1537
1538
1539 Table 3: Estimates of means, standard deviations, and coefficients of variation from *screened eastern WA*
1540 *data at terminal point types* with sample size approximations for four levels of relative precision.

Outcome	n	Est. Mean	SD	CV	*r* = 0.10	*r* = 0.15	*r* = 0.20	*r* = 0.30
Reach gradient (%) above LF point	126	20.25	14.64	0.72	141	63	35	16
Reach gradient (%) below LF point	108	11.30	4.81	0.43	49	22	12	5
Reach gradient difference (%)	108	5.87	8.92	1.52	624	277	156	69
Bankfull width (m) above LF point	123	2.57	1.52	0.59	95	42	24	11
Bankfull width (m) below LF point	110	2.43	1.28	0.53	75	34	19	8
Maximum change in distance (m)	108	73.40	195.84	2.67	1926	856	481	214

1541
1542
1543

1544   Table 4: Estimates of means, standard deviations, and coefficients of variation from *unscreened eastern*
1545   *WA data pooled across point types* with sample size approximations for four levels of relative precision
1546   (recommended eastern WA sample size in bold).

Outcome	n	Est. Mean	SD	CV	*r =* 0.10	*r =* 0.15	*r =* 0.20	*r =* 0.30
Reach gradient (%) above LF point	268	18.73	13.30	0.71	136	61	34	15
Reach gradient (%) below LF point	227	9.72	6.42	0.66	118	52	29	13
Reach gradient difference	227	8.13	10.23	1.26	428	190	107	48
Bankfull width (m) above LF point	282	2.02	1.47	0.73	**143**	63	36	16
Bankfull width (m)below LF point	264	1.59	1.30	0.81	179	79	45	20
Maximum change in distance (m)	153	74.21	172.56	2.33	1463	650	366	163

1547
1548

1549   Table 5: Estimates of means, standard deviations, and coefficients of variation from *unscreened eastern*
1550   *WA data at lateral point types* with sample size approximations for four levels of relative precision.

Outcome	n	Est. Mean	SD	CV	*r =* 0.10	*r =* 0.15	*r =* 0.20	*r =* 0.30
Reach gradient (%) above LF point	104	19.65	12.76	0.65	114	51	29	13
Reach gradient (%) below LF point	83	7.90	8.22	1.04	293	130	73	33
Reach gradient difference (%)	83	13.65	10.92	0.80	173	77	43	19
Bankfull width (m) above LF point	129	1.38	0.81	0.59	93	41	23	10
Bankfull width (m) below LF point	116	0.72	0.71	0.98	261	116	65	29
Maximum change in distance (m)	14	67.89	71.42	1.05	299	133	75	33

1551
1552

1553   Table 6: Estimates of means, standard deviations, and coefficients of variation from *unscreened eastern*
1554   *WA data at terminal point types* with sample size approximations for four levels of relative precision.

Outcome	n	Est. Mean	SD	CV	*r =* 0.10	*r =* 0.15	*r =* 0.20	*r =* 0.30
Reach gradient (%) above LF point	164	18.15	13.64	0.75	153	68	38	17
Reach gradient (%) below LF point	144	10.77	4.83	0.45	55	24	14	6
Reach gradient difference (%)	144	4.94	8.31	1.68	765	340	191	85
Bankfull width (m) above LF point	153	2.55	1.67	0.65	115	51	29	13
Bankfull width (m) below LF point	148	2.28	1.24	0.55	80	36	20	9
Maximum change in distance (m)	139	74.85	179.75	2.40	1561	694	390	173

1555
1556
1557

1558

1559 Table 7: Estimates of means, standard deviations, and coefficients of variation from *western Washington*
1560 *WTMF data pooled across point types* with sample size approximations for four levels of relative
1561 precision (recommended western WA sample size in bold).

Outcome	n	Est. Mean	SD	CV	r = 0.10	r = 0.15	r = 0.20	r = 0.30
Reach gradient (%) above LF point	1982	17.59	13.97	0.79	**171**	76	43	19
Reach gradient (%) below LF point	1512	5.96	4.13	0.69	130	58	32	14
Reach gradient difference (%)	1505	10.79	13.39	1.24	416	185	104	46
Bankfull width above LF point	1900	1.00	0.76	0.76	157	70	39	17
Bankfull width below LF point	1502	4.18	5.79	1.38	518	230	130	58

1562
1563
1564 Table 8: Estimates of means, standard deviations, and coefficients of variation from *western Washington*
1565 *WTMF data at lateral point types* with sample size approximations for four levels of relative precision.

Outcome	n	Est. Mean	SD	CV	r = 0.10	r = 0.15	r = 0.20	r = 0.30
Reach gradient (%) above LF point	1393	19.65	15.45	0.79	167	74	42	19
Reach gradient (%) below LF point	921	4.23	2.81	0.66	119	53	30	13
Reach gradient difference (%)	916	15.13	14.86	0.98	261	116	65	29
Bankfull width (m) above LF point	1318	0.81	0.54	0.67	121	54	30	13
Bankfull width (m) below LF point	913	5.90	6.86	1.16	367	163	92	41

1566
1567
1568 Table 9: Estimates of means, standard deviations, and coefficients of variation from *western Washington*
1569 *WTMF data at terminal point types* with sample size approximations for four levels of relative precision.

Outcome	n	Est. Mean	SD	CV	r = 0.10	r = 0.15	r = 0.20	r = 0.30
Reach gradient (%) above LF point	589	12.71	7.60	0.60	97	43	24	11
Reach gradient (%) below LF point	591	8.65	4.41	0.51	70	31	18	8
Reach gradient difference (%)	589	4.06	6.34	1.56	661	294	165	73
Bankfull width (m) above LF point	582	1.44	0.98	0.68	125	55	31	14
Bankfull width (m) below LF point	589	1.53	0.92	0.61	99	44	25	11

1570
1571

1572 Initial results from the sample size approximation (Tables 1 through 9) suggested to the ISAG subgroup
1573 that upstream metrics provided a robust basis for sample size approximation. Upstream gradient and
1574 bankfull width metrics were consistently measured and are ecologically meaningful for both point types,
1575 were available for both eastern and western WA data, and were the most precise among the channel
1576 characteristics examined. Furthermore, the subgroup also decided to use the unscreened data for sample
1577 size approximations based on eastern WA data because the metrics were slightly more variable in this
1578 data set and provide more conservative sample sizes.

1579 To obtain an overall statewide sample size that accounted for variation across the state, the unscreened
1580 eastern data and the western data were pooled. Coefficients of variation for estimates of means of both
1581 upstream metrics were computed to generate statewide sample sizes across both point types (Table 10),
1582 for lateral points (Table 11), and for terminal points (Table 12). From this analysis, a conservative
1583 statewide minimal sample size of surveyed sites to provide relative precision of 0.10 is obtained from the

1584 upstream bankfull width approximation of 190 sites (Table 10). Assuming that the proportion of sites
1585 classified as lateral points is similar to the proportion observed in the eastern WA data set (46%) and
1586 western WA data set (70%), we can expect roughly 87 to 133 lateral sites and 57 to 103 terminal sites
1587 from this sample of 190 sites. These sample sizes within each point type should be sufficient to obtain
1588 means of the two upstream metrics with at least 0.15 relative precision (Tables 11 and 12).

1589
1590 Table 10: Estimates of means, standard deviations, and coefficients of variation from *pooled eastern and*
1591 *western Washington data at all point types* with sample size approximations for four levels of relative
1592 precision.

Outcome	n	Est. Mean	SD	CV	*r =* 0.10	*r =* 0.15	*r =* 0.20	*r =* 0.30
Reach gradient (%) above LF point	2250	17.73	13.89	0.78	166	74	42	18
Bankfull width (m) above LF point	2182	1.13	0.95	0.84	190	84	47	21

1593
1594
1595 Table 11: Estimates of means, standard deviations, and coefficients of variation from *pooled eastern and*
1596 *western Washington data at lateral point types* with sample size approximations for four levels of relative
1597 precision.

Outcome	n	Est. Mean	SD	CV	*r =* 0.10	*r =* 0.15	*r =* 0.20	*r =* 0.30
Reach gradient (%) above LF point	1497	19.65	15.28	0.78	164	73	41	18
Bankfull width (m) above LF point	1447	0.86	0.59	0.69	129	57	32	14

1598
1599
1600 Table 12: Estimates of means, standard deviations, and coefficients of variation from *pooled eastern and*
1601 *western Washington data at terminal point types* with sample size approximations for four levels of
1602 relative precision.

Outcome	n	Est. Mean	SD	CV	*r =* 0.10	*r =* 0.15	*r =* 0.20	*r =* 0.30
Reach gradient (%) above LF point	753	13.90	9.52	0.69	127	56	32	14
Bankfull width (m) above LF point	735	1.67	1.24	0.74	149	66	37	17

1603

1604 This analysis provides guidance for establishing the sample size of sites for PHB surveys in eastern and
1605 western Washington. If the data sets that were provided are not representative of the larger population of
1606 PHBs in Washington, then variation may be underestimated causing approximated sample sizes to be
1607 lower than needed for the desired precision. The unscreened CMER data were used for the sample size
1608 approximation because they provided more conservative sample sizes than when the screened data were
1609 used. However, this application does not imply a preference for the unscreened data set relative to other
1610 analyses. Differences in site selection for eastern and western Washington data sets were not considered
1611 when pooling the data, but the combined data set provided an index of statewide variability that was not
1612 available otherwise. While the ultimate goal of this project is to identify criteria with which to identify
1613 PHBs, ensuring that the data collected on potential PHB criteria represent the range of conditions in the
1614 population will provide a robust basis for PHB modeling when three years of data are available.

1615
1616

1617   **Sampling Design Recommendations**

1618   Probabilistic selection of the sampling locations from the sampling frame is recommended to avoid
1619   selection bias and to provide a basis for inference to the larger population of interest (Lohr 2009). For
1620   ecological surveys, spatially-balanced sampling approaches provide methods to obtain probabilistic
1621   samples across large areas without risking selection of clustered points that are correlated and provide
1622   duplicate information. Several methods for selecting spatially-balanced samples are available and include
1623   generalized random tessellation stratified (GRTS) sampling (Stevens and Olsen 2003, 2004), balanced
1624   acceptance sampling (BAS; Robertson et al. 2013), and Halton iterative partitioning (HIP, Robertson et
1625   al. 2018). Data from samples selected with spatially-balanced sampling can be analyzed with design-
1626   based tools available in the *spsurvey* package (Dumelle et al. 2022). All three of the sampling techniques
1627   can be implemented in the *SDraw* package (McDonald and McDonald 2020). However, since the *SDraw*
1628   package is currently not maintained on the CRAN website (as of 12/6/21 and since 11/16/21), drawing
1629   GRTS samples with the *spsurvey* package is recommended to ensure that best practices for security
1630   protocols and package functionality are maintained.

1631   The sampling design for the PHB surveys will incorporate *a priori* geographic stratification by region
1632   (east or west WA) so that spatial balance is obtained for each region. Additionally, sampling effort will be
1633   apportioned among point types (terminal or lateral points) with "soft stratification" (Larsen et al. 2008,
1634   section 2). This approach is useful when the point types are not known for each site before the survey so
1635   no sampling frame is available to identify each subpopulation for a priori stratification. Survey crews will
1636   record the point type at the time of the survey and, when the desired sample size for a point type is
1637   satisfied, survey data from this point type will not be collected at subsequent points of this type. Because
1638   the point type is not known a priori so cannot be included as a survey design variable for stratification,
1639   employing this technique will require adherence to the spatially-balanced ordered list of sites to ensure
1640   that the obtained sample of sites within each point type is also spatially balanced. The point type should
1641   be recorded for each site so that inclusion probabilities for each site may be calculated prior to analysis
1642   for any design-based summaries such as means and totals (Larsen et al. 2008, section 2.4).

1643   Based on the sample size approximation for data pooled across region, the total sample size should be no
1644   less than 190 sites (Table 10) to obtain relation precision of 0.10 for the statewide estimates of mean
1645   channel characteristics. ISAG members expressed a desire to obtain estimates of means for channel
1646   characteristics with geographic stratum-level relative precision of 0.10. For the two metrics of interest
1647   (reach gradient above LF point and bankfull width above LF point), obtaining the more conservative
1648   sample size for each region is recommended. Therefore, the eastern WA sample should consist of 143
1649   sites (Table 4) and the western WA sample should consist of 171 sites (Table 7) for a total of 314 sites
1650   across the state.

1651   Given the ISAG statement that there are roughly five times more lateral points than terminal points, I
1652   examined methods to allocate sampling effort among the two point types. Proportional allocation of effort
1653   will favor lateral points since they exist more frequently throughout the landscape. Optimal allocation
1654   accounts for the relative precision of lateral and terminal points but is still influenced by the larger
1655   relative frequency of lateral points as compared to terminal points. The final sample sizes were based on
1656   reach gradient above LF point in eastern WA and bankfull width above LF point in eastern WA. The
1657   precision in the means for these two sets of estimates were similar between lateral and terminal point

1658   types. Therefore, I recommend an equal allocation of sampling effort among the two point types. Based
1659   on the sample size approximation of lateral and terminal points for eastern and western WA (Tables 5, 6,
1660   8, and 9), equal allocation of effort between the two point types should still provide channel characteristic
1661   means with relative precision between 0.10 and 0.15.

1662   Note that the suggested sample sizes are the numbers of sites where data are successfully collected. To
1663   account for inaccessible sites and sites that do not meet the definition of the target population (such as in
1664   reaches with no water), a larger sample of sites (perhaps three to five times larger than the desired sample
1665   size) should be drawn to successfully collect data at the desired number of sites. There is no penalty for
1666   selecting a much larger sample than needed, but the final set of surveyed sites should consist of a
1667   contiguous set of sites from the spatially-balanced randomized list of locations to avoid any sort of
1668   systematic or geographic bias in the sample locations caused by surveying a disproportionate number of
1669   sites in one area. For each site visited, notes on any frame error or nonresponse error should be recorded
1670   so that inclusion probabilities for each site can be accurately calculated. For model-based analysis
1671   approaches, incorporating design variables such as *a priori* and soft stratification variables such as region
1672   and point type (lateral or terminal) may account for the sampling design without directly incorporating
1673   inclusion probabilities.

1674

1675 **References**
1676 Cole, M.B. and J.L. Lemke. 2003. Eastern Washington Last Fish Variability Characterization Resurvey.
1677 Prepared by ABR, Inc. Environmental Research & Services, Forest Grove, Oregon for Washington
1678 Department of Natural Resources, Olympia, WA. 49 pp.
1679
1680 Cole, M.B. and J.L. Lemke. 2006. Annual and Seasonal Variability in the Upper Limit of Fish
1681 Distribution in Eastern Washington Streams. Prepared by ABR, Inc. Environmental Research & Services,
1682 Forest Grove, Oregon for Washington Department of Natural Resources, Olympia, WA. 49 pp.
1683
1684 Cupp, E. 2002. Data Collection for Development of Eastern Washington Water Typing Model.
1685 Unpublished report by Terrapin Environmental, Twisp, WA, for the Washington Department of Natural
1686 Resources, Olympia, WA. 11 pp.
1687
1688 Kincaid, T. M., Olsen, A. R., and Weber, M. H. (2019). spsurvey: Spatial Survey Design and Analysis. R
1689 package version 4.1.0.
1690
1691 Larsen, D.P., Olsen, A.R. and Stevens Jr, D.L. 2008. Using a master sample to integrate stream
1692 monitoring programs. Journal of Agricultural, Biological, and Environmental Statistics, pp.243-254.
1693
1694 Lohr, S.L. 2009. Sampling: Design and Analysis. Boston: Brooks/Cole.
1695
1696 McDonald, T. and A. McDonald. 2020. SDraw: Spatially Balanced Samples of Spatial Objects. R
1697 package version 2.1.13. https://CRAN.R-project.org/package=SDraw.
1698
1699 Robertson, B. L., J. A. Brown, T. McDonald, and P. Jaksons. 2013. BAS: Balanced acceptance sampling
1700 of natural resources. Biometrics 69:776-784.
1701
1702 Robertson, B., McDonald, T., Price, C. and Brown, J., 2018. Halton iterative partitioning: spatially
1703 balanced sampling via partitioning. Environmental and Ecological Statistics, 25(3), pp.305-323.
1704
1705 Stevens, D. L., and A. R. Olsen. 2003. Variance estimation for spatially balanced samples of
1706 environmental resources. Environmetrics 14:594-610.
1707
1708 Stevens, D. L., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. Journal of the
1709 American Statistical Association 99(465):262-278.
1710
1711 Thompson, S. K. 2002. *Sampling. Second edition*. John Wiley and Sons, Inc., Hoboken, New Jersey.

Identifying Potential Habitat Breaks in Washington Streams Using Random Forest Modeling

---

Prepared for:

Washington Department of Natural Resources

Olympia, Washington

Prepared by:

Jared Swenson and Leigh Ann Starcevich

Western EcoSystems Technology, Inc.
415 W. 17ᵗʰ Street, Suite 200
Cheyenne, Wyoming 82001

July 21, 2022

1792

1793   **LIST OF FIGURES**

1873
1874
1875   **LIST OF APPENDICES**

1876   Appendix A. Additional Figures

1877   Appendix B. Modeling Covariate Data Dictionary

1878
1879
1880

1881 **OBJECTIVES**

1882 The Washington Department of Natural Resources (WA DNR) is developing a survey protocol to
1883 identify physical characteristics associated with fish habitat breaks in Washington streams. In
1884 addition to developing criteria for identifying potential habitat breaks (PHBs), the Instream
1885 Scientific Advisory Group (ISAG) would like to evaluate criteria proposed by the Washington
1886 Forest Practices Board (FPB or Board). The goal of this analysis is to characterize the features
1887 associated with the end of fish occurrence in each stream. The goals of this pilot data analysis
1888 are to demonstrate methods for identifying PHBs and assessing FPB criteria. ISAG provided pilot
1889 data from streams in eastern and western Washington to facilitate an example analysis to identify
1890 the end of fish in each stream.

1891

1892 This pilot data analysis demonstrates several tools available for characterizing the end of fish
1893 based on stream segments classified as fish bearing (fish) or non-fish bearing (no fish). The end
1894 of fish is where electrofishing has identified the last fish segment, all waters upstream are thus
1895 non-fish bearing segments. The space between the sampling segment at the end of fish and the
1896 subsequent segment contains the potential habitat barrier, either as a segment level variable or
1897 a cumulative variable. A random forest analysis (Cutler et al. 2007) was applied to segment-level
1898 stream data to model fish presence as a function of habitat feature metrics. Random forest
1899 modeling generates a predictive model that can be accurately applied to novel datasets.
1900 Additionally, interaction forest models were applied to accommodate multivariate comparisons of
1901 habitat covariates that may exhibit relatively strong interactions. Random forest models were
1902 developed with R statistical software (2022) packages to evaluate the Board criteria that included
1903 binary categorical variables of stream characteristics, including gradient, width, obstacles, and
1904 other physical stream characteristics that affect or limit fish dispersal further upstream. For this
1905 objective, we trained a separate random forest model for each of three FPB-proposed PHB
1906 groups identifying criteria options for PHBs based on barrier, gradient, and width criteria, and a
1907 model for all seven unique criteria combined.

1908

1909 Random forest methodology does not explicitly identify the location of the end of fish nor exact
1910 thresholds, but stream metrics that are cumulative over multiple segments above or below a given
1911 segment can be used to explain habitat relationships with fish distribution at a broader scale rather
1912 than only at the segment scale. Additionally, Classification and Regression Tree (CART) models
1913 were developed based on the results of the random forest and interaction forest analyses and
1914 Board criteria to establish thresholds representing potential habitat barriers.

1915 **METHODS**

1916 **Pilot Data and Covariates**

1917 The pilot data set used for analysis included measurements from 2,313 stream segments
1918 representing 32 stream reaches across 11 basins, spanning western and eastern Washington
1919 and five ecoregions (Eastern: Canadian Rocky Mountains, East Cascades; Western: Northwest
1920 Coast Ecoregion [under the purview of WA DNR], Puget Trough, and West Cascades). Stream

1921   segments are defined as the stretch of stream between two survey stations, which are located at
1922   inflection points in the topography of the stream thalweg (Roni et al. 2018). Segment-level habitat
1923   metrics were provided for the random forest analysis. To expand the scale of habitat metrics for
1924   the predictive models, several covariates used in the analysis aggregate data from continuous
1925   groups of segments either upstream or downstream of the segment of interest. Examples include
1926   the maximum gradient upstream of a particular segment and the average sustained gradient of
1927   the 20 segments upstream from the segment of interest. We assessed the correlation between
1928   variables to eliminate covariate combinations that were highly correlated and redundant (Table 1)
1929   to avoid bias in variance importance metrics (Strobl et al. 2007, 2008), but retained all variables
1930   when not included in the same model. Individual stream segments were classified as fish bearing
1931   (Fish) or non-fish bearing (No-fish). The point at which the last fish was detected is the end of fish
1932   (EOF).
1933

Table 1. Details of which stream characteristics were correlated (>0.6). All characteristics were retained in this demonstration analysis to help determine which variables may be important for data collection.

Variable 1	Variable 2	Correlation
Eff.Step.Ht.m	Eff.Step.Ht.BFW	0.88
Eff.Grad	DelEff.Grad.Dn	0.72
Avg.Sus.Grad.Up	Del.Sus.Grad.UpDn	0.70
Avg.Sus.Grad.Dn	Max.Dn.Grad	0.65
Max.Dn.Grad	Max.Dn.Step.BFW10	0.63
Max.Up.Grad	Max.Up.Step.BFW10	0.63

Avg= average; BFW = bankfull width; BFW10 = ?; DelEff = change in effective; Dn = downstream; Eff = effective; Grad = gradient; Ht = height; m = meter; Step = Segment step:Sus = sustained; Up = upstream

1934

Random Forest Models

1935

1936   Random forest classification models can predict binary outcomes such as stream segments with
1937   fish or without fish, can accommodate both continuous and categorical (including binary)
1938   covariates, and are useful in identifying important covariates from covariates sets with substantial
1939   interactions (Cutler et al. 2007). Random forest does not explicitly identify the end of fish based
1940   on habitat characteristics, but provides a method for identifying variables that describe the binary
1941   state of a stream segment that does or does not contain fish. Here the random forest model is
1942   applied to determine variables of interest for use in the CART models and assess variation in
1943   variables of importance across the state of Washington.
1944

1945   Using a random forest model requires training and testing (validation) before applying the model
1946   to novel data sets. We trained a number of models and evaluated model performance to provide
1947   accurate prediction at different spatial scales. In this process, we used the full data set across
1948   Washington and split the data into east and west subsets to determine how transferrable the
1949   model might be across the entire state. For the first approach, we trained the model on a random
1950   subset of 80% of all stream segments across the Washington State dataset. The remaining
1951   segments were used for validation. This statewide *Full Random* model was compared to a model
1952   that was trained on all streams but one, which is referred to as *Full Random Leave One Out (LOO)*

1953 approach. The segments from the "left out" stream were used for model validation. We also
1954 compared the *Full Random* model performance to a model that incorporated geographic
1955 west/east as a predictor variable (*Full Random WE Predictor*). We performed the same routine
1956 for both the western (*Western Random and Western Random* LOO) and eastern (*Eastern*
1957 *Random* and *Eastern Random LOO*) regions in Washington. All models initially included
1958 categorical variables for streambed substrate and habitat unit type.
1959
1960 Random forest models cannot accommodate missing values in covariates. The *randomForest*
1961 package (Liaw and Wiener 2002) can impute these values based on the mean of other correlated
1962 covariates; however, this is not appropriate for this data set. Values were missing for the upstream
1963 gradient of the last segment along the stream and for step-related covariates where no step was
1964 observed. In order to include the last segment of each stream, the gradient was set to zero. This
1965 corresponded to the trajectory of most streams, and several segments had several zero values
1966 prior to the last segment. Missing values for step-related covariates were also set to zero following
1967 the logic that a stream missing a step has a step height of zero. The *Full Random* model includes
1968 these covariates, whereas the *Full Random Reduced Covariates* model excludes the variables
1969 with missing values. This comparison may help in determining the suite of variables important for
1970 future data collection. All eastern and western models included the same covariates as the *Full*
1971 *Random* model because the *Full Random* model performed better than the *Full Random Reduced*
1972 *Covariates* model (Table 2).
1973

Table 2.   Tuning parameters obtained from package *caret*. Model performance evaluated with validation
testing.

	mtry	Maxnodes	Number of Trees	AUC	Accuracy (PCC)	Sensitivity	Specificity	Kappa
Full Random Reduced Covariates	10	26	250	0.87	85.53%	0.92	0.82	0.69
Full Random	11	29	250	0.93	93.52%	0.91	0.96	0.87
Full Random (WE Predictor)	7	29	350	0.90	89.41%	0.91	0.88	0.78
Full Random (LOO)	12	24	250	0.86	82.14%	0.73	1.00	0.65
Eastern Random	6	24	250	0.93	92.83%	0.91	0.94	0.86
Eastern Random (LOO)	12	25	250	0.69	61.19%	0.8	0.58	0.20
Western Random	11	15	250	0.93	92.92%	0.91	0.94	0.85
Western Random (LOO)	5	19	250	0.86	86.08%	0.73	1.00	0.72

AUC = area under the curve; kappa = a measure of agreement between predicted presences and absences; LOO =
Leave One Out; Maxnodes = maximum number of nodes; mtry = optimum number of covariates; PCC = proportion
of presence correctly classified; sensitivity = proportion of presence correctly classified; specificity = the proportion
of absence correctly classified

1974
1975 Each model was built and tuned to maximize accuracy using the R package *caret* (Kuhn 2008)
1976 and trained and validated using *randomForest* (Liaw and Wiener 2002). We determined the
1977 optimum number of covariates allowed at each node (*mtry*), the number of trees, and the
1978 maximum number of nodes (*max nodes*) by comparing the accuracy of the model with varying
1979 values of *mtry, number of trees,* and *max nodes*. Parameters were tuned for each data subset
1980 described in the previous section.
1981

1982 For final model evaluation and comparison, we reported the area under the curve (AUC) to
1983 compare model performance, accuracy (overall percentage correctly classified), sensitivity
1984 (proportion of presence correctly classified), specificity (the proportion of absence correctly
1985 classified), and kappa (a measure of agreement between predicted presences and absences).
1986 Variables deemed important by random forest are displayed graphically along with partial
1987 dependency plots for all continuous variables. To further validate the variables deemed important
1988 in *randomForest*, we used the package *Boruta* as a secondary way to characterize important
1989 variables for each model (Kursa and Rudnicki 2010). To increase the utility of this demonstration,
1990 an appendix of box plots and violin plots were produced to qualitatively visualize potential criteria
1991 cutoffs for variables deemed important by random forest analyses (see Appendix A).

