Confidence Intervals for Run Composition of Returning Salmonids

Kirk Steinhorst1, Timothy Copeland2, Michael W. Ackerman3, Eric C. Anderson4, William C. Schrader2

1University of Idaho, Department of Statistical Science, 875 Perimeter Drive MS 1104, Moscow, Idaho, 83844-1104, USA

2Idaho Department of Fish and Game, Nampa Fisheries Research, 1414 East Locust Lane, Nampa, Idaho, 83686, USA

3Pacific States Marine Fisheries Commission & Idaho Department of Fish and Game, Eagle Fish Genetics Lab, 1800 Trout Road, Eagle, Idaho, 83616, USA

4Fisheries Ecology Division, Southwest Fisheries Science Center, 110 Shaffer Road, Santa Cruz, California, 95060, USA

Corresponding author: [kirk@uidaho.edu](mailto:kirk@uidaho.edu)

Keywords: Bootstrap, stratified sampling, salmonid escapement, simultaneous inference, genetic assignment test

ABSTRACT

We propose weighted estimators of fish composition for wild steelhead and Chinook salmon returning to a single point in a river system. Using data on overall fish numbers derived from 1) counts at an observation window located on a fish ladder at Lower Granite Dam, 2) trap data providing separation of wild and hatchery fish, and 3) data on sex, age, and genetic stock of wild fish, we propose one-at-a-time and simultaneous confidence intervals for sex or age or genetic stock. Further, for each genetic stock we estimate numbers for stock-sex and stock-age categories. Using two simulation, we demonstrate that estimators are generally unbiased with confidence intervals that have good coverage of the truth. We also show that if numbers returning for a given category are not too small, then we can estimate the true number returning within 10% with 90% confidence.

1. INTRODUCTION

Migratory life histories are important in fisheries management and conservation (Hilborn and Walters 1992). Given migratory characteristics, the spatial distribution of a stock or population is predictable, which facilitates exploitation and sampling for research and management. Sampling may be conducted by test fisheries (e.g., Flynn and Hilborn 2004, Beacham et al. 2012), hydroacoustics (e.g., Tarbox and Thorne 1996, Pritt et al. 2013), or various other means. A highly controlled sampling regime can be instituted by counting or sampling migrating fish as they move past barriers such as weirs or dams (e.g., Wagner 2007). In all of these sampling scenarios it is typical to estimate overall abundance and to subdivide that abundance into groups of management interest by applying compositional data (species, sex, age, stock) derived either from the primary sampling gear or a secondary sampling gear (e.g., using gill net biological samples to allocate total abundance derived from hydroacoustics to species, Rudstam et al. [2012]). The complexities of fisheries sampling programs and the relevant groups into which the fish are parsed present difficulties for estimating precision about the point estimates generated.

Here we present a case study with management and conservation applications. Populations of steelhead trout *Oncorhynchus mykiss* (hereafter steelhead) and Chinook salmon *O. tshawytscha* in the Snake River basin of Idaho, NE Oregon, and SE Washington declined substantially following the construction of hydroelectric dams in the Snake and Columbia rivers. Raymond (1988) documented a decrease in survival of juvenile Snake River steelhead and Chinook salmon emigrating to the Pacific Ocean following the construction of dams on the lower Snake River during the late 1960s and early 1970s. Survival improved and adult abundance rebounded slightly in the early 1980s, but then escapement over Lower Granite Dam into the Snake River basin declined again (Busby et al. 1996). As a result, Snake River steelhead were listed as threatened under the Endangered Species Act (ESA) in 1997; Snake River spring/summer Chinook salmon were listed as threatened in 1992. In recent years, abundances in the Snake River basin have slightly increased. However, the increase has been dominated by fish produced in hatcheries (intended to mitigate for lost fisheries and to supplement natural populations), while the returns of steelhead and Chinook salmon born in the natural environment remain low.

Fisheries biologists need to know how many wild versus hatchery steelhead and salmon return to properly manage fisheries and to assess the conservation status of wild populations. In our study, we are particularly interested in the number of wild fish returning for status and trends monitoring of ESA-listed steelhead and salmon. Further, for wild fish, we want to know the numbers of fish returning by sex, age, or genetic stock. And for each genetic stock, we want to know the numbers returning by stock-sex and stock-age; estimates of abundance by sex and age within a stock facilitate stock-specific productivity monitoring.

There are two approaches to determining run composition (Starr and Hilborn 1988): 1) Direct identification from tags, scales, genetic analysis, and/or morphology, and 2) Reconstruction of the run from catch data, adult escapement estimates (number of fish returning from the ocean to a given point), and migration paths (Prvost, Crozier, and Schn 2005, Flynn et al. 2006, Buckland et al. 2007, Chasco et al. 2007, Newman et al. 2009, Branch and Hilborn 2010). Run reconstruction uses maximum likelihood estimation of a product of probability density functions describing arrivals, escapement, harvest, and ages or it uses state-space models and Bayesian estimation via MCMC or related techniques. Run reconstruction methods have been widely applied for mixed stock fisheries in coastal fisheries although they have been applied to riverine settings as well (Eggers and Bernard 2011, Pritt et al. 2013).

McElhany et al. (2000), a document published by the National Oceanic and Atmospheric Administration (NOAA), provides guidance for status and trends monitoring of ESA-listed Pacific salmon populations based on Viable Salmonid Population (VSP) indicators. The VSP indicators include adult spawner abundance, productivity, spatial distribution, and diversity. McElhany et al. (2000) highlight the need for reliable population information to provide merit to regulatory and conservation decisions based on that information. As a result, NOAA suggests unbiased estimates with a coefficient of variation (CV) or 15% or less (Crawford and Rumsey 2011). Thus, an evaluation of accuracy and precision about estimators of run composition is warranted.

There is no statistical literature on properties of estimators and confidence intervals for composition determined from direct identification. This paper provides a statistical evaluation of estimators derived via direct identification methods. We observe adult steelhead and spring-summer Chinook salmon (hereafter Chinook salmon) as they migrate past Lower Granite Dam (LGD) on the Snake River 695 kilometers from the ocean. Adults returning from the Pacific Ocean to spawn in tributaries of the Snake River must ascend fish ladders at eight dams during their migration including four on the Columbia River and four on the Snake River. LGD is the final dam they encounter on the Snake River; fish migrating past LGD then disperse to tributaries throughout the Snake River basin to spawn. An observation window on the LGD fish ladder allows us to count fish by species as they migrate upstream. A trapping facility located above the observation window allows us to intercept fish as they migrate past to collect data and obtain biological samples. The counting and sampling of returning adult steelhead and Chinook salmon at the dam provides direct data for calculating composition.

We examine the properties of estimators of composition and their confidence intervals derived from weightings of window counts, wild versus hatchery data obtained from trap samples, and sex/age/genetic stock data collection from the biological sampling of wild fish subsampled at the trap. One-at-a-time and simultaneous confidence intervals are derived via a combination of bootstrap sampling of both the wild/hatchery trap data set and the sex/age/genetic stock data set produced by sampling wild fish. We consider counts at the observational window to be a census of fish passing the dam. Joint confidence intervals are found for sex, age, stock, stock-sex, and stock-age. In earlier work (Steinhorst et al. 2010), we found that confidence intervals for small groups of fish like those possible formed by combinations of stock-sex and stock-age were too wide to be useful. Joint confidence intervals on such groups would be even wider. We further address the dilemma posed by small groups in this paper.

