Estimating observer error and steelhead redd abundance using a modified Gaussian area-under-the-curve framework

Journal:	*Canadian Journal of Fisheries and Aquatic Sciences*
Manuscript ID	cjfas-2017-0335.R2
Manuscript Type:	Article
Date Submitted by the Author:	21-Feb-2018
Complete List of Authors:	Murdoch, Andrew; Washington Department of Fish and Wildlife Herring, Chad; Washington Department of Fish and Wildlife Frady, Charles; Washington Department of Fish and Wildlife See, Kevin ; Quantitative Consultants Inc. Jordan, Christopher; National Marine Fisheries Service - NOAA,
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	Steelhead, Redd estimation, Observer error

SCHOLARONE™
Manuscripts

Estimating observer error and steelhead redd abundance using a modified Gaussian area-under-the-curve framework

Andrew R. Murdoch, Chad J. Herring, Charles H. Frady, Kevin See and

Chris E. Jordan

Andrew R. Murdoch[1] (Andrew.Murdoch@dfw.wa.gov), Chad J. Herring

(Chad.Herring@dfw.wa.gov), Charles H. Frady (Charles.Frady@dfw.wa.gov). Washington

Department of Fish and Wildlife 600 Capitol Way North, Olympia, Washington 98501, USA

Kevin See (Kevin@qcinc.org). Quantitative Consultants Inc. 2725 Montlake Blvd. East, Seattle,

Washington 98112, USA

Chris E. Jordan (Chris.Jordan@noaa.gov). Northwest Fisheries Science Center, National Marine

Fisheries Service, National Oceanographic and Atmospheric Administration,2725 Montlake

Blvd. East., Seattle, Washington 98112, USA

[1] Corresponding author: Andrew.Murdoch@dfw.wa.gov 600 Capitol Way North, Olympia,

Washington 98501, USA. Phone (509) 664-3148 x278, FAX (509) 662-6606.

1

1   *Abstract.* – This study examined how a suite of habitat and environmental variables relate to

2   the ability of a stream surveyor to identify (observer efficiency) and distinguish (observer

3   accuracy) steelhead redds from other stream features. Two existing spawning survey protocols

4   that included one or two redd observers were used to develop models to estimate redd observer

5   error. In most cases, steelhead redd abundances using raw redd counts were underestimated.

6   Mean annual rates of observer efficiency ranged from 0.44 to 0.57 and observer accuracy ranged

7   from 0.67 to 0.83.   Regardless of the observer error model used, adjusted annual redd abundance

8   estimates were generally unbiased (range -1.6 – 0.6 redds). A Gaussian area-under-the-curve

9   methodology that incorporates redd count data and observer error rates was used to generate

10   unbiased estimates of steelhead redd abundance in the Wenatchee (170 redds, CV = 44%) and

11   Methow (106 redds, CV = 41%) rivers. Unbiased estimates of redd abundance will help inform

12   new population viability analyses to better prioritize those populations with the greatest

13   conservation need.

14   Estimates of redd abundance are a common metric used to monitor the status and trend of

15   adult salmonid populations (Dauble and Watson 1997; Jacobs et al. 2009; Howell and Sankovich

16   2012). When used in conjunction with other population specific data (e.g., spawners per redd)

17   redd abundance can be used to estimate spawner abundance (Gallagher and Gallagher 2005;

18   Murdoch et al. 2010). Raw counts of redd abundance are inherently inaccurate due to a

19   combination of observer experience, habitat, and environmental conditions that influence a redd

20   observer's ability to accurately detect and identify redds (Dunham et al. 2001; Muhlfeld et al.

21   2006; Howell and Sankovich 2012). Nevertheless, redd surveys are conducted in a wide range of

22   habitats for many salmonid species because redd counts are reported to be correlated with

23   spawner abundance (Dunham et al. 2001; Gallagher and Gallagher 2005; Gallagher et al. 2010;

24   Murdoch et al. 2010) and can be conducted at a relatively low cost (Chasco et al. 2014).

25   Historically, redd-based estimates of spawner abundance have adequately served as an index of

26   spawner abundance for most management objectives, but growing conservation concerns (e.g.,

27   populations listed under the United States Endangered Species Act) now require unbiased

28   estimates of spawner abundance with known levels of uncertainty (Crawford and Rumsey 2011).

29   Redd survey design and methodology are important considerations that have changed

30   through time in response to management questions. For some populations, aerial surveys were

31   abandoned for more accurate ground based surveys. Single pass peak redd counts were

32   eventually replaced with multiple pass index area redd counts which eventually lead to weekly

33   redd counts of all available spawning habitat or census redds counts. At spatial scales larger than

34   the stream or population level, more complex survey designs (e.g., generalized random

35   tessellation stratified) have also been successfully used (Stevens 2002; Boydstun and McDonald

36   2005). Chasco et al. (2014) retrospectively evaluated the accuracy and precision of three

3

37 common Chinook Salmon *Oncorhynchus tshawytscha* redd survey designs (i.e., single pass

38 aerial, multiple pass index area, and multiple pass census) in Johnson Creek, ID using a state-

39 space model. Spawner abundance estimates based on multiple pass census redd counts were

40 shown to be comparable to mark-recapture based estimates in terms of both accuracy and

41 precision, but implemented at a fraction of the cost.

42   Selecting the correct redd survey design minimizes sampling error (Liermann et al. 2014),

43 but estimating and adjusting for observation error in redd counts has only been recently

44 investigated. The accuracy of redds counts may be influenced by errors of omission (i.e., missed

45 redds) or commission (i.e., false identification). While these two sources of error can cancel each

46 other, if the two error rates are not identical the redd count will be biased in the direction of the

47 largest source of error. Muhlfeld et al. (2006) reported the Bull Trout *Salvelinus confluentus* redd

48 abundance was highly correlated with errors of omission, but unrelated to errors of commission

49 suggesting the processes behind the two sources of error are unrelated and unique. Surveyor

50 experience was found to be positively related to observer efficiency in all studies (Dunham et al.

51 2001; Muhlfeld et al. 2006; Howell and Sankovich 2012) except one (Dunham et al. 2001). In

52 that study, 82% of the redd observers had little to no experience conducting Bull Trout surveys.

53 Hence, large variation in redd observer efficiency and the limited range of surveyor experience

54 may explain why no relationship was detected. Dunham et al. (2001) did report that several redd

55 and habitat characteristics were found to be significantly correlated with error rates of both

56 omission and commission. Gallagher and Gallagher (2005) estimated mean redd observer

57 efficiency rates for steelhead *O. mykiss* in coastal Northern California streams based on the

58 proportion of flagged redds observed. Mean redd observer efficiency was negatively related to

59 weekly stream discharge and positively related to maximum visible water depth, but errors of

4

60   commission (false identification) or experience of the redd observers were not assessed. A

61   comprehensive evaluation of factors influencing redd observer error has not been conducted for

62   any anadromous salmonid species. Understanding the factors driving observer error can be used

63   to minimize (e.g., number or experience of observers) or account for the error in the estimate

64   (e.g., area-under-the-curve (AUC)). These relationships between observer experience, habitat

65   characteristics, environmental conditions, and observer error rates may be useful in developing

66   models to predict observer error for each river or river segment of interest.

67   Estimates of observer error coupled with redd life and count data are also required to use the

68   AUC methodology to estimate total redd abundance (Hilborn et al. 1999; Parken et al. 2003).

69   The AUC methodology or more accurately, the trapezoidal AUC (TAUC), involves plotting a

70   times series of count data (e.g., fish or redds) or an escapement curve and dividing the area under

71   the curve by the redd life or residence time and observer efficiency (Parken et al 2003, Millar et

72   al. 2012). The TAUC approach has been successfully adopted to various conditions, but

73   incorporating the uncertainty of each parameter into the estimate has been challenging. Parken

74   et al. (2003) developed a method for incorporating uncertainty into the escapement estimate

75   using costly replicate surveys. Conversely, Hilborn et al. (1999) developed a maximum

76   likelihood (ML) estimator with an estimate of standard error, but was computationally difficult to

77   generate and has not been widely adopted. In addition, the TAUC estimator generally

78   outperformed ML estimator when compared to weir counts (Hilborn et al. 1999). Millar et al.