Interaction Forest Models

1993 The random forest approach described above does not explicitly account for interactions between
1994 covariates that can influence categorical outcomes (Hornung and Boulesteix 2022). To investigate how
1995 interactions between stream features effect the predictive capacity of the model, we fit an interaction
1996 forest model using the *Full Random* training data set. We used the R package *diversityForest*
1997 (Hornung 2022) to train an interaction forest and R package *iml* (Molnar et al. 2018) to visualize
1998 interactions between covariates. The package *diversityForest* uses bivariate splitting to model
1999 quantitative and qualitative interaction effects. The effect importance measure (EIM) is produced to rank
2000 variable pairs with respect to their predictive importance. The pairs with the highest EIM are displayed
2001 through contour plots and cross section plots based on a 2-dimensional LOESS fit. Additionally,
2002 graphical output for the overall strength of interactions for all pairs was produced using the *iml* package
2003 in R. Overall interaction strength is calculated using Friedman's H-statistic (Friedman and Popescu
2004 2008). The H-statistic quantifies the share of variance that is explained by the interaction and represents
2005 the strength, but not the direction, of the interaction.

Evaluating Forest Practice Board proposed Potential Habitat Break Criteria

2007 To evaluate the FPB-proposed PHB criteria for end of fish habitat designation (Table 3), we used the
2008 pilot data to compare observed fish presence to predicted fish presence for four sets of criteria. The FPB
2009 criteria options A, B, and C consist of seven unique criteria overall. Each of the seven unique criteria
2010 was calculated from the pilot data as a binary indicator that the criterion was met. The FPB criteria
2011 options A, B, and C were based on the specific combinations of test criteria within each Fish Habitat
2012 Assessment Methodology (FHAM) Rule Option as outlined in Table 3. Additionally, a fourth criteria set
2013 that included all seven unique test criteria was examined. Each of the four criteria sets was used to
2014 predict fish presence and the results were compared to the observed fish data. A confusion matrix of
2015 results, AUC, accuracy, sensitivity, and specificity are reported for each of the criteria sets. See
2016 Appendix B for covariate definitions used in the assessment of FPB criteria.

Table 3.   List of draft Fish Habitat Assessment Methodology rule criteria (presented Washington Department of Natural Resources 2019) translated to metrics/variable names used for pilot analysis. The Forest Practices Board (FPB) Manual definition of bankfull width (BFW) as that for 10 times average BFW is used throughout unless specified otherwise. See Appendix B for variable definitions.

FHAM PHB Option	FHAM Draft Rule Line#	Criterion Type	FHAM Criterion Description	Criterion Description Translated to Pilot Data Variables	Test Criterion #
A	3-a-i	Gradient	Sustained gradient increase >= 5%; sustained = over 20*BFW	(AvgSusGradUpstrm- AvgSusGradDnstrm) >= 0.05	1
A	3-a-ii	Width	Bankfull width <= 2 feet (ft), sustained over 20*BFW	BFW_Up20_ft <= 2.0	2
A	3-a-iii-A	Obstacle	Vertical obstacle height >= BFW AND >= 3 ft	EffectiveGrad_pct > 150% AND EffectiveStepHeight_m >= (3*.3048) AND EffectiveStepHeight_BFW >=1.0	3
A	3-a-iii-B	Obstacle	Non-vertical step >= 30% AND elevation increase > 2*BFW	EffectiveGrad_pct >= 0.3 AND EffectiveStepHeight_BFW > 2.0	4
B	3-a	Gradient	Gradient >10%, sustained over 20 * BFW	AvgSusGradUpstrm > 10%	5
B	3-b (same as A 3-a-ii)	Width	Bankfull width <= 2 ft, sustained over 20*BFW	*See above*	
B	3-c-i (same as A 3-a-iii-A)	Obstacle	Vertical obstacle height >= BFW AND >= 3 ft	*See above*	
B	3-c-ii	Obstacle	Non-vertical step >= 20% gradient AND elevation increase >= upstream BFW	EffectiveGrad_pct >= 0.2 AND EffectiveStepHeight_m > BFW_Up10_m	6
C	3-i (same as A 3-a-i)	Gradient	Sustained gradient increase >= 5%; sustained for >= 20 * BFW	*See above*	
C	3-ii	Width	[Downstream to Upstream] BFW decrease >20%, sustained over 20 * BFW (at tributary junctions)	(BFW_Up20_m/BFW_Dn10_ m) < 0.8	7

Table 3. List of draft Fish Habitat Assessment Methodology rule criteria (presented Washington Department of Natural Resources 2019) translated to metrics/variable names used for pilot analysis. The Forest Practices Board (FPB) Manual definition of bankfull width (BFW) as that for 10 times average BFW is used throughout unless specified otherwise. See Appendix B for variable definitions.

FHAM PHB Option	FHAM Draft Rule Line#	Criterion Type	FHAM Criterion Description	Criterion Description Translated to Pilot Data Variables	Test Criterion #
C	3-iii-A (same as A 3-a-iii-A)	Obstacle	Vertical obstacle height >= BFW AND > 3 feet	*See above*	
C	3-iii-B (same as B 3-c-ii)	Obstacle	Non-vertical step >= 20% gradient, and elevation increase >= upstream BFW	*See above*	
A, B, C		Tributary Jctn	Tributary junctions must meet one of the other PHB criteria	*none*	

*(4) For purposes of this section:

(a) "Permanent Natural Obstacle" means a natural, non-deformable obstacle that completely blocks upstream fish movement. "Permanent natural obstacles" include vertical drops, steep cascades, bedrock sheets and bedrock chutes. A permanent natural obstacle excludes large woody debris and sedimentary deposits.

(b) "Potential Habitat Break" means a permanent, distinct and measurable change to in-stream physical characteristics. PHBs are typically associated with underlying geomorphic conditions and may consist of natural obstacles that physically prevent fish access to upstream reaches or a distinct measurable change in channel, bankfull width or a combination of the two.

BFW = bankfull width; FHAM = Fish Habitat Assessment Methodology; Jctn = junction; PHB = Potential Habitat Break; pct = Percent; Upstrm = Upstream.

2018

2019 As a more robust comparison, we trained and tested four separate random forest models using
2020 the *Full Random* training approach and validation datasets described above and in Table 3. For
2021 each of the four criteria sets the original dataset was altered to contain the fish/no-fish
2022 classification column and a binary feature column; one column for each of the criteria within each
2023 set as outlined in Table 3. The *Boruta* package was used to validate variable importance. The
2024 model AUC, accuracy, sensitivity, specificity, and kappa are reported to evaluate model
2025 performance.

2026

2027 **CART Analysis to Determine Thresholds Representing Potential Habitat Breaks**

2028 Classification and regression tree analysis (CART) was performed using the *rpart* package
2029 (Thernau and Atkinson 2022) in program R on the *Full Random* data set. A CART model was built
2030 for several combinations of variables to determine which set produces the highest prediction
2031 accuracy and enables comparison of model performance based on sensitivity, specificity, and
2032 Matthews Correlation Coefficient (MCC). Sensitivity represents the proportion of positive cases
2033 (fish) correctly classified whereas specificity represents the proportion of negative cases (no-fish)
2034 correctly classified. MCC is a statistical representation of all four confusion matrix categories (true
2035 positives, true negatives, false positives, and false negatives) that is a reliable and holistic
2036 indicator of model performance (Chicco and Jurman 2020). The data were split into a training and

2037 testing data set to assess the performance of CART models and produce a confusion matrix,
2038 prediction accuracy, sensitivity, specificity, and MCC. Additionally, a visual decision tree was
2039 generated for each model to identify potential thresholds for variables used at each node or
2040 decision point (Figure 1). The decision trees presented in this analysis includes a root node where
2041 a decision is made on a single variable forming a split and separate branches. Each subsequent
2042 node is a decision node where additional splits form new branches. The final node is the leaf node
2043 that is predicted on the outcome variable of interest. If a threshold at a split is true to the right
2044 branch the output is a "no fish" classification, if the threshold to the left represents "fish"
2045 classification. The classification rate (number of cases divided by total cases in that split) will be
2046 displayed below each leaf node.
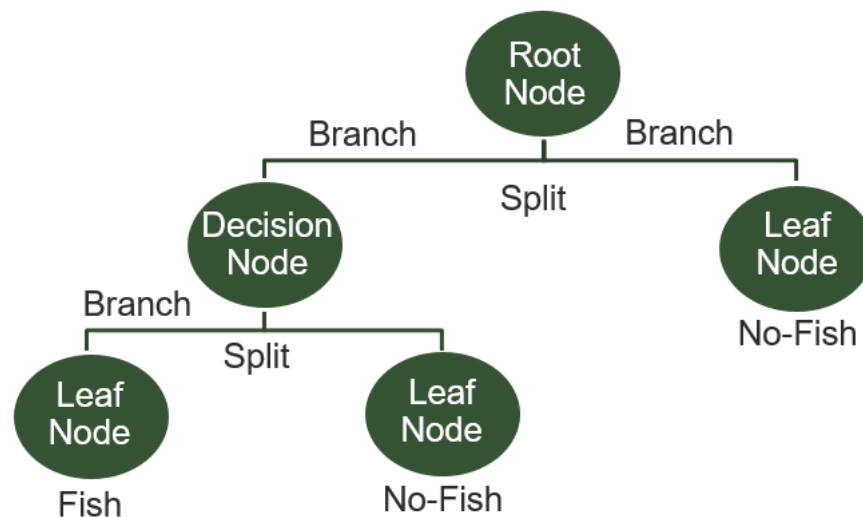
2047
2048 **Figure 1. Example labeled diagram for CART model decision tree output. Classification rates are not
2049 displayed but will be located below each leaf node.**
2050
2051 The CART models were informed by the random forest and interaction forest models and the
2052 criteria previously established by the Board. The CART model with the highest accuracy was
2053 manually pruned for improved clarity and utility by reducing the output to two and three splits. By
2054 comparing the accuracy, sensitivity, specificity, and MCC of the top model and the pruned models
2055 we can investigate trade-offs between model accuracy and complexity for establishing putative
2056 thresholds.

RESULTS

2057

Random Forest Models

2058

2059 Of the eight random forest models, the full random model was most accurate (Table 2). The *Full
2060 Random* model including step covariates exhibited an accuracy of 93.52%, whereas the *Full
2061 Random* model without step covariates demonstrated 85.53% accuracy. The random sampling of
2062 stream segments as opposed to the leave one-out approach of an entire stream performed better
2063 for all data set groupings. The difference between the accuracy of the *Western Random,*
2064 (92.92%), and the *Western Random LOO*, (86.08%) was 6.84%. The difference in accuracy

2065   between *Full Random,* (93.52%) and the *Full Random LOO* (82.14%), was 11.38%. However, the
2066   greatest difference in accuracy, 31.64%, occurred between the *Eastern Random* (92.83%), and
2067   the *Eastern Random LOO* (61.19%). The *Full Random WE Predictor* model exhibited an accuracy
2068   of 89.41%, which was higher than the *Full Random LOO* accuracy of 82.14% but lower than the
2069   *Full Random* (93.52%). Tuning parameters between model iterations appears to be an important
2070   procedure for these data as the *mtry*, *max nodes*, and *number of trees* values differed across
2071   models at the same spatial scale and across spatial scales (Table 2).
2072
2073   Across almost all model iterations, the maximum upstream gradient (Max.Up.Grad) and maximum
2074   downstream gradient (Max.Dn.Grad) exhibited the top two highest variable importance scores
2075   (Figure 2). However, the maximum upstream step bankfull width (Max.Up.Step.BFW10) was the
2076   most important variable for the *Western Random* model. Gradient and step-related characteristics
2077   exhibited the highest variable importance scores across all models. Substrate and UnitLabel
2078   exhibited small importance scores for all models. Violin plots and box plots in Appendix A provide
2079   a qualitative assessment for possible test criteria to define end of fish for several of these
2080   important variables. For example, the average values for maximum downstream gradient for fish
2081   segments is lower than the average at the end of fish segment and the segment just above the
2082   end of fish. The analysis using the *Boruta* package concluded that almost all variables were
2083   deemed important for each model iteration (Figure 3), and importance values followed a similar
2084   pattern as that reported by the *randomForest* output (Figure 3). Unit type (UnitLabel) for *Western*
2085   *Random LOO* was deemed tentatively important and unimportant for the *Western Random* model
2086   (Figure 3d). Effective step height in meters (Eff.StepHt.m) and effective step height at bankfull
2087   width (Eff.StepHt.BFW) for the *Eastern Random* models were deemed tentatively important
2088   (Figure 3c). The partial dependency plots (Figure 4) demonstrate the importance of maximum
2089   downstream gradient, maximum upstream gradient, and bankfull width at predicting fish presence
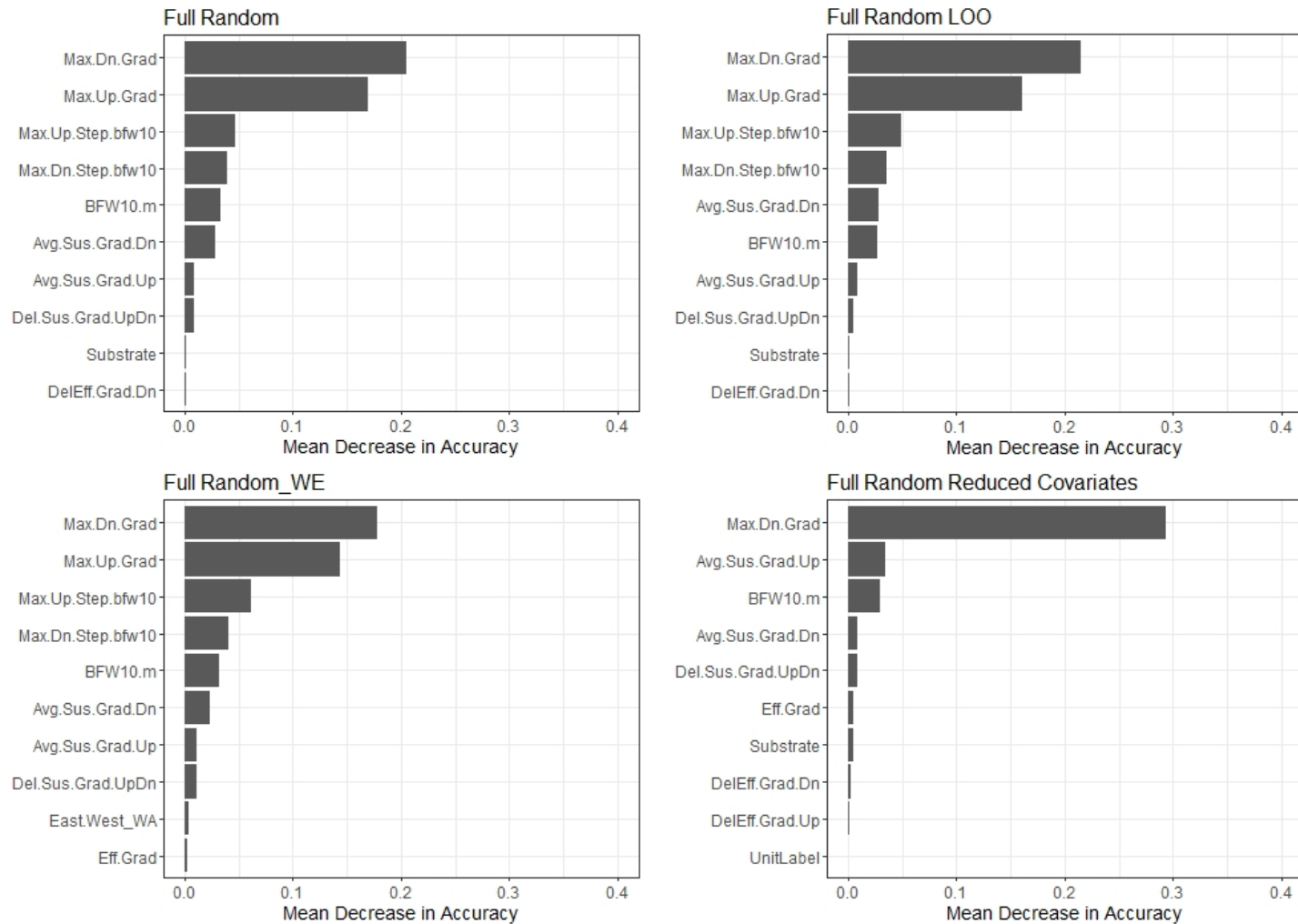2090   at a segment.

Figure 2a. Variable importance from random forest models using the *Full Random, Full Random Leave One Out (LOO), Full Random West/East (WE),* and *Full Random Reduced Covariates* data sets. Visualized using package vip (Greenwell and Boehmke 2020). Mean Decrease in Accuracy represents how much accuracy the model loses without the inclusion of that variable.
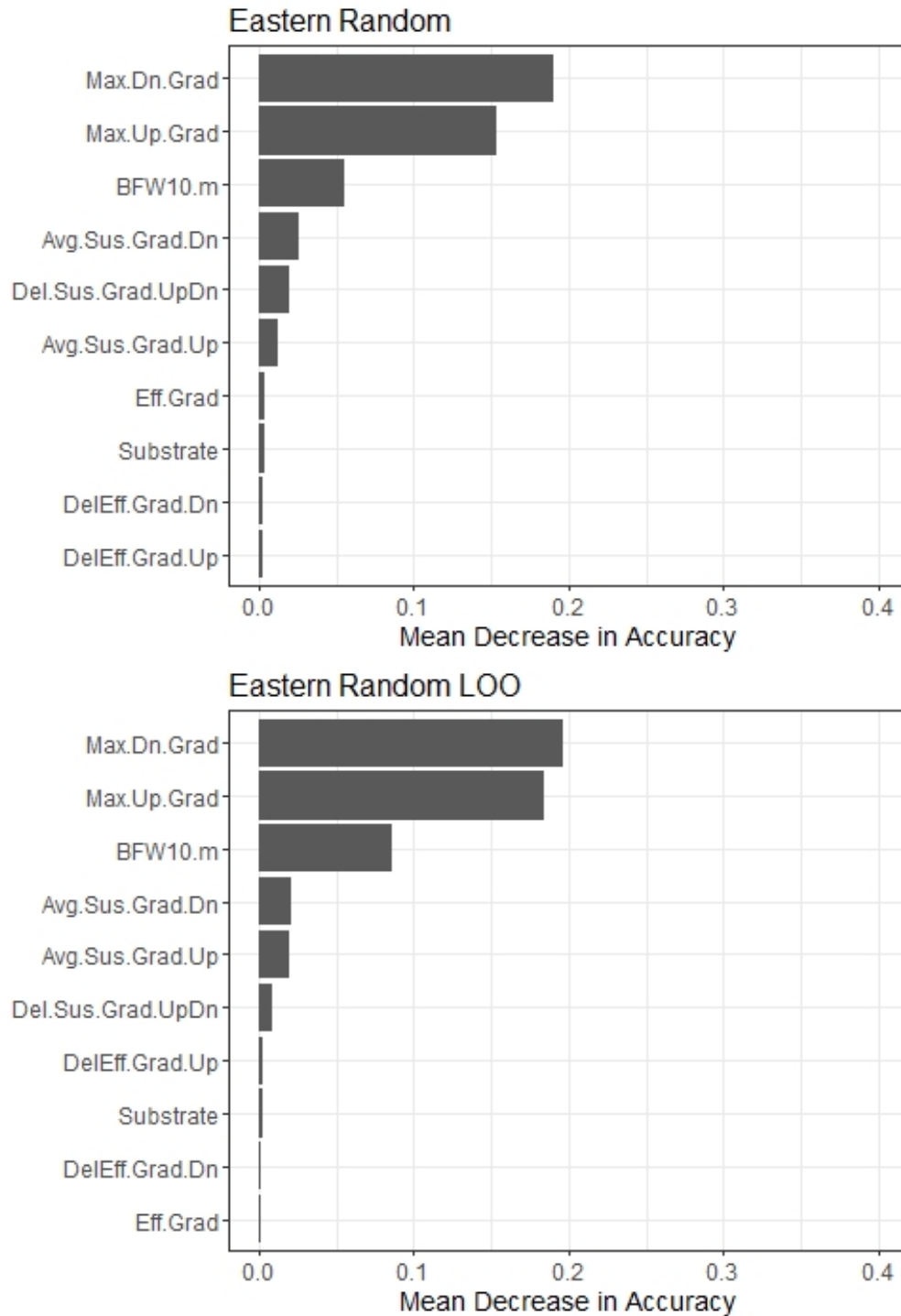
2091

Figure 2b. Variable importance (mean decrease in accuracy if the variable is removed) from random forest models using the *Eastern Random* and *Eastern Random Leave One Out (LOO)* data sets. Visualized using package vip (Greenwell and Boehmke 2020).

