NRC Research Press

ARTICLE

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

Abstract: This study examined how a suite of habitat and environmental variables relate to the ability of a stream surveyor to identify (observer efficiency) and distinguish (observer accuracy) steelhead (*Oncorhynchus mykiss*) redds from other stream features. Two existing spawning survey protocols that included one or two redd observers were used to develop models to estimate redd observer error. In most cases, steelhead redd abundances using raw redd counts were underestimated. Mean annual rates of observer efficiency ranged from 0.44 to 0.57, and observer accuracy ranged from 0.67 to 0.83. Regardless of the observer error model used, adjusted annual redd abundance estimates were generally unbiased (range 1.6–0.6 redds). A Gaussian area-under-the-curve methodology that incorporates redd count data and observer error rates was used to generate unbiased estimates of steelhead redd abundance in the Wenatchee (170 redds, coefficient of variation (CV) = 44%) and Methow (106 redds, CV = 41%) rivers. Unbiased estimates of redd abundance will help inform new population viability analyses to better prioritize those populations with the greatest conservation need.

Résumé : L'étude examine le lien entre un ensemble de variables reliées à l'habitat et au milieu et la capacité d'un agent d'évaluation des cours d'eau de déceler (efficacité de l'observateur) et de distinguer (exactitude de l'observateur) les nids de frai de truites arc-en-ciel (*Oncorhynchus mykiss*) d'autres éléments du cours d'eau. Deux protocoles existants d'inventaire de géniteurs qui comprennent un ou deux observateurs des nids de frai ont été utilisés afin de développer des modèles pour estimer l'erreur associée à l'observateur. Dans la plupart des cas, l'abondance des nids de frai de truites arc-en-ciel obtenue à partir des nombres bruts de nids dénombrés est sous-estimée. Les taux annuels moyens d'efficacité de l'observateur vont de 0,44 à 0,57 et l'exactitude de l'observateur va de 0,67 à 0,83. Peu importe le modèle d'erreur associée à l'observateur utilisé, les estimations ajustées de l'abondance annuelle des nids de frai sont généralement non biaisées (fourchette : 1,6–0,6 nid de frai). Une méthodologie de surface sous la courbe gaussienne qui incorpore des données de dénombrement des nids de frai et les taux d'erreur de l'observateur a été utilisée pour produire des estimations non biaisées de l'abondance de nids de frai de truites arc-en-ciel dans les rivières Wenatchee (170 nids, le coefficient de variation (CV) = 44 %) et Methow (106 nids, CV = 41 %). Des estimations non biaisées de l'abondance des nids de frai éclaireront de nouvelles analyses de la viabilité de populations pour permettre une meilleure priorisation des populations ayant les plus grands besoins en matière de conservation. [Traduit par la Rédaction]

Introduction

Estimates of redd abundance are a common metric used to monitor the status and trend of adult salmonid populations (Dauble and Watson 1997; Jacobs et al. 2009; Howell and Sankovich 2012). When used in conjunction with other population-specific data (e.g., spawners per redd), redd abundance can be used to estimate spawner abundance (Gallagher and Gallagher 2005; Murdoch et al. 2010). Raw counts of redd abundance are inherently inaccurate due to a combination of observer experience, habitat, and environmental conditions that influence a redd observer's ability to accurately detect and identify redds (Dunham et al. 2001; Muhlfeld et al. 2006; Howell and Sankovich 2012). Nevertheless, redd surveys are conducted in a wide range of habitats for many salmonid species because redd counts are reported to be correlated with spawner abundance (Dunham et al. 2001; Gallagher and Gallagher 2005; Gallagher et al. 2010; Murdoch et al. 2010) and can be conducted at a relatively low cost (Chasco et al. 2014). Historically, redd-based estimates of spawner abundance have adequately served as an index of spawner abundance for most management objectives, but growing conservation concerns (e.g., populations listed under the United States Endangered Species Act) now require unbiased estimates of spawner abundance with known levels of uncertainty (Crawford and Rumsey 2011).

Redd survey design and methodology are important considerations that have changed through time in response to management questions. For some populations, aerial surveys were abandoned for more accurate ground-based surveys. Single-pass peak redd counts were eventually replaced with multiple-pass index area redd counts, which eventually lead to weekly redd counts of all available spawning habitat or census redds counts. At spatial scales larger than the stream or population level, more complex survey designs (e.g., generalized random tessellation stratified) have also been successfully used (Stevens 2002; Boydstun and McDonald 2005). Chasco et al. (2014) retrospectively evaluated the accuracy and precision of three common Chinook salmon (*Oncorhynchus tshawytscha*) redd survey designs (i.e., single-pass aerial, multiple-pass index area, and multiple-pass census) in Johnson Creek, Idaho, using a state-space model. Spawner abundance estimates based on multiple-pass census redd counts were shown to be com-

Received 14 August 2017. Accepted 23 February 2018.

A.R. Murdoch, C.J. Herring, and C.H. Frady. Washington Department of Fish and Wildlife, 600 Capitol Way North, Olympia, WA 98501, USA.
K. See. Quantitative Consultants Inc., 2725 Montlake Blvd. East, Seattle, WA 98112, USA.
C.E. Jordan. Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, 2725 Montlake Blvd. East., Seattle, WA 98112, USA.

Corresponding author: Andrew R. Murdoch (email: Andrew.Murdoch@dfw.wa.gov).

parable to mark–recapture-based estimates in terms of both accuracy and precision, but implemented at a fraction of the cost.

Selecting the correct redd survey design minimizes sampling error (Liermann et al. 2015), but estimating and adjusting for observation error in redd counts has only been recently investigated. The accuracy of redds counts may be influenced by errors of omission (i.e., missed redds) or commission (i.e., false identification). While these two sources of error can cancel each other, if the two error rates are not identical, the redd count will be biased in the direction of the largest source of error. Muhlfeld et al. (2006) reported the bull trout (*Salvelinus confluentus*) redd abundance was highly correlated with errors of omission, but unrelated to errors of commission, suggesting the processes behind the two sources of error are unrelated and unique. Surveyor experience was found to be positively related to observer efficiency in all studies (Dunham et al. 2001; Muhlfeld et al. 2006; Howell and Sankovich 2012) except one (Dunham et al. 2001). In that study, 82% of the redd observers had little to no experience conducting bull trout surveys. Hence, large variation in redd observer efficiency and the limited range of surveyor experience may explain why no relationship was detected. Dunham et al. (2001) did report that several redd and habitat characteristics were found to be significantly correlated with error rates of both omission and commission. Gallagher and Gallagher (2005) estimated mean redd observer efficiency rates for steelhead (*Oncorhynchus mykiss*) in coastal Northern California streams based on the proportion of flagged redds observed. Mean redd observer efficiency was negatively related to weekly stream discharge and positively related to maximum visible water depth, but errors of commission (false identification) or experience of the redd observers were not assessed. A comprehensive evaluation of factors influencing redd observer error has not been conducted for any anadromous salmonid species. Understanding the factors driving observer error can be used to minimize (e.g., number or experience of observers) or account for the error in the estimate (e.g., area-under-the-curve (AUC)). These relationships among observer experience, habitat characteristics, environmental conditions, and observer error rates may be useful in developing models to predict observer error for each river or river segment of interest.