79   (2012) recently developed simpler analytical methods to estimate escapement and variance using

80   a Gaussian AUC (GAUC) methodology. While the traditional TAUC estimator performed

81   slightly better than the GAUC estimator, the later has an estimate of uncertainty that has been

82   lacking for decades. The GAUC methodology was adopted to fit with existing steelhead redd

5

83   survey protocols and redd count data sets, but estimates of observer efficiency were still needed.

84   The objectives of this study were to 1) measure redd observer error under an existing steelhead

85   redd survey protocol using two different levels of effort (i.e., one observer and two observers), 2)

86   develop predictive models to estimate observer error, 3) use those predictions of observer error

87   to subsequently estimate the total number of steelhead redds using a GAUC methodology and 4)

88   compare results of both survey protocols to inform the future direction of steelhead redd surveys.

89

90   **Methods**

91

92   *Study area.* – The Wenatchee and Methow rivers are located in north central Washington

93   State and drain 3,439 km^2 and 4,727 km2 of the Cascade Mountains eastern slope, respectively

94   (Figure 1). The Wenatchee watershed contains 332 km of major streams with 207 km accessible

95   to anadromous salmonids (Mullan et al. 1992). The Wenatchee River flows into the Columbia

96   River at rkm 754 and is located between Rock Island and Rocky Reach dams. The Methow

97   watershed contains 270 km of major streams with 239 km accessible to anadromous salmonids

98   (Mullan et al. 1992). The Methow River flows into the Columbia River at 843 rkm and is located

99   between Wells Dam and the confluence of Okanogan River. Steelhead in both rivers are

100   classified as summer-run and listed as threatened under the Endangered Species Act (NMFS

101   2016) as part of the Upper Columbia Distinct Population Segment (DPS). Summer-run steelhead

102   enter freshwater the year prior to spawning, typically June through October, become sexually

103   mature in their natal watershed and spawn the following spring between March and May.

104   Depending on the annual snowpack levels, stream discharge in the Wenatchee and Methow

105   rivers during the spawning period may increase over 200% and 600%, respectively.

6

106   *Study reach selection.* – Comprehensive steelhead redd surveys have been conducted in all

107   major spawning areas of both watersheds since 2004 (Hillman et al. 2011; Snow et al. 2011).

108   Steelhead spawn throughout the entire length of the Wenatchee and Methow rivers and in at least

109   some reaches of all the major tributaries. Given the diverse habitat and environmental conditions,

110   study reaches were selected to capture the maximum variability in habitat and environmental

111   conditions surveyors experience while conducting redd surveys. Four study reaches were

112   selected in each watershed that represented the range of survey conditions (Table 1). Korman et

113   al. (2002) reported that the variance of observer efficiency estimates (CV range 27 -100%)

114   increased as the proportion of steelhead observed (range 0.05 – 0.61) decreased. Hence, study

115   reaches were also selected to minimize the influence of the interannual variability in redd

116   abundance and the subsequent variance in observer error rates in order to facilitate model

117   development.

118   *Redd surveys.* – Redd observer error was assessed by comparing intensive census ground

119   redd counts to naïve observer redd counts similar to that described in Thurow and McGrath

120   (2010). All redd observers in the study received a similar, albeit limited, level of training that

121   included both a classroom and field portion. Census surveys were conducted every three days in

122   all study reaches by experienced surveyors (mean = 6 years; range 4 – 10) and began prior to the

123   start of spawning and ended when no new redds were found. Census surveyors did not change

124   reaches throughout the spawning period. It was assumed that redds found during census surveys

125   were an accurate unbiased count of the total number of redds in each reach. Features that could

126   be misinterpreted as steelhead redds (e.g., old salmon redds or hydrologic features) or errors of

127   commission were marked, including a unique geo-referenced waypoint and numbered

128   sequentially on aerial photographs. Aerial photographs were laminated for use in the field. The

7

129   average reach length per photograph was 445 m. Areas of localized high density spawning were

130   also annotated on detailed hand drawn maps. If redd superimposition was observed, the

131   proportion of the redd disturbed by another female steelhead was estimated during each survey.

132   During census surveys, the date when redds became no longer visible due to superimposition or

133   environmental conditions was recorded. Flagging was not attached to vegetation adjacent to

134   steelhead redds within study reaches in order to reduce potential bias during subsequent naïve

135   surveys. Surveys were conducted using one or two person catarafts depending on the size of

136   stream. When a two person cataraft was used the redd observer provided verbal direction to the

137   rower at all times. Surveys were generally conducted during periods of maximum potential

138   sunlight (1000 – 1500 h). All surveyors wore polarized sunglasses and a brimmed cap. If actively

139   spawning fish were observed, surveyors used caution to avoid disturbing fish.

140   Surveys were conducted near the peak of spawning by naïve observers to estimate redd

141   observer accuracy and efficiency. Consistent with the methodology typically used in each

142   respective watershed, a single redd observer was used in the Wenatchee watershed and two redd

143   observers in two catarafts were used in the Methow watershed. Naïve surveys were generally

144   conducted on the same day as a census survey, at a typical level of effort. However, because

145   variation in survey effort does exist among surveyors, the start and stop time of each survey was

146   recorded to assess the influence of effort on observer error. Naïve observers had no prior

147   knowledge of the number of redds present and used methods identical to those for census

148   surveys. All observers, both census and naïve, were instructed not to discuss the results of any

149   survey during the study period. The variability in experience of the naïve observers was

150   maximized to the extent possible to better understand the influence of experience on observer

8

151 error. The number of replicate naïve surveys conducted on each reach was dependent on

152 personnel availability and environmental conditions.

153   *Redd observer error model covariates.* – Experience of the redd observer may be an

154 important factor influencing observer error (Muhlfeld et al. 2006; Howell and Sankovich 2012).

155 Observer experience was quantified based on the species and frequency of surveys. Because all

156 salmonids that spawn in streams exhibit similar behavior during spawning (i.e., construct a redd),

157 experience was pooled across species. Experience was quantified based on the number of years

158 or seasons an individual had conducted redd surveys. For example, an individual that conducted

159 Chinook Salmon redd surveys one day a week during the entire spawning season for five years

160 received a value of five. However, if the same observer also conducted steelhead redd surveys

161 one day a week throughout the spawning season for three years, the surveyor would receive a

162 value of eight for experience conducting salmonid spawning ground surveys. Partial years or

163 single surveys, regardless of classification, were prorated based on a maximum of ten surveys

164 per season (i.e., one survey = 0.1 points). Because we believed there was a saturating effect of

165 experience, the log of the experience value was used in error models. Survey effort of the naïve

166 surveyors was standardized by dividing survey time by reach length. Observed redd density was

167 calculated by dividing the total length of each study reach by the total number of redds or redd-

168 like features identified by naïve surveyors.

169   Habitat variables describing the physical characteristics of each reach were measured at the

170 beginning of the spawning season along ten equidistant transects established perpendicular to the

171 flow or was summed across the entire reach (Bartley and Rutherfurd 2005). Mean stream width

172 and water depth for each study reach were calculated from the width and depth recorded at each

173 transect. Horizontal water visibility was estimated for each survey using a standard secchi disc.

9

174   At the start and end points of the study reach, a secchi disc was placed in approximately 0.5 m of

175   water with a laminar flow pattern. A depth of 0.5 m was selected to provide observers the ability

176   to remain standing in the stream channel throughout the spawning season as spring snow melt

177   increases river discharge and water velocity. The observer would walk upstream to a point when

178   the white and black sections of the disc were no longer discernible. The distance between the

179   observer and secchi disc was measured on the stream bank using a tape measure. The mean

180   distance of the two measurements was used as an index of horizontal water visibility for that

181   survey. River discharge measured at the nearest gauging station was recorded at the beginning

182   and end of the survey. The mean discharge for that survey day was used in the model.

