Trinity River Restoration Program

Juvenile Salmonid Outmigrant Monitoring Evaluation, Phase II

Final Technical Memorandum

December 21, 2009

Prepared for:

Trinity River Restoration Program
1313 S. Main Street
P.O. Box 1300
Weaverville, California 96093

Prepared by:

ESSA Technologies, Ltd.
1765 W 8th Avenue
Vancouver, British Columbia V6J 5C6

Simon Fraser University
Department of Statistics
and Actuarial Science
8888 University Drive
Burnaby, British Columbia V5A 1S6

North State Resources, Inc.
5000 Bechelli Lane, Suite 203
Redding, California 96002
(530) 222-5347

USBR TASK ORDER 06A0204097G
NSR# 10116-02

Citation: **Schwarz, C.J., D. Pickard, K. Marine and S.J. Bonner.** 2009. Juvenile Salmonid Outmigrant Monitoring Evaluation, Phase II– December 2009. Final Technical Memorandum for the Trinity River Restoration Program, Weaverville, CA. 155 pp. + appendices.

Executive Summary

The Trinity River Restoration Program (hereafter called the Program) was initiated in 1984 to restore and maintain the fish and wildlife stocks of the Trinity River Basin to levels that existed just prior to construction of the CVP Trinity River Division. Using an adaptive management framework the Program hopes to determine the most effective restoration strategies to restore the Trinity River's natural riverine processes and enhance fish and wildlife populations. The Program's Integrated Assessment Plan (IAP) identifies key assessments that can be used to evaluate long-term progress toward achieving Program goals and objectives and/or provide short-term feedback to improve Program management actions by testing key hypotheses and reducing management uncertainties. This report addresses juvenile salmonid outmigrant monitoring (a key component of the IAP), evaluating tradeoffs between alternative monitoring methods for assessing smolt abundance, run timing, and condition. In addition, we assess the ability of the monitoring data to inform Program restoration goals for affected salmon species.

Abundance and run timing

We propose a new Bayesian spline-based methodology to estimate salmon abundance and run timing which provides several compelling advantages over the more traditional pooled or stratified Peterson estimator. In particular, the ability to share data among weeks within the Bayesian approach allows greater flexibility in terms of handling missing data. This feature may enable the Program to reduce the frequency of required mark-recapture events during periods with fewer smolt outmigrants without affecting estimates of precision.

Flow-based methods using the fraction of discharge sampled appear to capture the general shape of the outgoing migration pattern quite well. However, estimated capture-efficiencies based on sampled flow volumes underestimate actual screw trap efficiencies as measured by mark-recapture methods. This implies that the flow-based methods may underestimate the actual number of outgoing migrants. Unfortunately, the relationship between the flow-based and mark-recapture efficiencies may vary considerably across years, even within the same study. Further work is needed to identify underlying reasons for such wide variation before flow-based estimates can reliably be applied to years lacking a supporting mark-recapture study.

Our results suggest that a hybrid approach may be most suitable for application in future years, as the relationship appears to remain fairly consistent across weeks within a year. This suggests that undertaking flow measurements over the entire season supplemented with mark-recapture experiments in a few weeks to calibrate screw traps and establish the relationship should work quite well, particularly in cases where continuous electronic flow monitoring is possible.

Estimates of smolt run timing can be obtained fairly easily from the spline-based methods based on the estimate of the population passing the screw-trap in each week. The method implicitly assumes that smolt passage is uniform within the week which is clearly not the case. Error introduced by this assumption is small, however, relative to sampling errors.

A key requirement to ensure that estimates of run times and abundance are sensible, is that the study covers the entire smolt migration period (or at least that the number of fish moving outside the study window is negligible). It is possible to use the spline-based methods to interpolate outside the study's temporal boundaries. Given that the tail-end of the study usually has few fish, interpolation after the study period is unlikely to be problematic. However, for several of the datasets large numbers of fish were

already passing the screw traps when the study began so that any interpolation prior to study initiation would be highly problematic.

Condition

Several potential fish size and condition metrics were initially considered for use in long-term evaluations of the program. Fork length has the longest time series available and was the only condition-related metric selected for detailed evaluation in this report. Assumptions around sampling fork length are not as rigorous as those for abundance and run timing. For example varying the sampling windows across years does not affect fork length estimates as much as it does the abundance and run timing estimates. There are substantial data for all three species of interest (i.e., coho, steelhead, and Chinook salmon); in fact fork length is the only dataset with sufficient data to complete any analyses for coho. A limitation in use of fork length data is that it is unclear how fork length would be expected to change in response to the restoration activities. We explored several hypotheses for the purpose of this report, but more effort should be invested to clarify specific hypotheses and associated metrics. The metrics should consider run timing in relation to fork length (e.g., is it more important to know the size of the early outmigrants or the late outmigrants?). We recommend using a smoothed metric to remove the day-to-day variability in the fork length data, rather than simply using the raw data.

Outmigrant data as a program tool

The key challenge in evaluating across-year trends in outmigrants is the small time series available for most measures. The limiting factor in detecting longer term trends is the process error. The sampling error can be controlled by adjusting effort within each year, but in many cases, sampling error is small relative to process error. Unless additional covariates can be obtained to remove some of the process error, this large variation apparent in the response measures between years results in studies with low power. Power can be increased appreciably where these studies can be continued for ten or more years.

Some of the variability across years, especially in terms of log(abundance) and run timing, results from the particular survey timing each year (e.g., the sampling window varies and in some years may miss the beginning of the run). Certain metrics are easier to measure and less sensitive to study design. For example, fish condition as measured by fork length is relatively easy and inexpensive to measure over the course of a year. However, it is uncertain how condition of outmigrating smolts relates to improvements in fish rearing due to habitat improvements. Abundance would seem to be a more direct measure of overall improvement, but has high process variation and is expensive to measure.

Table of Contents

List of Figures

List of Tables

Appendices

Acknowledgements

The project team who prepared this report, included Carl Schwarz, Simon Fraser University; Keith Marine, North State Resources, Inc; Simon Bonner, University of British Columbia; Darcy Pickard and Katy Bryan, ESSA Technologies Ltd. We would like to thank Dr. Nina Hemphill, Trinity River Restoration Program, for funding this work, arranging logistics, and providing many valuable insights along the way.

We would also like to thank the many scientists who generously shared their data and their time including: Paul Petros, Eric Logan, and George Kautsky, Hoopa Valley Tribe; Bill Pinnix, Joe Polos U.S. Fish and Wildlife Service; Tim Hayden and Shane Quinn, Yurok Tribe. These biologists provided constructive feedback throughout the project, helping to interpret the volumes of data, prioritize and define the most important questions, and refine hypotheses relating outmigrant data to the restoration activities.

1. Introduction

1.1 Background

In 2008, we completed an independent technical review of the Trinity River juvenile salmonid outmigrant monitoring program. During this review we: summarized the history of outmigrant monitoring, collected and summarized the available data, reviewed the field component of the monitoring program, and provided a high level review of the population estimation methods (North State Resources et al. 2008). A second phase of the project was initiated to address some of the recommendations from the initial review and the Science Advisory Board, as well as some of the Priority Items To Address identified in the Trinity River Restoration Program (TRRP) Integrated Assessment Plan (TRRP and ESSA 2008). During this second phase we work with the data directly to evaluate alternative methodologies and generate baseline data that may be used to evaluate the success of the TRRP.

1.2 Objectives

The overall objective of this phase of the project was to investigate and develop appropriate analytic approaches and methodologies for using juvenile salmonid outmigrant data to evaluate the success of the TRRP and to transfer this knowledge to the TRRP and program partners for future use.

There were four major components to this work: data extraction and synthesis, estimation of metrics of interest (abundance, run timing, and fish condition), tools for evaluating the success of the TRRP, and providing a technology transfer workshop for the TRRP and program partners.

1.2.1 Data extraction and synthesis (Section 2)

We extracted the necessary data from the historic datasets that were obtained in Phase 1 of the project. We then compiled the data from all sources (Hoopa Valley Tribe (HVT); US Fish and Wildlife Service (USFWS)) across all years into a single Access database, compatible with the USFWS Trinity River outmigrant database. This task took longer than expected due to a number of inconsistencies found in the raw datasets. The USFWS is working on a new version of the database which will likely resolve these issues in the future.

1.2.2 Population Estimation (Sections 3-7)

We evaluated alternative methods for generating estimates for abundance (Sections 3, 4, and 5), and characterizing run timing (Section 6), and fish condition (Section 7). Abundance estimates were the primary focus for this component of the project. We produced annual estimates of abundance with associated confidence intervals for all years and species where sufficient data were available. We used several different methodologies (e.g., pooled Peterson, stratified Peterson, and Bayesian spline) and compared the strengths and weaknesses of each method under different conditions. Ultimately, we found that the Bayesian spline method was comparable to the stratified Peterson under ideal conditions, but that it was better able to handle typical problems with the data that occur in the Trinity River (e.g., no sampling within a stratum). The pooled Peterson method had numerous limitations and generally underestimated the variance of the estimator (i.e. the reported standard errors are too small).

Outmigrant data were collected for many years prior to initiation of a mark-recapture protocol to estimate trap efficiencies. Historically, raw catch data or flow based expansions were reported as an index of outmigrant abundance in annual reports. In recent years both flow and mark-recapture data are available. If appropriate, the TRRP and the Program partners would like to be able to use the flow based estimates gathered pre-2002 to provide baseline data to the Program. We evaluated this question in Section 5 and found that while the flow based estimates and the mark-recapture estimates were strongly correlated among Julian weeks within a year, the proportionality constant was not as consistent across years. Some of the years where the constant of proportionality is different can be explainable by "one-of-a-kind" factors, but this would leave only 2 or 3 years on which to estimate the common proportionality constant. Several more years of data and analysis would be recommended to see if the pre-2002 flow based estimates are a reliable index of abundance. However it may be reasonable to use them to obtain run timing estimates as the pattern of the run seems to be well captured.

Run timing metrics were obtained directly from the abundance estimates. We did not produce run timing metrics for datasets where abundance estimates weren't feasible.

Various metrics related to fish condition were considered. Fork length measurements were the only data historically available to assess fish condition. More recently weights have also been measured. Development of condition indices is being addressed via an IAP technical working group and so our discussion is limited to how we might use the historical fork length data and a summary of our suggestions/questions and comments received by the TRRP and Program partners regarding metrics of interest going forward.

For all metrics, we propose methods for estimating the appropriate response from hatchery and natural fish separately. This question was of particular interest to the TRRP and Program partners and was identified as a priority issue to address in the IAP.

1.2.3 Outmigrant data as a tool to evaluate the TRRP (Section 8)

During this component of the project we investigated how we might use the various outmigrant population estimates to evaluate the success of the TRRP. We worked with the TRRP and Program partners to identify a sample of hypotheses related to each metric. Using the methods recommended in the estimation section, we generate annual estimates for a common time period to enable cross year comparisons. We then evaluate the process error and assess the ability to detect changes described in the hypotheses. We provide several detailed examples of different types of questions including: comparing between different types of years (e.g., water year types); before and after a particular event (e.g., change in flows); and trend detection across years. The TRRP provided us with a timeline of relevant events (e.g., water year types, flow management changes, restoration actions etc…) which frames these examples. We leave it to the TRRP, Program partners, and IAP technical working groups to use the strategies illustrated here to test additional hypotheses as they move forward.

1.2.4 Technology Transfer Workshop

The final component of this project was to provide a technology transfer workshop to ensure that the methods employed in this report are shared and that the scientists who will be tasked with completing these analyses in the future have the opportunity to ask questions and discuss the findings in a collaborative setting. The workshop is a stand alone deliverable and will not be considered further in this report.

2. Data Extraction

2.1 Data extraction

The first objective of this task was to compile all of the available raw rotary screw trap data from 1993 to the present for the Junction City, Pear Tree, and Willow Creek monitoring sites. The second objective was to extract the necessary information from the data to complete the analyses in Sections 3 - Section 7. In this section we describe the available data, some of the problems we encountered during the detailed data review and extraction process, and we make several recommendations for future data management.

2.2 Summary of available data

Table 2.1-Table 2.5 summarize the available data provided by the Program partners and the TRRP. The raw data from all sources and the synthesized data are provided in compact disc format. At the time of the review Willow Creek data up to 2006 and Pear Tree data up to 2007 were made available. In both cases we understand that at least the same level of effort has continued and so we assume that these datasets extend through 2009. The Junction City site is no longer in use.

Table 2.1. Summary of the years and locations where rotary screw traps were running and catch data are available.

Site	Chinook salmon	Coho	Steelhead
Willow Creek	1993-2006	1993-2006	1993-2006
Pear Tree	2003-2007	2003-2007	2003-2007
Junction City	1997-2004	1997-2004	1997-2004

Table 2.2. Summary of the years and locations where mark-recapture data are available and mark-recapture estimates are feasible.

Site	Chinook salmon	Coho	Steelhead
Willow Creek	2002-2005	-	-
Pear Tree	2003-2007	-	2007
Junction City	2002-2004	-	2002-2004

Table 2.3. Summary of the years and locations where flow and temperature data are available at the trap site and where the hours of trap operation were documented.

Site	Flow through the traps	River temperature at the trap	Hours of trap operation
Willow Creek	1998-2006	1998-2006	Roughly ½ of the records are missing
Pear Tree	2003-2004, half – ¾ of the records are missing 2005-2007, roughly ¼ of the records are missing	2003-2007	No data
Junction City	1997-2002, no flow data except the trap revolutions/second 2003-2004, roughly ¼ of the records are missing.	1997, 1999-2001, 2003-2004	1997-2001

Table 2.4. Summary of the years and locations where USGS flow gauges for the Trinity River are available (source: IIMS).

Location Code	Location Name	Start Date	End Date	Record Count
11527000	Trinity River near Burnt Ranch	01-Oct-31	01-Jan-09	22374
11525900	Browns Creek	01-Jan-57	03-Oct-05	4121
11523200	Coffee Creek	01-Oct-57	01-Jan-09	18719
11528700	South Fork Trinity River below Hyampom	01-Oct-65	01-Jan-09	15799
11525600	Grass Valley Creek at Fawn Lodge	17-Nov-75	30-Sep-05	10911
11525655	Trinity River below Limekiln Gulch	28-Apr-81	01-Jan-09	6091
11526250	Trinity River at Junction City	21-Jun-95	30-Sep-02	2659
11525854	Trinity River at Douglas City	04-Nov-95	30-Sep-02	2523
11525530	Rush Creek	01-Oct-96	30-Sep-02	2191
11525670	Indian Creek Near Douglas City	29-Jan-97	30-Sep-04	2802
11525655	Trinity River below Limekiln Gulch	01-Oct-97	30-Sep-02	1826
11525520	Deadwood Creek	01-Oct-97	30-Sep-04	2557
	Reading Creek	01-Oct-00	30-Sep-04	1461
11525750	Weaver Creek near Weaverville	01-Oct-00	01-Oct-04	1462
11526300	Canyon Creek	01-Oct-01	30-Sep-04	1096
11526250	Trinity River at Junction City	01-Oct-02	01-Jan-09	2285
11525854	Trinity River at Douglas City	01-Oct-02	01-Jan-09	2284
11525530	Rush Creek	01-Oct-02	01-Jan-09	2112
11525535	Trinity River below Rush Creek (Salt Flat)	05-May-04	01-Aug-04	89
11525670	Indian Creek Near Douglas City	01-Oct-04	01-Jan-09	1554
11526400	Trinity River above North Fork Trinity	29-Mar-05	01-Jan-09	1375
11525600	Grass Valley Creek at Fawn Lodge	01-Oct-05	04-Oct-05	4
11525540	Trinity River above Grass Valley Creek	02-Apr-06	31-Jul-08	243
11526500	North Fork Trinity River	01-Oct-11	31-Oct-08	10141
11530000	Trinity River at Hoopa	01-Oct-11	01-Jan-09	29679
11525500	Trinity River at Lewiston	01-Oct-11	01-Jan-09	35522

Table 2.5. Summary of the years and locations where condition related data are available.

Site	Fork length	Weight	Health observations
Willow Creek	1993-2006	2004-2006	-
Pear Tree	2003-2007	2006-2007	-
Junction City	1997-2004	-	-

2.3 Extraction problems and assumptions

Data management is a substantial task, especially for a long-term program with many partners and it is not surprising that some problems were encountered during the data extraction task. The objective of this task was to extract basic outmigrant data from the variety of sources provided to us across three sites and as many as 25 years. The raw data were provided to the review team during Phase 1 of the project and were summarized in Appendix A (North State Resources et al. 2008). The formats provided ranged from text files to databases. Each site's data are managed separately, and until recently they used different data management systems. The USFWS developed an access database for outmigrant trapping data in 2005. The Willow Creek data are stored in the USFWS database from 1993-current except for the recapture data which are being tracked separately. The Hoopa Valley Tribe adopted the database fully in 2006, and has transferred records to the USFWS database back to 2002. In order to improve future data collection and analyses we summarized problems encountered with the data here. The problems generally fell under four categories: 1) problems which occurred during transfer from one data management system to another, 2) cohort separation, 3) version control/communication, and 4) incomplete records.

2.3.1 Transfer between data management systems

In the process of extracting the Junction City & Pear Tree data we found a few discrepancies in the recapture data, particularly from earlier years (2002-2005). HVT scientists helped us resolve these issues and provided useful context explaining that historically all of the data had been documented and stored in spreadsheets and that the structure of the spreadsheets did not match the database. All of the analyses were completed from the spreadsheets not the database. After the fact, they went back and imported the 2002-2005 data into the database; this process was quite complicated due to the different data structures (Paul Petros, pers. comm.). In 2006 HVT began entering their rotary screw trap data directly into the Access database. This likely explains why we didn't find any problems with the 2006-2007 data.

There were a number of other discrepancies among the different datasets which had to be resolved including:

- in some years calendar weeks not Julian weeks were used for mark-recapture protocols
- different environmental or trap data was documented in different datasets
- different naming conventions for species, age-classes, and origin (Appendix E)

2.3.2 Cohort separation

Each of the species of interest (Chinook salmon, coho, and steelhead) has a different outmigration life history. As described in the Trinity River Flow Study (USFWS and HVT 1999) and reproduced in Appendix D, only a small percentage of juvenile Chinook salmon overwinter in the Trinity. Coho have substantial outmigration at both age 0 (young of year fish) and age 1 (over wintering fish). Steelhead may outmigrate at age 0, 1, or 2 (there are a small number of age 3 records as well). For many of the analyses

(e.g., fork length), it is more meaningful to consider the different ages separately. True aging methods such as scale readings are possible, but these are time consuming and expensive. In practice age is often determined by looking for different modes within length data and trying to identify to which cohort each fish belongs (Hilborn and Walters 1992). The majority of the Trinity River records clearly report the age of the fish. Chinook salmon and coho juveniles in the Trinity River have been aged based on observations of their size and development, but scales are used to partition out steelhead age-classes (Pinnix et al. 2007). The earliest records at Junction City (1997-2001) do not record age but rather group coho and steelhead into several different size categories. Based on feedback from the Hoopa Valley Tribe we translated those data into age classes (Appendix E). Except for the early Junction City data (1997-2001) we used the ages provided in the datasets. For the most part these seem reasonable. However, there were a couple of years in the Willow Creek database that did not make sense (Figure 2.1). This is likely due to the fact that we did not have the most recent quality controlled version of the Willow Creek data available for the review. The USFWS has been improving their database, particularly the quality control aspects and so this type of inconsistency may have already been addressed.

Table 2.6. Summary of datasets where the aging did not seem reasonable.

Year	Site	Species	Comment
2005	Willow Creek	Steelhead	The average fork length between Julian Day 100-150 seems unusually high; could there be some age-1 fish here?
2006	Willow Creek	Coho	The average fork length between Julian Day 140-180 seems unusually high; could there be some age-1 fish here?
2006	Willow Creek	Steelhead	The average fork length between Julian Day 140-180 seems unusually high; could there be some age-1 fish here?

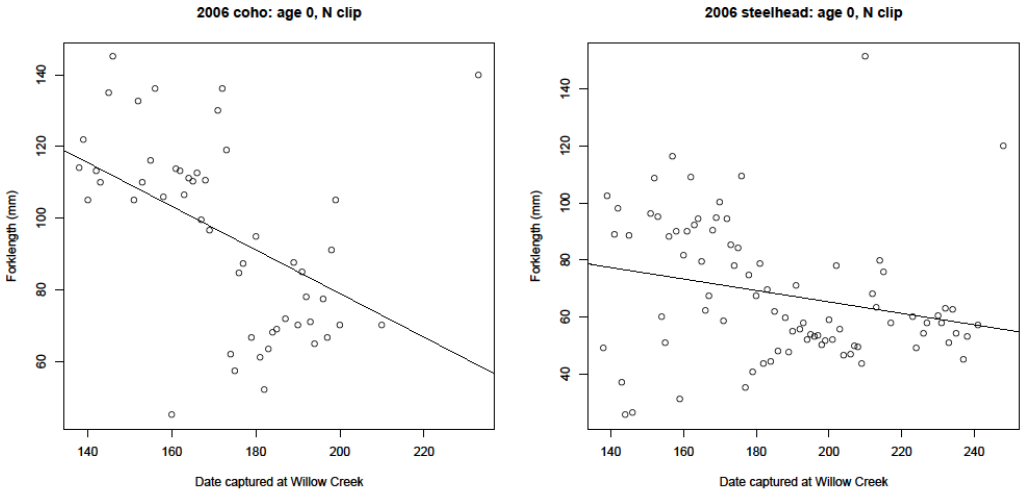


Figure 2.1. Two examples of years where the ages recorded don't appear accurate.

2.3.3 Version control/communication

Willow Creek Database:

- The original database (2005v1.1.mdb) provided to the review team was intended to be used as an example of the database's structural relationships only, not to be used for analyses. The USFWS has been working to quality control the data and improve the database itself. This misunderstanding led the review team to struggle with recapture data which was not sensible for a number of months before the USFWS was able to clarify that the database should not be used for analyses.
- The USFWS confirmed that the Willow Creek mark-recapture data have been tracked separately using excel spreadsheets. The rest of the data (catch and trap data) in the database were thought to be accurate (pers. comm. Bill Pinnix).
- The review team used the catch and trap data from the database, but used the summary mark-recapture data provided in the USFWS annual report appendices, which have been quality controlled.
- As a result of their concerns regarding the mark-recapture data entry system, the USFWS is working to improve the underlying code and data entry mechanics (which will result in stronger relationships between mark data and recaptures).

HVT database:

- The HVT is using an earlier version of the USFWS database (2005.mdb) which is similar but not identical to the Willow creek version we reviewed, and presumably will differ from the new version being developed by the USFWS.

2.3.4 Incomplete records

Missing data are a typical problem in data collection. This may happen for many reasons including: malfunctioning equipment, oversight, insufficient time, or an excess of data entry fields. During the course of this review we found many incomplete datasets including three of interest to the review team: 1) the number of hours the trap was running, 2) flow at the traps, and 3) water temperature.

2.4 Discussion and Guidance

The USFWS database overall appears to be a suitable tool for tracking the outmigrant data and has now been adopted by the HVT. There are a few minor suggestions to maintain transparency in the future:

- Complete database revisions to ensure the mark-recapture data are adequately quality controlled
- Ensure that everyone has the same, most current database.
- Use consistent naming practices across all sites or have clear conversions described
- Clarify how age-classes are defined in historical datasets (Appendix E)
- Summarize the assumptions that can be made when data are missing in different scenarios (e.g., interpolate between available flow data points). Provide clear instructions to the field technicians so that they understand when it is or is not ok to leave fields blank.

3. Population Estimation Methods

A key component of the Trinity River Restoration Project is estimating the number of outgoing young fish (Chinook salmon, coho, steelhead). The primary sampling protocol is the use of two-sample mark-recapture methods. In these methods, some fish are initially captured, marked, and released (traditionally called the "first" sample). A second sample is taken from the population and consists of both marked and unmarked fish. The ratio of marked-to-unmarked fish can be used to estimate the population size.

The simplest estimator for population size in two-sample mark-recapture experiments is the simple (pooled) Petersen estimator but this estimator requires that all individuals have the same probability of capture.

Stratification is commonly employed to avoid potential biases of abundance estimates and the corresponding measure of precision caused by heterogeneity in the capture probabilities among individuals. Instead of estimating the total abundance directly, the population is divided into groups of individuals, or strata, for which the capture probabilities can be assumed equal. Estimates are then computed separately for each stratum and summed to produce an estimate of the total population size.

A common stratification variable is by the *time* of release and recapture. These types of experiments are commonly used in fisheries research to monitor the number of salmon migrating along a river – either as juveniles moving from freshwater to the ocean (as in the TRRP) or as adults returning to the spawning grounds to breed. There are two common variants.

In the two-location variant, individuals are trapped, marked in some way, and returned to the population at the first location. Farther along the river (downstream for juveniles and upstream for adults) a second trap collects a sample which will contain both marked and unmarked fish (see Figure 3.1). The proportion of marked fish recaptured provides information about the capture probability in each stratum, which can be combined with the total number captured to estimate the population size.

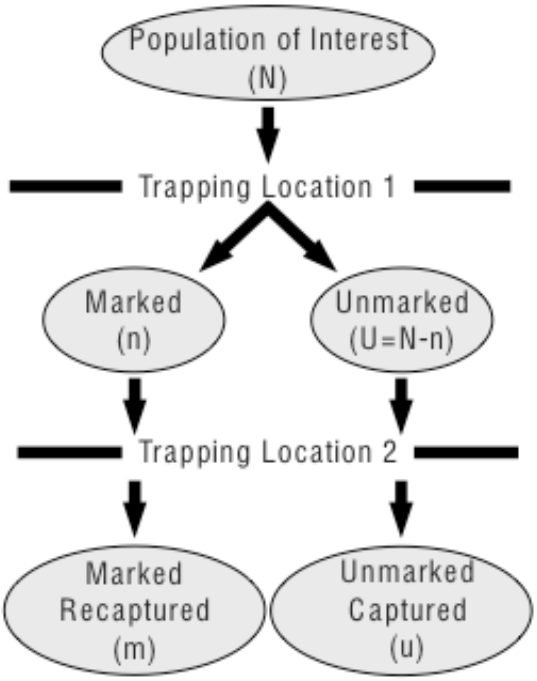


Figure 3.1. Schematic diagram of the standard two-sample capture-recapture experiment. A sample of size n_1 from the population of size *N* is captured at the first trapping location, marked and released. At the second location samples are obtained from the populations of both marked and unmarked fish. Notations for the number of fish in each group are included in brackets.

