

# Absolute abundance estimates of juvenile spring-run and winter-run Chinook salmon at Chipps Island

Funded by

Delta Science of the Delta Stewardship Council

(previously CALFED Bay-Delta Program)

Grant Agreement Number 1049

Awarded September 1, 2007

*Prepared by:*

**Brian Pyper<sup>1</sup>, Tommy Garrison, and Steve Cramer**

Cramer Fish Sciences

Gresham, OR

**Patricia L. Brandes**

U.S. Fish and Wildlife Service

Lodi, CA

**David P. Jacobson and Michael A. Banks**

Coastal Oregon Marine Experiment Station

Department of Fisheries & Wildlife

Oregon State University

Newport, Oregon

July 1, 2013

<sup>1</sup>Current address: Fish Metrics, 2027 SE Spokane St., Portland, OR



**U.S.  
Fish & Wildlife  
Service**



**Oregon State University**

## Acknowledgments

We would like to acknowledge Jonathan Speegle for database help, Denise Barnard and Jonathan Thompson for leading the organization and checking of DNA samples, and all the biological technicians and biologists who took tissue samples at Chipps Island as part of sampling by the U.S. Fish and Wildlife on behalf of the Interagency Ecological Program (IEP). We thank Eric Volkman, Pete Hrodey, Denise Barnard and John Netto who oversaw the tissue collections at Chipps Island. We would also like to thank the California Department of Fish and Wildlife's Central Valley tissue archive laboratory for splitting tissue samples prior to sending them to Oregon State University. California's Departments of Water Resources and Fish and Wildlife graciously funded the splitting of the tissue samples. We thank Ken Newman, mathematical statistician for the Stockton FWS office, for his invaluable support and suggestions throughout the project, and Casey Justice for his contributions in developing a statistical sampling plan. We thank Delta Science for funding the project and IEP for supporting the sampling at Chipps Island.

## Table of Contents

Executive Summary.....	9
Introduction.....	11
Methods.....	14
Data collection.....	14
Chippis Island trawl.....	14
Sampling plan.....	16
Genetic analyses and run assignments.....	17
Sample estimates of corrected assignments by run.....	18
Total catch estimates by run.....	25
Total abundance estimates by run.....	26
Estimates of efficiency.....	28
Results.....	29
Summaries of trawl effort, catch, and DNA samples.....	29
Observed DNA run assignments.....	30
Comparisons with assignments based on length-at-date criteria.....	35
Assignment corrections and total catch estimates by run.....	39
Fall run.....	39
Late-fall run.....	39
Butte Creek spring run.....	40
Mill-Deer spring run.....	40
Winter run.....	40
Negative estimates of corrected assignments.....	41
Total abundance estimates by run.....	47
Components of variance.....	48
Discussion.....	58
Comparison to length-at-date criteria.....	58
Corrections to DNA assignments based on blind-test data.....	59
Estimates of total abundance and trawl efficiency.....	61
Independent estimates of winter run abundance.....	63
Implications for past and future sampling.....	64

Related management implications .....	65
Summary of recommendations.....	66
References.....	68
Appendix A: Derivation of selected estimators .....	71
Estimates of total catch by run .....	71
Estimates of total abundance by run.....	72
Appendix B: Second most likely run assignments .....	78
Attachment A: TECHNICAL BREIF on Sample size allocation for DNA analysis of juvenile Chinook salmon captured in Chipps Island midwater trawl.....	79

## List of Figures

<b>Figure 1.</b> Map of San Francisco Bay and Delta showing location of Chipps Island where trawling for smolts has been conducted annually. ....	13
<b>Figure 2.</b> Overview of sampling process and estimators that lead to total abundance estimates. ....	15
<b>Figure 3.</b> Graphical display of Chipps Island trawl effort summaries by biweekly period and sampling year. The size of each rectangle is proportional to the value in Table 5. ....	32
<b>Figure 4.</b> Boxplots of fork length (mm) by period and sample year for DNA-analyzed and not-analyzed juvenile Chinook salmon caught in Chipps Island trawl. Boxplots show medians (horizontal lines), 25th-75th percentiles (boxes), $1.5 \times$ (75th- 25th percentiles) (whiskers). Data points beyond the end of the whiskers are outliers. ....	34
<b>Figure 5.</b> Scatterplot of juvenile Chinook salmon caught in Chipps Island trawl and DNA-assigned to run as a function of fork length and sample day (all four years of study data combined). Run abbreviations are fall (F), late-fall (LF), spring Butte (SB), spring Mill and Deer (SMD), and winter (W). ....	36
<b>Figure 6.</b> Comparison of run assignments based on length-at-date criteria versus DNA (observed and corrected) for juvenile Chinook salmon caught in Chipps Island trawl and DNA assigned to run. For each sample year and assignment method, the percentage of total juveniles assigned to each run is shown. ....	37
<b>Figure 7.</b> Scatterplot of juvenile Chinook salmon caught in Chipps Island trawl and DNA assigned to run (panels) as a function of fork length and sample day (all four years of study data combined). The color regions correspond to length-at-date criteria for run assignment. ....	38
<b>Figure 8.</b> Abundance estimates of winter-run juvenile Chinook salmon at Chipps Island by sample year for four different estimates of trawl efficiency (abundance estimates based on corrected DNA assignments). ....	56
<b>Figure 9.</b> Pie chart depicting the run composition of juvenile Chinook salmon abundance at Chipps Island by sample year. Abundance estimates were based on corrected DNA assignments using the Jersey Point estimate of trawl efficiency. Run abbreviations are fall (F), late-fall (LF), spring Butte (SB), spring Mill-Deer (SMD), and winter (W). ....	57

## List of Tables

<b>Table 1.</b> Classification table of blind-test data for the number of fish by true known run and DNA run assignment.....	22
<b>Table 2.</b> Modified classification table of the number of fish by true known run and DNA run assignment for two categories of ONCOR assignment probability.....	22
<b>Table 3.</b> Conditional probabilities of run assignments (for a fish of a given true run) for two categories of ONCOR assignment probability. Off-diagonal elements correspond to false positive error rates (rows: probability of wrongly assigning a different run to be the run of interest) and false negative error rates (columns: probability of wrongly assigning the run of interest to a different run). .....	22
<b>Table 4.</b> Estimates of Chipps Island trawl efficiency and standard errors (SE) as reported in Pyper et al. (2013). All estimates assume (i.e., are standardized to) a volume-sampled rate of 1000 m <sup>3</sup> /minute (based on volume measurements in the current trawl database). No standard error is provided for the fish flux method, which is based on a set of assumed constants. ....	28
<b>Table 5.</b> Chipps Island trawl effort summaries by biweekly period and sampling year. A “ – “ indicates that no trawling was conducted during this strata. Sample year 1 is defined as August 1 <sup>st</sup> , 2007 – July 31 <sup>st</sup> , 2008 and similarly for 2, 3 and 4. Sampling effort summaries are not presented from August 1 <sup>st</sup> – September 30 <sup>th</sup> , 2007, which preceded the onset of DNA sampling and from July 1- 30 <sup>th</sup> , 2011 when DNA sampling concluded. ....	31
<b>Table 6.</b> Raw catch and number of DNA samples (assigned to run) taken at Chipps Island by biweekly period and sampling year. A “ – “ indicates that no sampling (trawl or DNA) was conducted during this period. Blank entries indicate either zero catch or no DNA samples. Catch is not reported from August-September in sample year 2007 and July in sample year 2010 because no DNA samples were taken during these periods. ....	33
<b>Table 7.</b> Number of fish assigned to run by biweekly period and ONCOR assignment-probability category. All sample years are combined.....	35
<b>Table 8.</b> Number of fall-run juvenile Chinook by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. ....	42
<b>Table 9.</b> Number of late-fall-run juvenile Chinook by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. ....	43
<b>Table 10.</b> Number of spring-run (Butte Creek) juvenile Chinook by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for	

estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. .... 44

**Table 11.** Number of spring-run (Mill-Deer) juvenile Chinook by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. .... 45

**Table 12.** Number of winter-run juvenile Chinook by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. .... 46

**Table 13.** Absolute abundance estimates for juvenile fall-run Chinook by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. .... 49

**Table 14.** Absolute abundance estimates for juvenile late-fall-run Chinook by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. A “ – “ indicates that no trawl sampling was conducted during this strata. Time periods with zero catch or no fish sampled for DNA are left blank. .... 50

**Table 15.** Absolute abundance estimates for juvenile spring-run (Butte Creek) Chinook by biweekly period and sample year based based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. .... 51

**Table 16.** Absolute abundance estimates for juvenile spring-run (Mill-Deer) Chinook by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. .... 52

**Table 17.** Absolute abundance estimates for juvenile winter-run Chinook by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. .... 53

**Table 18.** Annual abundance estimates for juvenile Chinook by run and sample year based on observed and corrected DNA assignments. Annual abundances are shown for four alternative estimates of Chipps Island trawl efficiency (Jersey Point releases, paired-release tests, Pittsburg releases, and the fish flux method). Standard errors are shown in parentheses..... 54

**Table 19.** Percent difference between annual abundance estimates for juvenile Chinook by run and sample year based on observed versus corrected DNA assignments. Differences in abundance estimates were computed relative to abundances based on observed assignments (i.e., % difference = 100\*[corrected – observed]/observed)..... 55

**Table 20.** Coefficient of variation (CV) and components of variance (as a percentage of total variance) by run and sample year for annual abundance estimates based on corrected assignments and the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency.55

**Table B1.** Number of fish assigned to run by ONCOR first mostly likely run assignment and second most likely run assignment. .... 78

## Executive Summary

In this study, we estimated juvenile Chinook salmon abundance by run using catch in midwater trawling at the confluence of the Sacramento and San Joaquin Rivers and entrance to San Francisco Bay near Chipps Island. Juveniles captured since 1993 by this trawling effort have been assigned to run based on their length and date captured. Instead, we report the results of run assignments based on genetic (DNA) markers for trawl samples collected from October, 2007 to June, 2011. Independent blind-test data were used to estimate, and account for, error rates in DNA assignments. In a companion report, Pyper et al. (2013) examined alternative methods and data from the historical sampling period to estimate trawl capture efficiency. The results of that study were used in conjunction with the DNA run assignments to estimate total abundance of juveniles from each run reaching San Francisco Bay.

Across years, DNA assignments indicated that fall run composed between 84.0% and 92.8% of the annual juvenile abundance, late-fall run composed 1.9% to 4.4%, and Butte Creek spring run ranged between 3.9% and 9.0%. Mill-Deer creek spring run and winter run each composed less than 3% of the total abundance in 2008, and less than 2% in subsequent years. However, estimates of DNA assignments were highly uncertain for late-fall run and Mill-Deer spring run due to uncertainty in potential misclassifications of true fall run to these runs.

DNA-based estimates of race composition often differed substantially from those based on length-at-date criteria. Across all four years, more fish were assigned to fall run (+2.4%), and far fewer fish were assigned to spring and winter runs based on DNA compared to length criteria. Winter-run DNA assignments had the closest fit to their expected length-at-date range, with only a few fish overlapping the adjacent late-fall and spring ranges. However, relatively large numbers of fall, late-fall, and spring run fish overlapped with the winter-run length criteria. Similarly, large numbers of fall run (based on DNA) overlapped with the spring-run length criteria. Consequently, use of DNA assignments provided much more accurate, and reduced, annual estimates of run composition for the spring and winter runs, which were one half to one sixth of the run compositions based on length criteria across years.

Juvenile abundance was estimated assuming that trawl efficiency was constant across biweekly periods and years. Abundances were compared for four different efficiency estimates (Pyper et al. 2013), including three empirical estimates that were independently derived using coded-wire-tag release data (Pyper et al. 2013). The fourth approach examined was the “fish flux” method of Kimmerer (2008), which had an implied efficiency that was substantially higher than the empirical estimates, and was considered to be likely biased.

The ranges in annual abundances from 2008 (August 1, 2007 to July 31, 2008) to 2011 based on the trawl efficiency estimate for Jersey Point releases (the midrange of empirical estimates) were as follows: 1.4 million to 7.5 million for fall run; 71 thousand to 186 thousand for late-fall run; 67 thousand to 331 thousand for Butte Creek spring run; 36 thousand to 92 thousand for Mill-Deer

creek spring run; and 45 thousand to 63 thousand for winter run. Annual abundances were lowest for all runs in 2008, while the highest abundances were observed in 2011 for fall and late-fall runs, and in 2010 for spring and winter runs. The most precise estimates of annual abundance were for fall run, with coefficients of variation (CVs) of 21% or less. Abundances of were also relatively precise for Butte Creek spring run (CVs of 30% or less) and winter run (37% or less). Abundance estimates for late-fall run and Mill-Deer creek spring run were very imprecise (CVs > 75%) due to uncertainty in potential misclassifications of true fall run.

While the precision estimates for Butte Creek spring run and winter run are encouraging, they should be interpreted cautiously because abundances were sensitive to the choice of efficiency estimate (a roughly two-fold difference among the three empirical estimates), and because efficiency was assumed to be constant over time. It is currently unclear which of the efficiency estimates we examined is most accurate, and to what extent trawl efficiency may vary seasonally or among years. These and other uncertainties we identify warrant further investigation.

## Introduction

Two of the most important metrics for monitoring anadromous salmonid populations are the abundances of spawners and the number of juveniles they produce. In the Central Valley of California where adult Chinook salmon production supports major fisheries in the ocean and freshwater, the numbers of juveniles leaving freshwater during the spring has been sampled annually since 1978 by means of midwater trawling in the San Francisco Estuary near Chipps Island (Figure 1) (Brandes and McLain 2001). Chipps Island is located downstream from the junction of the Sacramento and San Joaquin rivers, and thus is located where all juvenile Chinook salmon produced in the two basins must pass enroute to the ocean. The area sampled near Chipps Island is relatively constricted (3/4 of a mile across the channel), which provides the most concentrated opportunity for sampling juveniles as they leave the Central Valley.

Since 1993, trawling has also occurred at Chipps Island in other months of the year to estimate juvenile abundance by run. The four runs in the Central Valley and more specifically the Sacramento River basin are fall, late-fall, spring, and winter run (Fisher 1994). The San Joaquin tributaries support only a fall-run population. These runs are named after the season in which adults return to freshwater. Winter run is listed as endangered and spring run is listed as threatened under the federal Endangered Species Act (ESA)(NMFS 1994), and thus distinction of these runs and estimation of their abundance is critical to gauging the success of management actions aimed at recovering these stocks. Abundance at Chipps Island has historically been estimated using two methods to expand catches: (1) using the proportion of time and channel width sampled to expand catches; and (2) using an estimate of trawl efficiency to expand catches (USFWS 1997). Trawl efficiency is based on the proportion of marked fish surviving to the trawl and recovered in the trawl from releases made upstream, corrected for sampling effort. Differences in abundance estimates between methods of catch expansion have raised uncertainty as to which method is most reliable.

In addition to uncertainty regarding catch expansion, genetic analyses indicate that length-at-date methods used to apportion total juvenile abundance into the various runs of Chinook salmon have been inaccurate. Those methods used length and date of capture to assign fish to a given race (Fisher, 1992 and S. Greene, California Department of Water Resources, pers. comm.). Because the fall run composes over 90% of adult Chinook returning to the Central Valley (CHINOOKPROD, [www.fws.gov/stockton/afpr/](http://www.fws.gov/stockton/afpr/), accessed 6/20/13), small errors in classification of individuals from this run can cause large errors in the numbers assigned to other runs.

In recent years, genetic markers have been developed that make it possible to distinguish race of Chinook with greater than 95% accuracy (Banks and Jacobson 2004). Fin tissue for DNA analysis was collected for 6 years from a subset of juveniles sampled at the Delta fish facilities, and results showed that true winter run (determined by DNA) composed between 4 to 84% (with an average of 49%) of the juvenile salmon that were designated as winter run based on length-at-date criteria (Hedgecock 2002). Although most genetic winter run were within their designated

length-at-date range (95.5%), roughly half the Chinook in that length range were actually of a different run (Hedgecock 2002). These results indicate that use of length-at-date criteria can result in large overestimates of juvenile winter-run abundance. The length-at-date method may be even less accurate for spring run because their length and time of juvenile migration overlap considerably with the fall run.

To reduce these sources of uncertainty, the study reported here was designed with two objectives: (1) to determine the most reliable methods for expanding trawl catches to total abundance; and (2) to sample genetic composition of the juvenile catches at Chipps Island and estimate the abundance that each genetically distinct group composed. Expansion of trawl catches to total abundance is based on estimates of capture efficiency (proportion of available fish captured). Chipps Island trawl efficiency was estimated using several alternative methods and is the focus of a separate report (Pyper et al. 2013). Here, we focus on the results of genetic sampling of juvenile salmon catches from October, 2007 to June, 2011 to estimate the abundance and proportionate contribution to total juvenile production from each run. Note that catches of fall, spring and late-fall run within each annual period likely incorporate progeny from two brood years. Although we report abundance estimates of all four runs, our focus is on spring and winter run because (1) the statistical power of individual-based genetic assignments of these runs is more established than for the other runs (Banks 2005), and (2) there is an urgent need for accurate estimates of their juvenile abundance to facilitate understanding of their population dynamics and status (Cramer et al. 2004).

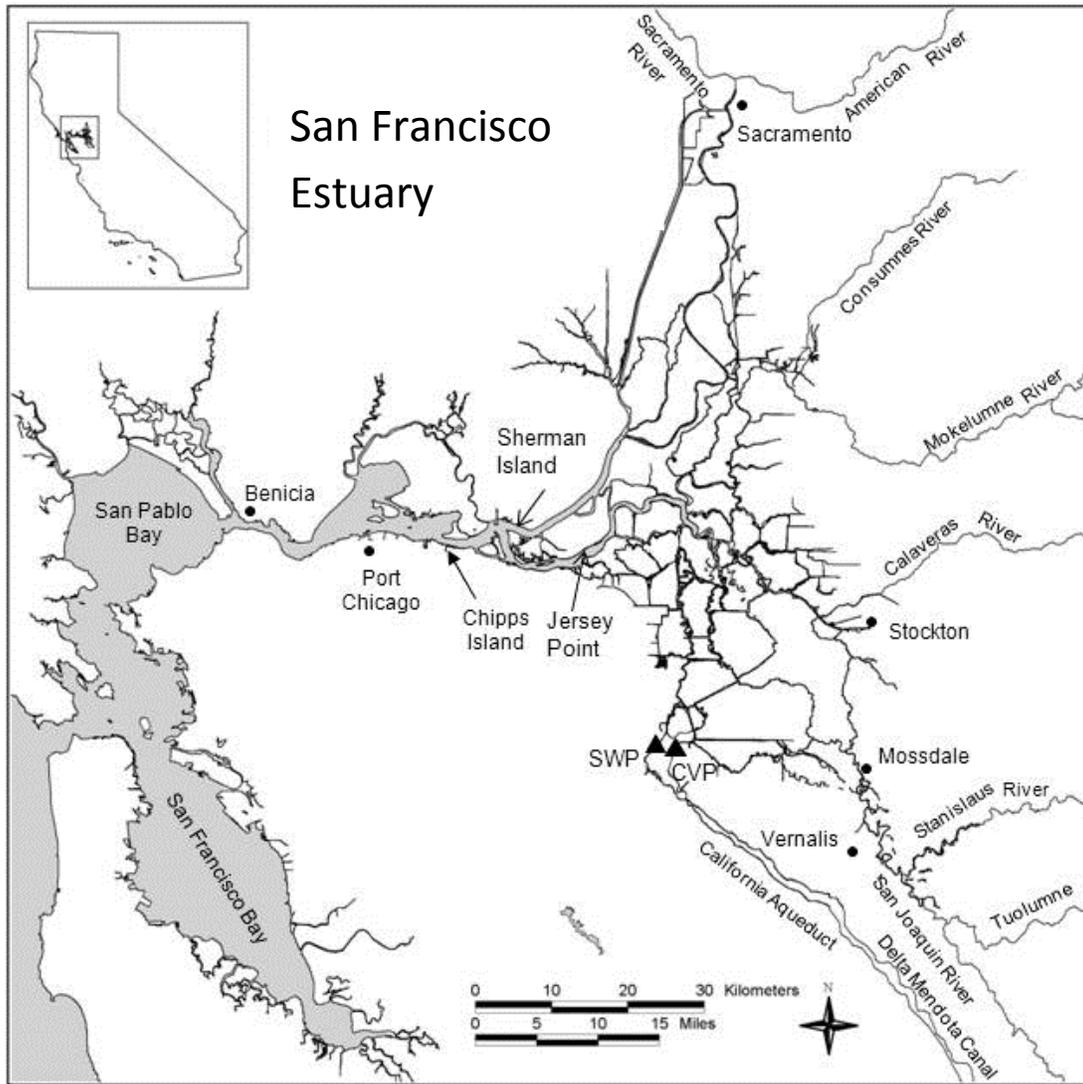


Figure 1. Map of San Francisco Estuary showing location of Chipps Island where trawling for smolts has been conducted annually.

## Methods

The following sections describe the steps we used to estimate absolute abundances ( $N$ ) by run of juvenile Chinook salmon migrating past Chipps Island. A simple overview of the key sampling processes and estimators is shown in Figure 2; actual estimates were more complicated. In short, there were three main steps to estimating absolute abundances: (1) estimation of “corrected” run assignments ( $x$ ) in samples of trawl catch based on observed DNA assignments ( $y$ ) and blind-test data; (2) estimation of total catch ( $X$ ) by run given the fraction ( $f$ ) of catch sampled; and (3) estimation of total abundance ( $N$ ) given estimates of trawl efficiency ( $E$ ) and trawl effort ( $p$ ).

### Data collection

#### Chipps Island trawl

Midwater trawling has been conducted at Chipps Island between April and June since 1978. This sampling was initiated to gain relative abundance and survival information on juvenile salmon emigrating from the Delta towards the Pacific Ocean (Brandes and McLain, 2001). In October, 1993, sampling was expanded to continue through June, 1994 and since October of 1994 year-round sampling has been conducted to better understand the temporal patterns of juvenile salmon emigration downstream. Generally, ten 20-minute tows were conducted three to seven days each week from April to June (Brandes and McLain, 2001). Sampling was conducted seven days per week during recovery of experimental releases of coded-wire-tagged (CWT) salmon (usually December-January and April-May) to increase the numbers recovered from these experimental fish released upstream and in the Delta.

Trawls were conducted within a 3 km section of river upstream of the western tip of Chipps Island (Brandes and McLain, 2001). Trawls were conducted in both directions (upstream and downstream) regardless of tide in three channel locations: north, south, and middle. Occasionally, inclement weather, mechanical problems, or excessive delta smelt or salmon catch reduced tow duration or number of tows per day.

Between October, 2007 and June, 2011, ten 20-minute tows were usually conducted two to three days per week but at times, tow duration was reduced or days were cancelled to stay within daily or annual delta smelt incidental take limits as managed through the Interagency Ecological Program. For instance, between February 5 and March 10 of 2008, trawling at Chipps Island was cancelled due to concerns about high delta smelt incidental take. A similar curtailment period occurred between June and October of 2007. During some periods, tows were limited to as little as 5 minutes to assess delta smelt take prior to conducting tows of 15 or 20 minutes.

Recent measurements conducted in 2009 determined that the trawl net fished at Chipps Island has a mean effective-fishing mouth size of 12.7 m<sup>2</sup> (Whitesel) or 13.0 m<sup>2</sup> (Confluence) depending on the vessel used (preliminary unpublished data). These values differ from the value of 18.5m<sup>2</sup> reported in Brandes and McLain (2001), which was based on fishing-net dimensions

reported in 1993 (USFWS 1994). Importantly, measurements of volume sampled in the current trawl database (and used in this report) do not reflect these changes in mouth size (i.e., database volumes and those reported here are based on an assumed mouth size of 18.5 m<sup>2</sup>). However, there was only one instance – in the estimation of abundance using the “fish flux” method discussed below – where modifications were required to incorporate the recent (improved) estimate of effective-fishing mouth size.

Fin tissue for DNA analysis was collected from juvenile Chinook salmon captured in the trawl sampling conducted at Chipps Island per a modified sampling plan.

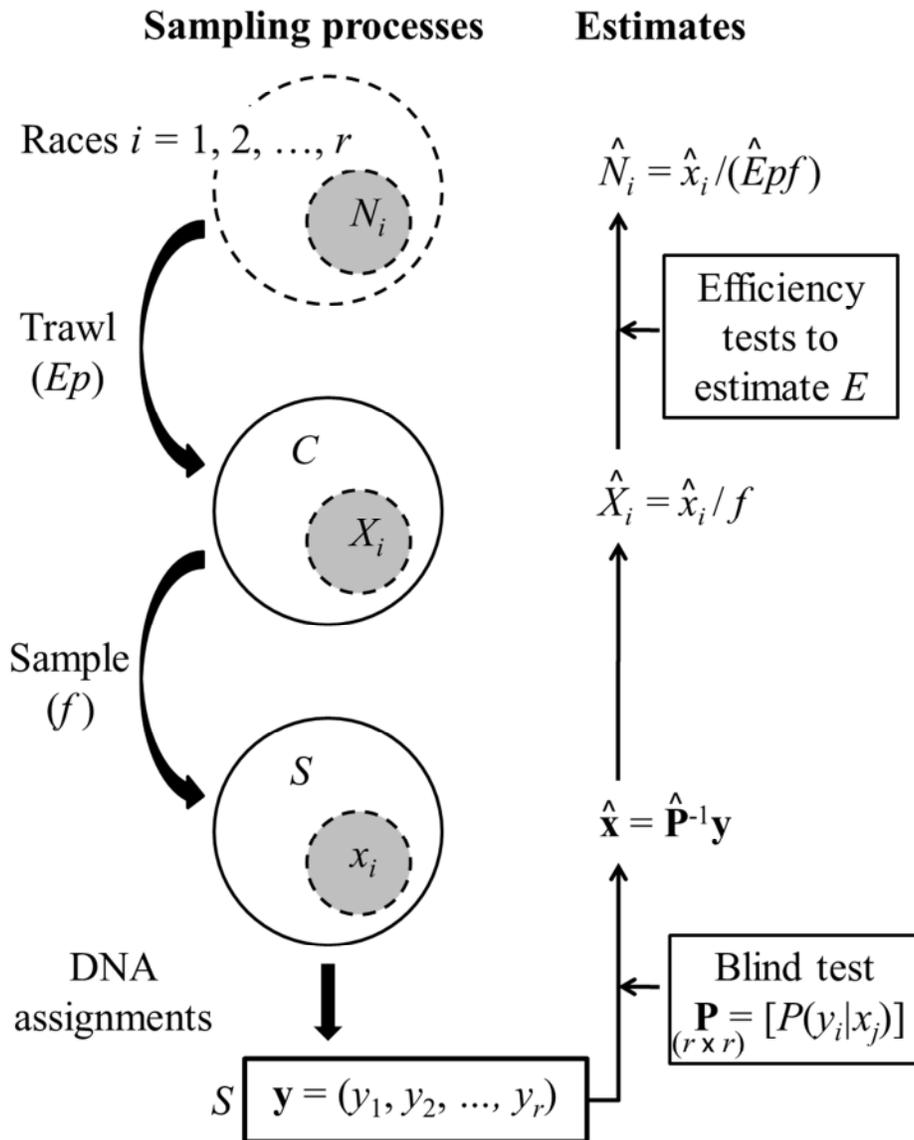


Figure 2. Overview of sampling process and estimators that lead to total abundance estimates. Race is the same as run (e.g. winter run).

### Sampling plan

The original sampling plan (Attachment A) recommended tissue collection from all unmarked juvenile salmon caught between December and June that had lengths either within or greater than the river model's length-at-date criteria for winter run (Fisher, 1992). Juvenile salmon within the spring run and fall run length-at-date criteria were also to be tissue sampled, but maximum sample sizes were specified depending on the length class and time of year. This original plan was designed to optimally allocate the annual target of 3000 samples for evaluating winter and spring run, recognizing that many fall run would likely be included in the spring length-at-date criteria, and a few true spring run would be included in the fall-run length-at-date criteria.

It should be noted that almost 100% of hatchery late-fall run, winter run, and spring run were marked with CWTs during these sample years and, thus, were not tissue sampled (Kormos et al. 2012; USFWS 2011, p.186). In addition, since 2007, a minimum of 25% of the fall-run hatchery production from the Central Valley hatcheries has been marked and tagged (USFWS, 2011). Thus some of the unmarked juvenile Chinook salmon sampled for tissue for this study were unmarked fall-run hatchery fish.