It is important to note that in our study, we assume that sex and age are determined without error (described below), and that bias and precision of our estimators is determined solely by sample size and bias. However, error does exist in the determination of stock; the stock of each fish is estimated via individual assignment (IA; Koljonen et al. 2005). The frequency of errors depends on a fish’s true stock of origin and the relative genetic distinctness of that stock relative to all other stocks. A fish originating from a genetically distinct stock will have a lower probability of misallocating to another stock. For any simulations involving stock, we evaluate our estimators under two scenarios: 1) assuming stock is known (no IA performed), and 2) using the inferred stock of each fish after performing IA. Bias and precision under the ‘known’ scenario would be due to sample size and bias, bias and precision under the ‘inferred’ scenario would be due to sample size and bias plus misallocation during IA.

It is useful to have one-at-a-time and joint confidence intervals for sex or age or stock since, for the most part, numbers of wild females and males or number of wild fish by various ages or stocks are large enough to provide meaningful intervals. In some cases, there may also be enough fish in categories formed by combinations of stock-sex and stock-age to make meaningful inferences, but this will depend on sample sizes and the frequency of stock misallocations from IA.

In section 2, we describe the setting and the data collection and processing methods. In section 3, we present the estimators and methods for obtaining the bootstrap confidence intervals. In section 4, we give the results for steelhead and Chinook salmon returning to spawn in 2011. Section 5 contains a description and results of simulations examining the properties of the estimators and confidence intervals. Our sampling goal is to know the true number of returning steelhead or Chinook salmon within categories within 10% with 90% confidence. If the distributions of our estimators are reasonably symmetric then the estimate will be close to the middle of the confidence interval. Since we are 90% confident that the true value is in the confidence interval, the differences between the estimator and the truth will be less than 10% of the truth if half the confidence interval is less than 10% of the estimator.

1. DATA COLLECTION

This paper summarizes the abundance and composition of wild adult steelhead and spring/summer Chinook salmon migrating past LGD during spawn year (SY) 2011. For steelhead, the SY2011 run is operationally defined as adults migrating past LGD between 07/01/10 and 06/30/11. Most steelhead migrate past LGD during the fall although a small proportion migrate the following spring; all steelhead spawn during the spring. For spring/summer Chinook salmon, the SY2011 is operationally defined as adults migrating past LGD during the 2011 calendar year prior to 08/17/11. Spring/summer Chinook begin arriving at the dam in April. Chinook migrating past LGD after 08/17/11 are considered fall Chinook, a separate lineage of Chinook salmon occurring within the Columbia River basin.

Adult steelhead and Chinook salmon returning from the ocean to the Snake River basin must pass LGD during their spawning migration; adults enter a fish ladder bypass system to ascend past the dam. An observation window, located on the fish ladder, allows United States Army Corps of Engineers (COE) personnel to tally and record the number of fish for each species that ascend the fish ladder daily. Snake River fisheries management agencies may assume the daily fish counts are recorded without error and represent the total escapement to LGD. Live window counts are conducted throughout a majority of the year and occur from 0400 to 2000 Pacific Time. Video counts are used in lieu of window counts in November, December, and March and occur from 0600-1600. Most fish pass the window during the 10 to 16 daylight hours when counts are made (Cassinelli et al. 2012). The ladder is closed in January and February. Count data were downloaded from the COE website: <http://www.nwp.usace.army.mil/Missions/Environment/Fish/Data.aspx>. The steelhead count consists of all fish > 12” identified as *O. mykiss*. For Chinook salmon, adults (≥ 56 cm) and jacks (males that return after only one year of saltwater residency, 30 – 56 cm) are counted separately; we combined the adult and jack counts to get total counts on a daily basis. Daily counts were aggregated on a weekly basis. Early and late in the season weeks were collapsed into longer strata if few fish were passing.

The fish ladder bypass system also contains a trapping facility, located above the observation window, designed to intercept adult steelhead and salmon at a predetermined sampling rate (Harmon 2003; Ogden 2010, 2011; Schrader et al. 2012, 2013). Trap sampling rates are determined by a committee of co-managers in an attempt to achieve sample requirements for multiple projects and to balance fish handling concerns; sample rates are typically 10 – 20%. The sample rate determines how long a trap gate remains open four times per hour; the trap is operational 24 hours per day. Additional details on the adult trap can be found in Harmon (2003) and Steinhorst et al. (2010).

Trapped fish are anesthetized and examined to determine whether they are of hatchery or wild origin. All wild fish have an intact adipose fin (a small, fleshy fin on the back posterior of the dorsal fin and anterior of the caudal fin) because they spend their entire life-cycle in the natural environment. Most hatchery origin fish have a clipped adipose fin so they can be identified for mark selective fisheries (i.e. recreational anglers may only keep fish with a clipped adipose fin). However, some hatchery fish may be released with an intact adipose fin to supplement natural populations or to bolster tribal harvest opportunities. Thus, fish with an intact adipose fin are further examined for a coded wire tag (CWT) or a passive integrated transponder (PIT) tag that were implanted into fish at the hatchery as juveniles. For adipose intact steelhead, hatchery origin may also be determined by the presence of dorsal and/or ventral fin erosion, which is assumed to occur only in hatchery-reared steelhead (Latremouille 2003). We also used the presence of a ventral fin clip to determine if an unclipped fish was of hatchery origin. For unclipped Chinook salmon, hatchery origin was determined solely by the presence of a CWT or ventral fin clip.

Scale and tissue samples are taken from trapped fish determined to be wild. Scale samples are used to determine age based on visual examination of scale annuli (Schrader et al. 2013). Scale samples were taken from above the lateral line and posterior to the dorsal fin. Samples were stored in coin envelopes for transport to the IDFG aging laboratory in Nampa, Idaho. Scales were mounted between two glass microscope slides and two technicians independently viewed each scale to assign ages without reference to fish length. If there was no age consensus among readers, a third reader viewed the image and all readers collectively examined the image to resolve their differences before a final age was assigned. If a consensus age was not attained, the sample was excluded from further age-specific analyses. Readers estimate both freshwater-age (time spent in freshwater prior to entering the ocean as juveniles) and saltwater-age (time spent in the ocean prior to upriver spawning migration as adults) for each fish. Ocean-age determination for steelhead is 98.8% accurate and for Chinook salmon is 98.1% accurate based on comparisons with known-age individuals. Currently, freshwater-age determination accuracy cannot be validated, but precision estimates based on the coefficient of variation between experienced readers were 11.9% for steelhead and 3.3% for Chinook (Schrader et al. 2013). Age data collected at LGD are used to assign returning adults back to specific brood years.

Tissue samples for DNA extraction and single nucleotide polymorphism (SNP) genotyping were taken from a small clip of the anal fin. Tissues were stored in a vial with 200-proof nondenatured ethyl alcohol for transport to the IDFG genetics laboratory in Eagle, Idaho. Tissue samples are used to determine sex and estimate genetic stock of origin. Sex determination is not based on phenotypic characteristics observed at the LGD adult trap; rather sex is determined *post hoc* using a sex-specific allelic discrimination assay (Campbell et al. 2012) that is screened during genetic analysis (described below). The sex-determination assay for steelhead is 99.3% accurate and for Chinook salmon is 99.0% accurate based on comparisons with known-sex individuals (C. Steele and J. McCane, Pacific States Marine Fisheries Commission, pers. comm.).