Estimates of observer error coupled with redd life and count data are also required to use the AUC methodology to estimate total redd abundance (Hilborn et al. 1999; Parken et al. 2003). The AUC methodology, or more accurately the trapezoidal AUC (TAUC), involves plotting a times series of count data (e.g., fish or redds) or an escapement curve and dividing the AUC by the redd life or residence time and observer efficiency (Parken et al 2003; Millar et al. 2012). The TAUC approach has been successfully adopted to various conditions, but incorporating the uncertainty of each parameter into the estimate has been challenging. Parken et al. (2003) developed a method for incorporating uncertainty into the escapement estimate using costly replicate surveys. Conversely, Hilborn et al. (1999) developed a maximum likelihood estimator with an estimate of standard error, but was computationally difficult to generate and has not been widely adopted. In addition, the TAUC estimator generally outperformed the maximum likelihood estimator when compared with weir counts (Hilborn et al. 1999). Millar et al. (2012) recently developed simpler analytical methods to estimate escapement and variance using a Gaussian AUC (GAUC) methodology. While the traditional TAUC estimator performed slightly better than the GAUC estimator, the later has an estimate of uncertainty that has been lacking for decades. The GAUC methodology was adopted to fit with existing steelhead redd survey protocols and redd count data sets, but estimates of observer efficiency were still needed. The objectives of this study were to (*i*) measure redd observer error under an existing steelhead redd survey protocol using two different levels of effort (i.e., one observer and two observers), (*ii*) develop predictive models to estimate observer error, (*iii*) use those predictions of observer error to subsequently estimate the total number of steelhead redds using a GAUC methodology, and (*iv*) compare results of both survey protocols to inform the future direction of steelhead redd surveys.

Methods

Study area

The Wenatchee and Methow rivers are located in north central Washington State and drain 3439 and 4727 km² of the Cascade Mountains eastern slope, respectively (Fig. 1). The Wenatchee watershed contains 332 km of major streams with 207 km accessible to anadromous salmonids (Mullan et al. 1992). The Wenatchee River flows into the Columbia River at river kilometre (rkm) 754 and is located between Rock Island and Rocky Reach dams. The Methow watershed contains 270 km of major streams with 239 km accessible to anadromous salmonids (Mullan et al. 1992). The Methow River flows into the Columbia River at rkm 843 and is located between Wells Dam and the confluence of Okanogan River. Steelhead in both rivers are classified as summer-run and are listed as threatened under the Endangered Species Act (NMFS 2016) as part of the Upper Columbia Distinct Population Segment. Summer-run steelhead enter fresh water the year prior to spawning, typically June through October, become sexually mature in their natal watershed, and spawn the following spring between March and May. Depending on the annual snowpack levels, stream discharge in the Wenatchee and Methow rivers during the spawning period may increase over 200% and 600%, respectively.

Study reach selection

Comprehensive steelhead redd surveys have been conducted in all major spawning areas of both watersheds since 2004 (Hillman et al. 2011; Snow et al. 2011). Steelhead spawn throughout the entire length of the Wenatchee and Methow rivers and in at least some reaches of all the major tributaries. Given the diverse habitat and environmental conditions, study reaches were selected to capture the maximum variability in habitat and environmental conditions that surveyors experience while conducting redd surveys. Four study reaches were selected in each watershed that represented the range of survey conditions (Table 1). Korman et al. (2002) reported that the variance of observer efficiency estimates (coefficient of variation (CV) range 27%–100%) increased as the proportion of steelhead observed (range 0.05–0.61) decreased. Hence, study reaches were also selected to minimize the influence of the interannual variability in redd abundance and the subsequent variance in observer error rates to facilitate model development.

Redd surveys

Redd observer error was assessed by comparing intensive census ground redd counts with naïve observer redd counts similar to that described in Thurow and McGrath (2010). All redd observers in the study received a similar, albeit limited, level of training that included both a classroom and field portion. Census surveys were conducted every 3 days in all study reaches by experienced surveyors (mean = 6 years; range 4–10 years) and began prior to the start of spawning and ended when no new redds were found. Census surveyors did not change reaches throughout the spawning period. It was assumed that redds found during census surveys were an accurate, unbiased count of the total number of redds in each reach. Features that could be misinterpreted as steelhead redds (e.g., old salmon redds or hydrologic features) or errors of commission were marked, including a unique georeferenced waypoint, and numbered sequentially on aerial photographs. Aerial photographs were laminated for use in the field. The mean reach length per photograph was 445 m. Areas of localized high-density spawning were also annotated on detailed hand-drawn maps. If redd superimposition was observed, the proportion of the redd disturbed by another female steelhead was estimated during each survey. During census surveys, the date when redds became no

Fig. 1. Vicinity map of steelhead survey area and study reaches in the Wenatchee and Methow watersheds.