The original sampling plan was modified during the first year of the study because of the unexpected curtailment (temporarily) in February, 2008, and reduction of sampling at Chipps Island thereafter, due to delta smelt take concerns. Given reduced sampling and a coincidental reduction in salmon abundance, attempts were made to tissue sample all juvenile salmon caught in the trawl, with some minor exceptions. Some sub-sampling was incorporated during late April and early May when many unmarked fall-run hatchery fish were assumed to be in the catch, based on the number of tagged hatchery fish being caught. During those times, 5 fish in the fall-run length-at-date category were to be tissue sampled per tow (50 per day). In addition, juvenile salmon under 50 millimeters were not sampled because it was determined that tissue sampling would cause mortality.

After sampling at Chipps Island was interrupted in February 2008, we obtained permission from California Bay-Delta Authority (our funders at the time) and modified our ESA take permit to allow tissue sampling of juvenile salmon collected near Sacramento in regular IEP trawling between March 2008 and June 2011. These tissue samples were processed for genetic run designation and will be the basis for future analyses.

Sampling protocols for collecting tissue from juvenile Chinook salmon captured in the trawl at Chipps Island were similar to those at the State Water Project (SWP) and Central Valley Project (CVP) fish facilities, although samples were placed on filter paper and air dried instead of using a buffer solution as has been the protocol at the SWP and CVP (Harvey et al. in press). A 1 X 2 mm or 2 X 4 mm triangular piece of tissue was taken in the field from the top or bottom lobe of the caudal fin shortly after a juvenile was caught. The tissue was placed on filter paper, folded over twice, and inserted into a labeled coin envelope for drying back in the laboratory, prior to

placing it in a plastic bag for longer-term storage. Samples were given a unique ID number and were linked to individual catches in the trawl catch database.

Once organized and checked, tissue samples were sent to the California Department of Fish and Game's (now California Department of Fish and Wildlife) Central Valley Archive lab for splitting. Samples were then sent to Michael Banks's genetics laboratory at Oregon State University for analyses and run determination. Note that some juveniles were tissue sampled but were not included in the final run-assignment data because their tissue samples were lost, damaged, or yielded inconclusive run assignments (i.e., no run assignment was given).

### **Genetic analyses and run assignments**

Samples were characterized using a 21 microsatellite panel named HMSC21 using protocols detailed in Banks et al. (in review). HMSC21 includes the following loci: *Ots-104*, *-107* (Nelson and Beacham 1999); *Ots-201b*, *-208b*, *-209* *-211*, *-212*, *-215* (Greig et al. 2003); *Ots-G78b*, *-G83b*, *-G249*, *-G253*, *-G311*, *-G422*, *-G409* (Williamson et al. 2002); *Ost515* (Naish and Park 2002); and five microsatellites derived from research characterizing alternate copies of the circadian rhythm transcription factor *Cryptochrome*, including *Cry2b.1*, *Cry2b.2*, *Cry3* (O'Malley et al. 2010), *Ots-701* (GeneBank accession # KF163438), and *Ots-702* (GeneBank accession # KF163440). Alternate microsatellite alleles were resolved through electrophoresis utilizing an Applied Biosystems (AB) 3730xl DNA analyzer and scored using AB GeneMapper software (Version 4).

Data for the 21 microsatellites for each sample were assessed against the Hatfield Marine Science Center HMSC21 baseline utilizing the "assign individual to baseline population" option available in the computer application ONCOR (Kalinowski 2008 [www.montana.edu/kalinowski/Software/ONCOR.htm](http://www.montana.edu/kalinowski/Software/ONCOR.htm)) to determine the most likely sub-population origin for each sample. Data in this baseline are comprised of five primary sub-populations as described in Banks et al. (2000). These sub-populations or runs are named fall, late-fall, winter, and two reproductively isolated spring runs: (1) spring run from Butte creek; and (2) spring run from neighboring Mill and Deer creeks. The fall run sub-population includes mainstem spawning populations from throughout the Sacramento and San Joaquin as well as both early (putative spring) and late (putative fall) returns to the Feather River (spring run) because of difficulty in resolving sub-structure among these stocks (Banks et al. 2000; Hedgecock et al. 2001).

ONCOR assigns individuals in a mixture sample to the baseline population that has the highest probability of producing the given genotype in the mixture. Emphasis is placed on the phrase "in the mixture" because ONCOR uses both genotype frequencies and mixture proportions when estimating the origin of individuals. ONCOR performs these calculations as follows. Let  $p_{ij}$  denote the probability that individual  $i$  (of unknown origin) belongs to baseline population  $j$ . This probability  $p_{ij}$  can be estimated from the genotype frequencies in each baseline population and an estimate of the stock composition of the fishery. Let  $f_{ij}$  be the frequency of the  $i$ th fish's

genotype in the  $j$ th population and let  $m_j$  be the estimated stock composition of the sample. Following Rannala and Mountain (1997), an estimate of  $p_{ij}$ , which we refer to as the run assignment probability, is given by

$$(1) \quad \hat{p}_{ij} = \frac{m_j f_{ij}}{\sum_j m_j f_{ij}}.$$

Typically when using genetic assignment data, an overall assignment error rate is controlled for by determining a run assignment probability threshold at which individual observations are discarded. This allows for determination of a false-positive detection rate (i.e., Type I error in statistical hypothesis testing). For instance, Harvey et al. (in press) used blind-test data (described below) and a bootstrap procedure in an attempt to control for assignment error rates by determining a threshold that would yield a desired test-wide error rate. However, their results were inconclusive because the desired error rate was either never achieved for a run (i.e., too few fish were miss-assigned), or error rates were only achieved at the highest assignment probability of 1.00, which would require exclusion of all fish with assignment probability  $\leq 1.00$ . Furthermore, the blind-test evaluations revealed that ONCOR assignment probabilities did not correlate well with actual assignment accuracy for some runs (Harvey et al. in press). Thus, we did not attempt to restrict assignment data based solely on a threshold value for ONCOR assignment probability. Instead, we used blind-test data to quantify and account for likely assignment errors, as outlined in the next section.

### **Sample estimates of corrected assignments by run**

We utilized blind-test data of 623 known origin Chinook adult salmon to account for false positive (wrongly assigning any other run to be run of interest) and false negative (wrongly assigning run of interest to be any other run) error rates when estimating abundances by run in samples of trawl catch. A complete description of the adult sampling and genetic analyses of the blind-test data is found in Banks et al. (in review). Determining assignment error rates from blind-test data allows us to compute “corrected” estimates that should more accurately reflect the “true” numbers of fish by run in a given field sample.

#### Run assignment corrections: Example with two runs

The following example uses blind-test data to correct a new sample of assignments when there are only two runs,  $a$  and  $b$ . Let  $n$  be the total number of fish in the sample, let  $y$  be the number assigned by run, and let  $x$  be the true number of fish by run. The expected number of fish assigned to run  $a$  in the sample is given by (Ken Newman, personal communication):

$$(2) \quad E(y_a) = x_a P(a|a) + x_b P(a|b),$$

where  $P(a|a)$  is the conditional probability that a fish of run  $a$  is correctly assigned as run  $a$  and  $P(a|b)$  is the probability that a fish of run  $b$  is incorrectly assigned as run  $a$  (i.e. the false positive

error rate). To express Equation (2) as function of the false negative error rate, the following two substitutions are made

$$(3) \quad \begin{aligned} x_b &= n - x_a \\ P(a|a) &= 1 - P(b|a), \end{aligned}$$

where  $P(b|a)$  is the false negative error rate. The method-of-moments approach (e.g., Mood et al. 1974, p. 274) can then be used to solve for  $x_a$ :

$$(4) \quad \hat{x}_a = \frac{y_a - nP(a|b)}{1 - P(a|b) - P(b|a)}.$$

The blind-test data is used to estimate  $P(a|b)$  and  $P(b|a)$  by constructing a  $2 \times 2$  table  $T$  where the rows are the numbers of fish assigned to runs  $a$  and  $b$  based on genetics and the columns are the true known numbers of run  $a$  and  $b$

		True Run	
		$a$	$b$
Genetic Assignment	$a$	$T_{aa}$	$T_{ab}$
	$b$	$T_{ba}$	$T_{bb}$

Conditioning on the true values (i.e., the column totals), the estimates of false positive and false negative error rates are then given by

$$(5) \quad \begin{aligned} P(a|b) &= T_{ab} / (T_{ab} + T_{bb}) \\ P(b|a) &= T_{ba} / (T_{aa} + T_{ba}). \end{aligned}$$

#### Run assignment corrections: Example with several runs

The approach above can be generalized for a set of runs  $i = 1, 2, \dots, r$ . In brief, for each run, we can specify an equation for the expected assignment ( $E[y_i]$ ) analogous to Equation (2). This provides a classic “system of linear equations” that has a vector-matrix form  $\mathbf{y} = \mathbf{P}\mathbf{x}$ , where the vector  $\mathbf{y}$  is the set of expected run assignments  $\{E[y_i]\}$ , the column vector  $\mathbf{x}$  is the set of true numbers by run  $\{x_j\}$ , and  $\mathbf{P}$  is an  $r \times r$  matrix (with rows  $i$  and columns  $j$ ) of conditional probabilities,  $P(i|j)$  (i.e., the probability that a fish of true run  $j$  is assigned to run  $i$ , where  $j \in i$ ). Given an estimate of  $\mathbf{P}$  from blind-test data and a new sample of assignments  $\{y_i\}$ , we can estimate the true numbers by run

$$(6) \quad \hat{\mathbf{x}} = \hat{\mathbf{P}}^{-1}\mathbf{y},$$

where the column vector  $\mathbf{y} = (y_1, y_2, \dots, y_r)$  holds the observed run assignments and  $\hat{\mathbf{P}}^{-1}$  denotes the matrix inverse of  $\hat{\mathbf{P}}$ . We refer to these estimates of  $x_i$  as “corrected run assignments.” As discussed below, negative estimates of  $x_i$  can occur and require special consideration.

The blind-test data are used to estimate each entry in the matrix  $\mathbf{P}$ . Let  $T$  be an  $r \times r$  table where the rows are the assigned numbers by run  $i$  and the columns are the true known numbers by run  $j$  ( $j \in i$ ). An estimate of  $P(i|j)$  is obtained by

$$(7) \quad \hat{P}(i|j) = T_{ij}/T_{.j} ,$$

where  $T_{.j}$  is the sum across all entries  $i$  in column  $j$ . The false positive and negative error rates are the off diagonal elements of  $\hat{\mathbf{P}}$ .

To estimate the variance of each estimate  $\hat{x}_i$ , we used a parametric bootstrap procedure. Specifically, we assumed that each column of  $T$  (the blind-test data) was an independent multinomial sample with probabilities  $\hat{P}(i|j)$ , and accordingly, generated 1,000 bootstrap replicates for  $\mathbf{P}$  and hence  $\mathbf{P}^{-1}$ . We then estimated the variance-covariance matrix of  $\mathbf{P}^{-1}$ , denoted  $\mathbf{Q}$ , and used the relevant component of this matrix (i.e., the  $i$ th row) to estimate the variance of  $\hat{x}_i$  given a new sample of assignments  $\mathbf{y}$ :

$$(8) \quad \hat{\sigma}_{\hat{x}_i}^2 = \mathbf{y}'\hat{\mathbf{Q}}_i\mathbf{y} .$$

Note that in this report, we unintentionally omitted an important source of variance in the corrected estimates  $\hat{x}_i$  (Ken Newman, U.S. Fish and Wildlife Service, pers. comm.). In the formulation above, the observed assignments  $\{y_i\}$  in a given sample are assumed to be multinomial variables conditional on the true numbers by run  $\{x_i\}$  and true probabilities  $\{P(i|j)\}$ . However, Equation (8) only accounts for uncertainty in estimates of  $P(i|j)$  derived from blind-test data, and hence, we ignored the multinomial variation or “sampling error” associated with each new sample  $\{y_i\}$  that should also be accounted for in the variance estimate for  $\hat{x}_i$ . As noted in the Discussion, we do not expect (in general) that this omission would have large effects on our estimates and conclusions regarding precisions of abundance estimates, in particular for annual estimates.

#### Application of blind-test data

The blind-test data provided by Banks et al. (in review) are shown in Table 1. Sample sizes were large for true fall and winter run, whereas very few spring run were collected, particularly from Mill and Deer Creek. After examining the data in relation to ONCOR assignment probabilities,

we determined that there was merit in splitting the blind-test data into two groups: (A) run assignments with ONCOR probabilities = 1; and (B) run assignments with ONCOR probabilities < 1. Modified versions of these two datasets, which form the basis for our run-assignment corrections, are shown in Table 2. The rationale for splitting the data is evident in the assignments for true fall run and true winter run (Table 2). For example, of the 295 true fall run with ONCOR assignment probabilities = 1, only one fish (0.3%) was misclassified (assigned as late-fall run). However, of the 46 true fall run with ONCOR probabilities < 1, five fish (10.9%) were misclassified. Similarly, assignments for true winter run were very accurate for ONCOR probabilities = 1, but were incorrect in all cases for ONCOR probabilities < 1. Thus, we split the data to account for these different error rates associated with ONCOR assignment probabilities (the data were insufficient to justify further stratification by ONCOR assignment probabilities).

Due to limited data, we modified the blind-test data as follows (see Table 2). First, there were insufficient numbers of true spring Mill-Deer creek fish ( $n = 2$ ; Table 1) and true spring-Butte Creek fish with assignment probabilities < 1 (only 1 fish) to provide meaningful estimates of error rates for these groups. We therefore pooled all true spring run assignments ( $n = 15$ ) and assumed they were equally applicable to both Butte Creek and Mill-Deer creek spring runs for both assignment-probability categories (Table 2). Second, for assignments with ONCOR probabilities < 1, no true winter-run fish was correctly identified as winter run and no fish of a different run was incorrectly assigned as winter run. This resulted in a row of zeros in the matrix  $\mathbf{P}$ , which violates a condition of matrix inversion. We therefore added one true winter run fish as being correctly assigned to winter run in this probability category (Table 2).

Table 3 shows the conditional probabilities  $P(i|j)$  (i.e., the probability that a fish of true run  $j$  is assigned to run  $i$ ) based on the two assignment tables (Table 2). The non-diagonal entries in Table 3 are the false positive error rates (when interpreted across a given row) and false negative error rates (when interpreted down a given column). The largest error rates were observed for assignments of true late-fall run, which were often misclassified as fall run regardless of ONCOR assignment probability.

In our application, we were interested in sample estimates of true numbers by run ( $\hat{x}_i$ ) summed across both categories of ONCOR assignment probability:

$$(9) \quad \hat{\mathbf{x}} = \hat{\mathbf{P}}_A^{-1} \mathbf{y}_A + \hat{\mathbf{P}}_B^{-1} \mathbf{y}_B ,$$

where subscripts “A” and “B” distinguish between ONCOR assignment probabilities = 1 and < 1, respectively. Thus, a sample of assignments  $\mathbf{y}$  provided two possible subsets of assignments ( $\mathbf{y}_A$  and  $\mathbf{y}_B$ ) depending on the data. The estimates of the conditional probability matrices ( $\mathbf{P}_A$  and  $\mathbf{P}_B$ ) are given in Table 3.

Table 1. Classification table of blind-test data for the number of fish by true known run and DNA run assignment.

		True run				
		Fall	Late-fall	Spring Butte	Spring Mill-Deer	Winter
DNA Assignment	Fall	333	34	1	1	2
	Late-fall	6	40	0	0	4
	Spring Butte	0	0	12	0	1
	Spring Mill-Deer	2	2	0	1	1
	Winter	0	1	0	0	168
	Total	341	77	13	2	176

Table 2. Modified classification table of the number of fish by true known run and DNA run assignment for two categories of ONCOR assignment probability.

		True run				
		Fall	Late-fall	Spring Butte *	Spring M-D *	Winter
<u>(A) ONCOR Assignment Probability = 1</u>						
DNA Assignment	Fall	294	13	2	2	1
	Late-fall	1	16	0	0	1
	Spring Butte	0	0	13	0	0
	Spring Mill-Deer	0	0	0	13	1
	Winter	0	1	0	0	168
	Total	295	30	15	15	171
<u>(B) ONCOR Assignment Probability &lt; 1</u>						
DNA Assignment	Fall	39	21	2	2	1
	Late-fall	5	24	0	0	3
	Spring Butte	0	0	13	0	1
	Spring Mill-Deer	2	2	0	13	0
	Winter	0	0	0	0	1**
	Total	46	47	15	15	6

\* Column data are combined true spring-run assignments (Butte and Mill-Deer, both ONCOR probability categories)

\*\* Not observed; one fish added to allow matrix inversions (see text)

Table 3. Conditional probabilities of run assignments (for a fish of a given true run) for two categories of ONCOR assignment probability. Off-diagonal elements correspond to false positive error rates (rows: probability of wrongly assigning a different run to be the run of interest) and false negative error rates (columns: probability of wrongly assigning the run of interest to a different run).

		True run				
		Fall	Late-fall	Spring Butte	Spring M-D	Winter
<u>(A) ONCOR Assignment Probability = 1</u>						
DNA Assignment	Fall	0.997	0.433	0.133	0.133	0.006
	Late-fall	0.003	0.533	0.000	0.000	0.006
	Spring Butte	0.000	0.000	0.867	0.000	0.000
	Spring Mill-Deer	0.000	0.000	0.000	0.867	0.006
	Winter	0.000	0.033	0.000	0.000	0.982
	<u>(B) ONCOR Assignment Probability &lt; 1</u>					
DNA Assignment	Fall	0.848	0.447	0.133	0.133	0.167
	Late-fall	0.109	0.511	0.000	0.000	0.500
	Spring Butte	0.000	0.000	0.867	0.000	0.167
	Spring Mill-Deer	0.043	0.043	0.000	0.867	0.000
	Winter	0.000	0.000	0.000	0.000	0.167

The variance estimate for a given run-specific estimate  $\hat{x}_i$  is given by

$$(10) \quad \hat{\sigma}_{\hat{x}_i}^2 = \mathbf{y}'_A \hat{\mathbf{Q}}_{Ai} \mathbf{y}_A + \mathbf{y}'_B \hat{\mathbf{Q}}_{Bi} \mathbf{y}_B ,$$

where estimates of the variance-covariance matrices ( $\mathbf{Q}_A$  and  $\mathbf{Q}_B$ ) were derived using the bootstrap procedure described above.

As an example of the magnitude and direction of run-assignment corrections, assume we had a sample of observed assignments with ONCOR probabilities = 1 and assignment numbers {100, 10, 5, 5, 5} for fall, late-fall, Butte Creek, Mill-Deer creek, and winter run, respectively. The corrected estimates, which are given by the first term in Equation (9) and then rounded to whole numbers, are {91, 18, 6, 6, 4} fish respectively. In this example, using the blind-test data (Table 3) to correct observed assignments had the largest numerical effect on assignments for fall and late-fall-runs, though proportional effects were also large for the numerically small assignments (e.g., the winter-run assignment changed from 5 to 4 fish, a 20% decline).

#### Treatment of negative estimates

Estimates of corrected assignments were often negative for the late-fall and spring Mill-Deer creek runs, in particular when these runs had very low observed assignments ( $y_i$ ) relative to fall run. In such cases, we (1) set negative values of  $\hat{x}_i$  to zero, (2) computed the sum  $\sum \hat{x}_i$ , and (3) multiplied the corrected assignments by the ratio  $\sum y_i / \sum \hat{x}_i$ . This procedure ensured that the total number of final corrected assignments was equal to the number of observed assignments ( $\sum y_i$ ). For example, suppose a set of 24 observed assignments yielded corrected estimates {20, -8, 4, 4, 4}. After setting -8 to zero,  $\sum \hat{x}_i = 32$  (i.e., the non-negative corrections contain eight more fish than were actually observed). Multiplying  $\{ \hat{x}_i \}$  by  $\sum y_i / \sum \hat{x}_i (= 24/32 = 0.75)$  yields the final corrected estimates {15, 0, 3, 3, 3}. In sum, we set negative corrections to zero and scaled the remaining corrections in a manner that retained their relative proportions and ensured consistency between total numbers of observed and corrected assignments.

#### Stratification and sums across strata

To estimate abundances by run, we first stratified catch and sample data into biweekly time periods and several fork-length strata. This was done for the following reasons. First, because not all juveniles caught in the Chipps Island trawl were tissue sampled and DNA-assigned to run, some level of temporal stratification was required to expand sample estimates to the total catch. After inspecting the data, we chose biweekly periods for stratification because they provided a reasonable balance between ensuring sufficient sample sizes (i.e., to reliably apply assignment corrections and estimate run components in the unanalyzed catch) and depicting seasonal migration patterns by run. Second, with respect to fork length, we expect differences in juvenile length by run at Chipps Island (e.g., Fisher 1992). Consequently, the original sampling plan (Attachment A) recommended targeting larger juveniles during specific periods to improve

estimates for the winter and spring runs. Although such length-selective tissue sampling was largely abandoned due to limited catches, it is still useful from a statistical perspective to stratify by length because we expect differences across length strata in both run composition and sampling fractions (due to chance or non-random length sampling), in particular for larger length classes where we expect few fish but potentially high proportions of spring, winter, or late-fall run.

Two “biweekly” periods were defined for each month, with days 1-15 forming the first period and the remaining days forming the second period (e.g., days 16-30 for April or days 16-31 for May). For fish length, we defined six strata: < 80 mm; 80-89 mm; 90-99 mm; 100-109 mm; 110-119 mm; and  $\geq 120$  mm. In addition, data and estimates were summarized for each of four “sampling years” (2008-2011), which were defined from August 1 of the previous sampling year through July 31 of the sampling year (e.g., the 2008 sampling year comprised the period from August 1, 2007 through July 31, 2008).

Estimates of corrected run assignments ( $\hat{x}_i$ ; Equation 9) were computed for a given biweekly period and length stratum using the sample DNA assignments for that period-length combination. We then estimated biweekly totals across length strata for each run; these are the primary estimates we report. Let  $t$  denote the  $t$ th biweekly period and  $k$  denote the  $k$ th length stratum ( $k = 1, 2, \dots, 6$ ). For a given run  $i$ , we are interested in the biweekly sum of corrected assignments across length strata:

$$(11) \quad \hat{x}_{it.} = \sum_k \hat{x}_{itk} \quad ,$$

which has a variance estimate given by (e.g., Mood et al. 1974, p. 178)

$$(12) \quad \begin{aligned} \hat{\sigma}_{\hat{x}_{it.}}^2 &= \sum_k \hat{\sigma}_{\hat{x}_{itk}}^2 + 2 \sum_k \sum_{l \neq k} \text{cov}[\hat{x}_{itk}, \hat{x}_{itl}] \\ &= \sum_k \hat{\sigma}_{\hat{x}_{itk}}^2 + 2 \sum_k \sum_{l \neq k} (\mathbf{y}'_{Atk} \hat{\mathbf{Q}}_{Ai} \mathbf{y}_{Atl} + \mathbf{y}'_{Btk} \hat{\mathbf{Q}}_{Bi} \mathbf{y}_{Btl}) . \end{aligned}$$

The second term of Equation (12) is the sum of estimated covariances (across all combinations of length strata  $k$  and  $l$ , where  $l \neq k$ ) that arise because all assignment corrections ( $\hat{x}_{itk}$ ) are based on the same estimates of  $\mathbf{P}_A$  and  $\mathbf{P}_B$  (i.e., they are not independent because they are based on the same blind-test data).

We also computed sums of corrected assignments across biweekly periods for each sample year. These estimates were analogous to those above, but with summations across all period and length stratum combinations.

### Total catch estimates by run

Estimates of total catch by run were obtained by expanding the corrected assignments by the fraction of catch that was sampled and DNA analyzed (e.g., Figure 2). Let  $C_{tk}$  be the trawl catch for biweekly period  $t$  and length stratum  $k$ , let  $S_{tk}$  be the sample of catch that was DNA assigned to run, and let  $f_{tk} (= S_{tk} / C_{tk})$  be the fraction of catch that was sampled and assigned to run. For cases where  $f_{tk} < 1$  (i.e., not all of the catch was assigned to run), we want to estimate the true abundance of each run in the catch. We denote this “total catch” for the  $i$ th run as  $X_{itk}$ . When deriving estimators for  $X_{itk}$  and its variance, there are two processes to consider: (1) the sampling of catch, which determines the distribution of true abundances,  $x_{itk}$ , in the sample; and (2) the estimation of  $x_{itk}$  given the observed run assignments  $\{y_{itk}\}$  and blind-test data. As detailed in Appendix A, the combination of these processes provides the following estimate of  $X_{itk}$ :

$$(13) \quad \hat{X}_{itk} = \frac{\hat{x}_{itk}}{f_{tk}}$$

with an approximate variance given by (Appendix A)

$$(14) \quad \hat{\sigma}_{\hat{X}_{itk}}^2 = \frac{1}{f_{tk}^2} \left( \hat{\sigma}_{\hat{x}_{itk}}^2 + \hat{X}_{itk} f_{tk} (1 - f_{tk}) \frac{C_{tk} - \hat{X}_{itk}}{C_{tk} - 1} \right).$$

Again, we were primarily interested in biweekly totals (by run) across all length strata:

$$(15) \quad \hat{X}_{it.} = \sum_k \hat{X}_{itk} .$$

As was the case for sums of assignment corrections (see Equation (12)), the variance estimate for  $\hat{X}_{it.}$  needs to account for covariances among estimates due to the use of blind-test data:

$$(16) \quad \begin{aligned} \hat{\sigma}_{\hat{X}_{it.}}^2 &= \sum_k \hat{\sigma}_{\hat{X}_{itk}}^2 + 2 \sum_k \sum_{l \neq k} \text{cov}[\hat{X}_{itk}, \hat{X}_{itl}] \\ &= \sum_k \hat{\sigma}_{\hat{x}_{itk}}^2 + 2 \sum_k \sum_{l \neq k} \frac{1}{f_{tk}} \frac{1}{f_{tl}} \text{cov}[\hat{x}_{itk}, \hat{x}_{itl}] \\ &= \sum_k \hat{\sigma}_{\hat{x}_{itk}}^2 + 2 \sum_k \sum_{l \neq k} \frac{1}{f_{tk}} \frac{1}{f_{tl}} (\mathbf{y}'_{Atk} \hat{\mathbf{Q}}_{Ai} \mathbf{y}_{Atl} + \mathbf{y}'_{Btk} \hat{\mathbf{Q}}_{Bi} \mathbf{y}_{Btl}). \end{aligned}$$

Similar estimators were used for annual sums of total catch estimates by run, but with summations across all biweekly periods as well as length strata.

As a baseline for comparison, we also computed total catch estimates based on observed run assignments. To obtain these estimates, we replaced  $\hat{x}_{itk}$  with  $y_{itk}$  in Equation (13), and removed the variance and covariance terms for  $\hat{x}$  in Equations (14) and (16).

### Total abundance estimates by run

Estimates of the total (or “absolute”) abundance of juveniles passing Chipps Island were computed for biweekly periods and then summed for each sampling year. Derivations and assumptions for abundance estimates, in particular for variances, are detailed in Appendix A. Here, we outline the essential steps and equations.

To estimate abundance, trawl catches are expanded to account for trawl efficiency (the proportion of migrating fish that is captured when the trawl is operating) and trawl effort (e.g., the proportion of time trawled within a given period). For example, in USFWS (2006), abundances were estimated on a monthly basis by dividing total catches of juveniles by an estimate of trawl efficiency and the proportion of time trawled. In our application, this is analogous to the following equation for a given run  $i$  (as depicted in Figure 2):

$$(17) \quad \hat{N}_i = \frac{\hat{X}_i}{\hat{E}p} = \frac{\hat{x}_i}{\hat{E}pf},$$

where  $N$  denotes total abundance,  $E$  denotes trawl efficiency, and  $p$  is the proportion of time sampled. The use of Equation (17) is only illustrative (e.g., it implies a generic period with no length stratification).

In this report, we take additional steps to better account for trawl effort, and hence, we use a different notation for abundance estimates. In descriptive terms, we computed the total number of fish (all runs) that would have been caught had the trawl operated continuously throughout a biweekly period, and multiplied this amount by the estimated proportion of fish composed of run  $i$  during that period. This provided an estimate of catch for run  $i$ , expanded to account for trawl effort. Abundance was estimated by dividing this expanded catch by the trawl efficiency.