The genetic stock of each fish is estimated using IA, a type of genetic stock identification (GSI; Pella and Milner 1987, Shaklee et al. 1999), using multi-locus SNP genotype data. IA allocates each fish to the population and genetic stock in which the probability of its genotype occurring is the greatest. Two sets of data are required for IA: 1) allele frequency estimates from multi-locus genotypes obtained from samples taken from all potentially contributing stocks (the baseline), and 2) corresponding multi-locus genotype data from unknown origin individuals whose stock is to be estimated (Koljonen et al. 2005). Ackerman et al. (2014) present the Snake River SNP baselines for both steelhead and Chinook salmon used for this study; unknown stock fish trapped at LGD were genotyped using the same SNP loci presented in that study. Each individual’s population was estimated and then summed into genetic stock estimates (allocate-sum procedure; Wood et al. 1987). Genetic stocks (reporting groups) are assemblages of reference (baseline) populations grouped primarily by genetic and geographic similarities and secondarily by political boundaries and/or management units (Ackerman et al. 2011). IA was performed using the Bayesian version of the program gsi\_sim (Anderson et al. 2008, Anderson 2010). The Bayesian gsi\_sim uses Markov chain Monte Carlo (MCMC) to computer posterior probabilities of stock membership conditional on the allele frequencies estimated from the baseline. The likelihood that a fish originates from a stock is computed using the compound Dirichlet-multinomial formulation of Rannala and Mountain (1997) conditional on the baseline samples and these likelihoods remain fixed throughout the MCMC simulation. To perform the MCMC, gsi\_sim uses a Gibbs sampler (Casella and George 1992) in which alternately: 1) the stock assignments of the fish in the mixture are updated as a multinomial draw from their posterior probabilities given the current estimate of the stock proportions and the stock-likelihoods of the fish; and 2) the stock proportions are updated as a draw from a Dirichlet distribution given a unit-information prior and the current values of the stock assignments of all the fish in the mixture.

Ackerman et al. (2014) defined the genetic stocks used for IA at LGD. Ten genetic stocks were defined for steelhead. Genetic stocks include: 1) UPSALM: the upper Salmon River; 2) MFSALM: Middle Fork Salmon River (including Chamberlain and Bargamin creeks); 3) SFSALM: South Fork Salmon River; 4) LOSALM: lower Salmon River; 5) UPCLWR: upper Clearwater River (Lochsa and Selway rivers); 6) SFCLWR: South Fork Clearwater River (including Clear Creek); 7) LOCLWR: lower Clearwater River; 8) IMNAHA: Imnaha River; 9) GRROND: Grande Ronde River; and 10) LSNAKE: Tucannon River, Asotin Creek and tributaries to the Snake River downstream of the Clearwater River confluence.

Seven genetic stocks were defined for Chinook salmon. Genetic stocks include: 1) UPSALM: upper Salmon River; 2) MFSALM: Middle Fork Salmon River; 3) CHMBLN: Chamberlain Creek; 4) SFSALM: South Fork Salmon River; 5) HELLSC: an aggregate stock that includes the Little Salmon, Clearwater, Grande Ronde, and Imnaha rivers and small tributaries of the Lower Snake River; 6) TUCANO: Tucannon River, and 7) FALL: Snake River fall Chinook salmon. The FALL genetic stock represents a separate lineage of Chinook salmon in the Snake River basin and is managed as a separate species under the ESA. Fall Chinook are genetically highly differentiated from spring/summer Chinook salmon and can be identified using IA with 100% accuracy (Ackerman et al. 2014). Genomic DNA extraction and amplification and SNP genotyping using multiplex 5’-nuclease reactions follow methods described in Seeb et al. (2009). DNA was extracted using nexttecTM Genomic DNA Isolation Kits from XpressBio (Thurmont, Maryland) or QIAGEN DNeasy Tissue Kits (Valencia, California). Ackerman et al. (2014) fully describes DNA amplification and SNP genotyping protocols; SNP genotyping was performed using the Fluidigm EP1TM platform. All adults queued for genotyping were screened at 191 SNPs and the sex-specific allelic discrimination assay (Campbell et al. 2012). Samples were processed at either the IDFG genetics laboratory in Eagle, Idaho, or the Columbia River Inter-Tribal Fish Commission genetics laboratory in Hagerman, Idaho.

The above sampling design produces three datasets: 1) a census of fish numbers returning to and migrating past the dam (window counts), 2) a hatchery versus wild dataset for all trapped fish, and 3) a dataset containing sex, age, and stock for a sample of wild fish that were trapped. These three data sets are used to produce the size of the run and estimates and confidence intervals for number of fish by sex or age or genetic stock, and further, combinations of stock-sex and stock-age.

1. ESTIMATES AND CONFIDENCE INTERVALS

The proportions of wild and hatchery steelhead or Chinook salmon changes over the season, but within a week are relatively constant. Given weekly counts* ,* the number of wild salmon or steelhead is estimated as



where are estimates of proportion wild by week from the trap data. Since the window counts are assumed to be fixed, we find a confidence interval for the number of wild fish using a parametric bootstrap to produce B bootstrap sets of  yielding . That is, for week *h* we assume the bootstrap number of wild fish has a binomial distribution,  where  is the number of fish trapped for that week and . The and percentiles of give us the  confidence interval for the true number of wild steelhead or Chinook passing Lower Granite Dam for the year.

Define . Then we have as the conditional probabilities for wild females and males for weeks 1,…,*n.* Similarly for A ages, we have conditional probabilities .For g stocks, we have conditional probabilities . Given estimates of these probabilities from wild fish examined in the trap, we obtain





##### and



and



That is, the number of wild females or males is a weighted sum of the weekly window counts where the weights are estimates of the probability of being wild and female (or male). Likewise for ages 1,…,*A* or genetic stocks 1, . . . , *g ,* the estimate is a weighted sum of the weekly counts with weights that are an estimate of the probability of being wild of that age or genetic stock.

To find confidence intervals for these estimates we have to account for the variability of both the trap data and the wild fish sample data. We do this by extending the parametric bootstrap process used above for obtaining the confidence interval for number wild by adding a conditional parametric bootstrap based on sex, age, and genetic stock data from wild fish in the trap.

For sex define pseudoreplicates using



where is the number of wild fish sexed. Bootstrap values for the number of wild females,  follow from  *.* The percentile bootstrap confidence interval for true number of wild females,  is given by finding the and  percentiles. Similarly, we find a bootstrap confidence interval for the number of wild males. Changing the binomial above to a multinomial, we generate B sets of  to obtain bootstrap confidence intervals for the true number of wild fish of ages 1,...,A. We follow the same procedure for genetic stocks 1,... ,g.

Simultaneous confidence intervals for the number of females and males is found by expanding the rectangle formed by  and  until  of all of the pairs are inside the expanded rectangle (REF here). Likewise simultaneous confidence intervals for ages are found by expanding the A dimensional hypercube formed from the one-at-a-time intervals until  of the B A-tuples, , are inside of the expanded hypercube. The same expansion leads to simultaneous confidence intervals for genetic stock.