2092
2093
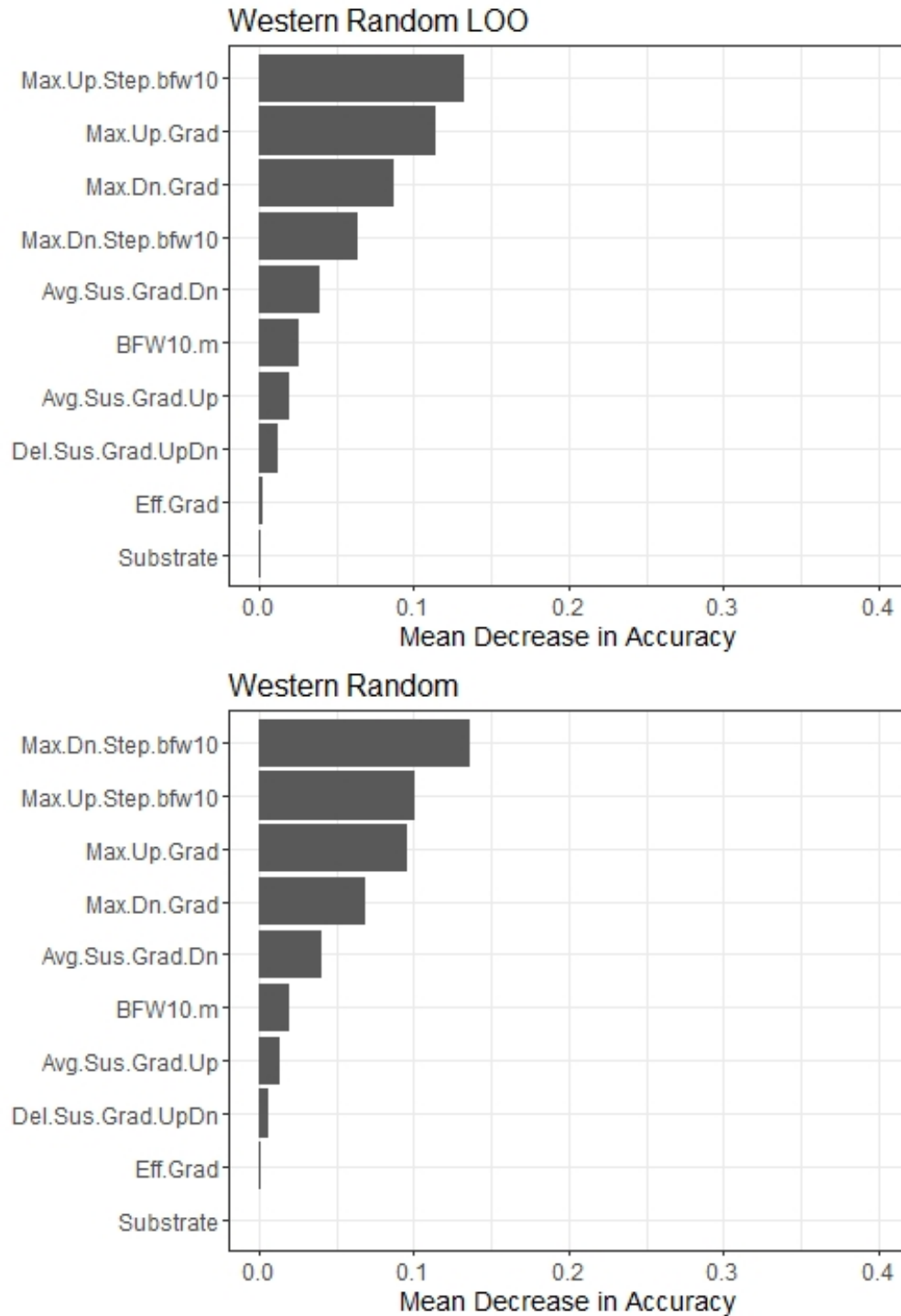
Figure 2c. Variable importance (mean decrease in accuracy if the variable is removed) from random forest models using the *Western Random* and *Western Random Leave One Out (LOO)* data sets. Visualized using package vip (Greenwell and Boehmke 2020).
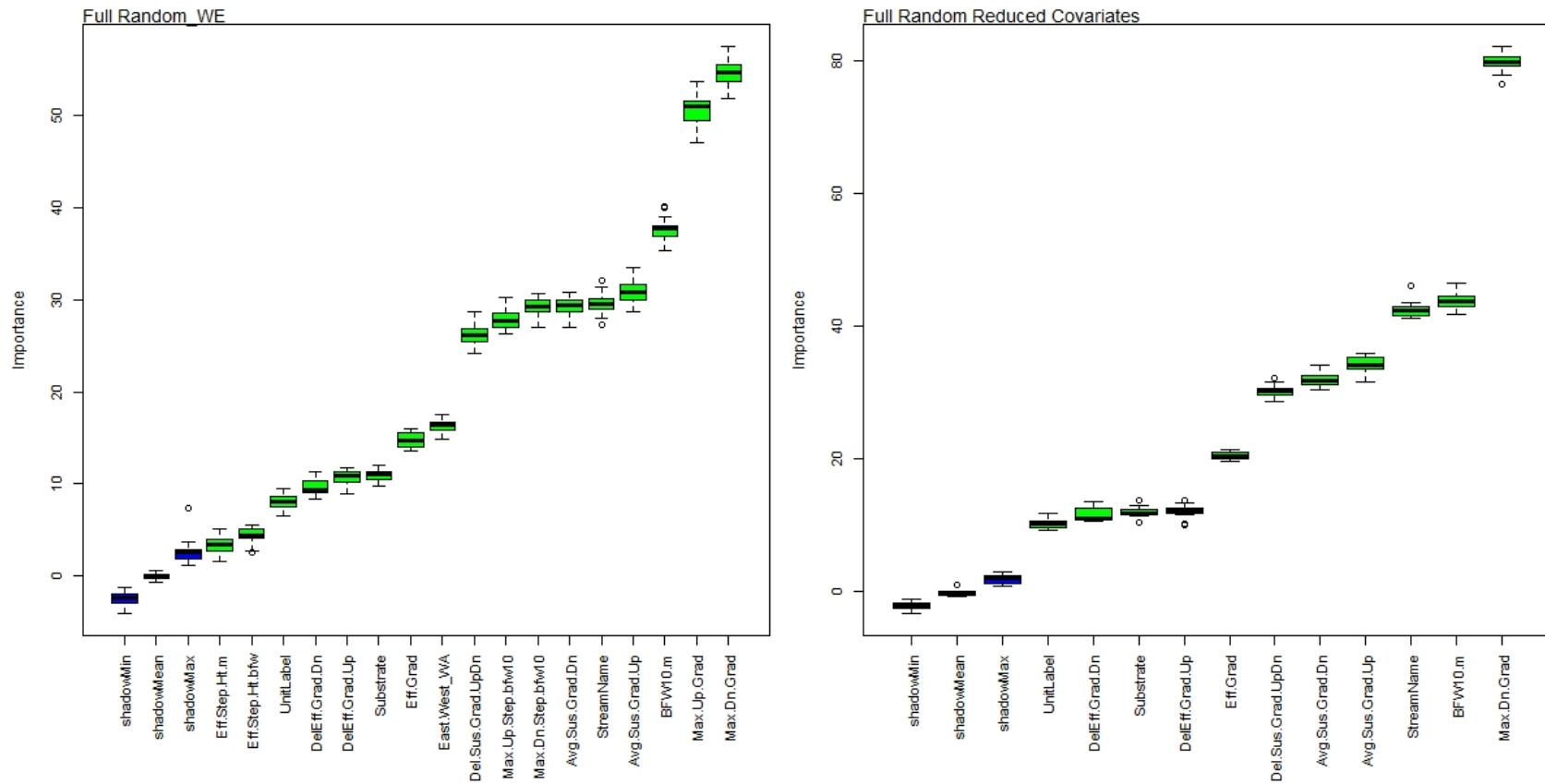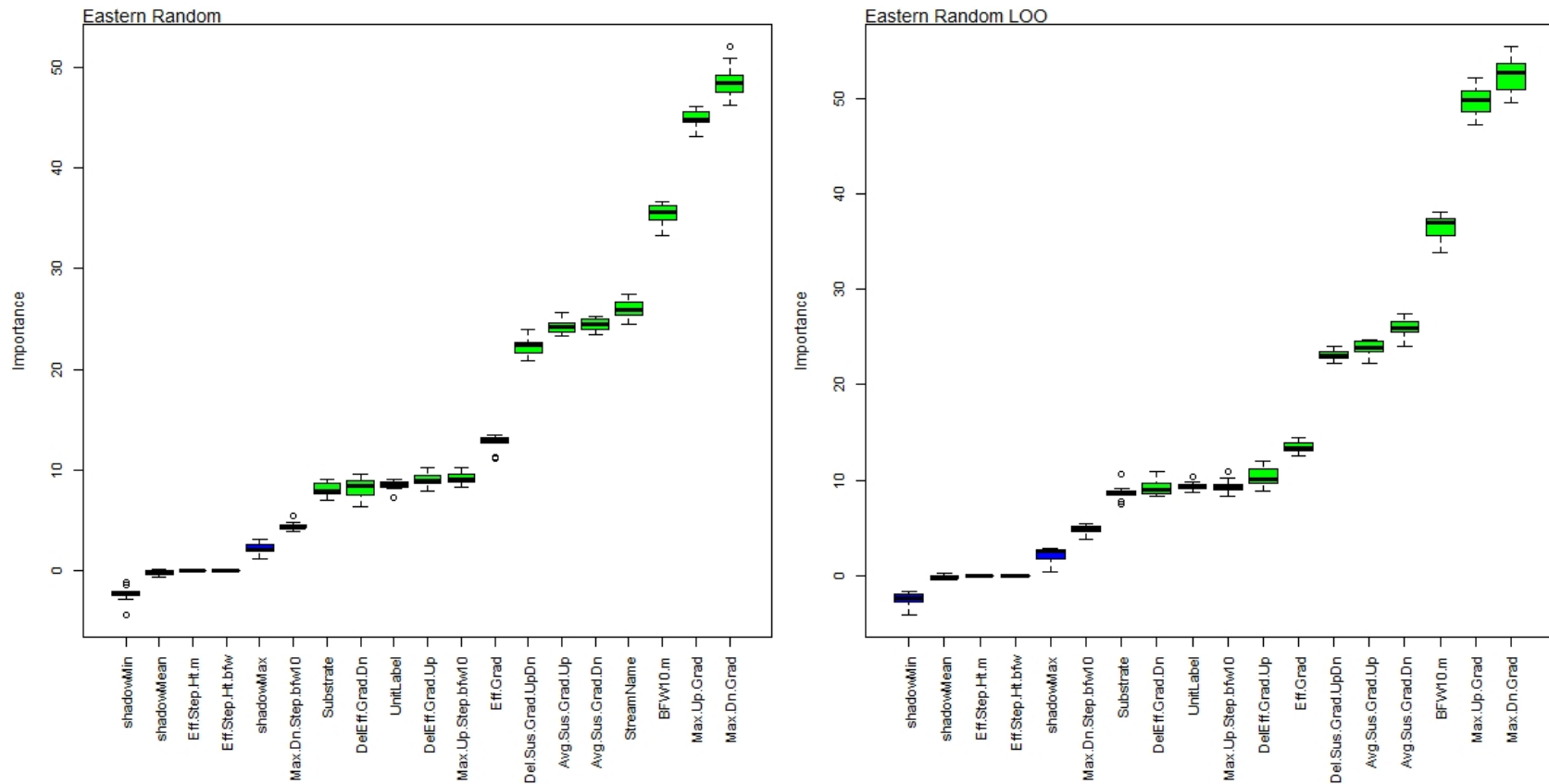
2094
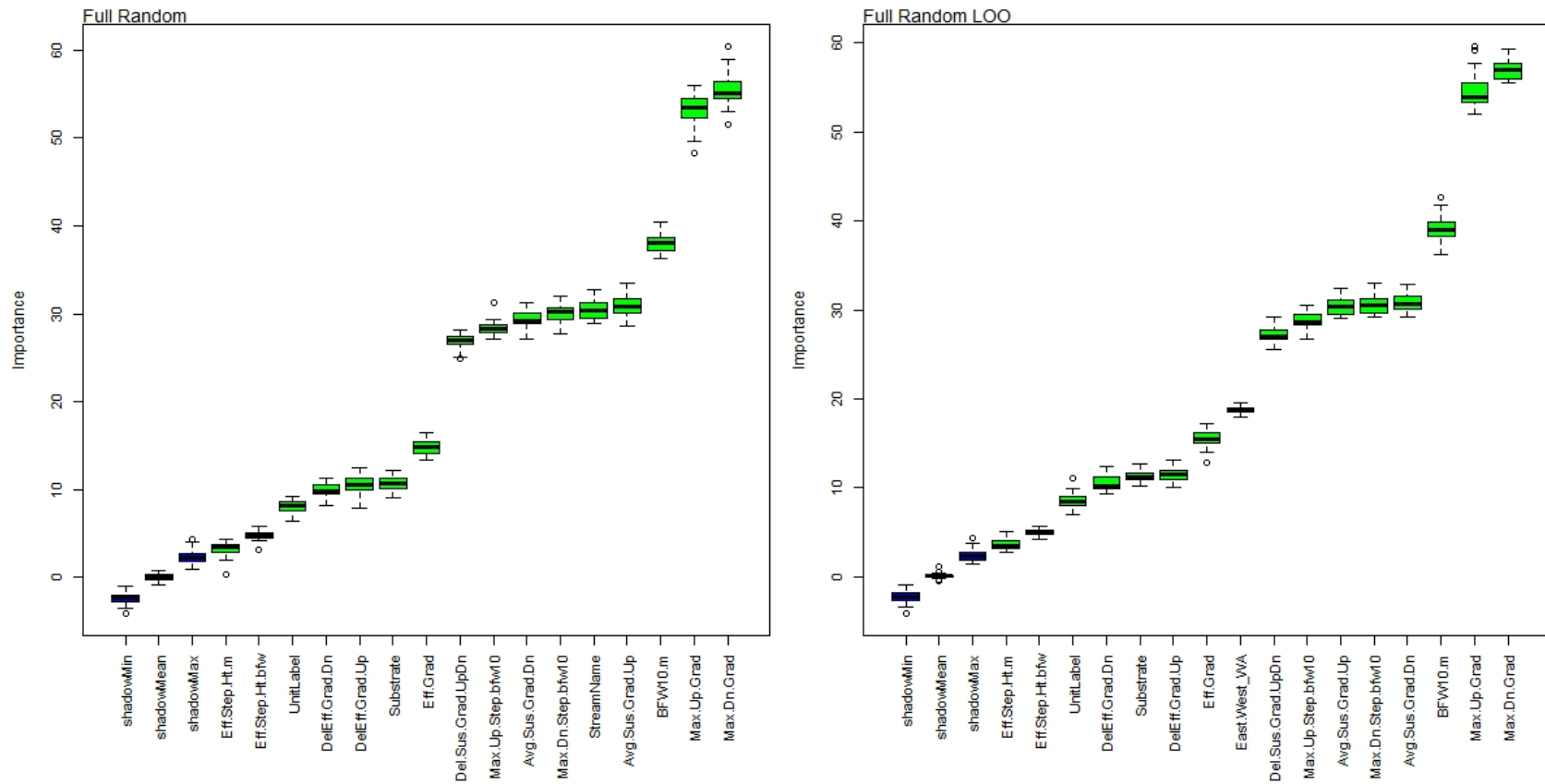
Figure 3a. Confirming variable importance with *Boruta* package using the *Full Random* and the *Full Random Leave One Out* data sets. Features in green were deemed important by *Boruta,* yellow are tentatively important, red are unimportant, and blue are called shadow features from *Boruta.* Shadow features are shuffled copies of all features to add randomness to the *Boruta* algorithm.

2095

Figure 3b. Confirming variable importance with *Boruta* package using the *Full Random West/East (WE),* and the *Full Random Reduced Covariates* data sets. Features in green were deemed important by *Boruta,* yellow are tentatively important, red are unimportant, and blue are called shadow features from *Boruta.* Shadow features are shuffled copies of all features to add randomness to the *Boruta* algorithm.

2096
2097

Figure 3c. Confirming variable importance with *Boruta* package using the *Eastern Random* and *Eastern Random Leave One Out* data sets. Features in green were deemed important by *Boruta,* yellow are tentatively important, red are unimportant, and blue are called shadow features from *Boruta.* Shadow features are shuffled copies of all features to add randomness to the *Boruta* algorithm.
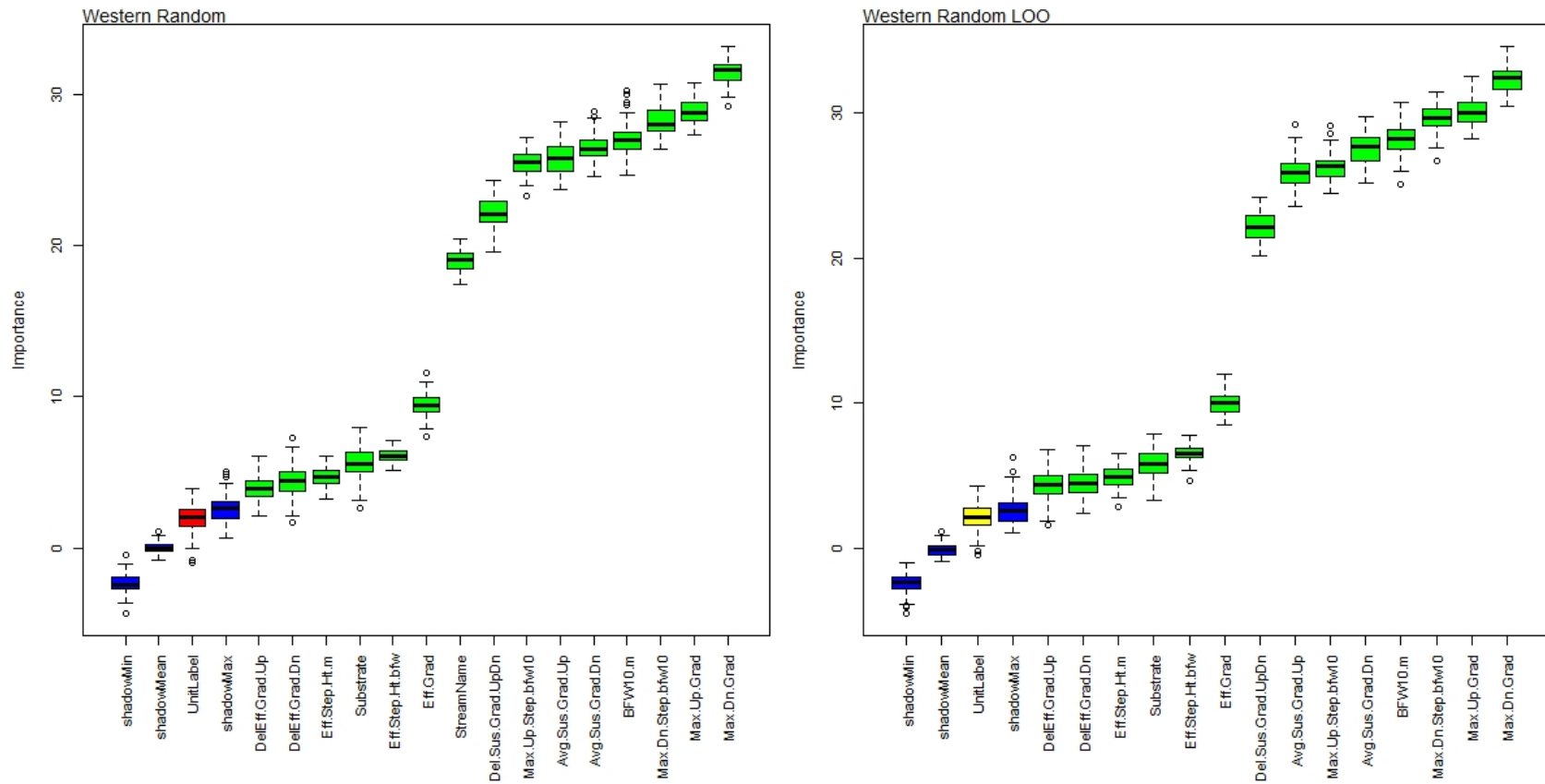
2098

2099

Figure 3d. Confirming variable importance with *Boruta* package using the *Western Random* and *Western Random Leave One Out* data sets. Features in green were deemed important by *Boruta,* yellow are tentatively important, red are unimportant, and blue are called shadow features from *Boruta.* Shadow features are shuffled copies of all features to add randomness to the *Boruta* algorithm.
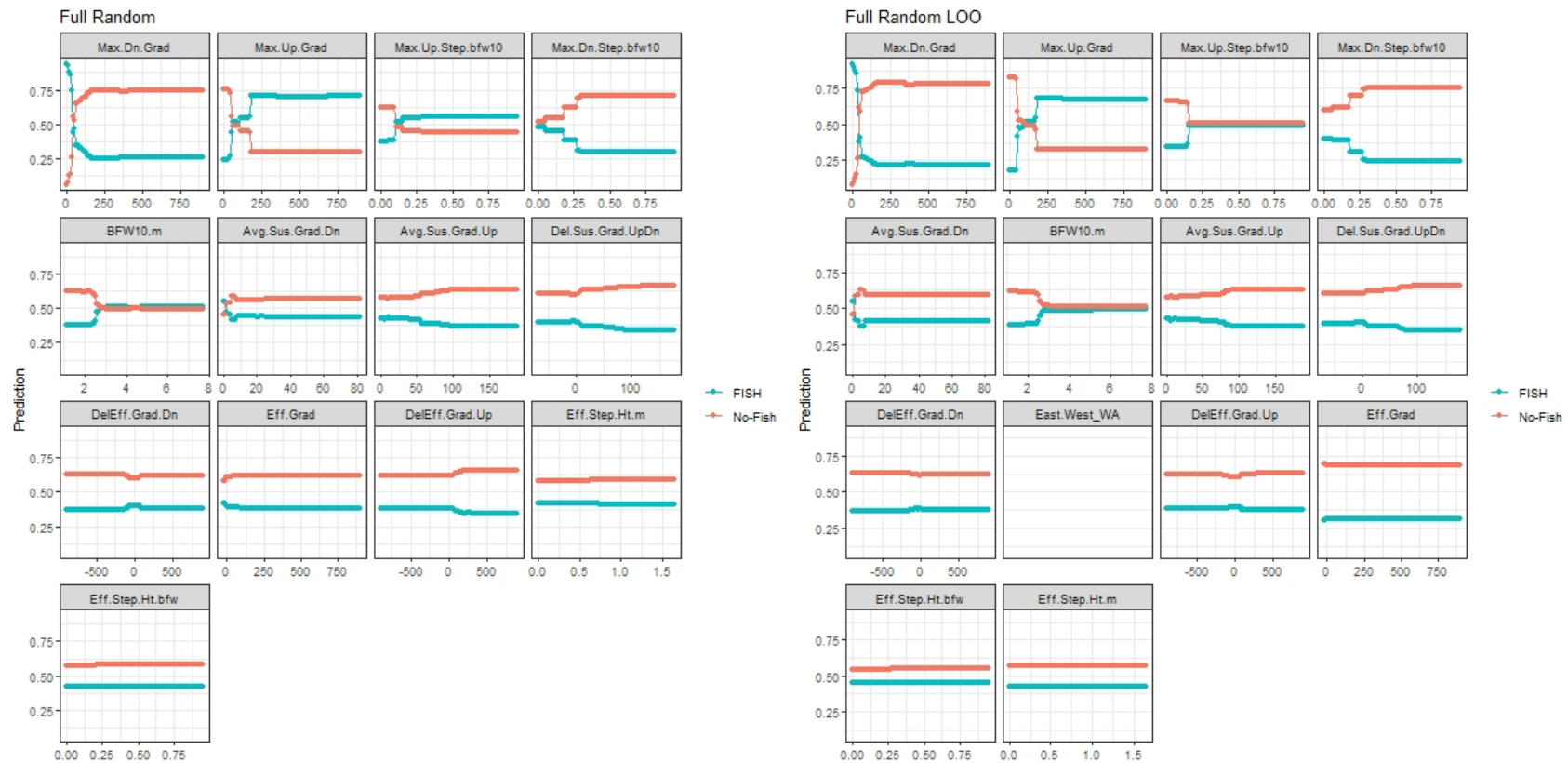
2100

Figure 4a. **Partial dependency plots in order of importance to the random forest model for *Full Random* and *Full Random Leave One Out (LOO)*. The y-axis represents the probability of prediction into a particular class based on the value (x axis) for that particular feature. X-axis labels are in the gray text box above each graph. Substrate and unit are not displayed. *Full Random West/East Predictor* model output is not displayed because it follows the same pattern as the *Full Random* model.**