In the one location variant (as practiced in the TRRP), trapping is conducted at only one location in order to reduce the cost and effort required. Fish are captured at the trap location. They are marked, moved upstream of the trap, and then released into the river. Samples are then captured from both the population of interest and the marked fish. Figure 3.2 shows how this emulates the second trapping location in the experiment with two traps. The difference in analyzing the data from this experiment is that the population of interest does not normally include the marked fish. Because these fish were previously captured then they have already been included in the estimate of population size and do not need to be counted again.

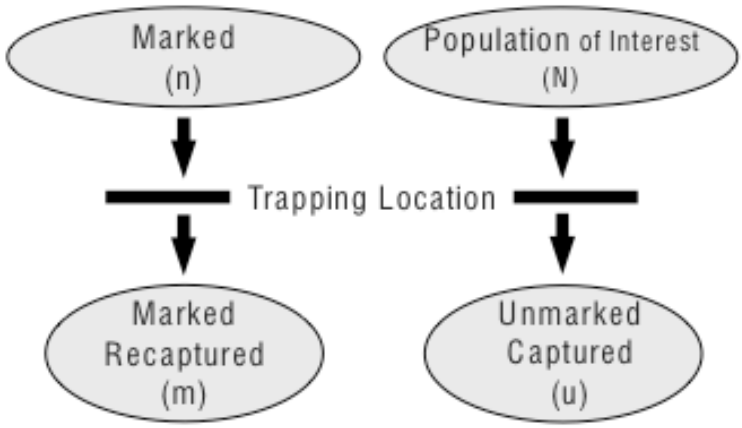


Figure 3.2. Schematic diagram of the modified two-sample capture-recapture experiment with only one trapping location. At the single location, fish are captured from the populations of both marked and unmarked individuals. Notations for the number of fish in each group are included in brackets. Marked fish are, most often, excluded from the population of interest to avoid double counting.

Under either variant, if it is reasonable to believe that the probability of being captured is the same for all fish, then no stratification by time is necessary as one of the critical assumptions of the simple Petersen estimator is then met. Assuming that the population is closed, i.e. that no fish enter or leave the population between the release and recapture locations, and that individuals behave independently, the probability of capture at the recapture-trap can be estimated by the proportion of marked individuals that were recaptured during the entire experiment and the simple Petersen estimator can be used.

However, fish migrations often last for several weeks and the capture probabilities may vary considerably over this time. For this reason, it is common to stratify the population by time (e.g. into weekly strata), estimating capture probabilities and population sizes separately for each stratum of the experiment. A challenge that arises in most experiments of this type is that fish marked and released during one stratum do not necessarily all pass the recapture site in the same stratum. Instead, some fish may migrate very quickly so that they pass the recapture-trap in the same time-stratum that they are released, while other fish may move more slowly and don't pass the recapture-site until later time-strata after being marked and released.

Note, that if a simple batch mark that is common for all fish has been applied, then it is not possible to know the stratum of release for recaptured fish and stratification is not feasible. A crucial aspect of the experimental protocol that allows such stratification is the use of individually numbered tags or batch marks (e.g. dye spots) that are specific to each stratum, so that the stratum of release and recovery of the marked fish can be determined. If unique marks are applied in each release stratum, it is possible to know when a recaptured fish was originally marked, and data from these experiments are commonly summarized by a matrix whose i, j entry indicates the number of fish marked on stratum i and recaptured in stratum j.

There are two cases. The matrix of recaptures will be diagonal if it is known that fish released in stratum i always pass the recapture location in stratum i and experiments generating such data will be referred to as

Time Stratified Petersen with Diagonal Recaptures Experiments (TSPDE). For example, data from a diagonal experiment with three strata might be displayed as:

Marked	Recaptured		
10	1		
10		1	
10			1
Unmarked	100	100	100

This indicates that in each stratum 10 fish were marked and released in each stratum and 1 of these was recaptured at the recapture location along with 100 more unmarked fish. This type of data might arise if marks and recaptures are stratified by week and the release and recapture locations are close enough to each other so that fish take at most a few hours to move from the first location to the second.

The second case is the more general experiment in which fish marked in stratum *i* may be recaptured outside of the stratum of release and will be referred to as the **Time Stratified Petersen with NonDiagonal Recaptures Experiment (TSPNDE)**. Sample data might be displayed as:

Marked	Recaptured		
10	1	1	1
10		1	1
10			1
Unmarked	100	100	100

This indicates that 10 fish were marked and released in each stratum. One fish from each was recaptured at the recapture location in each of the subsequent strata and a further 100 unmarked fish were captured in each stratum at the recapture location. The empty cells indicate structural zeros in the data; when the strata are based on time it is not possible for fish marked and released in one stratum to be recaptured in an earlier stratum.

The TRRP outmigrant monitoring program uses weekly batch marks, meaning it is only possible to identify the week in which a recaptured fish was marked, not the actual day. The majority of the TRRP recaptures occur during the same week as they are marked and released (i.e., the majority of the studies follow a diagonal experiment). As a result we will focus on the diagonal case in this report. If the TRRP were to begin using unique daily marks or unique individual marks (e.g., PIT tags) then the data could be examined on a finer scale and it is likely that the off diagonal cells will be non-zero. The analysis for a NonDiagonal recaptures experiment is similar but substantially more complex. The computer code has been provided to address this situation, but a detailed example isn't shown here because the diagonal structure is sufficient for the existing recapture data.

3.1 Introduction

In the simplest stratified experiment, fish cannot change strata between the time of release and the time of recapture and so the matrix of recapture is diagonal. Suppose that fish are marked and released for a total of *s* consecutive strata. Let

- n_i denote the number of fish marked and released in stratum *i*,

- m_{ii} the number of these fish that are released in stratum *i* and recaptured in stratum *i*, [1] and
- u_i is the number of unmarked fish captured at the recapture site.

The data from this experiment can be arranged in an array (with sample entries)

Marked	Recaptured		
n_{11}=10	m_{11}=1		
n_{12}=15		m_{22}=2	
n_{13}=20			m_{33}=3
Unmarked	u_1=100	u_2=100	u_3=100

Notice that the recapture matrix (the set of m_{ij}) is diagonal with blank entries representing zeroes.

Of interest is the total number of fish passing the recapture location in each of the strata. Let U_i denote the total number of unmarked individuals which pass the recapture site on day *i*.

The objective is to estimate U_1, K, U_s from which the total population size can be computed as $U_{total} = \sum_{i=1}^{s} U_i$. Note that the number of marked fish recaptured in a stratum, m_{ii}, is not included in the population size. It is not necessary to count these fish in the one trap design because they have already passed the trap once and have been counted in the quantity u_i. The number of marked fish does need to be accounted for in a two trap design. If "outside" fish (e.g. hatchery fish) are released upstream, then they have not been previously counted and the number of "outside" fish are added to U_{total} after the experiment is finished to estimate the total population size.

3.2 Pooled Petersen estimators

The simplest possible estimator is the (completely) pooled-Petersen estimator where the total number of marked fish, total number of recaptures, and total number of unmarked fish are found over all strata, and the usual Lincoln-Petersen estimator is computed as:

$$\hat{U}_{pp} = \frac{\left(\sum n_i\right)\left(\sum u_i\right)}{\sum m_{ii}}.$$

The pooled-Petersen estimator will be appropriate under the following assumptions:

- the population is closed (i.e., no fish enter or leave the population between the release and recapture locations);
- marks are not lost between the point of release and recapture;
- all fish over the entire experiment have the same probability of being captured at the recapture site (homogeneity);

[1] Double subscripts will be used on the marked fish recaptures because this notation easily extends to the non-diagonal case.

- and whether or not any individual is captured at the recapture site is independent of the capture of all other individuals.

The assumption of homogeneity is the key problematic assumption for the Trinity River projects – it is unlikely that the probability of capture is the same in all weeks of the study. If this assumption is violated, the pooled-Petersen estimator is relatively unbiased, but the reported standard error is unrealistically too small, i.e. the estimates appear to be more certain that they really are.

The assumption of homogeneity of capture in all strata can be assessed using a traditional chi-square test for homogeneity of proportions. A *2xs* contingency table is created:

Stratum	**Recaptured**	**Not-recaptured**
1	m_{11}	$n_1 - m_{11}$
2	m_{22}	$n_2 - m_{22}$
s	m_{ss}	$n_s - m_{ss}$

The test-statistic is computed as

$$X^2 = \sum_{i=1}^{s} \frac{\left(m_{ii} - E\left[m_{ii}\right]\right)^2}{E\left[m_{ii}\right]} + \sum_{i=1}^{s} \frac{\left(n_i - m_{ii} - E\left[n_i - m_{ii}\right]\right)^2}{E\left[n_i - m_{ii}\right]}$$

where the expected number of marks returned in stratum *j* is found as $E\left[m_{ii}\right] = n_j \frac{\sum_{i=1}^{s} m_{ii}}{\sum_{i=1}^{s} n_i}$ (i.e. using the average recapture rate over all strata). The test-statistic can be compared to χ^2_{s-1} distribution to see if it is unusually large which would indicate that pooling is not advisable.

The usual cautions must be employed when using this test when some of the expected counts are small (i.e. less than 3) as this can inflate the test statistic.

The pooled-Petersen estimator is also problematic when problems occur in the study. For example, there may be weeks in the study when no fish are marked and released (it will be necessary to assume that the capture rate for these weeks is the same as other weeks), or worse, there may be weeks in the study when no unmarked fish are captured (estimates will be biased low).

3.3 The Stratified-Petersen estimator

If the assumption of homogeneity is not appropriate, a stratified-Petersen estimator can be computed. In essence, the stratified-Petersen estimator is equivalent to computing the simple-Petersen estimator for each of the strata and then adding the results to estimate the overall run size.

The standard assumptions are that the samples of marked and unmarked fish captured at the recapture location in each stratum form a simple random sample from the sets of marked and unmarked fish available, with the same sampling probability for both. More formally, it is assumed that:

- the population is closed in each stratum (i.e., no fish enter or leave the population between the release and recapture locations);
- marks are not lost between the point of release and recapture;
- all fish in one stratum (marked and unmarked) have the same probability of being captured at the recapture site (homogeneity on an individual stratum level);
- and whether or not any individual is captured at the recapture site is independent of the capture of all other individuals.

Unfortunately, it is impossible to test these assumptions if batch marks are used in each stratum and so these assumptions usually must be assessed on biological grounds. If individually-numbered tags are used, then it possible to assess some of the assumptions but this would require much more data than normally collected by the TRRP protocol. With only two samples in each stratum, it is impossible to assess the assumption of closure. The assumption of no-mark loss could be examined by double-tagging some fish (Seber and Felton 1981). The assumption of homogeneity could be assessed by comparing the subsequent recapture rates of the daily batches of fish released within each stratum using a variant of the chi-square test in the previous section. The assumption of independence of capture could be assessed by computing the over-dispersion factor (ratio of chi-square test statistic to degrees-of-freedom) from the recaptures of daily batches of fish using a variant of the chi-square test in the previous section.

If these assumptions do hold, the numbers of marked and unmarked fish captured at the recapture site in stratum *i* have independent binomial distributions:

$$m_{ii} : \ Binomial(n_i, p_i)$$
$$u_i \sim Binomial(U_i, p_i)$$

where p_i is the probability that any individual in the i^{th} stratum is captured at the recapture site. The likelihood for the two sets of parameters $\{p_i\}$ and $\{U_i\}$ is then constructed by multiplying the contributions from the marked and unmarked fish for each stratum:

$$L(\mathbf{p},\mathbf{U} \mid \mathbf{n},\mathbf{m},\mathbf{u}) = \prod_{i=s}^{s} \left[\binom{n_i}{m_{ii}} p_i^{m_{ii}} (1-p_i)^{n_i - m_{ii}} \right] \left[\binom{U_i}{u_i} p_i^{u_i} (1-p_i)^{U_i - u_i} \right]$$

Given this model, the simplest method for estimating the total population size is to estimate each U_i separately given n_i, m_{ii}, and u_i and then to sum these values. The intuitive estimator of p_i, and also maximum likelihood estimator (MLE), is the proportion of marked fish recaptured in stratum *i*, $\hat{p}_i = \frac{m_{ii}}{n_i}$. Equating the expected and observed number of unmarked fish captured in stratum *i* and substituting $\hat{p}_i$ then yields the estimator $\hat{U}_i = \frac{n_i u_i}{m_{ii}}$ which is the simple Lincoln-Petersen estimator of abundance computed for each stratum individually. The grand total is found by summing the $\hat{U}_i$. Because each stratum estimator is a Lincoln-Petersen estimator, the same problems that arise with the simple Petersen can arise, particularly when the capture probabilities are small. First, the estimate does not exist when $m_{ii} = 0$ since this entails division by 0. Second, the estimator is biased for all experiments – in fact, $E\left[\hat{U}_i\right] = \infty$ because there is always a non-zero probability that $m_{ii} = 0$. To avoid problems with small numbers of recaptures, the simple estimator above is modified using the Chapman estimator (Seber 1982)

$$\hat{U}_i = \frac{(n_i + 1)(u_i + 1)}{m_{ii} + 1} - 1$$

which essentially adds 1 to the counts of marked, recaptured and unmarked individuals caught in order to avoid division by 0, and then subtracts 1 from the final estimate to account for the change. This estimator can always be computed and is known to be unbiased when $n_i + m_{ii} + u_i > U_i$ a condition which ensures that at least one marked individual was recaptured in each stratum. In the case where $n_i + m_{ii} + u_i < U_i$, Seber (1982, pg. 60) notes that the relative bias is negligible (< .02, 95% of the time) provided that $m_{ii} \geq 7$. The estimate of the total population size is again found by adding the estimates from the individual strata.

The estimated variance of the Chapman estimator for each stratum is (Seber 1982):

$$\widehat{V(\hat{U}_i)} = \frac{(n_i + 1)(m_{ii} + 1)(n_i - m_{ii})u_i}{(m_{ii} + 1)^2 (m_{ii} + 2)}$$

By independence, the variance of the estimator of total abundance is found as:

$$\widehat{V(\hat{U}_{total})} = \sum_{i=1}^{s} \widehat{V(\hat{U}_i)}$$

Steinhorst et al. (2004) examined several variants of the individual stratum estimators, and also derived methods for computing bootstrap and profile intervals for the total run size.

While this strategy does allow heterogeneity to be accounted for, the numbers of individuals marked and/or recaptured in some strata may be very small. This can lead to numerical problems in computing estimates for some strata or produce estimates with very low precision. Additionally, this simple method cannot easily deal with strata for which no recapture data or no unmarked fish counts are available because of logistical or other problems.

To overcome these limitations, various methods have been proposed for pooling strata (e.g. Darroch 1961; Schwarz and Taylor 1998; Bjorkstedt 2000). These methods essentially look at the results of chi-square tests for homogeneity of capture-probability for the strata to be pooled. However, as noted by Steinhorst et al. (2004), it is not clear how many ultimate strata are needed, nor which adjacent strata should be pooled and the small sample sizes in each stratum will lead to tests-of-pooling with small power to detect differences in catchability. The final estimate of precision after pooling has been finalized will not include any contribution from the pooling decisions that were made nor will it reflect the variability in capture rates among the pooled strata. At the moment there is no objective way to proceed. Finally, pooling will reduce the ability to produce estimates of the run-size at the stratum (weekly) level and some assumptions of how the run occurred during the pooled strata will be needed to disaggregate the total run for that stratum.

3.4 Bayesian methods – sharing information

The above method treats the counts in each stratum as completely independent of counts in all other strata. This is too general for studies of migrating salmon and other temporally-stratified mark-recapture data. While fluctuations in the counts from day-to-day will always occur, migrations tend to follow a fairly predictable pattern: few fish pass on the days early in the migration period, the numbers grow fairly steadily to one or two peaks in the middle of the migration and then drop back down at the end of the period. The result is that the abundance of fish in one stratum is strongly associated with the abundance in

the neighboring strata.

Hierarchical Bayesian modeling (Appendix B) presents an alternative to pooling that draws on the similarities between strata without assuming exact equality of the capture probabilities. In the simplest, non-hierarchical Bayesian model of the TSPDE, each p_i and U_i would be assigned independent prior distributions, and the posterior distribution for each would depend only on the prior and the data collected for each stratum *i*:

$$m_{ii} : Binomial(n_i, p_i)$$
$$u_i \sim Binomial(U_i, p_i)$$
$$p_i : Beta(\alpha_i, \beta_i) \quad \text{Prior on capture probability}$$
$$U_i : Poisson(\lambda_i) \quad \text{Prior on abundance}$$

Note that a separate prior distribution is specified for each stratum.

In cases where the number of recaptures is small and good prior information is available to guide estimation of the recapture probabilities in the absence of such data, this simple Bayesian model may overcome some of the difficulties in a fully stratified analysis. However, this Bayesian model presents NO advantages over that simple likelihood approach above unless good prior information is available on some of the parameters. Mantyniemi and Romakkaniemi (2002) developed a hierarchical model for non-diagonal data which can also be applied for diagonal experiments. In the hierarchical model, the parameters from different days are linked by assuming that **p** and **U** form random draws from some hyperpriors. Mantyniemi and Romakkaniemi (2002) suggest a normal model for the logit transformed capture probabilities and a multinomial prior for the stratum population sizes where the parameters of these distributions, including the total population size, are assigned hyperpriors with fixed parameter values. Their model for the diagonal capture case is:

$$m_{ii} : Binomial(n_i, p_i)$$
$$u_i \sim Binomial(U_i, p_i)$$
$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) : Normal(\eta_p, \sigma_p^2) \quad \text{Prior on capture probability.}$$
$$U_i : Multinomial(U, \{a_i; i = 1,\text{K } s\}) \quad \text{Prior on the run available in each week.}$$
$$a_i : Dirichlet(\{b_i; i = 1,\text{K } s\}) \quad \text{Prior on the prop. of the total run in each week.}$$
$$U : Uniform[1, \text{LargeValue}] \quad \text{Prior on the total run size.}$$

Appropriate priors on the hyperparamaters η_p, σ_p^2 are also imposed.

The advantage of the hierarchical model is that inference for the parameters in one stratum will depend on the data from all strata. For example, if the data suggests that all of the strata have p_i's close to .025, then this information is used to improve estimates for strata with poor data (or no data at all). In this way, data are shared among the strata, but the parameters are still allowed to vary among strata. The disadvantage of the hierarchical approach is that it makes no adjustment for the ordering of the data – the same amount of information is shared between strata 1 and 2 as strata 1 and 10 or strata 1 and 100.

While the hierarchical model allows for data to be shared among strata, it does not consider the natural ordering of the data when strata are based on time. By assuming that each p_i and U_i form independent

draws from their respective hyper-priors, the amount of information shared between (p_i, U_i) and (p_j, U_j) is the same regardless of whether $|i-j|$ is 1, 5 or *s*−1. In the case of temporal stratification, and particularly the application to salmon migrations, it seems intuitive that the number of fish passing the recapture trap will be more similar for strata that are close together and less similar for strata that are further apart.

This can be achieved through several different extensions of the hierarchical Bayesian model. For example, by assuming that the correlation between U_i and U_j follows some decreasing function of $|i-j|$ or by defining an autoregressive model on the parameters as in time series analysis.

Bonner (2008) chose to model U_i explicitly as a smooth curve using the Bayesian penalized spline (P-spline) method of Lang and Brezger (2004). As discussed in Appendix A, two factors control the smoothness of a spline: the number and locations of the knots and the variation in the coefficients of the basis-function expansion. The classical P-spline method of Eilers and Marx (1996) approaches this dichotomy by fixing a large number of knot points and then penalizing the first or second order differences of the coefficients. In the original implementation, the spline curve is fit by minimizing a target function which adds the sum-of-squared residuals and a penalty term formed as the product of a smoothing parameter and the sum of the differences of the spline coefficients. Increasing the smoothing parameter places more weight on the penalty term and results in a smoother curve. Decreasing the smoothing parameter places more weight on the sum-of-squared residuals and produces a fit that comes closer to interpolating the data.

Lang and Brezger (2004) develop a Bayesian formulation of the P-spline method which replaces the penalty term by a particular prior distribution for the coefficients. Let

$$B_1(x), \mathsf{K}\ , B_{K+q}(x)$$

denote the *B-spline* basis functions for a spline of order *q* with *K* knots and $b_1, \mathsf{K}\ , b_{K+q}$ the corresponding coefficients. To ensure that the population size in each stratum remains positive, the $\log(U_i)$ are modeled and the fitted spline has the form:

$$\log(U_i) = \sum_{k=1}^{K+q} b_k B_k(j)$$

where it is assumed that the time strata are equally sized and spaced through the run. Knots are chosen approximately every fourth stratum following the recommendation of Lang and Brezger (2004).

To enforce smoothness on the spline, the prior distribution for the parameters $b_1, \mathsf{K}\ , b_{K+q}$ models the differences in these coefficients as a second order random walk as suggested by Lang and Brezger (2004), i.e.

$$b_{k+1} = b_k + (b_k - b_{k-1}) + \varepsilon_k^{spline}$$

where ε_k^{spline} is assumed to have a $Normal(0, \sigma_U^2)$ distribution, The initial coefficients, b_1 and b_2 are assigned improper flat priors.

This model implies that the successive coefficient is found taking the previous coefficient plus the change seen from the yet earlier coefficient with the random error term allowing for some flexibility. If the variance of the ε_k^{spline} (σ_U^2) is small, then the spline is very smooth as the coefficients must follow a strict

progression. If the variance of the ε_k^{spline} is large, the spline is less smooth. The variance of ε_k^{spline} is determined by the data, i.e. how much the observed *U*'s vary from the smooth spline fit. In the Bayesian P-spline approach, the variance parameter (σ_U^2) plays the same role as (though opposite in direction to) the smoothing parameter in the classical method. Rather than fixing the value of σ_U^2, this parameter is assigned a prior distribution which incorporates uncertainty in the amount of smoothing required. Following Lang and Brezger (2004), an inverse gamma prior is used:

$$1/\sigma_U^2 : \ Gamma(\alpha, \beta)$$

with the parameters α and β chosen so that E(σ_U^2) is small (which favors a smooth fit), but V(σ_U^2) is large, which allows a less smooth curve if the data requires. Lang and Brezger (2004) suggest α = 1 with β = .05, .005, or .0005. By default, the computer program developed for this project uses α = 1 and $\beta = .05$, but these can be easily changed in the computer program developed for this report.

Although the spline model may reflect the trend in the daily population size, similar to a running mean, it is unlikely that U_j will exactly follow a smooth curve. If the deviations from a smooth curve are small then it seems reasonable that forcing U_j to be smooth will not have a large impact on the estimation of U_{total}. However, if there are large deviations from the smooth curve then forcing the U_j to be smooth may severely bias the estimate of the total population size. To allow for roughness in the U_j over-and-above the spline fit, the spline model is extended with an additional "error" (extra variation) term:

$$\log(U_i) = \sum_{k=1}^{K+q} b_k B_k(j) + \varepsilon_j^{\log U}$$

where the $\varepsilon_i^{\log U}$ are assumed independent and normally distributed with mean 0 and variance $\sigma_{\log U}^2$. If this variance is small, then the individual U_j will lie very close to the spline; if the variance is large, the individual U_j will vary considerably above and below the spline. As for the variation in the spline coefficients, the computer program uses an inverse gamma prior distribution

$$1/\sigma_{\log U}^2 : \ Gamma(1, .05)$$

for the variance parameter $\sigma_{\log U}^2$

There are no closed form solutions for the posterior distribution of the parameters, and samples from the posterior distributions of these models were generated through MCMC sampling using the WinBugs/OpenBUGS software package (Thomas et al. 2006; Lunn et al. 2000). Computer code for the Bayesian P-spline models are available in an R-package called BTSPAS (Bayesian Time Stratified Petersen Analysis System) that can be downloaded from the CRAN and R-forge libraries, detailed instructions are provided in Appendix C.

Bonner (2008, Section 2.2.4) conducted an extensive simulation study to compare the Bayesian P-spline method with the stratified-Petersen estimator. Generally, the Bayesian P-spline method had negligible bias and its precision was at least as good as the stratified-Petersen estimator. When "perfect" data were available (i.e., many marked fish were released and recaptured in each stratum) the results from the

Bayesian P-spline and the stratified-Petersen were very similar. When few marked fish were released or recaptured in each stratum, the performance of the Bayesian P-spline model depended on the amount of variation between the capture probabilities and the pattern of abundance over time. In the worst case, with large variations between the capture probabilities and abundances that followed no regular patterns, the two models continued to perform similarly. However, when the variation between the capture probabilities were smaller and the abundances followed close to a smooth curve the Bayesian P-spline produced much more precise estimates of the total population size.

The above development assumes that a single spline curve can be fit across all strata, but problems arise in fitting this model to the data from the TRRP. In some cases, there are obvious breaks in the pattern of abundance over time that need to be accounted for in the model. For example, the number of Chinook salmon passing the trapping sites jumps dramatically after the hatchery releases which occur twice each season. Attempting to fit a smooth curve across all strata ignores these jumps and so it is necessary to allow for breaks in the fitted spline. Bonner (2008, Section 2.4.2) shows how to generalize the above spline model by allowing for 2 or more spline-segments whose parameters are estimated separately, but whose smoothness parameters are assumed to be equal across the segments.

An alternative way to account for the mixture of wild and hatchery fish is to estimate a single spline curve for each stock, but assuming that all fish for all stocks have the same catchability. In the Trinity River experiments, hatchery-raised steelhead are all marked with adipose fin clips and so wild and hatchery fish can readily be distinguished. Only a portion of the Chinook salmon hatchery stocks are clipped (about 25%), which means that wild and unmarked hatchery fish cannot be distinguished but it still is possible to obtain estimates of separate runs if the proportion of hatchery fish clipped is known, as outlined below.