1. RESULTS FOR SPAWN YEAR 2011

We consider two empirical examples to illustrate the methods outlined in section 3 – steelhead and Chinook salmon returning to spawn in 2011. This provides two differing examples; there were over 200,000 steelhead returning to LGD and slightly less than 135,000 Chinook salmon. A total of 208,296 steelhead returned during SY2011. Of these, we estimate that 44,452 were wild (21.3%, Table 1). The 90% confidence interval is (43454,45473). We found 90% one-at-a-time and simultaneous confidence intervals on the numbers of each sex, on numbers by age (brood year), and numbers by genetic stock. The simultaneous confidence intervals were found separately for sex, age, and stock. In addition to female and male, there are five ages, and ten genetic stocks for steelhead. The simultaneous confidence intervals are proportionally wider for age and stock than for sex. Columns P1 and Psim give half the width of the confidence interval as a percentage of the estimate. If the estimate is near the center of the confidence interval then we are 90% confident that we know the truth within this percentage of the estimate. Confidence intervals for total wild, sex, BY05, BY06, BY07, GRROND, LSNAKE, and UPSALM meet our 10% goal for confidence interval width. BY04 and BY08 are not well estimated. The other genetic stocks are reasonably well estimated; the percent half widths fall in the range 10.4 to 15.2%.

[Table 1 about here]

A total of 134,594 Chinook were counted at the observation window at the dam. Of those, 26,673 were estimated to be wild (19.8%, Table 2). Based on B = 5,000 bootstrap iterations, we obtained a 90% confidence interval of (25937, 27400). In addition to female and male, there are five ages, and seven genetic stocks. Note that one of the genetic stocks identified is FALL Chinook (4.2%) which are managed as a different species under the ESA. If more than 1,129 spring/summer Chinook arrived after August 18th, then our window count is an underestimate of spring/summer Chinook salmon escapement and vice-versa. For total wild, the truth should be within 2.7% of the estimate. When the groups are quite small like BY05, BY09, and the TUCANO stock, then the truth might be as much as 35-50% away from the estimate. Note that the TUCANO stock (Tucannon River) is actually located below LGD; the TUCANO estimate represents fish that strayed above LGD or ascended LGD and fell back to the Tucannon River later. If you compare the P1 and Psim columns, you see that the widths of the simultaneous confidence intervals for female and male are not markedly wider than the one-at-a-time intervals. For age and genetic stock the simultaneous confidence intervals can be double the width of the one-at-a-time confidence intervals. Confidence intervals for total wild, sex, BY06, BY07, BY08, and 4 of 7 genetic stocks (UPSALM, MFSALM, SFSALM, HELLSC) meet our 10% goal for confidence interval width. Categories that don’t achieve our 10% goal were categories with a decreased abundance of 1,129 or less.

[Table 2 about here]

For steelhead, there were nearly twice as many females as males. For Chinook salmon, the reverse was true. These sex ratios vary annually, but the ratios seen for both species in 2011 are typical. For both steelhead and Chinook the middle age groups had more fish returning than the younger and older age groups. Steelhead genetic stocks are fairly evenly represented, although the GRROND stock has almost four times the number as the SFSALM stock (Table 1). Note that the LSNAKE stock includes some populations below LGD (Tucannon River and small tributaries to the lower Snake River). For Chinook, the HELLSC stock (Little Salmon, Clearwater, Grande Ronde, and Imnaha rivers and small tributaries of the Lower Snake River) was more than twice as abundant as any other stock. For stocks within the Salmon River, CHMBLN produced 616 spring/summer Chinook summer and the remaining 3 (UPSALM, MFSALM, SFSALM) all produced between 3,800 – 4,800 (Table 1).

1. SIMULATIONS

We used a parametric bootstrap for this problem because the variables in both the trap dataset and the wild dataset all follow binomial or multinomial distributions. Using the observed weekly proportions, we generated bootstrap proportions for both the hatchery/wild and categories of wild (sex, age, stock, stock-sex, stock-age) weekly proportions. We then combined these bootstrap samples with the window counts to produce estimates and confidence intervals. While these estimators and confidence intervals are intuitive, we do not know their statistical properties.

We set up two simulations; one similar to the steelhead and one similar to the Chinook salmon empirical scenarios presented above. We established parameter values for all binomial and multinomial distributions that are similar to weekly estimates obtained from the SY2011 datasets. We also set trap sample rates (PTrp) and wild subsampling rates (PWhandled) similar to those used in practice at LGD.

We simulated 500 data sets as follows. Simulate the number of fish trapped each week by generatingbinomial samples for each time stratum (weeks) with the number of binomial trials equal to population size SimPop and probabilities PTrp. For our simulation there were 27 time stratum for steelhead and 14 for Chinook salmon. Simulate the number of trapped fish that are wild for each stratum by generating binomial samples with the number of trials equal to the simulated number of trapped fish and with probabilities equal to PWild. The remaining trapped fish are hatchery. From these numbers, we can generate a simulated data set for trapped fish with 2 columns: time stratum and wild/hatchery. The length of this data set is the sum of the numbers of wild fish trapped across time stratum .

We can now calculate the random number of wild fish sexed, aged, and genotyped to estimate stock for each week by multiplying the simulated number of wild fish trapped by PWhandled. These numbers by stratum are the number of binomial or multinomial trials for sex or age or genetic stock, *mh*. For example, we can find the random number of wild/handled fish that are female by generating binomial trials of size *mh* with probabilities Pfemale. The remainder each week are male. Knowing the random number of wild females and males each week, we can put together the simulated data set for sexed fish, say, by forming a data set with columns equal to stratum and female/male. The length of this data set is equal to the sum across time strata of the numbers of handled fish, .

Given the numbers of fish arriving at the dam each week, some percent of them will be trapped. For steelhead, the trapping rate PTrp ranges from 3 – 14% (Table 3); for Chinook salmon PTrp ranges from 7 – 11% (Table 6). The trapping data set will have about 20,000 lines for steelhead (10% of 200,000) and about 13,500 lines for Chinook salmon (10% of 135,000). Depending on the week (time stratum) 10 to 55% percent of these will be wild (Pwild) and approximately 50% of wild steelhead and 75% of wild Chinook salmon will be handled (PWhandled) so the sex or age or genetic stock data set will be approximately 2,500 long for both species.

Note that the simulated genetic stock data represents the ‘known’ genetic stock of each fish and assumes that genetic stock can be determined via IA without error. However, for each simulated known stock fish we want to simulate SNP genotype data using allele frequencies from genetic baselines presented in Ackerman et al. (2014) and then estimate and ‘inferred’ genetic stock of each fish using gsi\_sim. Doing so allows us to evaluate both uncertainty in trap sampling and uncertainty in genetic assignments. To do this…

Now we have three random data sets: 1) random numbers of fish arriving each week, 2) a randomly generated trap data set with a random number of wild and hatchery fish, and 3) a randomly generated wild fish data set with randomly generated numbers of female/male or fish of various ages or fish of various genetic stocks including both the known and inferred genetic stock of each fish. We use the window counts and the trap and wild data sets to obtain estimates and confidence intervals for that iteration. We save the estimates and confidence intervals for subsequent analysis. We used B = 500 bootstrap iterations for finding the confidence intervals. After 500 simulation iterations, we have 500 estimates of total wild, female/male wild, wild by age, and wild by both known and inferred genetic stock. Within each genetic stock, we are also interested in combinations of stock-sex and stock-age. We also have 500 one-at-a-time and simultaneous confidence intervals for each estimate. Since we know the true number of wild fish and numbers of females/males and fish of different ages and different stocks, we compute the bias as the mean of the estimates minus the truth. We also compute variances of the estimates and the standard errors.