2101
2102

Figure 4b. Partial dependency plots in order of importance to the random forest model for *Eastern Random* and *Eastern Random Leave One Out (LOO)*. The y-axis represents the probability of prediction into a particular class based on the values (x-axis) for that particular feature. X-axis labels are in the gray text box above each graph. Substrate and unit are not displayed.
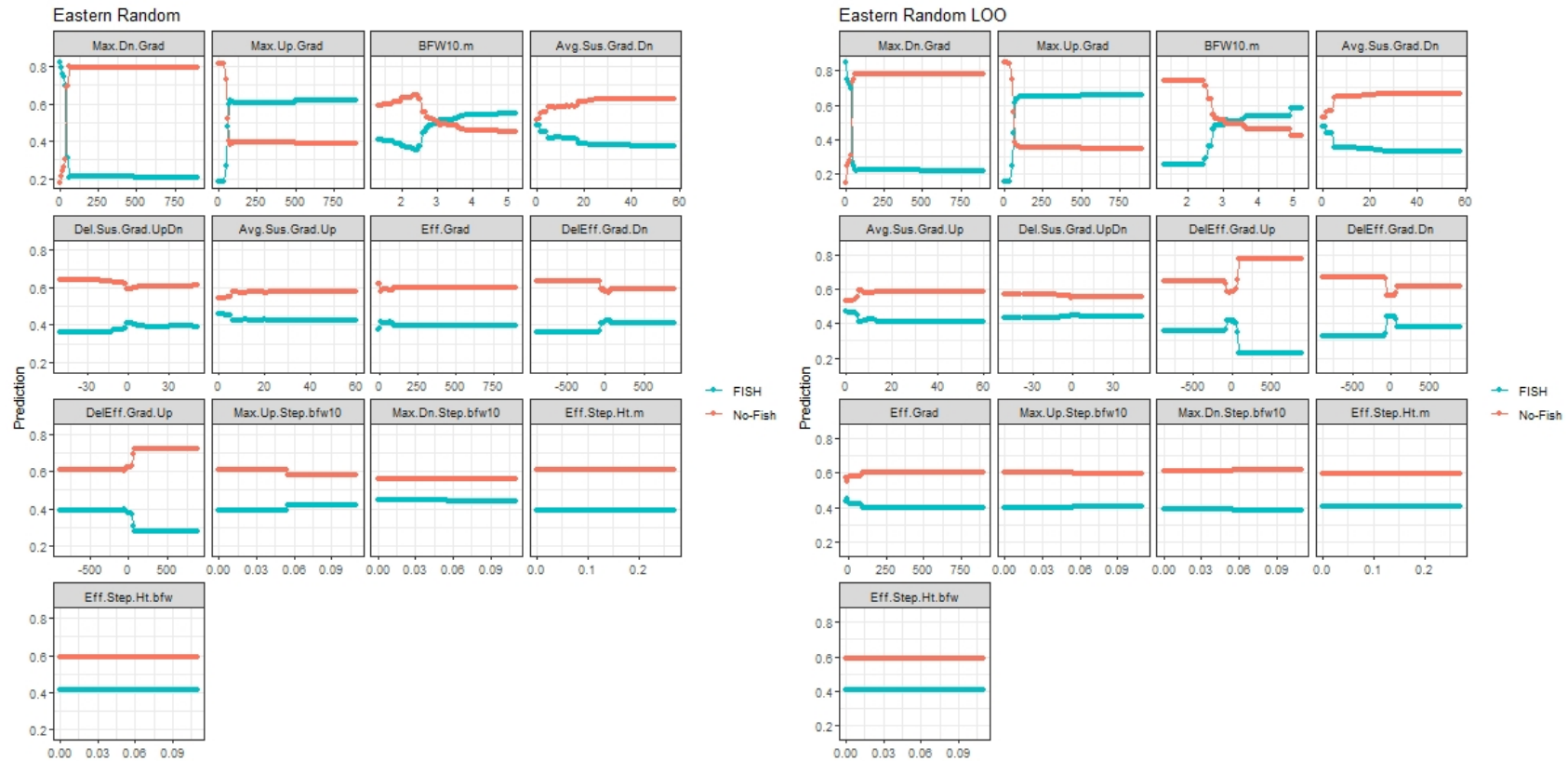
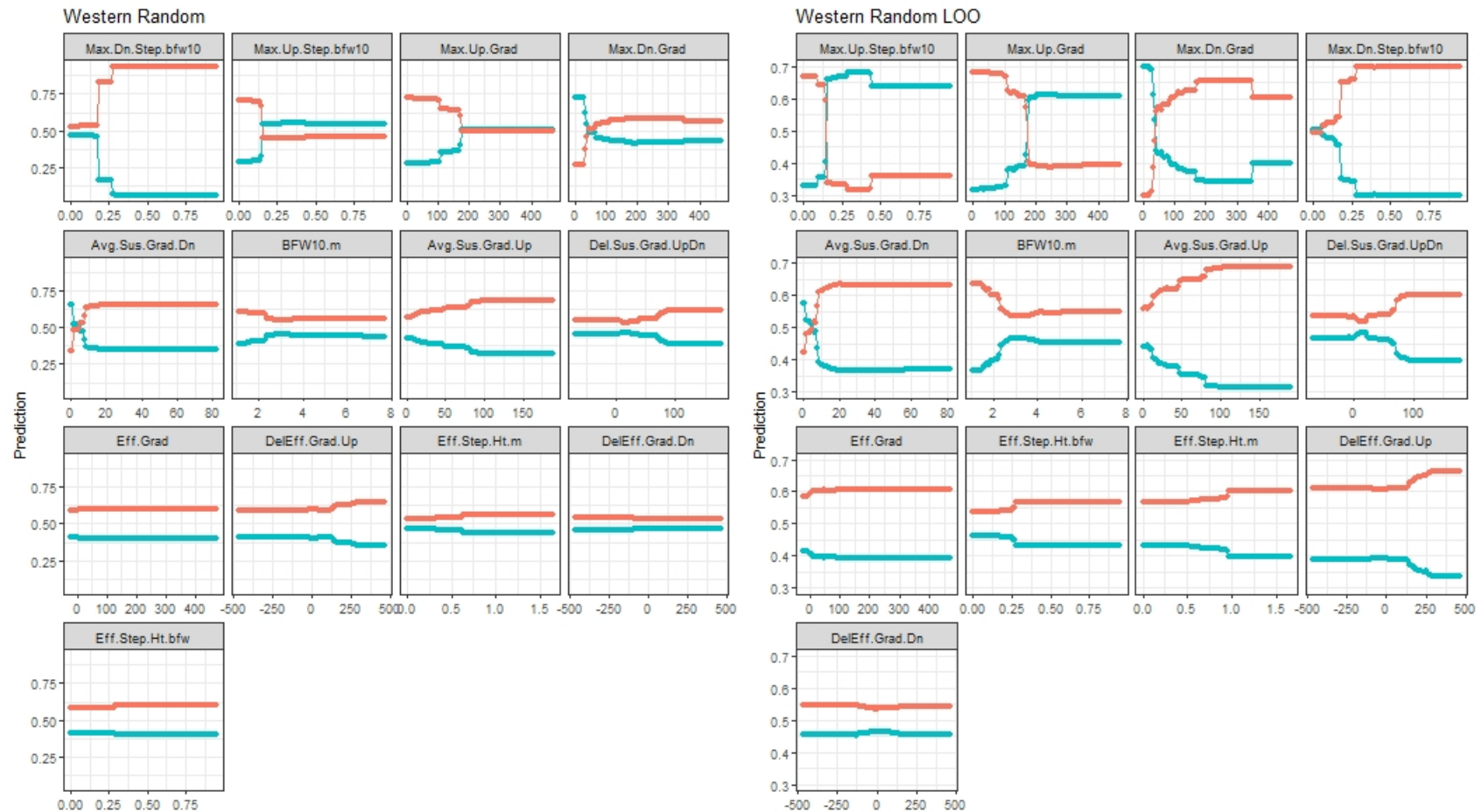2103

2104

Figure 4c. Partial dependency plots in order of importance to the random forest model for *Western Random* and *Western Random Leave One Out (LOO)*. The y-axis represents the probability of prediction into a particular class based on the values (x-axis) for that particular feature. X-axis labels are in the gray text box above each graph. Substrate and unit are not displayed.

2105

2106 **Interaction Forest Models**

2107 Using the pilot dataset, the interaction forest model produced a more accurate prediction (97.17%)
2108 than the random forest model, (89.63%; Table 4). The accuracy was primarily a function of higher
2109 specificity with the interaction forest model as compared to the random forest model which
2110 demonstrated higher sensitivity. This result would imply that the random forest model was more
2111 adept at identifying physical characteristics associated with the segments below a PHB while the
2112 interaction forest identified features associated with segments above the PHB. The pairwise
2113 interaction strength for the five covariate pairs with the highest EIM (Table 5) are displayed as
2114 contour maps (Figure 5). The contour maps display the probability of predicting fish presence
2115 given particular pairwise relationships. For example, a segment where the maximum upstream
2116 gradient is greater than 200% and the maximum downstream step (bankfull widths) is lower than
2117 0.38m has a high (90-100%) probability of being classified as containing fish. Additionally, the
2118 logistic regression test for interaction effects between pairs of covariates demonstrates that
2119 segments with a maximum downstream gradient greater than 71% and a low maximum upstream
2120 step bankfull width has a low probability of being classified as containing fish (Figure 5). The
2121 highest effect importance measure for maximum upstream gradient and maximum downstream
2122 step (bankfull width) was 0.007 (Table 5; Figure 5). While effective gradient had an overall low
2123 interaction strength, near zero (Figure 6), the interaction between effective gradient and maximum
2124 downstream gradient was one of the highest at 0.005 (Table 5). Maximum downstream gradient,
2125 maximum upstream gradient, maximum step bankfull width, bankfull width (BFW10.m), and the
2126 average sustained upstream gradient had the highest overall interaction strengths of all covariates
2127 (Figure 6).

2128

Table 4. Comparison between the full random sample using random forest and interaction forest. Interaction forest performed marginally better.

Model Type	Number of Trees	AUC	Accuracy (PCC)	Sensitivity	Specificity	Kappa
Random Forest[†]	300	0.90	89.63%	0.94	0.87	0.79
Interaction Forest	300	0.94	94.17%	0.90	0.98	0.88

[†]Random forest model tuning parameters and performance metrics using the *Random Full* data set with substrate and unit features removed.

AUC = area under the curve; kappa =a measure of agreement between predicted presences and absences; PCC = proportion of presence correctly classified; sensitivity = proportion of presence correctly classified; specificity = the proportion of absence correctly classified

2129
2130

Table 5. Effect importance measure (EIM) values for the interaction between variable pairs (A and B).

Variable A	Variable B	EIM
Max.Up.Grad	Max.Dn.Step.BFW10	0.007
Max.Dn.Grad	Max.Up.Step.BFW10	0.005
Eff.Grad	Max.Dn.Grad	0.005
Max.Up.Grad	Max.Up.Step.BFW10	0.004
Avg.Sus.Grad.Up	Max.Dn.Grad	0.004

Avg= average; BFW = bankfull width; BFW10 = BFW for 5 segments below, the current segment, and four segments above; DelEff = Change in effective; Dn = downstream; Eff = effective; Grad = gradient; Ht = height; m = meter; Step = ?:Sus = sustained; Up = upstream

2131

Figure 5. Interaction contour and cross section plots for pairs of variables with the highest effect importance measure values. P-values on each cross-section plot are overly optimistic according to the *diversityForest* manual. Since both predictors are continuous and the outcome is categorical, *diversityForest* employs a 2-dimensional LOESS regression. The color gradient in the contour plot ranges from purple at 0 (no-fish) to yellow at 1 (fish).

2132

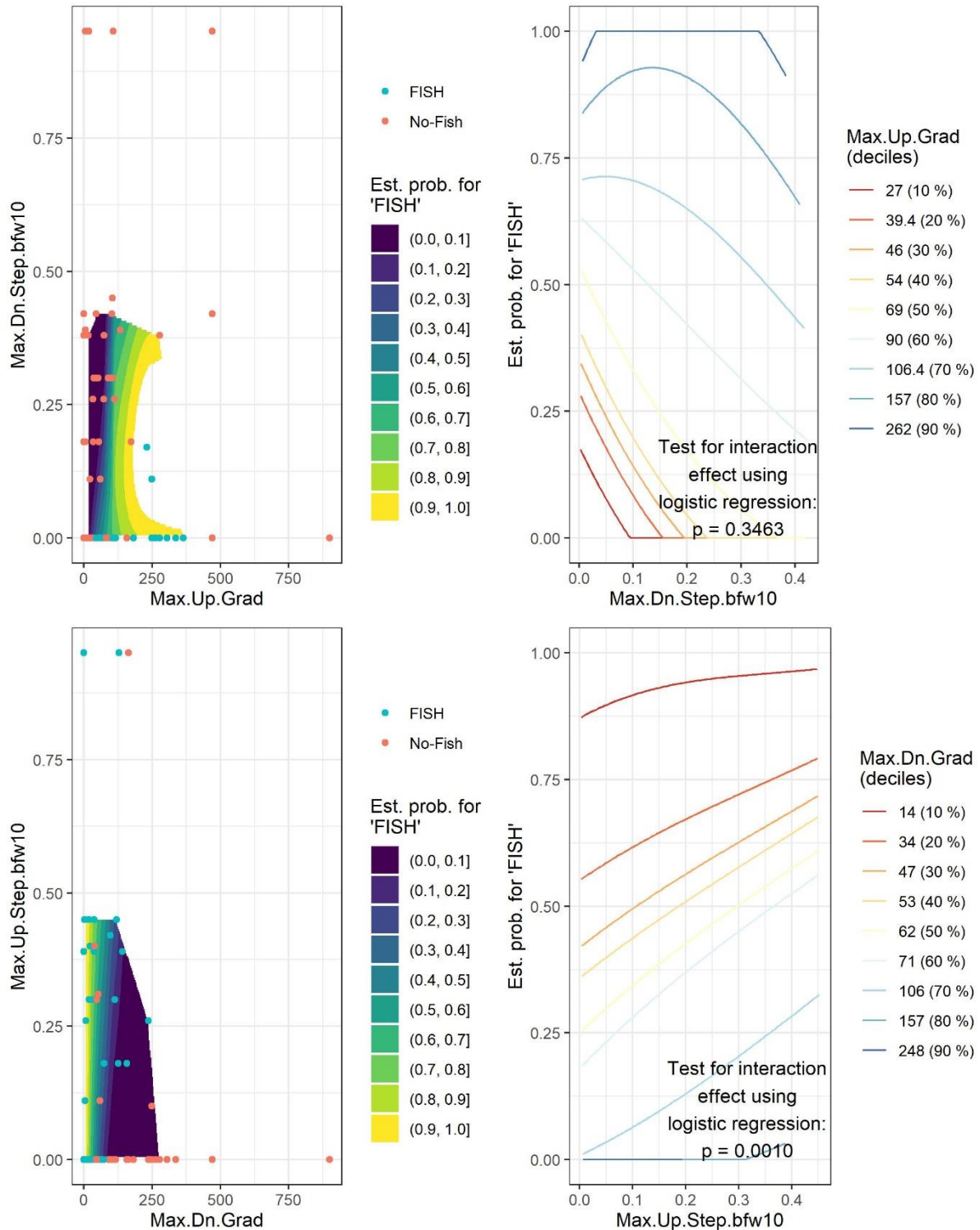**Figure 5.** (*continued*) Interaction contour and cross section plots for pairs of variables with the highest effect importance measure values. P-values on each cross-section plot are overly optimistic according to the *diversityForest* manual. Since both predictors are continuous and the outcome is categorical, *diversityForest* employs a 2-dimensional LOESS regression. The color gradient in the contour plot ranges from purple at 0 (no-fish) to yellow at 1 (fish).
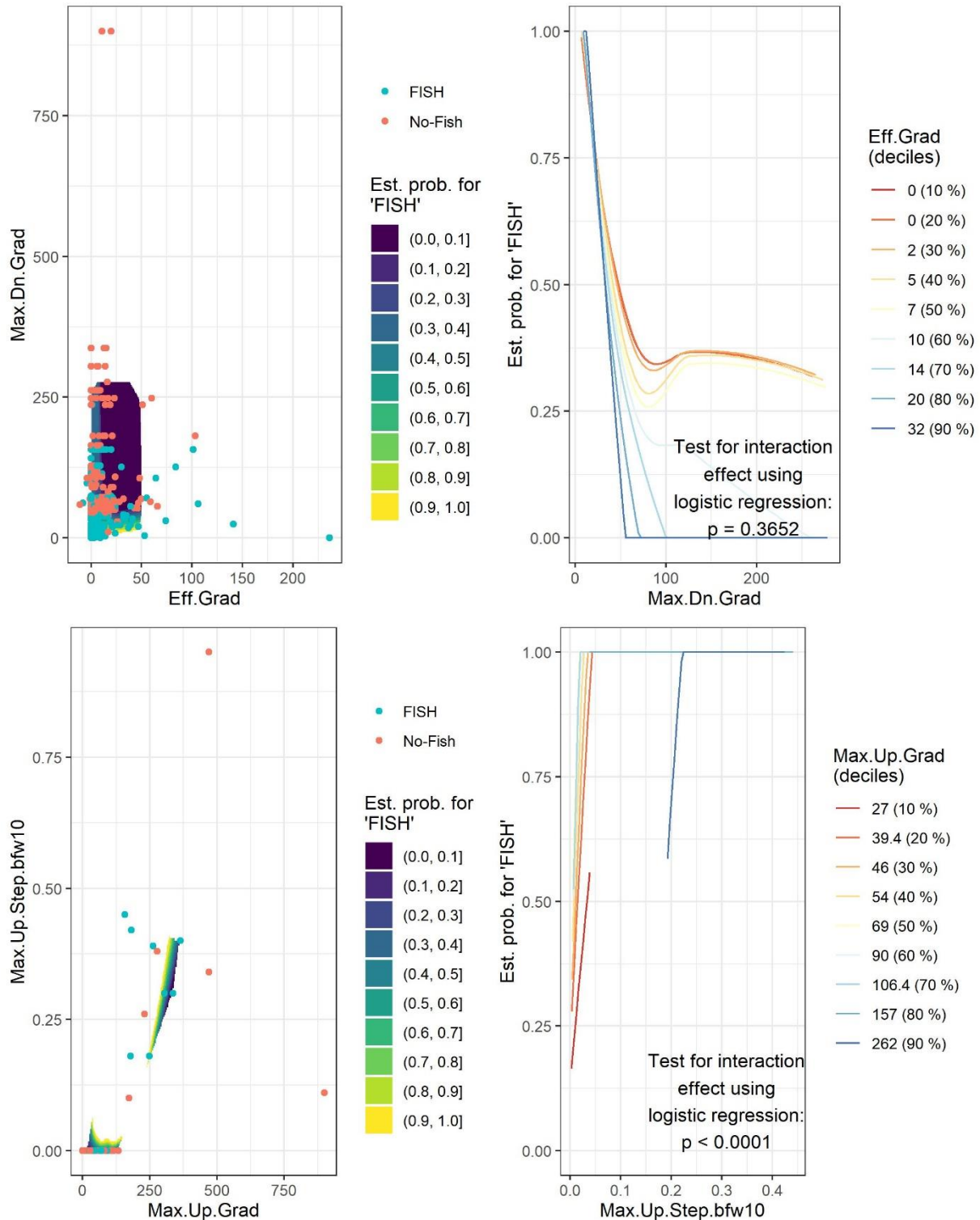
Figure 5. (*continued*) **Interaction contour and cross section plots for pairs of variables with the highest EIM values. P-values on each cross-section plot are overly optimistic according to the *diversityForest* manual. Since both predictors are continuous and the outcome is categorical, *diversityForest* employs a 2-dimensional LOESS regression. The color gradient in the contour plot ranges from purple at 0 (no -fish) to yellow at 1 (fish).**
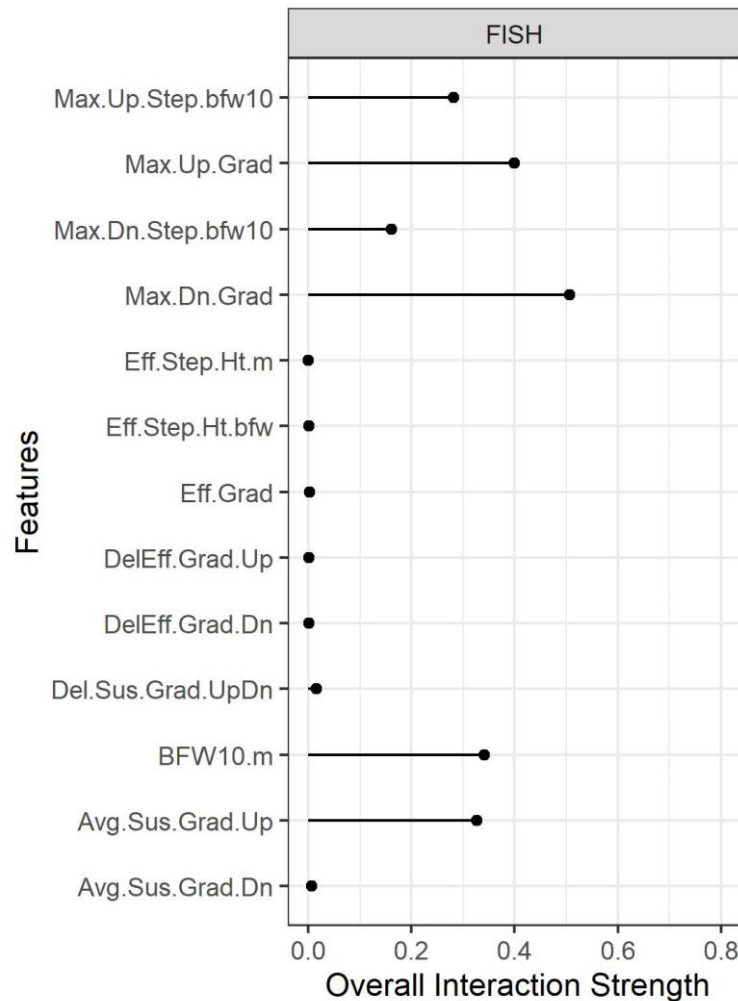
2133
2134

Figure 6. Overall interaction strength using package *iml* for each stream characteristic.

2135

Evaluating Forest Practice Board proposed Potential Habitat Break Criteria

2136

2137 Four criteria sets were examined related to the FPB criteria: options A, B, and C and the combined set
2138 of unique criteria used in the All Criteria model. Because no stream segments in the pilot data set met
2139 TestCriterion2 or TestCriterion3, these criteria were not included in the evaluations of options A, B, or C.
2140 Similarly, the random forest model for All Criteria combined contained only the five criteria that were met
2141 by any segments in the pilot data set (TestCriterion1, TestCriterion4, TestCriterion5, TestCriterion6, and
2142 TestCriterion7).

2143

2144 Predicting fish presence using the four criteria sets resulted in low accuracy, sensitivity, specificity, and
2145 kappa parameters (Table 6). This was most notable for Option B that exhibited an accuracy of 48.36%.
2146 The confusion matrices in Table 7 display the comparisons of observed fish presence versus. The fish
2147 presence based on FPB criteria. This result seems largely driven by the large number of false negative
2148 results (observed = fish; prediction = no-fish) for Option A, and false positives (observed = no-fish,
2149 prediction = fish) for All Criteria and Option B. Option C had nearly equal numbers of false negatives

2150   and false positives. Evaluating the FPB criteria using random forest models resulted in low accuracies
2151   and poor model performance (Table 8).

2152

Table 6.   Prediction evaluation of the four criteria compared to observed fish presence.

	AUC	Accuracy (PCC)	Sensitivity	Specificity	Kappa
All Criteria*	0.54	49.28%	0.84	0.24	0.07
Option A	0.60	62.52%	0.40	0.79	0.20
Option B	0.52	48.36%	0.74	0.29	0.03
Option C	0.59	59.8%	0.52	0.65	0.18

* Includes only TestCriterion1, TestCriterion4, TestCriterion5, TestCriterion6, and TestCriterion7 because no stream segments met the condition for TestCriterion2 or TestCriterion3.

AUC = area under the curve; kappa = a measure of agreement between predicted presences and absences; PCC = proportion of presence correctly classified; sensitivity = proportion of presence correctly classified; specificity = the proportion of absence correctly classified

2153
2154

Table 7.   Confusion matrices for each of the four criteria sets and the observed data.