An obvious extension of the Bayesian P-spline method is to simultaneously smooth both the capture-probabilities and the run size.. Unfortunately, such doubly smoothed models do not perform well in practice. In reality, neither the capture rates, nor the run sizes actually follow any smooth curve. Fitting a spline to the run sizes but not the capture rates provides enough flexibility in the capture rates to accommodate the variations in run size around the spline. Similarly, a model where the spline is fit to the capture rates, but not to the run sizes would provide enough flexibility in the run size to accommodate the variations in the capture probabilities around the spline. However, Bonner (2008) found if splines were fit to both sets of parameters, then the model is rarely flexible enough to accommodate the variations from smoothness. Model selection criteria rarely selected models where both parameters were smoothed.

3.5 Assessing goodness-of-fit.

The primary tools to assess the goodness-of-fit of the model to the data are visual inspections of the fit and predictive posterior values (also known as Bayesian p-values) (see Gellman, et al. 2004, Section 6.2 for an overview). Various plots derived from the final model can be generated to quickly check for discrepancies between the model and the data. For example, the estimates of abundance in each stratum from the Bayesian P-spline model should be compared with the estimates derived from a simple stratified-Petersen model (e.g. Figure 3.3) to see if the fitted model generally follows the pattern observed in the data. Similarly, the simple estimates of catchability vs. the modeled values should also be examined (e.g. Figure 3.5) to check for discrepancies. However, such visual assessments can be difficult to interpret or misleading when the data are sparse.

A plot of the autocorrelation function for the population total (example presented in Appendix C) will show if successive posterior values are highly correlated – while high autocorrelation is not technically a problem, it may indicate that the MCMC algorithm is not traversing the posterior parameter space well.

Trace plots of the successive values generated from the posterior distribution should be examined to see if the chains are adequately sampling from the posterior distribution. As there are many parameters in these models, it is often not feasible to examine the individual plots. A summary measure of mixing is the Brooks-Rubin-Gelman statistic (BRG, Brooks and Gelman (1998)). The intuitive idea of this statistic is to fit the model using multiple dispersed starting points. If the model fitting has converged, the results from the individual chains should all give similar results. The BRG statistic computes the ratio of the among-chain variation in the posterior values to the within-chain variation. The BRG statistic should be close to 1 (in practice, values smaller than 1.3 are acceptable). The BRG *R*-statistic is automatically computed for every estimate by the computer program and samples of the output are found in Appendix C.

Bayesian p-values provide a more objective method for assessing the fit of a model. First, some measure of discrepancy must be defined. This can be quite general and could be tailored to investigate specific model failings. For example, a generic method often used in classical methods is a chi-square type statistics comparing observed and expected counts. This is computed for a sample of parameter values from the posterior distribution. Second, for each set of parameters from the posterior distribution, a set of simulated data is generated. The discrepancy measure is computed for this simulated dataset. Third, the distribution of the discrepancy measure based on the real data and the simulated data are compared. If the model fits well, then the discrepancy measures for the real data should have a similar distribution to the discrepancy measure from the simulated data. The Bayesian p-value is computed as the proportion of times the discrepancy measure from the real data exceeds the corresponding discrepancy measure computed on the simulated data using the same value of the posterior. A Bayesian p-value near 0.5 indicates no evidence of lack-of-fit. Bayesian p-values near 0 or 1 indicate potential problems.

Note that a Bayesian p-value CANNOT be interpreted in the same way as goodness-of-fit statistics from likelihood methods. In particular, the distribution of Bayesian p-values if the model is adequate does NOT have a uniform [0, 1] distribution so standard rules such a rejecting the fit if the p-value is <0.05 does not apply. As noted by Gellman et al. (2004, p. 175), extreme values of the Bayesian p-value are indicative of a lack-of-fit, but moderate deviations (e.g. a p-value of 0.20) are difficult to interpret.

For this project, two different discrepancy measures were used for each of two aspects of the model. The first discrepancy measure is the Freeman-Tukey (Freeman and Tukey 1950) statistic that compares observed and expected counts:

$$D_1(\mathbf{x};\theta) = \sum\left(\sqrt{x_j} - \sqrt{e_j}\right)^2$$

where $\sqrt{x_j}$ and $\sqrt{e_j}$ are the observed and expected counts respectively. As indicated by Brooks et al. (2000), this discrepancy measure is less sensitive to very small expected counts and removes the need for pooling cells often found in the classical Pearson chi-square goodness-of-fit.

The second discrepancy measure is the deviance *(-2 x log-likelihood*) for a Binomial distribution

$$D_2(\mathbf{x};\theta) = -2\sum \log\left[\binom{n_j}{x_j} p_j^{x_j}(1-p_j)^{n_j - x_j}\right]$$

where n_j and x_j are the index and observed count for the jth binomial distribution.

Each of the discrepancy measures was applied to the $\{n_i, m_{ii}\}$ (the releases and marks recovered) and the $\{U_i, u_i\}$ (unmarked populations size and unmarked fish captured) sets. For example,

$$D_1(\{n_i, m_{ii}\}; \{p\}) = \sum \left(\sqrt{m_{jj}} - \sqrt{n_j p_j} \right)^2$$

and

$$D_2(\{n_i, m_{ii}\}; \{p\}) = -2 \sum \log \left[\binom{n_j}{m_{jj}} p_j^{m_{jj}} (1 - p_j)^{n_j - m_{jj}} \right]$$

with similar measures defined for the $\{U_i, u_i\}$ set.

An overall discrepancy measure for each type is also found by summing the two discrepancy measures from the two sets.

A Bayesian goodness-of-fit plot is constructed by plotting the discrepancy measures from the real and simulated data against each other and overlaying the line Y=X. About ½ of the points should lie on either side of the line (e.g. Figure 3.6).

Note that the Bayesian p-values measure the fit of the entire model including the prior distribution. A poorly fitting model may be a consequence of an ill-chosen prior. It is important then to try various model fits with different priors to assess if the final results differ materially.

3.6 Dealing with problems in the data

The Bayesian P-spline model can easily deal with many problems encountered with real data.

No marked fish released in a stratum

It may occur that no marked fish (i.e. n_i =0 for some *i*) are released in a particular stratum because no fish were available or because of logistical constraints. This would imply that a simple Petersen estimator for this stratum cannot be computed because no estimate of the capture probability is available. However, the Bayesian method will impute a range of capture-probabilities based on the capture probabilities in other strata, the shape of the spline curve, and the observed number of unmarked fish captured. The final estimate of U_{total} will (automatically) incorporate the uncertainty for this imputed capture probability. The computer program will automatically detect such circumstances.

No sampling during a stratum

If no data could be collected for a particular stratum (i.e. all of $n_i = 0, m_{ii} = 0, u_i = 0$), the spline will "impute" a value for the run size and capture probability in that stratum given the shape of the spline and the variability in individual run sizes about the spline; and will "impute" a value for the capture probability given the range of capture probabilities in the other strata. The final estimate of U_{total} will (automatically) incorporate the uncertainty for this imputed values. The computer program will automatically detect such circumstances (the sampling fraction will be zero).

While it is possible to interpolate for several strata in a row, there is of course, no information on the shape of the underlying spline during these missed strata, and so the results should be interpreted with care.

Less than 100% sampling during a stratum

The time strata in the TRRP are generally one week long. In most strata, sampling took place during all 7 days. However, in some strata, sampling may take place in only 3 of the 7 days. This causes no theoretical problem for estimation of the capture probability as the capture probability is assumed to be equal over the entire week so estimates based on 3 days may have poor precision but remain unbiased. However, the number of unmarked fish needs to be adjusted for the partial sampling during the stratum. At the moment, a simple moment correction is automatically applied, i.e. $u_i^* = u_i \frac{7}{days\ sampled}$. This should lead to sensible estimates of the total population size for this stratum, but no adjustment as been made for the additional uncertainty in the total population size induced by this expansion. In most cases, the number of strata with less than 100% sampling is small and so the underestimate of uncertainty should be small.

Varying effort during a stratum.

In some cases, the number of screw traps varies over the course of a week. For example, two traps may be operating on Monday, and then only trap is operating on Tuesday, etc. Unfortunately, this type of problem CANNOT be adequately dealt with by the spline (or any other method) that uses batch marks. The problem is that differing effort during a stratum (week) results in heterogeneity of catchability during the week, e.g. the catchability on days when two traps are operating is likely larger than the catchability on days when only one trap is operating. If a batch mark is used, the data are pooled over the stratum (week) and it is impossible to separate out catches according to how many traps are operating.

As for the pooled-Petersen estimator, this will likely result in estimates with low bias, but the precision of the estimates for these strata will be overestimated, i.e. the standard deviations of the estimated run sizes for these strata will be understated. While there is no way to assess the extent of the problem (other than via simulations), it is hoped that the stratification into weekly strata will resolve most of the underreporting of the precision by the pooled-Petersen estimator and that any remaining understatement is not material.

3.7 Using Covariates

The Bayesian P-spline model assumes that the variation in catchability is a random process over weeks. However, in some cases, additional covariates may be present to try to "explain" some of the variation in catchability. For example, Figure 5 of Green et al. (2007) appears to show a relationship between catchability and log(flow).

The above model is easily extended to allow for covariates. Let X_{ij} represent the value of the jth covariate in week *i*. Then catchability is modeled as:

$$\text{logit}(p_i) = \beta_0 + \sum_{j=1}^{J} \beta_j^p X_{ji} + \varepsilon_i^p$$

where ε_i^p has a $N(0, \sigma_p^2)$ distribution. Vague prior (e.g. N(0, 100)) are placed on the coefficient for the covariates.

The importance of a covariate in explaining variation of *p* can be assessed in two ways. First, the posterior distribution for the coefficient of the covariate can be examined to see if the 95% confidence interval excludes the value of 0 - indicating good evidence that the covariate appears to be useful. Second, models with and without the covariate can be compared using the Deviance Information Criteria (DIC, Gelman et

al. 2004, Section 6.6). Differences in DIC of more than about 4 or 5 indicate good evidence that the covariate is useful.

3.8 Separating Hatchery and Wild stocks

3.8.1 Chinook salmon

There are four components to the outgoing unmarked Chinook salmon population:

Wild Young-of-Year (YOY) fish which have no external modifications (i.e., not adipose-fin clipped).

Hatchery YOY. Standard practice is to clip the adipose-fin on 25% of the hatchery fish prior to release. These are usually released starting in about Julian week 11.

Wild Age 1+ fish which have no external marking. The number of Wild Age 1+ fish is negligible in the studies examined in this report and will be ignored.

Hatchery age 1+ fish. Standard practice is to clip the adipose fin on 25% of the hatchery fish. These are usually released starting in about Julian week 40.

Estimation of the number of hatchery aged 1+ fish that pass the screw-trap is straightforward. These occur sufficiently late in the season so that the number of wild YOY fish migrating concurrently with the age 1+ fish is negligible and by definition, there are no longer any Hatchery YOY fish left. As will be seen in the example, it is quite evident from the summary data when the hatchery age 1+ fish start to pass the traps and it will be assumed that from this point onwards the population consists only of hatchery age 1+ fish. Consequently, estimates of the outgoing fish numbers in these weeks to the end of the experiment from the models of the previous section will be extracted and used to estimate the number of hatchery age 1+ fish. For the remainder of this section, estimation is directed at the two components of the YOY fish.

Prior to the release of Hatchery age 1+ fish, outgoing fish that pass the screw-trap are a mixture of wild and hatchery YOY fish. Separation of the components is more complicated because only a portion of the hatchery YOY fish are adipose-fin clipped. The data for this part of the study consists of:

n_i - number of marked fish released in stratum *i* (as before)

m_{ii} - number of marked fish recaptured in stratum *i* (as before)

u_i^A - number of unmarked YOY fish captured in stratum *i* that have the adipose-fin clipped;

u_i^N - number of unmarked YOY fish captured in stratum *i* that do not have the adipose-fin clipped.

Note that $u_i = u_i^A + u_i^N$ which was used in the previous sections to estimate the total number of YOY fish that passed the screw-trap in stratum *i*.

As before, the recaptures of marked fish is modeled as

$$m_{ii} : Binomial(n_i, p_i)$$

It is implicitly assumed that the capture probabilities estimated by the recapture of marked fish is applicable to all components of the population as the marked fish are not separated by component.

The total number of unmarked YOY fish passing the trap in stratum *i* is broken into the wild and hatchery components:

$$U_i = U_i^W + U_i^H$$

The number of unclipped YOY fish captured in stratum *i* is a mixture of hatchery and wild fish. Prior to the known release of hatchery YOY fish, the unclipped fish are all wild fish and are modeled as:

$$u_i^N : Binomial(U_i^W, p_i)$$ for strata prior to hatchery YOY release

There are no fish captured with ad-clips prior to the hatchery YOY release.

Because only a portion *c* (currently 25%) of hatchery fish is adipose-fin clipped, the number of clipped YOY fish captured is modeled as:

$$u_i^A \ : \ Binomial(U_i^H, cp_i)$$

The number of unclipped YOY fish (which is a mixture of wild and hatchery fish) is modeled in a three-step process. First, a latent variable for the number of clipped YOY fish passing the trap in stratum *i* is modeled:

$$U_i^C \ : \ Binomial(U_i^H, c)$$

Second, then the number of unclipped YOY fish passing the screw-trap in stratum *i* is defined as:

$$U_i^{NC} = U_i^W + U_i^H - U_i^C$$

Finally, the number of unclipped YOY fish captured is modeled as:

$$u_i^N \ : \ Binomial(U_i^{NC}, p_i)$$

[These are approximations to the actual hypergeometric distributions that should be used, but given the large number of YOY fish passing the screw-traps, the binomial distributions used should be very close approximations.]

Splines are used to model the U_i^W and U_i^H YOY fish in much the same way as in the earlier sections but the exact methodology is not detailed here. Similar priors are used for the spline coefficients and parameters of the capture probabilities as before.

Assessment of goodness-of-fit is similar to the previous section and again not detailed here. It should be noted that the deviance cannot be (easily) computed for these (mixture) models for the reasons summarized by Gelman et al. (2004).

Similar strategies for dealing with problems in the data (e.g. missing sampling on some strata; partial sampling in some strata; etc.) are used as outlined earlier.

3.8.2 Steelhead

There are three components to the outgoing steelhead population:

- Wild Young-of-Year (YOY) fish which have no external modifications (i.e., not adipose-fin clipped).
- Wild Age 1+ fish which have no external modifications. This component consists of age 1 and age 2+ fish, but the number of age 2+ fish is negligible and is combined into the age 1+ group.
- Hatchery age 1+ fish. Standard practice is to clip the adipose fin on 100% of the hatchery steelhead fish. These are usually released starting in about Julian week 11. This component includes hatchery aged 2+ fish that residualized from previous years of releases, but again the numbers of such fish are negligible.

There are no hatchery-raised YOY steelhead fish released.

Because 100% of hatchery-raised steelhead are adipose-fin clipped, modeling is much easier in this case compared to the Chinook salmon separation because each fish can be unambiguously assigned to one of the three components. The data are now:

n_i - number of marked fish released in stratum *i* (as before)

m_{ii} - number of marked fish recaptured in stratum *i* (as before)

$u_i^{W.YoY}$ - number of wild YOY unmarked fish captured in stratum *i* ;

$u_i^{W.1+}$ - number of wild age 1+ unmarked fish captured in stratum *i*;

$u_i^{H.1+}$ - number of hatchery 1+ unmarked fish (but adipose-fin clipped) captured in stratum *i*.

Note that $u_i = u_i^{W.YoY} + u_i^{W.1+} + u_i^{H.1+}$ which was used in the previous sections to estimate the total number of fish that passed the screw-trap in stratum *i*.

As before, the recaptures of marked fish is modeled as

$$m_{ii} : Binomial(n_i, p_i)$$

It is implicitly assumed that the capture probabilities estimated by the recapture of marked fish is applicable to all components of the population as the marked fish are not separated by component.

The total number of unmarked fish passing the trap in stratum *i* is broken into the wild and hatchery components:

$$U_i = U_i^{W.YoY} + U_i^{W.1+} + U_i^{H.1+}$$

The number captured from each component is modeled as:

$$u_i^{W.YoY} : Binomial(U_i^{W.YoY}, p_i)$$
$$u_i^{W.1+} : Binomial(U_i^{W.1+}, p_i)$$
$$u_i^{H.1+} : Binomial(U_i^{H.1+}, p_i)$$

Splines are used to model the $U_i^{W.YoY}$, $U_i^{W.1+}$ and $U_i^{H.1+}$ in much the same way as in the earlier sections and is not detailed here. Similar priors are used for the spline coefficients and parameters of the capture probabilities as before.

Assessment of goodness-of-fit is similar to the previous section and again not detailed here. Because of the unambiguous separation of fish into the three components, the deviance is now readily computed and can be used for model selection and goodness-of-fit testing.

Similar strategies for dealing with problems in the data (e.g. missing sampling in some strata) are used as outlined earlier.

3.9 Example – Junction City Chinook salmon 2003

In this experiment, the migration of Chinook salmon smolts was monitored along the Trinity River in Northern California in 2003. The experiment was performed by the Hoopa Valley Tribal Fisheries Department and details are reported in Green et al. (2007). Data from the experiment are reproduced in Table 3.1 below.[2]

Note because of stochastic variability in MCMC methods, the results reported in this report may differ slightly from those on the web-site.

[2] Data below differs slightly from Appendix A-4 of Green et al. (2007). The data in this report was extracted from the common database.

Table 3.1. Summary of the 2003 Junction City Chinook salmon data.

Julian week	Days operating	Marked n_i	Recoveries in week			Total recovered	Unmarked u_i
			i	*i+1*	*i+2*		
9	3	0	0	0	0	0	4,135
10	8[1]	1,465	32	19	0	51	10,452
11	6	1,106	121	0	0	121	2,199
12	7	229	25	0	0	25	655
13	7	20	0	0	0	0	308
14	7	177	17	0	0	17	719
15	7	702	74	0	0	74	973
16	7	633	94	0	0	94	972
17	7	1,370	62	0	0	62	2,386
18	7	283	7	3	0	10	469
19	7	647	32	0	0	32	897
20	7	276	11	0	0	11	426
21	7	277	12	1	0	13	407
22	7	333	14	1	0	15	526
23	7	3,981	242	0	0	242	39,969
24	7	3,988	55	0	0	55	17,580
25	7	2,889	114	1	0	115	7,928
26	7	3,119	197	0	0	198	6,918
27	7	2,478	80	0	0	80	3,578
28	7	1,292	71	0	0	71	1,713
29	6	2,326	153	0	0	153	4,212
30	7	2,528	156	0	0	156	5,037
31	7	2,338	275	0	0	275	3,315
32	7	1,012	101	0	0	101	1,300
33	7	729	65	1	0	66	989
34	7	333	44	0	0	44	444
35	7	269	33	0	0	33	339
36	7	77	7	0	0	7	107
37	7	62	9	0	0	9	79
38	7	26	3	0	0	3	41
39	7	20	1	0	0	1	23
40	7	4,757	188	0	0	188	35,118
41	7	2,876	8	0	0	8	34,534
42	7	3,989	81	0	0	81	14,960
43	7	1,755	27	0	0	27	3,643
44	7	1,527	30	0	0	30	1,811
45	7	485	14	0	0	14	679
46	5	115	4	0	0	4	154
Total		**50,489**	**2459**			**2,486**	**209,995**

[1] The database records 8 days operating in this week and 6 days in the next. The same batch mark was used for 8 days rather than 7 days. This is adjusted for in the usual fashion.

In this year, smolts were trapped at the site for 38 consecutive weeks in 2003 starting on 27 February 2003. Along with the wild smolts migrating downstream, smolts were released from a hatchery above Junction City in Julian weeks 23 and 39. This example will analyze the total number of Chinook salmon pooling both hatchery and wild fish, and young of year and age 1+ fish.

The challenge in analyzing this data set is not the sparsity of the data. Over 50,000 salmon were marked over the 38 week period and almost 2500 of these were recaptured for an average recapture rate of about 5%. Moreover, the data was almost diagonal; fewer than 30 smolts were recovered outside of the week in which they were marked and these were ignored during the analysis, i.e. the experiment was treated as TSPDE.

The pooled-Petersen estimate is 4.3 (standard error .08) million fish for a relative standard error of 2%. However, the weekly recapture rates vary considerably between 0% and 15%. The chi-square test for no difference in capture rates strongly rejects ($p<.0001$) the hypothesis of equal capture rates over time which indicates that the key assumption of homogeneity is not tenable. The pooled-Petersen estimator may remain nearly unbiased, but the estimated precision is severely negatively biased.

Individual weekly estimates of population size can be formed using the weekly Chapman estimates of the run size. This gave an estimate of 16 (standard error 3.7) million smolts migrating over the entire period. However, no estimate can be obtained for Julian week 9 (no marks released which implies no estimate of catchability is available), and the estimate is problematic for Julian week 41. In Julian week 41, the capture probability was apparently much lower than in the other weeks. According to the data, just under 3000 salmon were marked in Julian week 41 and only 8 (0.3%) recovered so that the proportion of marked fish recovered in this week was 6 times lower than the lowest recapture rate in the remaining 37 weeks. The Chapman estimate of population size in this week, 11 (standard error 3.6) million fish, was more than 2/3 of the estimate of the total population size and it's uncertainly overwhelms the uncertainty from the other strata. If the data from Julian weeks 9 and 41 is discarded, the stratified estimate is reduced to 5.0 (standard error .21) million fish, now much more in line with the pooled-Petersen estimator. It should be noted as well, that by Julian week 41, the run is composed mainly of the second release of hatchery fish of which only about 1.5 million were released. The estimate in Julian week 41 is clearly unrealistic.

It is clear that the numbers of smolts marked and recaptured in Julian week 41 are not reliable. Either one of these numbers has been recorded incorrectly (e.g., the number of recaptures should be 80 not 8), or something happened during the experiment that greatly reduced the recapture rate (e.g., a large proportion of the marked fish were killed as they were transported to the release site). Additionally, there are several weeks where either no or only a small number of fish were released and recaptured giving very poor estimates of the capture rate (and the population size) for those weeks.

For this analysis, the data were modified to ignore the number of recaptures in Julian week 41 (specifically, the value of $m_{41,41}$ was set to a missing value), and to treat the number of releases and recaptures in Julian week 9 also as missing values, (i.e. n_9 and $m_{9,9}$ were set to missing values).

The run size could not be expected to be smooth over all 38 weeks because of the two introductions of hatchery fish and so the Bayesian P-spline model allowed for two jumps in the run size. To accommodate the introduction of hatchery fish before Julian weeks 23 and 40 the spline was broken into 3 segments: the first modeled the run over Julian weeks 9 to 23 with 3 knots (automatically chosen at weeks 12, 17, and 21), the second modeled Julian weeks 23 to 39 also with 3 knots (automatically chosen at Julian weeks 27, 31, 35), and the third models Julian weeks 40 to 46 with only 1 knot (automatically chosen at Julian

week 43). Prior distributions for the coefficients of each segment of the spline were defined separately, but the hyper-parameters were forced to be the same in order to achieve a constant degree of smoothness.

The *R*-wrapper[3] (Appendix C) for the analysis of this dataset and detailed results are available at the web site **http://www.stat.sfu.ca/~cschwarz/Consulting/Trinity/Phase2/**. A total of 200,000 iterations of the MCMC sampler were run with the first 100,000 discarded as burn-in[4], and a thinning[5] was used to reduce autocorrelation in the sampled values of the total population size. Approximately 6000 samples from the posterior distributions were retained. The run took approximately 30 minutes to complete running on Window's XP under Parallels (V.4) on a 2.66 GHz Intel-chip MacPro.

The total population size was estimated to be 5.3 (SD 0.18; 95% c.i. ranging from 5.0 to 5.7) million fish.

This estimate is similar to the Pooled-Petersen estimate (pooling over heterogeneous catchability usually leads to nearly unbiased estimates), but the reported precision is about twice as wide. The precision of the spline estimate accounts for the missing data in Julian weeks 9 and 41. The estimate and reported precision is also similar to the estimate from the stratified-Petersen after dropping Julian weeks 9 and 41. These results are not dramatically different when compared to the results from the stratified-Petersen because the data are fairly rich in most of the strata and there is little gain in "sharing" information from other strata to estimate the catchabilities. If the mark-recapture data had been sparser, the spline method is expected to produce estimates with better precision compared to the stratified-Petersen estimator.

A plot of the fitted spline curve and the estimate individual runs (allowing for error) is found in Figure 3.3 (below).

[3] R wrapper: This refers to the computer program that has been provided, which is written in the 'R' statistical software package. This program calls a second software package 'WinBugs/OpenBugs 'wrapped' within the R code (see Appendix C for details).

[4] Burn-in: This refers to the early part of the Bayesian algorithm. It takes the algorithm some time to move from the initial starting values to converge on a consistent posterior distribution. These early iterations are discarded so they don't affect the reported posterior distribution (See Appendix B for details)..

[5] Thinning is used to remove autocorrelation from the realized posterior distributions (See Appendix B for details).