183   The influence of channel or habitat complexity was quantified using the coefficient of

184   variation (CV) in the thalweg profile (Madej 1999; Bartley and Rutherford 2005). Because the

185   influence of channel complexity associated with spawning sites was the metric of interest,

186   steelhead redd locations were used to select sample locations for the thalweg profile. The redd

187   location was used as the midpoint of the sample to ensure representation of channel complexity

188   at the spawning site. A longitudinal profile of water depth was measured every 3m along the

189   thalweg, extending 150m upstream and 150m downstream from each of three randomly selected

190   redds in each study reach. The thalweg CV was calculated for each 300m sample location. The

191   mean thalweg CV for each study reach was used as an index of channel complexity. Data on

192   possible covariates collected during census surveys are summarized in Table 1.

193   *Redd observer error rates.* – Individual redds identified during naïve surveys were

194   compared to redds identified in census surveys using georeferenced redd locations, aerial

195   photographs, and hand drawn maps. Redd observer error may occur when redds are falsely

196   identified (commission) or when redds are missed (omission) and were based only on redds that

10

197 were still visible during the naïve surveys (Thurow and McGrath 2010). Redd observer accuracy

198 was defined as the proportion of features identified by an observer that were steelhead redds and

199 not confused with other features created hydraulically (i.e., scour and deposition) or redds from

200 other species:

201

202 (1) $A_i = R_i/F_i$

203

204 where A is the accuracy of naïve redd survey i, R_i is the number of correctly identified steelhead

205 redds found by both the naïve and census surveyors during naïve redd survey i and F_i is the total

206 number of redd-like features found by the naïve redd observer i. Redd observer efficiency was

207 defined as the proportion of steelhead redds correctly identified during a survey:

208

209 (2) $E_i = R_i/V_{Cen}$

210

211 where E is the efficiency of naïve survey i, R_i is the number of correctly identified steelhead

212 redds found by both the naïve and census surveyors during naïve redd survey i, and V_{Cen} is the

213 number of redds found by the census surveyor that were still visible on the day of naïve survey i.

214 Hence, modeling the error structure of steelhead redd observers requires a two stage process that

215 includes; 1) estimating the proportion of features identified during a survey that were steelhead

216 redds and 2) estimating the proportion of steelhead redds that were correctly identified during the

217 survey. Alternatively, when total redd abundance, not specific redd locations or distribution, is

218 the primary metric of interest, the sum of both sources of error (i.e., net error rate) may be

219 modeled more simply in a single step. Net error rates were calculated by dividing the number of

11

220   redd-like features identified during naïve survey *i* by the number of redds found by the census

221   surveyor that were still visible on the day of naïve survey *i*:

222

223   (3) $NE_i = F_i/V_{Cen}$

224

225   where *NE* is the net error of naïve survey *i*. Net error rates may be less than or greater than one,

226   depending on which source of error (i.e., omission or commission) was the greatest. Correlations

227   among error rates (commission, omission, and net error) were examined using a Spearman

228   correlation coefficient.

229   *Model selection.* – Observer efficiency, observer accuracy and net error rates were modeled

230   separately. All covariates were centered to have a mean of zero. Separate models were also

231   developed based on the number of observers (i.e., Wenatchee = 1; Methow = 2). Observer

232   efficiency and accuracy error rates were modeled with a logit link using a binomial error

233   structure, while net error was modeled as a linear model with Gaussian error structure, with a list

234   of possible covariates including habitat metrics, redd density, survey effort, and observer

235   experience. The model selection process involved fitting a series of models with all possible

236   combinations of covariates, including an intercept-only model, using Akaike information criteria

237   corrected for small sample size (AIC_c) to identify the most parsimonious model (Burnham and

238   Anderson 2002). In addition, covariates with a correlation > 0.5 were not included in the same

239   candidate model. The AIC_c score of each model was recorded, and the model with the lowest

240   AIC_c score was considered the most parsimonious model supported by the data (Akaike 1974).

241   All models with a $\Delta AIC_c < 2$ were equally supported by the data (Burnham and Anderson

242   2002). Therefore, model averaging was conducted with all models with a $\Delta AIC_c < 2$ to make

12

243   predictions of error rates. All statistical analyses were conducted using R software (R Core Team

244   2015).

245   *Model performance.* – A leave-one-out cross validation (LOOCV) approach was used to

246   assess the predictive performance of the different models (Hastie et al. 2009). To perform a

247   LOOCV, we took all the models with a Δ AICc of less than 2, and re-fit them to the entire

248   dataset except for a single data point. We then predicted the error rates for that point using model

249   averaged coefficients, and then predicted the number of redds for that data point as well. We

250   cycled through the entire dataset, leaving each point out once. Comparisons (i.e., mean bias, root

251   mean squared error (RMSE), weighted correlation and coverage probability) between those

252   predictions and the observed values were made to assess model performance. Mean bias

253   describes the average difference between observed and predicted values and can be used to

254   assess whether a model's predictions are unbiased. The RMSE describes how far off in either

255   direction the predictions are from the truth, on average. For predicted redds, a normalized RMSE

256   (NRMSE) value divides the RMSE by the average number of observed redds, providing the

257   average proportion of redds that the prediction was off by. Weighted correlations describe the

258   correlation between observed and predicted values, weighted by the relative number of observed

259   redds so that predictions for surveys that encountered a large number of redds are more heavily

260   weighted than surveys that encountered few redds. Because surveys encountered many fewer

261   redds in the Methow than the Wenatchee, the relative number of observed redds was calculated

262   by dividing the number of observed redds by the maximum number of redds observed in that

263   watershed, so as to not overly discount the Methow surveys. Coverage probabilities assess the

264   rate at which the 95% confidence interval encompasses the true value. A good model should

13

265   have a coverage probability near 95%, whereas one with a low coverage probability is not

266   capturing enough of the uncertainty.

267   *Application of GAUC to steelhead redd survey data.* – Millar et al. (2012) presented a less

268   computational demanding analytical approach to estimating AUC abundance and variance. As

269   traditionally applied, AUC uses a time series plot of spawner count data to estimate spawner-

270   days, which is then divided by average spawner life and observer efficiency to estimate spawner

271   abundance. The key difference between spawner counts and redd surveys is that it is impossible

272   to determine which spawners were counted in the previous survey, whereas redds can be geo-

273   referenced. This eliminates the need to account for redd life because a time-series of previously

274   unseen redds or new redds can be used instead of total visible redds, using survey number

275   instead of date as the point of reference. However, the GAUC methodology still requires an

276   estimate of observer error, with standard errors. Based on the suite of covariates included in the

277   top models, estimates of redd observer error were generated for each survey, and then reach

278   averages and standard deviations were calculated to be used in the GAUC calculations.

279   In 2014, steelhead redd observers conducted weekly surveys of the major spawning reaches

280   in both the Wenatchee and Methow rivers. Selected model covariates were collected as

281   previously described. Because the steelhead spawning distribution is known to be clumped

282   (Hillman et al. 2011; Snow et al. 2011), presumably due to heterogeneous spawning habitat

283   quality, minor spawning reaches were only surveyed at peak spawning, thereby significantly

284   decreasing costs of conducting redd surveys. For these minor spawning reaches, the number of

285   visible redds at peak spawning was adjusted for the predicted observer error of the adjacent

286   major spawning reach to generate an estimate of total redd abundance. Estimates of redd

14

287   abundance and variance were summed across all major and minor spawning reaches to generate

288   a total redd estimate for each river.

289

290   **Results**

291   *Single redd observer error rates (Wenatchee Basin)*

292   During the study period, 52 naïve surveys were conducted in order to estimate observer

293   error (Table 2). The mean (standard deviation (SD)) number of naïve surveys conducted in a

294   single day was 2.4 (1.0). On average, naïve observers identified (i.e., efficiency) only 48% (SD =

295   23%) of the steelhead redds visible on the day of the survey indicating that a majority of redds

296   were missed. However, redd observers were more accurate (mean = 79%; SD = 17%) than

297   efficient, meaning that while observers missed a large proportion of visible redds, it was more

298   likely that those redds identified were true steelhead redds and not confused with redd-like

299   features. Accuracy was not significantly correlated with observer efficiency (r_s = 0.20, P = 0.16)

300   or net error rates (r_s = - 0.14, P = 0.33), but was higher than observer efficiency in every study

301   reach, resulting in intermediate rates of net error. However, because efficiency was generally

302   smaller than accuracy, the mean net error (63%; SD = 31%) was less than one. As expected,

303   observer efficiency was correlated with net error rate (r_s = 0.86, P < 0.001) and a majority of

304   naïve surveys (88%, N = 46) underestimated the redd abundance regardless of observer

305   efficiency.