All Criteria*		Observed	
		Fish	**No-Fish**
Prediction	Fish	811	997
	No-Fish	160	313

Option A		Observed	
		Fish	**No-Fish**
Prediction	Fish	391	275
	No-Fish	580	1,035

Option B		Observed	
		Fish	**No-Fish**
Prediction	Fish	721	928
	No-Fish	250	382

Option C		Observed	
		Fish	**No-Fish**
Prediction	Fish	509	455
	No-Fish	462	855

* Includes only TestCriterion1, TestCriterion4, TestCriterion5, TestCriterion6, and TestCriterion7 because no stream segments met the condition for TestCriterion2 or TestCriterion3.

2155

Table 8.   Parameters from model tuning in *caret* and model performance from validation testing for TestCriterion1, TestCriterion2, and TestCriterion3.

	mtry	Maxnodes	Number of Trees	AUC	Accuracy (PCC)	Sensitivity	Specificity	Kappa
All Criteria*	5	5	250	0.58	59.73%	0.53	0.62	0.13
Option A	1	5	250	0.64	64.77%	0.61	0.66	0.24
Option B	1	5	250	NA	57.77%	NA	0.58	0
Option C	2	5	250	0.62	62.14%	0.62	0.62	0.21

* Includes only TestCriterion1, TestCriterion4, TestCriterion5, TestCriterion6, and TestCriterion7 because no stream segments met the condition for TestCriterion2 or TestCriterion3.

AUC = area under the curve; kappa = a measure of agreement between predicted presences and absences; Maxnodes = maximum number of nodes; mtry = optimum number of covariates; PCC = proportion of presence correctly classified; sensitivity = proportion of presence correctly classified; specificity = the proportion of absence correctly classified

2156

2157 Variables of importance differed little between each for each criteria set. TestCriterion1, the barrier
2158 cutoff of 20%, was the most useful predictor for the models for All Criteria, Option A, and Option C
2159 (Figure 7). TestCriterion5 and TestCriterion6, followed by the gradient of 10%, exhibited low
2160 variable importance in the All Criteria and Option C models (Figure 7), but was deemed
2161 unimportant for the Option B model by the *Boruta* algorithm (Figures 7 and 8). Similarly,
2162 TestCriterion7 was deemed important in the All Criteria model by random forest and *Boruta*, but
2163 unimportant for the Option C model (Figures 7 and 8).

2164

2165 TestCriterion1 relates to sustained stream gradient and parallels the results from the random
2166 forest *Full Random* model (Figure 2) where variables related to gradient were deemed most
2167 important and the interaction forest model (Figure 6) where gradient variables had strongest
2168 interaction strength. TestCriterion5 is also related to gradient but did not emerge as strong of a
2169 predictor as TestCriterion1. TestCriterion6 relates to obstacles and step heights and was found
2170 most important when paired with TestCriterion1 (Figure 8). This finding is corroborated in both
2171 the random forest models and the interaction forest model. Step-related variables were
2172 consistently in the top five most important variables (Figures 2–4), and the strongest interaction
2173 strength existed between gradient-related variables and step variables (Figure 5). More
2174 specifically, the interaction strengths were strongest for maximum upstream or downstream
2175 gradient variables and the bankfull width at the step. Width changes are encapsulated in
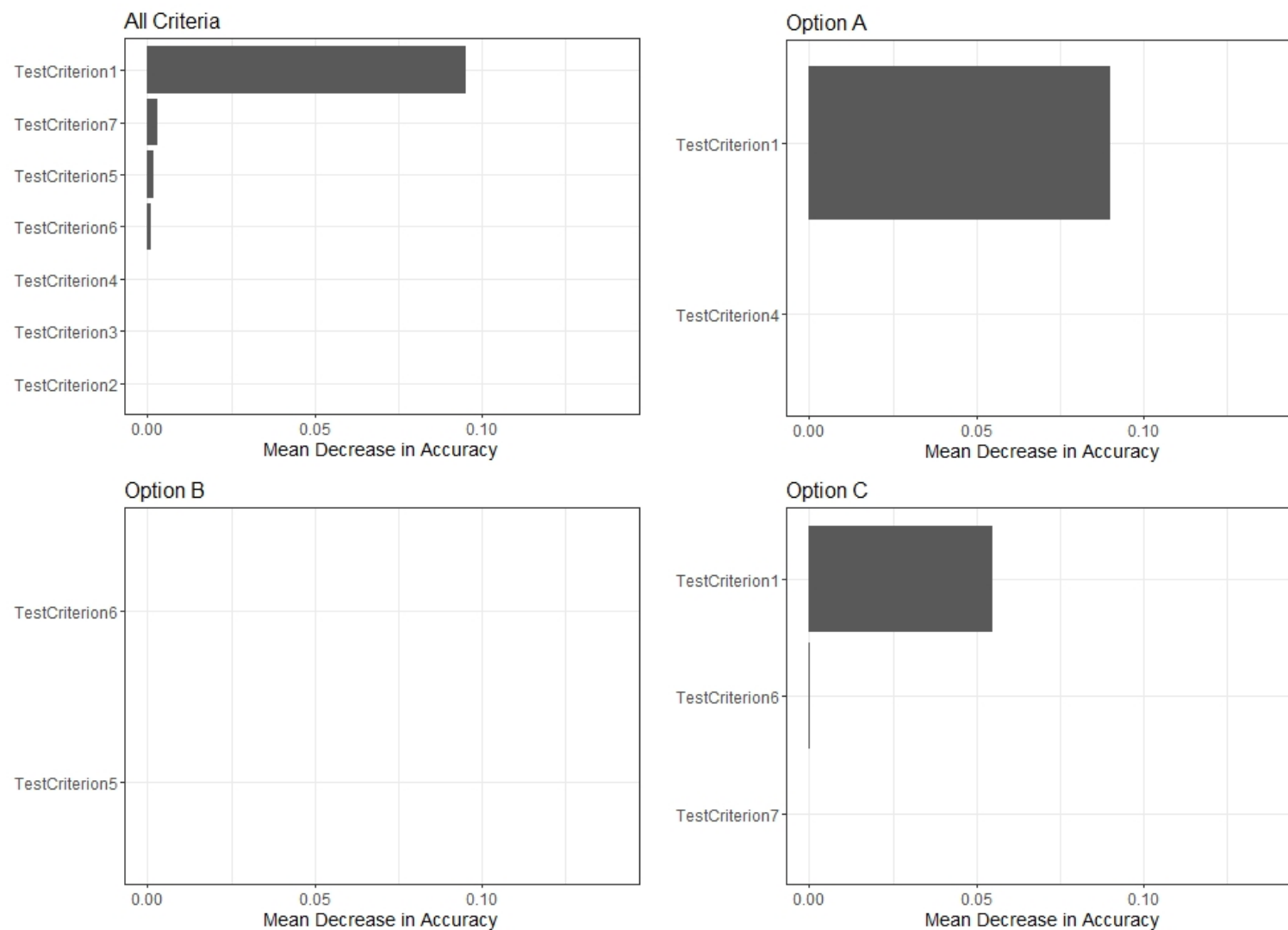2176 TestCriterion7, and the width criteria were deemed important for the All Criteria model.

Figure 7. Variable importance from random forest models for criteria sets based on options from the Washington Forest Practices Board outlined in Table 3. Visualized using package vip (Greenwell and Boehmke 2020).

2177
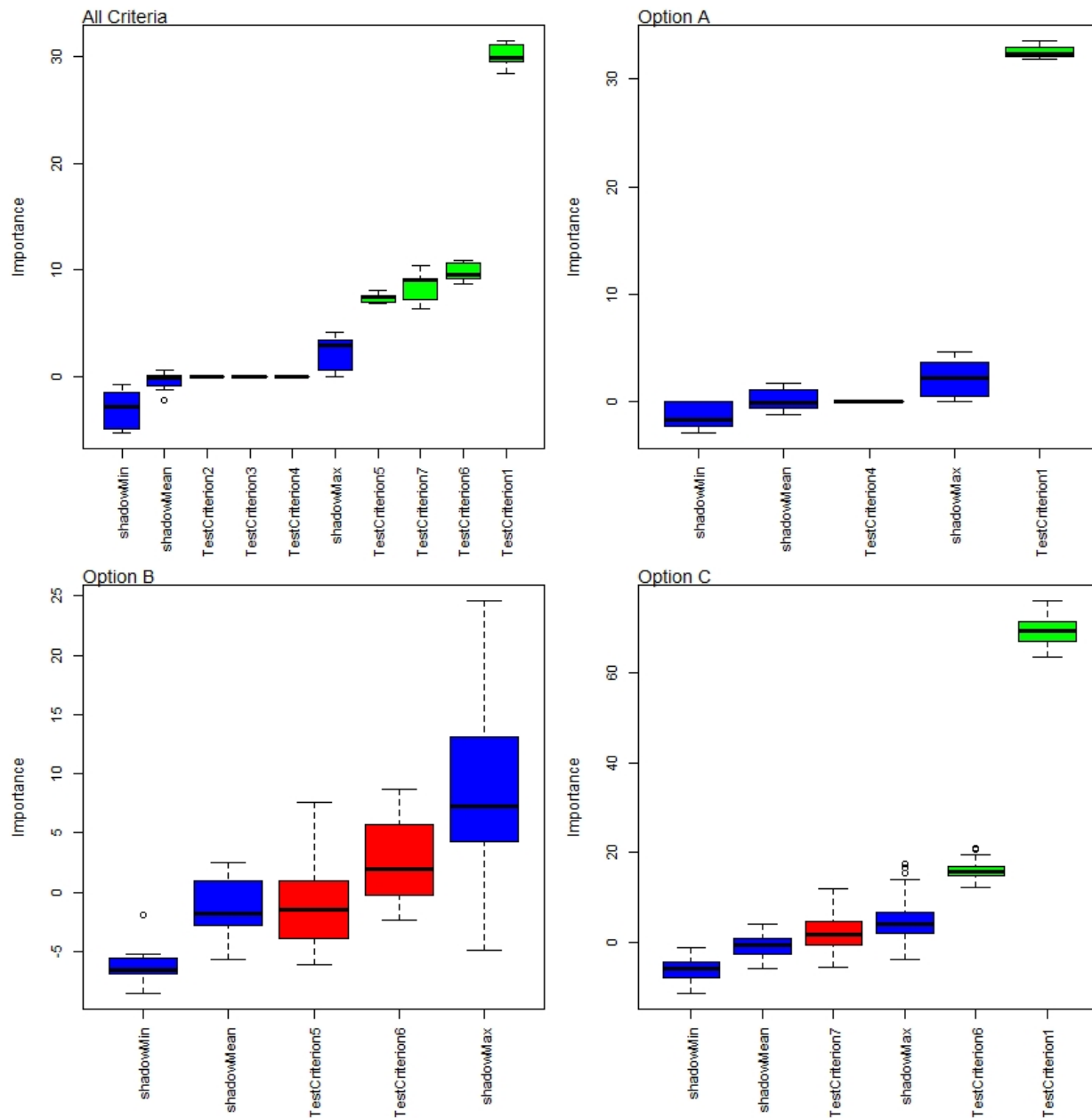
Figure 8. Variable importance validation using *Boruta*. Variable importance is displayed for each criterion described by the Washington Forest Practices Board in Table 3. Features in green were deemed important by *Boruta,* yellow are tentatively important, red are unimportant, and blue are called shadow features from *Boruta.* Shadow features are shuffled copies of all features to add randomness to the *Boruta* algorithm.

2178

2179 **CART Analysis to Determine Thresholds Representing Potential Habitat Breaks**

2180 The CART model derived from the random forest analysis included the six most important variables
2181 (Figure 2; Table 9) and the CART model derived from the interaction forest analysis included the top
2182 three interaction pairs (Table 5; Table 9).

2183

2184 **Table 9.** **CART model summaries. The basis for the variables included in the CART models and comparison**
2185 **metrics are included in the table.**

| Informed CART Model	Variables Included	Prediction Accuracy	Sensitivity	Specificity	MCC
Random Forest	BFW10.m Avg.Sus.Grad.Dn Max.Dn.Grad Max.Up.Grad Max.Dn.Step.bfw10 Max.Up.Step.bfw10	90.06% [86.97-92.63%]	0.84	0.95	0.80
Interaction Forest (Pairs)	Eff.Step.Ht.m/Eff.Step.Ht.bfw Eff.Grad/Del.Eff.Grad.Dn Avg.Sus.Grad.Up/Del.Sus.Grad.UpDn	67.82% [63.35-72.06%]	0.43	0.86	0.32
Board Criteria	Avg.Sus.Grad.Up Avg.Sus.Grad.Dn BFW10.m Eff.Grad Eff.Step.Ht.m Eff.Step.Ht.bfw	71.27% [66.92-75.36%]	0.52	0.86	0.40
Random Forest 3 splits	BFW10.m Avg.Sus.Grad.Dn Max.Dn.Grad Max.Up.Grad Max.Dn.Step.bfw10 Max.Up.Step.bfw10	89.2% [86.01-91.88%]	0.80	0.96	0.78
Random Forest 2 splits	BFW10.m Avg.Sus.Grad.Dn Max.Dn.Grad Max.Up.Grad Max.Dn.Step.bfw10 Max.Up.Step.bfw10	88.34% [85.06-91.12%]	0.80	0.94	0.76

2186
2187

2188 The CART model informed by the random forest analyses produced an accuracy of 90.06% (Table 9). The
2189 high accuracy is mostly due to the correct classification of no-fish stream segments (specificity = 0.95),
2190 and limited by the correct classification of fish bearing segments predicted as non-fish (sensitivity =
2191 0.84). All six of the included variables were deemed important by the CART model (Table 10). The overall
2192 classification capacity of the model, as represented by MCC, is relatively high at 0.80. The root node
2193 splits at a maximum downstream gradient of 39%; segments with a maximum downstream gradient less
2194 than 39 were split into a final leaf node classified as fish (Figure 9). Segments with a maximum
2195 downstream gradient greater than 39% were further split at a decision node for maximum upstream
2196 gradient of 175%. Segments with a maximum upstream gradient greater than or equal to 175% were
2197 further split by Max.Dn.Step.bfw.10 of 0.28 m. Overall there are eight splits in this decision tree
2198 representing putative thresholds for PHBs if the thresholds are considered relative to the root node and
2199 the decision nodes above each leaf. For example, if a segment has less than an average sustained
2200 gradient of 8.5% it should only be used as a threshold if the decision nodes above it are considered,
2201 including a maximum upstream gradient of less than 85%, maximum downstream gradient of less than
2202 62%, maximum upstream gradient of greater than or equal to 50%, but less than 175%, and a maximum
2203 downstream gradient of 39%. A decision tree of this length, while more accurate, may be impractical for
2204 application in the field and a pruned model may be more beneficial.

2205
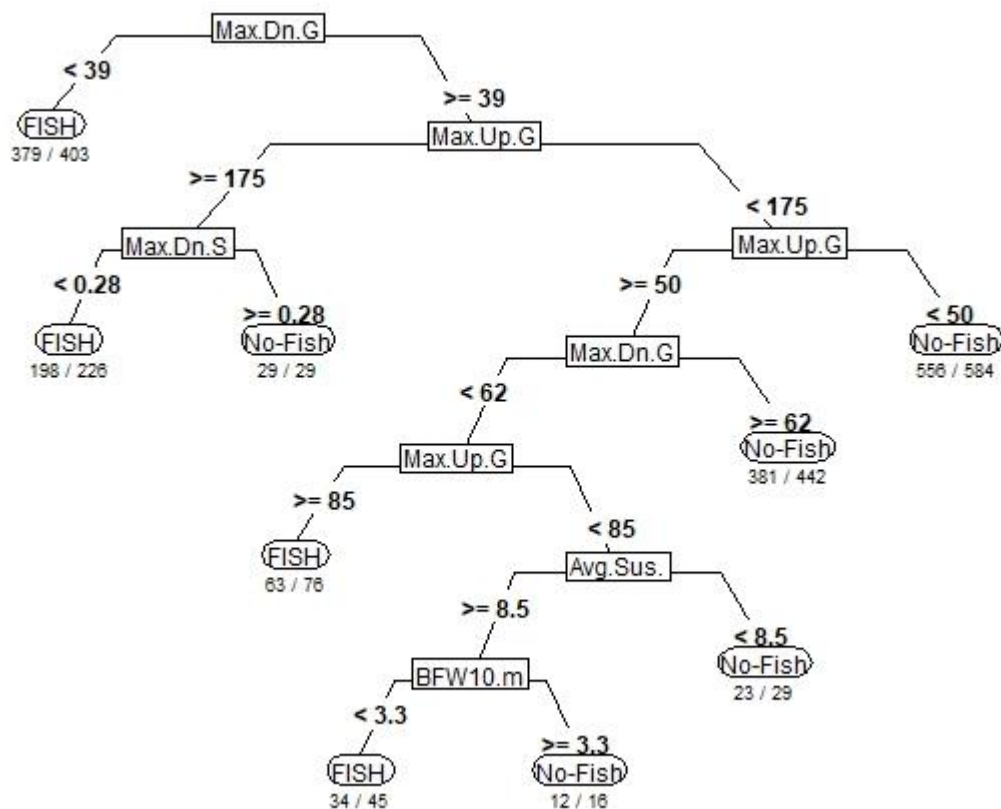
Variable Name	Variable Abbreviation
Max.Dn.Grad	Max.Dn.G
Max.Up.Grad	Max.Up.G
Max.Up.Step.bfw10	Max.up.S
Avg.Sus.Grad.Dn	Avg.Sus.
Max.Dn.Step.BFW10	Max.Dn.S
BFW10.m	BFW10.m

2206　**Figure 9. Decision tree for the CART model informed by the most influential variables in the random forest**
2207　　　**analysis.**
2208
2209

Table 10. Variable Importance for the Random Forest informed CART model.

Variable	Importance Value
Max.Dn.Grad	41
Max.Up.Grad	23
Max.Up.Step.bfw10	16
Avg.Sus.Grad.Dn	10
Max.Dn.Step.BFW10	5
BFW10.m	5

Table 11. Confusion Matrix for the Random Forest informed CART model.

	Reference	
Prediction	Fish	No-Fish
Fish	166	14
No-Fish	32	251

2210
2211
2212
2213　The CART model incorporating the three sets of variables with the highest interaction strength produced
2214　a substantially lower accuracy of 67.82% when compared to the random forest, 90.06% (Table 9).
2215　Although six variables were included in this model only three were deemed important, and the two
2216　most important variables (Del.Sus.Grad.UpDn/Avg.Sus.Grad.Up) were, surprisingly, not the two with the
2217　strongest interaction strength (Eff.Step.Ht.m/Eff.Step.Ht.bfw) (Table 12). The low sensitivity (0.43) for

2218   this model indicates that the correct classification of fish bearing segments was low. Only 83 out of 194
2219   fish bearing segments were correctly classified (Table 13). The low sensitivity may have impacted the
2220   MCC score of 0.32 demonstrating an overall poor classification performance. However, the specificity
2221   was relatively high at 0.86, demonstrating that the non-fish bearing segments were classified correctly.
2222
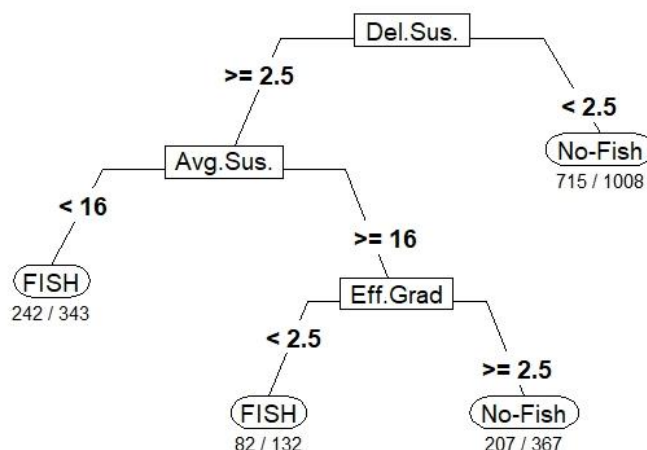
Variable Name	Variable Abbreviation
Del.Sus.Grad.UpDn	Del.Sus
Avg.Sus.Grad.Up	Avg.Sus
Eff.Grad	Eff.Grad

2223   **Figure 10. Decision tree for the CART model informed by the most influential variables in the interaction**
2224   **forest analyses.**
2225
2226

Table 12. Variable Importance for the Interaction Forest informed CART model.

Variable	Importance Value
Del.Sus.Grad.UpDn	56
Avg.Sus.Grad.Up	34
Eff.Grad	9

Table 13. Confusion Matrix for the Interaction Forest informed CART model.

	Reference	
Prediction	Fish	No-Fish
Fish	83	38
No-Fish	111	231

2227
2228
2229   The CART model incorporating the variables used for the Board Criteria produced an accuracy, 71.27%,
2230   similar to that of the Interaction Forest, 67.82% (Table 9). Like the other models, the sensitivity was
2231   lower (0.52) indicating poor performance for classifying fish bearing segments correctly, but specificity
2232   was high (0.86). The root node splits for BFW10.m at 2.7 m (Figure 11). The subsequent decision nodes
2233   for those segments greater than or equal to 2.7 m BFW10.m are further split by an average sustained
2234   downstream gradient of 12% whereas those segments less than 2.7 m BFW10.m are split by an average
2235   sustained gradient of 2%. Four variables were deemed important including Avg.Sus.Grad.Dn, BFW10.m,
2236   Avg.Sus.Grad.Up, and Eff.Grad (Table 14), however, only the top three variables were influential enough
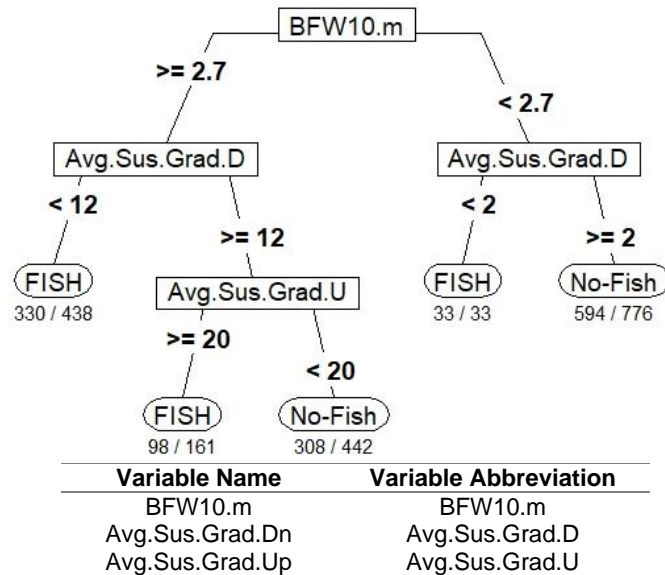2237   to warrant a split in the decision tree (Figure 11).
2238

Variable Name	Variable Abbreviation
BFW10.m	BFW10.m
Avg.Sus.Grad.Dn	Avg.Sus.Grad.D
Avg.Sus.Grad.Up	Avg.Sus.Grad.U

2239
2240 **Figure 11. Decision tree for the CART model informed by the Board criteria.**
2241
2242

**Table 14. Variable Importance for the Board criteria informed CART model.**

Variable	Importance Value
Avg.Sus.Grad.Dn	44
BFW10.m	33
Avg.Sus.Grad.Up	19
Eff.Grad	4

Table 15. Confusion Matrix for the Board criteria informed CART model.

|	Reference	
Prediction	Fish	No-Fish
Fish	100	39
No-Fish	94	230

2243
2244 Given the strong performance of the Random Forest informed CART model we wanted to test how well
2245 the model performed if pruned to three or two splits instead of allowing the CART model to select the
2246 optimal number of splits. The pruned models performed similarly to the full Random Forest informed
2247 CART model with accuracies of 88.34% for the two split and 89.2% for the three split models (Table 9).
2248 Additionally, the MCC scores remained similar to the overall model (0.8) with 0.78 for the three split and
2249 0.76 for the two split. The two and three split models have the same sensitivity (0.8) but differ in
2250 specificity by the correct classification of four additional stream segments in the three split (Table 16) vs.
2251 the two split model (Table 17). The thresholds established in the Random Forest informed CART model
2252 are the same; only the number of splits (Figure 12 & 13) and, therefore, the distribution of segment
2253 classification has changed (Table 16 & 17).
2254
2255
2256
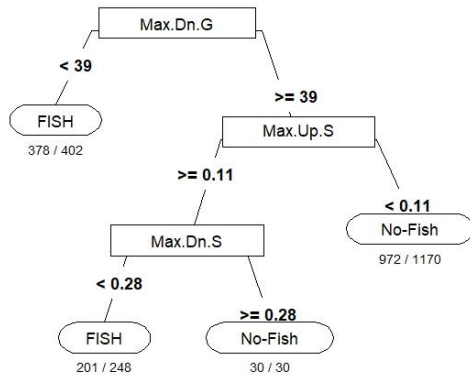
Variable Name	Variable Abbreviation
Max.Dn.Grad	Max.Dn.G
Max.Up.Step.bfw10	Max.up.S
Max.Dn.Step.BFW10	Max.Dn.S

Figure 12. Decision tree for the CART model informed by the Random Forest model with only three splits.
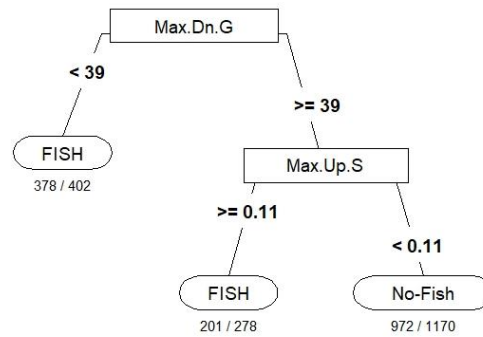
Variable Name	Variable Abbreviation
Max.Dn.Grad	Max.Dn.G
Max.Up.Step.bfw10	Max.up.S

Figure 13. Decision tree for the CART model informed by the Random Forest model with only two splits.