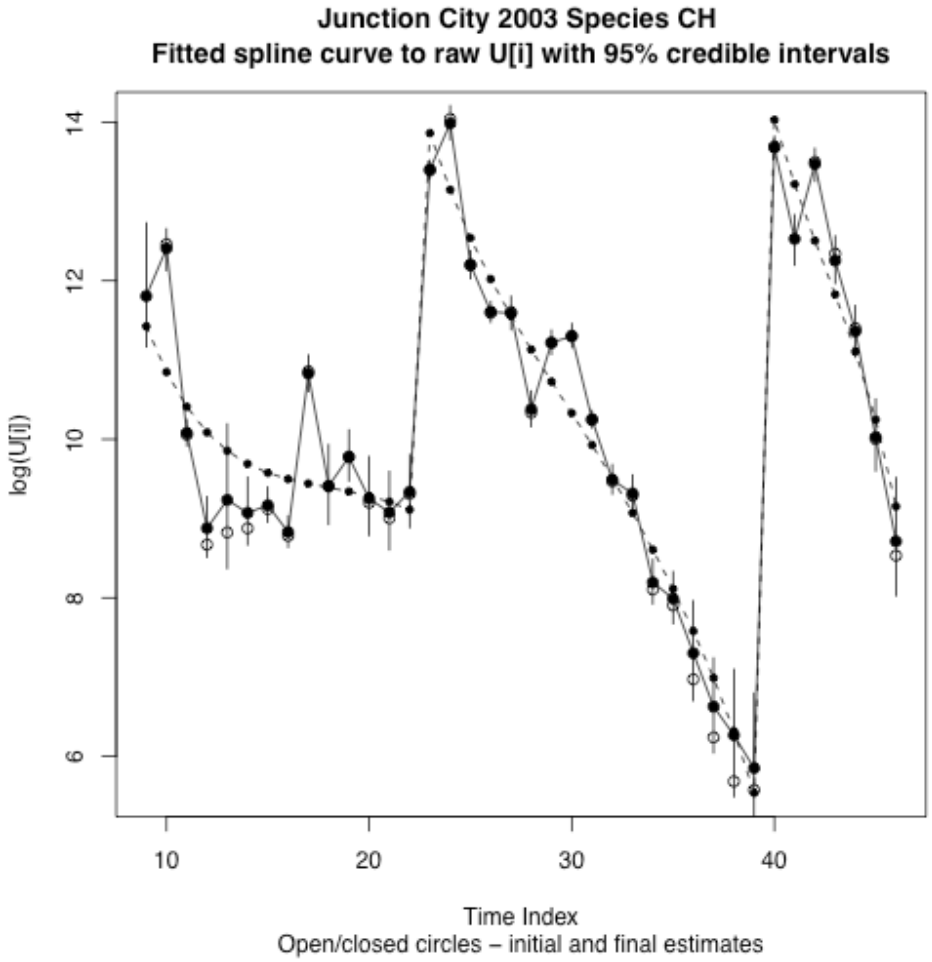


Figure 3.3. Estimates of abundance and the underlying spline curve used to impute abundances for weeks with problem data.

The fitted spline curve follows the trends in abundance fairly well. Notice how the uncertainty in the run size in Julian weeks 9 and 41 (when no data on the recapture rate was used) is larger than for the other weeks but the spline interpolated well. Of course, it is implicitly assumed that the unrealistically high estimate from the simple stratified-Petersen in Julian week 41 is not valid and that the spline gives a more realistic value for this week. The fitted curve allows for variation above and below the spline curve because again the mark-recapture data are fairly rich for each week.

The three separate splines show the break in the run sizes after Julian weeks 23 and 40 to account for the pulse of hatchery fish.

The posterior distribution of the total run size is presented in Figure 3.4 (below).

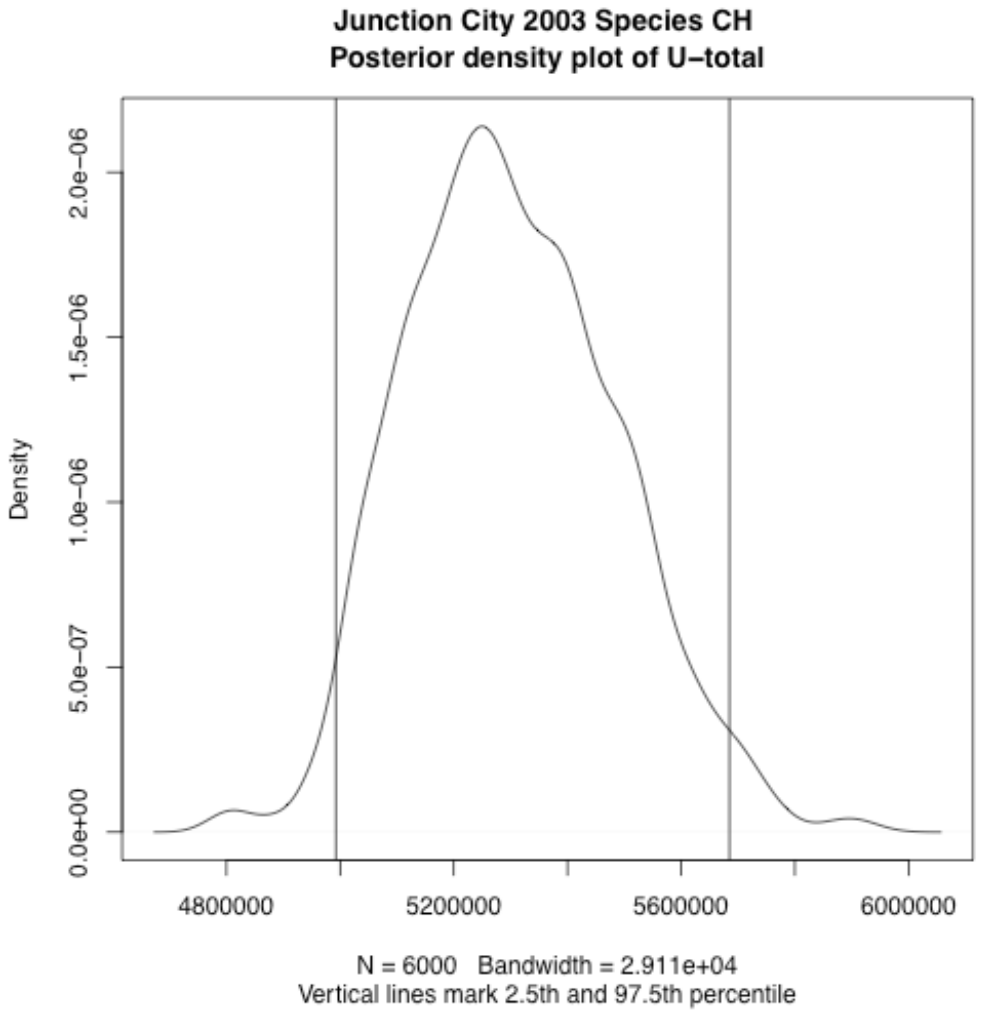


Figure 3.4. Posterior Distribution for the estimated number of unmarked fish.

[The "bumps" in the posterior are artifacts of retaining only 6000 sampled points from the posterior.] The posterior distribution for the total number of fish is right skewed to allow for the weeks with uncertain recapture rates. The autocorrelation function (acf) plot for the total run size showed no evidence of autocorrelation among successive posterior values after thinning (not shown, but see Appendix C).

Posterior summaries for the stratum specific capture probabilities (on the logit scale[6]) are shown in Figure 3.5 below.

[6] $\theta = \text{logit}(p) = \log\left(\frac{p}{1-p}\right)$. $p = \frac{1}{1+\exp(-\theta)}$. A p=.5 corresponds to a logit of 0. A logit of -3 corresponds to $p = \frac{1}{1+\exp(-(-3))} = .047$

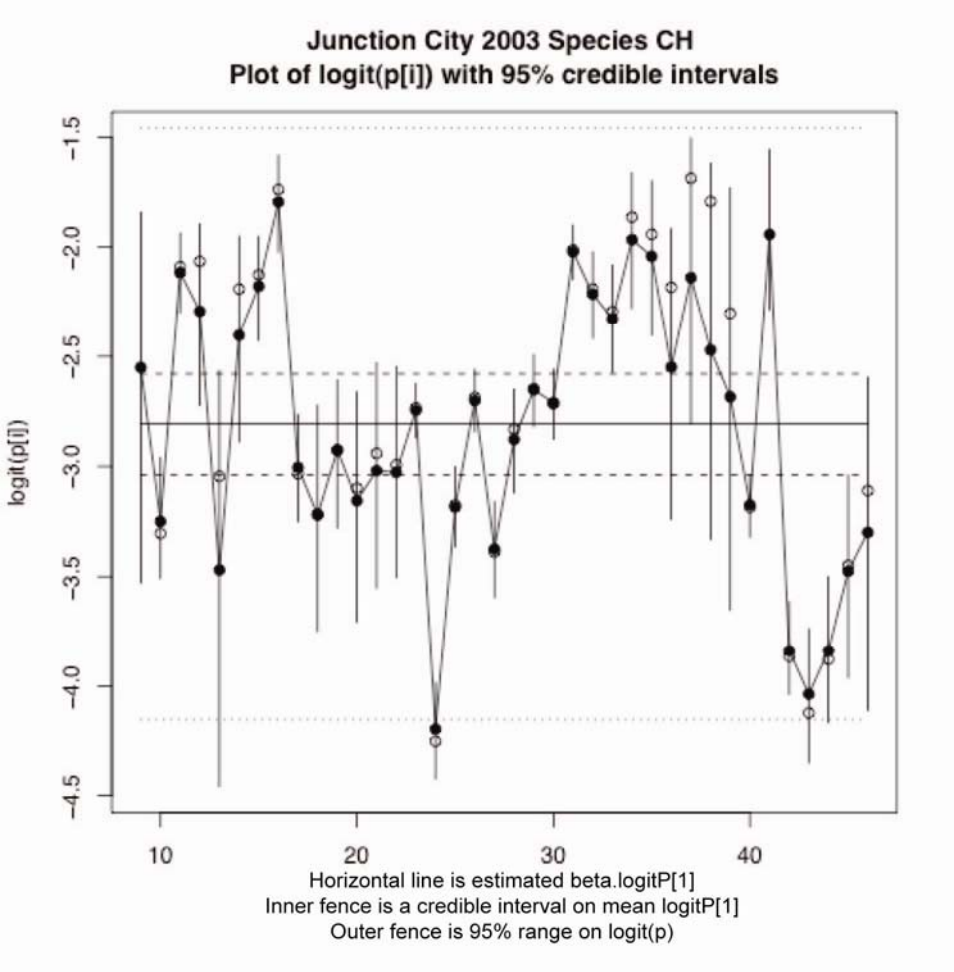


Figure 3.5. Plot of logit(p) vs Julian week.

Again note that larger uncertainty in the estimated capture probabilities in Julian week 9 when no fish were marked and released. In Julian week 41, the estimated capture probability is constrained to be consistent with the other weeks and with the observed number of unmarked fish seen in the trap given the smoothed curve on the population run size for that week. There appears to be evidence of a systematic trend in the p's (e.g. the estimated p's in Julian weeks 28-38 are all above the average). This suggests that an external covariate (such as flow) may be useful to try and explain this pattern.

The two previous plots indicated that the model seemed to fit reasonably well. The formal Bayesian p-value plots are shown in Figure 3.6.

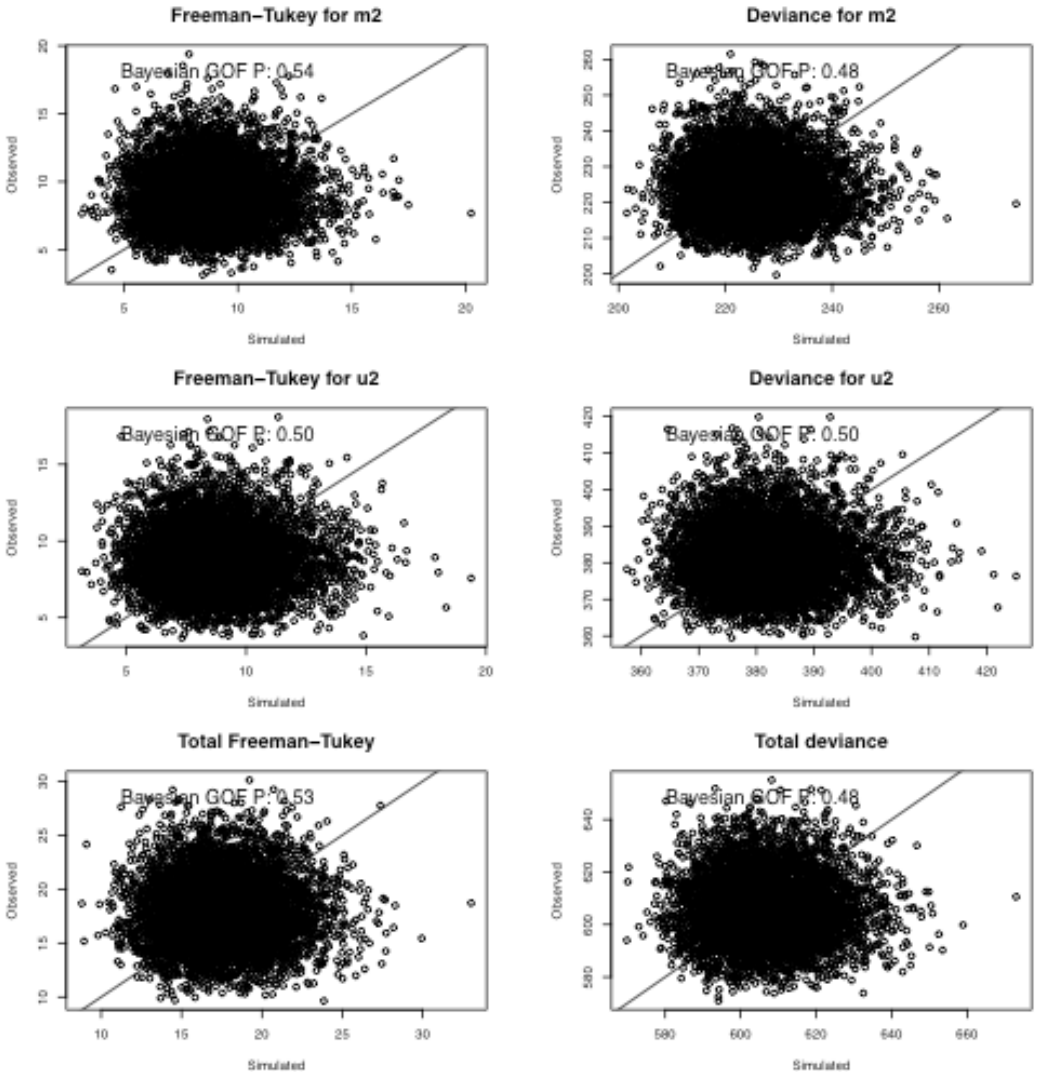


Figure 3.6. The Bayesian goodness-of-fit plots.

All of the plots had goodness-of-fit p-values around 0.50 and there is no evidence of lack-of-fit with the pattern of deviances from the observed and simulated data showing a roughly symmetric form around the line Y=X. The BRG statistics for the parameters of interest are all close to 1 again providing no indication of problems in the fit.

Estimates (SD) of the run timing quantiles (in Julian weeks) are presented in Table 3.2. For example, it is estimated that 50% of the run had passed the trap by Julian week 26.3 (SD .70).

Table 3.2. Estimated of selected quantiles of run timing (Julian weeks).

	Quantile										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Mean	9	19.4	23.7	24.3	24.7	26.3	38.6	40.7	41.9	42.7	47
Sd	0	4.00	0.13	0.07	0.07	0.70	3.16	0.09	0.21	0.06	0

Figure 5 of Green et al. (2007) appeared to show that log(flow) may be a useful covariate for explaining variation in catchability. A second model was fit with flow as measured at the USGS Junction City gauge (11526250) as the covariate. A plot of the fitted p's along with the covariate value is shown in Figure 3.7. There may be a very weak relationship between logit(catchability) and log(flow) with a large amount of scatter.

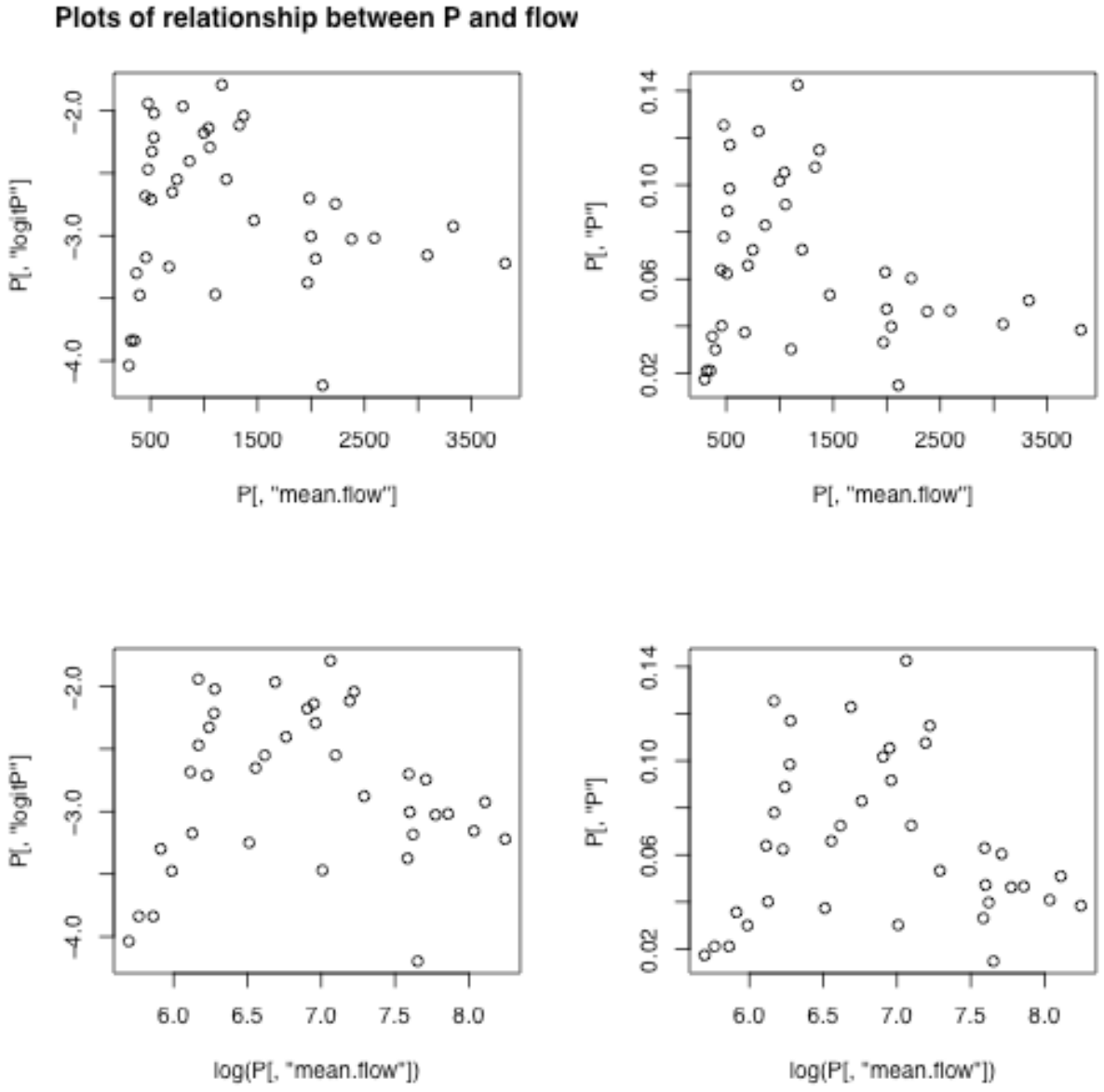


Figure 3.7. Resulting plot of estimated catchability from model with no covariates against flow and log(flow). Left column has *logit(p)* on the *Y* axis, right column has *p*. Top row has mean weekly flow on the *X* axis; bottom row has *log(mean weekly flow)* on the *X* axis.

Summary statistics for the two models (with and without using log(flow) as a covariate are presented in Table 3.3. There is actually no evidence that log(flow) is a useful covariate because the 95% c.i. for the

coefficient associated with log(flow) includes the value of 0. If the raw data are inspected more closely, the apparent decline in catchability at low flows may be an artifact of the data as these week coincide with low number of fish being marked and released and so estimates of catchability have poor precision.

Table 3.3. Summary of model fits with and without flow covariates.

	TSPDE – no covariate	**TSPDE – log(flow) as covariate**
Coefficient of log(flow)	-	-0.051 (SD .12) 95% c.i. (-.29, .18)
U-total	5.3 (SD .18) million fish	5.3 (SD .19) million fish
DIC	673.6	670.6
pD (effective # of parameters)	66.6	63.1

The two estimates of total run size are virtually identical. There is some evidence from the DIC that the model using flow is slightly preferred to the model without flow as a covariate, however, the evidence is not overwhelming.

Figure 3.8 shows a plot of the fitted logit(p) overlain with the log(flow) while Figure 3.9 shows a plot of the estimated *logit(p)* vs. the covariate directly.. While in general there appears to be a moderate relationship between flow and catchability, the two variable often move in opposite directions (e.g. Julian weeks 12, 24-29. The large numbers of fish marked and releases and recaptured in these Julian weeks overrides any relationship with log(flow) that presumes to exist.

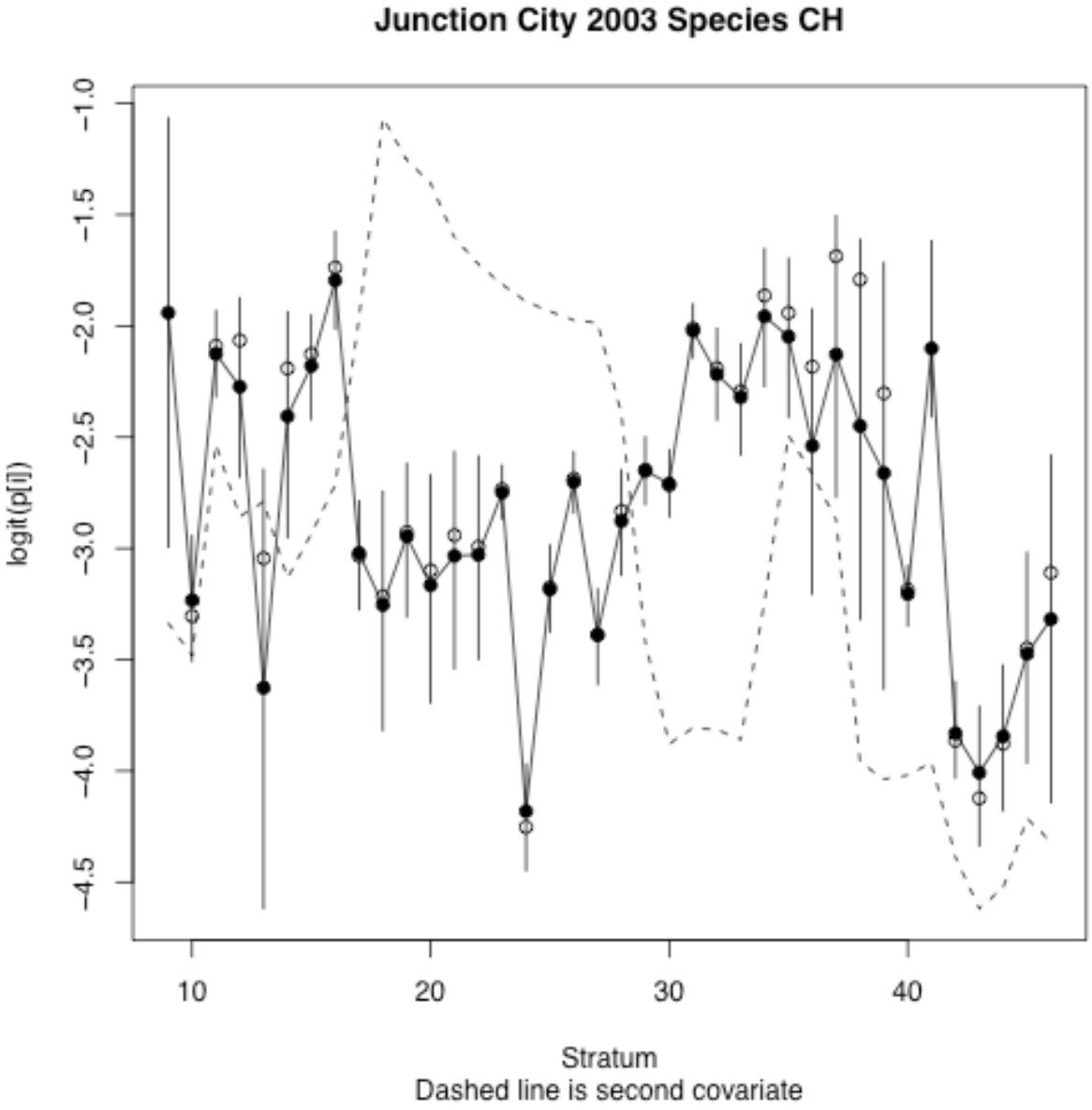


Figure 3.8. Estimated logit(p) (solid line) overlaid with the log-flow (dashed line, no scale).

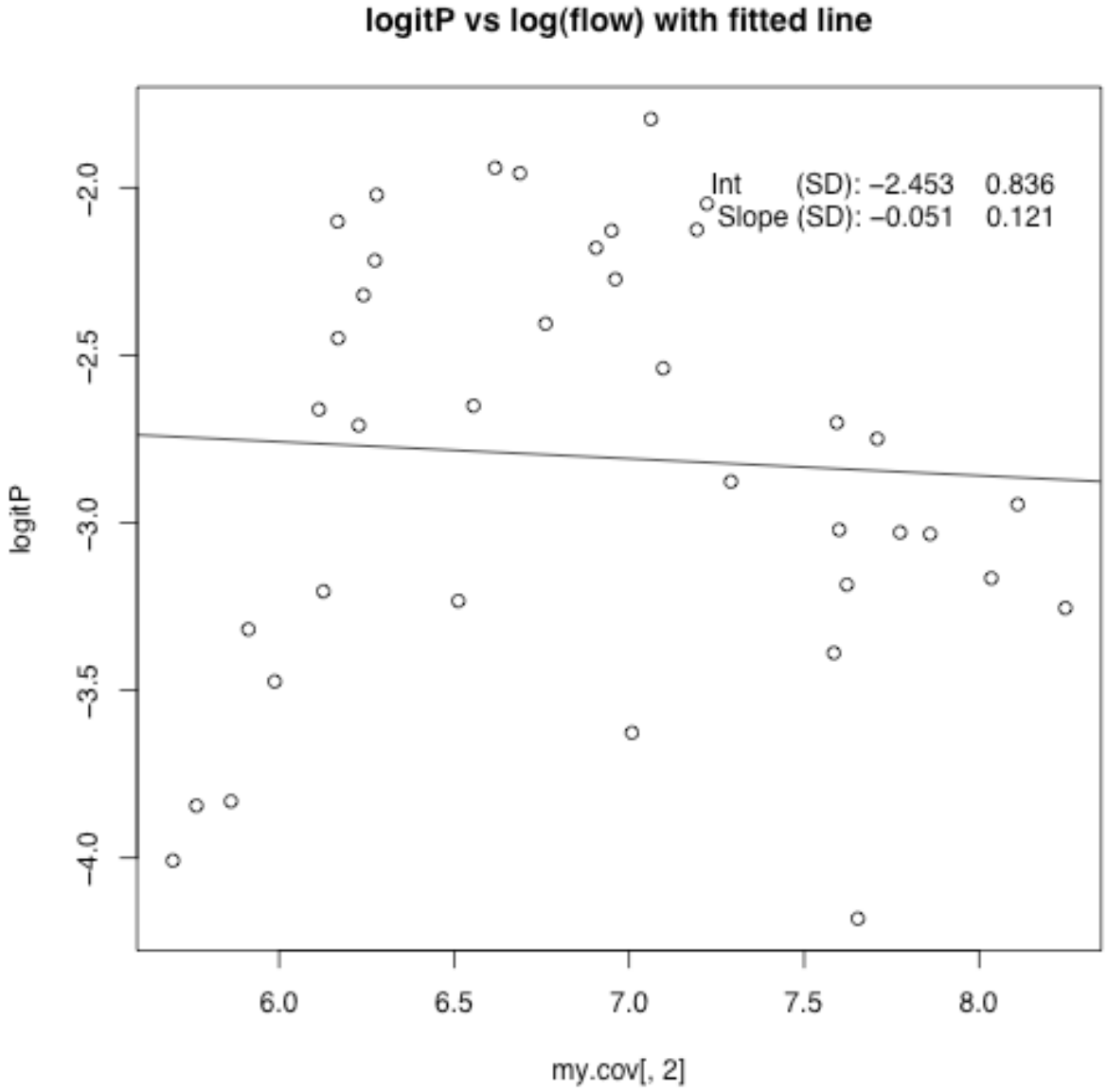


Figure 3.9. Shows the fitted relationship between the covariate (log(flow)) and the logit(P). There is no evidence of a linear relationship. A quadratic relationship could be explored, but this was not done in this report.