306   When modeling observer efficiency, stream depth, observed redd density, and the mean

307   thalweg CV were all included in the top model (Table 3). The second-best model included the

308   same covariates except for observed redd density, and had a difference of over 3 AICc units.

309   After that, no other model was within 10 AICc units. The effects of these main covariates

                                                   15

310   suggest that efficiency improves when stream complexity or redd density increases, while it

311   declines when the stream depth increases (Table 4). For observer accuracy, the only covariate

312   included in the top model was standardized effort (Table 3). As expected, accuracy improved

313   when effort increased, suggesting when surveyors spend more time evaluating redd-like features

314   they are less likely to make errors of commission (Table 4). The mean CV of thalweg depth and

315   observed redd density were also relatively important, but models containing these covariates had

316   a Δ AICc greater than 2, suggesting most of the variation in observer accuracy is captured by

317   effort. The same three covariates (stream depth, the observed redd density, and the mean thalweg

318   CV) most important in explaining observer efficiency were also the most important for net error

319   (Table 3 and 4). Given the correlation between observer efficiency and net error rates, it was

320   expected that the same covariates would also be important in predicting net error rates.

321   Model averaged predictions of error rates had moderate weighted correlation coefficients

322   with observed rates in a leave-one-out cross validation analysis (Efficiency = 0.68, Accuracy =

323   0.48, Net Error = 0.59). The RMSE of the LOOCV exercise was 17% for observer efficiency,

324   15% for observer accuracy and 25% for net error.

325

326   *Two redd observer error rates (Methow watershed)*

327   During the study period, 43 naïve surveys were conducted in order to estimate redd observer

328   error (Table 2). The mean (SD) number of naïve surveys conducted in a single day was 1.7 (0.5).

329   On average, naïve observers identified (i.e., efficiency) only 51% (SD = 27%) of the steelhead

330   redds visible on the day of the survey or they found only slightly more than they missed. As was

331   the case under a single observer protocol, a majority of the steelhead redds found were not

332   misidentified and were true steelhead redds (mean = 77%, SD = 14%). Observer accuracy was

16

333   not significantly correlated with observer efficiency (r_s = -0.08, P = 0.62), but was modestly

334   correlated with net error (r_s = -0.39, P = 0.01). Similar to the single observer error rates in the

335   Wenatchee watershed, rates of omission were generally larger than rates of commission resulting

336   in a mean net error (68%, SD = 38%) less than one. Similarly, observer efficiency was also

337   correlated with net error rate (r_s = 0.92, P < 0.0001) and a majority of naïve observers (95%, N =

338   41) underestimated the redd abundance regardless of observer efficiency.

339   When modeling observer efficiency, redd density and the mean thalweg CV were the most

340   important covariates, being included in every model with Δ AICc < 2 (Table 3). Other covariates

341   that also contributed some information included experience, gradient, mean depth, discharge and

342   effort. Under a two-observer model, unlike the one-observer model, increases in thalweg CV had

343   a negative effect on efficiency, while increases in water depth had a positive effect (Table 4). As

344   expected, increase in surveyor experience had a positive influence on observer efficiency.

345   Observer accuracy was modeled best with gradient and mean thalweg CV, as well as

346   observer experience as all models with Δ AICc < 2 contained some combination of these three

347   covariates (Table 3). As gradient or mean thalweg CV increases, the predicted accuracy

348   improves, although it decreases slightly as experience increases (Table 4). Numerous other

349   covariates received some support, with Δ AICc in the 2 – 7 range. Although the RMSE was only

350   14%, the predictions were all nearly identical, regardless of which data point was left out. The

351   intercept-only model had a Δ AICc of 2.74, suggesting some support for just using the average

352   observer accuracy rate without any covariates.

353   The most important covariates for modeling net error were mean discharge, mean thalweg

354   CV, redd density, and salmonid redd survey experience. There was some support for a few other

355   covariates, but those four covariates dominated the Akaike weights (Table 3). These covariates

17

356   also had a similar direction of influence on net error as was observed in modeling observer

357   efficiency (Table 4).

358   In the leave-one-out cross validation, the weighted correlation coefficients were fairly high

359   for efficiency and net error, but low for accuracy (Efficiency = 0.78, Accuracy = 0.2, Net Error =

360   0.74). The RMSE of the LOOCV exercise was 18% for observer efficiency, 14% for observer

361   accuracy and 27% for net error.

362

363   *Comparison of redd observer error models using one or two observers*

364   Comparing model covariates and direction of influence on efficiency or accuracy error rates

365   may provide insights as to possible mechanisms or causation. Redd density was the only

366   covariate that was consistent across all efficiency and net error models (Table 4). Conversely,

367   surveyor experience was included in all of the two-observer error models and none of the single

368   observer models. Naïve surveyors used in the development of the Wenatchee or single observer

369   models had a significantly lower mean experience level (-32%) compared to the Methow or two-

370   observer model (t-test, $P < 0.01$), and they also included a smaller range of experience levels.

371   The index of channel complexity or thalweg CV also had inconsistent effects on model

372   performance, but was not significantly different between the two watersheds (t-test, $P = 0.27$).

373   A single redd observer floating downstream must choose where to look for redds within the

374   stream channel. Hence, as channel complexity increases and areas of gravel sorting (e.g., pool

375   tailouts, log jams, split channels) are more readily identified, a single observer may make better

376   decisions where to look for redds. Conversely, with two redd observers the decision of where to

377   look for redds is less important because each observer is only looking in half the channel.

378   However, as potential redd locations increase with greater channel complexity, the probability of

18

379 identifying redds decreases. Hence, the difference in the number of observers and how they react

380 to channel complexity during surveys best explains the different effects that channel complexity

381 had on observer efficiency and net error.

382   Gradient had a positive influence, but not significant, in both the efficiency and accuracy

383 models in the Methow watershed, but no models in the Wenatchee watershed. Because reaches

384 in the Wenatchee watershed had a significantly greater mean gradient than those in the Methow

385 (t-test, $P < 0.001$) inclusion in the Methow models was likely attributed to the physical

386 difference between the watersheds, not related to the number of observers, similar to the

387 experience covariate.

388   While effort (min/m) was also not significantly different between the two watersheds (t-test,

389 $P = 0.33$), it was only included in the accuracy model for the Wenatchee watershed and likely

390 related to differences in the number of observers. Two redd observers provide sufficient

391 coverage within the stream channel such that any additional effort does not result in greater

392 efficiency or accuracy. Water depth was not significantly different between the two watersheds

393 (t-test, $P = 0.87$), but had contrasting influence on observer efficiency, potentially due to

394 differences in how redd observers navigate rivers and examine spawning habitat when there are

395 one versus two boats.

396

397 *Model performance*

398   The models all made unbiased predictions of error rates, based on the LOOCV results

399 (Table 5). Weighted R^2 values were fairly high for the efficiency and net error models,

400 particularly the two-observer model in the Methow, but were low for both accuracy models.

401 Across all six models, the RMSE ranged from $0.14 - 0.27$. Based on the weighted correlations,

19

402   the Methow models using two redd observers outperformed the Wenatchee models using a single

403   redd observer for efficiency and net error, but both accuracy models had low weighted R^2 values.

404   When used to predict the true number of redds in a reach, the average bias of predicted

405   redds compared to observed in the LOOCV was generally less than one redd (Table 6). Other

406   metrics of performance were generally comparable between the efficiency/accuracy predictions

407   and the net error predictions, at least within a basin. The NRMSE indicates that the predicted

408   redds could be off by 30-39%, but the unbiasedness means they were equally likely to be an

409   overestimate as an underestimate. In contrast, assuming that the observed redds were a census

410   would lead to an underestimate of redds by 27 – 37%, a consistently negative bias.