We compute the coverage of any confidence interval by tallying the number of times the true population number falls inside the confidence interval. We get the simultaneous coverage by tallying the number of times the true population numbers fall within the expanded hypercube. For example, for steelhead we know the true number of females is 28,972 and the true number of males is 14,784. The joint coverage is the proportion of times that the simultaneous female confidence interval contains 28,972 AND the simultaneous male confidence interval contains 18,375.

Our original goal was to find one-at-a-time and simultaneous confidence intervals for major group of fish including sex, age, and genetic stock. But we are also interested in the numbers of fish by sex and age within each stock (i.e. stock-sex and stock-age combinations) and their confidence intervals. For steelhead there are potentially 20 stock-sex and 50 stock-age combinations. For Chinook salmon there are potentially 14 stock-sex and 35 stock-age combinations. Some of these groups will have a fair number of fish and finding confidence intervals for them should be reasonable. However, some groups have so few fish it would be futile to try to find confidence intervals and attempting to find simultaneous confidence intervals for a large number of groups would be a fool’s errand. In addition to the sex, age, and stock simulations reported above, we also simulated estimates and one-at-a-time confidence intervals for stock-sex and stock-age combinations for both steelhead and Chinook salmon. For this simulation, we assume that sex, age and stock are independent so that we can multiply the probabilities in Tables 3, 4, 6, and 7 to get the joint probabilities for the true number of fish.

*Steelhead*

Input parameters for the steelhead simulation are in Table 3 with genetic stock proportions shown separately in Table 4. For the steelhead simulation, we set the total adult escapement to LGD as 200,000. Across stratum, the proportion of fish trapped (PTrp) ranged from 4 to 14%. For our known population, 43,756 (22%) of the total escapement was wild. The proportion of wild fish by stratum (Pwild) ranged from 20 to 50%. And of the wild fish intercepted at the trap, 35 to 100% are handled (PWhandled).

[Table 3 about here.]

The numbers of female fish by stratum are found by multiplying the number wild (PopWild) column by the proportion of females (Pfemale) for that stratum. Summing the number of females across strata gives us the true number of females. Of the wild steelhead, 28,972 were female (66%). The remaining fish were males. The numbers of fish of each age (brood year) were found by multiplying the number wild (PopWild) column by the corresponding age column proportions (BY04…BY08). Summing those numbers by age over strata (weeks) gives the true numbers of fish of each age. True numbers by age were: BY04 = 911 (2%), BY05 = 10,695 (24%), BY06 = 20,842 (48%), BY07 = 10,302 (24%), BY08 = 1,007 (2%). Similarly, the numbers of wild fish of each genetic stock by stratum were found by multiplying the number of wild (PopWild) by the corresponding genetic stock proportion columns and summing over strata gives the true numbers of fish of each genetic stock. True numbers by genetic stock were: UPSALM = 6,624 (15.1%), MFSALM = 3,872 (8.8%), SFSALM = 2,324 (5.3%), LOSALM = 1,551 (3.5%), UPCLWR = 4,123 (9.4%), SFCLWR = 4,315 (9.9%), LOCLWR = 1,779 (4.1%), IMNAHA = 2,455 (5.6%), GRROND = 7,106 (16.2%), LSNAKE = 9,607 (22.0%).

[Table 4 about here]

For the steelhead simulation, we find that the estimates of total wild, number of females, and number of males are essentially unbiased (Table 5). The coverage of one-at-a-time confidence intervals for females and males were 89.4% and 88%, respectively. The joint coverage for females and males was 90.2%. The expected widths for the joint confidence intervals for females and males were 1.3 and 1.2 times the expected widths for the one-at-a-time confidence intervals. The one-at-a-time coverages for BY04, BY05, BY06, BY07, and BY08 were all very near the nominal 90%. The expected half widths as a percentage of the truth were less than 10% for BY05, BY06, and BY07, but were 24.5% and 23.8% for BY04 and BY08, respectively. The joint coverage for ages was 90%. The expected widths of the simultaneous confidence intervals for ages were 1.2 to 2.2 times the expected widths of the one-at-a-time confidence intervals. We do a very good job of estimating numbers of females and males and a modestly good job of estimating numbers by age.

[Table 5 about here]

The true numbers of steelhead by stock are shown in Figure 1 along with the percent bias and confidence interval coverage of estimates for each stock. Results are shown for both scenarios in which 1) we assume stock is known without error (Known) and 2) where we perform IA to infer the stock of each fish (Inferred). Again, the Known results reflect uncertainty due to sample size and sampling bias; the Inferred results reflect uncertainty in the Known simulations PLUS the uncertainty due to misallocation or error from IA.

When we assume that genetic stock is known without error, we find that estimates of numbers by genetic stock are largely unbiased; the largest bias is of the IMNAHA stock (+2.6%). We only achieve our goal of 90% one-at-a-time coverage for one genetic stock, LOSALM; however, all stocks exceed a one-at-a-time coverage of 85%. The expected half widths as a percentage of the truth for the three most abundant genetic stocks (UPSALM, GRROND, LSNAKE) are less than 10%. The percent half-width for the least abundant stock, LOSALM, is 18.9%.

Results differ when we simulate SNP genotype data for each simulated fish and infer their genetic stock (Inferred simulation). Eight of the genetic stocks are biased by <10%; however, LOSALM is biased by -19.4% and LSNAKE is biased by 11.5%. None of the genetic stocks achieve our goal of 90% one-at-a-time coverage, but five stocks (MFSALM, SFSALM, UPCLWR, SFCLWR, and IMNAHA) have a one-at-a-time coverage that exceeds 77% and four of those exceed 84%. Again, the expected half widths as a percentage of the truth are still less than 10% for the three most abundant stocks. Stock-specific differences between the ‘known’ and ‘inferred’ simulation results reflect the additional uncertainty due to misallocation during IA; misallocation is a related to the true genetic stock origin of each fish and the genetic distinction of that stock relative to all other stocks present in the analysis (Ackerman et al. 2014, Anderson et al. 2008).

[Figure 1 about here]

The true numbers of steelhead by stock-sex and stock-age combinations are shown in Figures 2 and 3, respectively. Figures 2 and 3 also present the percent bias and confidence interval coverage of estimates for each combination present in the simulations.

For the Known simulations, we find that estimates of numbers for stock-sex combinations are largely unbiased (Figure 2). The largest percent bias occurs among estimates of GRROND males (-4.5%). We achieve greater than 90% one-at-a-time confidence interval coverage for 6 of the 20 stock-sex combinations (UPSALM..F, MFSALM..M, SFSALM..F, LOSALM..F, UPCLWR..F, LSNAKE..M); all stock-sex combinations have >80% one-at-a-time coverage. The expected half widths as a percentage of the truth is less than 10% for only the most abundant stock-sex combination (GRROND..F). The largest expected half-width occurs for the least abundant category (SFSALM..M, 30.7%). When we simulate SNP genotype data for each fish and infer genetic stock from IA, we generally observe increased bias and decreased precision in estimates. Fourteen of the 20 stock-sex combinations have a percent bias of <10%. The six stock-sex combinations whose bias exceeds 10% all occur for both males and females for the LOSALM, GRROND, and LSNAKE stocks which exhibit decreased genetic structure relative to other stocks in the Snake River (Ackerman et al. 2014, Matala et al. 2014). None of the stock-sex combinations achieve our goal of 90% one-at-a-time coverage, but 11 of the 20 have >80% coverage (UPSALM..M, MFSALM..F, MFSALM..M, SFSALM..F, SFSALM..M, UPCLWR..F, UPCLWR..M, SFCLWR..F, SFCLWR..M, IMNAHA..F, IMNAHA..M). Similar to the Known simulation, GRROND females is the only combination where the expected half width as a percentage of the truth is <10%.