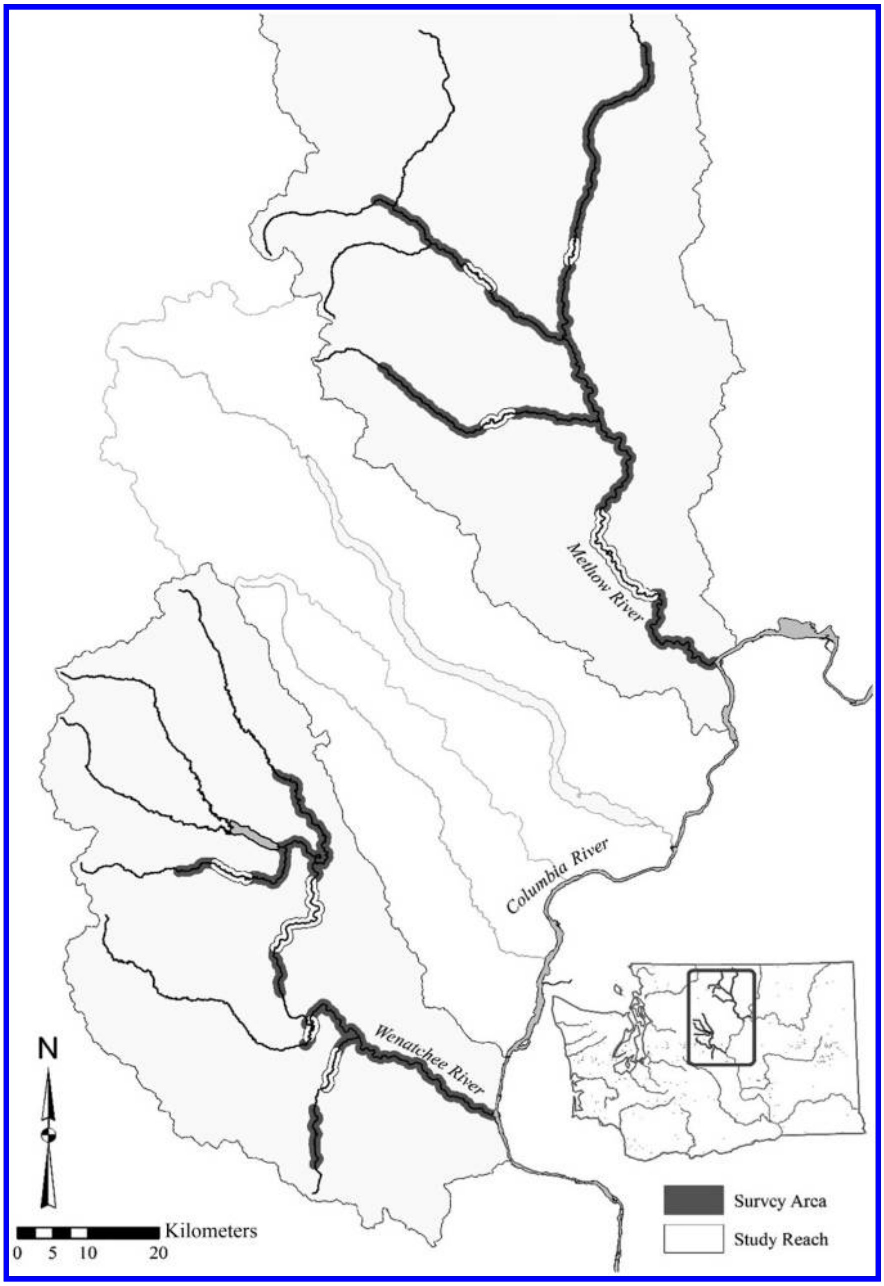


Table 1. Mean and standard deviation (SD) of reach characteristics and covariates included in observer error model selection.

	Wenatchee		Methow	
Variable	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 (no.·km^{-1})	22.99	38.08	1.90	1.80
Water visibility (m)	15.01	3.00	13.57	5.28
Discharge (CFS)*	1103.0	1257.0	1069.0	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^{-1})	0.03	0.01	0.03	0.01

*1 ft^3·s^{-1} = 28.317 L·s^{-1}.

longer visible due to superimposition or environmental conditions was recorded (Gallagher et al. 2007). Flagging was not attached to vegetation adjacent to steelhead redds within study reaches to reduce potential bias during subsequent naïve surveys. Surveys were conducted using one- or two-person catarafts depending on the size of stream. When a two-person cataraft was used, the redd observer provided verbal direction to the rower at all times. Surveys were generally conducted during periods of maximum potential sunlight (1000–1500 h). All surveyors wore polarized sunglasses and a brimmed cap. If actively spawning fish were observed, surveyors used caution to avoid disturbing fish.

Surveys were conducted near the peak of spawning by naïve observers to estimate redd observer accuracy and efficiency. Consistent with the methodology typically used in each respective watershed, a single redd observer was used in the Wenatchee watershed and two redd observers in two catarafts were used in the Methow watershed. Naïve surveys were generally conducted on the same day as a census survey, at a typical level of effort. However, because variation in survey effort does exist among surveyors, the start and stop time of each survey was recorded to assess the influence of effort on observer error. Naïve observers had no prior knowledge of the number of redds present and used methods identical to those for census surveys. All observers, both census and naïve, were instructed not to discuss the results of any survey during the study period. The variability in experience of the naïve observers was maximized to the extent possible to understand better the influence of experience on observer error. The number of replicate naïve surveys conducted on each reach was dependent on personnel availability and environmental conditions.

Redd observer error model covariates

Experience of the redd observer may be an important factor influencing observer error (Muhlfeld et al. 2006; Howell and Sankovich 2012). Observer experience was quantified based on the species and frequency of surveys. Because all salmonids that spawn in streams exhibit similar behavior during spawning (i.e., constructing a redd), experience was pooled across species. Experience was quantified based on the number of years or seasons an individual had conducted redd surveys. For example, an individual that conducted Chinook salmon redd surveys 1 day a week during the entire spawning season for 5 years received a value of five. However, if the same observer also conducted steelhead redd surveys 1 day a week throughout the spawning season for 3 years, the surveyor would receive a value of eight for experience conducting salmonid spawning ground surveys. Partial years or single surveys, regardless of classification, were prorated based on a maximum of ten surveys per season (i.e., one survey = 0.1 points). Because we believed there was a saturating effect of experience, the log of the experience value was used in error models. Survey effort of the naïve surveyors was standardized by dividing survey time by reach length. Observed redd density was calculated by dividing the total length of each study reach by the total number of redds or redd-like features identified by naïve surveyors.

Habitat variables describing the physical characteristics of each reach were measured at the beginning of the spawning season along ten equidistant transects established perpendicular to the flow or was summed across the entire reach (Bartley and Rutherfurd 2005). Mean stream width and water depth for each study reach were calculated from the width and depth recorded at each transect. Horizontal water visibility was estimated for each survey using a standard Secchi disc. At the start and end points of the study reach, a Secchi disc was placed in approximately 0.5 m of water with a laminar flow pattern. A depth of 0.5 m was selected to provide observers the ability to remain standing in the stream channel throughout the spawning season, as spring snow melt increases river discharge and water velocity. The observer would walk upstream to a point when the white and black sections of the disc were no longer discernible. The distance between the observer and Secchi disc was measured on the stream bank using a tape measure. The mean distance of the two measurements was used as an index of horizontal water visibility for that survey. River discharge measured at the nearest gauging station was recorded at the beginning and end of the survey. The mean discharge for that survey day was used in the model.

The influence of channel or habitat complexity was quantified using the CV in the thalweg profile (Madej 1999; Bartley and Rutherfurd 2005). Because the influence of channel complexity associated with spawning sites was the metric of interest, steelhead redd locations were used to select sample locations for the thalweg profile. The redd location was used as the midpoint of the sample to ensure representation of channel complexity at the spawning site. A longitudinal profile of water depth was measured every 3 m along the thalweg, extending 150 m upstream and 150 m downstream from each of three randomly selected redds in each study reach. The thalweg CV was calculated for each 300 m sample location. The mean thalweg CV for each study reach was used as an index of channel complexity. Data on possible covariates collected during census surveys are summarized in Table 1.