411   All four models failed to capture all the appropriate uncertainty, as reflected in the 95%

412   coverage probabilities. That is a measure of how frequently the 95% confidence interval

413   contained the true value, and should be close to 95%. The Wenatchee models made predictions

414   that were far too precise, so even though the CV was low, the predictions were often inaccurate

415   with inappropriately small confidence intervals. The Methow models did better, with the net

416   error model having the highest coverage.

417

418   *Application of GAUC to steelhead redd survey data.*

419   The two-observer net error model was selected because our analysis indicted that it had a 1)

420   high weighted correlation between observed and predicted redds, 2) produced a reasonably

421   unbiased estimate of redds, 3) provided the most appropriate 95% coverage interval and 4) the

422   net error model involves only one model, and so is more parsimonious compared to using an

423   efficiency and accuracy model. In 2014, two observers conducted steelhead redd surveys in both

424   the Wenatchee and Methow rivers and likewise the two-observer net error model was used to

20

425   estimate observer error. Because there were no data points with zero redds in the original dataset,

426   we could not fit a two-stage or hurdle model to this data, which leaves us with no way to adjust

427   for observed zeros. Therefore, the estimates of total redds will potentially be conservative, but

428   hopefully with minimal bias. The mean net error in the Wenatchee (0.58) was larger than the

429   Methow (0.48) and the coefficients of variation for net error within a reach were generally larger

430   in the Wenatchee as well. Although there were some general differences in several covariates

431   between the two rivers (i.e., greater redd density in the Wenatchee River and more experienced

432   surveyors in the Methow River) the difference in net error estimates was mainly driven by larger

433   discharge in the Wenatchee River (mean = 109 m^3/s) compared to the Methow River (mean = 42

434   m^3/s).

435   GAUC curves were fit to time-series of counts of new redds for all major spawning reaches

436   (Figure 2). Surveys conducted in minor spawning areas were conducted near peak based on the

437   GAUC spawning curve of an adjacent major spawning reach. The observed number of redds in

438   the Methow (N = 41) and Wenatchee (N = 97) rivers were 39% and 57% of the predicted

439   number of redds, respectively (Methow = 106, Wenatchee = 170). The estimated number of

440   steelhead redds in the Methow and Wenatchee had similar CVs of 41% and 44%, respectively.

441

442   **Discussion**

443   Using a two-observer net error model under a modified GAUC framework, steelhead redd

444   surveys conducted on a typical weekly schedule underestimated redd abundance by 61% and

445   43% in the Methow and Wenatchee rivers, respectively. Given the magnitude of negative bias

446   associated with steelhead redd counts, the status of populations based solely on raw redd count

447   data should be reevaluated. The complex error structure associated with redd surveys (i.e.,

448 omission and commissions) did result in a small number of cases (12%) that resulted in

449 overestimates in redd abundance. It is also noteworthy that error rates as defined in this study

450 were based on redd counts from census surveys (i.e., every three days by experienced surveyors)

451 assuming no error. The findings of the study demonstrate redd observer error is present and

452 prevalent. Therefore, it must be recognized that both census and naïve surveyors could have

453 missed redds (i.e., omission) and the false identification of redds or errors of commission by

454 census surveyors was also possible as none of the redds found during census surveys were

455 physically disturbed to confirm the presence of eggs. It was precisely because of these potential

456 sources of process error that census surveys were conducted every three days by highly

457 experienced surveyors. Hence, process error was not eliminated, but rather minimized such that

458 the results of the study remain valid.

459   High variability in redd observer efficiency was attributed to multiple factors including river

460 characteristics, habitat complexity, redd density and experience conducting salmonid redd

461 surveys. Predicting observer accuracy rates was generally more difficult and less precise than

462 predicting observer efficiency. An intercept-only model was included in the model selection

463 process, and although it had a Δ AICc near 40 for the Wenatchee data, the Δ AICc for the

464 Methow data was only 2.74, suggesting the covariates that were included may not provide much

465 information about observer accuracy. Variability in the accuracy of a redd observer was mostly

466 attributed to surveyor effort or channel gradient and habitat complexity. It is interesting that

467 greater effort by a single redd observer improves their ability to correctly identify redds (i.e.,

468 accuracy), but not find more redds (i.e., efficiency). Greater surveyor effort may translate into

469 how intensively each redd like feature was examined rather than how extensively the surveyor

470 examined possible spawning habitat. Increases in gradient and channel complexity may simply

22

471   reduce the probably of a redd-like feature being present thereby reducing the probability of

472   making errors of commission. There also may be human behavioral traits (e.g., dependability,

473   independence or routine-oriented) that would assist in predicating redd observer accuracy, but

474   none were included in this study.

475   The top net error models included similar covariates as observer efficiency for each

476   respective redd survey protocol. Both observed redd density and thalweg CV were included

477   regardless of survey protocol. Additionally, a measure of stream condition on the day of the

478   survey was also included (i.e., mean water depth for a single redd observer and mean discharge

479   for two redd observers). As expected, the efficiency and net error of a single redd observer

480   decreased with an increase in water depth presumably due to a greater difficulty in observing

481   contrast in the substrate within the redd. Conversely, greater water depth increased the

482   efficiency, although not significantly, when two observers were used. Because steelhead redds

483   are generally associated with edge habitat, greater water depth may provide two observers

484   greater access to edge habitat via rafts thereby allowing better identification of redds. The

485   positive effect of increasing discharge, albeit small, in the net error models further supports the

486   hypothesis that greater access to edge habitat by each of the two observers improves their ability

487   to identify redds.

488   In general terms, a single steelhead redd observer's ability to correctly identify redds is a

489   function of their ability to observer the substrate (i.e., shallow water depth), identify probable

490   redd locations (i.e., greater channel complexity or thalweg CV) and to a lesser degree other redds

491   (i.e., greater redd density). Conversely, two redd observers identify more redds correctly as redd

492   density increases, habitat is more readily observed (i.e., decreases in channel complexity or

493   thalweg CV), as experience increases and to a lesser extent as discharge increases.

23

494   Surveyor experience was included in the top net error models under the two redd observer

495   protocol. For reasons previously discussed, the top net error models under two redd observer

496   protocol was applied to redd data in 2014. The covariates included in the top models included

497   surveyor experience, an index of stream condition on the day surveys were conducted (i.e.,

498   discharge), an index of habitat complexity of each reach (i.e., thalweg CV), and an index of

499   spawning habitat quality (i.e., observed redd density). These covariates can be easily measured

500   before, during, and after the spawning season thereby increasing the application of the model to

501   most river systems.

502   Between the two basins and methodologies, parameter estimates sometimes showed

503   discordant results. Determining whether this is due to difference between the basins, the

504   methodologies, or other unknown factors is difficult to do with certainty. For example, surveyor

505   experience was not considered an important covariate in any single observer models, but was

506   included in the top two observer models. This phenomenon is counter intuitive and may be an

507   artifact of the experience of the naïve observer used in each watershed. As shown previously, the

508   mean experience level of naïve redd observers in the Wenatchee (27.91) was lower than the

509   Methow (41.48). Hence, if surveyors with a greater range of experience were used in the

510   Wenatchee watershed, we would expect experience to be an important covariate. Ideally, a

511   comparison of protocols would have observers with a similar range of experience. Once a two-

512   observer protocol has been adopted in the Wenatchee Basin, we recommend gathering additional

513   data points in that basin to include in and update the observer error models. This may lead to a

514   larger difference in the performance of the efficiency/accuracy models compared to the net error

515   model, making it more straightforward to choose between the two approaches.

516   Currently, the use of either approach could be justified and would be an improvement over

517   raw redd counts, since the predicted number of redds from either protocol was only slightly

518   biased compared to census counts (Table 6). Although the coverage probability was less than

519   desired, the resulting estimates are certainly more accurate than raw redd counts. Another benefit

520   of using the GAUC methodology was quantifying the uncertainty associated with redds surveys.