Table 16. Confusion Matrix for the Random Forest informed CART model with only three splits.

Prediction	Reference	
	Fish	No-Fish
Fish	155	11
No-Fish	39	258

Table 17. Confusion Matrix for the Random Forest informed CART model with only two splits.

Prediction	Reference	
	Fish	No-Fish
Fish	155	15
No-Fish	39	254

DISCUSSION

In this example analysis with pilot data, we demonstrated that random forest models and interaction forest models can classify presence of fish on stream segments in Washington State with greater than 90% accuracy. More importantly, random forest and interaction forest enabled a multivariate analysis to determine which variables best described areas with fish and without fish, including stream gradient, steps or barrier height, bankfull width, and other characteristics. Interaction forests outperformed random forest models based on model accuracy, kappa, and specificity, and helped identify key parameters that in combination influence end of fish. These results correspond with findings from a comparison of random forest and interaction forest classification models on 220 different data sets (Hurnung and Boulesteix 2022). Given the lower accuracy of classifying eastern Washington stream segments, a larger sample in conjunction with an interaction forest approach may improve model performance in future analyses.

Applying a CART model to the variables identified in the random forest, interaction forest, and the board criteria demonstrated that thresholds for potential habitat breaks can be established using a decision tree with relatively high accuracy even when pruned. Notably, using the variables from the random forest model in a CART analysis resulted in an accuracy of 90.06% and accuracy was only reduced to 88.34% when the decision tree was pruned to only two splits. Including the

2275 variables, but not the thresholds, selected by the Board resulted in higher accuracies in the CART
2276 analysis 71.27% than applying the Board criteria to the available data set (48.36%-62.52%; Table
2277 6). Across all CART models, root nodes included maximum downstream gradient, change in the
2278 sustained gradient, and bankfull width measurements. Decision nodes (below the root) included
2279 maximum downstream and upstream gradients, average sustained upstream and downstream
2280 gradient, effective gradient at the segment, and bank full width. Further investigation into the
2281 distribution of these thresholds on stream longitudinal profiles will be essential for their utility. The
2282 threshold values described at each split should not be extracted nor viewed in isolation from the
2283 previous nodes. Doing so may lead to misinterpretation when the same variable is used at
2284 different nodes on the same tree.
2285
2286 Evaluating the FPB criteria by comparing observed fish presence for sets of criteria with random
2287 forest models resulted in relatively low accuracies. Reducing a continuous habitat covariate to a
2288 binary indicator may reduce the predictive power of the random forest model if the cutoff point
2289 used to create the binary indicator is not closely associated with the end of fish. TestCriteria 2
2290 and TestCriteria3 were not met by any segments in the pilot data set, but we anticipate that these
2291 criteria will be incorporated into future analyses. To more adequately evaluate the criteria
2292 following additional sampling, we recommend measuring all steps, not just those presumed to
2293 cause a barrier to reduce bias in the gradient and barrier parameters.
2294
2295 The random forest and interaction forest analyses demonstrated that certain stream features are
2296 useful predictors of fish versus non-fish habitat. Application of the random forest results to the
2297 CART analysis gets us closer to the ultimate objective of describing the inflection point or
2298 transition at the end of fish. Box and violin plots in Appendix A were added to qualitatively assess
2299 the stepwise progression from average fish habitat, habitat near end of fish, and habitat without
2300 fish. These plots in conjunction with the CART models provide an empirical basis for establishing
2301 criteria for habitat covariates.
2302

2303 # REFERENCES

2304 Benhaiem, S., L. Marescot, H. Hofer, M. L. East, J.-D. Lebreton, S. Kramer-Schadt, and O. Gimenez. 2018.
2305 Robustness of Eco-Epidemiological Capture-Recapture Parameter Estimates to Variation in
2306 Infection State Uncertainty. Frontiers in Veterinary Science 5: 197. doi: 10.3389/fvets.2018.00197.

2307 Chicco, D., and Jurman, G. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1
2308 score and accuracy in binary classification evaluation. BMC Genomics. 21(6).

2309 Cutler, D. R., T. C. Edwards, Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007.
2310 Random Forests for Classification in Ecology. Ecology 88(11): 2783-2792. doi: 10.1890/07-0539.1.

2311 Friedman, J. H. and B. E. Popescu. 2008. Predictive Learning Via Rule Ensembles. Annals of Applied
2312 Statistics 2(3): 916-954. doi: 10.1214/07-AOAS148.full.

2313 Greenwell, G.M and B. C. Boehmke. 2020. Variable Importance Plots—An Introduction to the vip Package.
2314 The R Journal, 12(1), 343--366. URL https://doi.org/10.32614/RJ-2020-013.

2315   Hornung, R. 2022. Diversity Forests: Using Split Sampling to Enable Innovative Complex Split Procedures
2316   in Random Forests.SN Computer Science 3:1. doi: 10.1007/s42979-021-00920-1.

2317   Hornung, R. and L. A. Boulesteix. 2022. Interaction Forests: Identifying and Exploiting Interpretable
2318   Quantitative and Qualitative Interaction Effects. Computational Statistics and Data Analysis 171:
2319   107460. doi: 10.1016/j.csda.2022.107460.

2320   Johnstone, R. P. H. 2021. Landscape Effects on Breeding Habitat Selection and Incubation Behaviour in
2321   Boreal Nesting Ducks. Thesis. University of Waterloo, Waterloo, Ontario, Canada.

2322   Khodadadi, A. and M. Asgharian. 2008 Change-Point Problem and Regression: An Annotated Bibliography.
2323   COBRA Preprint Series. Working Paper 44. Available online: https://biostats.bepress.com/
2324   cobra/art44

2325   Killick, R., P. Fearnhead, and I. A. Eckley. 2012. Optimal Detection of Changepoints with a Linear
2326   Computational Cost. Journal of the American Statistical Association 107(500): 1590-1598. doi:
2327   10.1080/01621459.2012.737745.

2328   Kuhn, M. 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software,
2329   28(5): 1-26. doi: 10.18637/jss.v028.i05.

2330   Kursa, M. B. and W. R. Rudnicki. 2010. Feature Selection with the Boruta Package. Journal of Statistical
2331   Software 36(11): 1-13. doi: 10.18637/jss.v036.i11.

2332   Large, S. I., G. Fay, K. D. Friedland, and J. S. Link. 2013. Defining Trends and Thresholds in Responses
2333   of Ecological Indicators to Fishing and Environmental Pressures. ICES Journal of Marine Sciences
2334   70(4): 755-767. doi: 10.1093/icesjms/fst067.

2335   Lewis, M. C. 2021. Out-Migration Dynamics of Juvenile Adfluvial Bull Trout in Tributaries to the Lower Clark
2336   Fork River, Montana. Dissertation. Montana State University-Bozeman, Montana.

2337   Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. R News 2(3): 18-22.

2338   Lloyd, K. J., W. C. Oosthuizen, M. N. Bester, and P. N. de Bruyn. 2020. Trade-Offs Between Age-Related
2339   Breeding Improvement and Survival Senescence in Highly Polygynous Elephant Seals: Dominant
2340   Males Always do Better. Journal of Animal Ecology 89: 897-909.

2341   Mahmud, S., M. A. Islam, and S. S. Hossain. 2020. Analysis of Rainfall Occurrence in Consecutive Days
2342   Using Markov Models with Covariate Dependence in Selected Regions of Bangladesh. Theoretical
2343   and Applied Climatology 140(3): 1419-1434. doi: 10.1007/s00704-020-03159-7.

2344   McClintock, B. T., R. Langrock, O. Gimenez, E. Cam, D. L. Borchers, R. Glennie, and T. A. Patterson. 2020.
2345   Uncovering Ecological State Dynamics with Hidden Markov Models. Ecology Letters 23(12):
2346   1878-1903. doi: 10.1111/ele.13610,

2347   Molnar, C. 2022. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
2348   Second edition. Independent publisher. Available online: http://christophm.github.io/interpretable-
2349   ml-book/

2350   Molnar, C., G. Casalicchio, and B. Bernd. 2018. iml: An R Package for Interpretable Machine Learning.
2351   Journal of Open Source Software 2.26: 786.

2352   Muggeo, V. M. 2008. segmented: An R Package to Fit Regression Models with Broken-Line Relationships.
2353   R News 8(1): 20-25.

2354 Patterson, T. A., M. Basson, M. V. Bravington, and J. S. Gunn. 2009. Classifying Movement Behaviour in
2355   Relation to Environmental Conditions Using Hidden Markov Models. Journal of Animal Ecology 78:
2356   1113-1123. doi: 10.1111/j.1365-2656.2009.01583.x.

2357 R Core Team. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical
2358   Computing, Vienna, Austria. Available online: https://www.R-project.org/

2359 Roni, P., K. Ross, H. Berge, P. Bisson, J. Kershner, J. Maroney, P. Trotter, J. Walter. 2018. Potential Habitat
2360   Break (PHB) pilot study final report. Prepared for Washington Department of Natural Resources,
2361   Olympia, WA.

2362 Schmidt, J. H., D. S. Johnson, M. S. Lindberg, and L. G. Adams. 2015. Estimating Demographic Parameters
2363   Using a Combination of Known-Fate and Open *N*-Mixture Models. Ecology 96: 2583-2589. doi:
2364   10.1890/15-0385.1.

2365 Shogren, A. J., J. P. Zarnetske, B. W. Abbott, S. Bratsman, B. Brown, M. P. Carey, R. Fulweber, H. E.
2366   Greaves, E. Haines, F. Iannucci, and J. C. Koch. 2022. Multi-Year, Spatially Extensive,
2367   Watershed-Scale Synoptic Stream Chemistry and Water Quality Conditions for Six Permafrost-
2368   Underlain Arctic Watersheds. Earth System Science Data 14(1): 95-116. doi: 10.5194/essd-14-95-
2369   2022.

2370 Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and Zeileis, A. 2008. Conditional Variable Importance for
2371   Random Forests. BMC Bioinformatics 9: 307. doi: 10.1186/1471-2105-9-307.

2372 Strobl, C., A. L. Boulesteix, A. Zeileis, and T. Hothorn, 2007. Bias in Random Forest Variable Importance
2373   Measures: Illustrations, Sources and a Solution. BMC Bioinformatics 8: 25. doi: 10.1186/1471-
2374   2105-8-25.

2375 Therneau T., Atkinson B. 2022. rpart: Recursive Partitioning and Regression Trees. R package version
2376   4.1.19, https://CRAN.R-project.org/package=rpart.

2377 Washington Department of Natural Resources. 2019. Draft Rule Proposal for a Permanent Water Typing
2378   System, dated April 22, 2019. *In:* Forest Practices Board meeting materials for Board meeting of
2379   May 8, 2019. Available online: https://www.dnr.wa.gov/publications/bc_fpb_mtgmaterial
2380   _20190508.pdf

2381 Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and Management of Animal Populations.
2382   Academic Press, San Diego, California.

2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393   **Appendix A. Additional Figures**
2394

2395   Box and violin plots for the distribution of variables deemed important by the random forest analyses. The
2396   plots include stream segments designated as fish, end of fish (EOF), one segment above end of fish
2397   (EOF+1; EOF_Plus_1), and no-fish. Segments at EOF and EOF+1 were not double counted, and thus
2398   represent the average for a particular value at the potential habitat break. Figures are in the order of variable
2399   importance based on the *Full Random* model.
2400
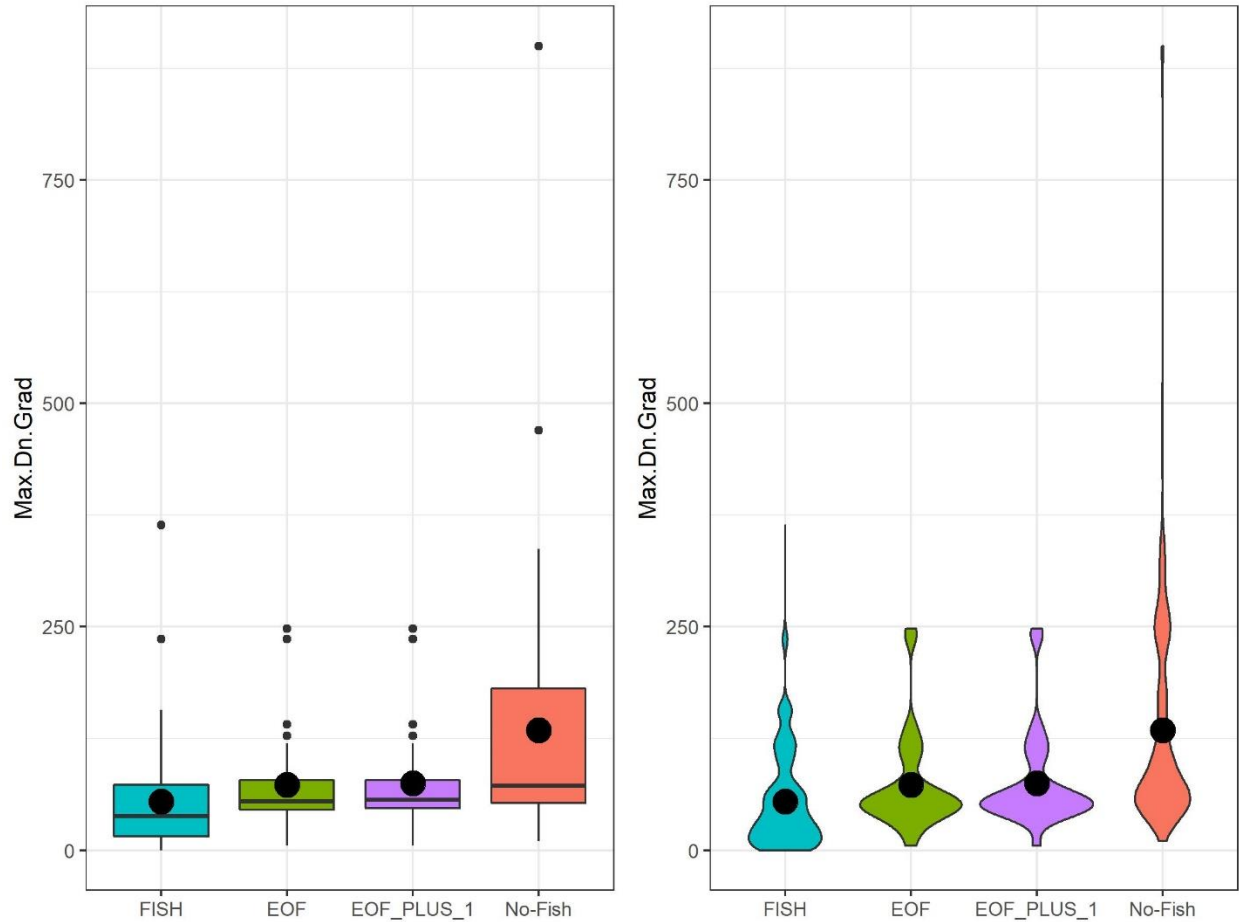
2401

2402

2403

2404

2405
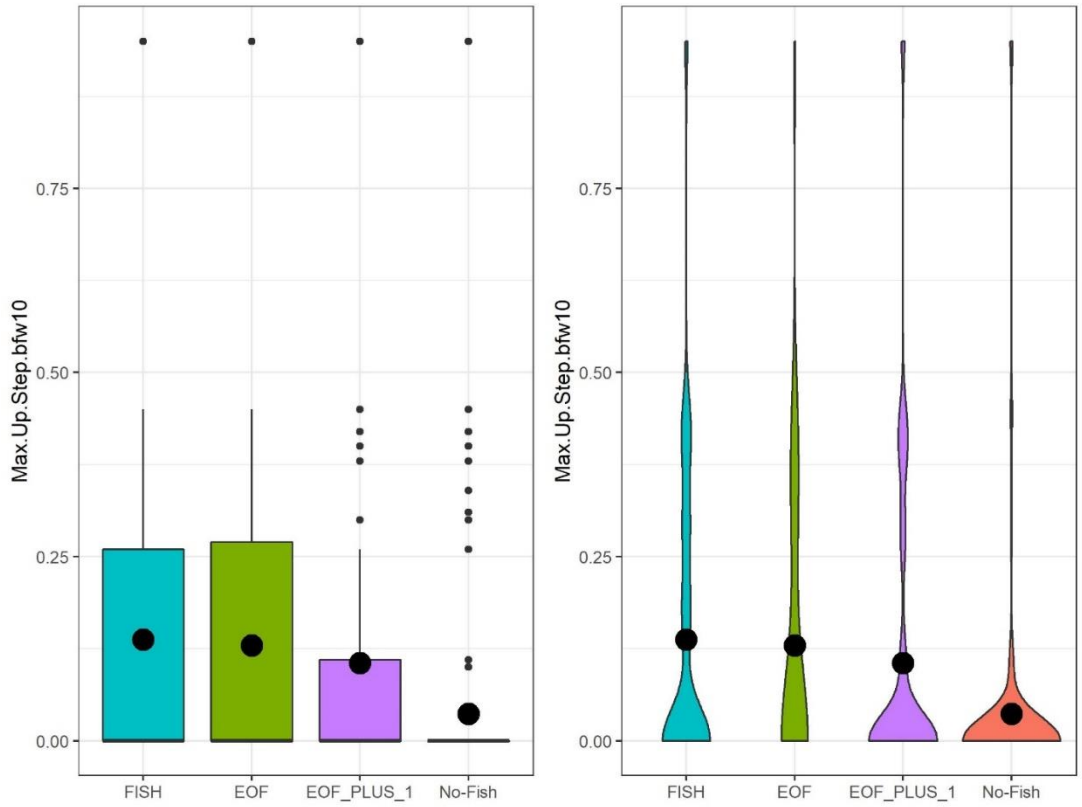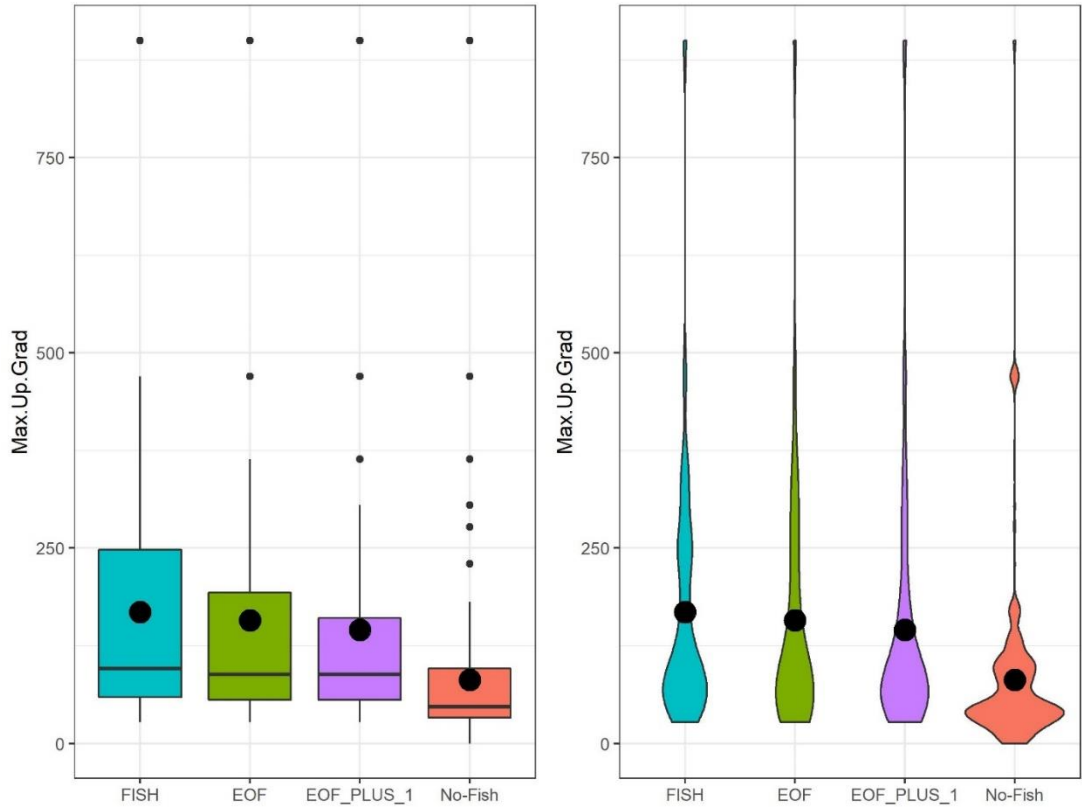
2406
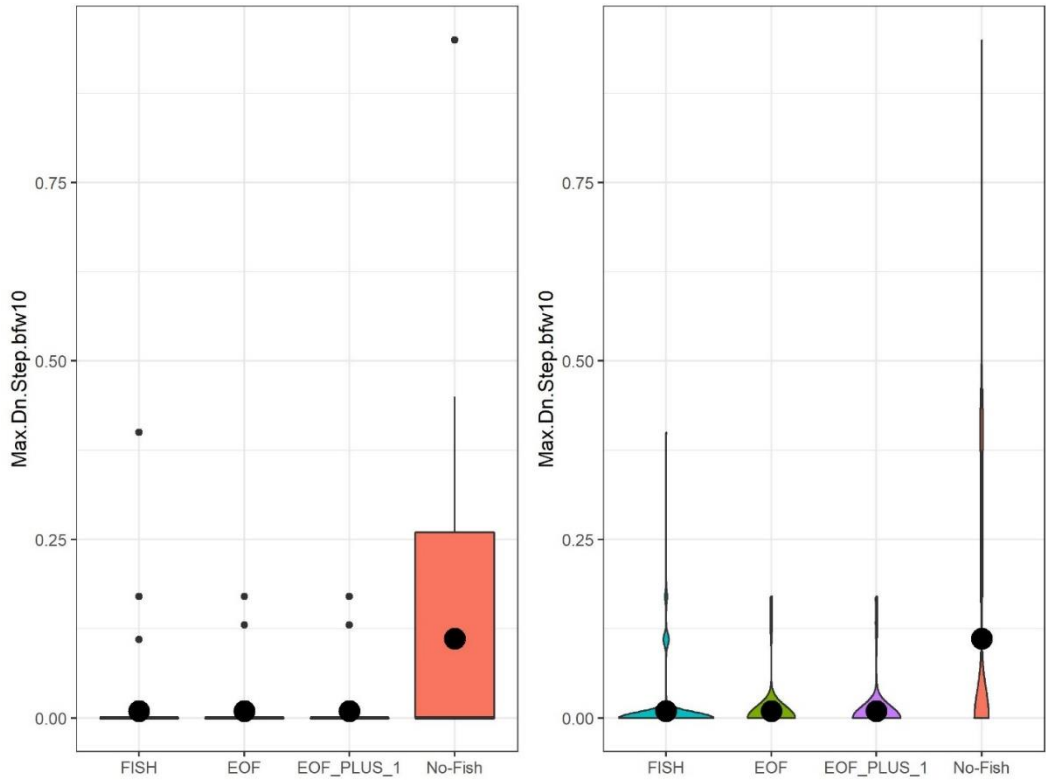
2407