Specifically, biweekly abundances by run (across length strata) were estimated as

$$(18) \quad \hat{N}_{it} = \frac{\hat{\rho}_{it}\hat{\gamma}_t}{\hat{E}},$$

where  $\hat{\rho}_{it}$  is an estimate of the proportion of migrating juveniles composed of run  $i$  in period  $t$ , and  $\hat{\gamma}_t$  is an estimate of the total catch of juveniles (all runs) that would have been observed had the trawl operated continuously throughout the period. The estimate  $\hat{\rho}_{it}$  was given by

$$(19) \quad \hat{\rho}_{it} = \frac{\hat{X}_{it.}}{C_t},$$

where  $\hat{X}_{it.}$  is the estimated total catch for run  $i$  (across length strata; see Equation (15)) and  $C_t$  is the total observed catch of all juveniles (summed across length strata).

The estimate  $\hat{\gamma}_t$  accounted for missing days (i.e., days with no trawling) as well as variation in catch per unit effort among days. Let subscript  $d$  denote day, let  $D_t$  be the total days in biweekly period  $t$ , and let  $M_t$  be the number of missing days (where  $M_t < D_t$ ). For each day of trawling, there is an observed total catch,  $C_{td}$  (across all runs and length strata), and a computed proportion of the day trawled,  $p_{td}$ . The estimate  $\hat{\gamma}_t$  was given by

$$(20) \quad \hat{\gamma}_t = \sum_d^{D_t - M_t} \frac{C_{td}}{p_{td}} + M_t \bar{c}_t = D_t \bar{c}_t,$$

where  $\bar{c}_t$  is mean of the set of  $(D_t - M_t)$  daily observations  $\{C_{td} / p_{td}\}$ , which are analogous to catch-per-unit-effort data. Note that in our application, we estimated  $p_{td}$  as a standardized proportion of water volume trawled (see Appendix A), which matched the definition for efficiency estimates.

An approximate variance estimator for the biweekly abundance estimate (Equation (18)) is given by

$$(21) \quad \hat{\sigma}_{\hat{N}_{it}}^2 \cong \frac{\hat{\gamma}_t^2}{\hat{E}^2} \hat{\sigma}_{\hat{\rho}_{it}}^2 + \frac{\hat{\rho}_{it}^2}{\hat{E}^2} \hat{\sigma}_{\hat{\gamma}_t}^2 + \frac{\hat{\rho}_{it}^2 \hat{\gamma}_t^2}{\hat{E}^4} \hat{\sigma}_{\hat{E}}^2.$$

Expressions for the variances of  $\hat{\rho}_{it}$  and  $\hat{\gamma}_t$  are provided in Appendix A. Estimates of trawl efficiency,  $\hat{E}$ , and its variance were obtained from Pyper et al. (2013), as discussed below.

It is useful to decompose Equation (21) further with respect to the variance for  $\hat{\rho}_{it}$  (the estimate of the proportion of migrating juveniles composed of run  $i$  in period  $t$ ):

$$(22) \quad \hat{\sigma}_{\hat{N}_{it}}^2 \cong \frac{\hat{\gamma}_t^2}{\hat{E}^2 C_t^2} \left( \sum_k \sum_l \frac{1}{f_{tk}} \frac{1}{f_{tl}} \left( \mathbf{y}'_{Aik} \hat{\mathbf{Q}}_{Ai} \mathbf{y}_{Ail} + \mathbf{y}'_{Bik} \hat{\mathbf{Q}}_{Bi} \mathbf{y}_{Bil} \right) + \sum_k \frac{\hat{x}_{itk}}{f_{tk}^2} \right) + \frac{\hat{\rho}_{it}^2}{\hat{E}^2} \hat{\sigma}_{\hat{\gamma}_t}^2 + \frac{\hat{\rho}_{it}^2 \hat{\gamma}_t^2}{\hat{E}^4} \hat{\sigma}_{\hat{E}}^2$$

The four additive terms in Equation (22) correspond respectively to (1) measurement error in DNA assignment corrections; (2) sampling variation in trawl captures of the run of interest; (3) variance in biweekly catch due to missing days; and (4) variation in the estimate of trawl efficiency. Potentially large components of variation have been omitted, specifically, temporal variation in efficiency and/or catch (e.g., overdispersion due to clumpy spatial and/or temporal patterns of fish migration).

Last, we computed annual sums of abundance by run across biweekly periods. The estimators for annual abundance are provided in Appendix A.

As a baseline for comparison, we also computed abundance estimates based on observed run assignments. To obtain these estimates, we replaced corrected assignments ( $\hat{x}$ ) with observed assignments ( $y$ ) and removed all variance and covariance terms for  $\hat{x}$ .

Estimates of efficiency

Total abundances were estimated using four different estimates of Chipps Island trawl efficiency as reported in Pyper et al. (2013). In Pyper et al. (2013), data for paired-release tests (across numerous years) were used to examine relationships between estimated efficiency and potential covariates (e.g., run, fork length, temperature, turbidity and flow). They found little evidence of such relationships, and concluded that variation in efficiencies (across tests and years) was largely driven by confounding effects of differing ocean recovery rates (e.g., survival rates) between control releases and fish of upstream releases passing Chipps Island. Following their recommendation, we assumed that trawl efficiency was constant across periods and years. Thus, one estimate of efficiency we used was the mean efficiency across paired-release tests (Table 4).

Pyper et al. (2013) also evaluated efficiency estimates based on proximal releases (i.e., releases made close to Chipps Island) for the Jersey Point and Pittsburg release locations. These estimates differed, but both were considerably higher than the mean efficiency for paired-release tests (Table 4). As discussed in Pyper et al. (2013), it was unclear which efficiency estimate was preferable; there were advantages and disadvantages to each of three datasets (paired-release tests, Jersey Point, and Pittsburg) and the different methods used to estimate efficiency. We therefore compared abundance estimates based on each of the three efficiency estimates. However, we chose the Jersey Point estimate (the midrange) as the baseline for comparisons.

Table 4. Estimates of Chipps Island trawl efficiency and standard errors (SE) as reported in Pyper et al. (2013). All estimates assume (i.e., are standardized to) a volume-sampled rate of 1000 m<sup>3</sup>/minute (based on volume measurements in the current trawl database). No standard error is provided for the fish-flux method, which is based on a set of assumed constants.

	Estimate	SE
Paired-release tests	0.0064	0.0007
Jersey Point releases	0.0088	0.0018
Pittsburg releases	0.0124	0.0016
Fish flux	0.04	-

The fourth approach we used to estimate abundance was the “fish flux” method of Kimmerer (2008). This is an expansion method in which trawl catch is expanded by the ratio of the volume of water trawled versus the (assumed) volume of water that a migrating fish occupies. As detailed in Pyper et al. (2013), the fish-flux method has an implied (constant) efficiency, which was estimated to be 0.041 or 0.042 depending on the vessel used (Confluence or Whitesel). For simplicity, we assumed an efficiency of 0.04 for the fish-flux method (Table 4). Because the fish-flux method is based on a set of assumed constants (e.g., average migration speed and the

channel width and water depth occupied by a migrating fish), there is no variance for this efficiency.

An important detail for all efficiency estimates (Table 4) is that they are standardized by volume sampled, that is, they are estimates of efficiency when the trawl is fishing at a rate of 1000 m<sup>3</sup>/minute. This convention was adopted by Pyper et al. (2013) to standardize trawl effort as a function of volume sampled rather than time sampled (see Pyper et al. 2013 for the rationale and evidence in support of using volume rather than time), and was accounted for when estimating abundances because daily values of trawl effort ( $p_{td}$ ) were standardized accordingly (see Appendix A).

As noted earlier, the effective-fishing mouth size of the trawl net fished at Chipps Island was recently estimated to be 12.7 m<sup>2</sup> (Whitesel) or 13.0 m<sup>2</sup> (Confluence), as opposed to the value of 18.5m<sup>2</sup> that is currently used to compute volume sampled in the trawl database (the data used here). However, such a change would not affect our abundance estimates, except in the case of the fish-flux method. The paired- and proximal-release efficiencies were derived using the same volumes and standardized effort as used here, so a simple scalar change to all trawl volumes would not affect abundance estimates (only the definition of “standardized effort” would change). In contrast, the fish-flux method depends on actual volumes sampled. To correctly apply the fish-flux method to current database volumes, Pyper et al. (2013) adjusted implied efficiency by the ratio of recent (“correct”) versus database (“incorrect”) estimates of net-mouth areas. (Note that similar three-decimal values for implied efficiency are obtained for both the recent estimate of 12.7 m<sup>2</sup> for the Whitesel (0.041) and 13.0m<sup>2</sup> (0.042) for the Confluence.)

## Results

### Summaries of trawl effort, catch, and DNA samples

Trawl effort at Chipps Island, as measured by hours sampled, was reasonably similar across biweekly periods for sample years 2009-2011 (Table 5, Figure 3). By comparison, effort in the 2008 sample year was relatively high during December/January, but low during February/March (there was no trawling from February 5 through March 10 due to concerns for delta smelt). In most biweekly periods and sample years, trawling occurred in less than half of the total days available.

Total catches and DNA sample numbers are reported for biweekly periods in Table 6. Catches were low in all biweekly periods from August through March, with a high of 27 fish caught in the March 16-31, 2010 period. Beginning in April, catches steadily increased in most years and peaked in either the April 16-30 or May 1-15 period. Total annual catches increased in each sample year. From June through mid-April, the fraction of catch that was tissue sampled and DNA-assigned to run was typically greater than 80%. However, during the periods of relatively high catch from mid-April through mid-May, the fraction of catch that was DNA analyzed was

often less than 50%. Overall, a much larger fraction of catch was DNA analyzed in sample years 2008 and 2011 compared to 2009 and 2010.

Fork length distributions for DNA-analyzed and non-analyzed fish were similar for most biweekly periods in which there was partial sampling of catch (Figure 4). However, notable differences were evident in some March and April periods, and in general, the largest fish caught (i.e., high outliers in fork length) were predominantly found among DNA-analyzed fish. This result is not surprising given that the smaller, fall run were sometimes subsampled whereas the larger fish were not.

### **Observed DNA run assignments**

There were stark differences among runs in the ONCOR probability categories ( $P < 1$  or  $P = 1$ ) for DNA assignments. Observed DNA run assignments, tabulated by biweekly period and ONCOR probability category, are shown in Table 7 for all sample years combined. Assignments for winter-run had the highest proportion (98%) of ONCOR probabilities = 1, followed by Butte Creek spring run (70%) and fall run (63%). In contrast, proportions of ONCOR probabilities = 1 were very low for assignments of late-fall run (5%) and Mill-Deer creek spring run (18%). In the case of fall run, there was a much lower proportion of ONCOR probabilities = 1 from July to March (i.e., 23 of 88 assignments, or 26%) than from April to June (64%), suggesting that fall-run assignments were less certain in months outside the peak migration period (April to June). Across years, total DNA assignments ( $n = 5104$ ) were dominated by fall run (4,326; 84.8%), followed by Butte Creek spring run (301; 5.9%), late-fall run (272; 5.3%), winter run (105; 2.1%), and Mill-Deer creek spring run (100; 2.0%).

Relationships between fork length and capture date of fish that were DNA assigned to run are shown in Figure 5 (the top panel highlights fall and late-fall runs; the bottom panel highlights spring and winter runs). Fish assigned to the fall and late-fall runs prior to the spring emigration period had relatively high fork lengths. Winter-run assignments were more confined in length and time, and were generally larger in size, than spring-run assignments. A greater number of spring Mill-Deer creek assignments overlapped in length and time with winter run than did Butte Creek assignments. The peak for winter-run assignments occurred in mid-March, whereas the peak for the both spring runs occurred in April.

Table 5. Chippis Island trawl effort summaries by biweekly period and sampling year. A “-” indicates that no trawling was conducted during this strata. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011. Sampling effort summaries are not presented from August 1<sup>st</sup> – September 30<sup>th</sup>, 2007, which preceded the onset of DNA sampling and from July 1- 30<sup>th</sup>, 2011 when DNA sampling concluded.

Period	Days	Days of trawling				Minutes trawled				Volume trawled (thousands of m <sup>3</sup> )			
		2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15	15		4	6	6		780	1,200	1,120		846	1,376	1,249
Aug 16-31	16		4	7	7		793	1,400	1,277		829	1,595	1,437
Sep 1-15	15		5	6	7		1,000	1,100	1,320		1,104	1,202	1,365
Sep 16-30	15		4	7	6		799	1,400	1,140		847	1,521	1,175
Oct 1-15	15	6	4	6	7	1,196	780	1,100	1,286	1,296	806	1,204	1,366
Oct 16-31	16	7	5	7	6	1,400	1,000	1,360	1,200	1,413	980	1,458	1,289
Nov 1-15	15	7	4	6	7	1,340	800	1,182	1,355	1,330	824	1,245	1,296
Nov 16-30	15	6	4	7	6	1,100	700	1,300	1,100	1,139	664	1,426	1,067
Dec 1-15	15	12	6	6	7	2,210	1,080	1,140	1,280	2,232	1,048	1,262	1,483
Dec 16-31	16	15	5	8	6	3,015	920	1,506	1,100	3,087	889	1,619	1,097
Jan 1-15	15	5	6	6	6	960	1,160	1,200	1,155	919	1,194	1,347	1,283
Jan 16-31	16	15	7	4	6	2,772	1,360	795	1,080	2,885	1,417	854	1,190
Feb 1-15	15	4	6	6	7	520	1,200	1,200	1,400	571	1,329	1,313	1,565
Feb 16-28	13	-	6	6	6	0	1,200	1,200	1,200	-	1,290	1,271	1,392
Mar 1-15	15	2	6	7	6	315	1,215	1,400	1,200	341	1,264	1,561	1,327
Mar 16-31	16	5	7	7	7	980	1,400	1,396	1,400	1,045	1,382	1,571	1,440
Apr 1-15	15	4	8	6	7	800	2,280	1,181	1,400	900	2,262	1,302	1,426
Apr 16-30	15	4	7	7	6	780	1,830	1,380	1,200	853	1,813	1,482	1,323
May 1-15	15	5	8	6	6	1,000	1,865	1,180	1,200	1,053	1,893	1,300	1,160
May 16-31	16	4	6	6	7	745	1,140	1,120	1,260	810	1,259	1,159	1,405
Jun 1-15	15	6	7	7	7	1,200	1,360	1,398	1,400	1,308	1,430	1,532	1,521
Jun 16-30	15	7	6	7	6	1,380	1,200	1,298	1,220	1,543	1,326	1,515	1,374
Jul 1-15	15	4	7	6		660	1,300	1,120		749	1,476	1,297	
Jul 16-31	16	5	7	6		810	1,380	912		914	1,571	1,030	
Total	365	123	139	153	142	23,183	28,542	29,468	27,293	24,388	29,743	32,442	29,230

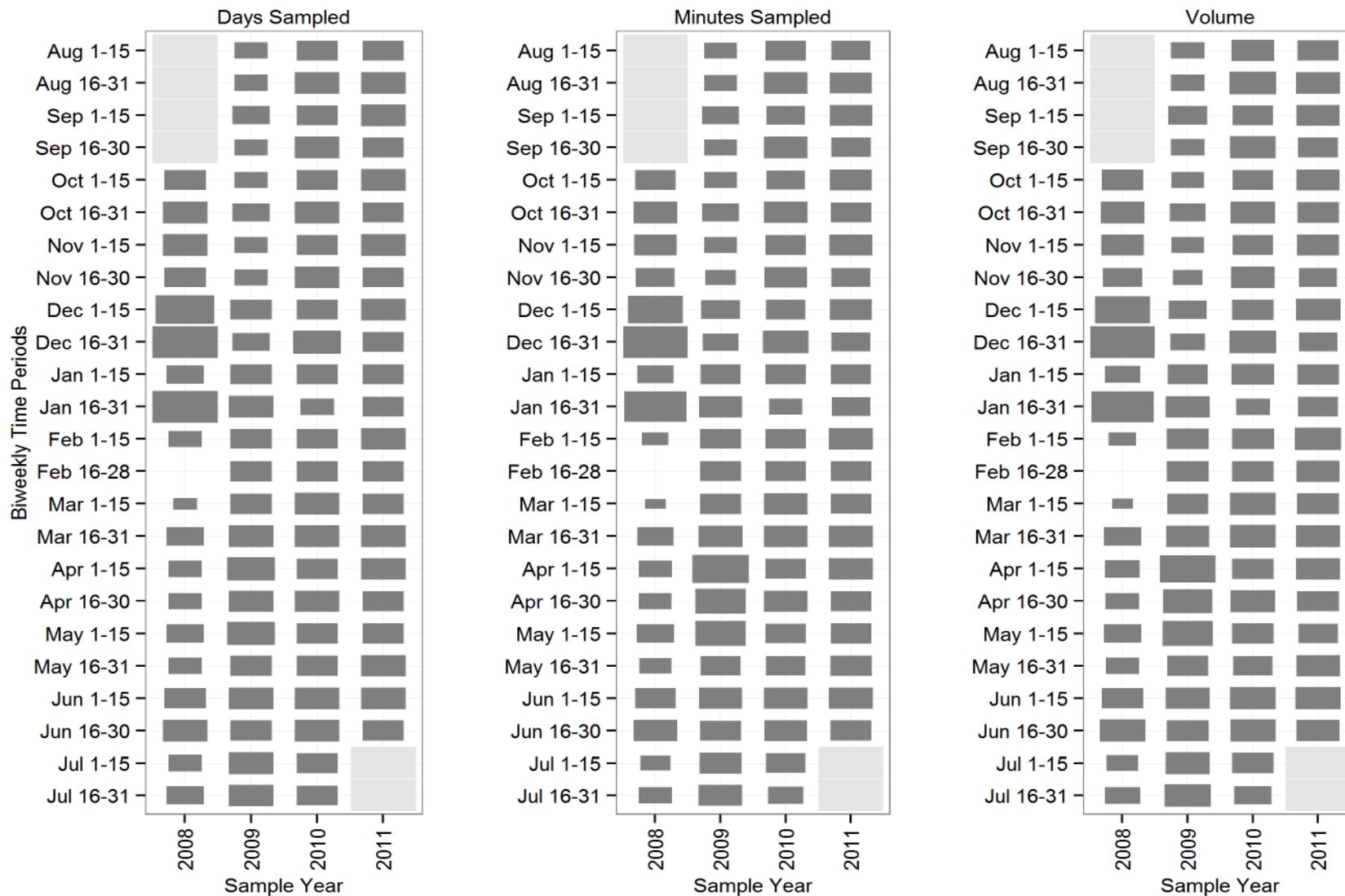


Figure 3. Graphical display of Chipps Island trawl effort summaries by biweekly period and sampling year. The size of each rectangle is proportional to the value in Table 5. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

Table 6. Raw catch and number of DNA samples (assigned to run) taken at Chipps Island by biweekly period and sampling year. A “-” indicates that no sampling (trawl or DNA) was conducted during this period. Blank entries indicate either zero catch or no DNA samples. Catch is not reported from August-September in sample year 2008 and July in sample year 2011 because no DNA samples were taken during these periods.

	Catch ( <i>C</i> )				DNA samples analyzed ( <i>S</i> )				Fraction analyzed ( $f = S/C$ )			
	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		1	0	2	-	1		2	-	1.00		1.00
Aug 16-31		0	0	1	-			1	-			1.00
Sep 1-15		0	1	2	-		1	2	-		1.00	1.00
Sep 16-30		0	1	0	-		1		-		1.00	
Oct 1-15	1	0	0	1	0			1	0.00			1.00
Oct 16-31	0	2	0	0		0				0.00		
Nov 1-15	1	2	3	3	1	1	3	3	1.00	0.50	1.00	1.00
Nov 16-30	1	0	0	1	1			1	1.00			1.00
Dec 1-15	9	1	0	1	6	1		1	0.67	1.00		1.00
Dec 16-31	10	0	2	4	10		2	4	1.00		1.00	1.00
Jan 1-15	6	6	1	3	5	6	1	3	0.83	1.00	1.00	1.00
Jan 16-31	14	0	4	1	8		4	0	0.57		1.00	0.00
Feb 1-15	0	1	6	1		1	6	1		1.00	1.00	1.00
Feb 16-28	-	12	7	6	-	10	7	6	-	0.83	1.00	1.00
Mar 1-15	5	24	24	7	5	16	23	7	1.00	0.67	0.96	1.00
Mar 16-31	17	24	27	18	17	23	24	17	1.00	0.96	0.89	0.94
Apr 1-15	27	122	62	67	25	111	57	66	0.93	0.91	0.92	0.99
Apr 16-30	134	690	1,740	845	124	268	462	699	0.93	0.39	0.27	0.83
May 1-15	316	1,059	764	1,763	141	198	307	790	0.45	0.19	0.40	0.45
May 16-31	45	71	371	687	33	50	235	447	0.73	0.70	0.63	0.65
Jun 1-15	40	37	194	641	39	32	170	424	0.98	0.86	0.88	0.66
Jun 16-30	28	16	37	144	26	16	33	86	0.93	1.00	0.89	0.60
Jul 1-15	1	11	12		1	10	12	-	1.00	0.91	1.00	-
Jul 16-31	1	6	2		1	6	2	-	1.00	1.00	1.00	-
Total	656	2,085	3,258	4,198	443	750	1,350	2,561	0.68	0.36	0.41	0.61

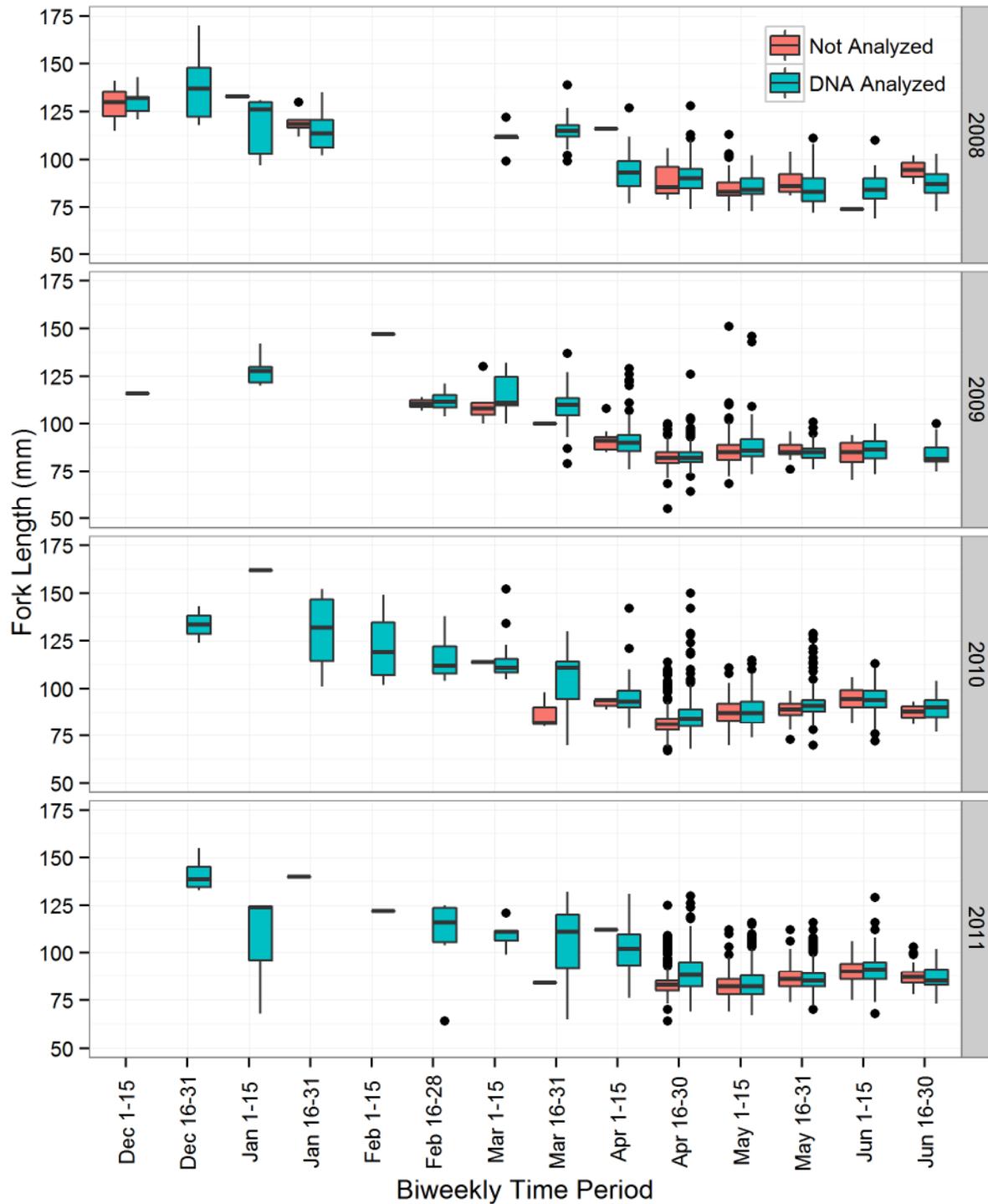


Figure 4. Boxplots of fork length (mm) by period and sample year for DNA-analyzed and not-analyzed juvenile Chinook salmon caught in Chipps Island trawl. Boxplots show medians (horizontal lines), 25th-75th percentiles (boxes),  $1.5 \times (75\text{th} - 25\text{th percentiles})$  (whiskers). Data points beyond the end of the whiskers are outliers. Sample years include December from the previous year (2008, includes December of 2007, etc.).

Comparisons with assignments based on length-at-date criteria

Total run assignments based on DNA often differed substantially with those based on the river model’s length-at-date criteria (Fisher 1992). Across all four years, more fish were DNA assigned to the fall and late-fall runs, and far fewer fish were DNA assigned to spring and winter runs, than compared to length-at-date assignments (Figure 6). These trends were consistent for both observed and corrected DNA assignments (Figure 6) (corrected assignments are reported below). Scatter plots of length and capture date for observed DNA assignments illustrate the differences between DNA assignments and their expected length-at-date ranges (Figure 7). For example, many fish that were DNA assigned to fall run were contained within the length-at-date range for spring run, and to a lesser extent, within the late-fall and winter run ranges. The length-at-date distributions for late-fall DNA-assignments were similar to those for fall-run assignments, and were spread across the length-at-date ranges for all runs. Spring-run DNA assignments were centered in the spring-run length-at date range, but also overlapped considerably with the adjacent winter and fall-run length-at-date ranges. Winter-run DNA assignments had the closest fit to their expected length-at-date range, with only a few fish overlapping the adjacent late-fall and spring-run ranges (Figure 7).

Table 7. Number of fish assigned to run by biweekly period and ONCOR assignment-probability category. All sample years are combined.

	ONCOR assignment probability (P < 1 or P = 1)									
	Fall		Late-Fall		Spring Butte		Spring Mill-Deer		Winter	
	<1	=1	<1	=1	<1	=1	<1	=1	<1	=1
Aug 1-15	1		2							
Aug 16-31	1									
Sep 1-15	1	1	1							
Sep 16-30		1								
Oct 1-15			1							
Oct 16-31										
Nov 1-15	3	2	3							
Nov 16-30	1		1							
Dec 1-15	2	1	4							1
Dec 16-31	6	2	6	2						
Jan 1-15	5	1	5	1						3
Jan 16-31	7		3							2
Feb 1-15	2	1	2							3
Feb 16-28	1		3				1		1	17
Mar 1-15	9		5				5	3		29
Mar 16-31	12	10	6	1	1	3	10	5	1	32
Apr 1-15	61	92	8		24	54	9	2		9
Apr 16-30	486	825	55		42	108	28	2		7
May 1-15	480	839	52		11	33	20	1		
May 16-31	221	507	21	1	6	6	2	1		
Jun 1-15	216	379	49	2	6	5	5	3		
Jun 16-30	53	79	21	3	1	1	2	1		
Jul 1-15	9	4	7	3						
Jul 16-31	5		3	1						
Total	1,582	2,744	258	14	91	210	82	18	2	103
% of total	37%	63%	95%	5%	30%	70%	82%	18%	2%	98%



Figure 5. Scatterplot of juvenile Chinook salmon caught in Chipps Island trawl and DNA-assigned to run as a function of fork length and sample day (all four years of study data combined). Run abbreviations are fall (F), late-fall (LF), spring Butte (SB), spring Mill and Deer (SMD), and winter (W).