521   Millar and Jordan (2013) extended the GAUC methodology to incorporate more complex curves,

522   but the underlying need for an estimate of observer error remains. Unbiased precise estimates of

523   spawner escapement are recommended for ESA listed salmon and steelhead populations

524   (Crawford and Rumsey 2011). While more work is needed to better understand the relationship

525   of steelhead redds and spawner abundance, methods of estimating fish per redd values have been

526   suggested (See Murdoch et al. 2010).

527   This study was conducted in the Wenatchee and Methow rivers, in part, due to a ten-year

528   history of successful implementation of steelhead redd surveys. Successful steelhead redd

529   surveys are defined as being able to conduct weekly surveys throughout the spawning period

530   (10-12 weeks) by raft and observe all potential steelhead spawning substrate. While spring

531   freshets are common, they are short in duration and river conditions are acceptable again in only

532   a few days. If the magnitude and variability of steelhead redd observer error reported in this

533   study is representative of other populations, the status of steelhead populations based on redd

534   counts are negatively biased. Trends in population status could also be biased if spawner

535   distribution (i.e., thalweg CV) or spawn timing (i.e., water depth or discharge) has changed or

536   the abundance is highly variable (i.e., redd density).

537   Estimates of adult spawners is the most important metric in assessing extinction risk

538   (Crawford and Rumsey 2011). The results of this study suggest that steelhead redd surveys

25

539   conducted with two experienced observers produced the least biased redd estimates. Application

540   of this methodology and models can assist managers in generating unbiased and precise

541   estimates of redd abundance to better understand the risk to populations with high conservation

542   value. Training of redd surveyors can include suggestions of how stream flow, complexity and

543   redd density may influence successful surveys. Managers can also take steps to reduce bias in

544   steelhead redds surveys by using two experienced observers and establish survey reaches that

545   promote diligence or thoroughness of the observers and not speed or constrained by time (i.e.,

546   use as much effort as needed).

547   In addition, application of the GAUC methodology as described with parameters derived

548   from typical steelhead redd survey data can be generated for little additional cost. A comparison

549   of redd based spawner abundance estimates (i.e., number of redds multiplied by fish per redd

550   value) to other independent methods of estimation must be performed to better understand the

551   usefulness of these data. However, unlike other potential methods of estimation, redd surveys

552   also provide critical data on spawn timing and location that are of value to managers for either

553   restoration or protection purposes.

554

555   **Acknowledgements**

566   **References**

567   Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on

568   Automatic Control 19(6):716–723.

569

570   Bartley, R. and I. Rutherfurd. 2005. Measuring the reach scale geomorphic diversity of streams:

571   application to a stream disturbed by a sediment slug. River Research and Applications

572   21:39-59.

573

574   Boydstun, L. B., and T. McDonald. 2005. Action plan for monitoring California's coastal

575   salmonids. Final report to NOAA Fisheries, Contract Number WASC-3-1295, Santa

576   Cruz, California.

577

578   Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference: a

579   practical information-theoretic approach. Springer-Verlag, New York.

580

581   Chasco, B.E., E.J. Ward, J.A. Hesse, C. Rabe, R. Kinzer, J.L. Vogel, and R. Orme. 2014.

582   Evaluating the accuracy and precision or multiple abundance estimators using state-space

583   models: a case study for a threatened population of Chinook Salmon in Johnson Creek,

584   Idaho. North American Journal of Fisheries Management 34:5, 945-954.

27

585

586   Crawford B.A. and S. M. Rumsey. 2011. Guidance for the monitoring recovery of Pacific

587   Northwest salmon & steelhead listed under the federal endangered species act. National

588   Marine Fisheries Service, NW Region, Portland, Oregon.

589

590   Dauble, D.D., and D.G. Watson. 1997. Status of fall Chinook salmon populations in the mid-

591   Columbia River, 1948-1992. North American Journal of Fisheries Management 17:283-

592   300.

593

594   Dunham, J. B. Rieman, and K. Davis. 2001. Sources and magnitude of sampling error in redd

595   counts for bull trout. North American Journal of Fisheries Management 21:343-352.

596

597   Gallagher, S.P., P.A. Adams, D.W. Wright, and B.W. Collins. 2010. Performance of spawner

598   survey techniques at low abundance levels. North American Journal of Fisheries

599   Management 30:1086-1097.

600

601   Gallagher, S.P., and C.M. Gallagher. 2005. Discrimination of Chinook salmon, coho salmon, and

602   steelhead redds and evaluation of the use of redd data for estimating escapement in

603   several unregulated streams in northern California. North American Journal of Fisheries

604   Management 25:284-300.

605

606   Gallagher, S.P., P.K.J. Hahn, and D. H. Johnson. 2007. Redd counts. Pages 197 - 234 *in* Johnson,

607   D.H., B.M. Shier, J.S. O'Neal, J.A. Knutzen, X. Augerot, T.A. O'Neil, T.N. Pearsons

28

608   plus 37 contributing authors. Salmonid field protocols handbook: techniques for

609   assessing status and trends in salmon and trout populations. American Fisheries

610   Society, Bethesda, Maryland.

611

612   Hastie T., R. Tibshirani, and J. Friedman. 2009. Model Assessment and Selection. *In* The

613   elements of statistical learning. Springer, New York, NY.

614

615   Hilborn, R., Bue, B.G., and Sharr, S. 1999. Estimating spawning escapements from periodic

616   counts: a comparison of methods. Canadian Journal of Fisheries and Aquatic Sciences

617   56: 888–896.

618

619   Hillman, T., M. Miller, T. Miller, M. Tonseth, M. Hughes, A. Murdoch, J. Miller, and B. Kessee.

620   2011. Monitoring and evaluation of the Chelan County PUD hatchery programs: 2010

621   annual report. Report to the HCP Hatchery Committee, Wenatchee, WA.

622

623   Howell, P. J., and P. M. Sankovich. 2012. An evaluation of redd counts as a measure of bull trout

624   population size and trend. North American Journal of Fisheries Management. 32:1-13.

625

626   Jacobs, S.E., W. Gaeuman, M.A. Weeber, S. L. Gunckel, and S. J. Starcevich. 2009. Utility of

627   probabilistic sampling design to determine bull trout population status using redd counts

628   in basins of the Columbia River Plateau. North American Journal of Fisheries

629   Management 29:1590-1604.

630

631 Korman, J., R.N.M. Ahrens, P.S. Higgins, and C.J. Walters. 2002. Effects of observer efficiency,

632   arrival timing, and survey life on estimates of escapement for steelhead trout

633   (Oncorhynchus mykiss) derived from repeat mark-recapture experiments. Canadian

634   Journal of Fisheries and Aquatics Sciences 59:1116-1131.

635

636 Liermann, M.C., Rawding, D., Pess, G.R. and Glaser, B., 2014. The spatial distribution of

637   salmon and steelhead redds and optimal sampling design. Canadian Journal of Fisheries

638   and Aquatic Sciences, 72:434-446.

639

640 Madej M.A. 1999. Temporal and spatial variability in thalweg profiles of a gravel-bed river.

641   Earth Surface Processes and Landforms 24:1153–1169.

642

643 Millar, R.B. S. McKechnie and C.E. Jordan. 2012. Simple estimators of salmonid escapement

644   and its variance using a new area-under-the-curves method. Canadian Journal of Fisheries

645   and Aquatics Sciences 69:1002-1015.

646

647 Millar, R. B. and Jordan, C. E. 2013. A simple variance estimator for the trapezoidal area-under-

648   the-curve estimator of the spawner abundance of pacific salmon. Canadian Journal of

649   Fisheries and Aquatic Sciences, 70(8):1231–1239.

650

651 Muhlfeld, C.C., M.L. Taper, D.F. Staples, and B.B. Shepard. 2006. Observer error structure in

652   bull trout redd counts in Montana streams: implications for inference on true redd

653   numbers. Transactions of the American Fisheries Society 135:643-654.