2408

2409

2410

2411

2412
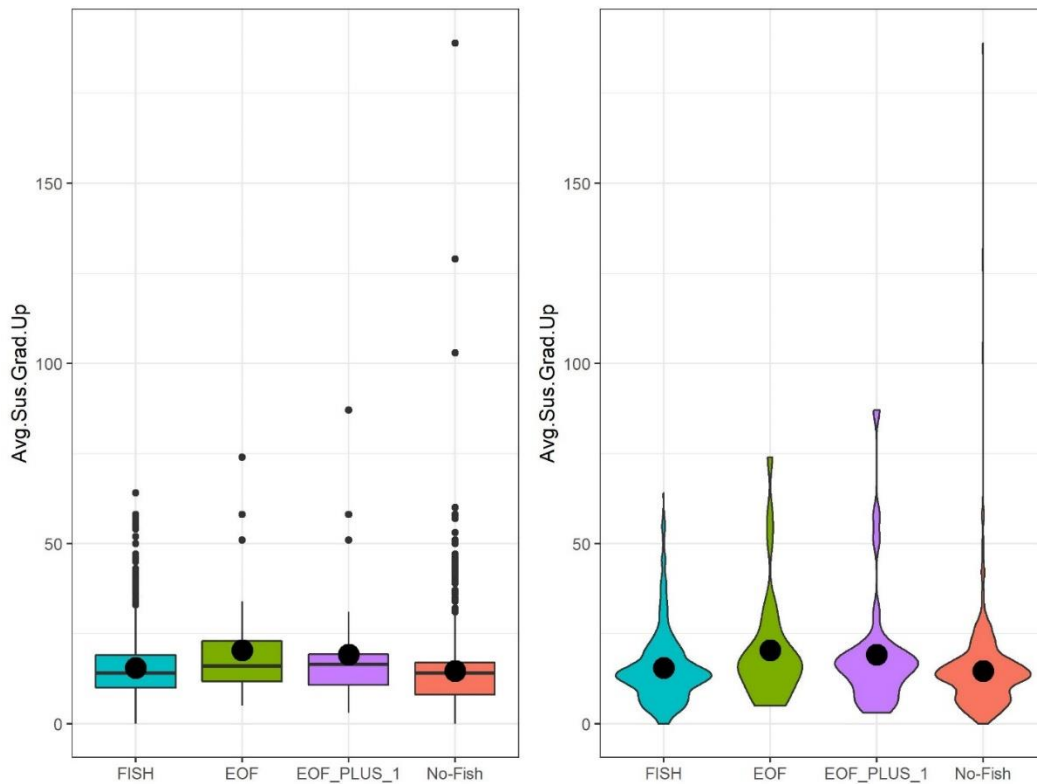
2413

2414
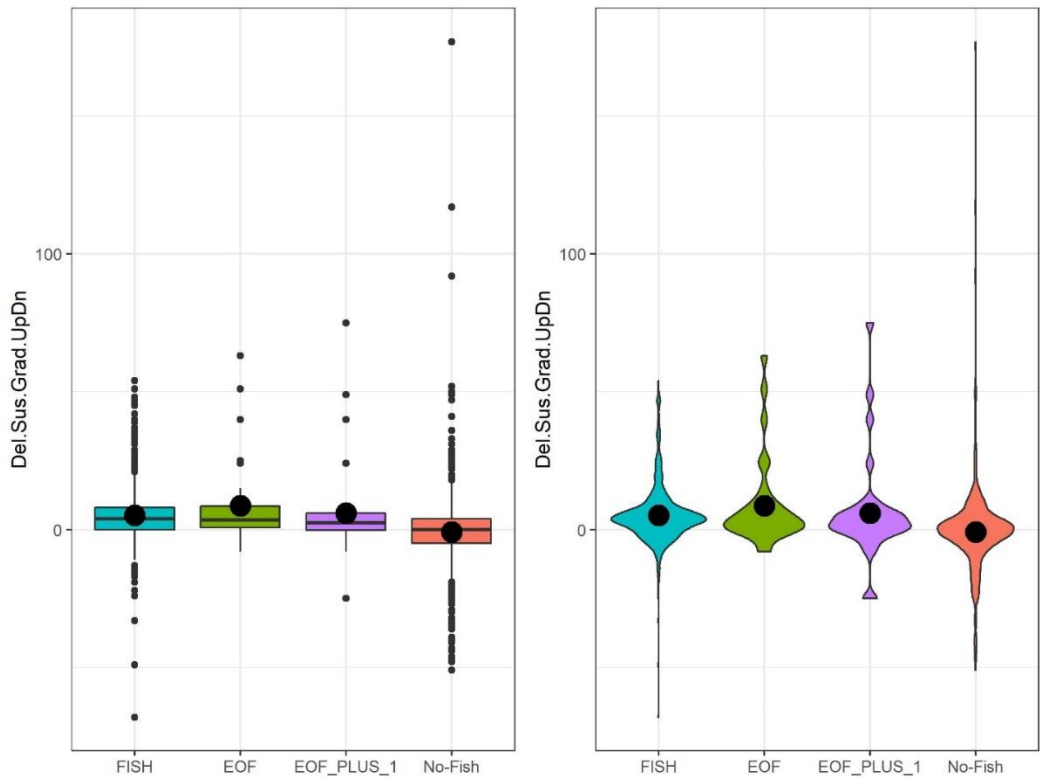
2415

2416
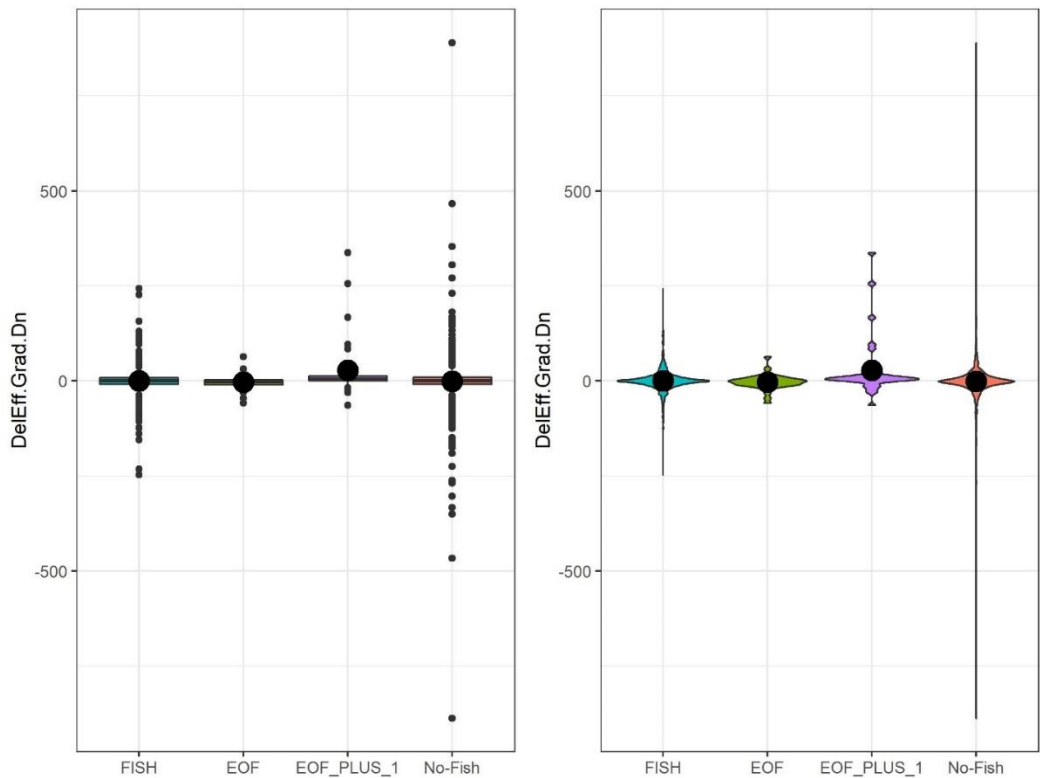
2417

2418
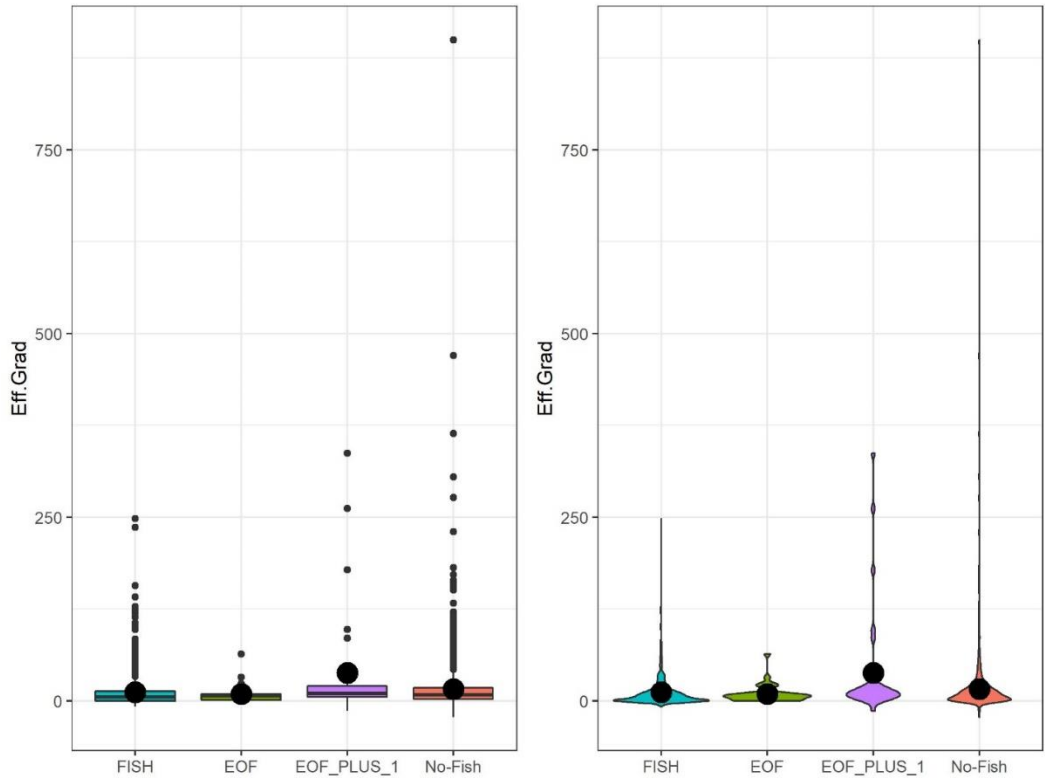
2419

2420
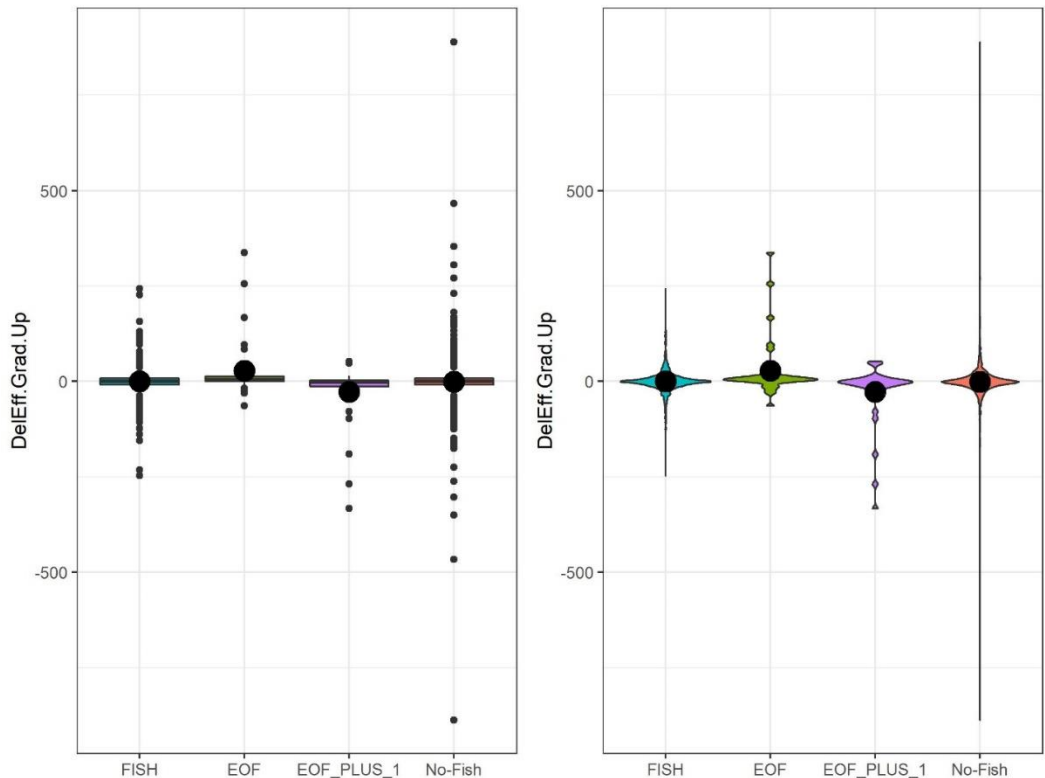
2421

2422
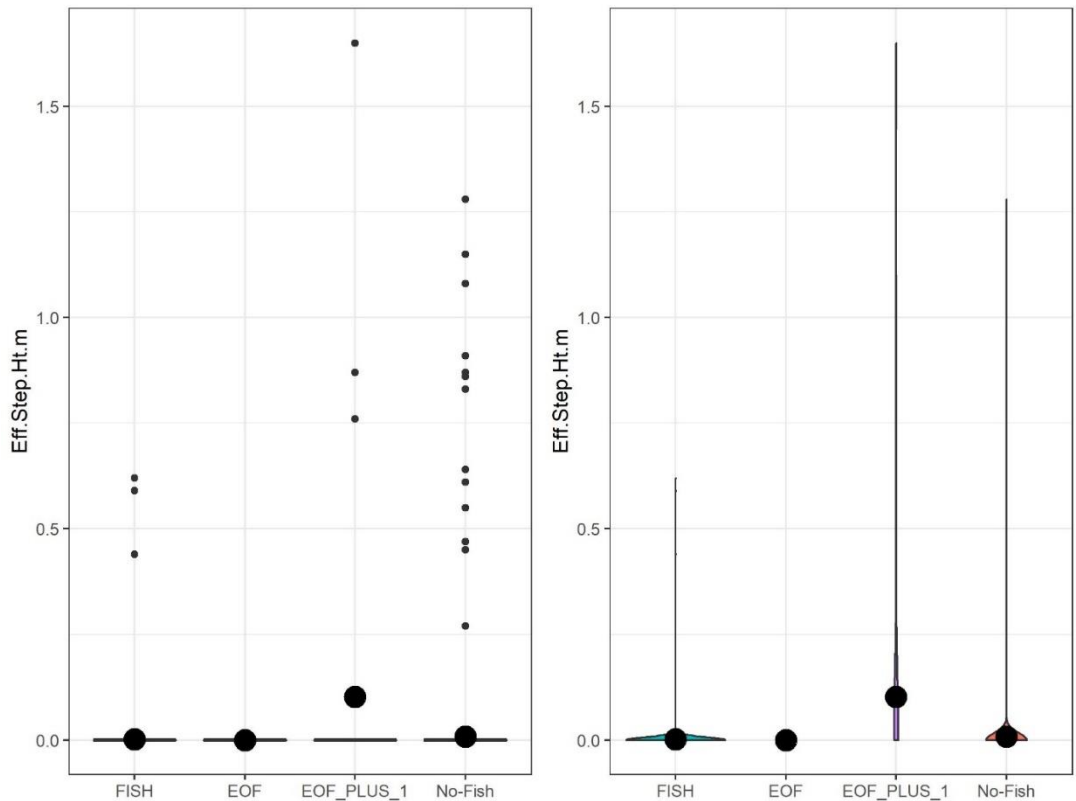
2423
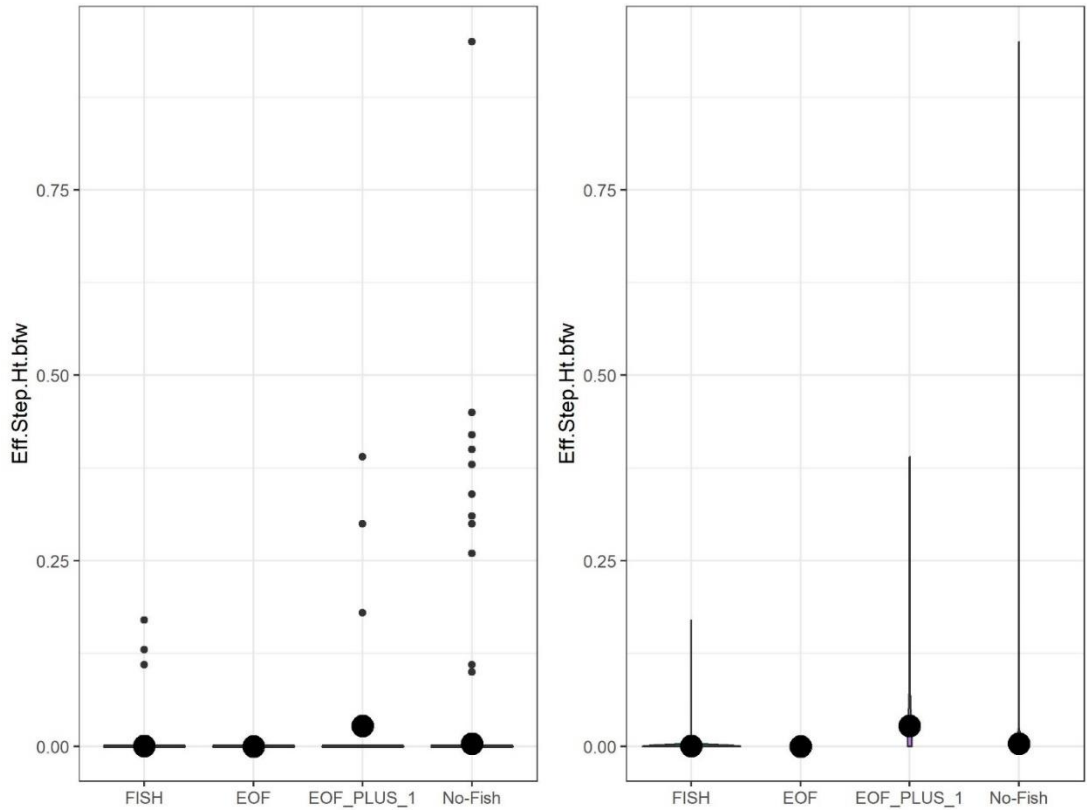
2424

2425
2426
2427
2428
2429
2430
2431
2432
2433
2434

2435   **Appendix B. Modeling Covariate Data Dictionary**
2436

2437

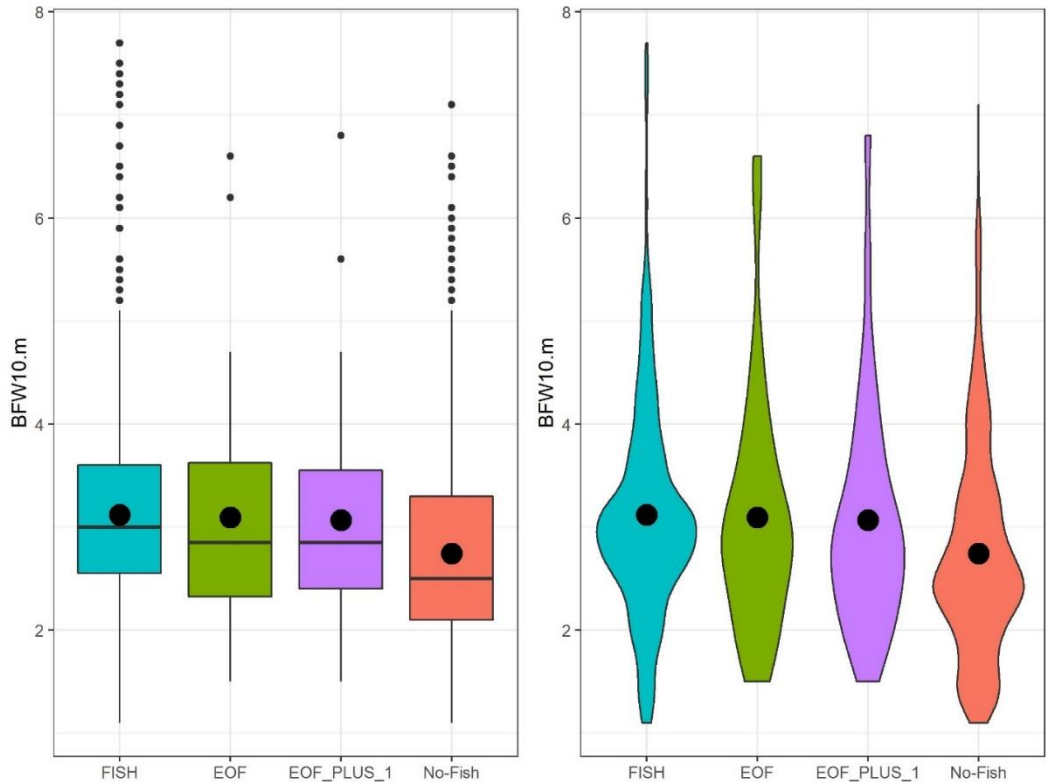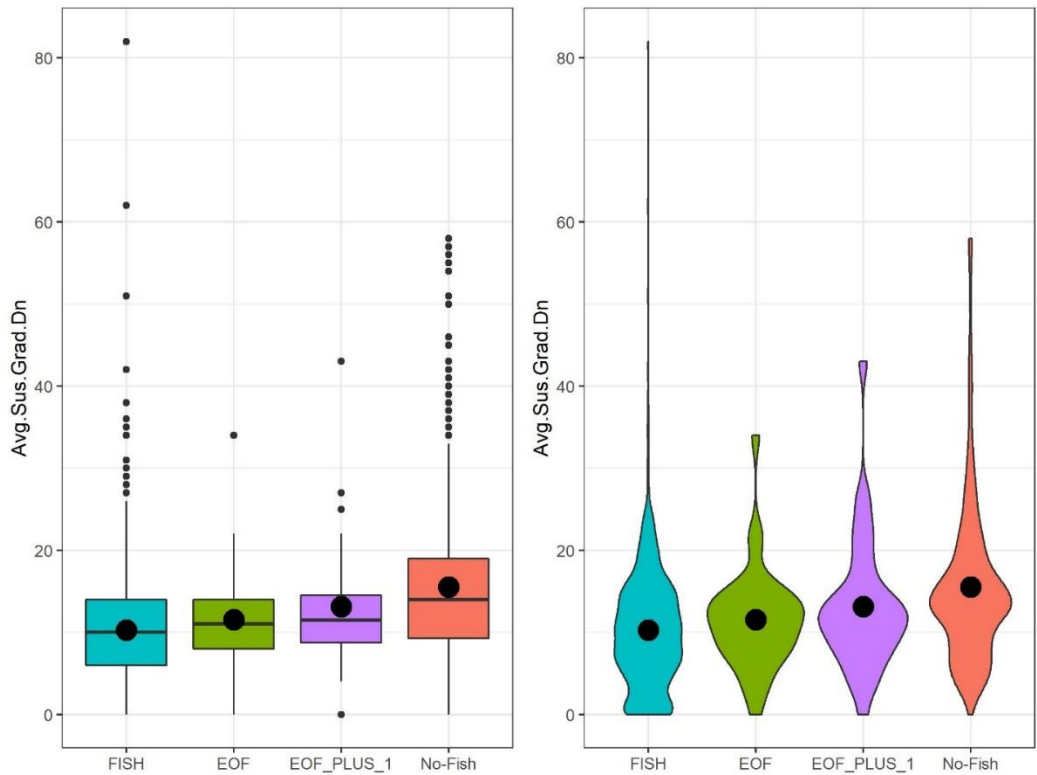| Variable	Definition
StreamName	copied main stream data to each tributary for separate evaluation, consistent with pilot study analysis.
Station	Survey station
DelDistance	Length of segment (m)
CumulativeDistance	Distance from start of survey (m)
Substrate	
Comments	
EOFpt	In methodology, be clear that this is the last segment WITH fish; EOF pt is at top of segment.
FISH/NO-FISH	fish are assumed to use all segments below the EOF station
Flow Condition	flowing/dry
UnitLabel	Unit type modified for use in PHB analysis Riffle/Pool/Step Step defined as >150% gradient based on pilot study. Step-Pool is when gradient is >8% and Substrate = Fines or Sand (not implemented) If Unit = Riffle but elevation change is <= 0, Unit was changed to Pool
EffectiveGrad_pct	Based on Effective Elevation Change, which sets pool elevations to the elevation of the tail-out (riffle or step downstream of pool) Add in functionality to figure out (presumed) head of pool and calculate gradient above that only? Subgroup decided 6/16/2022 not to bother for the purposes of this pilot, but real study must.