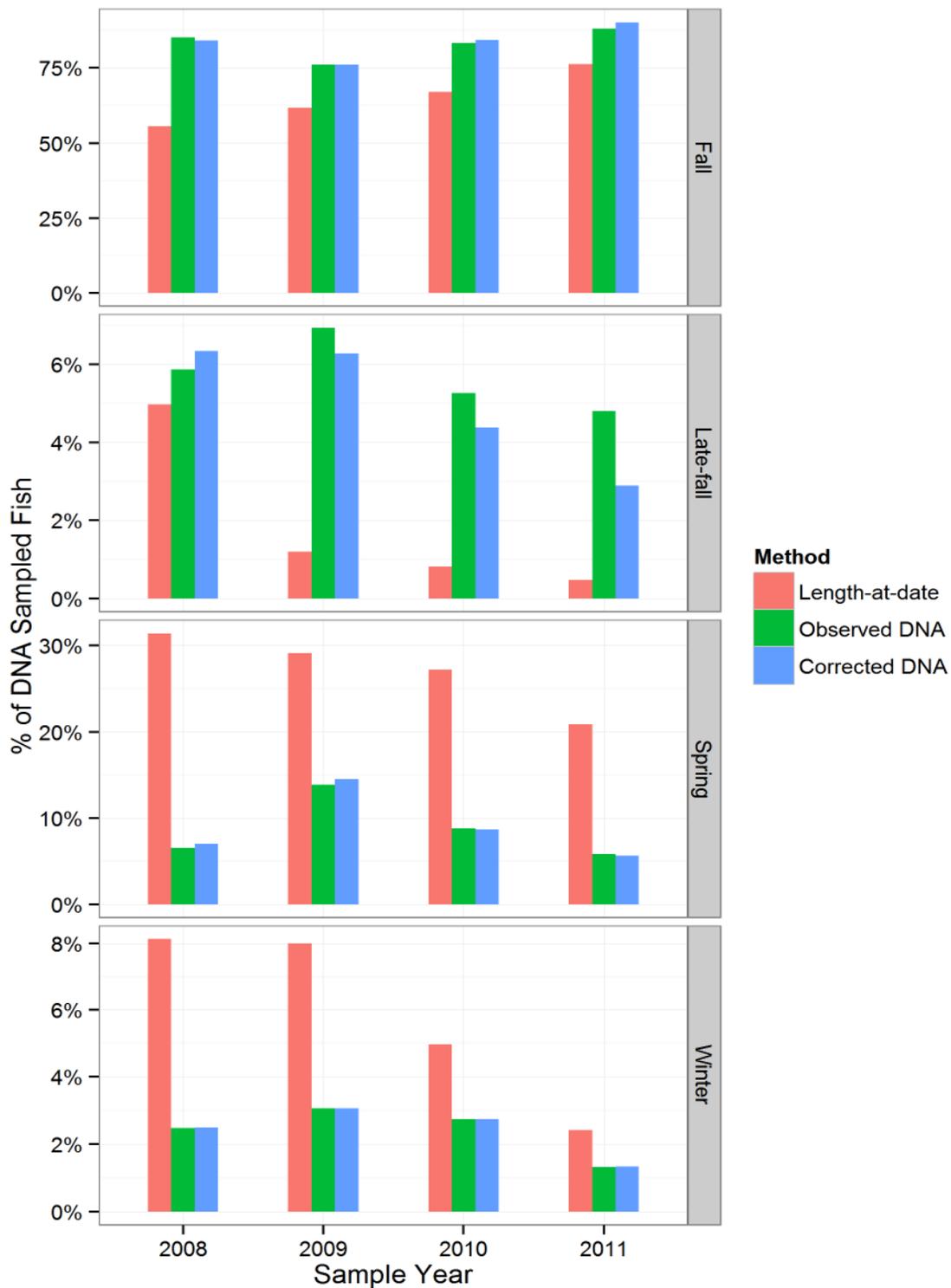


Figure 6. Comparison of run assignments based on length-at-date criteria versus DNA (observed and corrected) for juvenile Chinook salmon caught in Chipps Island trawl and DNA assigned to run. For each sample year and assignment method, the percentage of total juveniles assigned to each run is shown. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

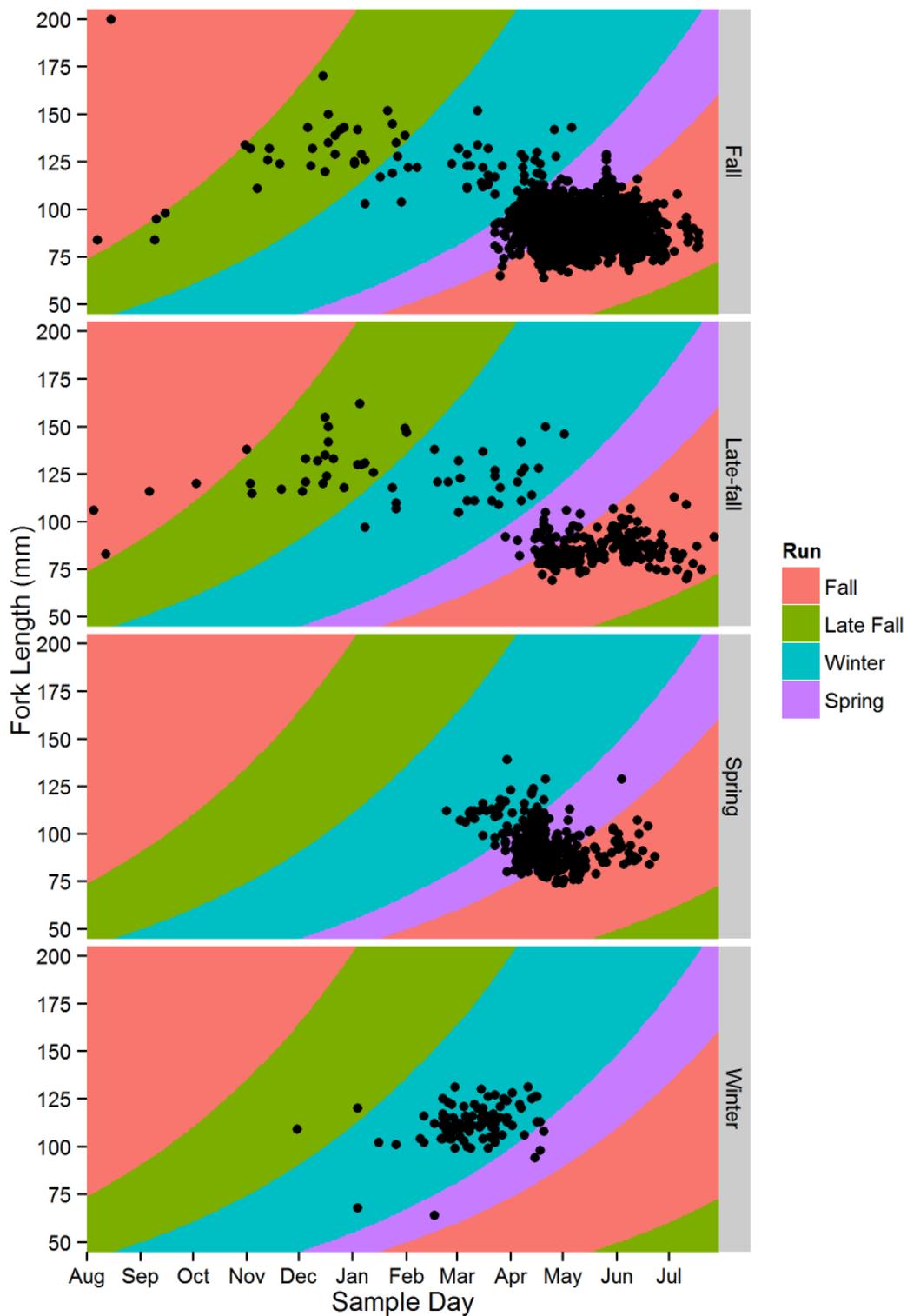


Figure 7. Scatterplot of juvenile Chinook salmon caught in Chipps Island trawl and DNA assigned to run (panels) as a function of fork length and sample day (all four years of study data combined). The color regions correspond to length-at-date criteria for run assignment.

### **Assignment corrections and total catch estimates by run**

Observed and corrected DNA assignments, and their corresponding total catch estimates, are presented for each run in Table 8-Table 12. Comparisons of observed DNA assignments ( $y$ ) and corrected assignments ( $\hat{x}$ ) show the effects of using blind-test data to “correct” assignments (i.e., accounting for empirical error rates in DNA assignments). The total catch estimates ( $\hat{X}$ ) are simple expansions that account for the proportion of trawl catch that was not DNA assigned to run. Standard errors, shown in parentheses, reflect two sources of uncertainty. First, there is estimation error in assignment corrections, and second, there is additional (hypergeometric) sampling error in estimates of total catch (i.e., for those periods in which some catch was not DNA assigned to run).

Note that because the reported estimates for biweekly periods are *sums* across length strata, the apparent expansion factors to total catch may differ for observed and corrected assignments. For example, during the May 1-15 period in 2009, there were eight observed assignments for late-fall run with a corresponding total catch estimate of 37 fish, a more than four-fold expansion (Table 9). In contrast, the corrected assignments for this period summed to only two fish, with a corresponding total catch estimate of only three fish (after rounding). This difference in the observed and corrected expansions to total catch was due to the differing proportions of catch sampled within length strata (i.e., observed assignments were more numerous in length strata with low proportions of sampled catch).

#### Fall run

Fish that were DNA assigned to fall run were observed throughout the sampling year, with the largest numbers observed in the April 16-30 and May 1-15 periods (Table 8). Two year classes are evident, with the older young of the year from the previous brood year migrating between August and April, and the current brood year migrating between April and July (Figure 7). Across years, observed and corrected assignments for fall run were reasonably similar (Table 8; see also Figure 6). The largest annual difference was in 2011, with 2,255 observed assignments and 2,309 corrected assignments (i.e., a 2.4% increase in corrected assignments relative to observed).

Estimates of total annual catch increased each year from roughly 600 fall run in 2008 to almost 4,000 in 2011. Total catch estimates based on observed and corrected assignments were similar, though standard errors were typically much larger for estimates based on corrected assignments.

#### Late-fall run

Late-fall fish were also identified throughout the sampling year, with relatively high numbers from mid-April to late June in most years (Table 9). Like fall run, late-fall run display two year classes, with the older young of the year from the previous brood year migrating between August and May, while the current brood year migrates between April and July (Figure 7). However, there were often large differences between observed and corrected DNA assignments of late-fall run. In particular, corrected assignments tended to be much lower than observed assignments in

the three biweekly periods extending from mid-April to the end of May, when fall-run abundances were high. Outside these periods, corrected assignments were typically equal to, or somewhat larger than, the observed assignments. The largest annual difference was in 2011, with 123 observed and 74 corrected assignments (a 39.8% decrease in corrected assignments relative to observed).

For sample years 2009-2011, estimates of total annual catch based on corrected assignments were considerably lower (roughly half) compared to estimates based on observed assignments. Standard errors for the corrected catch estimates were very large, reflecting the high uncertainty in estimates of corrected assignments for late-fall run.

#### Butte Creek spring run

Butte Creek spring run were identified from March through June, with relatively high numbers in the three biweekly periods extending from April to mid-May (Table 10). It is likely that spring run outmigrants during this period contain both yearlings and young-of-the-year. Corrected DNA assignments tended to be slightly higher than observed DNA assignments (e.g., annual totals of corrected assignments were roughly 10% higher than for observed assignments). Similar differences were observed for total catch estimates. For example, annual totals of corrected catch estimates were roughly 10% larger than estimates based on observed assignments for years 2009-2011. Estimates of corrected assignments and total catch for Butte Creek spring run were relatively precise (i.e., had low standard errors).

#### Mill-Deer spring run

Assignments of Mill-Deer creek spring run (Table 11) were far less numerous than for Butte Creek. The timing of peak counts for Mill-Deer creek assignments varied across years, ranging from late March (2008) to early May (2011). Spring run outmigrants during this period likely consist of both yearlings and young of the year. Corrected assignments declined each year in comparison to observed assignments. There was no difference in 2008, while the largest annual difference was in 2011, with 30 observed and 10 corrected assignments (a 63% decrease in corrected assignments relative to observed). Larger differences were observed for total catch estimates across years, with reductions in corrected catch estimates ranging from 9% (2008) to 72% (2011) relative to estimates based on observed assignments. As was the case for late-fall run, estimates of corrected assignments and total catch for Mill-Deer creek spring run were highly uncertain (large standard errors).

#### Winter run

Most assignments of winter run occurred in three biweekly periods from late February through March, though 12 of the 34 observed assignments in 2011 were identified in April (Table 12). In all periods and years, corrected assignments and total catch estimates (after rounding) were very precise and were identical to the observed values. Because assignments of winter run occurred in time periods and length strata with high proportions of sampled catch (typically 100%), total catch estimates were equal to, or only slightly larger than, the corresponding assignment

numbers. Annual total catch estimates for winter run were low, ranging from just 12 fish in 2008 to a high of 38 fish in 2010.

#### Negative estimates of corrected assignments

As detailed in the Methods section, when corrected run assignments were negative for a given biweekly period and length stratum, we set negative estimates to zero and scaled the remaining run corrections to equal the total number of observed assignments. Negative estimates can be indication of biased or insufficient blind-test data, so we note their prevalence here for each run.

The frequency of negative estimates was much greater for assignments with ONCOR probabilities  $< 1$  than for probabilities  $= 1$ . For probabilities  $< 1$ , there was a total of 180 combinations of period-length strata (across years) with at least one observed DNA assignment. Among these 180 cases, negative corrections occurred in 34 (19% of cases) for fall run, 54 (30%) for late-fall run, 2 (1%) for Butte Creek spring run, and 26 (14%) for Mill-Deer creek spring run. The negative estimates for fall run tended to occur when observed fall-run assignments were low relative to late-fall assignments (e.g., December through March), while negative estimates for the late-fall and Mill-Deer creek runs occurred when observed assignments of these runs were low relative to fall-run assignments (e.g., April through June). In contrast, among the 178 period-length combinations with ONCOR assignment probabilities  $= 1$ , negative corrections occurred in only 8 cases (4.5%) for fall run and 10 cases (5.6%) for late-fall run. There were no negative estimates for winter run in either ONCOR probability category.

Table 8. Number of fall-run juvenile Chinook salmon by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “–” indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Observed assignments				Corrected assignments				Total catch based on observed assignments				Total catch based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		1		0		1 (0.2)		0 (1.2)		1 (0)		0 (0)		1 (0.2)		0 (1.2)
Aug 16-31				1				1 (0.2)				1 (0)				1 (0.2)
Sep 1-15			1	1			1 (0)	1 (0.4)			1 (0)	1 (0)			1 (0)	1 (0.4)
Sep 16-30			1				1 (0)				1 (0)				1 (0)	
Oct 1-15				0				0 (0.6)				0 (0)				0 (0.6)
Oct 16-31																
Nov 1-15	1	1	1	2	1 (0)	1 (0.2)	0 (1)	2 (0.4)	1 (0)	2 (0)	1 (0)	2 (0)	1 (0)	2 (0.5)	0 (1)	2 (0.4)
Nov 16-30	1			0	1 (0.2)			0 (0.6)	1 (0)			0 (0)	1 (0.2)			0 (0.6)
Dec 1-15	3	0		0	1 (1.4)	0 (0.6)		0 (0)	4 (0.9)	0 (0)		0 (0)	1 (2)	0 (0.6)		0 (0)
Dec 16-31	7		1	0	6 (0.9)		0 (0.4)	0 (1.8)	7 (0)		1 (0)	0 (0)	6 (0.9)		0 (0.4)	0 (1.8)
Jan 1-15	2	2	0	2	1 (1.6)	0 (0.9)	0 (0.4)	2 (0.5)	2 (0.6)	2 (0)	0 (0)	2 (0)	1 (1.9)	0 (0.9)	0 (0.4)	2 (0.5)
Jan 16-31	4		3		2 (1.2)		3 (0.7)		7 (1.5)		3 (0)		4 (2.3)		3 (0.7)	
Feb 1-15		0	2	1		0 (0.6)	1 (0.4)	1 (0.2)		0 (0)	2 (0)	1 (0)		0 (0.6)	1 (0.4)	1 (0.2)
Feb 16-28	-	0	0	1	-	0 (0.6)	0 (2.5)	1 (0.2)	-	0 (0)	0 (0)	1 (0)	-	0 (0.6)	0 (2.5)	1 (0.2)
Mar 1-15	0	6	2	1	0 (0.2)	5 (1.1)	1 (0.7)	0 (0.4)	0 (0)	8 (1.2)	2 (0)	1 (0)	0 (0.2)	7 (2.1)	1 (0.7)	0 (0.4)
Mar 16-31	4	5	6	7	3 (0.7)	4 (2.3)	6 (0.8)	7 (3.6)	4 (0)	5 (0.4)	8 (0.5)	8 (0)	3 (0.7)	4 (2.5)	8 (1)	8 (3.7)
Apr 1-15	17	56	43	37	16 (1.1)	50 (6.2)	43 (3.4)	35 (2.7)	18 (0)	61 (1.7)	47 (1)	37 (0.5)	17 (1.1)	54 (7.1)	48 (3.8)	36 (2.8)
Apr 16-30	112	231	389	579	113 (9.1)	240 (16.1)	400 (25.1)	590 (37.9)	121 (1)	596 (11.3)	1,531 (21.6)	706 (4.2)	123 (9.8)	620 (42.3)	1,589 (101)	721 (46.6)
May 1-15	132	175	275	737	134 (8.2)	182 (12.4)	284 (19.6)	771 (47.2)	295 (4.8)	930 (23.9)	686 (10.3)	1,648 (11.4)	301 (18.6)	967 (67.1)	708 (50)	1,728 (105.3)
May 16-31	32	44	225	427	33 (2.5)	43 (3.2)	229 (12.1)	438 (23.6)	44 (0.7)	62 (1.9)	355 (3.1)	656 (4.1)	44 (3.5)	61 (5)	362 (19.9)	673 (36)
Jun 1-15	38	29	143	385	38 (3.2)	28 (2.2)	138 (8.2)	390 (20.4)	39 (0)	34 (0.6)	163 (2)	583 (5.2)	39 (3.2)	32 (2.7)	157 (9.6)	589 (31.2)
Jun 16-30	24	10	24	74	23 (0.9)	7 (2.2)	21 (1.9)	69 (4.4)	26 (0.4)	10 (0)	27 (0.9)	123 (3.6)	25 (1.2)	7 (2.2)	24 (2.3)	115 (8.5)
Jul 1-15	0	7	6		0 (0.6)	6 (1.2)	6 (1.6)		0 (0)	8 (0)	6 (0)		0 (0.6)	7 (1.2)	6 (1.6)	
Jul 16-31	0	4	1		0 (0.6)	4 (0.8)	1 (0.4)		0 (0)	4 (0)	1 (0)		0 (0.6)	4 (0.8)	1 (0.4)	
Total	377	571	1,123	2,255	372 (23.2)	571 (37.6)	1,136 (69)	2,309 (133.8)	569 (5.3)	1,723 (26.6)	2,835 (24.2)	3,770 (14.3)	566 (34.9)	1,766 (117.4)	2,910 (182.4)	3,878 (224.7)

Table 9. Number of late-fall-run juvenile Chinook salmon by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “–” indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Observed assignments				Corrected assignments				Total catch based on observed assignments				Total catch based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0		2		0 (0.2)		2 (1.2)		0 (0)		2 (0)		0 (0.2)		2 (1.2)
Aug 16-31				0				0 (0.2)				0 (0)				0 (0.2)
Sep 1-15			0	1			0 (0)	1 (0.4)			0 (0)	1 (0)			0 (0)	1 (0.4)
Sep 16-30			0				0 (0)				0 (0)				0 (0)	
Oct 1-15				1				1 (0.6)				1 (0)				1 (0.6)
Oct 16-31																
Nov 1-15	0	0	2	1	0 (0)	0 (0.2)	3 (1)	1 (0.4)	0 (0)	0 (0)	2 (0)	1 (0)	0 (0)	0 (0.4)	3 (1)	1 (0.4)
Nov 16-30	0			1	0 (0.2)			1 (0.6)	0 (0)			1 (0)	0 (0.2)			1 (0.6)
Dec 1-15	3	1		0	5 (1.4)	1 (0.6)		0 (0)	4 (0.9)	1 (0)		0 (0)	7 (2)	1 (0.6)		0 (0)
Dec 16-31	3		1	4	4 (0.9)		2 (0.4)	4 (1.8)	3 (0)		1 (0)	4 (0)	4 (0.9)		2 (0.4)	4 (1.8)
Jan 1-15	3	2	1	0	4 (1.6)	4 (0.9)	1 (0.4)	0 (0.4)	4 (0.6)	2 (0)	1 (0)	0 (0)	5 (1.9)	4 (0.9)	1 (0.4)	0 (0.4)
Jan 16-31	3		0		5 (1.2)		0 (0.7)		6 (1.5)		0 (0)		9 (2.3)		0 (0.7)	
Feb 1-15		1	1	0		1 (0.6)	2 (0.4)	0 (0.2)		1 (0)	1 (0)	0 (0)		1 (0.6)	2 (0.4)	0 (0.2)
Feb 16-28	-	1	2	0	-	1 (0.6)	2 (3.4)	0 (0.2)	-	1 (0)	2 (0)	0 (0)	-	1 (0.6)	2 (3.4)	0 (0.2)
Mar 1-15	0	3	1	1	0 (0.1)	4 (1.1)	2 (0.5)	2 (0.4)	0 (0)	5 (1.7)	1 (0)	1 (0)	0 (0.1)	6 (2.6)	2 (0.5)	2 (0.4)
Mar 16-31	1	5	1	0	2 (0.4)	6 (2.3)	2 (0.5)	0 (4.2)	1 (0)	5 (0.4)	1 (0.4)	0 (0)	2 (0.4)	6 (2.4)	2 (0.7)	0 (4.3)
Apr 1-15	1	2	3	2	1 (0.9)	3 (3.5)	2 (3.1)	3 (1.9)	1 (0)	2 (0)	3 (0.4)	2 (0.3)	1 (0.9)	3 (3.9)	2 (3.4)	3 (1.9)
Apr 16-30	2	11	15	27	0 (8.6)	1 (15.1)	5 (23.4)	11 (35)	2 (0.4)	29 (6.7)	53 (12.6)	33 (2.7)	0 (9.3)	4 (38.7)	5 (94)	14 (43.1)
May 1-15	5	8	10	29	2 (7.8)	2 (11.7)	1 (18.6)	1 (44.8)	12 (4)	37 (12.6)	25 (6)	66 (9.1)	5 (17.4)	3 (60.1)	2 (46.7)	3 (99.9)
May 16-31	1	4	5	12	0 (2.4)	5 (3)	0 (11.6)	2 (22.5)	1 (0.7)	6 (1.6)	9 (2.5)	19 (3.2)	1 (3.4)	7 (4.6)	0 (18.8)	3 (34.2)
Jun 1-15	0	3	17	31	0 (3)	4 (2.1)	22 (7.8)	30 (19.4)	0 (0)	3 (0.6)	19 (1.6)	46 (4.7)	0 (3.1)	5 (2.6)	25 (9.1)	46 (29.6)
Jun 16-30	2	6	5	11	3 (0.9)	9 (2.2)	8 (1.8)	15 (4.2)	2 (0.4)	6 (0)	6 (0.8)	19 (3.5)	3 (1.1)	9 (2.2)	9 (2.2)	27 (8.3)
Jul 1-15	1	3	6		1 (0.6)	4 (1.2)	6 (1.7)		1 (0)	3 (0)	6 (0)		1 (0.6)	4 (1.2)	6 (1.7)	
Jul 16-31	1	2	1		1 (0.6)	2 (0.8)	1 (0.4)		1 (0)	2 (0)	1 (0)		1 (0.6)	2 (0.8)	1 (0.4)	
Total	26	52	71	123	29 (21.9)	46 (34.5)	58 (64.8)	74 (126.1)	38 (4.5)	103 (14.5)	131 (14.3)	196 (11.6)	39 (32.8)	56 (107.4)	64 (171.1)	108 (212.4)

Table 10. Number of spring-run (Butte Creek) juvenile Chinook salmon by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “–” indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Observed assignments				Corrected assignments				Total catch based on observed assignments				Total catch based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0		0		0 (0)		0 (0)		0 (0)		0 (0)		0 (0)		0 (0)
Aug 16-31				0				0 (0)				0 (0)				0 (0)
Sep 1-15			0	0			0 (0)	0 (0)			0 (0)	0 (0)			0 (0)	0 (0)
Sep 16-30			0				0 (0)				0 (0)				0 (0)	
Oct 1-15				0				0 (0)				0 (0)				0 (0)
Oct 16-31																
Nov 1-15	0	0	0	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Nov 16-30	0			0	0 (0)			0 (0)	0 (0)			0 (0)	0 (0)			0 (0)
Dec 1-15	0	0		0	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)
Dec 16-31	0		0	0	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)
Jan 1-15	0	0	0	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Jan 16-31	0		0		0 (0)		0 (0)		0 (0)		0 (0)		0 (0)		0 (0)	
Feb 1-15		0	0	0		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)
Feb 16-28	-	0	0	0	-	0 (0)	0 (1)	0 (0)	-	0 (0)	0 (0)	0 (0)	-	0 (0)	0 (1)	0 (0)
Mar 1-15	0	0	0	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Mar 16-31	1	2	1	0	1 (0.1)	2 (0.3)	1 (0.1)	0 (1)	1 (0)	2 (0)	1 (0.4)	0 (0)	1 (0.1)	2 (0.3)	1 (0.4)	0 (1)
Apr 1-15	5	46	9	18	5 (0.5)	52 (4.7)	10 (0.9)	20 (1.6)	5 (0)	50 (1.7)	9 (0.8)	18 (0.5)	5 (0.5)	57 (5.5)	11 (1.3)	20 (1.7)
Apr 16-30	9	21	41	79	10 (0.9)	24 (2.3)	46 (4)	89 (7.9)	10 (0.9)	53 (8.8)	101 (13.3)	90 (3)	10 (1.3)	61 (11)	112 (16.7)	101 (9.4)
May 1-15	4	10	16	14	5 (0.4)	11 (1.2)	18 (1.7)	16 (1.3)	9 (3.1)	64 (18.3)	41 (7.8)	26 (4.8)	9 (3.2)	74 (20.9)	46 (9.2)	29 (5.6)
May 16-31	0	2	5	5	0 (0)	2 (0.3)	5 (0.5)	6 (0.5)	0 (0)	3 (1.1)	8 (2.2)	8 (2.1)	0 (0)	3 (1.1)	8 (2.3)	9 (2.3)
Jun 1-15	0	0	9	2	0 (0)	0 (0)	10 (0.8)	2 (0.2)	0 (0)	0 (0)	11 (1.2)	3 (1)	0 (0)	0 (0)	11 (1.6)	3 (1.1)
Jun 16-30	0	0	1	1	0 (0)	0 (0)	1 (0.1)	1 (0.1)	0 (0)	0 (0)	1 (0)	2 (1.3)	0 (0)	0 (0)	1 (0.1)	2 (1.3)
Jul 1-15	0	0	0		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)	
Jul 16-31	0	0	0		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)	
Total	19	81	82	119	21 (1.8)	92 (8.7)	92 (7.9)	133 (11.3)	25 (3.2)	172 (20.4)	172 (15.6)	147 (6.3)	25 (4)	197 (29)	190 (22.9)	164 (15.2)