Example continued with separation of wild and hatchery YOY for Chinook salmon.

The data for the separation of the wild and hatchery Chinook salmon components of the run is presented in Table 3.4.

Table 3.4. Summary of raw data to separate out wild and hatchery components for Junction City 2003 Chinook salmon.

Julian Week	Days Operating	YOY Ad clipped	YOY Non-clipped	Age 1+ Ad clipped	Age 1+ Non-clipped	Marked n_i	Total recovered m_{ii}
9	3	0	4135	0	0	0	0
10	8	0	10452	0	0	1465	51
11	6	0	2181	0	18	1106	121
12	7	0	652	0	3	229	25
13	7	0	308	0	0	20	0
14	7	0	719	0	0	177	17
15	7	0	973	0	0	702	74
16	7	0	972	0	0	633	94
17	7	0	2386	0	0	1370	62
18	7	0	469	0	0	283	10
19	7	0	897	0	0	647	32
20	7	0	426	0	0	276	11
21	7	0	407	0	0	277	13
22	7	0	526	0	0	333	15
23	7	9427	26987	0	3555	3981	242
24	7	4243	12902	0	435	3988	55
25	7	1646	6179	0	103	2889	115
26	7	1359	5552	7	0	3119	198
27	7	619	2954	0	5	2478	80
28	7	258	1455	0	0	1292	71
29	6	637	3575	0	0	2326	153
30	7	753	4284	0	0	2528	156
31	7	412	2903	0	0	2338	275
32	7	173	1127	0	0	1012	101
33	7	91	898	0	0	729	66
34	7	38	406	0	0	333	44
35	7	22	317	0	0	269	33
36	7	8	99	0	0	77	7
37	7	2	77	0	0	62	9
38	7	3	37	1	0	26	3
39	7	1	22	0	0	20	1
40	7	136	0	8276	26706	4757	188
41	7	0	0	7703	26831	2876	8
42	7	0	0	3651	11309	3989	81
43	7	0	0	966	2677	1755	27
44	7	1	5	467	1338	1527	30
45	7	0	1	160	518	485	14
46	5	0	0	24	130	115	4

Notice that the hatchery release of age 1+ fish occurred in Julian week 40 and that there were negligible numbers of YOY fish captured after Julian week 40 or thereafter. Also notice that no marked fish were released in Julian week 9, and the number of fish recaptured in Julian week 41 is unusually low.

It seems unreasonable that the 4000 age 1+ fish captured in Julian weeks 23-25 are in fact ages 1+ and not simply YOY fish given that virtually none were seen elsewhere. These may simply be misclassifications of larger YOY fish and will be added to the corresponding number of YOY fish.

Similarly, the 136 YOY fish recorded as being captured in Julian week 40 are again unlikely to be YOY fish and may be smaller fish from the age 1+ category. These fish are combined with the age 1+ fish in Julian week 40.

It is clear that fish seen in Julian weeks 40 onwards are virtually all hatchery age 1+ fish. Consequently, the estimate of hatchery age 1+ fish passing the screw-trap in these weeks is found from the estimates of the total numbers of fish regardless of age in previous sections. The estimates are simply extracted from the posterior samples from the previous fit. The estimated number of hatchery aged 1+ fish in Julian weeks 40 onwards is 2.2 (SD .10) million fish.

Separation of hatchery and wild YOY fish was done for Julian weeks 9 to 39. It was assumed that 25% of hatchery fish were clipped. It should be noted that if one assumes that only hatchery fish of age 1+ are present in Julian weeks 40 onwards, then the observed clipping rate can be estimated as 23.5%. [These fish are 1 year older than the YOY fish released, so this may represent additional mortality due to clipping of the adipose-fin.]

The *R*-wrapper to separate the wild and hatchery YOY components is found in Appendix C for the analysis of this dataset and detailed results are available at the web site

http://www.stat.sfu.ca/~cschwarz/Consulting/Trinity/Phase2/.

A total of 200,000 iterations of the MCMC sampler were run with the first 100,000 discarded as burn-in, and a thinning was used to reduce autocorrelation in the estimates of the total population sizes. Approximately 6000 samples from the posterior distributions were retained. The run took approximately 20 minutes to complete running using Window's XP under Parallels (V.4) on a 2.66 GHz Intel-chip MacPro.

The pooled-Petersen estimate for the total Wild and Hatchery YOY fish is 2.0 (standard error .04) million fish. Applying the pooled-Petersen estimator to the ad-clipped YOY fish and then expanding to account for the 25% clipping rate gives an estimated hatchery YOY total population of 1.3 (standard error .03) million fish and an estimated wild YOY total population of 0.74 (standard error .02) million fish. However, as seen in earlier sections, the test for equal catchability over time is rejected which implies the estimate may be biased and the precision of the pooled-Petersen estimates is likely understated.

The stratified-Petersen estimate of the total population size (hatchery and wild YOY) is 3.0 (standard error .18) million fish. Again by assuming a 25% clip rate, the stratified-Petersen method gives an estimate of 2.3 (standard error .17) million fish for the hatchery YOY component and .71 (standard error .04) million fish for the wild YOY component.

The spline methods gave an estimate of 2.3 (SD .12) million fish for the hatchery YOY component; .87 (standard error .11) million fish for the wild YOY component ; and 3.1 (SD .15) million fish for both components. As in the previous analysis, the spline estimate is very close to the stratified-Petersen estimate because of the unequal catchability that appears to be present.

A plot of the fitted spline curve and the estimated individual components is found in Figure 3.10.

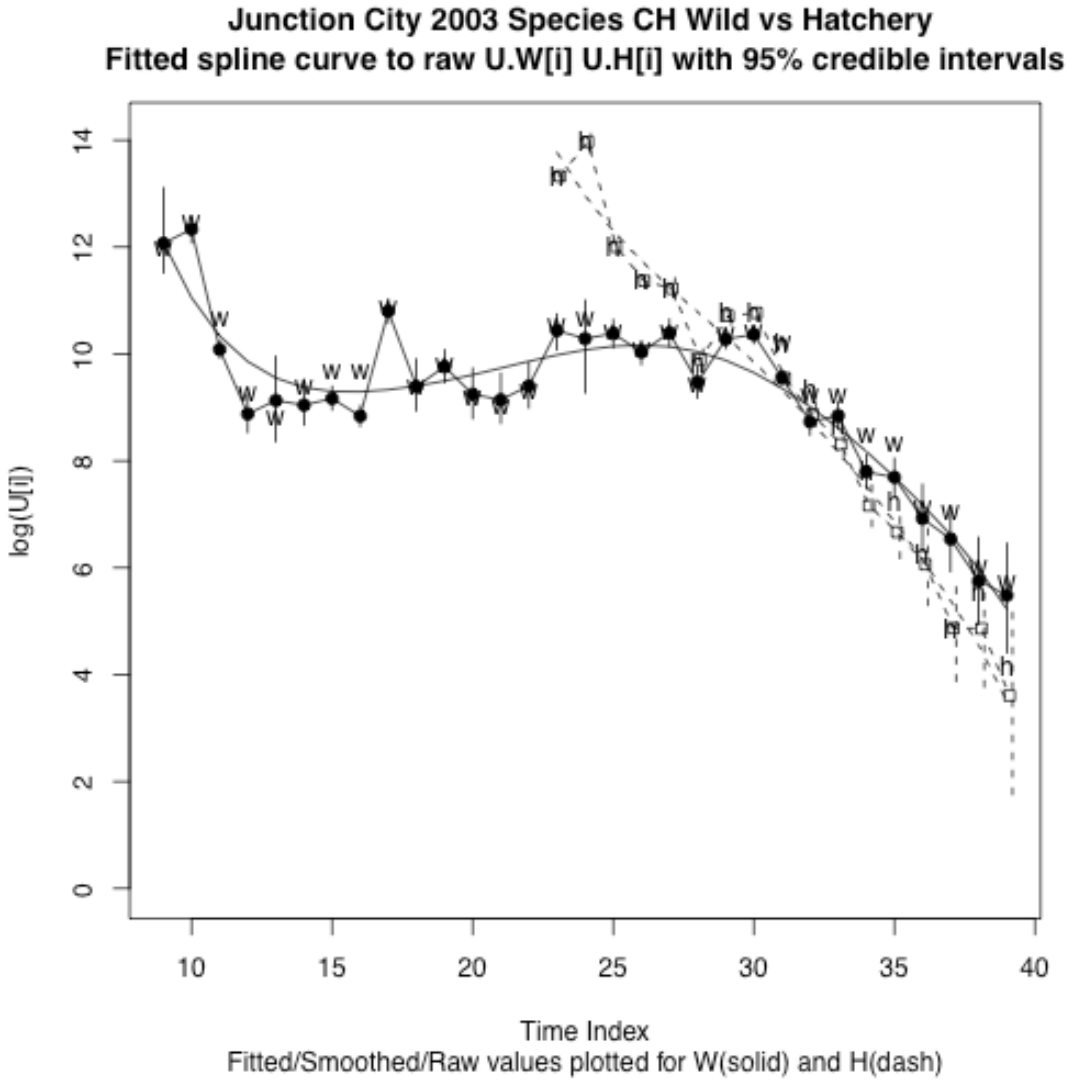


Figure 3.10. Fitted spline curves for hatchery and wild YOY components.

The fitted spline curve follows the trends in abundance fairly well. Notice how the uncertainty in the run size in Julian week 9 (when no data on the recapture rate was used) is larger than for the other weeks but the spline interpolated well.

Plots of the posterior distributions for the total run size of both components and the plot of the estimated capture probabilities are available on the web site.

The goodness-of-fit statistics showed no evidence of a lack-of-fit and plots are available on the web site. Note that goodness-of-fit tests based on the deviance are not available for these models.

Estimates (SD) of the run timing quantiles (in Julian weeks) are presented in Table 3.5 for the YOY components. The run timing for the hatchery age 1+ component is of limited interest.

Table 3.5. Estimated of selected quantiles of run timing (Julian weeks).

	Quantile										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Wild YOY											
Mean	9.0	9.5	10.0	10.3	10.7	13.0	17.3	22.5	25.8	29.2	40.0
Sd	0.0	0.2	0.2	0.3	0.3	2.1	2.6	2.0	0.8	0.6	0.0
Hatchery YOY											
Mean	23.0	23.4	23.7	24.1	24.2	24.4	24.6	24.8	25.1	26.9	40.0
Sd	0.00	0.02	0.04	0.03	0.02	0.02	0.02	0.02	0.13	0.17	0.00

Separation of Wild vs Hatchery components for steelhead

The data for the separation of the wild and hatchery steelhead components of the run is presented in Table 3.6 for the Junction City 2003 study.

Table 3.6. Summary of raw data to separate out wild and hatchery components for Junction City 2003 steelhead.

Julian Week	Days Operating	YOY Ad clipped	YOY Non-clipped	Age 1 Ad clipped	Age 1 Non-clipped	Age 2+ Ad clipped	Age 2+ Non-clipped	Marked n_i	Total recovered m_{ii}
9	3	0	0	0	58	0	0	0	0
10	8	0	0	2	357	0	0	0	0
11	6	0	0	0	701	0	19	0	0
12	7	0	0	4643	678	0	172	999	5
13	7	0	11	5706	585	52	0	1707	13
14	7	0	0	4220	532	0	0	1947	39
15	7	0	0	2328	796	0	77	2109	7
16	7	0	0	1474	249	0	54	972	1
17	7	0	1	875	230	0	61	687	0
18	7	0	33	39	12	0	0	0	0
19	7	0	31	14	92	1	9	0	0
20	7	0	11	13	43	0	4	0	0
21	7	0	78	26	46	0	3	0	0
22	7	0	46	22	38	0	6	0	0
23	7	0	35	59	44	0	6	0	0
24	7	0	30	15	30	0	8	0	0
25	7	0	309	8	53	0	5	0	0
26	7	0	278	4	33	0	3	3	0
27	7	0	207	2	13	0	0	0	0
28	7	0	196	0	5	0	0	0	0
29	6	0	613	0	12	0	0	0	0
30	7	0	764	0	15	0	0	0	0
31	7	0	556	0	11	0	0	0	0
32	7	0	250	0	12	0	0	0	0
33	7	0	106	0	12	0	1	0	0

Julian Week	Days Operating	YOY Ad clipped	YOY Non-clipped	Age 1 Ad clipped	Age 1 Non-clipped	Age 2+ Ad clipped	Age 2+ Non-clipped	Marked n_i	Total recovered m_{ii}
34	7	0	413	0	8	0	4	0	0
35	7	0	995	1	12	0	16	0	0
36	7	0	357	0	8	0	2	0	0
37	7	27	181	0	7	0	1	0	0
38	7	0	53	0	3	2	0	0	0
39	7	0	29	0	2	0	0	0	0
40	5	0	3	0	0	0	0	0	0
41	7	0	5	0	0	0	0	0	0
42	7	0	14	0	3	0	1	0	0
43	7	0	8	0	3	0	7	0	0
44	7	0	19	0	1	0	6	0	0
45	7	0	46	0	1	0	3	0	0

Notice that releases of hatchery age 1+ fish arrived at the screw trap in Julian week 12. The 27 YOY ad-clipped fish in Julian week 37 are assumed to be a recording error and are ignored. There are negligible numbers of age 2+ fish and these have been combined with the age 1 fish. Marking is very restricted and basically only happens in 6 weeks starting in Julian week 12. Recapture rates vary considerably over the 6 week marking period.

The *R*-wrapper to separate out the 3 components is found in Appendix C and detailed results are available at the web site:

http://www.stat.sfu.ca/~cschwarz/Consulting/Trinity/Phase2/.

A total of 200,000 iterations of the MCMC sampler were run with the first 100,000 discarded as burn-in[7], and a thinning[8] was used to reduce autocorrelation in the estimates of the total population size. Approximately 6000 samples from the posterior distributions were retained. The run took approximately 20 minutes to complete running using Window's XP under Parallels (V.4) on a 2.66 GHz Intel-chip MacPro.

The pooled-Petersen, stratified-Petersen, and spline methods estimates for the three components are presented in Table 3.7.

[7] Burn-in: This refers to the early part of the Bayesian algorithm. It takes the algorithm some time to move from the initial starting values to converge on a consistent posterior distribution. These early iterations are discarded so they don't affect the reported posterior distribution (See Appendix B for details)..

[8] Thinning is used to remove autocorrelation from the realized posterior distributions (See Appendix B for details).

Table 3.7. Estimates of steelhead using three methods.

	Pooled Petersen (millions)	Stratified Petersen (millions)	Spline Method (millions)
Wild YOY:	.78 (SE .10)	.01 (SE .001)	1.27 (SD .40)
Wild 1+:	.68 (SE .08)	.82 (SE .25)	1.17 (SD .23)
Hatchery 1+:	2.50 (SE .31)	3.62 (SE .90)	3.59 (SD .50)
Total:	3.95 (SE .48)	4.44 (SE 1.1)	6.04 (SD .94)

This experiment is difficult to analyze because of the very limited amount of marking done. The pooled-Petersen estimates assume homogeneity of catchability, but a test for equal catchability shows strong evidence of unequal catchability. The stratified-Petersen estimates are nonsensical because for most of the strata, no marking took place which makes it impossible to determine a sensible estimate. The spline methods have poor precision because of the need to extrapolate from the few strata where marking took place to all of the other strata.

A plot of the fitted spline curve for the three components is found in Figure 3.11.

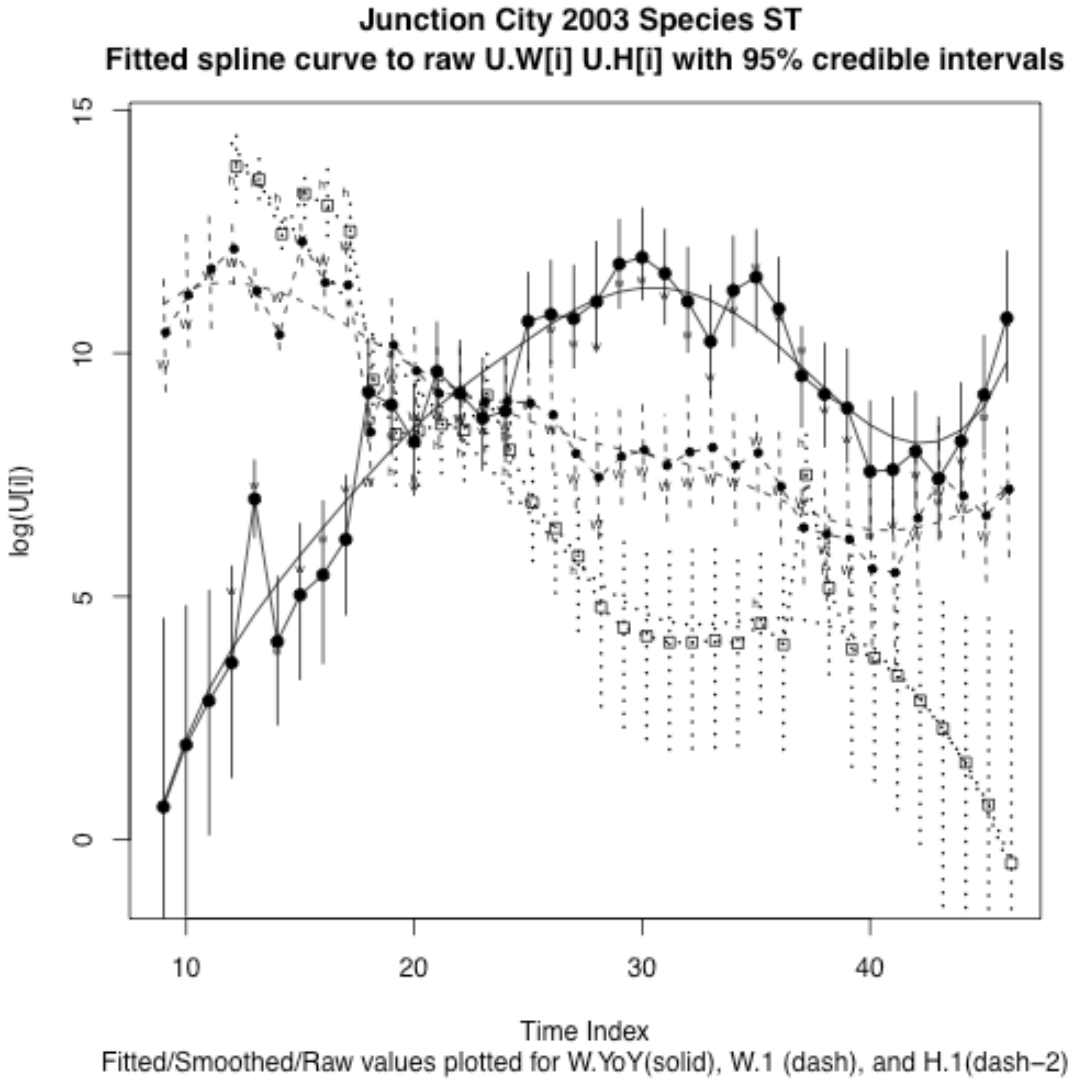


Figure 3.11. Spline fit to the steelhead data for Junction City 2003.

Notice the very wide confidence bounds for strata with no mark-recapture data. Plots of the posterior distributions for the total run size of all components are available on the web site.

A plot of the estimated capture rates is found in Figure 3.12:

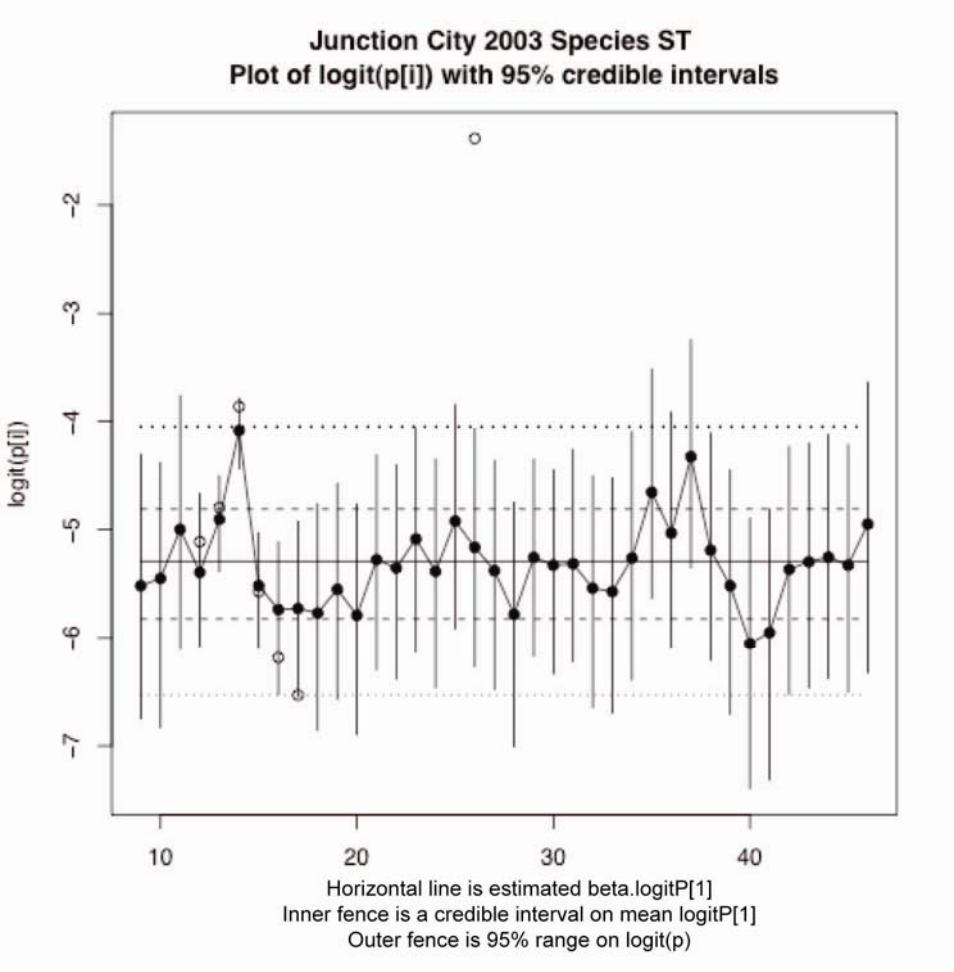


Figure 3.12. Estimated capture probabilities for 2003 Junction City steelhead dataset.

Again notice the very wide credible intervals for all strata where no-recaptures took place. Estimates are close to the mean capture rate except where the spline curve force them above or below the mean.

The goodness-of-fit measures showed no evidence of lack of fit, but with limited marking effort, likely have poor power to detect all but the grossest problems.

Finally, estimates of run timing of the 3 components is presented in Table 3.8:

Table 3.8. Estimated of selected quantiles of run timing (Julian weeks) for Junction City 2003 steelhead.

	Quantile										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Wild YOY											
Mean	9.3	26.2	28.3	29.5	30.3	31.1	32.2	33.8	35.4	37.8	47.0
Sd	0.6	0.7	0.8	0.5	0.5	0.6	1.0	1.1	0.8	2.7	0.0
Wild 1+											
Mean	9.0	11.0	11.8	12.4	13.2	14.4	15.4	16.1	17.1	19.9	47.0

Sd	0.0	0.4	0.4	0.4	0.7	0.9	0.5	0.4	0.6	1.5	0.0
Hatchery 1+											
Mean	12.0	12.4	12.7	13.1	13.5	14.0	14.8	15.6	16.2	17.0	47.0
Sd	0.0	0.1	0.2	0.3	0.4	0.6	0.6	0.4	0.4	0.3	0.0

4. Population Estimation Results

4.1 Applying the spline models to past data

The methods of the previous section were applied to the Junction City (2002-2004), Pear Tree (2003-2007), and Willow Creek (2002-2005) mark-recapture data extracted as outlined in Section 2. The data were sufficiently rich so that estimates of the total outmigration of Chinook salmon could be obtained for all studies (Table 4.1). Data for steelhead were sparser and estimates could only be obtained for fewer studies (Table 4.4). Data for Coho were extremely sparse with only a handful of fish captured and even fewer marked coho recaptured. Unfortunately, data were too sparse for the methods of this report and no estimates of abundance are reported for coho.

If one were willing to assume that the catchability of steelhead and coho were the same as Chinook salmon, then similar methods as used for separating the wild vs. hatchery components of Chinook salmon could be used to obtain estimates of abundance (and other parameters) for steelhead and coho.

In order to save space, only summary results are presented in the tables below. The full set of results (tables, plots, raw data, computer code, etc) is available at the web pages at:
http://www.stat.sfu.ca/~cschwarz/Consulting/Trinity/Phase2/.