Variable	Definition
EffectiveStepHeight_m	Change in effective elevation (elevation - previous elevation or pool residual elevation) for a segment having gradient >=150%
EffectiveStepHeight_BFW	Change in effective elevation (elevation - previous elevation or pool residual elevation) for a segment having gradient >=150% reported in multiples of the BFW10 at each station (col BA)
DelEffectiveGradFromDnstrmSeg	Change in effective gradient from downstream segment
DelEffectiveGradToUpstrmSeg	Change in effective gradient to next segment upstream
BFW10_m	includes 10 stations, per WAC definition (as close as we can reasonably get); five stations below, the present station, and four stations above; bedrock units excluded from average calculation
AvgSusGradDnstrm	includes 20 segments downstream (19 stations below plus this one) stations, per WAC definition (as close as we can reasonably get)
AvgSusGradUpstrm	includes 20 segments upstream (20 stations above) stations, per WAC definition (as close as we can reasonably get)
MaxDnstrmGrad	Requires that data be ordered by StreamName and Station Maximum segment effective gradient downstream of each station

Variable	Definition
MaxUpstrmGrad	Requires that data be ordered by StreamName and Station Maximum segment gradient upstream of each station
MaxDnstrmStep_BFW10	The maximum step downstream of the present station, in multiples of BFW10
MaxUpstrmStep_BFW10	The maximum step upstream of the present station, in multiples of BFW10
BFW_Dn10	Average of the BFW for the 10 segments downstream of current station (m)
BFW_Up10_m	Average of the BFW of the 10 segments upstream of the current station (m)
BFW_Up20_m	Average of the BFWs for the 20 segments upstream of the current station (m)
BFW_Up20_ft	Average of the BFWs for the 20 segments upstream of the current station (ft)

2438
2439

Appendix D. Potential for a concurrent eDNA study

2440

2441 The original study design (PHB Science Panel 2019) included a proposed collaborative
2442 complementary study with the U.S. Forest service to compare environmental DNA (eDNA) and
2443 electrofishing to identify fish habitat. A separate pilot for that proposed complementary study
2444 was completed in 2020 (Penaluna 2020).

2445 The project team explored ways to include further eDNA components into this study design.
2446 The team determined that the best option would be to recommend that an additional
2447 complementary study is developed by the Adaptive Management Program that utilizes the
2448 sample sites and the fish location data that are collected in this study. This companion study
2449 can further compare electrofishing and eDNA as methods for determining the location of the
2450 upper extent of fish use, as well as different methods for eDNA collection and analysis, and can
2451 take advantage of the lessons learned from the pilot study. Conducting a complementary study
2452 in conjunction with the PHB study might save time, money, and resources.

2453 # Appendix E. Budget

2454 Budget estimate from DNR PM Anna Toledo as of February 18, 2022. Estimates are based on figures updated from the FY19 study design,
2455 expenditures from the FY19 pilot study, and existing contract budgets for similar work. These estimates may change based on revisions
2456 made during CMER, ISAG, and ISPR reviews.

Task	Expenditures FY17-FY21	FY22	FY23	FY24	FY25	FY26	FY27	FY28	FY29	Total
Study design, coordination, site reconnaissance, permitting, crew training		31,247	69,250	163,679	114,167	30,512	30,918	N/A	N/A	439,773
Field sampling – Spring/summer (350 sites)					723,697	723,433	737,901	N/A	N/A	2,185,031
Field sampling – Fall/winter (175 sites: fixed + alternating panels)					N/A	176,389	179,917	183,515	N/A	539,821
Crew variability (10% of sites – all crews)					57,944	55,028	56,129	25,505	N/A	194,606
Data collection equipment					183,600	27,540	27,540	27,540	N/A	266,220
Data analysis and reporting				12,485	39,202	67,832	69,189	94,796	61,229	344,733
Project Management				9,364	15,918	16,236	16,561	10,930	4,460	73,469
Total	**398,702**	**31,247**	**69,250**	**185,528**	**1,134,529**	**1,096,970**	**1,118,155**	**342,286**	**65,689**	**4,442,355**

2457

2458 **Budget Comparison**

2459 Comparison of original study design and revised study design budgets. Original study design

2460 budget and tasks in grey.

Task	Original Study Design Totals	Revised Study Design Totals	Notes
Study design, coordination, site reconnaissance, permitting, crew training	421,900	439,773	Revised budget accounts for a 2% yearly increase for inflation/COLA throughout all line items, which was not accounted for in the original budget.
Field sampling – Spring (245 sites)	1,519,000		Total site visits (original): 529 Total site visits (revised): 525
Field sampling – Spring/summer (350 sites)		2,185,031	
Field sampling – Summer (82+60)	460,151		
Field sampling – Fall (82+60); pilot in FY 19	581,151		
Field sampling – Fall/winter (175 sites: fixed + alternating panels)		539,821	
Crew variability (10% of sites – all crews)	115,000	194,606	
Data collection equipment		266,220	Data collection equipment was not a separate line item in original budget.
eDNA sampling (82 sites 3 times)	50,000		eDNA recommended as a complementary study, removed from revised budget.
eDNA Lab Analysis and reporting	164,000		
Data analysis and reporting	180,163	344,733	Budget updated to reflect updated time estimate for analysis and reporting.
Project Management	72,669	73,469	
Total	3,564,034	4,442,355	

2461

2462   # Appendix F.   Data Tables and Attribute Descriptions

2463   **Table F-1. Site selection initial fish survey start point attributes – GIS-derived**

Attribute	Source	Units	Description
SiteID	GIS		Identifier from DNR hydro layer
Stream Name	GIS		Local name
Stream Order	GIS		Strahler Stream Order #
Ecoregion	GIS		DNR Natural Heritage Level III [Northwest Coast, Puget Trough, North Cascades, West Cascades, East Cascades, Okanogan, Canadian Rocky Mountains, Blue Mountains]
Side of State	GIS		Location relative to cascade crest [East, West]
Latitude of currently mapped F/N break	GIS	dd	WGS1984
Longitude of currently mapped F/N break	GIS	dd	WGS1984
Elevation of currently mapped F/N break	GIS	m	
Currently mapped F/N break point type	GIS		Terminal or Lateral
Broad-scale land use class	GIS		Industrial timberland, USFS, small private timberland, conservation forest, residential, other forestry, other non-forest
30-year annual and seasonal normal precipitation	GIS	mm	PRISM model and data from neighborhood reference rain gauges
30-year annual and seasonal normal flows for one or more neighboring gauged streams	Calculated	cms	30-year or as close to that as possible; the point is to be able to place the survey year flow levels in the broader long-term flow context
Seasonal Sampling Scheme	Assigned		Fixed or alternating panel, and if alternating, which of (3) years
Optimal Spring Survey Timing	Assigned		Based on information provided by local/regional experts
Optimal Seasonal Survey Timing	Assigned		Based on information provided by local/regional experts

2464

2465   **Table F-2. Site field attribute table**

Attribute	Source	Units	Description (detail in Methods Manual)
SiteID	GIS		Identifier from DNR Hydro layer
Landscape Reference Point (LRP)	Field		Narrative description of a permanent topographic/physical feature used to help locate the FRPs and LFPs
LRP Latitude	Field	dd	Decimal degrees; WGS 1984
LRP Longitude	Field	dd	Decimal degrees; WGS 1984
Fixed Reference Point (FRP)	Field		Narrative description of FRP closest to initial LF point relative to permanent topographic/physical feature such as a confluence point with mainstem, tributary junction, etc.
FRP Latitude	Field	dd	Decimal degrees; WGS 1984
FRP Longitude	Field	dd	Decimal degrees; WGS 1984
FRP Elevation	Field	m	Will be baseline from which habitat surveys are conducted
Notes	Field		Any features significant at a site level

2466

2467   **Table F-3. Uppermost fish survey data for each survey event; Uppermost fish point (EOF) will be**
2468   **baseline from which habitat surveys are conducted.**

Attribute	Source	Units	Description (detail in Methods Manual)
SiteID	GIS		Identifier from DNR Hydro layer
SurveyID	Assigned		Which survey (year/season)
Date			
Weather Conditions	Field		sunny, rainy, snowy, cloudy
Air Temp	Field	C	
Field Crew			
Fish Survey Start Point	Field	dd, m	Lat, Long, Elev at fish survey start point
Fish Survey Start Water Temp	Field	C	
Stream Conductivity	Field	uS/cm	
Electrofisher Setting	Field		
Fish Survey End Point	Field	dd, m	Lat, Long, Elev at fish survey end point
Fish Survey End Water Temp	Field	C	
EOF Latitude	Field	dd	Decimal degrees; WGS 1984
EOF Longitude	Field	dd	Decimal degrees; WGS 1984
EOF Elevation_GPS	Field	m	NAD83

Attribute	Source	Units	Description (detail in Methods Manual)
EOF Stream Distance From Topographic Reference Point (RP)	Field	m	EOF point field-identifiable location relative to a permanent topographic or physical feature such as a confluence point with mainstem, tributary junction, etc., if feasible Also identify reference objects to help locate
EOF Date-Time	Field		YYYY-MM-DD-24-hour; Standard Time;
EOF WaterTemp	Field	C	To nearest 0.5 C
Upstream-Most Fish Species/Family	Field		When it can be determined (salmonid; sculpin (cottid); stickleback; mudminnow; etc)
Fish Size Category	Field	mm	<25mm, 25-75mm, 75-150mm, >150mm
EOF Point Type	Field		Terminal or Lateral
EOF Flow Status	Field		Flowing, Dry
EOF Habitat Unit Type	Field		Pool, Riffle, Step-Pool, Step (>=2' vertical)
EOF Measurement Point Type	Field		e.g. crest of tailout; bottom of pool; head of pool
Potential Reason (Feature) for Uppermost Fish	Field		If present and identifiable; eg – deformable obstacle/debris jam; dry channel; falls; other; etc
Vertical/Near-vertical Obstacle(s) present?	Field	Yes/No	
Lateral/Terminal Stream	Field		May vary based on uppermost fish location
EOF Riparian Stand Type (RB)	Field		Watershed Analysis methods
EOF Riparian Stand Type (LB)	Field		Watershed Analysis methods
Streamside Land Use Class at EOF	Field		Industrial timberland, USFS, small private timberland, conservation forest, agriculture, residential, other forestry, other non-forest
Notes	Field		Include potential explanatory features (CMZ, alluvial fan, debris flow, end of channel)
EOF Elevation_GIS	GIS	m	Lidar-based
EOF Drainage Area	GIS	km^2	
EOF Distance-From-Divide	GIS	m	
EOF Valley Aspect	GIS		Compass points [N, NE, E, SE, S, SW, W, NW]
EOF Valley Width	GIS	m	
EOF Valley Confinement	Calculated		Valley Width/Channel Width ratio
EOF Geologic Competence	GIS		Resistant or Erodible, based on classifications provided for Hard/Soft Rock Type N studies [Competent/Medium/Incompetent]

Attribute	Source	Units	Description (detail in Methods Manual)
Total Annual Precipitation for Current Hydrologic Year	nearby reference rain gauges	mm	from nearby reference rain gauges (see Table F-1)
Total Seasonal Precipitation for Survey Season	nearby reference rain gauges	mm	from nearby reference rain gauges
% of AnnualNormal Precipitation	Calculated	%	Total annual P for survey year/annual Normal
% of Seasonal Normal Precip	Calculated	%	Total seasonal P for survey season/seasonal Normal
Total Annual Streamflow for Current Hydrologic Year	nearby reference stream gauges	cms	from nearby reference stream gauges (see Table F-1)
Total Seasonal Streamflow for Survey Season	nearby reference stream gauges	cms	from nearby reference stream gauges (see Table F-1)
% of AnnualNormal Streamflow	Calculated	%	Total annual Q for survey year/annual Normal
% of Seasonal Normal Streamflow	Calculated	%	Total seasonal Q for survey season/seasonal Normal

2469

2470   **Table F-4. Habitat survey site field attributes**

Attribute	Source	Units	Description
SiteID	GIS		Identifier from DNR Hydro layer
SurveyID	Assigned		e.g., 2024-spring; 2025-fall, etc.; precise form of survey ID to be determined
Survey Date	Field		
Weather	Field		sunny, rainy, snowy, cloudy
Field Crew	Field		
Bottom of Survey (BOS) Latitude	Field, GPS	dd	WGS84
BOS Longitude	Field, GPS	dd	WGS84 (Negative dd for west)
BOS Elevation	Field, GPS	m	NAD83

Attribute	Source	Units	Description
Top of Survey (TOS) Latitude	Field, GPS	dd	WGS84
TOS Longitude	Field, GPS	dd	WGS84 (Negative dd for west)
TOS Elevation	Field, GPS	m	NAD83
Turnpoint Numbers and Locations	Assigned during survey		Turnpoints may be set on a Station, in which case the station can be identified as the location, or may be set outside of the channel thalweg, in which case the location relative to the previous turnpoint must be recorded.

2471

2472 **Table F-5. Habitat Survey Channel Survey Station Measured Attributes**

Attribute	Source	Units	Description
SiteID	GIS		Identifier from DNR Hydro layer
SurveyID			
Station Number	Assigned during survey		sequential numbering of survey stations from Bottom of Survey
Turnpoint Number	Assigned		Turnpoint ID (see Table F-4) from which station location is measured
Station Distance from Turnpoint	Measured	m	
Station Azimuth from Turnpoint	Measured	deg	
Station Elevation from Turnpoint	Measured	m	
Uppermost Fish Segment	Observation of Monument	LF	Observation of Uppermost Fish monument from Fish Survey occurs within measurement segment; not necessarily at the surveyed station if LF is monumented within a homogeneous segment
Water Depth	Measured	m	Instantaneous depth at station along thalweg (not BFD)
Channel Width	Measured	m	At bankfull elevation
Wetted Width	Measured	m	Water's edge
Flow Status	Observation		Dry, Flowing
Dominant Substrate	Ocular estimate	Categ.	Categorical (e.g. sand, gravel, cobble, boulder, bedrock, silt/clay/fines, wood)
Habitat Unit Type	Ocular estimate	Categ.	Pool, Riffle, Step, Step-Pool, Obscured
Station Point Type	Ocular estimate	Categ.	e.g. crest of tailout; bottom of pool; head of pool (may be blank)

Attribute	Source	Units	Description
Obstacle Type	Ocular estimate	Categ.	Vertical/Non-Vertical
Step Forming Medium	Ocular estimate	Categ.	Categorical (e.g. wood (log, debris, roots), hardpan, boulder, bedrock)
Tributary Junction	Observation	1	Flag if present; place station at point
Vertical Step Height	Measured	m	Continuous variable with 0 as an allowable value

2473

2474 **Table F-6. Stream habitat survey segment calculated attributes**

Attribute	Source	Units	Description
SiteID			
SurveyID			
Station #			
Segment Length [m]	Calculated	m	Calculated distance from Station n-1 to Station n; segment data relate to the segment below the station (i.e., "stations" are the upstream point of the segment)
Distance from Bottom of Survey			Running total of segment lengths from BOS (BOS = Station 0)
Above, at, or Below Uppermost Fish Segment	Calculated	US/DS/LF	Calculated based on location of LF segment from Table F-5; required for calculation of other attributes
Fish Presence	Calculated	FISH/NO-FISH	Assigned to segments based on location relative to LF point; needed for random forest models
Bankfull Width 10 (=bfw10)	Calculated	m	Average of bankfull widths from 4 stations downstream, current station, and 5 stations upstream, in approximate conformance with Forest Practices rule
Average BFW for 10 * bfw10 upstream	Calculated	m	Average of bankfull widths for a distance of 10*bfw10 upstream Required to test for FPB criteria
Average BFW for 20 * bfw10 upstream	Calculated	m	Average of bankfull widths for a distance of 20*bfw10 upstream Required to test for FPB criteria
Average BFW for 10 * bfw10 downstream	Calculated	m	Average of bankfull widths for a distance of 10*bfw10 downstream Required to test for FPB criteria
Segment Thalweg Bed Rise (Vertical Distance)	Calculated	m	Vertical Distance from Beg to End of Segment; calculated as change in elevation from station n-1 to station n
Thalweg Bed Gradient	Calculated	%	Segment Thalweg Bed Elevation Change/Segment Length

Attribute	Source	Units	Description
Effective Elev	Calculated	m	Calculated for pools based on pool tailout elevation; that (residual pool) elevation is translated to the segment upstream of the pool to determine the "effective" bottom elevation of the next (n+1) stream segment, for the purpose of calculating "effective, fish-eye" gradient of the n+1 segment
Effective Segment Rise		m	elevation of segment end minus the Effective Elevation, if there is one; otherwise, equals segment thalweg bed rise
Effective Segment Gradient		%	Effective Segment Rise/Segment Length
Effective Gradient Change From Downstrm Segment			Effective Gradient change from n-1 to n
Effective Gradient Change To Upstrm Segment			Effective Gradient difference from n to n+1
Maximum Effective Gradient Downstream from EOF	Calculated	%	Calculated from segment data using effective gradients
Length of Max Dnstrm Gradient Feature	Calculated	m	Calculated from segment data using effective gradients
Max sustained5 gradient downstrm	Calculated		Max of the running Minimum gradient feature over 5 cw; using effective gradients
Sustained Gradient Downstream	Calculated	%	Minimum gradient feature over 20 cw downstream of station n (including segment n); using effective gradients
Maximum Gradient Upstream of EOF	Calculated	%	Calculated from segment data; using effective gradients
Length of Max upstrm Gradient	Calculated	m	Calculated from segment data
Max sustained5 gradient upstrm	Calculated		Max of the running Minimum gradient feature over 5 cw; using effective gradients
Sustained upstream gradient	Calculated	%	Minimum gradient feature over 20 cw upstream of station n; using effective gradients
Delta Sustained Gradient upstrm	Calculated	%	Sustained upstream gradient – Sustained downstream gradient
Maximum Step Height Upstream	Calculated	bfw10s	

Attribute	Source	Units	Description
Maximum Step Height Downstream	Calculated	bfw10s	
Pool Frequency Upstream of Segment	Calculated	pool count/ bfw10	Calculated over 20*bfw10 upstream of current station
Pool Spacing Upstream of Segment	Calculated	m	Calculated over 20*bfw10 upstream of current station
Pool Frequency Downstream of Segment	Calculated	pool count/ bfw10	Calculated over 20*bfw10 downstream of current station
Pool Spacing Downstream of Segment	Calculated	m	Calculated over 20*bfw10 downstream of current station

2475

2476   **Table F-7. Habitat survey attributes calculated for stream at each survey**

Attribute	Source	Units	Description
SiteID	GIS		Identifier from DNR Hydro layer
SurveyID			
LF Distance from BOS	Calculated	m	
LF Elevation_GIS	GIS	m	Lidar-based
LF Drainage Area	GIS	km^2	
LF Distance-From-Divide	GIS	m	
LF Valley Aspect	GIS		Compass points [N, NE, E, SE, S, SW, W, NW]
LF Valley Width	GIS	m	
LF Valley Confinement	Calculated		Valley Width/Channel Width ratio
LF Geologic Competence	GIS		Resistant or Erodible, based on classifications provided for Hard/Soft Rock Type N studies [Competent/Medium/Incompetent]
Total Annual Precipitation for Current Hydrologic Year	nearby reference rain gauges	mm	from nearby reference rain gauges (see Table F-1)
Total Seasonal Precipitation for Survey Season	nearby reference rain gauges	mm	from nearby reference rain gauges
% of AnnualNormal Precipitation	Calculated	%	Total annual P for survey year/annual Normal

Attribute	Source	Units	Description
% of Seasonal Normal Precip	Calculated	%	Total seasonal P for survey season/seasonal Normal
Total Annual Streamflow for Current Hydrologic Year	nearby reference stream gauges	cms	from nearby reference stream gauges (see Table F-1)
Total Seasonal Streamflow for Survey Season	nearby reference stream gauges	cms	from nearby reference stream gauges (see Table F-1)
% of AnnualNormal Streamflow	Calculated	%	Total annual Q for survey year/annual Normal
% of Seasonal Normal Streamflow	Calculated	%	Total seasonal Q for survey season/seasonal Normal
Habitat Unit Upstream of LF	Calculated		
Effective Gradient of Segment Upstream of LF	Calculated	%	
BFW of segment Upstream of LF	Calculated	m	
Delta Sustained Gradient upstrm of LF	Calculated	%	Sustained upstream gradient – Sustained downstream gradient
Maximum Gradient Downstream from LF	Calculated	%	Calculated from segment data
Length of Max Dnstrm Gradient Feature	Calculated	M	Calculated from segment data
Maximum Sustained Gradient Downstream from LF	Calculated	%	Defined based on 20 bfw (multiple versions)
Length of Max Sustained Dnstrm Gradient Feature	Calculated	Multiples of bfw (m)	Calculated from segment data
Max Gradient Change Downstream of LF	Calculated	%	Calculated from segment data
Maximum Gradient Upstream of LF	Calculated	%	Calculated from segment data
Length of Max upstrm Gradient	Calculated	m	Calculated from segment data

Attribute	Source	Units	Description
Max sustained upstream gradient	Calculated	%	Sustained for minimum of 20*bfw10 to be in line with PHB proposals
Length of Max sustained upstream gradient	Calculated	m, bfw10	Length of the above in meters and also in multiples of bfw10
Max Sustained Gradient Change upstrm of LF	Calculated	%	Calculated from segment data; each gradient sustained for 20* bfw10
Maximum Step Height Upstream of LF	Calculated	bfw10s	
Maximum Step Height Downstream of LF	Calculated	bfw10s	
Pool Frequency Upstream of Segment	Calculated	count/ bfw10	Calculated over 20*bfw10 upstream of current station
Pool Spacing Upstream of Segment	Calculated	m	Calculated over 20*bfw10 upstream of current station
Pool Frequency Downstream of Segment	Calculated	pool count/ bfw10	Calculated over 20*bfw10 downstream of current station
Pool Spacing Downstream of Segment	Calculated	m	Calculated over 20*bfw10 downstream of current station

2477

2478 Appendix G. Glossary

2479 Concurred F/N Breaks: Supported by approved Water Type Modification Form

2480 Cumulative Metrics (defined in the data tables): Those metrics averaged or calculated over
2481 greater than one measurement

2482 Default Physical Criteria (DPC): Ranges of values for physical stream attributes presumed to
2483 represent fish use in the absence of protocol surveys

2484 Distance-From-Divide: The distance from the watershed divide downstream along the flow
2485 path to the point of interest on the stream. Where there are tributaries upstream of the point
2486 of interest, the distance-from-divide is through the longest channel path.

2487 Lateral (end of fish/end of habitat points): Sites where a stream without fish intersects a fish-
2488 bearing stream reach with fish both upstream and downstream of the junction with the fishless
2489 stream (Fransen et al 2006)

2490 Legacy Water Type (from DNR Hydrolayer but not based on the model): See data dictionary
2491 (https://www.dnr.wa.gov/publications/fp_fpamt_wt_defn_viewingguide.pdf)

2492 Region: East vs. west of the Cascade crest

2493 Terminal (end of fish/end of habitat points): Sites where fish occurrence terminates within a
2494 continuous reach of stream or at the junction of two or more fishless streams (Fransen et al
2495 2006)

2496

2497

2498

2499 **EndDocument**