[Figure 2 about here]

Estimates of stock-age combinations for the Known simulations are also largely unbiased. Note that 5 stock-age combinations were not present in our SY2011 empirical dataset, and thus, only 45 (10 stocks x 5 ages minus 5stock-age combinations not present) stock-age combinations are represented in the simulation. Percent bias among estimates for 43 of the 45 stock-age combinations were <10% and were <5% for 38 of 45 combinations (Figure 3). Only 4 stock-age (LOSALM..BY05, LOSALM..BY06, SFCLWR..BY05, LSNAKE..BY06) combinations achieve the nominal 90% one-at-a-time coverage, but 39 combinations were >80%. No stock-age categories achieved our goal of <10% expected half-width as a percentage of the truth.

As expected, the Inferred stock-age simulation incorporating uncertainty in IA demonstrates increased bias and variance among estimates (Figure 3). Percent bias among estimates for 26 of the 45 stock combinations were less than 10% and were less than5% for 17 combinations. Percent bias exceeded 20% for 12 stock-age combinations. No stock-age combinations achieved the nominal 90% one-at-a-time coverage, but 23 combinations achieved a one-at-a-time coverage greater than 80%. Similar to the Known simulation, no stocks achieved our goal of less than 10% expected half-width as a percentage of the truth.

[Figure 3 about here]

*Chinook salmon*

Input parameters for the Chinook salmon simulation are in Table 6 with genetic stock proportions shown separately in Table 7. For the Chinook salmon simulation, we set the total adult escapement at LGD as 135,000. Across stratum, the proportion of fish trapped (PTrp) ranged from 7 to 11%. For our known population, 27,243 (20.2%) of the total escapement was wild. The proportion of wild fish by stratum (Pwild) ranged from 10 to 55%. And of the wild fish intercepted at the trap, 68 to 76% are handled (PWhandled).

[Table 6 about here]

Of the wild Chinook salmon, the true number of females was 8,867 (32.5%). True numbers by age were: BY05 = 184 (0.7%), BY06 = 8,271 (30.4%), BY07 = 14,660 (53.8%), BY08 = 3,975 (14.6%), BY09 = 153 (0.6%). True numbers by genetic stock were: UPSALM = 4,833 (17.7%), MFSALM = 4,519 (16.6%), CHMBLN = 623 (2.3%), SFSALM = 3,944 (14.5%), HELLSC = 11,937 (43.8%), TUCANO = 272 (1.0%), FALL = 1,114 (4.1%).

[Table 7 about here]

Results for Chinook salmon were similar to results from the steelhead simulation. For the Chinook salmon simulation, we find that the estimates of total wild, number of females, and number of males are essentially unbiased (Table 8). The coverages of the one-at-a-time confidence intervals are very close to the nominal 90%. For females and males, our goal of estimating true numbers within 10% with 90% confidence was met (column P1). The expected half widths as a percentage of the truth were 5.8% and 3.6%, respectively. The coverage of the joint confidence intervals for numbers of females and males was 93%. For ages, there was some bias in estimating the least abundant ages BY05 (3.50%) and BY09 (-1.97%). For the three most abundant ages (BY06, BY07, BY08) our goal of estimating true numbers within 10% with 90% confidence was met; the half widths as a percentage of the truth were 8.9% or smaller. However, P1 for BY08 was 44.0% and for BY09 was 45.7% and coverage of the one-at-a-time confidence interval for BY09 was slightly less than nominal (85.6%). The joint coverage for the age simultaneous confidence intervals was slightly below nominal (87%). The simultaneous confidence intervals were 1.2 to 3.3 times the corresponding one-at-a-time confidence intervals.

[Table 8 about here]

The true numbers of Chinook salmon by genetic stock are shown in Figure 1 with the percent bias among estimates and the confidence interval coverage for each stock. Similar to steelhead, results are shown for both Known and Inferred simulations. When we assume that genetic stock is known, we find that estimates of numbers by genetic stock are unbiased; PctBias for all 7 stocks are less than 1%. Despite none of the genetic stocks achieving the nominal 90% one-at-a-time coverage, all stocks exceeded 88.5%. The expected half-widths as a percentage of the truth for the four most abundant stocks (UPSALM, MFSALM, SFSALM, HELLSC) were all less than 10%. The percent half-width for the least abundant stock, TUCANO, was 36.2%.

Results from the Inferred stock simulation do not differ substantially from the Known simulation. Estimates of number by stock for the Inferred simulation are also largely unbiased; PctBias for TUCANO was -4.6%, all other stock were less than 2.5%. The FALL genetic stock achieved the nominal 90% one-at-a-time coverage. Two genetic stocks had one-at-a-time coverage less than 80% (UPSALM = 78.2%, SFSALM = 76.4%), but all others exceeded 80%. Similar to the Known simulation, the expected-half widths as a percentage of the truth for the four most abundant stocks were all less than 10%.

[Figure 4 about here]

The true numbers of Chinook salmon by stock-sex and stock-age combinations are shown in Figures 5 and 6, respectively. Figures 5 and 6 also show the percent bias among estimates and one-at-a-time coverage for each combination present. Estimates of numbers for stock-sex combinations in the Known simulation are unbiased; the greatest bias is for CHMBLN females (-2.0%). We achieve greater than 90% one-at-a-time coverage for 6 of the 14 stock-sex combinations (UPSALM..F, MFSALM..F, CHMBLN..M, HELLSC..M, FALL..F, FALL..M) and the lowest coverage (UPSALM..M = 85.2%) was still reasonable. Results are similar for most combinations when we simulate SNP genotype data and infer the genetic stock of simulated fish. The PctBias is less than 5% for 19 of the 20 stock-sex combinations and less than 2% for 16 combinations. TUCANO females exhibit the largest bias (-7.5%). For both the Known and Inferred simulations, the expected half-widths as a percentage of the truth was less than 10% for three stock-sex combinations (UPSALM..M, HELLS..F, HELLSC..M).

[Figure 5 about here]

Estimates of stock-age combinations for the Known simulations are also largely unbiased. Note that 11 stock-age combinations were not present in the SY2011 Chinook salmon empirical dataset, and thus, only 24 (7 stocks x 5 ages minus 11 stock-ages not present) stock-age combinations were present in the simulation. The largest bias occurred for estimates of FALL..BY09 (-2.3%); PctBias for 17 of the 24 stock-age combinations were less than 1%. Six stock-age combinations (UPSALM..BY06, UPSALM..BY08, MFSALM..BY07, MFSALM..BY08, SFSALM..BY08, HELLSC..BY08) achieved the nominal 90% one-at-a-time coverage and 23 of 24 combinations had greater than 80% coverage.

Estimates from the Inferred simulations, in terms of PctBias and CI coverage, also performed reasonably well. All 24 stock-age combinations had a PctBias of less than 10%; 23 of those had a PctBias less than 5%. Only one stock-age combination (FALL..BY07) achieved the nomimal 90% one-at-a-time confidence interval coverage, but 23 of the 24 stock-age combinations had coverage greater than 80%. For both the Known and Inferred simulations, only the most abundant combination, HELLSC..BY07, achieved our goal of less than 10% expected half-width as a percentage of the truth.