No data are perfect and each of the studies presented some problems for a straightforward application of the TSPDE as indicated in the tables. These problems include:

(a) Weeks where no fish are marked and released, i.e. $n_i = m_{ii} = 0$ but $u_i \neq 0$. In this case information is available on the number of outgoing smolt for that week if some estimate of recapture can be obtained from other weeks.
(b) Weeks where no fish were marked and released, nor any unmarked fish captured, i.e. $n_i = m_{ii} = u_i = 0$. In this case, there is no direct information on the number of outgoing smolts. Pooled Petersen or Stratified-Petersen methods cannot estimate the number of outgoing smolt for these weeks, but the spline methods can use the trend in adjoining weeks to obtain an estimate.
(c) Large jumps in the number of outgoing smolt. In most cases, these are obvious and correspond to the release of hatchery fish. The Pooled-Petersen and Stratified-Petersen estimators do not have to adjust. The spline models allow for a sudden jump in the outmigrant population numbers at these points. The inflection points are identified by hand prior to running the spline models.
(d) Weeks were the number of recaptures appears to be unusually low, i.e. m_{ii} / n_i is unusually low. These weeks are problematic. If they correspond to known causes, e.g. a known fish kill, then these mark and recapture values are discarded (set to 0) and the previous results hold. If there is no apparent cause, then human judgment should be exercised to selectively exclude the m_{ii} value (i.e. set to missing). These cases are usually quite apparent because the Stratified-Petersen estimate of the outmigrant population size for that week is usually high.

Three types of estimators are presented.

(a) Pooled Petersen estimator (after the Chapman modification). Up to three variants of the estimator are presented. First, the Pooled Petersen is computed pooling over all releases, recaptures, and unmarked fish regardless of any problems in a particular week. For example, there may be weeks where no fish are marked, released and recaptured (i.e. $n_i = m_{ii} = 0$) but some unmarked fish are

captured ($u_i \neq 0$). If the probability of capture is approximately equal in all week, little bias would be introduced as the other weeks' recapture data would be used to estimate the recapture rate and would inflate the number of unmarked fish to estimate the number of outgoing smolt for that week. However, if no data were available on the number of unmarked fish, then this estimator will be biased downwards as no information on the number of outgoing smolt is available.

The Pooled Petersen estimator is also computed after dropping all weeks where $n_i = m_{ii} = 0$. In these cases, the estimator is expected to be biased downwards as no information on the number of outgoing smolt for those missing weeks is available.

Thirdly, in some cases weeks were the number of recaptures is abnormally low are identified. The Pooled-Petersen is then computed dropping those weeks along with weeks where $n_i = m_{ii} = 0$. This will tend to reduce the positive bias introduced by including weeks with unusually low recapture rates.

The Pooled-Petersen estimator makes the implicit assumption that capture-rates are equal for all fish. This is likely to untrue. This violation of the key assumption usually results in estimators that are nearly unbiased, but the reported standard errors can be severely biased downwards.

(b) Stratified-Petersen estimator. Up to two variants of this estimator are presented. First, all data available are used. In cases where $n_i = m_{ii} = 0$, the Chapman-modification for those strata will return the observed number of captured unmarked fish, which is clearly a lower bound to the number of outgoing fish. The reported standard error for these strata are also 0. The estimate of total outmigration will be biased downwards and estimates of precision will also be biased downwards. In cases where $n_i = m_{ii} = u_i = 0$, no estimate of the outmigrant population is available.

Second, the weeks that are problematic (i.e. $n_i = m_{ii} = 0$), or week where the number of recaptures appears to be odd (i.e. much lower than expected) are excluded. This version tries to avoid including estimates that are severely positively biased by very small recapture rates.

(c) Spline estimator (TSPDE). This estimator is computed adjusting for problematic weeks. In cases where $n_i = m_{ii} = 0$, the hierarchical model borrows information from the other weeks as to the range in possible recapture rates that are used to inflate the number of unmarked fish captured. This population estimate must also be consistent with the pattern in adjoining weeks as fitted by the spline. In weeks where $u_i = 0$, there is no direct information about the outmigrating population size for that week, but again neighboring weeks provide information as to the likely value from the fitted spline. In cases where the observed recapture rate is very small, this is treated identically to the case of $n_i = m_{ii} = 0$, i.e. the hierarchical model borrows information from the other weeks as to the recapture rates and from the spline for the population size.

4.1.1 Results for Chinook salmon

In all cases, the number of recaptures of marked fish outside the week of release is very small (generally less than 5% of all recaptures for Junction City site studies and less than 10% of all recaptures for Pear Tree site studies; no data available from the Willow Creek site studies on recaptures outside of the week of release). Consequently, the TSPDE estimator of Section 3 is generally applicable. As a rough rule of

thumb, the small number of recaptures outside the week of release will tend to bias downwards the recapture rate for that stratum which will tend to increase the estimated population size by about the same proportion of marks that occur outside the week of release, e.g. if about 5% of recapture takes place outside the week of release, estimates of the recapture rates will be biased low by about 5% and estimates of population size will be biased upwards by about 5%. This bias is acceptable as it is well below the relative standard error of the estimates.

In most of the cases in Table 4.1, the Pooled-Petersen estimator is much lower than that Stratified-Petersen estimator especially when some weeks have problematic data. The Pooled-Petersen estimator tends to be comparable to the estimator from the spline method, but with a much smaller reported standard error. If we have perfect weekly data, there is often no clear advantage to using the spline method over the Stratified-Petersen estimator if the mark-recapture data are rich, but in weeks with little, missing or problematic mark-recapture data the ability of the spline method to share information from neighboring weeks makes the spline method more reliable.

In some cases, the simple Stratified-Petersen estimator is severely biased upwards with one or two weeks accounting for the enormous bias. Then both the spline estimator and the Pooled-Petersen estimator tend to be much smaller. It should be noted that in the Junction City 2004 study, there are three weeks late in the study with low, but plausible recapture rates. These three weeks have estimates for the outmigrant population numbers that are very large. These weeks for this study need to be examined in more detail to determine if there is a known cause for the lower recapture rates.

The Willow Creek studies demonstrate the real advantage of the spline method. In all years, the mark-recapture experiment was conducted only in a small number of weeks in the middle of the study, but counts of unmarked fish were obtained for a large number of weeks prior to and after the mark-recapture window. Additionally, the number of fish marked and recaptured is very low. In these cases, the spline method tends to borrow information about the catchability from all of the weeks where mark-recapture took place. In many cases, the estimates were very similar to the pooled-Petersen estimator but now report a more realistic measure of uncertainty than the Pooled-Petersen estimator. [The Pooled-Petersen estimator will underestimate the standard error when the assumption of equal catchability across strata is violated.]

Table 4.1. Summary of estimators for the number of outmigrating Chinook salmon - All ages – hatchery and wild combined.

Study	**Estimator**	**Estimate (SE/SD) Millions of fish**	**Comments**
Junction City 2002	37 sample weeks ranging from Julian week 9 to 45. Spline jumps after Julian weeks 22, 40. Unusual m_{ii} values in Julian weeks: none. No/few marked fish released in Julian weeks: 12, 13, 16, 39. Same mark used in Julian weeks 41 and 42. Data from Julian week 42 was split in half over Julian weeks 41 and 42. Only 62/2869 recaps occurred outside of week of release.		
	Pooled Petersen using ALL data	4.5 (.081)	
	Pooled Petersen using part of data	4.3 (.079)	Excludes data from Julian weeks 16 and 39 because n_i =0.

	Stratified Petersen using ALL data	6.2 (.24)	Excludes Julian weeks 16 and 39 because n_i =0. Test for pooling has p < .001
	TSPDE	6.3 (.23)	
Junction City 2003	38 sample weeks ranging from Julian weeks 9 to 46. Spline jumps after Julian weeks 22, 39. Unusual m_{ii} values in Julian weeks: 41 (too few). No/few marked fish released in Julian weeks: 9. Same mark was used for 8 days in Julian week 10 and 6 days in Julian week 11. The number of unmarked in Julian week 8 will be reduced by a factor of 7/8. Only 26/2486 recaps occurred outside of week of release.		
	Pooled Petersen using ALL data	4.4 (.085)	
	Pooled Petersen using partial data	4.2 (.081)	Excludes data from Julian weeks 9 because n_i =0.
	Pooled Petersen using partial data	4.2 (.083)	Excludes data from Julian weeks 9 and 41 because n_i =0 or m_{ii} is unusual.
	Stratified Petersen using ALL data	16.0 (3.7)	Julian week 41 accounts for 11 million fish.
	Stratified Petersen dropping bad strata	5.0 (.21)	Julian week 41 dropped. Test for pooling has p < .0001
	TSPDE	5.3 (.18)	
Junction City 2004	41 sample weeks ranging from Julian weeks 6 to 46. Spline jumps after Julian weeks 22, 39. Unusual m_{ii} values in Julian weeks: none. No/few marked fish released in Julian weeks: 6, 7, 8, 15, 20, 21. No data for Julian week 7. Number of recaps in Julian week 12 is larger than number released but the data appears to be commingled with Julian week 15. Only 5/2363 recaptures occurred outside of week of release (after adjusting for coding error).		
	Pooled Petersen using ALL data	5.7 (.11)	
	Pooled Petersen using partial data	5.1 (.11)	Excludes data from Julian weeks 6, 7, 8, 15, 20, 21 because n_i =0.
	Stratified Petersen using ALL data	12.6 (1.0)	Julian weeks 40, 41, 42 accounts for 7.5 million fish but have "sensible" numbers of recoveries. Test for pooling has p < .0001
	TSPDE	13.6 (.76)	
Pear Tree	36 sample weeks ranging from Julian weeks 10 to 46.		

2003	Spline jumps after Julian weeks 22, 39. Unusual m_{ii} values in Julian weeks: none. Marked fish released in only 9 weeks of the program! This is a VERY sparse dataset. Over 100/500 recaps occurred outside of week of release. However, the data looks anomalous and will be treated as diagonal.		
	Pooled Petersen using ALL data	1.7 (.074)	
	Pooled Petersen using partial data	1.0 (.043)	Excludes data from 27 Julian weeks where n_i =0.
	Stratified Petersen using ALL data	2.7 (.39)	For 27 Julian weeks with no releases of marked fish, the observed value of u2 was used as the estimated population size. Test for pooling has p < .0001
	TSPDE	3.8 (.32)	
Pear Tree 2004	35 sample weeks ranging from Julian weeks 12 to 46. Spline jumps after Julian weeks 22, 39. Unusual m_{ii} values in Julian weeks: 16, 27, 30, 42, 43. Only 30/404 recaptures took place outside week of release.		
	Pooled Petersen using ALL data	5.7 (.28)	
	Pooled Petersen using partial data	5.2 (.25)	Excludes Julian weeks 12 and 20 when no releases took place.
	Pooled Petersen using partial	3.2 (.16)	Excludes Julian weeks 12,16, 20, 27, 30, 42, 43 where no releases too place or very low recaptures
	Stratified Petersen using ALL data	41 (25.0)	Julian week 42 contributed 25 million fish. Other weeks with poor recaptures also contributed large amounts to estimate.. Test for pooling has p < .0001
	Stratified Petersen using partial data	11 (3.5)	Excluded Julian weeks 16, 27, 30, 42, 43. Test for pooling has p < .0001.
	TSPDE	13 (3.0)	
Pear Tree 2005	27 sample weeks ranging from Julian weeks 1 to 27. NO Spline jumps. No marks released in Julian weeks 1, 2, 3, 4, 5, 19, 27. Unusual m_{ii} values in Julian weeks: none. Only 22/210 recaptures took place outside week of release.		
	Pooled Petersen using ALL data	5.1 (.35)	
	Pooled Petersen using partial data	4.6 (.32)	Excludes Julian weeks 1, 2, 3, 4, 5, 19, 27 where no releases too place.
	Stratified Petersen using ALL data	8.2 (1.7)	In Julian weeks 1, 2, 3, 4, 5, 19, 27 where no releases took place, observed numbers of unmarked fish used as estimate of stratum

			population size. Test for pooling has p < .0001
	TSPDE	7.6 (1.3)	
Pear Tree 2006	30 sample weeks ranging from Julian weeks 8 to 52. However, after Julian week 33, marks only released in Julian weeks 38, 40, 42, 52. These latter 4 weeks will be dropped. Spline jumps after Julian week 23. No marks released in Julian weeks 8, 9, 10, 20, 21, 22, 23, 38, 40, 42, 52. Unusual m_{ii} values in Julian weeks: none. Only 6/141 recaptures took place outside week of release.		
	Pooled Petersen using ALL data	.55 (.045)	Data after Julian week 33 discarded.
	Pooled Petersen using partial data	.44 (.036)	Excludes Julian weeks 8, 9, 10, 20, 21, 22, 23 where no releases took place as well as data after Julian week 33.
	Stratified Petersen using ALL data	.86 (.35)	In Julian weeks 8, 9, 10, 20, 21, 22, 23 where no releases took place, observed numbers of unmarked fish used as estimate of stratum population size. Data after Julian week 33 also discarded. Test for pooling has p < .0001.
	TSPDE	1.1 (.28)	
Pear Tree 2007	32 sample weeks ranging from Julian weeks 1 to 32. Spline jumps after Julian weeks 14 and 22. No marks released in Julian weeks 1, 2. 3. 4. 5. 6. Unusual m_{ii} values in Julian weeks: none. Only 126/1934 recaptures took place outside week of release.		
	Pooled Petersen using ALL data	2.0 (.045)	
	Pooled Petersen using partial data	2.0 (.044)	Excludes Julian weeks 1, 2, 3, 4, 5, 6 where no releases took place.
	Stratified Petersen using ALL data	3.1 (.23)	In Julian weeks 1, 2, 3, 4, 5, 6 where no releases took place, observed numbers of unmarked fish used as estimate of stratum population size. Test for pooling has p < .0001.
	TSPDE	3.0 (.15)	
Willow Creek 2002	37 sample weeks ranging from Julian weeks 11 to 47. Spline jumps after Julian weeks 22 and 40. No marks released in Julian weeks 11-16, 28-47. Unusual m_{ii} values in Julian weeks: none. An unknown number of recaptures took place outside week of release.		
	Pooled Petersen using ALL data	2.0 (.14)	
	Pooled Petersen	1.4 (.091)	Excludes Julian weeks 11-16 and 28-47

	using partial data		where no releases took place.
	Stratified Petersen using ALL data	1.3 (.13)	In Julian weeks 11-16 and 28-47 where no releases took place, observed numbers of unmarked fish used as estimate of stratum population size. Test for pooling has p < .0001.
	TSPDE	2.0 (.25)	
Willow Creek 2003	38 sample weeks ranging from Julian weeks 10 to 47. Spline jumps after Julian weeks 24 and 40. No marks released in Julian weeks 10-23, 31-47. Unusual m_{ii} values in Julian weeks: none. (but very small in all weeks) An unknown number of recaptures took place outside week of release.		
	Pooled Petersen using ALL data	1.2 (.22)	
	Pooled Petersen using partial data	.58 (.10)	Excludes Julian weeks 10-23 and 31-47 where no releases took place.
	Stratified Petersen using ALL data	.68 (.22)	In Julian weeks 10-23 and 31-47 where no releases took place, observed numbers of unmarked fish used as estimate of stratum population size. Test for pooling has p =.022.
	TSPDE	1.1 (.23)	
Willow Creek 2004	31 sample weeks ranging from Julian weeks 12 to 42. Spline jumps after Julian weeks 24 and 40. No marks released in Julian weeks 12, 20, 31-42. Unusual m_{ii} values in Julian weeks: none. (but very small in all weeks). An unknown number of recaptures took place outside week of release.		
	Pooled Petersen using ALL data	2.1 (.37)	
	Pooled Petersen using partial data	.83 (.15)	Excludes Julian weeks 10-23 and 31-47 where no releases took place.
	Stratified Petersen using ALL data	.44 (.07)	In Julian weeks 10-23 and 31-47 where no releases took place, observed numbers of unmarked fish used as estimate of stratum population size. Test for pooling has p =.91.
	TSPDE	1.27 (.32)	
Willow Creek 2005	27 sample weeks ranging from Julian weeks 10 to 36. Spline jumps after Julian weeks 24 and 40. No marks released in Julian weeks 10-13, 20, 28-36. Unusual m_{ii} values in Julian weeks: none. An unknown number of recaptures took place outside week of release.		
	Pooled Petersen	2,6 (.10)	

	using ALL data		
	Pooled Petersen using partial data	1.7 (.07)	Excludes Julian weeks 10-13, 20, 28-36where no releases took place.
	Stratified Petersen using ALL data	2.3 (..14)	In Julian weeks 10-23 and 31-47 where no releases took place, observed numbers of unmarked fish used as estimate of stratum population size. Test for pooling has p =.91
	TSPDE	3.7 (.38)	

Table 4.2 summarizes the mean catchability (on the logit scale) across the studies. The average catchability appears to be fairly consistent across years within the Junction City and Willow Creek sites, but there appears to be a large variability for the Pear Tree site.

Table 4.2. Comparison of mean catchability.

Site	Year	Population Estimate (millions)		Total unmarked fish	Mean logit(catchability)	
		Mean	SD		Mean	SD
Junction City	2002	6.33	0.23	230,056	-2.92	0.18
Junction City	2003	5.30	0.18	215,332	-2.80	0.12
Junction City	2004	13.63	0.76	256,410	-3.26	0.23
Pear Tree	2003	3.18	0.41	95,016	-2.99	0.26
Pear Tree	2004	13.50	3.03	79,199	-4.10	0.27
Pear Tree	2005	7.60	1.27	42,242	-4.90	0.17
Pear Tree	2006	1.15	0.28	13,072	-4.20	0.35
Pear Tree	2007	3.00	0.15	82,543	-3.56	0.17
Willow Creek	2002	1.97	0.25	67,242	-3.51	0.23
Willow Creek	2003	1.15	0.23	17,165	-4.11	0.27
Willow Creek	2004	1.26	0.32	22,087	-4.30	0.23
Willow Creek	2005	3.73	0.38	52,958	-4.32	0.27

As expected, there is a strong correspondence between the estimated population size from the spline method and a simple "Petersen-type" estimate of population size found by expanding the total unmarked fish by the average catchability (Figure 4.1). This indicates that if some independent method to estimate

the average catchability could be derived, it may be possible to obtain reasonable population estimates (but no measure of precision) for years where no capture-recapture experiments took place.

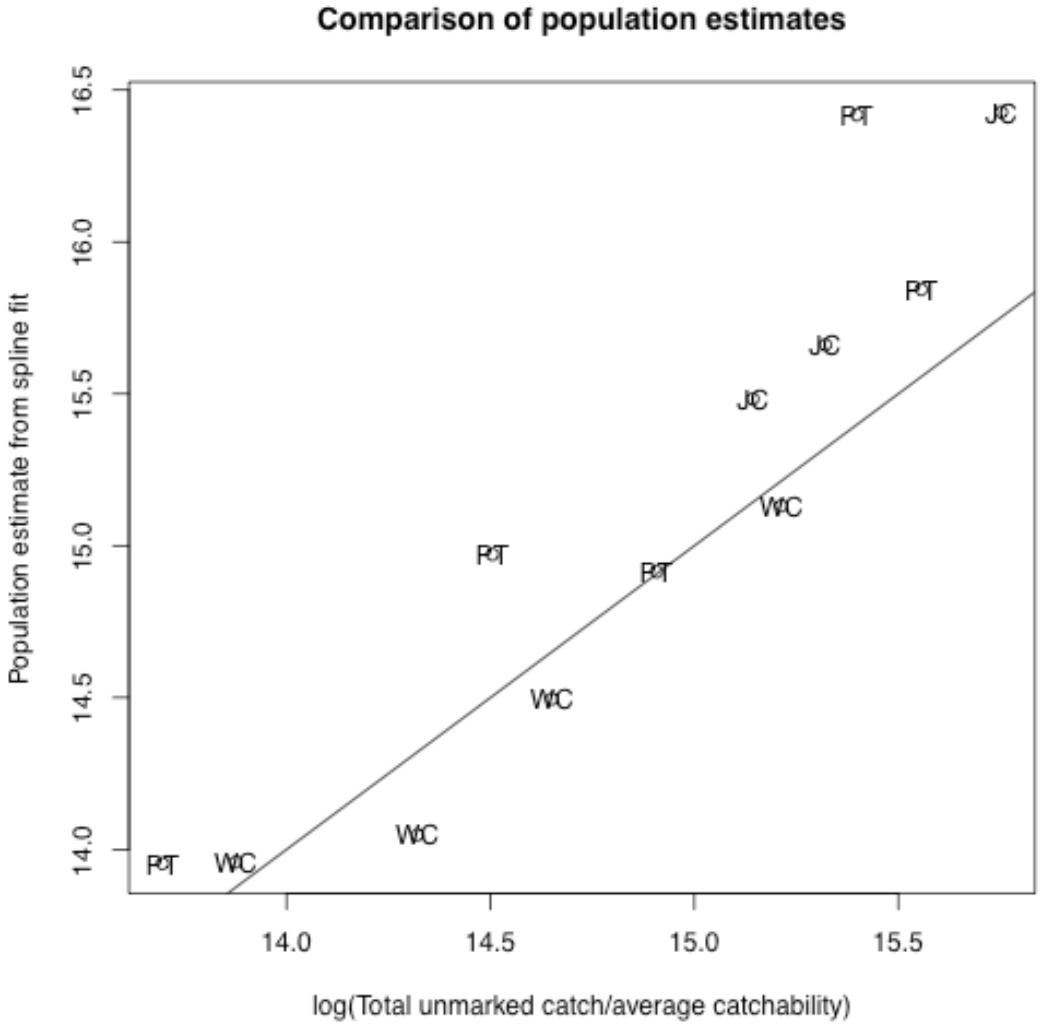


Figure 4.1. Comparison of the population estimates from spline method and average catchability method.

The spline methods were used to separate the total outmigrating Chinook salmon population into three components as presented in Table 4.3.

Table 4.3. Estimates (SD) of total Chinook salmon outmigration in the three components.

Study	Total[1] (millions)	Wild YOY (millions)	Hatchery YOY (millions)	Hatchery 1+ (millions)[2]
Junction City 2002	6.3 (.23)	2.2 (.13)	3.2 (.15)	.97 (.07)
Junction City 2003	5.3 (.18)	.87 (.11)	2.3 (.12)	2.20 (.10)

Junction City 2004	13.6 (.76)	2.4 (.37)	3.7 (.26)	7.90 (.51)
Pear Tree 2003	3.8 (.32)	.60 (.11)	.70 (.07)	1.75 (.31)
Pear Tree 2004	13.0 (3.0)	1.9 (.26)	4.0 (.37)	7.63 (2.5)
Pear Tree 2005	7.6 (1.3)	4.4 (.49)	4.7 (.32)	Not separated[3] because there does not appear to be hatchery age 1+ releases.
Pear Tree 2006	1.1 (.28)	.63 (.10)	.38 (.10)	Not separated[3] because second hatchery release of age 1+ fish occurs in last week.
Pear Tree 2007	3.0 (.15)	1.8 (.11)	1.1 (.05)	Not separated[3] because there does not appear to be hatchery age 1+ releases.
Willow Creek 2002	2.0 (.25)	1.1 (.09)	.48 (.03)[4]	Not separated[3] but appear to be after Julian week 40?
Willow Creek 2003	1.1 (.23)	.43 (.05)	.37 (.07)[4]	Not separated[3] but appear to be after Julian week 40?
Willow Creek 2004	1.3 (.32)	.74 (.20)	.19 (.06)[4]	Not separated[3] but appear to be after Julian week 40?
Willow Creek 2005	3.7 (.38)	3.0 (.49)	.64 (.04)[4]	Not separated[3] but appear to be after Julian week 40?

1 The estimates from the three columns to the right may not add exactly to this column.
2 Determined by total outmigration starting when second hatchery release reaches the screw traps.
3 Data files did not record the number of hatchery age 1+ fish.
4 It was assumed that hatchery YOY fish were present only up to Julian week 40.

There appears to be some ambiguity among the studies on the classification of hatchery fish as either YOY or age 1+ and it unknown if it is useful to separate out these two releases rather than simply estimating the total of the hatchery releases.

This analysis also assumed a known clip-rate of 25%. If the data from later in the study (typically past Julian week 40) is assumed to be ONLY hatchery age 1+ fish, the clip rate estimated from these fish varies from 22% to 30% - such variation is far in excess of the sampling variation around 25% expected from regular sampling. This may be an indication of a differential mortality rate or differential catchability of clipped vs unclipped fish. A version of the software that estimates the clip-fraction has also been developed, but this version is not recommended for general usage because it requires a very (!) long simulation series to converge. If the clipping-fraction is unknown, future studies should take a sample of the outgoing fish to estimate the clipping rate directly – this would speed convergence considerably of the latter program.

4.1.2 Results for steelhead

The mark-recapture data for steelhead was much sparser, particularly for the Pear Tree sites where sufficient (or any data!) was only available for the 2007 study. No mark-recapture data steelhead is available for the Willow Creek studies. In studies with sufficient data, the number of recaptures of marked fish outside the week of release is very small (generally less than 5% of all recaptures for Junction City studies and less than 10% of all recaptures for Pear Tree studies). Consequently, the TSPDE estimator of Section 3 is generally applicable. Results are presented in Table 4.4.

In most of the studies, the mark-recapture effort was concentrated in a small number of weeks but captures of unmarked fish took place over a much longer period of time. In these cases, the Pooled-Petersen and the TSPDE estimator assume that the capture rates from the short mark-recapture period is also applicable to the weeks where no recapture effort took place. The TSPDE estimator accounts for the variability in recapture rates observed during the mark-recapture phase, while the Pooled-Petersen estimator does not. This usually results in a much larger estimate of precision for the spline method. The Stratified-Petersen is unable to borrow information from the weeks with recaptures and so cannot estimate the number of outgoing fish for weeks without recapture effort.

Table 4.4 Summary of estimators for the number of outmigrating steelhead - All ages – hatchery and wild combined.