Table 11. Number of spring-run (Mill-Deer creek) juvenile Chinook salmon by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Observed assignments				Corrected assignments				Total catch based on observed assignments				Total catch based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0		0		0 (0.1)		0 (0.2)		0 (0)		0 (0)		0 (0.1)		0 (0.2)
Aug 16-31				0				0 (0.1)				0 (0)				0 (0.1)
Sep 1-15			0	0			0 (0)	0 (0.1)			0 (0)	0 (0)			0 (0)	0 (0.1)
Sep 16-30			0				0 (0)				0 (0)				0 (0)	
Oct 1-15				0				0 (0.1)				0 (0)				0 (0.1)
Oct 16-31																
Nov 1-15	0	0	0	0	0 (0)	0 (0.1)	0 (0.2)	0 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.1)	0 (0.2)	0 (0.1)
Nov 16-30	0			0	0 (0.1)			0 (0.1)	0 (0)			0 (0)	0 (0.1)			0 (0.1)
Dec 1-15	0	0		0	0 (0.2)	0 (0.1)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0.3)	0 (0.1)		0 (0)
Dec 16-31	0		0	0	0 (0.2)		0 (0.1)	0 (0.3)	0 (0)		0 (0)	0 (0)	0 (0.2)		0 (0.1)	0 (0.3)
Jan 1-15	0	0	0	0	0 (0.3)	0 (0.2)	0 (0)	0 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.3)	0 (0.2)	0 (0)	0 (0.1)
Jan 16-31	0		0		0 (0.2)		0 (0.2)		0 (0)		0 (0)		0 (0.4)		0 (0.2)	
Feb 1-15		0	0	0		0 (0.1)	0 (0.1)	0 (0.1)		0 (0)	0 (0)	0 (0)		0 (0.1)	0 (0.1)	0 (0.1)
Feb 16-28	-	1	0	0	-	1 (0.2)	0 (0.2)	0 (0.1)	-	1 (0.4)	0 (0)	0 (0)	-	1 (0.5)	0 (0.2)	0 (0.1)
Mar 1-15	2	2	4	0	2 (0.2)	2 (0.3)	4 (0.4)	0 (0.1)	2 (0)	4 (1.5)	4 (0.3)	0 (0)	2 (0.2)	3 (1.5)	4 (0.5)	0 (0.1)
Mar 16-31	5	6	4	0	5 (0.5)	6 (0.7)	3 (0.6)	0 (0.6)	5 (0)	6 (0)	4 (0.5)	0 (0)	5 (0.5)	6 (0.7)	4 (0.8)	0 (0.6)
Apr 1-15	1	6	2	2	1 (0.3)	6 (1.1)	2 (0.9)	2 (0.6)	2 (0.7)	6 (0.7)	2 (0.5)	2 (0)	1 (0.8)	6 (1.4)	2 (1.1)	2 (0.6)
Apr 16-30	1	3	17	9	1 (2.2)	1 (4.1)	11 (6.5)	3 (9.6)	1 (0.3)	8 (3.6)	57 (12.7)	11 (1.5)	1 (2.4)	2 (10.7)	34 (27.3)	4 (11.8)
May 1-15	0	5	6	10	0 (2.1)	2 (3.2)	3 (5)	2 (12.2)	0 (0)	27 (11.2)	15 (4.5)	23 (5.3)	0 (4.7)	14 (18.3)	7 (12.8)	3 (27.2)
May 16-31	0	0	0	3	0 (0.6)	0 (0.9)	0 (3.1)	1 (6)	0 (0)	0 (0)	0 (0)	5 (1.7)	0 (0.9)	0 (1.2)	0 (5)	2 (9.2)
Jun 1-15	1	0	1	6	1 (0.8)	0 (0.6)	0 (2.2)	3 (5.6)	1 (0)	0 (0)	1 (0.4)	9 (2.2)	1 (0.8)	0 (0.7)	0 (2.5)	3 (8.5)
Jun 16-30	0	0	3	0	0 (0.2)	0 (0.4)	3 (0.5)	0 (1.2)	0 (0)	0 (0)	3 (0.5)	0 (0)	0 (0.3)	0 (0.4)	3 (0.8)	0 (1.9)
Jul 1-15	0	0	0		0 (0.1)	0 (0.2)	0 (0.2)		0 (0)	0 (0)	0 (0)		0 (0.1)	0 (0.2)	0 (0.2)	
Jul 16-31	0	0	0		0 (0.1)	0 (0.2)	0 (0.1)		0 (0)	0 (0)	0 (0)		0 (0.1)	0 (0.2)	0 (0.1)	
Total	10	23	37	30	10 (6.2)	18 (9.9)	27 (18.1)	10 (35)	11 (0.8)	52 (11.9)	86 (13.5)	50 (6.1)	10 (9.3)	32 (31.1)	54 (47.9)	14 (58.9)

Table 12. Number of winter-run juvenile Chinook salmon by biweekly period and sample year that were DNA assigned to run (observed and corrected assignments) and corresponding estimates of total catch based on observed and corrected assignments. Standard errors for estimates are reported in parentheses. A “–” indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Observed assignments				Corrected assignments				Total catch based on observed assignments				Total catch based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0		0		0 (0)		0 (0)		0 (0)		0 (0)		0 (0)		0 (0)
Aug 16-31				0				0 (0)				0 (0)				0 (0)
Sep 1-15			0	0			0 (0)	0 (0)			0 (0)	0 (0)			0 (0)	0 (0)
Sep 16-30			0				0 (0)				0 (0)				0 (0)	
Oct 1-15				0				0 (0)				0 (0)				0 (0)
Oct 16-31																
Nov 1-15	0	0	0	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Nov 16-30	0			0	0 (0)			0 (0)	0 (0)			0 (0)	0 (0)			0 (0)
Dec 1-15	0	0		1	0 (0)	0 (0)		1 (0)	0 (0)	0 (0)		1 (0)	0 (0)	0 (0)		1 (0)
Dec 16-31	0		0	0	0 (0.1)		0 (0)	0 (0.1)	0 (0)		0 (0)	0 (0)	0 (0.1)		0 (0)	0 (0.1)
Jan 1-15	0	2	0	1	0 (0)	2 (0)	0 (0.1)	1 (0)	0 (0)	2 (0)	0 (0)	1 (0)	0 (0)	2 (0)	0 (0.1)	1 (0)
Jan 16-31	1		1		1 (0)		1 (0)		1 (0)		1 (0)		1 (0)		1 (0)	
Feb 1-15		0	3	0		0 (0)	3 (0)	0 (0)		0 (0)	3 (0)	0 (0)		0 (0)	3 (0)	0 (0)
Feb 16-28	-	8	5	5	-	8 (0.1)	5 (1.7)	5 (0.1)	-	10 (0.4)	5 (0)	5 (0)	-	10 (0.4)	5 (1.7)	5 (0.1)
Mar 1-15	3	5	16	5	3 (0)	5 (0.1)	16 (0.2)	5 (0.1)	3 (0)	7 (1.8)	17 (0.3)	5 (0)	3 (0)	7 (1.8)	17 (0.3)	5 (0.1)
Mar 16-31	6	5	12	10	6 (0.1)	5 (0.1)	12 (0.1)	10 (1.7)	6 (0)	6 (0.6)	12 (0)	10 (0)	6 (0.1)	6 (0.6)	12 (0.1)	10 (1.7)
Apr 1-15	1	1	0	7	1 (0)	1 (0)	0 (0)	7 (0.1)	2 (0.7)	1 (0)	0 (0)	7 (0.4)	2 (0.7)	1 (0)	0 (0)	7 (0.4)
Apr 16-30	0	2	0	5	0 (0)	2 (0.1)	0 (0.1)	5 (0.2)	0 (0)	3 (1.6)	0 (0)	5 (0.8)	0 (0)	3 (1.6)	0 (0.4)	5 (0.8)
May 1-15	0	0	0	0	0 (0)	0 (0)	0 (0.1)	0 (0.2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.1)	0 (0.2)	0 (0.2)	0 (0.4)
May 16-31	0	0	0	0	0 (0)	0 (0.1)	0 (0.1)	0 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.1)	0 (0.1)	0 (0.2)
Jun 1-15	0	0	0	0	0 (0)	0 (0)	0 (0)	0 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.2)
Jun 16-30	0	0	0	0	0 (0)	0 (0.1)	0 (0.1)	0 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0.1)	0 (0.1)	0 (0.1)
Jul 1-15	0	0	0		0 (0)	0 (0)	0 (0.2)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0.2)	
Jul 16-31	0	0	0		0 (0)	0 (0)	0 (0.1)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0.1)	
Total	11	23	37	34	11 (0.1)	23 (0.3)	37 (1.8)	34 (1.8)	12 (0.7)	29 (2.5)	38 (0.3)	34 (0.9)	12 (0.7)	29 (2.5)	38 (1.9)	34 (2.1)

### **Total abundance estimates by run**

Biweekly estimates of total (absolute) abundance, both for observed and corrected DNA assignments, are presented for each run in Table 13 - Table 17. These biweekly abundances, which were derived using the Jersey Point estimate of Chipps Island trawl efficiency, followed similar seasonal patterns to those discussed above for DNA assignments and total catch. In the remainder of this section, we focus on comparisons of annual abundance estimates.

Annual abundance estimates varied considerably across years for most runs. For example, based on estimates derived using corrected assignments and the Jersey Point efficiency, the ranges in annual abundances by run were as follows (summarized in Table 18): 1.4 million (in 2008) to 7.5 million (2011) for fall run; 71 thousand (2008) to 186 thousand (2011) for late-fall run; 67 thousand (2008) to 331 thousand (2010) for Butte Creek spring run; 36 thousand (2008) to 92 thousand (2010) for Mill-Deer creek spring run; and 45 thousand (2008) to 63 thousand (2010) for winter run. Regardless of how estimates were derived, annual abundances were lowest for all runs in sample year 2008, while the highest abundances were observed in sample year 2011 for fall and late-fall runs, and in sample year 2010 for spring and winter runs (Table 18).

Abundance estimates were strongly influenced by the choice of efficiency estimate (Table 18). Note that because abundance is inversely proportional to trawl efficiency (e.g., a low efficiency yields a high abundance estimate), and because efficiency was assumed to be constant, abundances based on different efficiencies had the same relative differences regardless of year, run, or assignment type. Thus, the largest abundance estimates were based on the lowest, paired-release estimate of efficiency (0.0064; see Table 4), while the lowest abundances were based on the high efficiency (0.04) implied by the fish-flux method. Relative to Jersey Point estimates, abundances were always 38% higher based on the paired-release efficiency, 29% lower for the Pittsburg efficiency, and 78% lower for the fish-flux method (Table 18). These large and consistent differences in abundance estimates based on the differing efficiencies apply to biweekly estimates as well, as illustrated for winter run in Figure 8.

Annual abundances based on corrected DNA assignments were often much lower than those based on observed assignments for late-fall run and Mill-Deer creek spring run (e.g., Table 18). Percentage differences in corrected versus observed estimates are presented in Table 19 (these were the same regardless of the efficiency estimate used). In sample years 2009-2011, abundances based on corrected assignments for the late-fall and Mill-Deer creek runs were 35% to 73% lower than those based on observed assignments (these fish were mostly transferred to fall run, accounting for the small percentage increases in corrected abundances of fall run). Note that the slight annual differences shown for winter run (Table 19) were not due to differences in observed and corrected assignments, which were always equal (see Table 12); rather, these differences were due to rounding errors in sums of total (corrected) catch estimates across runs, which affected estimates of winter-run proportions in a few biweekly periods.

Across years, fall run composed between 84.0% and 92.8% of the total annual abundance across runs, based on corrected assignments (Figure 9). Late-fall run composed 1.9% to 4.4%, while Butte Creek spring run ranged between 3.9% and 9.0%. Mill-Deer creek spring run and winter run each composed less than 3% of the total abundance in sample year 2008, and less than 2% in subsequent years. Abundance proportions based on observed assignments (not shown) were notably higher for the late-fall and Mill-Deer creek runs in sample years 2009-2011.

#### Components of variance

The precision of annual abundance estimates, as reflected by their standard errors, varied considerably depending on run, assignment type (observed or corrected), and efficiency estimate (Table 18). Variances for abundances based on corrected assignments were larger than those for observed assignments because of measurement errors in assignment corrections. Note that variances for abundances based on the fish-flux method were biased low because there was no estimate of precision associated with this efficiency (see Table 4).

An assessment of the variance components for annual abundances is presented in Table 20. These results are for abundances based on corrected assignments and the Jersey Point efficiency (similar results were found for the paired-release and Pittsburg efficiencies). Note that corrected assignments provide a much better reflection of true uncertainty than observed assignments, which assume no error in DNA assignments. The four variance components in Table 20 were defined in Equation 22. In addition, as a relative measure of precision, Table 20 reports the coefficient of variation ( $CV = \text{standard error}/\text{estimate}$ ) for abundance estimates.

The most precise estimates were for fall run, with CVs of 21% or less (Table 20). Variances for fall-run abundances were largely driven by two components: variances in catch (due to missing sampling days) and efficiency. Abundances for Butte Creek spring run and winter run were also reasonably precise, though their variance components differed. For Butte Creek estimates, all four components contributed 20% or more of the variance, depending on the year. For winter run, sampling error (i.e., low numbers of assignments) was the dominant source of variance (50% or more), while measurement error in corrected assignments contributed little (6% or less). In contrast, abundance estimates for late-fall run and Mill-Deer creek spring run were very imprecise ( $CVs > 75\%$ ) because of measurement error in corrected assignments.

Table 13. Absolute abundance estimates for juvenile fall-run Chinook salmon by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “–” indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Abundance based on observed assignments				Abundance based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		3,370 (4,490)		0 (0)		3,370 (4,554)		0 (3,752)
Aug 16-31				3,287 (3,714)				3,287 (3,788)
Sep 1-15			4,509 (5,118)	1,567 (1,763)			4,509 (5,118)	1,567 (1,890)
Sep 16-30			3,252 (3,656)				3,252 (3,656)	
Oct 1-15				0 (0)				0 (930)
Oct 16-31								
Nov 1-15	1,562 (1,961)	5,684 (7,573)	1,913 (2,197)	3,670 (2,985)	1,562 (1,961)	5,684 (7,682)	0 (1,914)	3,670 (3,090)
Nov 16-30	3,007 (3,854)			0 (0)	3,007 (3,913)			0 (1,169)
Dec 1-15	5,918 (3,422)	0 (0)		0 (0)	1,479 (2,958)	0 (1,259)		0 (17)
Dec 16-31	6,156 (2,709)		1,613 (1,808)	0 (0)	5,277 (2,579)		0 (702)	0 (3,438)
Jan 1-15	5,873 (5,560)	4,389 (3,657)	0 (0)	3,814 (3,409)	2,936 (6,541)	0 (1,911)	0 (654)	3,814 (3,516)
Jan 16-31	6,378 (3,643)		9,303 (6,629)		3,644 (3,419)		9,303 (6,954)	
Feb 1-15		0 (0)	3,790 (2,991)	1,655 (2,077)		0 (1,149)	1,895 (2,172)	1,655 (2,110)
Feb 16-28	-	0 (0)	0 (0)	1,536 (1,723)	-	0 (1,060)	0 (4,152)	1,536 (1,759)
Mar 1-15	0 (0)	15,358 (7,424)	3,142 (2,355)	1,772 (1,870)	0 (1,703)	14,022 (7,689)	1,571 (1,945)	0 (775)
Mar 16-31	10,533 (5,800)	9,273 (5,359)	13,613 (7,532)	14,169 (6,666)	7,899 (5,271)	7,419 (6,473)	13,109 (7,589)	14,169 (9,348)
Apr 1-15	47,159 (17,584)	75,467 (23,450)	92,678 (26,429)	66,292 (24,263)	47,965 (17,771)	66,254 (22,481)	91,645 (27,269)	62,603 (23,637)
Apr 16-30	346,378 (106,015)	781,219 (189,985)	2,502,152 (674,205)	1,371,401 (422,320)	352,104 (111,272)	811,500 (204,293)	2,599,928 (718,837)	1,400,539 (440,542)
May 1-15	698,262 (245,006)	1,214,466 (296,072)	1,306,498 (431,479)	3,455,630 (972,025)	714,726 (254,279)	1,262,784 (318,272)	1,355,467 (457,132)	3,623,379 (1,042,399)
May 16-31	146,679 (59,541)	128,054 (41,796)	752,419 (225,346)	1,261,971 (323,185)	146,679 (60,650)	125,989 (42,265)	771,403 (234,651)	1,296,559 (339,010)
Jun 1-15	72,503 (21,591)	56,061 (21,686)	261,213 (77,549)	936,173 (220,114)	72,503 (22,404)	52,764 (20,988)	252,901 (76,632)	945,808 (227,770)
Jun 16-30	42,160 (13,318)	18,742 (8,827)	54,928 (21,362)	196,979 (60,615)	40,539 (13,030)	13,119 (7,916)	48,825 (19,805)	184,168 (58,196)
Jul 1-15	0 (0)	14,872 (6,934)	11,090 (5,424)		0 (2,004)	13,013 (6,758)	11,090 (6,167)	
Jul 16-31	0 (0)	6,603 (4,442)	1,891 (2,441)		0 (1,397)	6,603 (4,616)	1,891 (2,572)	
Total	1,392,577 (275,465)	2,333,565 (356,147)	5,024,010 (835,987)	7,319,925 (1,131,567)	1,400,328 (286,062)	2,382,526 (382,209)	5,166,794 (887,687)	7,542,762 (1,204,820)

Table 14. Absolute abundance estimates for juvenile late-fall-run Chinook salmon by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. A “–” indicates that no trawl sampling was conducted during this strata. Time periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Abundance based on observed assignments				Abundance based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0 (0)		6,468 (5,312)		0 (732)		6,468 (6,516)
Aug 16-31				0 (0)				0 (714)
Sep 1-15			0 (0)	1,567 (1,763)			0 (30)	1,567 (1,894)
Sep 16-30			0 (0)				0 (21)	
Oct 1-15				1,603 (2,012)				1,603 (2,219)
Oct 16-31								
Nov 1-15	0 (0)	0 (0)	3,826 (3,462)	1,835 (1,977)	0 (10)	0 (1,234)	5,739 (5,022)	1,835 (2,137)
Nov 16-30	0 (0)			2,016 (2,583)	0 (653)			2,016 (2,838)
Dec 1-15	5,918 (3,422)	2,171 (2,781)		0 (0)	10,357 (5,496)	2,171 (3,056)		0 (23)
Dec 16-31	2,638 (1,635)		1,613 (1,808)	7,735 (5,187)	3,518 (2,078)		3,226 (2,895)	7,735 (6,240)
Jan 1-15	11,746 (9,289)	4,389 (3,657)	1,800 (2,307)	0 (0)	14,683 (12,416)	8,778 (6,165)	1,800 (2,412)	0 (829)
Jan 16-31	5,467 (3,454)		0 (0)		8,200 (4,790)		0 (2,021)	
Feb 1-15		1,980 (2,537)	1,895 (2,008)	0 (0)		1,980 (2,788)	3,790 (3,107)	0 (359)
Feb 16-28	-	1,653 (1,766)	3,313 (2,778)	0 (0)	-	1,653 (2,046)	3,313 (6,280)	0 (346)
Mar 1-15	0 (0)	9,599 (6,062)	1,571 (1,618)	1,772 (1,870)	0 (531)	12,019 (7,764)	3,142 (2,465)	3,544 (2,887)
Mar 16-31	2,633 (2,702)	9,273 (5,359)	1,701 (1,848)	0 (0)	5,266 (4,083)	11,128 (7,525)	3,277 (2,830)	0 (7,607)
Apr 1-15	2,619 (2,807)	2,474 (1,856)	5,915 (3,723)	3,583 (2,814)	2,821 (3,654)	3,680 (5,268)	3,818 (7,172)	5,216 (4,907)
Apr 16-30	5,725 (4,522)	38,012 (14,468)	86,619 (33,262)	64,102 (23,158)	0 (26,650)	5,235 (50,690)	8,181 (153,807)	27,194 (84,480)
May 1-15	28,403 (15,944)	48,317 (21,657)	47,612 (21,556)	138,392 (46,448)	11,872 (41,767)	3,917 (78,510)	3,829 (89,294)	6,290 (209,515)
May 16-31	3,333 (4,186)	12,392 (7,074)	19,075 (9,985)	36,551 (13,916)	3,333 (11,779)	14,457 (11,863)	0 (39,955)	5,779 (66,007)
Jun 1-15	0 (0)	4,946 (3,488)	30,448 (11,461)	73,866 (21,629)	0 (5,742)	8,244 (6,291)	40,270 (20,208)	73,866 (51,674)
Jun 16-30	3,243 (2,526)	11,245 (6,040)	12,206 (6,685)	30,427 (12,625)	4,864 (3,571)	16,867 (9,162)	18,309 (9,856)	43,239 (20,149)
Jul 1-15	3,454 (4,603)	5,577 (3,537)	11,090 (5,424)		3,454 (5,025)	7,436 (4,725)	11,090 (6,228)	
Jul 16-31	2,408 (3,166)	3,301 (2,767)	1,891 (2,441)		2,408 (3,464)	3,301 (3,038)	1,891 (2,585)	
Total	77,594 (21,392)	155,335 (29,947)	230,582 (43,982)	369,923 (59,991)	70,782 (54,026)	100,874 (96,100)	111,681 (184,217)	186,359 (242,200)

Table 15. Absolute abundance estimates for juvenile spring-run (Butte Creek) Chinook salmon by biweekly period and sample year based based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “ – “ indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Abundance based on observed assignments				Abundance based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0 (0)		0 (0)		0 (0)		0 (0)
Aug 16-31				0 (0)				0 (0)
Sep 1-15			0 (0)	0 (0)			0 (0)	0 (0)
Sep 16-30			0 (0)				0 (0)	
Oct 1-15				0 (0)				0 (0)
Oct 16-31								
Nov 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Nov 16-30	0 (0)			0 (0)	0 (0)			0 (0)
Dec 1-15	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)
Dec 16-31	0 (0)		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)
Jan 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Jan 16-31	0 (0)		0 (0)		0 (0)		0 (0)	
Feb 1-15		0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)
Feb 16-28	-	0 (0)	0 (0)	0 (0)	-	0 (0)	0 (1,739)	0 (0)
Mar 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Mar 16-31	2,633 (2,702)	3,709 (2,935)	1,701 (1,848)	0 (0)	2,633 (2,721)	3,709 (2,974)	1,638 (1,859)	0 (1,859)
Apr 1-15	13,099 (7,026)	61,858 (19,616)	17,746 (7,403)	32,250 (13,005)	14,107 (7,275)	69,935 (22,832)	21,002 (8,584)	34,779 (14,230)
Apr 16-30	28,626 (12,580)	69,470 (22,320)	165,066 (51,925)	174,824 (56,855)	28,626 (12,861)	79,841 (25,995)	183,254 (58,760)	196,192 (65,689)
May 1-15	21,302 (12,607)	83,576 (33,080)	78,085 (32,020)	54,518 (21,280)	21,370 (12,777)	96,634 (37,755)	88,067 (36,215)	60,809 (23,703)
May 16-31	0 (0)	6,196 (4,620)	16,955 (9,065)	15,389 (7,821)	0 (0)	6,196 (4,676)	17,047 (9,285)	17,338 (8,489)
Jun 1-15	0 (0)	0 (0)	17,627 (7,593)	4,817 (3,431)	0 (0)	0 (0)	17,719 (7,756)	4,817 (3,468)
Jun 16-30	0 (0)	0 (0)	2,034 (2,142)	3,202 (3,166)	0 (0)	0 (0)	2,034 (2,159)	3,202 (3,185)
Jul 1-15	0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)	
Jul 16-31	0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (0)	
Total	65,663 (19,337)	224,811 (44,803)	299,219 (62,643)	285,004 (62,750)	66,738 (19,723)	256,317 (51,510)	330,764 (70,679)	317,140 (71,952)

Table 16. Absolute abundance estimates for juvenile spring-run (Mill-Deer) Chinook salmon by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “ - ” indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Abundance based on observed assignments				Abundance based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0 (0)		0 (0)		0 (183)		0 (660)
Aug 16-31				0 (0)				0 (179)
Sep 1-15			0 (0)	0 (0)			0 (0)	0 (119)
Sep 16-30			0 (0)				0 (0)	
Oct 1-15				0 (0)				0 (163)
Oct 16-31								
Nov 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (309)	0 (329)	0 (140)
Nov 16-30	0 (0)			0 (0)	0 (163)			0 (206)
Dec 1-15	0 (0)	0 (0)		0 (0)	0 (430)	0 (221)		0 (14)
Dec 16-31	0 (0)		0 (0)	0 (0)	0 (176)		0 (123)	0 (592)
Jan 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (967)	0 (336)	0 (1)	0 (208)
Jan 16-31	0 (0)		0 (0)		0 (386)		0 (507)	
Feb 1-15		0 (0)	0 (0)	0 (0)		0 (202)	0 (150)	0 (90)
Feb 16-28	-	1,653 (1,890)	0 (0)	0 (0)	-	1,653 (1,920)	0 (325)	0 (101)
Mar 1-15	15,059 (11,238)	7,679 (5,809)	6,284 (3,538)	0 (0)	15,059 (11,326)	6,009 (5,054)	6,284 (3,610)	0 (150)
Mar 16-31	13,166 (6,625)	11,128 (6,020)	6,806 (4,024)	0 (0)	13,166 (6,745)	11,128 (6,146)	6,554 (4,104)	0 (1,042)
Apr 1-15	5,239 (4,913)	7,422 (3,750)	3,943 (3,041)	3,583 (2,758)	2,821 (3,506)	7,361 (4,022)	3,818 (3,547)	3,477 (2,915)
Apr 16-30	2,862 (3,078)	10,486 (6,462)	93,156 (34,692)	21,367 (9,597)	2,862 (7,573)	2,617 (14,125)	55,630 (48,084)	7,769 (23,421)
May 1-15	0 (0)	35,258 (18,231)	28,567 (14,732)	48,227 (20,133)	0 (11,219)	18,282 (24,772)	13,401 (25,449)	6,290 (57,159)
May 16-31	0 (0)	0 (0)	0 (0)	9,618 (5,951)	0 (2,939)	0 (2,513)	0 (10,646)	3,853 (18,049)
Jun 1-15	1,859 (1,916)	0 (0)	1,602 (1,760)	14,452 (6,809)	1,859 (2,406)	0 (1,162)	0 (3,941)	4,817 (13,902)
Jun 16-30	0 (0)	0 (0)	6,103 (4,209)	0 (0)	0 (463)	0 (715)	6,103 (4,386)	0 (3,117)
Jul 1-15	0 (0)	0 (0)	0 (0)		0 (353)	0 (433)	0 (431)	
Jul 16-31	0 (0)	0 (0)	0 (0)		0 (246)	0 (279)	0 (103)	
Total	38,188 (14,405)	73,629 (21,490)	146,465 (38,463)	97,249 (24,225)	35,769 (19,634)	47,053 (30,086)	91,793 (56,134)	26,209 (65,994)

Table 17. Absolute abundance estimates for juvenile winter-run Chinook salmon by biweekly period and sample year based on observed and corrected DNA assignments and using the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency. Standard errors for estimates are reported in parentheses. A “–” indicates that no trawl sampling was conducted during this strata. Periods with zero catch or no fish sampled for DNA are left blank. Sample year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

	Abundance based on observed assignments				Abundance based on corrected assignments			
	2008	2009	2010	2011	2008	2009	2010	2011
Aug 1-15		0 (0)		0 (0)		0 (0)		0 (0)
Aug 16-31				0 (0)				0 (0)
Sep 1-15			0 (0)	0 (0)			0 (1)	0 (0)
Sep 16-30			0 (0)				0 (1)	
Oct 1-15				0 (0)				0 (0)
Oct 16-31								
Nov 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Nov 16-30	0 (0)			0 (0)	0 (0)			0 (0)
Dec 1-15	0 (0)	0 (0)		2,000 (2,511)	0 (0)	0 (0)		2,000 (2,511)
Dec 16-31	0 (0)		0 (0)	0 (0)	0 (67)		0 (0)	0 (148)
Jan 1-15	0 (0)	4,389 (3,657)	0 (0)	1,907 (2,173)	0 (1)	4,389 (3,658)	0 (138)	1,907 (2,173)
Jan 16-31	911 (932)		3,101 (3,360)		911 (932)		3,101 (3,360)	
Feb 1-15		0 (0)	5,685 (3,841)	0 (0)		0 (0)	5,685 (3,841)	0 (0)
Feb 16-28	-	16,533 (8,506)	8,284 (5,257)	7,684 (5,201)	-	16,533 (8,508)	8,284 (5,970)	7,684 (5,202)
Mar 1-15	22,589 (14,112)	13,438 (7,362)	26,707 (9,411)	8,860 (4,963)	22,589 (14,114)	14,022 (7,442)	26,707 (9,415)	8,860 (4,964)
Mar 16-31	15,799 (7,409)	11,128 (6,244)	20,420 (8,561)	17,712 (7,365)	15,799 (7,411)	11,128 (6,247)	19,664 (8,387)	17,712 (7,964)
Apr 1-15	5,239 (4,913)	1,237 (1,265)	0 (0)	12,541 (6,272)	5,643 (4,945)	1,226 (1,265)	0 (18)	12,172 (6,195)
Apr 16-30	0 (0)	3,932 (3,255)	0 (0)	9,712 (5,466)	0 (82)	3,926 (3,261)	0 (639)	9,712 (5,478)
May 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (182)	0 (297)	0 (318)	0 (888)
May 16-31	0 (0)	0 (0)	0 (0)	0 (0)	0 (34)	0 (214)	0 (213)	0 (356)
Jun 1-15	0 (0)	0 (0)	0 (0)	0 (0)	0 (18)	0 (11)	0 (63)	0 (302)
Jun 16-30	0 (0)	0 (0)	0 (0)	0 (0)	0 (12)	0 (141)	0 (172)	0 (195)
Jul 1-15	0 (0)	0 (0)	0 (0)		0 (0)	0 (2)	0 (424)	
Jul 16-31	0 (0)	0 (0)	0 (0)		0 (0)	0 (0)	0 (145)	
Total	44,540 (16,705)	50,659 (13,825)	64,199 (14,682)	60,420 (13,646)	44,943 (16,719)	51,228 (13,877)	63,442 (14,884)	60,051 (13,987)

Table 18. Annual abundance estimates for juvenile Chinook salmon by run and sample year based on observed and corrected DNA assignments. Annual abundances are shown for four alternative estimates of Chipps Island trawl efficiency (Jersey Point releases, paired-release tests, Pittsburg releases, and the fish-flux method). Standard errors are shown in parentheses. Year 2008 is defined as August 1<sup>st</sup>, 2007 – July 31<sup>st</sup>, 2008 and similarly for 2009, 2010 and 2011.