Redd observer error rates

Individual redds identified during naïve surveys were compared with redds identified in census surveys using georeferenced redd locations, aerial photographs, and hand-drawn maps. Redd observer error may occur when redds are falsely identified (commission) or when redds are missed (omission) and were based only on redds that were still visible during the naïve surveys (Thurow and McGrath 2010). Redd observer accuracy was defined as the proportion of features identified by an observer that were steelhead redds and not confused with other features created hydraulically (i.e., scour and deposition) or redds from other species:

$$A_i = R_i/F_i \tag{1}$$

where A is the accuracy of naïve redd survey i, R_i is the number of correctly identified steelhead redds found by both the naïve and census surveyors during naïve redd survey i, and F_i is the total number of redd-like features found by the naïve redd observer i. Redd observer efficiency was defined as the proportion of steelhead redds correctly identified during a survey:

$$E_i = R_i/V_{\mathrm{Cen}} \tag{2}$$

where E is the efficiency of naïve survey i, R_i is the number of correctly identified steelhead redds found by both the naïve and census surveyors during naïve redd survey i, and V_{Cen} is the number of redds found by the census surveyor that were still visible on the day of naïve survey i. Hence, modeling the error structure of steelhead redd observers requires a two-stage process that includes (*i*) estimating the proportion of features identified during a survey that were steelhead redds and (*ii*) estimating the proportion of steelhead redds that were correctly identified during the survey. Alternatively, when total redd abundance, not specific redd locations or distribution, is the primary metric of interest, the sum of both sources of error (i.e., net error rate) may be modeled more simply in a single step. Net error rates were calculated by dividing the number of redd-like features identified during naïve survey i by the number of redds found by the census surveyor that were still visible on the day of naïve survey i:

$$\mathrm{NE}_i = F_i/V_{\mathrm{Cen}} \tag{3}$$

where NE is the net error of naïve survey i. Net error rates may be less than or greater than one, depending on which source of error (i.e., omission or commission) was the greatest. Correlations among error rates (commission, omission, and net error) were examined using a Spearman correlation coefficient.

Model selection

Observer efficiency, observer accuracy, and net error rates were modeled separately. All covariates were centered to have a mean of zero. Separate models were also developed based on the number of observers (i.e., Wenatchee = 1; Methow = 2). Observer efficiency and accuracy error rates were modeled with a logit link using a binomial error structure, while net error was modeled as a linear model with Gaussian error structure, with a list of possible covariates including habitat metrics, redd density, survey effort, and observer experience. The model selection process involved fitting a series of models with all possible combinations of covariates, including an intercept-only model, using Akaike information criteria corrected for small sample size ($\mathrm{AIC_c}$) to identify the most parsimonious model (Burnham and Anderson 2002).

In addition, covariates with a correlation >0.5 were not included in the same candidate model. The AIC_c score of each model was recorded, and the model with the lowest AIC_c score was considered the most parsimonious model supported by the data (Akaike 1974). All models with a $\Delta AIC_c < 2$ were equally supported by the data (Burnham and Anderson 2002). Therefore, model averaging was conducted with all models with a $\Delta AIC_c < 2$ to make predictions of error rates. All statistical analyses were conducted using R software (R Core Team 2015).

Model performance

A leave-one-out cross-validation (LOOCV) approach was used to assess the predictive performance of the different models (Hastie et al. 2009). To perform a LOOCV, we took all the models with a ΔAIC_c of less than 2 and refit them to the entire data set except for a single data point. We then predicted the error rates for that point using model-averaged coefficients and then predicted the number of redds for that data point as well. We cycled through the entire data set, leaving each point out once. Comparisons (i.e., mean bias, root mean squared error (RMSE), weighted correlation, and coverage probability) between those predictions and the observed values were made to assess model performance. Mean bias describes the average difference between observed and predicted values and can be used to assess whether a model's predictions are unbiased. The RMSE describes how far off in either direction the predictions are from the truth, on average. For predicted redds, a normalized RMSE value divides the RMSE by the mean number of observed redds, providing the mean proportion of redds that the prediction was off by. Weighted correlations describe the correlation between observed and predicted values, weighted by the relative number of observed redds, so that predictions for surveys that encountered a large number of redds are more heavily weighted than surveys that encountered few redds. Because surveys encountered many fewer redds in the Methow than in the Wenatchee, the relative number of observed redds was calculated by dividing the number of observed redds by the maximum number of redds observed in that watershed, so as to not overly discount the Methow surveys. Coverage probabilities assess the rate at which the 95% confidence interval encompasses the true value. A good model should have a coverage probability near 95%, whereas one with a low coverage probability is not capturing enough of the uncertainty.

Application of GAUC to steelhead redd survey data

Millar et al. (2012) presented a less computationally demanding analytical approach to estimating AUC abundance and variance. As traditionally applied, AUC uses a time series plot of spawner count data to estimate spawner-days, which is then divided by mean spawner life and observer efficiency to estimate spawner abundance. The key difference between spawner counts and redd surveys is that it is impossible to determine which spawners were counted in the previous survey, whereas redds can be georeferenced. This eliminates the need to account for redd life because a time series of previously unseen redds or new redds can be used instead of total visible redds, using survey number instead of date as the point of reference. However, the GAUC methodology still requires an estimate of observer error, with standard errors. Based on the suite of covariates included in the top models, estimates of redd observer error were generated for each survey, and then reach means and standard deviations (SDs) were calculated to be used in the GAUC calculations.

In 2014, steelhead redd observers conducted weekly surveys of the major spawning reaches in both the Wenatchee and Methow rivers. Selected model covariates were collected as previously described. Because the steelhead spawning distribution is known to be clumped (Hillman et al. 2011; Snow et al. 2011), presumably due to heterogeneous spawning habitat quality, minor spawning reaches were only surveyed at peak spawning, thereby substantially decreasing costs of conducting redd surveys. For these minor spawning reaches, the number of visible redds at peak spawning was adjusted for the predicted observer error of the adjacent major spawning reach to generate an estimate of total redd abundance. Estimates of redd abundance and variance were summed across all major and minor spawning reaches to generate a total redd estimate for each river.

Table 2. Mean observer error rates and standard deviation (in parentheses) of naïve surveys conducted in the selected reaches of major steelhead spawning areas in the Wenatchee and 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)

Results

Single redd observer error rates (Wenatchee Basin)

During the study period, 52 naïve surveys were conducted to estimate observer error (Table 2). The mean number of naïve surveys conducted in a single day was 2.4 (SD = 1.0). On average, naïve observers identified (i.e., efficiency) only 48% (SD = 23%) of the steelhead redds visible on the day of the survey, indicating that a majority of redds were missed. However, redd observers were more accurate (mean = 79%; SD = 17%) than efficient, meaning that while observers missed a large proportion of visible redds, it was more likely that those redds identified were true steelhead redds and not confused with redd-like features. Accuracy was not significantly correlated with observer efficiency ($r_s = 0.20$, $P = 0.16$) or net error rates ($r_s = -0.14$, $P = 0.33$), but was higher than observer efficiency in every study reach, resulting in intermediate rates of net error. However, because efficiency was generally smaller than accuracy, the mean net error (63%; SD = 31%) was less than one. As expected, observer efficiency was correlated with net error rate ($r_s = 0.86$, $P < 0.001$), and a majority of naïve surveys (88%, $N = 46$) underestimated the redd abundance regardless of observer efficiency.

When modeling observer efficiency, stream depth, observed redd density, and the mean thalweg CV were all included in the top model (Table 3). The second-best model included the same covariates except for observed redd density and had a difference of over 3 AIC_c units. After that, no other model was within 10 AIC_c units. The effects of these main covariates suggest that efficiency improves when stream complexity or redd density increases, while it declines when the stream depth increases (Table 4). For observer accuracy, the only covariate included in the top model was standardized effort (Table 3). As expected, accuracy improved when effort increased, suggesting when surveyors spend more time evaluating redd-like features, they are less likely to make errors of commission (Table 4). The mean CV of thalweg depth and observed redd density were also relatively important, but models containing these covariates had a ΔAIC_c greater than 2, suggesting most of the variation in observer accuracy is captured by effort. The same three covariates (stream depth, the observed redd density, and the mean thalweg CV) most important in explaining observer efficiency were also the most important for net error (Tables 3 and 4). Given the correlation between observer efficiency and net error rates, it was expected that the same covariates would also be important in predicting net error rates.