Study	**Estimator**	**Estimate (SE/SE) Millions of fish**	**Comments**
Junction City 2002	37 sample weeks ranging from Julian week 9 to 45. No spline jumps required. Unusual m_{ii} values in Julian weeks: none. No/few marked fish released in Julian weeks: 16, 39, 41. Same mark used in Julian weeks 41 and 42. Data from Julian week 42 was split in half over Julian weeks 41 and 42. Only 16/491 recaptures occurred outside of week of release.		
	Pooled Petersen using ALL data	.33 (.015)	
	Pooled Petersen using part of data	.26 (.011)	Excludes data from Julian weeks 16, 39, 41 because n_i =0
	Stratified Petersen using ALL data	.33 (.034)	Estimate for Julian weeks 16, 39, and 41 is observed number of unmarked fish because n_i =0.

			Test for pooling has p < .001.
	TSPDE	.41 (.047)	
Junction City 2003	38 sample weeks ranging from Julian weeks 9 to 46. No spline jumps. Fish marked and released ONLY in Julian weeks 12 to 17 with no/few recaptures from fish released in weeks 16 and 17. Only 2/65 recaptures occurred outside of week of release.		
	Pooled Petersen using ALL data	4.0 (.48)	
	Pooled Petersen using partial data	3.0 (.36)	Excludes data from ALL Julian weeks except 12 to 17 because n_i =0
	Stratified Petersen using ALL data	4.4 (1.1)	Estimates of outgoing fish in all Julian weeks except 12 to 17 are the observed number of unmarked fish. Test for pooling has p < .0001.
	TSPDE	6.2 (1.3)	Extensive interpolation required outside of Julian weeks 12 to 17.
Junction City 2004	41 sample weeks ranging from Julian weeks 6 to 46. No spline jumps. Fish marked and releases ONLY in Julian weeks 12, 13, 14, 16, 17, 18, and 19. Only 7/66 recaptures occurred outside of week of release.		
	Pooled Petersen using ALL data	1.9 (.28)	
	Pooled Petersen using partial data	1.3 (.16)	Excludes data from ALL Julian weeks except 12-14 and 16-19 because n_i =0.
	Stratified Petersen using ALL data	1.7 (.66)	Estimates of outgoing fish in all Julian weeks except 12-14 and 16-19 are the observed number of unmarked fish. Test for pooling has p = .17.
	TSPDE	3.5 (1.3)	
Pear Tree 2003	36 sample weeks ranging from Julian weeks 10 to 46. Only 1 fish marked and released. NO mark-recapture estimate is sensible.		
Pear Tree 2004	35 sample weeks ranging from Julian weeks 12 to 46. Only 8 fish recaptured from 239 releases. Mark-recapture estimates not computed.		
Pear Tree 2005	27 sample weeks ranging from Julian weeks 1 to 27. Only 7 fish recaptured from 824 releases. Mark-recapture estimates not computed.		
Pear Tree 2006	30 sample weeks ranging from Julian weeks 8 to 52. NO fish recaptured from 212 releases, NO mark-recapture estimate is sensible.		
Pear Tree 2007	32 sample weeks ranging from Julian weeks 1 to 32. No spline jumps.		

	No marks released in Julian weeks 1, 2. 3. 4. 5. 6. Unusual m_{ii} values in Julian weeks: none. Only 7/88 recaptures took place outside week of release.		
	Pooled Petersen using ALL data	.33 (.034)	
	Pooled Petersen using partial data	.32 (.033)	Excludes Julian weeks 1, 2, 3, 4, 5, 6 where no releases took place.
	Stratified Petersen using ALL data	.38 (.066)	In Julian weeks 1, 2, 3, 4, 5, 6 where no releases took place, observed numbers of unmarked fish used as estimate of stratum population size. Test for pooling has p < .0001
	TSPDE	.44 (.066)	
Willow Creek 2003	No mark recapture data available.		
Willow Creek 2004	No mark recapture data available.		
Willow Creek 2005	No mark recapture data available.		
Willow Creek 2006	No mark recapture data available.		

The spline methods were used to separate the total outmigrating steelhead population into three components as presented in Table 4.5.

Table 4.5 Estimates (SD) of total steelhead outmigration in the three components.

Study	**Total (millions)**	**Wild YOY (millions)**	**Wild 1+ (millions)**	**Hatchery 1+ (millions)**
Junction City 2002	.39 (.05)	.12 (.02)	.10 (.02)	.16 (.03)
Junction City 2003	6.0 (.95)	1.3 (.40)	1.2 (.23)	3.6 (.50)
Junction City 2004	2.6 (.47)	1.4 (.23)	.70 (.13)	.48 (.23)
Pear Tree 2003	Insufficient Data			
Pear Tree 2004	Insufficient Data			
Pear Tree 2005	Insufficient Data			
Pear Tree 2006	Insufficient Data			
Pear Tree 2007	,46 (.07)	.19 (.03)	.11 (.02)	.16 (.03)

Willow Creek 2002	Insufficient Data
Willow Creek 2003	Insufficient Data
Willow Creek 2004	Insufficient Data
Willow Creek 2005	Insufficient Data

Unfortunately, separation of wild and hatchery stocks for steelhead often has insufficient mark-recapture data to provide useful estimates for the Pear Tree or Willow Creek studies. In the Junction City studies, the separation was straightforward, but has poor precision because of the limited marking that was done. This required that the variation in the capture rates observed during the limited marking period had to be extended to the other strata to account for potential variation in catchability.

4.2 Discussion and guidance

The spline-based TSPDE methodology has several compelling advantages over the pooled- or stratified-Petersen estimator:

- it automatically adjusts the estimates of precision for abundance for heterogeneity in catchability among weeks, The pooled-Petersen will produce estimates of precision that are biased low, i.e. the estimates look more precision than they are in reality.
- it shares information on the weekly catchability from other weeks particularly when the number of fish marked and released is small. For example, the stratified-Petersen method cannot produce a sensible estimate when the number of recaptures of marked fish is zero.
- it automatically imputes estimates of catchability for weeks where no marking effort took place based on the catchability in the other weeks. The stratified-Petersen method cannot produce any estimate when there is no marking nor any capture of unmarked fish.
- it automatically imputes estimates of abundance based on the underlying spline when data are missing on the number of unmarked fish captured by the trap.
- it automatically adjusts the estimates of precision for abundance for weeks where no mark-recapture data are available or when interpolation from the spline is used when no unmarked fish are counted.
- it can readily incorporate covariate information on the catchability such as flow or temperature.
- because individual weekly estimates of abundance are provided, estimates of run timing are easily computed along with measures of precision unlike in the Petersen estimators.

The spline-based methods is also self-calibrating. Unless traditional mark-recapture models a wide variety of potential models are fit (e.g. catchability varying by week, catchability constant over weeks, etc) and information-theoretic methods (e.g. AIC) are used to select the "best" model and to average results over several competing models. The spline-based model will automatically fit a more complex model as more (and richer) data become available and will fit a "simpler" model in cases with sparse data. For example, if the number of fish marked and releases is very small in each stratum, the spline-based model will automatically "choose" a model where the catchability could be constant over the weeks (i.e. tending towards a pooled-Petersen). However, if the number of fish marked and released is very large in each week, the spline-based model will automatically fit a model that allows the catchability to vary among weeks (i.e. tending towards a stratified-Petersen estimator). The spline-based model also allows the number of unmarked fish to vary around the smooth spline depending on the amount and quality of data

available. Large variations are allowed when the data are rich, but smoothness is enforced when data are sparse.

The spline-based methods also provide a way to modify the sampling protocol without incurring large penalties in precision or bias. For example, during the later part of the run when the number of fish passing the trap is small, sampling could be reduced to biweekly rather than weekly. Or in some week, counts of unmarked fish are still obtained, but no marking is done.

Of course no method is perfect. The spline-based TSPDE will have difficulties in providing estimates under the following conditions:

- catchability among weeks does not vary around a common mean. For example, if the number of screw traps in operation changed each week then the hierarchical model will try and fit a common mean and variance when in fact a common mean is not appropriate. The hierarchical model will still "work", but now the estimated precision will be larger than it needs be. However, if the number of traps in operation in each week were known, this could be used a covariate to group weeks into sets with common numbers of traps operating.
- interpolation will work well when the underlying spline is a realistic approximation to the run. Interpolations on the declining arm of the spline are well buttressed by the trend on either side of the gap. However, it could be very dangerous to extrapolate before the start of sampling where the shape of the underlying run is unknown. Consequently it is important to start sampling before a substantial portion of the run has passed the trap.
- the method implicitly assumes that catchability is constant within a week. If the number of traps changes within a week, the resulting heterogeneity will likely result in a underestimate of the precision, but based on extensive simulations, it is far more important to account for heterogeneity among weeks rather than heterogeneity within weeks. If different marks were used within a week, it would be possible to investigate how much catchability varies within a week. The computer-programs could be modified to include this additional information, but this has not been done.
- the computations for the methods are more complex and very "black-boxish", i.e. it is extremely difficult to follow intermediate computations during the MCMC stage. Consequently, it is important that a good intuitive understanding of the roles of hierarchical model for catchability (sharing information on catchability) and the spline (sharing information on abundance trends) is essential to avoid using the method in inappropriate situations.

It is not surprising that the estimates from the spline-based methods are very similar to the estimates from the Stratified-Petersen in cases where the data are complete (i.e. few missing weeks) and large number of fish are marked and released. In these circumstances, the spline-based method is essentially fitting a stratified-Petersen model and little sharing of information takes place. However, as soon as there are strata with missing data, the spline-based methods provide estimates where the other methods cannot.

Neither the spline-based, nor the Petersen variants can deal with cases where the data do not appear to be unusually, but the estimates lead to unrealistic estimates of abundance. For example, in the Pear Tree 2005 example, all of the method estimate that about 3 million fish passed the trap from the second hatchery release, but this release was only 1.5 million fish! The raw data appears to be consistent with other weeks, so it is unclear what went wrong in the sampling protocol to arrive at this unrealistic estimate.

5. Evaluation of Flow Based Population Estimates

5.1 Methods

The evaluation of the usefulness of using flow as a covariate to model catchability was performed in three ways. First, models with and without using flow as a covariate were fit to the Chinook salmon data sets. [The steelhead and coho datasets were too sparse to be useful.] Second, an empirical relationship across all years and datasets was used to estimate population numbers without using the mark-recapture data, mimicking a proposed approach for the pre-2002 data where no mark-recapture data were collected. Finally, we used the method described by Pinnix et al. (2007) to expand the catch estimates based on the relative volume of flow through the traps compared to the flow from the entire river.

5.2 Obtaining flow information

Flow information was available from two sources. First, average daily flow (cfs) information was available from the HVT/USGS gauge (
Table 2.4) at Junction City (reference number 11526250; 1995-2009), USGS gauge at Pear Tree (reference number 11526400; 2005-2009), and the USGS gauge at Willow Creek (reference number 11530000; 1911-2009). This record of flows was complete with no missing values. However, the Pear Tree USGS gauge was not in operation in 2003 and 2004. A plot of the USGS Pear Tree readings vs. the USGS Junction City readings showed a very strong linear relationship between the two (upper right of Figure 5.1) and so the Junction City flow reading was used for the Pear Tree flow in 2003 and 2004.

Second, daily flow measurements were taken on the screw traps at the three locations (Table 2.3) and a database was created from information provided. This information is not complete with many missing values when the traps were not in operation or the gauges were not working.

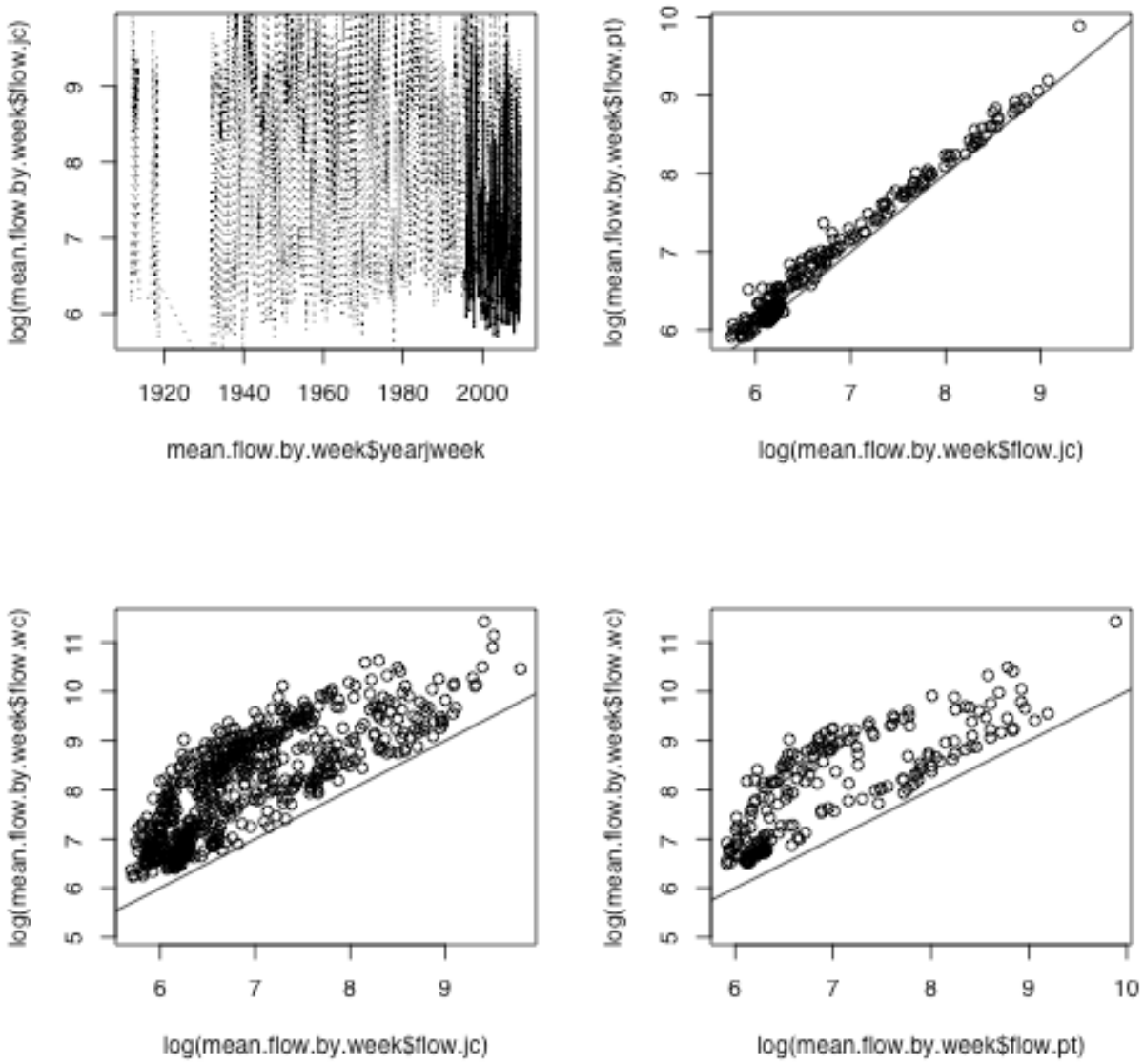


Figure 5.1. Comparison of the average weekly flows among the Junction City, Pear Tree, and Willow Creek USGS/HVT gauges. The upper left plot is a plot of the three readings over time – the wide variation in Junction City flows makes the graph difficult to read.

5.3 Using flow as a covariate in conjunction with mark-recapture data

Preliminary plots (not shown) showed that log(flow) was a suitable covariate for modeling logit(p). Table 5.1 compares the summary results for most study sites when models with and without the log-flow covariate (as measured at the HVT/USGS gauges) are fit in conjunction with the mark-recapture data as outlined in Section 3.7. Models incorporating the log(flow) as a covariate were only fit to the Chinook salmon datasets as the steelhead and coho datasets are too sparse for such models to be useful.

There was no consistent pattern in the estimated slope associated with log(flow). The 95% c.i. for the coefficient associated with log(flow) often included 0 or one of the bounds was very close to zero. A comparison of the DICs did not show any overwhelming preference for the model with the flow covariate. Estimated population sizes were very similar under the two approaches (with or without the flow covariate included).

There is little evidence to suggest that using log(flow) will either improve estimation of population numbers or could serve as a surrogate for catchability. This is somewhat good news as it implies that there is no strong evidence that the efficiency of the screw traps varies with the flow of the river and that the screw-traps may be sampling at a consistent rate regardless of the flow. This does NOT imply that the screw traps are sampling a constant fraction of the river. For example, at high flows, the trap is anchored on the edge of the stream, fish may swim near the edges of the river and still be sampled at the same rate despite the fact that the screw traps are sampling a smaller fraction of the river's flow.

Table 5.1. Summary of model fits with and without flow covariates for Chinook salmon (all ages, wild and hatchery combined).

	TSPDE – no covariate	**TSPDE – log(flow) as covariate**
Junction City 2002 Chinook salmon		
Coefficient of log(flow)	-	-0.37 (SD .16) 95% c.i. (-.68, -.05)
U-total	6.3 (SD .27) million fish	6.3 (SD .25) million fish
DIC	645.4	643.4
pD (effective # of parameters)	66.9	65.7
Junction City 2003 Chinook salmon		
Coefficient of log(flow)	-	-0.051 (SD .12) 95% c.i. (-.29, .18)
U-total	5.3 (SD .18) million fish	5.3 (SD .19) million fish
DIC	673.6	670.6
pD (effective # of parameters)	66.6	63.1
Junction City 2004 Chinook salmon		
Coefficient of log(flow)	-	-0.47 (SD .14) 95% c.i. (-.73, -.37)
U-total	13.6 (SD .76) million fish	13.3 (SD 1.2) million fish
DIC	654.1	630.8
pD (effective # of parameters)	42.8	19.5
Pear Tree 2003 Chinook salmon		
Coefficient of log(flow)	-	-0.22 (SD .17) 95% c.i. (-.78, -.14)
U-total	3.0 (SD .32) million fish	3.3 (SD .33) million fish
DIC	4385	438.4
pD (effective # of parameters)	35.0	33.3
Pear Tree 2004 Chinook salmon		
Coefficient of log(flow)	-	-0.47 (SD .17) 95% c.i. (-.53, 11)

U-total	13 (SD 3) million fish	18.(SD 9) million fish
DIC	443.1	444.3
pD (effective # of parameters)	11.1	13.0
Pear Tree 2005 Chinook salmon		
Coefficient of log(flow)	-	-0.57 (SD 0.14) 95% c.i. (-.86, -.30)
U-total	7.6 (SD 1.2) million fish	7.5 (SD .81) million fish
DIC	403.0	403.6
pD (effective # of parameters)	40.5	38.6
Pear Tree 2006 Chinook salmon		
Coefficient of log(flow)	-	-0.46 (SD 0.15) 95% c.i. (-.74, -.17)
U-total	1.1 (SD .28) million fish	1.3 (SD .32) million fish
DIC	316.8	326.0
pD (effective # of parameters)	24.3	33.7
Pear Tree 2007 Chinook salmon		
Coefficient of log(flow)	-	-.54 (SD .14) 95% c.i. (-.81, -.25)
U-total	3.0 (SD .15) million fish	3.0 (SD .16) million fish
DIC	533.4	533.2
pD (effective # of parameters)	53.1	52.8
Willow Creek 2002 Chinook salmon		
Coefficient of log(flow)	-	-.35 (SD .12) 95% c.i. (-.57, .12)
U-total	2.0 (SD .25) million fish	1.8 (SD .14) million fish
DIC	436.9	437.5
pD (effective # of parameters)	42.2	42.5
Willow Creek 2003 Chinook salmon		
Coefficient of log(flow)	-	-.26 (SD .13) 95% c.i. (-.51, -.02)
U-total	1.1 (SD .23) million fish	1.6 (SD .48) million fish
DIC	366.6	367.2
pD (effective # of parameters)	39.2	39.3
Willow Creek 2004 Chinook salmon		
Coefficient of log(flow)	-	-.40 (SD .14) 95% c.i. (-.65, .12)
U-total	1.3 (SD .32) million fish	1.2 (SD .28) million fish
DIC	337.4	337.2
pD (effective # of parameters)	34.5	34.5
Willow Creek 2005 Chinook salmon		
Coefficient of log(flow)	-	-.37 (SD .14)

		95% c.i. (-.63, -.11)
U-total	3.7 (SD .38) million fish	3.6 (SD .39) million fish
DIC	352.1	351.8
pD (effective # of parameters)	35.5	35.3

If flow doesn't appear to affect catchability, this could imply that the catchability is roughly constant over time. If so, then a simple index of the total of the u_i vs. estimated population size may prove to be used as in Pinnix et al. (2007).

5.4 Using an empirical relationship of flow with catchability across years and datasets.

The previous section allowed for a separate relationship between flow and catchability for each dataset. This was possible because of the presence of the mark-recapture data. However, in order to use a flow relationship for the pre-2002 data to predict the run size by week, an empirical relationship between flow and catchability that can be used in the absence of mark-recapture data must be available.

Figure 5.2, Figure 5.3 and Figure 5.4 present summary plots of the relationship between the empirical logit of catchability ($\hat{p}_i = \frac{m_{ii}}{n_i}; \text{elogit}(\hat{p}_i) = \log\left(\frac{\frac{m_{ii}+.5}{n_i+1}}{1-\frac{m_{ii}+.5}{n_i+1}}\right)$) and the log(flow) as measured at the USGS/HVT flow gauges for the Chinook salmon datasets based on weeks with sufficient mark-recapture data ($n_i \geq 20$) along with a fitted line (generalized linear model) between the two variables for each year of the study.

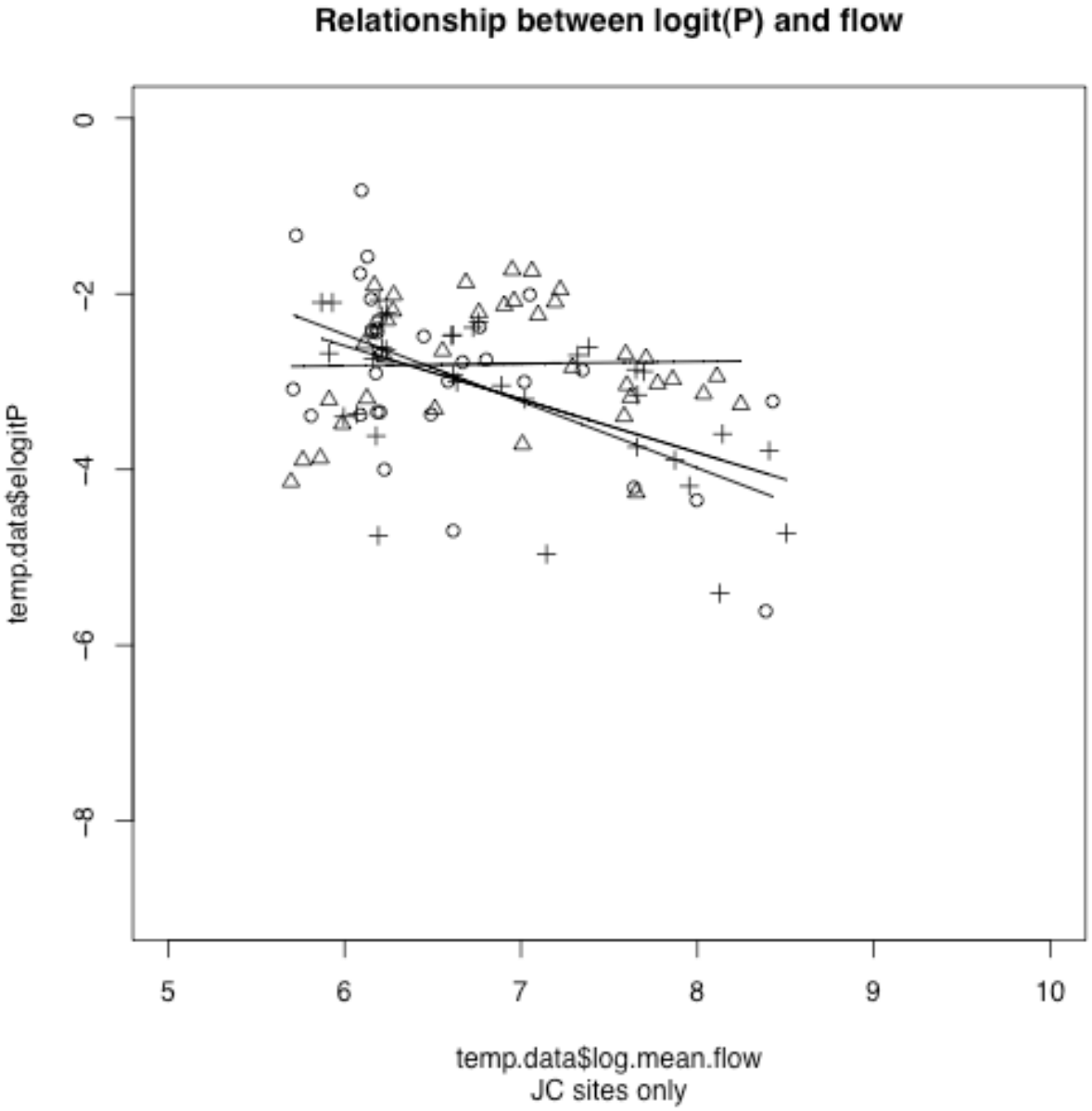


Figure 5.2. Empirical relationships between logit(P) and log(flow) as measured at the USGS/HVT Junction City gauge for Junction City Chinook salmon datasets. Each line and symbol represents a different year.