Run	Year	Annual abundance based on observed assignments (thousands)				Annual abundance based on corrected assignments (thousands)			
		Jersey Point	Paired release	Pittsburg	Fish flux	Jersey Point	Paired release	Pittsburg	Fish flux
Fall	2008	1392.6 (275.5)	1914.8 (327.7)	988.3 (173.5)	306.4 (48.8)	1400.3 (294.2)	1925.5 (355.0)	993.8 (187.5)	308.1 (53.3)
	2009	2333.6 (356.1)	3208.7 (347.2)	1656.1 (192.6)	513.4 (43.2)	2382.5 (399.8)	3276.0 (416.6)	1690.8 (227.2)	524.2 (55.9)
	2010	5024.0 (836.0)	6908.0 (913.6)	3565.4 (492.6)	1105.3 (128.0)	5166.8 (924.3)	7104.3 (1044.8)	3666.8 (559.2)	1136.7 (150.3)
	2011	7319.9 (1131.6)	10064.9 (1224.0)	5194.8 (661.6)	1610.4 (170.0)	7542.8 (1256.9)	10371.3 (1409.5)	5352.9 (755.7)	1659.4 (201.5)
Late-fall	2008	77.6 (21.4)	106.7 (28.3)	55.1 (14.7)	17.1 (4.5)	70.8 (84.2)	97.3 (115.6)	50.2 (59.7)	15.6 (18.5)
	2009	155.3 (29.9)	213.6 (38.1)	110.2 (19.9)	34.2 (5.9)	100.9 (145.1)	138.7 (199.4)	71.6 (102.9)	22.2 (31.9)
	2010	230.6 (44.0)	317.0 (54.9)	163.6 (28.8)	50.7 (8.4)	111.7 (304.8)	153.6 (418.9)	79.3 (216.2)	24.6 (67.0)
	2011	369.9 (60.0)	508.6 (71.0)	262.5 (37.7)	81.4 (10.5)	186.4 (416.0)	256.2 (571.6)	132.3 (295.0)	41.0 (91.4)
Spring Butte	2008	65.7 (19.3)	90.3 (25.0)	46.6 (13.0)	14.4 (3.9)	66.7 (20.3)	91.8 (26.4)	47.4 (13.7)	14.7 (4.1)
	2009	224.8 (44.8)	309.1 (53.9)	159.5 (28.5)	49.5 (8.1)	256.3 (55.5)	352.4 (68.2)	181.9 (35.9)	56.4 (10.4)
	2010	299.2 (62.6)	411.4 (74.1)	212.3 (39.3)	65.8 (11.0)	330.8 (73.9)	454.8 (89.1)	234.7 (47.0)	72.8 (13.4)
	2011	285.0 (62.7)	391.9 (74.0)	202.3 (39.3)	62.7 (11.0)	317.1 (74.5)	436.1 (89.6)	225.1 (47.3)	69.8 (13.4)
Spring Mill-Deer	2008	38.2 (14.4)	52.5 (19.2)	27.1 (9.9)	8.4 (3.0)	35.8 (26.9)	49.2 (36.7)	25.4 (18.9)	7.9 (5.8)
	2009	73.6 (21.5)	101.2 (28.0)	52.3 (14.6)	16.2 (4.4)	47.1 (43.4)	64.7 (59.4)	33.4 (30.7)	10.4 (9.5)
	2010	146.5 (38.5)	201.4 (47.5)	103.9 (25.0)	32.2 (7.2)	91.8 (87.2)	126.2 (119.1)	65.1 (61.5)	20.2 (19.0)
	2011	97.2 (24.2)	133.7 (30.6)	69.0 (16.0)	21.4 (4.7)	26.2 (115.1)	36.0 (158.2)	18.6 (81.7)	5.8 (25.3)
Winter	2008	44.5 (16.7)	61.2 (22.0)	31.6 (11.4)	9.8 (3.5)	44.9 (16.7)	61.8 (22.0)	31.9 (11.4)	9.9 (3.5)
	2009	50.7 (13.8)	69.7 (18.1)	36.0 (9.4)	11.1 (2.8)	51.2 (13.9)	70.4 (18.1)	36.4 (9.4)	11.3 (2.8)
	2010	64.2 (14.7)	88.3 (18.4)	45.6 (9.6)	14.1 (2.8)	63.4 (14.9)	87.2 (18.7)	45.0 (9.8)	14.0 (2.9)
	2011	60.4 (13.6)	83.1 (17.7)	42.9 (9.2)	13.3 (2.8)	60.1 (14.0)	82.6 (18.3)	42.6 (9.5)	13.2 (2.9)

Table 19. Percent difference between annual abundance estimates for juvenile Chinook salmon by run and sample year based on observed versus corrected DNA assignments. Differences in abundance estimates were computed relative to abundances based on observed assignments (i.e., % difference = 100\*[corrected – observed]/observed).

Sampling year	Fall	Late-fall	Spring Butte	Spring Mill-Deer	Winter
2008	0.6	-8.8	1.6	-6.3	0.9
2009	2.1	-35.1	14.0	-36.1	1.1
2010	2.8	-51.6	10.5	-37.3	-1.2
2011	3.0	-49.6	11.3	-73.0	-0.6

Table 20. Coefficient of variation (CV) and components of variance (as a percentage of total variance) by run and sample year for annual abundance estimates based on corrected assignments and the Jersey Point (proximal release) estimate of Chipps Island trawl efficiency.

Run	Year	CV (%)	Components of variance in annual abundance (% of total variance)			
			Assignment corrections	Sampling error (Poison)	Missing days (catch)	Efficiency estimate
Fall	2008	21.0	8.8	7.1	52.0	32.2
	2009	16.8	14.9	8.3	17.2	59.6
	2010	17.9	12.2	3.3	39.2	45.4
	2011	16.7	12.2	1.7	39.2	46.9
Late-fall	2008	118.9	95.5	2.9	1.2	0.4
	2009	143.8	98.3	1.1	0.3	0.2
	2010	272.9	99.5	0.2	0.1	0.1
	2011	223.2	99.4	0.3	0.1	0.2
Spring Butte	2008	30.4	8.8	58.0	18.2	15.0
	2009	21.7	20.1	39.7	12.1	28.1
	2010	22.3	14.3	26.4	26.8	32.5
	2011	23.5	12.8	14.7	39.5	33.0
Spring Mill-Deer	2008	75.1	73.0	23.7	0.8	2.4
	2009	92.1	86.9	10.9	0.9	1.2
	2010	95.0	91.0	5.7	1.4	1.9
	2011	439.3	99.4	0.5	0.0	0.0
Winter	2008	37.2	0.1	83.9	4.1	11.9
	2009	27.1	0.3	61.8	24.3	13.7
	2010	23.5	4.3	49.9	23.0	22.8
	2011	23.3	6.1	55.8	23.2	14.9

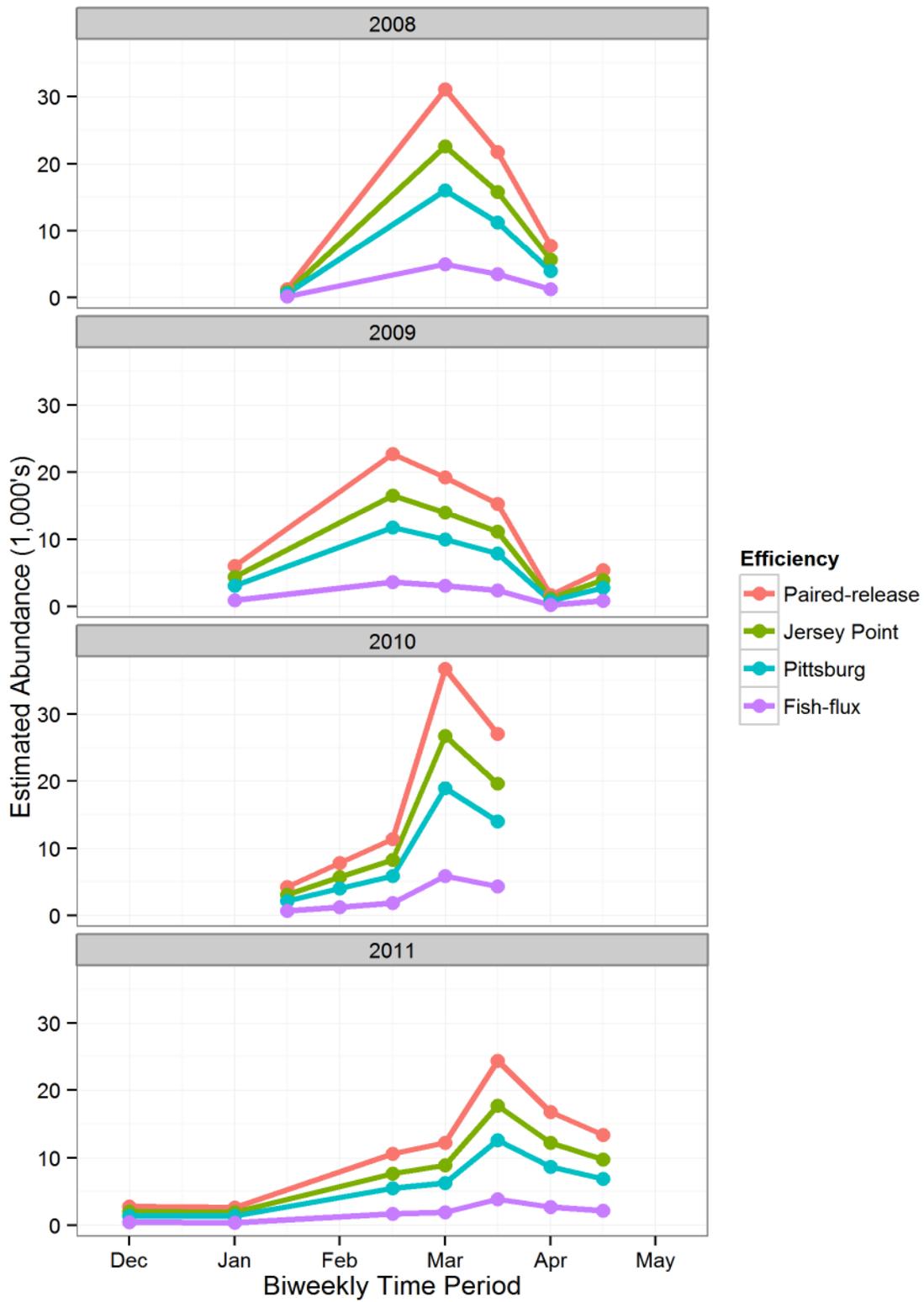


Figure 8. Abundance estimates of winter-run juvenile Chinook salmon at Chipps Island by sample year for four different estimates of trawl efficiency (abundance estimates based on corrected DNA assignments).



Figure 9. Pie chart depicting the run composition of juvenile Chinook salmon abundance at Chipps Island by sample year. Abundance estimates were based on corrected DNA assignments using the Jersey Point estimate of trawl efficiency. Run abbreviations are fall (F), lafe-fall(LF), spring Butte (SB), spring Mill-Deer (SMD), and winter (W).

## **Discussion**

In this study, we developed and applied an analytical framework for estimating juvenile abundances of genetically distinct Chinook salmon populations captured in trawl samples near Chipps Island. The results of four years of juvenile sampling indicate that DNA assignments are likely to be much more accurate than length-at-date criteria (the historical method) for distinguishing winter and spring-run Chinook salmon, which are both ESA-listed. However, there were two critical sources of uncertainty in DNA-based estimates. First, there is a lack of blind-test data to reliably determine DNA assignment errors and corrections for spring run, and assignment corrections for Mill-Deer creek spring run were highly uncertain. Second, estimates of total juvenile abundance (i.e., juveniles emigrating past Chipps Island) were strongly influenced by estimates of trawl efficiency. It is currently unclear which of the efficiency estimates we examined is most accurate, and to what extent trawl efficiency may vary seasonally or among years.

### **Comparison to length-at-date criteria**

A key limitation of the use of length-at-date criteria for run classification is that juveniles of the fall and late-fall runs may overlap with, and potentially dominate, those categories designated for spring and winter runs. Our results indicate that such overlap occurred in all four years examined, and when compared to length criteria, use of DNA assignments provided much more accurate, and reduced, estimates of run composition for the spring and winter runs (see Figure 6 and Figure 7 for results relevant to this discussion). Note that although we did not estimate total abundances based on length criteria, the raw assignments provide a good proxy for comparing relative differences in run compositions between methods.

In general, most fall run migrate to sea as subyearling juveniles in April and May, with small percentages migrating in other months or as yearlings in their second spring. Late-fall-run juveniles, as classified by their DNA, showed a life-history diversity and length-at-date distribution similar to fall run. These runs overlapped considerably with the length-at-date criteria for spring and winter runs. As a result, estimates of spring- and winter-run composition based on length-at-date assignments were roughly 2 to 6 times greater (i.e., overestimates) than compared to DNA assignments. In contrast, late-fall compositions were strongly underestimated in most years based on length criteria (i.e., relatively few fish were captured in the length criteria for late-fall run, while most DNA assignments of late-fall run occurred within the length criteria for other runs).

These findings have important implications. First, given the broad overlap in length-at-date distributions observed among runs (based on DNA assignments), it is clear that the length-at-date method cannot be substantively improved to better allocate runs. In short, the use of mutually exclusive run criteria is inappropriate, in particular given the low relative abundances of spring and winter run. Winter run was the only run for which the length-at-date criteria was accurate, but relatively high numbers of fall and late-fall DNA assignments were also present within the

winter length-at-date criteria. The length-at-date range for spring run was incomplete and included large numbers of fall-run DNA assignments in most years. Given the numerical dominance of fall run (e.g., 84% to 93% of the total abundance across years; Figure 9), it is inevitable that overlaps of fall run will inflate estimates of spring and winter run based on length criteria.

Second, the consistent and substantial overestimates in spring and winter-run compositions, when determined by length criteria, indicates that past estimates for these runs (all based on length criteria) were likely biased high. This conclusion should be applicable to all Delta sampling programs that have employed length criteria to estimate run compositions. The same tendency for length criteria to overestimate the number of winter run has been found through genetic sampling of juvenile salmon salvaged at the Delta Fish Facilities (Hedgecock 2002; B. Harvey, California Department of Water Resources, pers. comm.). Because length criteria consistently overestimated spring and winter assignments in the four years we examined, it is tempting to consider the development of possible correction factors to apply to length-based estimates in previous years. However, there was considerable variation in the overestimates (e.g., 2- to 6-fold changes), and we might expect much greater variation across years due to changes in the true relative compositions of the various runs. At a minimum, it is reasonable to conclude that past estimates based on length criteria would tend to strongly overestimate the spring run and winter-run components.

### **Corrections to DNA assignments based on blind-test data**

The methodology we used to generate “corrected” estimates of DNA run assignments appears intuitive and appropriate; however, it was beyond the scope of this project to conduct a thorough review of the relevant statistical literature. The premise of the approach is sound, that is, there will be assignment errors using DNA methods, and blind-test data provide estimates of those error rates that can be used to adjust or “correct” new samples of observed assignments. In our application, the correction method had important effects on estimates for some runs, but there were also limitations of the blind-test data and methodology that warrant further investigation.

To review, we divided the available blind-test data (Tables 1-3) into two ONCOR assignment probability categories ( $P = 1$  and  $P < 1$ ) to account for differing error rates between these categories. Due to a lack of true spring-run subjects, we combined subjects for Butte Creek ( $n = 13$ ) and Mill-Deer creek ( $n = 2$ ), and applied these data to both runs and both probability categories. Application of the blind-test data had the largest effects on assignments for late-fall run and Mill-Deer creek spring run; the corrected estimates for these runs were often considerably lower than the observed assignments, and were highly uncertain.

Obviously, there is a pressing need for additional blind-test data for true spring run. The corrected spring-run estimates reported in this study should be interpreted cautiously because they are based on very limited blind-test data and arbitrary assumptions (i.e., we used predominantly Butte Creek subjects with  $P = 1$  to represent Mill-Deer creek run and the  $P < 1$

categories). However, it is instructive to understand the different results found for the Butte Creek and Mill-Deer creek runs. Even though the same blind-test data were used to represent the “true” subjects for both runs, their run corrections differed greatly. Compared to observed assignments, corrections for Butte Creek were slightly larger and reasonably precise, while corrections for Mill-Deer creek were often much lower than observed and very imprecise. These differences were due entirely to the false-negative error rates observed for true fall run in the  $P < 1$  category, in which two of 46 fish (4.3%) were incorrectly assigned to Mill-Deer creek but none to Butte Creek. Consequently, when observed assignments of Mill-Deer creek run were present with large numbers of fall run (a frequent scenario), the correction algorithm reallocated Mill-Deer creek assignments to the fall run. Because these reallocations were based on a small and uncertain binomial probability (2 of 46), variances for the Mill-Deer creek corrections were very large. A similar explanation underlies the results for late-fall run.

These results contain an important insight. Because of the numerical dominance of fall run, its false-negative error rates were the key determinants of change and uncertainty in the corrected estimates (and abundance estimates) for late-fall run and Mill-Deer creek spring run. Thus, while obtaining additional blind-test data for true spring run is important, it is more important (if possible) to eliminate false-negative error rates in assignments of true fall run.

In contrast, results for winter run were very encouraging. In blind tests, true winter-run were correctly identified 98.2% of the time when  $ONCOR P = 1$ , and roughly 98% of observed field assignments of winter run fell into this category. More importantly, true fall run were never incorrectly identified as winter run in blind tests. As a result, assignment corrections for winter-run (after rounding) always equaled the number of observed assignments, and were very precise.

Unfortunately, we omitted a potentially important source of variance in estimates of corrected DNA assignments (K.Newman, U.S. Fish and Wildlife Service, pers. comm.), which should be accounted for in future applications. Specifically, our Equation (8) accounts for uncertainty in estimates of assignment probabilities (error rates) derived from blind-test data, but ignores the multinomial variation or “sampling error” associated with each new sample of observed assignments. As a result, our reported variances for total abundance estimates were underestimated to some extent. However, we expect that this omission would have little effect on variance estimates for winter run and Butte Creek spring run, in particular for their annual abundance estimates, for two reasons: (1) multinomial sampling variation would be minimal when key error rates are close to 0, which was the case winter run and Butte Creek spring run; and (2) the relative importance of multinomial sampling variation, which is specific to each sample, would diminish when summing abundances across multiple periods and length strata, whereas variances in estimated assignment probabilities (blind-test data) apply to all samples and accumulative across strata via covariance terms. Larger implications would be expected for late-fall run and Mill-Deer creek spring run (due to false-negative error rates in assignments of true fall run), but variances in the abundance estimates for these runs are already dominated by uncertainty in corrected DNA assignments, so our general conclusions would not change.

Other improvements to our methodology could be explored. First, additional stratification of ONCOR assignment probabilities may be useful, in particular if more blind-test data are collected. While there were clear differences in error rates between the  $P = 1$  and  $P < 1$  categories, we only cursorily examined other subcategories (e.g.,  $P < 0.9$  or  $P < 0.8$ ) before concluding that there were insufficient data to warrant further stratification. Second, we used a bootstrap procedure to estimate variances for assignment corrections. However, if an analytical variance estimator could be developed, it would greatly simplify computations and make variance estimation possible in a spreadsheet, for example. Third, the correction algorithm often produced negative estimates of corrected assignments. As detailed above, we dealt with this using a simple procedure in which negative estimates were set to zero and the remaining estimates were scaled so their sum equaled the total observed assignments. A more sophisticated approach would involve bounded likelihood or Bayesian models that explicitly incorporate the blind-test data, observed assignments, and parameter constraints (i.e.,  $\hat{x}_i \geq 0$ ) to estimate corrected assignments. However, such nonlinear models can be a challenge to fit, and separate fits would be required for each stratification that resulted in negative estimates (more than 100 cases in our application). It seems doubtful that such an approach would lead to substantively different estimates worthy of the effort.

### **Estimates of total abundance and trawl efficiency**

The accuracy and precision of abundance estimates depend critically on estimates of trawl efficiency. For a given efficiency estimate, annual abundances were reasonably precise (e.g., CV = 20% to 30%) for the fall, Butte-Creek spring, and winter runs. While these precision estimates are encouraging, they should be interpreted cautiously for three reasons, which we discuss in turn: (1) abundances were sensitive to the choice of efficiency estimate; (2) efficiency was assumed to be constant over time; and (3) sampling error in catch assignments may be greater than assumed.

We used three efficiency estimates that were independently derived using CWT-release data (Pyper et al. 2013). The lowest efficiency (0.0064 for paired-release tests) produced abundances that were roughly two times greater than those based on the highest efficiency (0.0124 for Pittsburg releases). This two-fold difference implies considerably greater uncertainty than indicated by the standard errors for abundances based on any single efficiency estimate. Thus, determining the most appropriate data and methodology for estimating Chipps Island trawl efficiency is of high priority. As discussed in Pyper et al. (2013), there were advantages and disadvantages to the data and methods used to develop each efficiency estimate. Speculatively, they suggest that the paired-release estimate is biased low, while the Jersey Point estimate appears most defensible (the Pittsburg estimate was based on only three releases). In any case, additional releases and analyses such as those recommended in Pyper et al. (2013) may help to resolve these uncertainties.

By comparison, the fish-flux method (Kimmerer 2008) produced by far the lowest abundance estimates. We have little confidence in these estimates. The fish-flux method, which had an implied efficiency (0.04) that was roughly four times greater than the empirical estimates, is a simple conceptual model that relies on several key assumptions (e.g., a fixed migration speed and random trawl and/or fish distributions). Most critically, the method does not account for avoidance behaviors that would likely reduce fish vulnerability to trawl capture (see Pyper et al. 2013 for further discussion). For Chipps Island trawl, it is clear that the fish-flux method likely produces substantial underestimates of abundance, and we would expect similar biases to occur at other trawl-sampling locations.

The second key source of uncertainty relates to possible temporal variation in trawl efficiency. In our application, we assumed that efficiency was constant across years, and the variance terms we used for efficiency (Table 4) reflected only statistical error in estimates of “mean efficiency.” However, additional temporal variation is expected, and hence, the standard errors we reported for abundance estimates (and variance components due to efficiency) are likely biased low.

Temporal variation in efficiency may occur at short time scales (e.g., daily) and longer time scales (e.g., seasonal or annual differences). For example, daily fluctuations in efficiency could result from daily differences in trawl operation (e.g., time of day and location of tows). As noted below, additional analyses could be pursued to better characterize variation in daily catch and efficiency. However, the relative importance of short-term variation would be diminished or “averaged out” to a large extent when estimating annual abundances across numerous days.

Seasonal or annual variation in efficiency is of greater concern. For example, a 30% difference in annual efficiency (relative to the assumed constant value) would result in a roughly 30% bias in the abundance estimate, and such variation would not be adequately captured in the variance estimator we used. Seasonal or annual changes in efficiency could result from seasonal/annual variation in conditions affecting fish behavior (e.g., fish size, water flow, temperature, turbidity, etc.). Although Pyper et al. (2013) found little evidence of such relationships, the paired-release data they examined were highly variable. Furthermore, it is very difficult to quantify temporal variation in efficiency using available CWT-release data because of the confounding effects of variation in survival rates. As discussed in Pyper et al. (2013), further efforts should be considered to either develop reliable, year-specific estimates of efficiency, or better quantify potential inter-annual variation in efficiencies from past data.

The third source of uncertainty relates to sampling error. We accounted for sampling error in abundance estimates by assuming catch assignments followed Poisson distributions. However, daily trawl catches are likely “overdispersed” (i.e., have more patchy or clumpy distributions with higher variances than assumed under the Poisson model). Numerous factors could result in overdispersion, including spatial and temporal patchiness in daily fish migration, and daily differences in trawl operation (i.e., analogous to short-term variation in efficiency noted above). It should be possible to quantify (or approximate) levels of overdispersion in daily catches by

analyzing tow-specific catch data (and simultaneously assess potential covariates such as time of day, channel location, trawl direction, etc., that may affect trawl efficiency). Catch assignments could then be modeled as overdispersed Poisson variables, for example, in the variance estimator for abundance. The implications of overdispersion could be large for the spring and winter runs because these runs had low numbers of catch assignments. A low number of assignments translates into high sampling error (e.g., see results in Table 20 for winter run). Thus, if sampling-error variances increased considerably due to overdispersion, we would expect potentially large reductions in the precision of abundance estimates for Butte Creek spring run and winter run (Mill-Deer creek precisions were largely driven by assignment errors). In sum, the potential for overdispersion in daily catch warrants further investigation.

### **Independent estimates of winter run abundance**

Noble Hendrix of R2 Resources Consultants, Inc., developed the *Oncorhynchus* Bayesian ANalysis (OBAN) model, which is a statistical life-cycle model for winter-run salmon in the Sacramento River. The OBAN model was developed to evaluate factors influencing the returning numbers of winter-run salmon to the upper Sacramento River. It contains two estimates of survival: 1) the survival from adults on the spawning grounds to juveniles at Red Bluff Diversion Dam (RBDD) and 2) the survival from juveniles at RBDD to adults on the spawning ground. In order to identify the survival rate in the Delta, Hendrix concluded that an index of abundance was needed at Chipps Island to separate the survival in the Delta from survival in the ocean (N.Hendrix, R2 Resources Consultants, Inc., pers. comm.). Estimates of survival to age 2 could also be obtained by using run reconstruction information for age 3 and age 4 winter-run Chinook salmon (e.g. O'Farrell et al. 2012). Hendrix also concluded that the data obtained at Chipps Island, although imprecise, could be quite influential on the results of the OBAN model (N.Hendrix, R2 Resources Consultants, Inc., pers. comm.).

When Hendrix initially used the Chipps Island estimates of winter-run abundance based on the length-at-date criteria, the OBAN model consistently under-estimated these abundances (N.Hendrix, R2 Resources Consultants, Inc., pers. comm.). Hendrix then re-ran the OBAN model to determine the expected abundance of winter run at Chipps Island, and he estimated abundances of winter run between 33,506 and 37,398 for the years between 2008 and 2011, although these abundances were confounded by the inability to differentiate survival in the Delta from survival in the ocean (N. Hendrix, R2 Resources Consultants, Inc., pers. comm.). Results from our study, using corrected DNA assignments and the Jersey Point efficiency estimate, suggest winter-run abundance at Chipps Island ranged from 45 thousand to 63 thousand between 2008 and 2011. Although the OBAN abundance estimates incorporate mortality after Chipps Island, they are within the relative magnitude of our abundance estimates (using DNA and Jersey Point efficiency), but much lower than winter-run estimates at Chipps Island obtained using the length-at-date criteria (~200,000 for all four years; Speegle et al. 2013). The results from re-running the OBAN model appear to support the relative magnitude of the estimates of winter-run

abundances we obtained using corrected DNA assignments and the Jersey Point efficiency estimate.