Model-averaged predictions of error rates had moderate weighted correlation coefficients with observed rates in a LOOCV analysis (efficiency = 0.68, accuracy = 0.48, net error = 0.59). The RMSE of the LOOCV exercise was 17% for observer efficiency, 15% for observer accuracy, and 25% for net error.

Table 3. Fixed effects variables from top candidate redd observer error models (only models with $\Delta AIC_c < 2$ are shown) when using a single redd observer in the Wenatchee Basin and two redd observers in the Methow Basin.

Model	*N*	AIC_c	ΔAIC_c	AIC_c 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	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	122.0	1.17	0.10
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

Note: *N* is the number of parameters estimated.

Table 4. Model parameters and statistical test results (standard error (SE) and *P* value) from general linear regression models describing the relationship between observer error rates in the Wenatchee Basin (single redd observer) and Methow Basin (two redd observers) and variables from top models based on AIC scores (mean of all models < 2 ΔAIC_c).

	Wenatchee			Methow		
Parameter	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
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**			—

Note: Parameters with $P < 0.05$ are in bold.

Two redd observer error rates (Methow watershed)

During the study period, 43 naïve surveys were conducted to estimate redd observer error (Table 2). The mean number of naïve surveys conducted in a single day was 1.7 (SD = 0.5). On average, naïve observers identified (i.e., efficiency) only 51% (SD = 27%) of the steelhead redds visible on the day of the survey, or they found only slightly more than they missed. As was the case under a single observer protocol, a majority of the steelhead redds found were not misidentified and were true steelhead redds (mean = 77%, SD = 14%). Observer accuracy was not significantly correlated with observer efficiency ($r_s = -0.08$, $P = 0.62$), but was modestly correlated with net error ($r_s = -0.39$, $P = 0.01$). Similar to the single observer error rates in the Wenatchee watershed, rates of omission were generally larger than rates of commission resulting in a mean net error (68%, SD = 38%) less than one. Similarly, observer efficiency was also correlated with net error rate ($r_s = 0.92$, $P < 0.0001$), and a majority of naïve observers (95%, $N = 41$) underestimated the redd abundance regardless of observer efficiency.

When modeling observer efficiency, redd density and the mean thalweg CV were the most important covariates, being included in every model with $\Delta AIC_c < 2$ (Table 3). Other covariates that also contributed some information included experience, gradient, mean depth, discharge, and effort. Under a two-observer model, unlike the one-observer model, increases in thalweg CV had a negative effect on efficiency, while increases in water depth had a positive effect (Table 4). As expected, increase in surveyor experience had a positive influence on observer efficiency.

Observer accuracy was modeled best with gradient and mean thalweg CV, as well as observer experience, as all models with $\Delta AIC_c < 2$ contained some combination of these three covariates (Table 3). As gradient or mean thalweg CV increases, the predicted accuracy improves, although it decreases slightly as experience

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

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

Model	Mean no. of redds	Mean no. 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

increases (Table 4). Numerous other covariates received some support, with ΔAIC_c in the 2–7 range. Although the RMSE was only 14%, the predictions were all nearly identical, regardless of which data point was left out. The intercept-only model had a ΔAIC_c of 2.74, suggesting some support for just using the mean observer accuracy rate without any covariates.

The most important covariates for modeling net error were mean discharge, mean thalweg CV, redd density, and salmonid redd survey experience. There was some support for a few other covariates, but those four covariates dominated the Akaike weights (Table 3). These covariates also had a similar direction of influence on net error as was observed in modeling observer efficiency (Table 4).

In the LOOCV, the weighted correlation coefficients were fairly high for efficiency and net error, but low for accuracy (efficiency = 0.78, accuracy = 0.2, net error = 0.74). The RMSE of the LOOCV exercise was 18% for observer efficiency, 14% for observer accuracy, and 27% for net error.

Comparison of redd observer error models using one or two observers

Comparing model covariates and direction of influence on efficiency or accuracy error rates may provide insights as to possible mechanisms or causation. Redd density was the only covariate that was consistent across all efficiency and net error models (Table 4). Conversely, surveyor experience was included in all of the two-observer error models and none of the single-observer models. Naïve surveyors used in the development of the Wenatchee or single-observer models had a significantly lower mean experience level (–32%) compared with the Methow or two-observer models (t test, $P < 0.01$), and they also included a smaller range of experience levels.

The index of channel complexity or thalweg CV also had inconsistent effects on model performance, but was not significantly different between the two watersheds (t test, $P = 0.27$). A single redd observer floating downstream must choose where to look for redds within the stream channel. Hence, as channel complexity increases and areas of gravel sorting (e.g., pool tailouts, log jams, split channels) are more readily identified, a single observer may make better decisions where to look for redds. Conversely, with two redd observers, the decision of where to look for redds is less important because each observer is only looking in half the channel. However, as potential redd locations increase with greater channel complexity, the probability of identifying redds decreases. Hence, the difference in the number of observers and how they react to channel complexity during surveys best explains the different effects that channel complexity had on observer efficiency and net error.

Gradient had a positive influence, but not significant, in both the efficiency and accuracy models in the Methow watershed, but no models in the Wenatchee watershed. Because reaches in the Wenatchee watershed had a significantly greater mean gradient than those in the Methow (t test, $P < 0.001$), inclusion in the Methow models was likely attributed to the physical difference between the watersheds, not related to the number of observers, similar to the experience covariate.