[Figure 6 about here]

1. DISCUSSION

Steinhorst et al. (2010) recommend that group sizes should be at least 300 to be confident that you can estimate the true numbers within 10% with 90% confidence. In the tables above, we see that if the group size is 100 or less the coverage may be low and the expected confidence interval wide. In this study, we also found that coverage could be low and/or expected confidence intervals could be wide when a group is smaller than other groups in the set. See Table 5 where female coverage is low and brood years BY04 and BY08 have low coverage and wide confidence intervals even though all groups are bigger than 100.

Our goal of knowing the true group size within 10% with 90% confidence may be too restrictive for our biological application. If there are only a few fish in a category, we would be happy with a more lenient rule. For example, if our estimate is 50 and the confidence interval is (20, 80) then the percent half widths is 60% but knowing that the true number is between 20 and 80 might be sufficient.

When the numbers for a group are relatively small, then we worry about the distribution of the pseudovalues being skewed. If so, the estimate may not be near the middle of the confidence interval. For the Chinook salmon data set, the estimate of number of fish from BY05 is 148. We captured 5,000 pseudovalues for this case and looked at the distribution (Figure 3). Even though this group (and BY09) are much smaller than the other groups, the distribution of pseudovalues is pretty symmetric. The midpoint of the confidence interval (68, 240) is 154. The problem, then, is not asymmetry, but rather coverage and/or confidence interval width.

[Figure X about here]

Overall, our estimators and confidence intervals have good properties for major groups of fish. When the absolute number of fish in a group is low or relatively low compared to other groups, then there may be a problem with coverage and/or confidence interval width. Our joint confidence intervals for females and males are quite good. In our simulations, age and stock had confidence interval coverages of 0.856 or higher, but in some cases the confidence intervals were wide. When we try to find confidence intervals for combinations of sex, age, and genetic stock, the results are mixed depending on the size of the groups formed.

Our results have implications for how fish populations are monitored. If the interest is on the relative composition of the total (Tim, I am lost here), then most sampling programs will yield sufficient results. However, weak stocks are frequently a problem in conservation and fisheries management (CITE). We should make some positive statements about how to sample and to analyze monitoring data. Obviously there is a tradeoff between sample size and number of subdivisions that can be supported. Previous work (Gerritsen) shows this but the criteria for success is on overall (average) coverage. Thompson 1987 found that a sample size of 510 should suffice under a worst case scenario for alpha=0.05 (even proportions among groups but number of groups doesn’t matter). Our results show something else. What can we recommend?

In our scenario, subsample size is a function of 1) total fish escapement, 2) trap sample rate, and 3) handling rate. Total fish escapement is out of our hands, but sample size can effectively be increase by increasing the trap sample rate or the handling rate. For SY2011 steelhead, the trap rate averaged 10% and the handling rate was slightly over 50%. For Chinook salmon, the trap rate also averaged 10% and the handling rate was about 75%. We reran both the steelhead and Chinook salmon simulations increasing trap rates to 20% and the handling rate to 100%. This doubles the number of fish trapped in both cases. For steelhead, the number of fish for which sex, age, and stock was determined is 4 times the original simulation. For Chinook salmon, the number for which sex, age, and stock was determined is 2 2/3 times the original simulation. In practice, it is possible that this level of sampling would swamp the trapping crew (which can generally handle up to 800 fish a day due to logistical constraints), but we wanted to examine the effects of increasing sample sizes to this level. For both species, the percent half widths relative to the true population sizes were reduced in proportion to the reciprocal of the square root of the increase in sample size: = 0.5 for steelhead and = 0.61 for Chinook salmon. That is, the steelhead confidence intervals were half their original width and the Chinook salmon confidence intervals were 61% their original size. When the increased sampling leads to sample sizes of 300 or more, we meet our sampling goals. In other cases, sample sizes remain too small.

Estimates of abundance by sex and by age within a stock of fish allow biologists to generate brood tables for given stocks. Brood tables summarize the number of offspring that return as adults (or recruit to the population) in subsequent years for a given spawn or brood year (recruits per female or recruits per spawner) and provide a metric of productivity. Time-series estimates of recruits per spawner allow us to fit stock-recruitment curves to the data. Stock-recruitment curves are an important tool in fisheries management and conservation (Ricker 1973, Hilborn 2003). Estimates of productivity also facilitate extinction risk analysis. Reasonable estimates of stock-sex and stock-age combinations for genetic stocks at LGR would allow us to generate brood tables for genetic stocks and facilitate stock-recruitment and productivity monitoring at a scale and coverage recently unavailable to managers. For steelhead, if we use a more lenient rule of less than 10% PctBias and greater than 80% confidence interval coverage, we demonstrate that we can get reasonable estimates of male and female abundance for 5 genetic stocks (MFSALM, SFSALM, UPCLWR, SFCLWR, IMNAHA) under the simulation input parameters we used similar to SY2011. Further, we show that for the 3 most prevalent age classes (BY05, BY06, BY07) within these 5 stocks we are able to achieve these goals for 10 of the 15 stock-age combinations. For spring-summer Chinook salmon, we achieve less than 10% PctBias and greater than 80% confidence interval coverage for 13 of the 14 stock-sex combinations (UPSALM..M = 74.2% CI Cover) and for the 3 most prevalent ages (BY06, BY07, BY08) we achieve these goals for all 21 stock-age combinations. Stock-recruit analyses and productivity analyses within the Snake River basin would largely be driven by these prevalent age classes. The LGR sampling program combined with our time-stratified parametric bootstrapping program plus individual genetic assignments should allow us to perform productivity analyses for 5 of the steelhead stocks and all 7 Chinook salmon stocks delineated in Ackerman et al. (2014) and provide a powerful tool for managers for conservation assessments.

Based on the Inferred simulations, we found that we cannot likely obtain reasonable stock, stock-sex, or stock-age abundance estimates for 5 steelhead stocks: UPSALM, LOSALM, LOCLWR, GRROND, LSNAKE. The increased bias and decreased confidence interval coverage for estimates of these stocks is likely primarily due to stock misallocations. These 5 stocks exhibit decreased genetic structure (Ackerman et al. 2014, Matala et al. 2014) relative to the other stocks in the Snake River basin, and thus, increased genetic misallocation occurs among these 5 stocks. Four of these 5 stocks occur in lower mainstem reaches of the Snake River basin and show evidence of increased straying (gene flow) among stocks (); hatcheries in the remaining stock, UPSALM, were founded by fish captured in Hells Canyon that are genetically similar to fish in the LOSALM, LOCLWR, and LSNAKE stocks (). There are a number of potential solutions to reduce or eliminate stock misallocation. First, you can add collections or populations to the baseline used for genetic stock identification if it is believed that contributing stocks may not be represented. However, the Snake River genetic baselines have dense representation across the basin (Ackerman et al. 2014) and this is likely not a problem for IA at LGD. A second solution is to ‘collapse’ genetically similar stocks, a common practice in genetic stock identification. Collapsing genetincally similar stocks would effectively eliminate genetic misallocations that occur between those stocks (). However, collapsing of two or more genetic stocks into geographically larger stocks may not be desirable to managers. A final solution would be to correct for genetic misallocations given what we know about genetic structure among stocks and misallocation during simulations…

Although our case study is characterized by a high degree of sampling control (near census with systematic compositional samples), there are lessons for other fisheries applications. Certainly SY2011 had more control than other years at Lower Granite Dam (cite some Schrader reports) in that the run was as expected and there were no unexpected gaps in sampling. Frequently, sampling rate changes due to logistical concerns (e.g., more fish than expected can overwhelm trap and processing capacity) or environmental issues such as high summer water temperatures which can prohibit fish handling due to stress concerns. These types of problems will be greater in other cases where total abundance estimates have considerable variance or the sampling rate of the population for compositional data is small or unknown. The value of our case study is that is provides a best-case scenario to illustrate some basic statistical issues.