### **Implications for past and future sampling**

Tissue samples of juvenile Chinook salmon captured in Chipps Island trawl were collected from 1996 to 2002, but not all samples were DNA analyzed. We reviewed these data to determine if reliable DNA-based estimates of abundance could be obtained for the winter and/or spring runs. Tissue sampling was sparse and sporadic from 1996 to 2000 (e.g., annual sample sizes ranged from 15 to 272), with negligible sample numbers and/or poor temporal coverage within the length criteria for the winter and spring runs. However, useful estimates for winter-run may be obtained for 2001 and 2002, with minimal effort, assuming that most migrants were within the winter-run length-at-date criteria (consistent with observed assignments for sample years 2008-2011). In 2001, tissue was collected from all 102 juvenile salmon caught in the winter-run length criteria. Although only five of these were DNA analyzed, they were all assigned to winter run. Speculatively, genetic analysis of the remaining tissue samples could provide enough DNA assignments of winter-run (e.g., > 30) to achieve precisions for 2001 abundance estimates similar to those for 2010 and 2011. In 2002, there were 71 juvenile salmon caught in the winter-run length criteria, of which 63 were tissue sampled, 48 were DNA analyzed, and 13 were assigned to winter run. Thus, genetic analysis of the remaining 15 tissue samples would likely yield only a few more winter-run assignments. With or without additional genetic analysis, we would expect precisions for 2002 abundance estimates to be somewhere between those for 2008 and 2009.

The methods used in this study could be readily applied (or adapted) to the DNA assignments of juveniles collected in Sacramento trawl sampling from 2008 to 2011. At a minimum, expansions could be estimated to account for sampling fractions and trawl effort, thereby providing estimates of run composition. However, it is unclear if trawl efficiency at Sacramento can be reliably estimated to provide meaningful abundance estimates. In addition, genetic sampling and analysis could be used in similar ways to estimate run compositions of migrating juvenile Chinook salmon at other locations in the Central Valley. Future DNA sampling could be particularly advantageous at Sacramento trawl or at the Knights Landing screw traps, which are intended to monitor juvenile Chinook produced in the Sacramento Basin as they enter the Sacramento Delta. As demonstrated for Chipps Island trawl, location-specific estimates of winter and spring-run compositions (relative abundances) based on DNA would likely be much more accurate than estimates based on length criteria.

The results of this study can also be used to improve the sampling design for tissue collection and DNA analysis at Chipps Island trawl in future years. Presumably, there will be numerous constraints, objectives, and alternative sampling designs to consider. Constraints may include trawl restrictions to reduce take of Delta smelt, and budget restrictions that limit trawl effort and/or tissue sampling for DNA analysis. Key objectives may include the run types to target

(e.g., winter and/or spring run) and the desired precision of estimates. Given such objectives and constraints, alternative sampling designs could be explored to best allocate trawl effort and tissue sampling across discrete time periods and fish-length strata. The original sampling plan for this study (Attachment A) was based on analyses of optimal sample allocations; however, the statistical framework and assignment data in this report would provide a much stronger basis for determining optimal sampling designs.

Obviously, from a sampling perspective, the key to improving estimates for winter and spring run is to increase the number of true juveniles of these runs that are DNA analyzed. This will be most easily achieved for winter run because of the relatively constrained time and length distribution of juvenile migrants. To improve estimates, trawl sampling would be maximized during the peak winter-run period, with tissue collection targeting the larger length classes in which winter run are predominantly found. Speculatively, trawl effort would be directed toward more days of sampling, which would increase winter-run catches as well as reduce uncertainty due to missing trawl days, as opposed to increasing minutes fished on selected days (also useful, of course). By comparison, it would likely be much more costly to increase observed DNA-assignments of Butte Creek spring run due to their broad overlap with fall run. Nevertheless, the same principles would generally apply: target trawl sampling and tissue collection at the periods and length strata of peak juvenile migration.

### **Related management implications**

Improved estimates of juvenile run composition and abundance at Chipps Island, as provided by DNA sampling, would improve our understanding of the population dynamics of the ESA-listed Chinook runs. In the case of winter run, the estimates of fry passing Red Bluff Diversion Dam (RBDD) can be compared to the abundance estimates of juveniles passing Chipps Island to estimate juvenile survival in freshwater. RBDD is located near the lower limit of winter-run spawning, and most winter-run juveniles pass RBDD as fry less than 45 mm in length during September through November, while winter-run juveniles pass Chipps Island from mid-February to mid-April at 100 to 125 mm in length. Thus, the abundance estimates for winter run at RBDD and Chipps Island trawl provide meaningful bookends for the freshwater rearing of the population. Accurate estimates of freshwater survival derived from these two sampling locations would greatly enhance assessments of effects of water management actions and other human activities on the freshwater production of winter run. The reduced estimate of winter run juvenile production that results from DNA-based run assignments means that freshwater survival is less than would have been calculated by methods employed before the 2008 sample year; however, such estimates were rarely used because of the low reliability of distinguishing winter-run juveniles based on length criteria.

The lack of reliable abundance estimates for Chinook salmon juveniles reaching San Francisco Bay each year has been an obstacle to resolving debate over management actions that should be taken to restore runs of ESA-listed winter and spring-run Chinook salmon. For example, there has been much speculation about the magnitude of mortality that water project operations impose

on these runs. In particular, the number of juvenile Chinook captured at the Delta Fish Facilities (DFF) has been used as the basis for calculating direct loss due to the water project pumps, and increases in this index of loss have been assumed to indicate increases in mortality rather than increases in abundance of juveniles passing through the Delta. Modeling studies are underway to estimate the influence of water project operations on juvenile survival through the Delta, and accurate estimates of juvenile numbers arriving at and leaving the Delta would be of great value for developing and validating such models.

### **Summary of recommendations**

1. Past estimates of juvenile abundance in the Delta for winter and spring runs of Chinook salmon based on length criteria should be regarded as unreliable and biased substantially high. Given the broad overlap in length-at-date distributions observed among runs (based on DNA assignments), it is clear that the length-at-date method will consistently produce high rates of error in non-fall-run assignments.
2. Determining the most appropriate data and methodology for estimating Chipps Island trawl efficiency is of high priority. Efforts should be made to better quantify mean trawl efficiency, as well as potential seasonal/annual variation in efficiency and overdispersion in trawl catch. Additional releases and analyses such as those recommended in Pyper et al. (2013) may help to resolve these uncertainties.
3. There will be assignment errors using DNA methods, and the use of blind-test data can provide estimates of those error rates to adjust or “correct” new samples of observed assignments. Application of the blind-test data had the largest effects on assignments for late-fall run and Mill-Deer creek spring run. Because of the numerical dominance of fall run, its false-negative error rates were the key determinants of change and uncertainty in the corrected estimates (and abundance estimates) for late-fall run and Mill-Deer creek spring run. Thus, while obtaining additional blind-test data for true spring run is important, it would be more valuable to develop genetic markers that could eliminate false-negative error rates in assignments of true fall run.
4. We omitted a component of variance (multinomial sample variation) for estimates of corrected DNA assignments that should be included in future applications (see Equation (8) and related text). Other improvements to our methods for handling assignment error rates could be explored, including stratification of ONCOR assignment probabilities, development of an analytical estimator for variances, and a better method for handling negative estimates of corrected DNA assignments.
5. The statistical framework and assignment data in this report would provide a strong basis for determining optimal designs for collecting genetic samples in future years. To improve estimates of non-fall runs, trawl effort and tissue sampling would focus on the peak periods of migration and length classes for the target run.
6. Future DNA sampling could be particularly advantageous at Sacramento trawl or at the Knights Landing screw traps, which are intended to monitor juvenile Chinook produced in the Sacramento Basin as they enter the Sacramento Delta. Comparison of size, time,

and abundance of juveniles entering the Delta to those leaving the Delta would be valuable for elucidating mechanisms that relate to through-Delta survival.

7. Future analysis could assess the utility of using DNA run assignments at the Delta Fish Facilities (DFF) as surrogate for estimating past run composition at the Chipps Island trawl. DNA sampling has been more extensive at salvage prior to this study, and may be useful for expanding past years of trawl catch at Chipps Island.
8. We recommend a comparison between abundance of winter run at Chipps Island relative to abundance at the Delta Fish Facilities to help determine if the direct loss estimated at the facilities is a function of higher mortality at the pumps or higher abundance using years where genetic composition at both locations was estimated with HMSC16 set of microsatellites.

## References

- Banks M.A., Rashbrook V.K., Calavetta M.J., Dean C.A. & Hedgecock D. 2000. Analysis of microsatellite DNA resolves genetic structure and diversity of Chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. *Canadian Journal of Fisheries and Aquatic Sciences* 57, 915-927.
- Banks M.A. & Jacobson. D.P. 2004. Which genetic markers and GSI methods are more appropriate for defining marine distribution and migration of salmon? *North Pacific Anadromous Fish Commission Technical Note* 5, 39-42.
- Banks, MA. 2005. Stock identification for the conservation of threatened or endangered species. In: *Stock identification methods*, eds: Cadrin, S. X., K. D. Friedland and J. R. Waldman. Elsevier Press. pp 609-629.
- Banks, MA., D. Jacobson, I. Meusnier, C. Greig, V. Rashbrook, W. Arden, J. Bernier-Latmani, J. Van Sickle, K. O'Malley. In review. Testing Advances of Molecular Discrimination among Chinook salmon Life Histories: Evidence from a Blind Test. Submitted to *Animal Genetics*. 5/31/13.
- Brandes, P. and J. McLain. 2001. Juvenile Chinook salmon abundance and distribution, and survival in the Sacramento-San Joaquin Estuary. In: Brown RL, editor. *Fish Bulletin* 179: Contributions to the Biology of Central Valley Salmonids. Volume 2. Sacramento (CA): California Department of Fish and Game.
- Cramer, S.P., M. Daigneault, and M. Teply. 2004. *Integrated Modeling Framework User's Guide*. S.P. Cramer & Associates prepared for California Urban Water Agencies and State Water Contractors, Sacramento, CA 105 pp.
- Fisher, F. W. 1992. Chinook salmon, *Oncorhynchus tshawytscha*, growth and occurrence in the Sacramento-San Joaquin River system. California Department Fish and Game.
- Fisher, F.W. 1994. Past and present status of Central Valley Chinook salmon. *Conservation Biology* 8(3):970-873.
- Greig C.A. & Banks. M.A. 1999. Five multiplexed microsatellite loci for rapid response run identification of California's endangered winter Chinook salmon. *Animal Genetics* 30, 318-320.
- Greig C.A. & Banks. M.A. 2003. New tetranucleotide microsatellites for fine-scale discrimination among endangered Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology Notes* 3, 376-379.
- Harvey, B. and C. Stroble (in press). Comparison of genetic versus Delta Model Length-at-Date run assignments for juvenile Chinook salmon at state and federal south Delta salvage facilities.

California Department of Water Resources. Submitted to Interagency Ecological Program for the San Francisco Estuary as a technical report. March 2013.

Hedgecock D., Banks M.A., Rashbrook V.K., Dean C.A. & Blankenship. S.M. 2001. Applications of population genetics to conservation of Chinook salmon diversity in the Central Valley In: *Contributions to the Biology of Central Valley Salmonids* (ed. By R.L. Brown) State of California Resources Agency Department of Fish and Game. *Fishery Bulletin* 179, 45-70.

Hedgecock D. 2002. Microsatellite DNA for the management and protection of California's Central Valley Chinook salmon (*Oncorhynchus tshawytscha*). Final Report for the Amendment to Agreement No. B-59638. University of California Davis, Bodega Marine Laboratory. 2099 Westside Road, Bodega Bay, CA 94923-0247.

Kalinowski S.T. 2008. ONCOR software for genetic stock identification.  
<http://www.montana.edu/kalinowski/Software/ONCOR.htm>

Kimmerer, W.J. 2008. Losses of Sacramento River Chinook salmon and delta smelt to entrainment in water diversions in the Sacramento-San Joaquin Delta. San Francisco Estuary Watershed Science. Available from: <http://escholarship.org/uc/item/7v92h6fs>

Kormos, B., M. Palmer-Zwahlen and A. Low. 2012. Recovery of Coded-Wire Tags from Chinook salmon in California's Central Valley Escapement and Ocean Harvest in 2010. California Department of Fish and Game. Fisheries Branch Administrative Report 20120-02. March 2012. 41pp.

Mood, A.M, Graybill, F.A and D.C. Boes. Introduction to the Theory of Statistics, 3rd Edition. New York: McGraw-Hill, 1974.

Naish K.A. & Park L.K. 2002. Linkage relationships for 35 new microsatellite loci in Chinook salmon *Oncorhynchus tshawytscha*. *Animal Genetics* 33, 316–318.

National Marine Fisheries Service. (NMFS) 1994. Endangered and threatened species; threatened status for two Chinook salmon Evolutionarily Significant Units (ESu's) in California; final rule Federal Register 50393.64 (179). 16 September, 1999.

Nelson R.J. & Beacham. T.D. 1999. Isolation and cross species amplification of microsatellite loci useful for study of Pacific salmon. *Animal Genetics* 30, 228–229.

O'Farrell, M.R., Mohr, M.S., Grover, A.M., Satterthwaite, W.H., 2012. Sacramento River winter Chinook cohort reconstruction: analysis of ocean fishery impacts. U.S. Department of Commerce, NOAA Technical Memorandum NMFS- SWFSC-491. Available from <http://swfsc.noaa.gov/publications/TM/SWFSC/> NOAA-TM-NMFS-SWFSC-491.pdf

O'Malley K., McClelland E.K. & Naish K.A. 2010. Clock genes localize to stage-specific quantitative trait loci for growth in juvenile coho salmon, *Oncorhynchus kisutch*. *Journal of Heredity* 101, 628-632.

Pyper, B., T. Garrison, and S. Cramer. 2013. Analysis of trawl efficiency at Chipps Island using coded-wire-tagged releases of juvenile Chinook salmon. Cramer Fish Sciences Technical Report for U.S. Fish and Wildlife Service, Lodi, CA. 97 pp.

Rannala, B., and Mountain, J.L. 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA*. 94: 9197–9201.

Speegle, J., J. Kirsch, and J. Ingram. 2013. Annual report: juvenile fish monitoring during the 2010 and 2011 field seasons within the San Francisco Estuary, California. Stockton Fish and Wildlife Office, United States Fish and Wildlife Service, Lodi, California.

U.S. Fish and Wildlife Service (USFWS). 1994. 1993 annual progress report: "Abundance and survival of juvenile Chinook salmon in the Sacramento-San Joaquin Estuary". Stockton, CA.

U.S. Fish and Wildlife Service (USFWS) 1997. 1994 annual progress report: "Abundance and survival of juvenile Chinook salmon in the Sacramento-San Joaquin Estuary". Stockton, CA.

U.S. Fish and Wildlife Service. (USFWS) 2006. 2000 annual progress report: "Abundance and survival of juvenile Chinook salmon in the Sacramento-San Joaquin Estuary". Stockton, CA.

U.S. Fish and Wildlife Service. (USFWS) 2011. Biological Assessment of Artificial Propagation at Coleman National fish hatchery and Livingston Stone National Fish Hatchery: program description and incidental take of Chinook salmon and steelhead. Prepared by U.S. Fish and Wildlife Service, Red Bluff Fish and Wildlife Office, Red Bluff, CA 96080 and U.S. Fish and Wildlife Service Coleman National Fish Hatchery Complex, Anderson, CA 96007. 372 p.

Williamson K.S., Cordes J.F. & May. B. (2002) Characterization of microsatellite loci in Chinook salmon (*Oncorhynchus tshawytscha*) and cross-species amplification in other salmonids. *Molecular Ecology Notes* 2, 17-19.

## Appendix A: Derivation of selected estimators

### Estimates of total catch by run

The following derivation applies to an estimate of total catch,  $X_{itk}$ , for a given run  $i$ , biweekly period  $t$ , and length stratum  $k$ . For simplicity, all subscripts are omitted until the final estimates of interest are obtained.

When deriving estimators for total catch ( $X$ ) and its variance, two processes are considered: (1) the sampling of catch ( $C$ ), which determines the distribution of true abundance,  $x$ , in the sample ( $S$ ); and (2) the estimation of  $x$  given the observed DNA-run assignments  $\{y\}$  and blind-test data. Let  $f (= S/C)$  be the fraction of catch that is sampled and assigned to run. Assuming all fish  $C$  have an equal probability ( $f$ ) of being sampled, the sampling of catch is “sampling without replacement” and thus  $x$  follows a hypergeometric distribution conditional on the values of  $C$ ,  $X$ , and  $S$  (e.g., Mood et al. 1974, p. 91). Formally,  $x \sim \text{Hypergeometric}(C, X, S)$  with expectation

$$(A1) \quad E[x] = X \left( \frac{S}{C} \right) = Xf$$

and variance

$$(A2) \quad \begin{aligned} V[x] &= X \left( \frac{S}{C} \right) \left( \frac{C-S}{C} \right) \left( \frac{C-X}{C-1} \right) \\ &= Xf(1-f) \left( \frac{C-X}{C-1} \right). \end{aligned}$$

We do not observe the true abundance  $x$  (unless DNA assignments are 100% accurate). Rather, we have an estimator  $\hat{x}$  that is conditional on  $x$  (as well as all other true run-specific abundances and the true error rates associated with DNA assignments). Assuming  $\hat{x}$  and  $x$  are jointly distributed random variables, we can express the expectation and variance of  $\hat{x}$  as (e.g., Mood et al. 1974, p. 158-159):

$$(A3) \quad E[\hat{x}] = E_x[E[\hat{x} | x]]$$

$$(A4) \quad V[\hat{x}] = E_x[V[\hat{x} | x]] + V_x[E[\hat{x} | x]],$$

where  $E[\hat{x} | x]$  and  $V[\hat{x} | x]$  denote the conditional expectation and variance of  $\hat{x}$ , respectively.

Further assuming that  $E[\hat{x} | x] = x$  (i.e.,  $\hat{x}$  is an unbiased estimate of  $x$ ) and  $E_x[V[\hat{x} | x]] = V[\hat{x} | x]$ , Equations (A3) and (A4) become:

$$(A5) \quad E[\hat{x}] = E[x]$$

$$(A6) \quad V[\hat{x}] = V[\hat{x}|x] + V[x].$$

Combining the definitions for Equations (A5) and (A2) and solving for  $X$  gives:

$$(A7) \quad X = \frac{E[\hat{x}]}{f}$$

with variance

$$(A8) \quad \begin{aligned} V[X] &= \frac{V[\hat{x}]}{f^2} \\ &= \frac{1}{f^2} (V[\hat{x}|x] + V[x]) \\ &= \frac{1}{f^2} \left( V[\hat{x}|x] + Xf(1-f) \left( \frac{C-X}{C-1} \right) \right). \end{aligned}$$

Finally, to obtain estimators, we substitute  $E[\hat{x}]$ ,  $X$ , and  $V[\hat{x}|x]$  in Equations (A7) and (A8) with the observed estimates  $\hat{x}$ ,  $\hat{X}$ , and  $\hat{\sigma}_{\hat{x}}^2$ , respectively. Returning subscripts, the final estimator of total catch by run is given by

$$(A9) \quad \hat{X}_{itk} = \frac{\hat{x}_{itk}}{f_{itk}}$$

with approximate variance

$$(A10) \quad \hat{\sigma}_{\hat{X}_{itk}}^2 = \frac{1}{f_{itk}^2} \left( \hat{\sigma}_{\hat{x}_{itk}}^2 + \hat{X}_{itk} f_{itk} (1-f_{itk}) \frac{C_{itk} - \hat{X}_{itk}}{C_{itk} - 1} \right).$$

### Estimates of total abundance by run

The derivation of abundance estimates proceeds stepwise from a simple conceptual model to the final variations used here. We first construct estimators that account for sampling variation and measurement error in corrected run assignments ( $\hat{x}$ ). We then account for uncertainty due to missing days of trawl sampling, followed by uncertainty in trawl efficiency estimates.

To begin, we ignore length stratification and assume that a known number ( $x_i$ ) of fish of a given run  $i$  are identified in a sample ( $S$ ) of trawl catch ( $C$ ) collected during a discrete time period. Let  $f (= S/C)$  be the fraction of catch that is sampled and assigned to run. We want to estimate the total abundance of juveniles  $N_i$  given the observation  $x_i$ . It is assumed that all fish are independent have the same probability ( $E$ ) of capture, where  $E$  is the trawl efficiency, and the

same probability ( $f$ ) of being sampled if caught. Given these assumptions, we could specify  $x_i \sim \text{Binomial}(N_i, Ef)$ ; however, because  $E$  is expected to be very low (e.g., 0.01 or 1%), we can simplify the model by specifying  $x_i \sim \text{Poisson}(N_iEf)$ .

The trawl does not operate continuously; rather, it is further assumed that trawl sampling provides a representative sample of the migration ( $N_i$ ) passing Chipps Island during a given period (e.g., a day). Let  $p$  denote the proportion of the period trawled. Assuming fish passage is random throughout the period, we can now specify:

$$(A11) \quad x_i \sim \text{Poisson}(N_iEpf); \quad E[x_i] = N_iEpf; \quad V[x_i] = N_iEpf$$

where  $Ep$  is the probability of capture across the full period.

In our application, we have an estimator  $\hat{x}_i$  (assignment correction) rather than a direct observation of the true abundance  $x_i$ . Given Equation (A11), and following the same steps outlined for total catch estimates (see Equations (A3)-(A8)), we obtain the following expression for  $N_i$ :

$$(A12) \quad N_i = \frac{E[\hat{x}_i]}{Epf}$$

and its variance

$$(A13) \quad \begin{aligned} V[N_i] &= \frac{1}{(Epf)^2} (V[\hat{x}_i | x_i] + V[x_i]) \\ &= \frac{1}{(Epf)^2} (V[\hat{x}_i | x_i] + N_iEpf). \end{aligned}$$

To obtain estimators from Equations (A12) and (A13), we substitute  $E[\hat{x}_i]$  and  $V[\hat{x}_i | x_i]$  with the observed estimates  $\hat{x}_i$  and  $\hat{\sigma}_{\hat{x}_i}^2$ , and  $N_iEpf (= E[x_i])$  with  $\hat{x}_i$  as well. This gives

$$(A14) \quad \hat{N}_i = \frac{\hat{x}_i}{Epf}$$

with variance estimator

$$(A15) \quad \hat{\sigma}_{\hat{N}_i}^2 = \frac{(\hat{\sigma}_{\hat{x}_i}^2 + \hat{x}_i)}{(Epf)^2}.$$

In summary, this simple variance estimator accounts for measurement error ( $\hat{\sigma}_{\hat{x}_i}^2$ ) in DNA assignment corrections and sampling variation in captures (i.e., assuming captures  $x_i$  follow a Poisson distribution).

In our application, we computed assignment corrections  $\hat{x}_{itk}$  for samples of catch that were stratified into biweekly periods ( $t$ ) and length strata ( $k$ ). Moreover, we were primarily interested in biweekly totals across length strata. We can modify Equation (A14) accordingly:

$$(A16) \quad \hat{N}_{it.} = \frac{1}{Ep_t} \sum_k \frac{\hat{x}_{itk}}{f_{tk}} = \frac{\hat{X}_{it.}}{Ep_t},$$

where  $\hat{N}_{it.}$  is the biweekly estimate of total abundance,  $\hat{X}_{it.}$  is the total catch estimate for run  $i$ , and  $p_t$  is a measure of the proportion of the period that was trawled. However, a potential problem with this formulation lies in the definition of  $p_t$ . In previous applications (e.g., USFWS 2006),  $p_t$  has been computed as an aggregate across days (e.g., the sum of minutes trawled in a month divided by the total minutes in a month). Such an approach will be biased when effort and catch per unit effort vary across days (a reasonable expectation). In addition, there were numerous missing days (i.e., days of no trawl sampling) in most of the biweekly periods we examined, and we wanted to account for uncertainty in catch when expanding observed catches to missing days.

To better account for missing days and variation in daily effort and catch, we estimated biweekly abundances by run (across length strata) as

$$(A17) \quad \hat{N}_{it.} = \frac{\hat{\rho}_{it}\hat{\gamma}_t}{E},$$

where  $\hat{\rho}_{it}$  is the estimated proportion of migrating juveniles composed of run  $i$  in period  $t$ , and  $\hat{\gamma}_t$  is an estimate of the expected total catch of juveniles (all runs) that would have been observed had the trawl operated *continuously* throughout the period. The estimate  $\hat{\rho}_{it}$  was given by

$$(A18) \quad \hat{\rho}_{it} = \frac{\hat{X}_{it.}}{C_t.} = \frac{1}{C_t.} \sum_k \frac{\hat{x}_{itk}}{f_{tk}},$$

where  $C_t.$  is the total observed catch of all juveniles (summed across length strata). The variance of  $\hat{\rho}_{it}$  can be expressed as

$$\begin{aligned}
\sigma_{\hat{\rho}_{it}}^2 &= \left( \frac{1}{C_t} \right)^2 \hat{\sigma}^2 \left[ \sum_k \frac{\hat{x}_{itk}}{f_{tk}} \right] \\
\text{(A19)} \quad &= \frac{1}{C_t^2} \left( \sum_k \frac{1}{f_{tk}^2} \left( \hat{\sigma}_{\hat{x}_{itk}}^2 + \hat{x}_{itk} \right) + 2 \sum_k \sum_{l \neq k} \frac{1}{f_{tk}} \frac{1}{f_{tl}} \text{cov}[\hat{x}_{itk}, \hat{x}_{itl}] \right) , \\
&= \frac{1}{C_t^2} \left( \sum_k \sum_l \frac{1}{f_{tk}} \frac{1}{f_{tl}} \left( \mathbf{y}'_{Atk} \hat{\mathbf{Q}}_{Ai} \mathbf{y}_{Atl} + \mathbf{y}'_{Btk} \hat{\mathbf{Q}}_{Bi} \mathbf{y}_{Btl} \right) + \sum_k \frac{\hat{x}_{itk}}{f_{tk}^2} \right).
\end{aligned}$$

Here, the variance of  $\hat{\rho}_{it}$  is computed with respect to total abundance, where the variance of each assignment correction  $\hat{x}_{itk}$  comprises sampling variation and measurement error (see Equation A15). Thus, in the final expression of Equation (A19), the first (double) summation is a compact expression for the sum of variances and covariances in assignment corrections (measurement error), while the second summation accounts for sampling variation (assumed to be a Poisson process).

The estimate  $\hat{\gamma}_t$  accounted for missing days (i.e., days with no trawling) as well as variation in catch per unit effort among days. Let subscript  $d$  denote day, let  $D_t$  be the total days in biweekly period  $t$ , and let  $M_t$  be the number of missing days (where  $M_t < D_t$ ). For each day of trawling, there is an observed total catch,  $C_{td}$  (across all runs and length strata), and a computed proportion of the day trawled,  $p_{td}$ . Assuming the days sampled ( $D_t - M_t$ ) represent a random sample within the period, we estimated  $\hat{\gamma}_t$  as

$$\text{(A20)} \quad \hat{\gamma}_t = \sum_d^{D_t - M_t} \frac{C_{td}}{p_{td}} + M_t \bar{c}_t = D_t \bar{c}_t$$

with variance given by

$$\text{(A21)} \quad \hat{\sigma}_{\hat{\gamma}_t}^2 = M_t^2 \left( \frac{s_{ct}^2}{D_t - M_t} \right) + M_t s_{ct}^2 ,$$

where  $\bar{c}_t$  and  $s_{ct}^2$  denote the sample estimates of mean and variance, respectively, of the set of ( $D_t - M_t$ ) daily observations  $\{C_{td} / p_{td}\}$ , which are analogous to catch-per-unit-effort data.