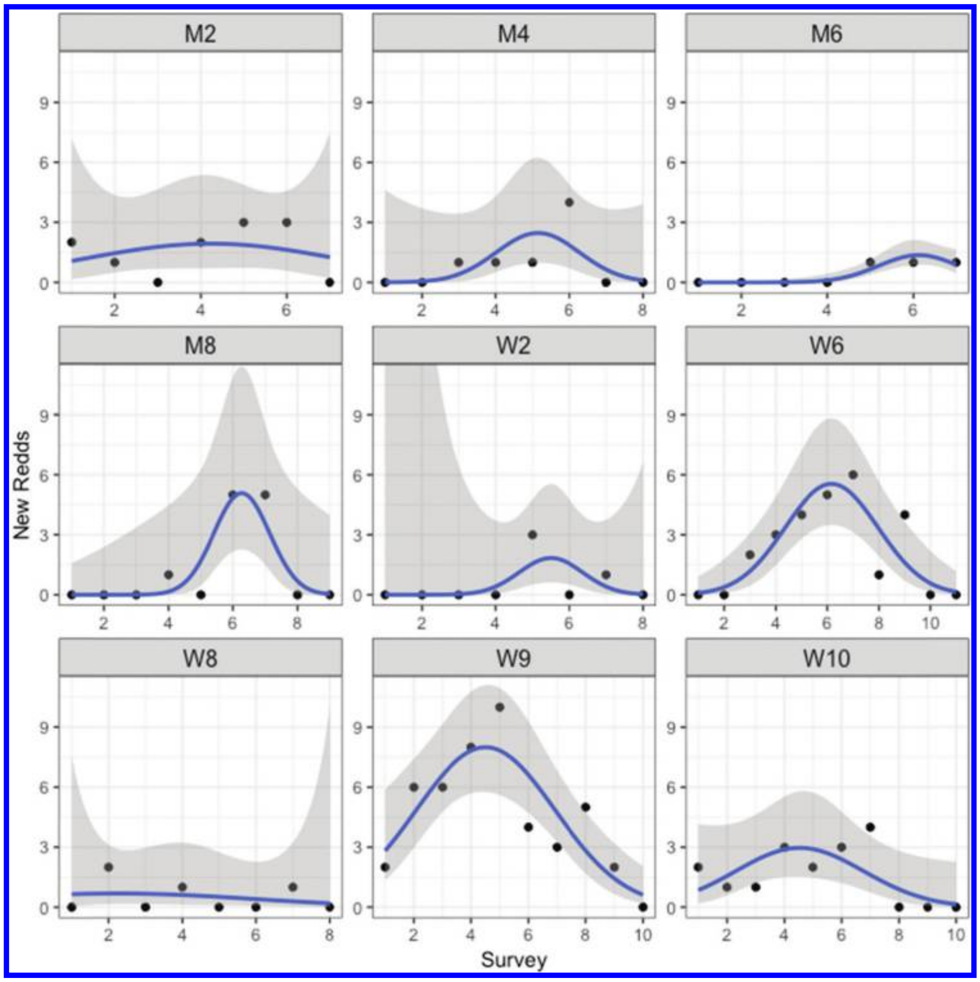
While effort (min·m^{-1}) was also not significantly different between the two watersheds (t test, $P = 0.33$), it was only included in the accuracy model for the Wenatchee watershed and likely related to differences in the number of observers. Two redd observers provide sufficient coverage within the stream channel such that any additional effort does not result in greater efficiency or accuracy. Water depth was not significantly different between the two watersheds (t test, $P = 0.87$), but had contrasting influence on observer efficiency, potentially due to differences in how redd observers navigate rivers and examine spawning habitat when there are one versus two boats.

Model performance

The models all made unbiased predictions of error rates, based on the LOOCV results (Table 5). Weighted R^2 values were fairly high for the efficiency and net error models, particularly the two-observer model in the Methow, but were low for both accuracy models. Across all six models, the RMSE ranged from 0.14 to 0.27. Based on the weighted correlations, the Methow models using two redd observers outperformed the Wenatchee models using a single redd observer for efficiency and net error, but both accuracy models had low weighted R^2 values.

When used to predict the true number of redds in a reach, the mean bias of predicted redds compared with observed in the LOOCV was generally less than one redd (Table 6). Other metrics of performance were generally comparable between the efficiency–accuracy predictions and the net error predictions, at least within a basin. The normalized RMSE indicates that the predicted redds

Fig. 2. Plots of observed redd counts (dots) through time for each major spawning reach and the fitted curve from the Gaussian area-under-the-curve (GAUC) model (line) with associated 95% confidence interval (gray shaded area).



could be off by 30%–39%, but the unbiasedness means they were equally likely to be an overestimate as an underestimate. In contrast, assuming that the observed redds were a census would lead to an underestimate of redds by 27%–37%, a consistently negative bias.

All four models failed to capture all the appropriate uncertainty, as reflected in the 95% coverage probabilities. That is a measure of how frequently the 95% confidence interval contained the true value and should be close to 95%. The Wenatchee models made predictions that were far too precise, so even though the CV was low, the predictions were often inaccurate with inappropriately small confidence intervals. The Methow models did better, with the net error model having the highest coverage.

Application of GAUC to steelhead redd survey data

The two-observer net error model was selected because our analysis indicted that it (*i*) had a high weighted correlation between observed and predicted redds, (*ii*) produced a reasonably unbiased estimate of redds, (*iii*) provided the most appropriate 95% coverage interval, and (*iv*) involves only one model and so is more parsimonious compared with using an efficiency and accuracy model. In 2014, two observers conducted steelhead redd surveys in both the Wenatchee and Methow rivers, and likewise the two-observer net error model was used to estimate observer error. Because there were no data points with zero redds in the original data set, we could not fit a two-stage or hurdle model to this data, which leaves us with no way to adjust for observed zeros. Therefore, the estimates of total redds will potentially be conservative, but hopefully with minimal bias. The mean net error in the Wenatchee (0.58) was larger than that of the Methow (0.48), and the CVs for net error within a reach were generally larger in the Wenatchee as well. Although there were some general differences in several covariates between the two rivers (i.e., greater redd density in the Wenatchee River and more experienced surveyors in the Methow River), the difference in net error estimates was mainly driven by larger discharge in the Wenatchee River (mean = 109 $m^3 \cdot s^{-1}$) compared with the Methow River (mean = 42 $m^3 \cdot s^{-1}$).

GAUC curves were fit to time series of counts of new redds for all major spawning reaches (Fig. 2). Surveys conducted in minor spawning areas were conducted near peak based on the GAUC spawning curve of an adjacent major spawning reach. The observed number of redds in the Methow (N = 41) and Wenatchee (N = 97) rivers were 39% and 57% of the predicted number of redds, respectively (Methow = 106, Wenatchee = 170). The estimated number of steelhead redds in the Methow and Wenatchee rivers had similar CVs of 41% and 44%, respectively.

Discussion

Using a two-observer net error model under a modified GAUC framework, steelhead redd surveys conducted on a typical weekly schedule underestimated redd abundance by 61% and 43% in the Methow and Wenatchee rivers, respectively. Given the magnitude of negative bias associated with steelhead redd counts, the status of populations based solely on raw redd count data should be reevaluated. The complex error structure associated with redd surveys (i.e., omission and commissions) did result in a small number of cases (12%) that resulted in overestimates in redd abundance. It is also noteworthy that error rates as defined in this study were based on redd counts from census surveys (i.e., every 3 days by experienced surveyors) assuming no error. The findings of the study demonstrate redd observer error is present and prevalent. Therefore, it must be recognized that both census and naïve surveyors could have missed redds (i.e., omission) and the false identification of redds or errors of commission by census surveyors was also possible, as none of the redds found during census surveys were physically disturbed to confirm the presence of eggs. It was precisely because of these potential sources of process error that census surveys were conducted every 3 days by highly experienced surveyors. Hence, process error was not eliminated, but rather minimized such that the results of the study remain valid.