654

655   Mullan, J.W., K.R. Williams, G. Rhodus, T.W. Hillman, and J.D. McIntyre. 1992. Production

656   and habitat of salmonids in mid-Columbia River tributary streams. United States Fish and

657   Wildlife Service Monograph 1, Leavenworth, Washington.

658

659   Murdoch, A.R., T.N. Pearsons, and T.W. Maitland. 2010. Estimating the spawning escapement

660   of hatchery and natural origin spring Chinook salmon using redd and carcass data. North

661   American Journal of Fisheries Management 30:361-375.

662

663   National Marine Fisheries Service (NMFS). 2016. 2016 5-year review: summary and evaluation

664   of Upper Columbia River Steelhead and Upper Columbia River Spring-run Chinook

665   Salmon. National Marine Fisheries Service West Coast Region, Portland, OR.

666

667   Parken, C. K, R.E. Bailey and J.R. Irvine. 2003. Incorporating uncertainty into area-under-the-

668   curve and peak count salmon escapement estimation. North American Journal of

669   Fisheries Management 23:78-90.

670

671   R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation

672   for Statistical Computing, Vienna, Austria.

673

674   Snow, C., C. Frady, A. Repp, A. Murdoch, M.P. Small, S. Bell, and C. Dean 2011. Monitoring

675   and evaluation of Wells and Methow hatchery programs in 2010. Report to the Douglas

31

676   County Public Utility district and the Wells Habitat Conservation Plan Hatchery

677   Committee, Wenatchee, WA.

678

679   Stevens, D. L. 2002. Sampling design and statistical methods for the integrated biological and

680   physical monitoring of Oregon streams. Oregon State University, Department of

681   Statistics, Corvallis, Oregon, and EPA National Health and Environmental Effects

682   Research Laboratory, Western Ecology Division, Corvallis, Oregon.

683

684   Thurow, R.F., and C.C. McGrath. 2010. Evaluating the bias and precision of Chinook Salmon

685   redd counts in the Middle Fork Salmon River basin, Idaho. Bonneville Power

686   Administration, Project 2002-049-00, Portland, Oregon.
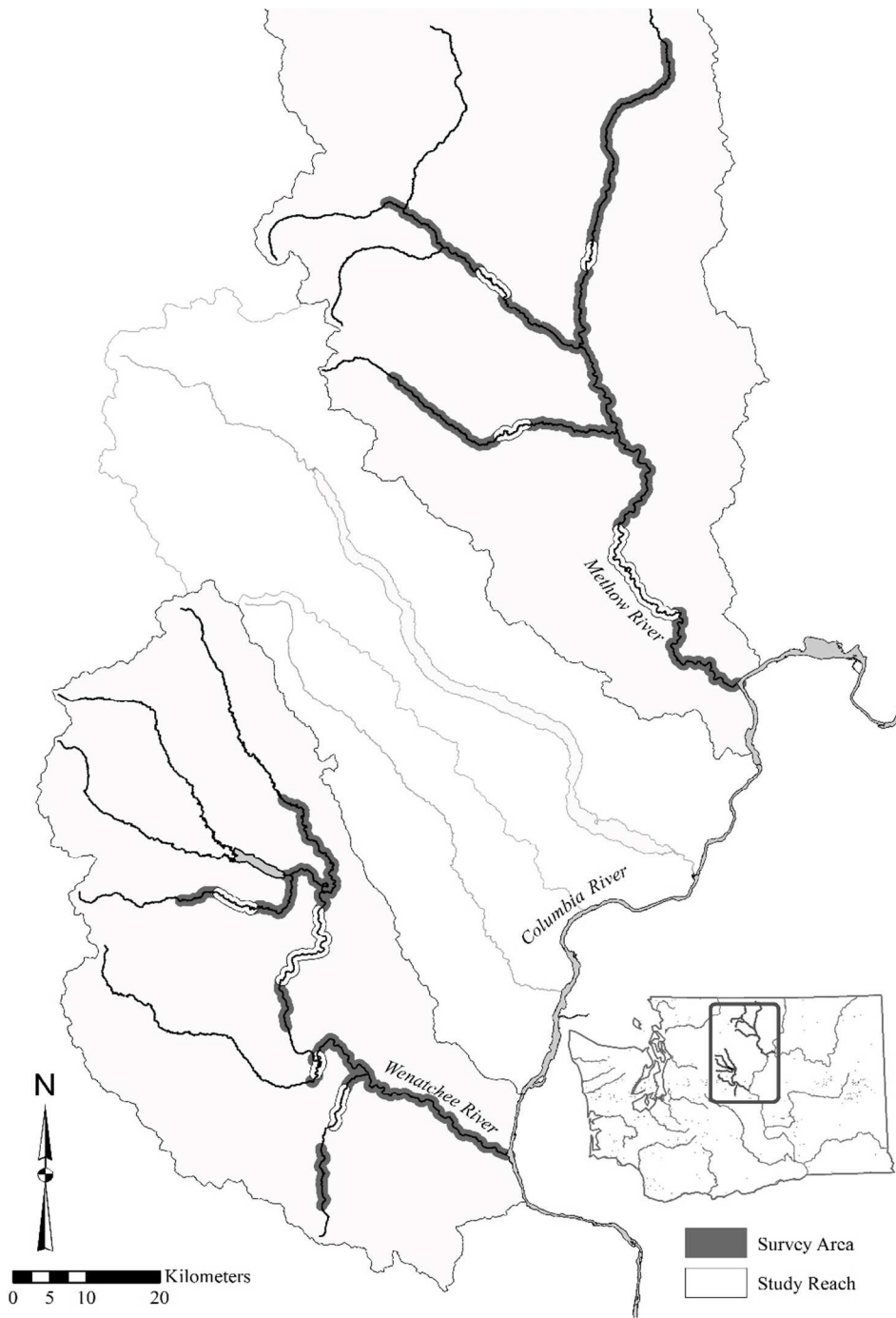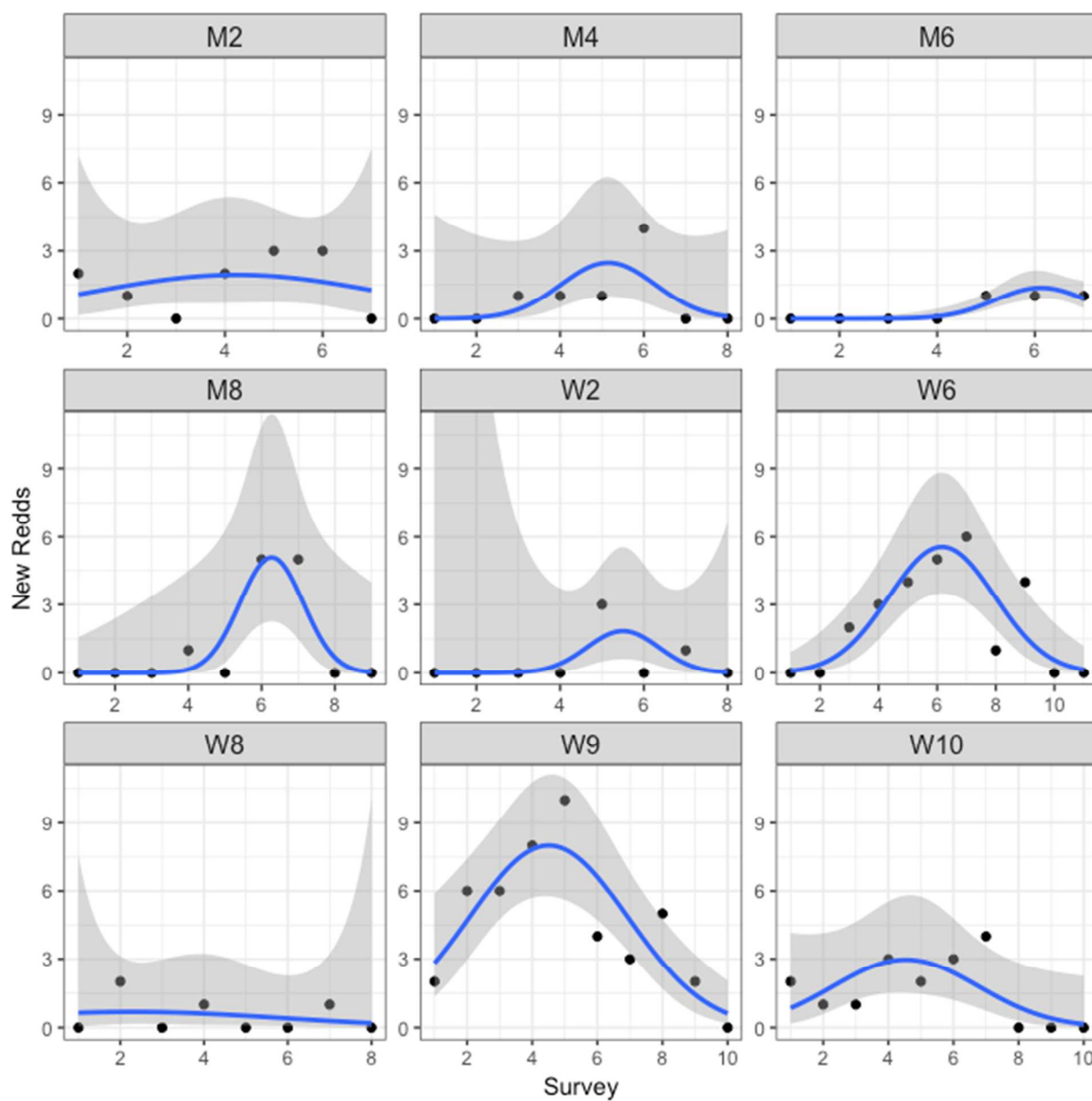
687