In short, the two data set bootstrap method for finding confidence intervals worked well. Our method for estimating numbers of steelhead and spring/summer Chinook salmon in compositional groups provides unbiased estimators and reasonably narrow confidence intervals if the group size is 300 or more. If the group size is 100 or more the confidence interval coverage is generally close to the nominal value (95%), but the confidence intervals may be wide. Estimators and confidence intervals for groups of less than 100 may not have good statistical properties.

Reasonable estimates of stock-sex and stock-age combinations would allow biologists to build brood tables for each stock and allow for productivity analyses…

REFERENCES

Ackerman, M. W., N. Vu, J. McCane, C. A. Steele, M. R. Campbell, A. P. Matala, J. E. Hess, and S. R. Narum. 2014. Chinook and Steelhead Genotyping for Genetic Stock Identification at Lower Granite Dam. Idaho Department of Fish and Game Report 14-01. Annual Report, Bonneville Power Administration Project 2010-026-00. Available at: <https://collaboration.idfg.idaho.gov/FisheriesTechnicalReports/Res14-01Ackerman2014%20Chinook%20and%20Steelhead%20Genotyping%20for%20GSI%20at%20LGR.pdf>

Ackerman, M. W., C. Habicht, and L. W. Seeb. 2011. Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. Transactions of the American Fisheries Society. 140(3):865-881. DOI: 10.1080/00028487.2011.588137

Anderson, E. C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Molecular Ecology Resources 10(4):701-710.

Anderson, E. C., R. S. Waples, and S. T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 65:1475-1486.

Busby, P. J., T. C. Wainwright, G. J. Bryant, L. J. Lierheimer, R. S. Waples, F. W. Wauneta, and I. V. Lagomarsino. 1996. Status review of West Coast steelhead from Washington, Idaho, Oregon, and California. NOAA Technical Memorandum NMFS-NWFSC-27.

Campbell, M. R., C. C. Kozfkay, T. Copeland, W. C. Schrader, M. W. Ackerman, and S. R. Narum. 2012. Estimating Abundance and Life History Characteristics of Threatened Wild Snake River Steelhead Stocks by Using Genetic Stock Identification. Transactions of the American Fisheries Society. 141(5):1310-1327. DOI: 10.1080/00028487.2012.690816

Casella, G. and E. I . George. Explaining the Gibbs Sampler. The American Statistician. 46(3):167-174.

Cassinelli, J., S. Rosenberger, and F. Bohlen. 2012. 2011 Calendar Year Hatchery Chinook Salmon Report: IPC and LSRCP Monitoring and Evaluation Programs in the State of Idaho. Idaho Department of Fish and Game Report Number 12-02. Available at: <https://collaboration.idfg.idaho.gov/FisheriesTechnicalReports/Res12-02Cassinelli2011%20Calendar%20Year%20Hatchery%20Chinook.pdf>

Harmon, J. R. 2003. A trap for handling adult anadromous salmonids at Lower Granite Dam on the Snake River, Washington. North American Journal of Fisheries Management 23:989-992.

Hilborn, R., and Walters, C. J. (1992), *Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty*, New York: Chapman and Hall.

Koljonen, M. L., J. J. Pella, and M. Masuda. 2005. Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo salar*) catches from DNA microsatellite data. Canadian Journal of Fisheries and Aquatic Sciences 62:2143-2158.

Latremouille, D. N. 2003. Fin erosion in aquaculture and natural environments. Reviews in Fisheries Science 11:315-335.

Matala, A. P., M. W. Ackerman, M. R. Campbell, and S. R. Narum. 2014. Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes. Evolutionary Applications.

McDowall, R. M. (1999) , "Tim I need the ref here," *I dont know,* 99, 1-999.

Ogden, D. A. 2012. Operation of the Lower Granite Dam adult trap. National Marine Fisheries Service. Annual report 2010, BPA project 2005-002-00.

Ogden, D. A. 2011. Operation of the Lower Granite Dam adult trap. National Marine Fisheries Service. Annual report 2009, BPA project 2005-002-00.

Pella, J. J. and G. B. Milner. 1987. Use of genetic marks in stock composition analysis. Pages 274-276 *in* N. Ryman and F. Utter, editors. Population genetics and fisheries management. University of Washington Press, Seattle.

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the USA 94:9197-9201.

Raymond, H. L. 1988. Effects of hydroelectric development and fisheries enhancement on spring and summer Chinook salmon and steelhead in the Columbia River basin. North American Journal of Fisheries Management 8:1-24.

Schrader, W. C., M. P. Corsi, P. Kennedy, M. W. Ackerman, M. R. Campbell, K. K. Wright, and T. Copeland. 2013. Wild adult steelhead and Chinook salmon abundance and composition at Lower Granite Dam, spawn year 2011. Idaho Department of Fish and Game Report 13-15. Annual report 2011, BPA Projects 1990-055-00, 1991-073-00, 2010-026-00. Available at: https://collaboration.idfg.idaho.gov/FisheriesTechnicalReports/Res13-15Schrader2013%20Wild%20Adult%20Steelhead%20and%20Chinook%20Salmon%20Abundance.pdf

Schrader, W. C. T. Copeland, P. Kennedy, M. W. Ackerman, K. K. Wright, and M. R. Campbell. 2012. Wild adult steelhead and Chinook salmon abundance and composition at Lower Granite Dam, spawn year 2010. Idaho Department of Fish and Game Report 12-16. Annual report 2010, BPA projects 1990-055-00, 1991-073-00, 2010-026-00. Available at: https://collaboration.idfg.idaho.gov/FisheriesTechnicalReports/Res12-16Schrader2012%20Wild%20adult%20steelhead%20and%20Chinook%20salmon%20abundance.pdf

Seeb, J. E., C. E. Pascal, R. Ramakrishnan, and L. W. Seeb. 2009. SNP genotyping by the 5’-nuclease reaction: advances in high-throughput genotyping with nonmodel organisms. Pages 277-292 in A. Komar, editor. Methods in molecular biology: single-nucleotide polymorphisms, 2nd edition. Humana Press, Totowa, New Jersey.

Shaklee, J. B., T. D. Beacham, L. Seeb, and B. A. White. 1999. Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fisheries Research 43:45-78.

Steinhorst, K., D. Milks, G. P. Naughton, M. Schuck, and B. Arnsberg. 2010. Use of Statistical Bootstrapping to Calculate Confidence Intervals for the Fall Chinook Salmon Run Reconstruction to Lower Granite Dam. Transactions of the American Fisheries Society 139:1792-1801. DOI: 10.1577/T09-200.1

Wood, C. C. S. McKinnell, T. J. Mulligan, and D. A. Fournier. 1987. Stock identification with the maximum-likelihood mixture model: sensitivity analysis and application to complex problems. Canadian Journal of Fisheries and Aquatic Sciences 44:866-881.