High variability in redd observer efficiency was attributed to multiple factors, including river characteristics, habitat complexity, redd density, and experience conducting salmonid redd surveys. Predicting observer accuracy rates was generally more difficult and less precise than predicting observer efficiency. An intercept-only model was included in the model selection process, and although it had a ΔAIC_c near 40 for the Wenatchee data, the ΔAIC_c for the Methow data was only 2.74, suggesting the covariates that were included may not provide much information about observer accuracy. Variability in the accuracy of a redd observer was mostly attributed to surveyor effort or channel gradient and habitat complexity. It is interesting that greater effort by a single redd observer improves their ability to correctly identify redds (i.e., accuracy), but not find more redds (i.e., efficiency). Greater surveyor effort may translate into how intensively each redd-like feature was examined rather than how extensively the surveyor examined possible spawning habitat. Increases in gradient and channel complexity may simply reduce the probably of a redd-like feature being present, thereby reducing the probability of making errors of commission. There also may be human behavioral traits (e.g., dependability, independence, or routine-oriented) that would assist in predicating redd observer accuracy, but none were included in this study.

The top net error models included similar covariates as observer efficiency for each respective redd survey protocol. Both observed redd density and thalweg CV were included regardless of survey protocol. Additionally, a measure of stream condition on the day of the survey was also included (i.e., mean water depth for a single redd observer and mean discharge for two redd observers). As expected, the efficiency and net error of a single redd observer decreased with an increase in water depth presumably due to a greater difficulty in observing contrast in the substrate within the redd. Conversely, greater water depth increased the efficiency, although not significantly, when two observers were used. Because steelhead redds are generally associated with edge habitat, greater water depth may provide two observers greater access to edge habitat via rafts, thereby allowing better identification of redds. The positive effect of increasing discharge, albeit small, in the net error models further supports the hypothesis that greater access to edge habitat by each of the two observers improves their ability to identify redds.

In general terms, a single steelhead redd observer's ability to correctly identify redds is a function of their ability to observe the substrate (i.e., shallow water depth), identify probable redd locations (i.e., greater channel complexity or thalweg CV), and to a lesser degree other redds (i.e., greater redd density). Conversely, two redd observers identify more redds correctly as redd density increases, habitat is more readily observed (i.e., decreases in channel complexity or thalweg CV), as experience increases, and to a lesser extent as discharge increases.

Surveyor experience was included in the top net error models under the two redd observer protocol. For reasons previously discussed, the top net error models under two redd observer protocol was applied to redd data in 2014. The covariates included in the top models included surveyor experience, an index of stream condition on the day surveys were conducted (i.e., discharge), an index of habitat complexity of each reach (i.e., thalweg CV), and an index of spawning habitat quality (i.e., observed redd density). These covariates can be easily measured before, during, and after the spawning season, thereby increasing the application of the model to most river systems.

Between the two basins and methodologies, parameter estimates sometimes showed discordant results. Determining whether this is due to difference between the basins, the methodologies, or other unknown factors is difficult to do with certainty. For example, surveyor experience was not considered an important covariate in any single observer models, but was included in the top two observer models. This phenomenon is counterintuitive and may be an artifact of the experience of the naïve observer used in each watershed. As shown previously, the mean experience level of naïve redd observers in the Wenatchee (27.91) was lower than that in the Methow (41.48). Hence, if surveyors with a greater range of experience were used in the Wenatchee watershed, we would expect experience to be an important covariate. Ideally, a comparison of protocols would have observers with a similar range of experience. Once a two-observer protocol has been adopted in the Wenatchee Basin, we recommend gathering additional data points in that basin to include and update the observer error models. This may lead to a larger difference in the performance of the efficiency–accuracy models compared with the net error model, making it more straightforward to choose between the two approaches.

Currently, the use of either approach could be justified and would be an improvement over raw redd counts, since the predicted number of redds from either protocol was only slightly biased compared with census counts (Table 6). Although the coverage probability was less than desired, the resulting estimates are certainly more accurate than raw redd counts. Another benefit of using the GAUC methodology was quantifying the uncertainty associated with redd surveys. Millar and Jordan (2013) extended the GAUC methodology to incorporate more complex curves, but the underlying need for an estimate of observer error remains. Unbiased, precise estimates of spawner escapement are recommended for Endangered Species Act-listed salmon and steelhead populations (Crawford and Rumsey 2011). While more work is needed to better understand the relationship of steelhead redds and spawner abundance, methods of estimating fish per redd values have been suggested (see Murdoch et al. 2010).

This study was conducted in the Wenatchee and Methow rivers, in part due to a 10-year history of successful implementation of steelhead redd surveys. Successful steelhead redd surveys are defined as being able to conduct weekly surveys throughout the spawning period (10–12 weeks) by raft and observe all potential steelhead spawning substrate. While spring freshets are common, they are short in duration, and river conditions are acceptable again in only a few days. If the magnitude and variability of steelhead redd observer error reported in this study is representative of other populations, the status of steelhead populations based on redd counts are negatively biased. Trends in population status could also be biased if spawner distribution (i.e., thalweg CV) or spawn timing (i.e., water depth or discharge) has changed or the abundance is highly variable (i.e., redd density).

Estimates of adult spawners is the most important metric in assessing extinction risk (Crawford and Rumsey 2011). The results of this study suggest that steelhead redd surveys conducted with two experienced observers produced the least biased redd estimates. Application of this methodology and models can assist managers in generating unbiased and precise estimates of redd abundance to better understand the risk to populations with high conservation value. Training of redd surveyors can include suggestions of how stream flow, complexity, and redd density may influence successful surveys. Managers can also take steps to reduce bias in steelhead redds surveys by using two experienced observers and establish survey reaches that promote diligence or thoroughness of the observers and not speed or constraint by time (i.e., use as much effort as needed).

In addition, application of the GAUC methodology as described with parameters derived from typical steelhead redd survey data can be generated for little additional cost. A comparison of redd-based spawner abundance estimates (i.e., number of redds multiplied by fish per redd value) with other independent methods of estimation must be performed to better understand the usefulness of these data. However, unlike other potential methods of estimation, redd surveys also provide critical data on spawn timing and location that are of value to managers for either restoration or protection purposes.

Acknowledgements

This study was funded by the Bonneville Power Administration and Chelan and Douglas County Public Utility districts. This project benefited greatly from the experience, knowledge, and conversations with Russell Thurow and Claire McGrath. We thank the numerous staff who participated throughout the study. Special thanks go to census surveyors Alix Blake, Brad Buechel, Clint Deason, Bobby Foster, Pat Hale, Jamie Hallman, Evan Haug, Ben Goodman, Sean Kramer, Joe Lemoine, Brady Miller, Jon Nichol, Brandon Sly, Charlie Snow, Kayla Truscott, Michael Vaughan, and Matt Young for their dedication in maintaining a strict survey schedule. We thank Tim Copeland, Alf Haukenes, and George Pess for valuable comments on earlier draft manuscripts.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control, **19**(6): 716–723. doi:10.1109/TAC.1974.1100705.

Bartley, R., and Rutherfurd, I. 2005. Measuring the reach scale geomorphic diversity of streams: application to a stream disturbed by a sediment slug. River Res. Appl. **21**: 39–59. doi:10.1002/rra.813.

Boydstun, L.B., and McDonald, T. 2005. Action plan for monitoring California's coastal salmonids. Final report to NOAA Fisheries, Contract Number WASC-3-1295, Santa Cruz, California.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York.