32

1

2   Figure 1. Vicinity map of steelhead survey area and study reaches in the Wenatchee and Methow

3   watersheds.

1

4

5

Figure 2. Plots of observed redd counts (black dots) through time for each major spawning reach,

and the fitted curve from the GAUC model (blue line) with associated 95% confidence interval

(gray).

9

1   Table 1. Mean and standard deviation (SD) of reach characteristics and covariates included in

2   observer error model selection.

Variable	Wenatchee		Methow	
	Mean	SD	Mean	SD
Reach Length (km)	8.27	4.84	9.75	7.74
Stream width (m)	30.83	19.21	33.87	11.49
Gradient (%)	0.82	0.98	0.29	0.18
Redd density (#/km)	22.99	38.08	1.9	1.8
Water Visibility (m)	15.01	3	13.57	5.28
Discharge (CFS)	1,103	1,257	1,069	704.4
Water depth (m)	64.49	13.98	0.65	0.34
Thalweg CV (%)	40.33	11.99	37.45	13.47
Experience	27.91	17.24	41.48	31.44
Effort (min/m)	0.03	0.01	0.03	0.01

3

4

5

6

7

8

1

9   Table 2. Mean observer error rates and standard deviation (in parentheses) of naïve surveys

10   conducted in the selected reaches of major steelhead spawning areas in the Wenatchee and

11   Methow watersheds.

Watershed	Year	n	Efficiency	Accuracy	Net Error
Wenatchee	2011	38	0.493 (0.230)	0.827 (0.172)	0.609 (0.309)
	2012	14	0.453 (0.270)	0.674 (0.245)	0.673 (0.319)
	All	52	0.482 (0.230)	0.786 (0.172)	0.626 (0.309)
Methow	2012	20	0.441 (0.257)	0.772 (0.144)	0.596 (0.371)
	2013	23	0.566 (0.267)	0.768 (0.135)	0.757 (0.377)
	All	43	0.508 (0.267)	0.770 (0.137)	0.682 (0.378)

12

13

14

15

16

17

18

19

20

21

22

23

2

24 Table 3. Fixed effects variables from top candidate redd observer error models (only models with

25 $\Delta AIC_c < 2$ are shown) when using a single redd observer in the Wenatchee Basin and two redd

26 observers in the Methow Basin. N is the number of parameters estimated.

Model	N	AICc	ΔAICc	AICc weight
Wenatchee Basin (single redd observer)				
Efficiency ~ Water depth + Thalweg CV + Redd density	4	487.2	0.00	0.87
Accuracy ~ Effort	2	284.5	0.00	0.57
Net Error ~ Water depth + Thalweg CV	4	8.1	0.00	0.51
Net Error ~ Water depth + Thalweg CV + Redd density	5	8.2	0.10	0.49
Methow Basin (two redd observers)				
Efficiency ~ Experience + Thalweg CV + Redd density	4	177	0.00	0.37
Efficiency ~ Gradient + Thalweg CV + Redd density	4	178.6	1.61	0.16
Efficiency ~ Thalweg CV + Redd density	3	178.7	1.72	0.16
Efficiency ~ Water depth + Thalweg CV + Redd density	4	178.8	1.82	0.15
Accuracy ~ Gradient	2	120.8	0.00	0.18
Accuracy ~ Thalweg CV	2	122	1.17	0.1
Accuracy ~ Experience + Thalweg CV	3	122.5	1.71	0.08
Accuracy ~ Gradient + Thalweg CV	3	122.7	1.92	0.07
Net Error ~ Experience + Thalweg CV + Redd density	5	16.6	0.00	0.62
Net Error ~ Discharge + Thalweg CV + Redd density	5	18.2	1.56	0.29

27

28

29   Table 4. Model parameters and statistical test results (standard error (SE) and p-value (P)) from

30   general linear regression models describing the relationship between observer error rates in the

31   Wenatchee Basin (single redd observer) and Methow Basin (two redd observers) and variables

32   from top models based on AIC scores (average of all models < 2 Δ AICc). Parameters with p-

33   value < 0.05 are in bold.

Parameter	Wenatchee			Methow		
	Estimate	SE	*P*	Estimate	SE	*P*
Efficiency						
(Intercept)	**-0.115**	**0.043**	**0.007**	-0.037	0.091	0.682
Experience			-	0.099	0.134	0.463
Gradient			-	0.193	0.482	0.688
Redd density	**0.002**	**0.001**	**0.014**	**0.436**	**0.066**	**<0.001**
Thalweg CV	**0.045**	**0.004**	**<0.001**	**-0.036**	**0.009**	**<0.001**
Water depth	**-0.045**	**0.004**	**<0.001**	0.084	0.226	0.709
Accuracy						
(Intercept)	**1.355**	**0.063**	**<0.001**	**1.163**	**0.122**	**<0.001**
Effort	**50.891**	**6.759**	**<0.001**			-
Experience			-	-0.036	0.101	0.724
Gradient			-	0.661	0.741	0.372
Thalweg CV			-	0.008	0.011	0.450
Net Error						
(Intercept)	**0.626**	**0.035**	**0.001**	**0.682**	**0.043**	**<0.001**
Discharge			-	0.0001	0.0001	0.545

4

Experience			-	0.097	0.079	0.218
Redd density	0.001	0.001	0.482	**0.135**	**0.041**	**0.001**
Thalweg CV	**0.012**	**0.003**	**<0.001**	-0.012	0.004	0.001
Water depth	**-0.014**	**0.003**	**<0.001**			-

34

35   Table 5. Summary statistics comparing observed and predicted values of various error rates.

Model	Mean observed	Mean bias	RMSE	Weighted R^2	Mean CV	95% coverage probability
Wenatchee Basin (single redd observer)						
Accuracy	0.79	0.00	0.15	0.23	0.02	0.17
Efficiency	0.48	-0.01	0.17	0.46	0.05	0.25
Net error	0.63	0.00	0.25	0.34	0.13	0.50
Methow Basin (two redd observers)						
Accuracy	0.77	-0.01	0.14	0.04	0.05	0.45
Efficiency	0.55	-0.05	0.18	0.61	0.11	0.52
Net error	0.73	-0.04	0.27	0.55	0.23	0.65

36

37

5

38   Table 6. Summary statistics comparing observed and predicted values of redds.

Model	Mean number of redds	Mean number of redds observed	Mean bias	RMSE	Normalized RMSE	R^2	Mean CV	95% coverage probability
Wenatchee Basin (single redd observer)								
Efficiency/ accuracy	54	34	0.19	20.32	0.37	0.60	0.05	0.25
Net error	54	34	-1.61	21.12	0.39	0.54	0.13	0.50
Methow Basin (two redd observers)								
Efficiency/ accuracy	15	11	0.50	4.49	0.30	0.46	0.12	0.62
Net error	15	11	0.61	5.56	0.37	0.27	0.23	0.70

39
40

6