In our application, we estimated  $p$  as a standardized proportion of water volume trawled, which matched the definition for trawl efficiencies as estimated by Pyper et al. (2013). Specifically, we computed  $p$  for a given day  $d$  as:

$$(A22) \quad p_d = \frac{V_d}{1440 * v},$$

where  $V_d$  was the total daily volume of water sampled and  $v$  was an arbitrary scalar defining a “standard” rate of volume sampled. As in Pyper et al. (2013),  $v$  was set equal to 1000 m<sup>3</sup>/minute, such that the “standardized” daily volume sampled (i.e., for continuous 24-hour trawl operation) was 1440 minutes/day \* 1000 m<sup>3</sup>/minute = 1,440,000 m<sup>3</sup>/day.

In summary, in our use of Equation (A17), we computed the total number of fish ( $\hat{\gamma}_t$ ) that would have been caught had the trawl operated continuously throughout a biweekly period, and multiplied this amount by the estimated proportion of fish composed of run  $i$  during that period ( $\hat{\rho}_{it}$ ). This provided an estimate of catch for run  $i$ , expanded to account for trawl effort (i.e.,  $\hat{\rho}_{it} \hat{\gamma}_t$ ). Abundance was estimated by dividing this expanded catch by the trawl efficiency. Note that for certain conditions, such as constant effort ( $p_{td}$ ) across days, it is easy to show that Equations (A16) and (A17) provide equivalent expressions for abundance.

From Equation (A17), the variance of  $\hat{N}_{it}$  is given by approximate variance of a product of two independent random variables (Mood et al. 1974, p. 180):

$$(A23) \quad \hat{\sigma}_{\hat{N}_{it}}^2 \cong \frac{\hat{\rho}_{it}^2 \hat{\sigma}_{\hat{\gamma}_t}^2 + \hat{\gamma}_t^2 \hat{\sigma}_{\hat{\rho}_{it}}^2}{E^2},$$

which now incorporates potential error due to missing days of trawl sampling.

To obtain our final estimator for biweekly abundance, we substitute  $E$  with an estimate of  $E$ :

$$(A24) \quad \hat{N}_{it} = \frac{\hat{\rho}_{it} \hat{\gamma}_t}{\hat{E}}.$$

The variance estimator is given by the approximate variance of a ratio (Mood et al. 1974, p. 181), where the numerator variance follows from Equation (A23):

$$(A25) \quad \begin{aligned} \hat{\sigma}_{\hat{N}_{it}}^2 &\cong \frac{(\hat{\rho}_{it} \hat{\gamma}_t)^2}{\hat{E}^2} \left[ \frac{\hat{\sigma}^2[\hat{\rho}_{it} \hat{\gamma}_t]}{(\hat{\rho}_{it} \hat{\gamma}_t)^2} + \frac{\hat{\sigma}_{\hat{E}}^2}{\hat{E}^2} \right] \\ &\cong \frac{(\hat{\rho}_{it} \hat{\gamma}_t)^2}{\hat{E}^2} \left[ \frac{\hat{\gamma}_t^2 \hat{\sigma}_{\hat{\rho}_{it}}^2 + \hat{\rho}_{it}^2 \hat{\sigma}_{\hat{\gamma}_t}^2}{(\hat{\rho}_{it} \hat{\gamma}_t)^2} + \frac{\hat{\sigma}_{\hat{E}}^2}{\hat{E}^2} \right] \\ &\cong \frac{\hat{\gamma}_t^2}{\hat{E}^2} \hat{\sigma}_{\hat{\rho}_{it}}^2 + \frac{\hat{\rho}_{it}^2}{\hat{E}^2} \hat{\sigma}_{\hat{\gamma}_t}^2 + \frac{\hat{\rho}_{it}^2 \hat{\gamma}_t^2}{\hat{E}^4} \hat{\sigma}_{\hat{E}}^2. \end{aligned}$$

It is useful to decompose Equation (A25) further with respect to the variance of  $\hat{\rho}_{it}$  (Equation A19) to isolate each component of variation:

$$(A26) \quad \hat{\sigma}_{\hat{N}_{it}}^2 \cong \frac{\hat{\gamma}_t^2}{\hat{E}^2 C_t^2} \left( \sum_k \sum_l \frac{1}{f_{tk}} \frac{1}{f_{tl}} (\mathbf{y}'_{A tk} \hat{\mathbf{Q}}_{Ai} \mathbf{y}_{A tl} + \mathbf{y}'_{B tk} \hat{\mathbf{Q}}_{Bi} \mathbf{y}_{B tl}) + \sum_k \frac{\hat{x}_{itk}}{f_{tk}^2} \right) + \frac{\hat{\rho}_{it}^2}{\hat{E}^2} \hat{\sigma}_{\hat{\gamma}_t}^2 + \frac{\hat{\rho}_{it}^2 \hat{\gamma}_t^2}{\hat{E}^4} \hat{\sigma}_{\hat{E}}^2.$$

The four additive terms in Equation (A26) correspond respectively to (1) measurement error in DNA assignment corrections; (2) sampling variation in trawl captures; (3) estimation variance in catch (extrapolation of missing days); and (4) variation in the estimate of trawl efficiency. Potentially large components of variation have been omitted, specifically, temporal variation in efficiency and/or catch (e.g., overdispersion due to clumpy spatial and/or temporal patterns of fish migration).

Last, we estimated annual abundances by sample year as sums across  $n_t$  biweekly estimates:

$$(A27) \quad \hat{N}_{i..} = \sum_{t=1}^{n_t} \hat{N}_{it.} = \frac{1}{\hat{E}} \sum_{t=1}^{n_t} \hat{\rho}_{it} \hat{\gamma}_t,$$

with an approximate variance estimator given by

$$(A28) \quad \hat{\sigma}_{\hat{N}_{i..}}^2 \cong \frac{1}{\hat{E}^2} \sum_{t=1}^{n_t} \sum_{u=1}^{n_t} \left( \frac{\hat{\gamma}_t}{C_t} \frac{\hat{\gamma}_u}{C_u} \sum_{k=1}^{n_k} \sum_{l=1}^{n_k} \frac{1}{f_{tk}} \frac{1}{f_{ul}} (\mathbf{y}'_{A tk} \hat{\mathbf{Q}}_{Ai} \mathbf{y}_{A ul} + \mathbf{y}'_{B tk} \hat{\mathbf{Q}}_{Bi} \mathbf{y}_{B ul}) \right) + \frac{1}{\hat{E}^2} \sum_{t=1}^{n_t} \left( \frac{\hat{\gamma}_t^2}{C_t^2} \sum_{k=1}^{n_k} \frac{\hat{x}_{itk}}{f_{kt}^2} \right) + \frac{1}{\hat{E}^2} \sum_{t=1}^{n_t} \hat{\rho}_{it}^2 \hat{\sigma}_{\hat{\gamma}_t}^2 + \frac{\hat{\sigma}_{\hat{E}}^2}{\hat{E}^4} \sum_{t=1}^{n_t} \sum_{u=1}^{n_t} \hat{\rho}_{it} \hat{\gamma}_t \hat{\rho}_{iu} \hat{\gamma}_u.$$

In Equation (A28), variances as well as covariances are computed across all combinations of biweekly strata ( $t$  and  $u$ ) to account for dependencies due to blind-test data (used in all assignment corrections) and the estimate of trawl efficiency.

## Appendix B: Second most likely run assignments

For DNA-run assignments with ONCOR probabilities less than one, a second most likely run assignment was usually provided (Table B1). These data provide an alternative measure to the blind-test data of the direction of run-assignment uncertainty. Most fish that were assigned to fall-run (with assignment probabilities < 1) had a second most likely run assignment of late-fall-run, followed by spring Mill-Deer creek. A similar pattern existed for late-fall-run. Fish first assigned to spring Butte Creek had approximately the same second most likely run assignments belonging to fall and spring Mill-Deer creek, and again a similar pattern existed for fish first assigned to Spring Mill-Deer. There were only two winter-run assignments with probabilities < 1, and neither was assigned to a second run.

Table B1. Number of fish assigned to run by ONCOR first mostly likely run assignment and second most likely run assignment.

		ONCOR 1 <sup>st</sup> Most Likely Assignment				
		Fall	Late-fall	Spring Butte	Spring Mill-Deer	Winter
ONCOR 2 <sup>nd</sup> Most Likely Assignment	Fall	-	181	11	43	0
	Late-fall	409	-	1	3	0
	Spring Butte	11	1	-	15	0
	Spring Mill and Deer	96	9	13	-	0
	Winter	1	0	0	0	-
Total		517	191	25	61	0

Attachment A:



## TECHNICAL BREIF

**TO: Pat Brandes, USFWS**

**FROM: Brian Pyper, Casey Justice, and Steve Cramer**

**DATE: January 30, 2008**

**SUBJECT:** Sample size allocation for DNA analysis of juvenile Chinook salmon captured in Chipps Island midwater trawl

### Summary

The following provides a summary of the statistical background, analyses, and recommendations for the 2008 DNA sampling plan for Chipps Island trawl. This memo synthesizes the key findings discussed in the two previous memos (dated January 18, 2008 and January 24, 2008).

### Objective

The goal of the DNA analysis is to identify juveniles that are either winter-run or spring-run Chinook salmon, such that the total juvenile abundance of each race passing Chipps Island can be estimated. The objective here is to allocate DNA samples over time to maximize the precision of total abundance estimates. We consider both winter-run and spring-run abundance estimates, bi-weekly sampling periods, length criteria for stratifying catches and samples, and a total sample size for DNA analysis of 3000 juvenile Chinook salmon.

### Statistical Framework for Optimal Sample Allocation

First, we consider estimates pertaining to a single race and time period, denoting variables as follows:

Variable	Description
$T$	Total abundance of Chinook juveniles passing Chipps Island
$N$	Abundance of the race of interest (e.g., winter-run or spring-run)
$C$	Trawl catch
$S$	Total number of juveniles sampled from the catch for DNA analysis
$X$	Number of juveniles in the sample identified as the race of interest
$p = N/T$	The proportion of total juveniles composed of the race of interest
$e = C/T$	Trawl efficiency
$f = S/C$	Fraction of catch that is sampled

Assuming that all juveniles have an equal probability of capture ( $e$ ), an equal probability of being sampled if caught ( $f$ ), and are sampled without replacement, then the number ( $X$ ) of juveniles of the race of interest identified in the sample follows the hypergeometric distribution with expectation

$$(1) \quad E[X] = N * e * f$$

and variance

$$(2) \quad \text{var}[X] = \frac{S * N * (T - N) * (T - S)}{T^2 * (T - 1)} .$$

With a little algebra, and replacing  $(T - 1)$  with  $T$ , the variance can be expressed as:

$$(3) \quad \begin{aligned} \text{var}[X] &\cong S * p * (1 - p) * (1 - e * f) \\ &\cong S * p * (1 - p) \end{aligned} .$$

The term  $(1 - e * f)$  can be ignored because trawl efficiencies are extremely low (e.g., 0.1%). Hence, the expected variance of  $X$  is essentially the same as assuming that  $X$  follows a binomial distribution with sample size  $S$  and binomial probability  $p$ .

From equation (1), it follows that an estimate of the total abundance ( $N$ ) of the race of interest is given by:

$$(4) \quad \hat{N} = \frac{X}{e * f} .$$

For simplicity, we assume that both the trawl efficiency ( $e$ ) and the sampling fraction ( $f$ ) are known. From equations (3) and (4), the variance of the abundance estimate can be approximated by:

$$(5) \quad \hat{\sigma}_{\hat{N}}^2 = \left( \frac{1}{e * f} \right)^2 \text{var}[X] = \frac{S * \hat{p} * (1 - \hat{p})}{(e * f)^2} = \frac{C^2 * \hat{p} * (1 - \hat{p})}{e^2 * S} .$$

where  $p$  is replaced by the proportion estimated from the DNA sample ( $\hat{p} = X / S$ ).

#### Multiple Time Periods, Length Classes and Races

Catches may be stratified by time and length criteria. For example, length-at-date criteria have been developed to provide rough designations of Chinook juveniles as winter-run, spring-run, fall-run, etc. (e.g., Figures 1 and 2). For a given race, the total abundance estimate across  $K$  discrete time periods  $t$ , and across  $L$  length classes  $i$ , is the sum:

$$(6) \quad \hat{N} = \sum_{t=1}^K \sum_{i=1}^L \hat{N}_{t,i} = \frac{1}{e} \sum_{t=1}^K \sum_{i=1}^L \frac{X_{t,i}}{f_{t,i}},$$

with variance

$$(7) \quad \hat{\sigma}_{\hat{N}}^2 = \frac{1}{e^2} \sum_{t=1}^K \sum_{i=1}^L \frac{C_{t,i}^2 * \hat{p}_{t,i} * (1 - \hat{p}_{t,i})}{S_{t,i}}.$$

Here, we assume that trawl efficiency ( $e$ ) is constant across time periods and length classes within a given season, which is consistent with current methods used to expand Chipps Island trawl estimates (USFWS 2000 and 2003). To determine the optimal allocation of sample sizes  $\{S_{t,i}\}$  that minimizes the variance, we take the derivative of equation (7) with respect to  $S$ , set the result equal to zero, and solve for  $S_{t,i}$ . The following “optimal” allocation results:

$$(8) \quad S_{t,i} = \frac{Y * C_{t,i} * \sqrt{p_{t,i} * (1 - p_{t,i})}}{\sum_{t=1}^K \sum_{i=1}^L C_{t,i} * \sqrt{p_{t,i} * (1 - p_{t,i})}},$$

where  $Y$  denotes the total sample-size constraint:

$$(9) \quad Y = \sum_{t=1}^K \sum_{i=1}^L S_{t,i}.$$

In addition, sample size ( $S_{t,i}$ ) must be less than or equal to the catch ( $C_{t,i}$ ) for each combination of time period ( $t$ ) and length class ( $i$ ).

We are interested in maximizing the precision of two estimates of total juvenile abundance (winter-run and spring-run Chinook salmon). Given the objective of minimizing the overall variance of the abundance estimates (i.e., the sum of their variances), the following allocation results:

$$(10) \quad S_{t,i} = \frac{Y * C_{t,i} * \sqrt{p_{1,t,i} * (1 - p_{1,t,i}) + p_{2,t,i} * (1 - p_{2,t,i})}}{\sum_{t=1}^K \sum_{i=1}^L C_{t,i} * \sqrt{p_{1,t,i} * (1 - p_{1,t,i}) + p_{2,t,i} * (1 - p_{2,t,i})}},$$

where the subscripts “1” and “2” distinguish the two races.

### **Application to Chipps Island trawl**

To provide a rough guide for allocating DNA samples, we examined catch data for Chipps Island trawl for 11 sampling seasons (the 1996 season through the 2007 season). Specifically, we used daily catches by

length category (1 mm intervals for fork length) as provided by Pat Brandes (“CHN Forked.xls”). Across years, length measurements were taken for approximately 92% of the total juvenile catch.

We used the catch-by-length data and daily length criteria (provided by Sheila Green) to determine catches for bi-weekly periods (e.g., March 1-15, March 16-31, April 1-15, April 16-30, May 1-15, etc.) for three length classes (winter length or greater; spring length; and fall length or less). The winter-length class we used was based on the length criteria for the salvage data (e.g., Figure 1) rather than Chipps, but this likely does not matter because the optimal sampling designs suggest that all winter-run length fish should be sampled (we don’t expect that to change because even with the lower length criteria, the expected number of captures will still be low, e.g. 400 or less). In any case, the analysis has been setup so that it can be quickly updated for different length criteria.

Table 1 shows the average Chipps Island trawl catches by length class across 1996-2007 for December through June. Refer to Figure 1 for length class designations.

**TABLE 1. Average catch (1996-2007) by length class of juvenile Chinook in Chipps Island trawl.**

		Catch by Length Class		
		BiWeek	Winter +	Spring
<b>Dec</b>	1-15	20	0	0
	16-31	16	0	0
<b>Jan</b>	1-15	10	0	1
	16-31	10	0	26
<b>Feb</b>	1-15	6	0	111
	16-28	15	1	44
<b>Mar</b>	1-15	23	15	25
	16-31	16	223	33
<b>Apr</b>	1-15	3	503	246
	16-30	1	1151	3132
<b>May</b>	1-15	0	287	4764
	16-31	0	80	2885
<b>Jun</b>	1-15	0	3	691
	16-30	0	0	126
<b>Total</b>		<b>121</b>	<b>2263</b>	<b>12084</b>

Note that the average for “fall length” in the first half of February (111, shaded cell) was driven by catches for one year (1996). Thus, the total catch for all length classes through the end of March is expected to be low (roughly 600), with roughly 400 of these being “spring length” or larger. Only two years had considerably higher catches through the end of March (1996 and 1998).

Catches of “spring length” fish peak in April, particularly the second half of April. Catches of “fall length” are expected to be high from mid-April through the end of May.

As discussed below, we computed optimal sample-size allocations based on the catch-by-length data and additional assumptions regarding proportions of winter-run and spring-run juveniles in the catch. Analyses were conducted separately for each year (1996-2007), as well as for the average catches across years. We found that sample allocations based on average catches provided a reasonable summary and generalization of the year-specific results; hence, we only report the allocation results for the average catches (Table 1). These analyses are contained in the attached spreadsheet (“Chipps\_DNA\_sample\_size.xls”).

### Sample sizes

The optimal sample sizes by length class and period depend critically on assumptions about the proportions of catch composed of either winter-run or spring-run juveniles (see equation 10). While additional data/analyses could be conducted in an attempt to better estimate such proportions, we have arbitrarily selected values as an example; these are shown in Table 2. The DNA analysis completed for salvage (e.g., Figure 1) suggests winter-run are largely found within the “winter” length class, and to a lesser extent in the “spring” length class, with few winter-run expected after April. Spring-run are expected as both yearlings and sub-yearlings, and show a variety of length/timing patterns across years for the DNA analysis completed for salvage (e.g., Figure 2). The implications of the proportions we selected (Table 2) in terms of the distributions of length classes and migration timing within a given race are computed and displayed within the spreadsheet provided (“Chipps\_DNA\_sample\_size.xls”).

Given the catches in Table 1, the proportions in Table 2, and a total sample size of 3,000 fish, we computed optimal sample sizes for each period and length class (winter, spring, and fall) (see equation 10). Sample sizes are shown in Table 3, and implied fractions of the catch sampled are shown in Table 4.

The optimal allocation suggests sampling 100% of winter-length fish throughout the season (expected sample size = 121). For spring-length fish, the allocation implied 100% sampling through mid-March (expected sample size = 16), 74% in late March (sample = 165), and roughly 50% of spring-length fish beginning in April onward. Although a much lower fraction of fall-length fish should be sampled, most of the samples are allocated to this length class. Because we assumed that the relative proportion of spring-run juveniles in the fall-length class would decline in late April, May and June (Table 2), a lower proportion of these catches were sampled (Table 4) compared to early April.

**TABLE 2. Assumed proportions of Chipps Island trawl catch by length class composed of winter-run and spring-run Chinook.**

	BiWeek	Proportion of Length Class Composed of <u>Winter-run</u>			Proportion of Length Class Composed of <u>Spring-run</u>		
		Winter +	Spring	Fall	Winter +	Spring	Fall
<b>Dec</b>	1-15	75.0%	--	--	10.0%	--	--
	16-31	75.0%	--	--	10.0%	--	--
<b>Jan</b>	1-15	75.0%	--	--	10.0%	--	--
	16-31	75.0%	--	--	10.0%	--	--
<b>Feb</b>	1-15	75.0%	--	1.0%	10.0%	--	1.0%
	16-28	75.0%	25.0%	1.0%	10.0%	10.0%	1.0%
<b>Mar</b>	1-15	75.0%	25.0%	1.0%	10.0%	10.0%	1.0%
	16-31	75.0%	5.0%	1.0%	10.0%	10.0%	1.0%
<b>Apr</b>	1-15	75.0%	2.0%	1.0%	10.0%	5.0%	1.0%
	16-30	75.0%	0.5%	0.1%	10.0%	5.0%	0.5%
<b>May</b>	1-15	--	0.0%	0.0%	--	5.0%	0.5%
	16-31	--	0.0%	0.0%	--	5.0%	0.5%
<b>Jun</b>	1-15	--	0.0%	0.0%	--	5.0%	0.2%
	16-30	--	0.0%	0.0%	--	5.0%	0.2%

**TABLE 3. Optimal sample allocations for DNA analysis by length class.**

		<b>Samples by Length Class</b>			
		<b>BiWeek</b>	<b>Winter +</b>	<b>Spring</b>	<b>Fall</b>
<b>Dec</b>	1-15	20	0	0	
	16-31	16	0	0	
<b>Jan</b>	1-15	10	0	0	
	16-31	10	0	0	
<b>Feb</b>	1-15	6	0	31	
	16-28	15	1	12	
<b>Mar</b>	1-15	23	15	7	
	16-31	16	165	9	
<b>Apr</b>	1-15	3	259	69	
	16-30	1	524	481	
<b>May</b>	1-15	0	124	668	
	16-31	0	35	405	
<b>Jun</b>	1-15	0	1	61	
	16-30	0	0	11	
<b>Total</b>		<b>121</b>	<b>1124</b>	<b>1756</b>	
<b>(All)</b>				<b>(3000)</b>	

**TABLE 4. Fraction of catch sampled under optimal sample allocations.**

		<b>Fraction of Catch Sampled</b>			
		<b>BiWeek</b>	<b>winter +</b>	<b>spring</b>	<b>fall</b>
<b>Dec</b>	1-15	100%	0%	0%	0%
	16-31	100%	0%	0%	0%
<b>Jan</b>	1-15	100%	0%	0%	0%
	16-31	100%	0%	0%	0%
<b>Feb</b>	1-15	100%	0%	28%	28%
	16-28	100%	100%	28%	28%
<b>Mar</b>	1-15	100%	100%	28%	28%
	16-31	100%	74%	28%	28%
<b>Apr</b>	1-15	100%	52%	28%	28%
	16-30	100%	46%	15%	15%
<b>May</b>	1-15	0%	43%	14%	14%
	16-31	0%	43%	14%	14%
<b>Jun</b>	1-15	0%	43%	9%	9%
	16-30	0%	43%	9%	9%
<b>Overall</b>		<b>100%</b>	<b>50%</b>	<b>15%</b>	<b>15%</b>

We examined various (but seemingly reasonable) values for the proportions in Table 2, and the allocation results did not change all that much (can be tested in the spreadsheet provided). Given the expected catches (Table 1) and the sampling results (Tables 3 and 4), we suggest the following guidelines for sampling (Table 5). It seems prudent to simplify the sampling “rules” as much as possible, while recognizing that each year shall present a different distribution of catches, and yet also recognizing that rough approximations to the optimal design should still yield “almost” optimal results.

**TABLE 5. Guidelines for Chipps Island trawl DNA sampling plan.**

<b>Period</b>	<b>Winter-length (plus)</b>	<b>Spring-length</b>	<b>Fall-length</b>
Dec-March	All	All (max 250)	All (max 100)
April 1-15	All	250	100
April 16-30	All	500	500*
May 1-15	All	150	500*
May 16-31	All	50	500*
June	All	All (expect < 10)	100
<b>Total</b>	<b>Expect ~120 (up to 200)</b>	<b>1,200</b>	<b>1,800</b>

\* If possible, collect additional tissue samples (e.g., 1,000 total) for post-season sub-sampling.

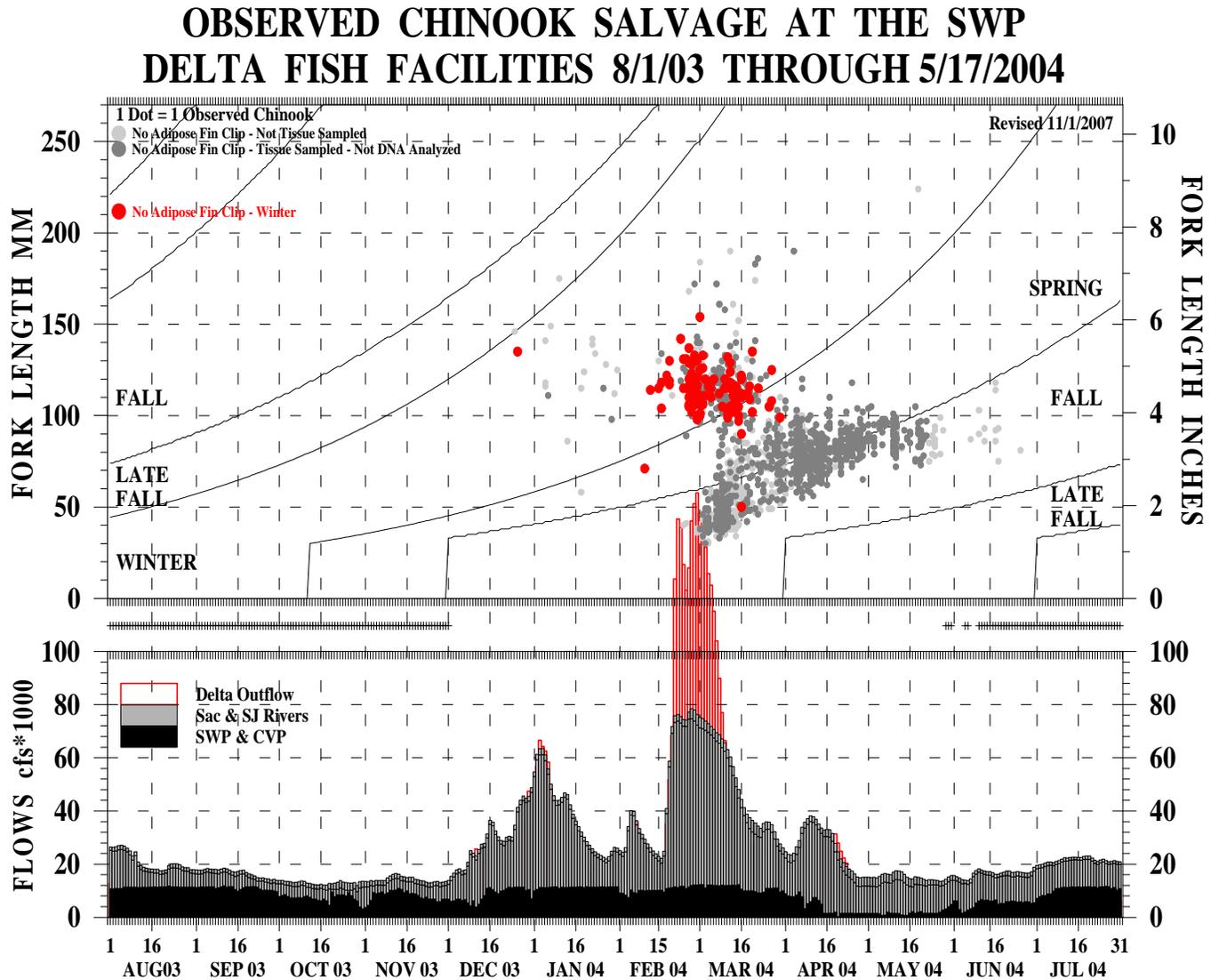
Note, the expected numbers of winter-length fish and sample sizes for spring-run assume the salvage winter-length criteria; we’ve mimicked the Chipps length-criteria and found that the average annual catch of winter-length fish (Table 1) increased from roughly 120 to 215 fish. This has little effect on the optimal design.

*Additional Considerations*

We want to obtain 3,000 samples, so if numbers of samples are falling short due to lack of spring-length fish, for example, we should update the sampling plan as needed within the sampling season. We recommend reviewing the status of samples every two weeks, beginning at the end of March.

By examining the year-specific sample allocations (1996-2007), we found that catches (and therefore optimal sample sizes) were quite variable for fall-length juveniles across three bi-weekly periods: April 16-30; May 1-15; and May 16-31 (marked with “\*” in Table 5). Ideally, a similar fraction or proportion of each of these catches should be sampled (Table 4), but we cannot know the distribution of catches ahead of time. Thus, the sample allocation for these catches would be enhanced if a “surplus” of tissue samples could be collected. For example, if 1000 tissue samples were collected in each period, these could be sub-sampled at the end of the season such that allocation was in proportion to late-April/May catches and in accordance with the overall sample constraint of 3000.

**Figure 1.** Juvenile Chinook recovered at SWP Delta fish facilities that were identified via DNA analysis as winter-run Chinook (figure provided by Sheila Green, CA Department of Water Resources).



**Figure 2.** Juvenile Chinook recovered at SWP Delta fish facilities that were identified via DNA analysis as spring-run Chinook (figure provided by Sheila Green, CA Department of Water Resources).