Chasco, B.E., Ward, E.J., Hesse, J.A., Rabe, C., Kinzer, R., Vogel, J.L., and Orme, R. 2014. Evaluating the accuracy and precision of multiple abundance estimators using state-space models: a case study for a threatened population of Chinook Salmon in Johnson Creek, Idaho. N. Am. J. Fish. Manage. **34**(5): 945–954. doi:10.1080/02755947.2014.926302.

Crawford, B.A., and Rumsey, S.M. 2011. Guidance for the monitoring recovery of Pacific Northwest salmon & steelhead listed under the federal endangered species act. National Marine Fisheries Service, NW Region, Portland, Oregon.

Dauble, D.D., and Watson, D.G. 1997. Status of fall Chinook salmon populations in the mid-Columbia River, 1948–1992. N. Am. J. Fish. Manage. **17**: 283–300. doi:10.1577/1548-8675(1997)017<0283:SOFCSP>2.3.CO;2.

Dunham, J., Rieman, B., and Davis, K. 2001. Sources and magnitude of sampling error in redd counts for bull trout. N. Am. J. Fish. Manage. **21**: 343–352. doi:10.1577/1548-8675(2001)021<0343:SAMOSE>2.0.CO;2.

Gallagher, S.P., and Gallagher, C.M. 2005. Discrimination of Chinook salmon, coho salmon, and steelhead redds and evaluation of the use of redd data for estimating escapement in several unregulated streams in northern California. N. Am. J. Fish. Manage. **25**: 284–300. doi:10.1577/M04-016.1.

Gallagher, S.P., Hahn, P.K.J., and Johnson, D.H. 2007. Redd counts. *In* Salmonid field protocols handbook: techniques for assessing status and trends in salmon and trout populations. D.H. Johnson, B.M. Shier, J.S. O'Neal, J.A. Knutzen, X. Augerot, T.A. O'Neil, T.N. Pearsons, plus 37 contributing authors. American Fisheries Society, Bethesda, Maryland. pp. 197–234.

Gallagher, S.P., Adams, P.A., Wright, D.W., and Collins, B.W. 2010. Performance of spawner survey techniques at low abundance levels. N. Am. J. Fish. Manage. **30**: 1086–1097. doi:10.1577/M09-204.1.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. Model assessment and selection. *In* The elements of statistical learning. Springer, New York.

Hilborn, R., Bue, B.G., and Sharr, S. 1999. Estimating spawning escapements from periodic counts: a comparison of methods. Can. J. Fish. Aquat. Sci. **56**(5): 888–896. doi:10.1139/f99-013.

Hillman, T., Miller, M., Miller, T., Tonseth, M., Hughes, M., Murdoch, A., Miller, J., and Kessee, B. 2011. Monitoring and evaluation of the Chelan County PUD hatchery programs: 2010 annual report. Report to the HCP Hatchery Committee, Wenatchee, Wash.

Howell, P.J., and Sankovich, P.M. 2012. An evaluation of redd counts as a measure of bull trout population size and trend. N. Am. J. Fish. Manage. **32**: 1–13. doi:10.1080/02755947.2011.649192.

Jacobs, S.E., Gaeuman, W., Weeber, M.A., Gunckel, S.L., and Starcevich, S.J. 2009. Utility of a probabilistic sampling design to determine bull trout population status using redd counts in basins of the Columbia River Plateau. N. Am. J. Fish. Manage. **29**: 1590–1604. doi:10.1577/M08-210.1.

Korman, J., Ahrens, R.N.M., Higgins, P.S., and Walters, C.J. 2002. Effects of observer efficiency, arrival timing, and survey life on estimates of escapement for steelhead trout (*Oncorhynchus mykiss*) derived from repeat mark–recapture experiments. Can. J. Fish. Aquat. Sci. **59**(7): 1116–1131. doi:10.1139/f02-081.

Liermann, M.C., Rawding, D., Pess, G.R., and Glaser, B. 2015. The spatial distribution of salmon and steelhead redds and optimal sampling design. Can. J. Fish. Aquat. Sci. **72**(3): 434–446. doi:10.1139/cjfas-2014-0181.

Madej, M.A. 1999. Temporal and spatial variability in thalweg profiles of a gravel-bed river. Earth Surf. Process. Landf. **24**: 1153–1169. doi:10.1002/(SICI)1096-9837(199911)24:12<1153::AID-ESP41>3.0.CO;2-8.

Millar, R.B., and Jordan, C.E. 2013. A simple variance estimator for the trapezoidal area-under-the-curve estimator of the spawner abundance of pacific salmon. Can. J. Fish. Aquat. Sci. **70**(8): 1231–1239. doi:10.1139/cjfas-2012-0499.

Millar, R.B., McKechnie, S., and Jordan, C.E. 2012. Simple estimators of salmonid escapement and its variance using a new area-under-the-curve method. Can. J. Fish. Aquat. Sci. **69**(6): 1002–1015. doi:10.1139/f2012-034.

Muhlfeld, C.C., Taper, M.L., Staples, D.F., and Shepard, B.B. 2006. Observer error structure in bull trout redd counts in Montana streams: implications for inference on true redd numbers. Trans. Am. Fish. Soc. **135**: 643–654. doi:10.1577/T05-129.1.

Mullan, J.W., Williams, K.R., Rhodus, G., Hillman, T.W., and McIntyre, J.D. 1992. Production and habitat of salmonids in mid-Columbia River tributary streams. United States Fish and Wildlife Service Monograph 1, Leavenworth, Washington.

Murdoch, A.R., Pearsons, T.N., and Maitland, T.W. 2010. Estimating the spawning escapement of hatchery- and natural-origin spring Chinook salmon using redd and carcass data. N. Am. J. Fish. Manage. **30**: 361–375. doi:10.1577/M09-071.1.

NMFS. 2016. 2016 5-year review: summary and evaluation of Upper Columbia River Steelhead and Upper Columbia River Spring-run Chinook Salmon. National Marine Fisheries Service West Coast Region, Portland, Ore.

Parken, C.K., Bailey, R.E., and Irvine, J.R. 2003. Incorporating uncertainty into area-under-the-curve and peak count salmon escapement estimation. N. Am. J. Fish. Manage. **23**: 78–90. doi:10.1577/1548-8675(2003)023<0078:IUIAUT>2.0.CO;2.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Snow, C., Frady, C., Repp, A., Murdoch, A., Small, M.P., Bell, S., and Dean, C. 2011. Monitoring and evaluation of Wells and Methow hatchery programs in 2010. Report to the Douglas County Public Utility district and the Wells Habitat Conservation Plan Hatchery Committee, Wenatchee, Wash.

Stevens, D.L. 2002. Sampling design and statistical methods for the integrated biological and physical monitoring of Oregon streams. Oregon State University, Department of Statistics, Corvallis, Oregon, and EPA National Health and Environmental Effects Research Laboratory, Western Ecology Division, Corvallis, Oregon.

Thurow, R.F., and McGrath, C.C. 2010. Evaluating the bias and precision of Chinook Salmon redd counts in the Middle Fork Salmon River basin, Idaho. Bonneville Power Administration, Project 2002-049-00, Portland, Oregon.