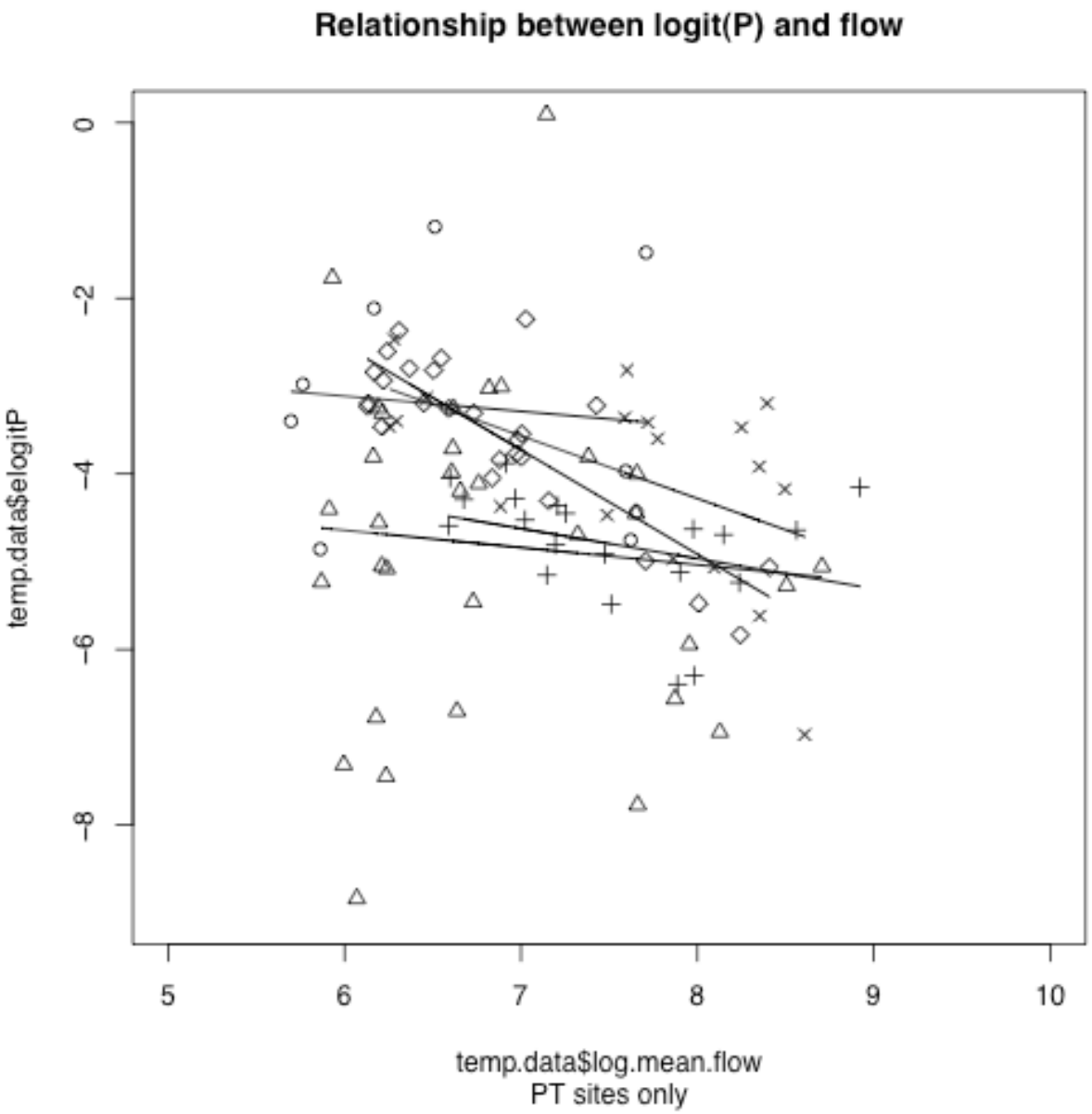


Figure 5.3. Empirical relationships between logit(P) and log(flow) as measured at the USGS Pear Tree gauge for Pear Tree Chinook salmon datasets. Each line and symbol represents a different year. In 2002 and 2003, the Junction City flow measurements were used because the Pear Tree flow data were not available.

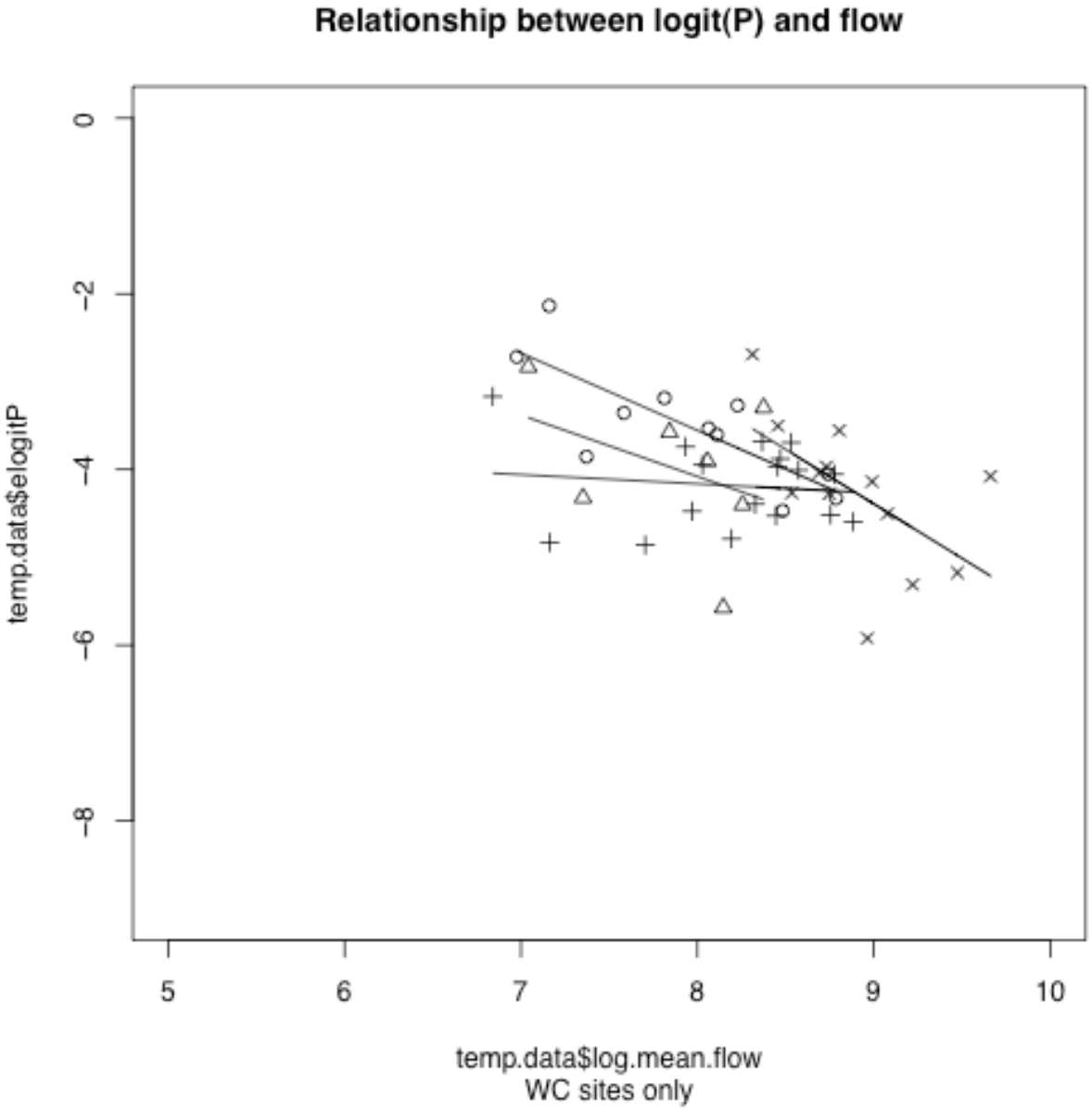


Figure 5.4. Empirical relationships between logit(P) and log(flow) as measured at the Willow Creek USGS gauge for Willow Creek Chinook salmon datasets. Each line and symbol represents a different year.

These plots show that the relationship between catchability and flow is variable across years within a site, but roughly comparable across sites. Table 5.2 summarizes the formal fitting process along with tests of hypotheses that a single relationship between logit(p) and log(flow) is tenable across all years for each site. For the Junction City site, there is strong evidence that a separate relationship is needed for each year; for the Pear Tree site, there is evidence that parallel lines across years are satisfactory; for the Willow Creek site a single relationship across all years may be tenable. The estimated common line in the last column is consistent with the estimates reported in Table 5.1 where log(flow) was used directly as a covariate in the spline-model.

Table 5.2. Summary of model fitting to the relationship between logit(p) and log(flow).

Site	Summary of Hypothesis Tests.[a] The model **Year LogFlow Year*LogFlow** implies separate lines for each year of the study. The model **Year LogFlow** implies parallel lines across years of the study. The model **LogFlow** implies a single line over all years.	Estimated line common to all years., i.e. the model logit(p)=**LogFlow** Residual overdispersion:
Junction City	Year LogFlow Year*LogFlow vs Year LogFlow p=.005 Year LogFlow vs LogFlow p=.15	Logit(p)=-.05-.42(LogFlow) $\sigma = .80$
Pear Tree	Year LogFlow Year*LogFlow vs Year LogFlow p=.27 Year LogFlow vs LogFlow p=.002	Logit(p)=-.79-.48(LogFlow) $\sigma = 1.38$
Willow Creek	Year LogFlow Year*LogFlow vs Year LogFlow p=..13 Year LogFlow vs LogFlow p=.12	Logit(p)=.61-.56(LogFlow) $\sigma = .66$

[a] The first test is of the hypothesis that the slopes are parallel with separate intercepts for each year vs. separate lines for each year. The second test is of the hypothesis that a single line is sufficient for all years vs separate lines with parallel lines across years.

The utility of a simple flow-based model without the mark-recapture data was evaluated for each site as follows. A single model over all years for each site (third column of Table 5.2) was used to predict the catchability at different flows. [This common model is only tenable for the Willow Creek site.] The generalized linear model (logistic regression) also found substantial overdispersion (in the relationship between logit(p) and log(flow) indicating that the points in Figure 5.2, Figure 5.3 and Figure 5.4 are scattered about the common line much more that expected from simple binomial variation.

The performance of this flow-based model when applied to data without mark-recapture information was evaluated by applying the following empirical-bootstrap procedure to the existing data:

(a) The third column of Table 5.2 was used to estimate the mean logit(probability) of capture for each Julian week based on the log(flow) for that week.
(b) Random noise consistent was added to each logit(probability) of capture using the standard deviation in the third column of Table 5.2.

(c) The probability of capture was obtained from the logit values.

(d) The estimated population size in this stratum was found by $\hat{U}_i = \frac{u_i}{\hat{p}_i}$

(e) The estimated total population size was found by summing the individual weeks' population sizes.

(f) This was repeated for 1000 times and the mean and standard deviation of the population estimates were obtained.

For example, Figure 5.5 shows the mean weekly population sizes (and 95% intervals) for the Junction City 2003 Chinook salmon data. The estimated population size is 6.0 (SD 1.9) million fish. The estimate must necessarily follow the pattern of the unmarked fish and the weeks with very large number of unmarked Chinook salmon passing have very large estimated population sizes. The precision of each week's estimate is very poor because of the overdispersion seen in the earlier plots. The poor precision for the grand total (see right margin of plot) is a consequence of 3 or 4 weeks with very large estimates of abundance each of which has poor precision.

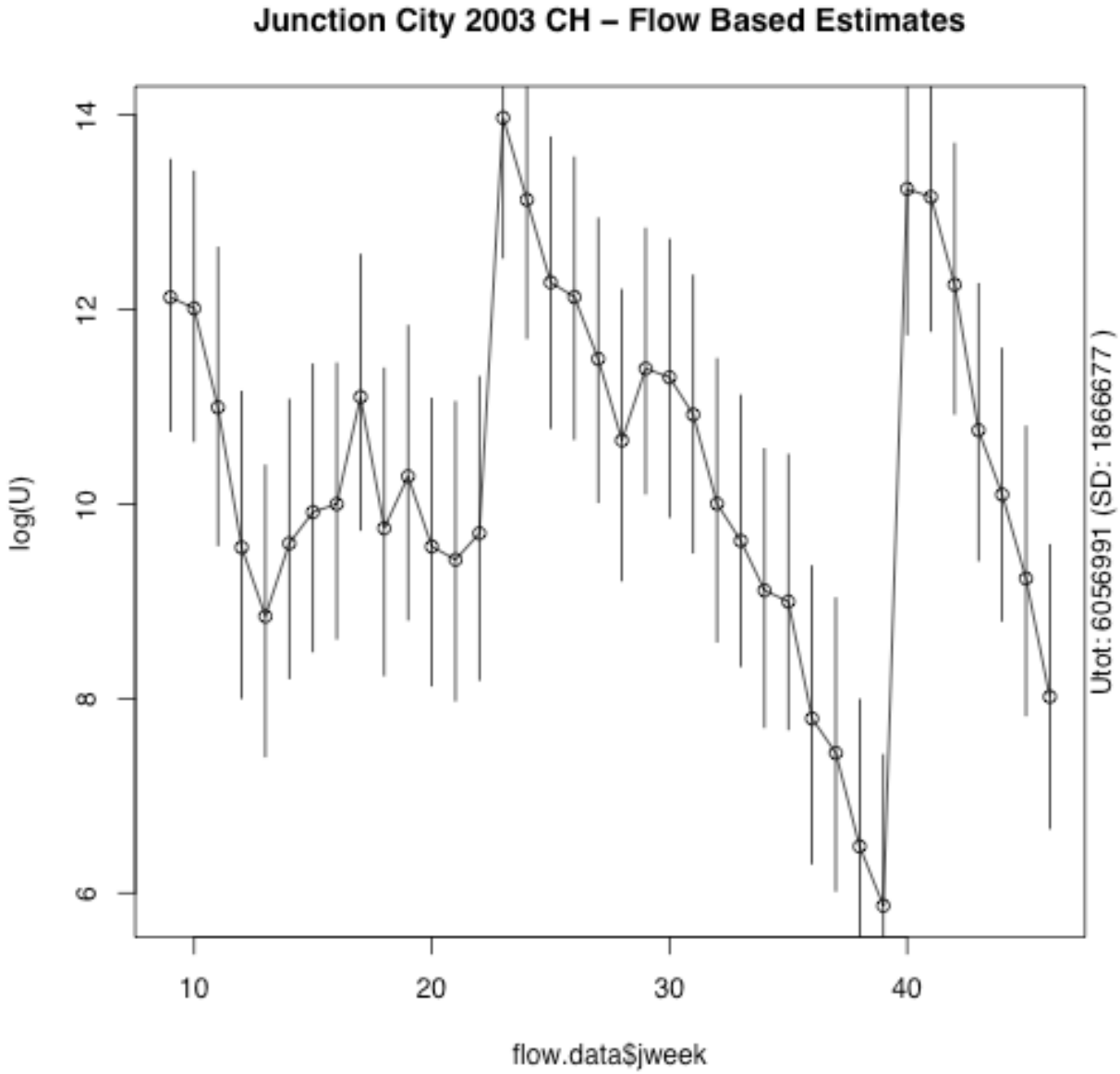


Figure 5.5. Estimated weekly based population values (and 95% confidence limits) using a flow based estimate for the Junction City 2003 Chinook salmon data.

A summary of the flow based estimates when applied to the existing datasets is found in Table 5.3. The flow-based estimates are consistent with the spline-based estimates but the precision of the flow-based estimates is considerably poorer than the precision from the spline-based estimates. The flow-based estimates do not "share" information across weekly strata so each individual weekly estimate has poor precision to account for the range in catchabilities at the flow for that week. When the weekly estimates are summed over weeks, the grand total is mainly driven by weeks with very large estimates – consequently, the precision of the grand total is driven by the poor precision in a few weeks of the run.

Table 5.3. Summary of pure flow-based estimates applied to existing datasets.

Site	Flow-based estimate (millions of fish)	Mark-recapture estimate from spline model without flow covariate (millions of fish)
Junction City 2002 Chinook salmon	5.5 (SD 1.7)	6.3 (SD .27)
Junction City 2003 Chinook salmon	6.0 (SD 1.8)	5.3 (SD .18)
Junction City 2004 Chinook salmon	7.5 (SD 2.0)	13.6 (SD .76)
Pear Tree 2003 Chinook salmon	2.4 (SD .56)	3.0 (SD .32)
Pear Tree 2004 Chinook salmon	12.0 (SD 8.0)	13.0 (SD 3.0)
Pear Tree 2005 Chinook salmon	9.1 (SD 7.9)	7.6 (SD 1.2)
Pear Tree 2006 Chinook salmon	3.1 (SD 1.7)	1.1 (SD .28)
Pear Tree 2007 Chinook salmon	12.6 (SD 7.3)	3.0 (SD .15)
Willow Creek 2002 Chinook salmon	2.9 (SD .67)	2.0 (SD .25)
Willow Creek 2003 Chinook salmon	.9 (SD .19)	1.1 (SD .23)
Willow Creek 2004 Chinook salmon	.9 (SD .23)	1.3 (SD .32)
Willow Creek 2005 Chinook salmon	4.4 (SD .96)	3.7 (SD .38)

This flow-based method used the flows as the HVT/USGS gauges which may not reflect the actual flows at the screw-traps. This uncertainty in the actual flow at the trap is a case of the error-in-variables problem where both the Y and X values are measured with error. In cases of simple regression, large error in the measured X values leads to attenuation, i.e. the observed slopes are pulled towards 0 (positive slopes are pulled down; negative slopes are pulled upwards). This could introduce a negative bias into the estimates of run size (predicted capture rates are too large which leads to a negative bias when the unmarked number of fish is expanded to the run size); however, given the generally poor precision seen in Table 5.3, this bias is expected to be small. A similar analysis was done using the actual flows measured at the traps. This data has many missing values and required much interpolation for the missing flow measurements. The results were comparable to those seen in Table 5.3 and are not reported.

5.5 Evaluation of using the sampled discharge to estimate abundance of outmigrating salmonids

Pinnix et al. (2007) report on using the fraction of the Trinity River discharge that is sampled in the screw-traps as an expansion factor for the number of unmarked fish to estimate the total number of outmigrating fish. This section will examine the method in more detail.

5.5.1 Estimating the fraction of the discharge that is sampled.

As outlined in Pinnix et al. (2007) measurements of flow at the screw-traps were taken using a torpedo meter. For the 8 ft screw-traps, six readings were taken (3 at a depth of .8 times the radius and 3 at a depth of .2 times the radius); for the 5 ft traps, three readings were taken at a depth of .6 times the radius. The number of revolutions was standardized to the number of revolutions per minute. The database of the standardized readings on a daily basis for each screw trap at the three sites was provided for this project (Section 2.2).

Not all sites had flow measurements taken in all years, and there are a considerable number of missing values within each year (Table 2.3;
Table 2.4). In particular, there were insufficient flow readings for Junction City 2002, Pear Tree 2003, and Pear Tree 2004.

A preliminary pair-wise plot (i.e. each of the 6 readings against each other) of the standardized readings is found in Figure 5.6.

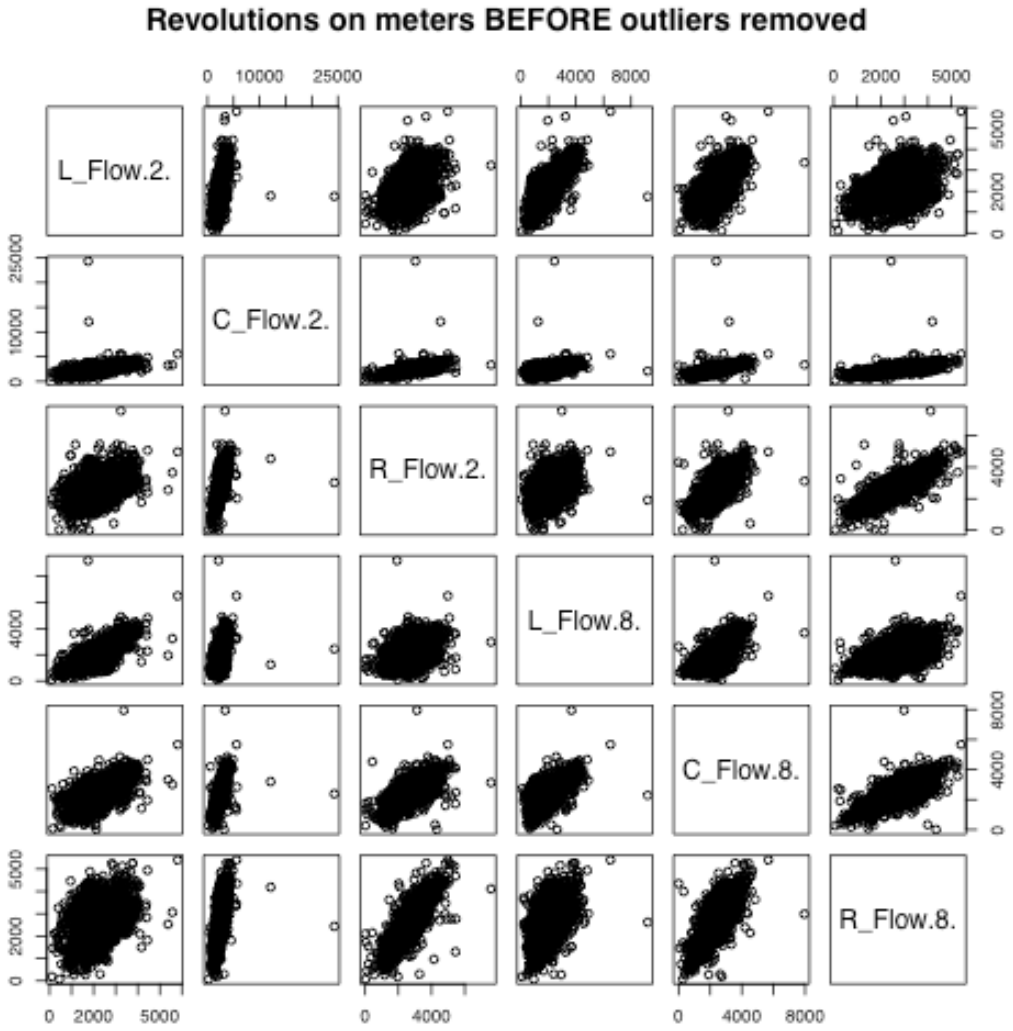


Figure 5.6. Pair-wise plot of flow meter readings combined over all sites and years.

Figure 5.6 shows that generally the flow meters are consistent among themselves (the flow is not expected to differ greatly across the different positions on the screw traps), but there are some anomalous data values. All flow readings below 1000 revolutions/minute or above 6000 revolutions/minute were removed as outliers. Figure 5.7 repeats the plot after removing these anomalous points.

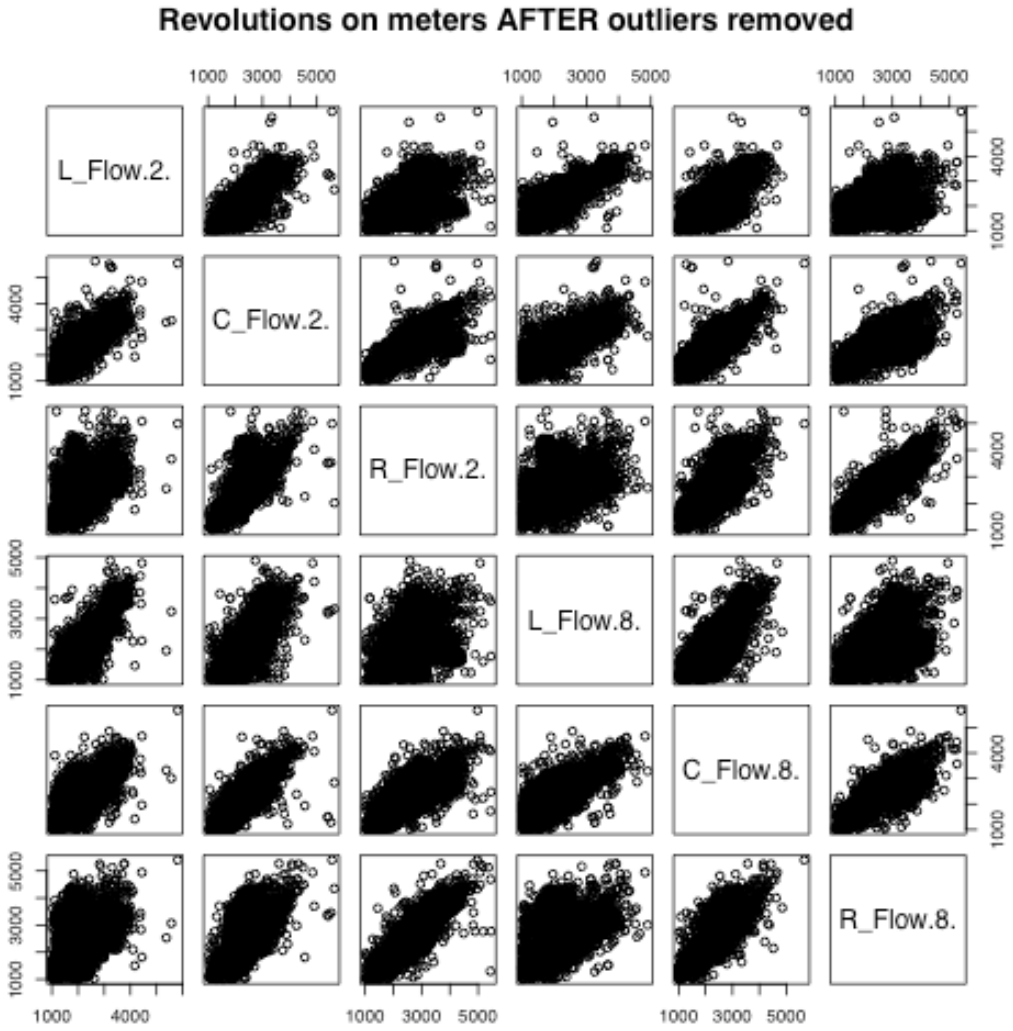


Figure 5.7. Pair-wise plot of flow meter readings combined over all sites and years after removing anomalous data values.

The remaining values for a particular site-day-trap combination (i.e. up to 6 measurements) were averaged to get the average revolutions/minute for the trap. The average revolutions was converted to an estimate of the stream velocity (ft/second) in site *s*, date *d*, and trap *t* using the following conversion (P. Petros, personal communication, March 12, 2009).

$$\overline{velocity}_{sdt} = \overline{rev}_{sdt} \times \frac{26873}{999999} \times \frac{100}{2.54(12)60}.$$

The second conversion factor converts meters/minute to feet/second.]

In some cases at the Pear Tree site, both an 8 ft and 5 ft screw trap were operating simultaneously, but readings on the flow were not obtained from both traps on the same day. A plot of the estimated velocity from the two traps showed a generally positive relationship (Figure 5.8) with the velocity at the 8 ft traps being and average of 1.13 times the velocity at the 5 ft traps.

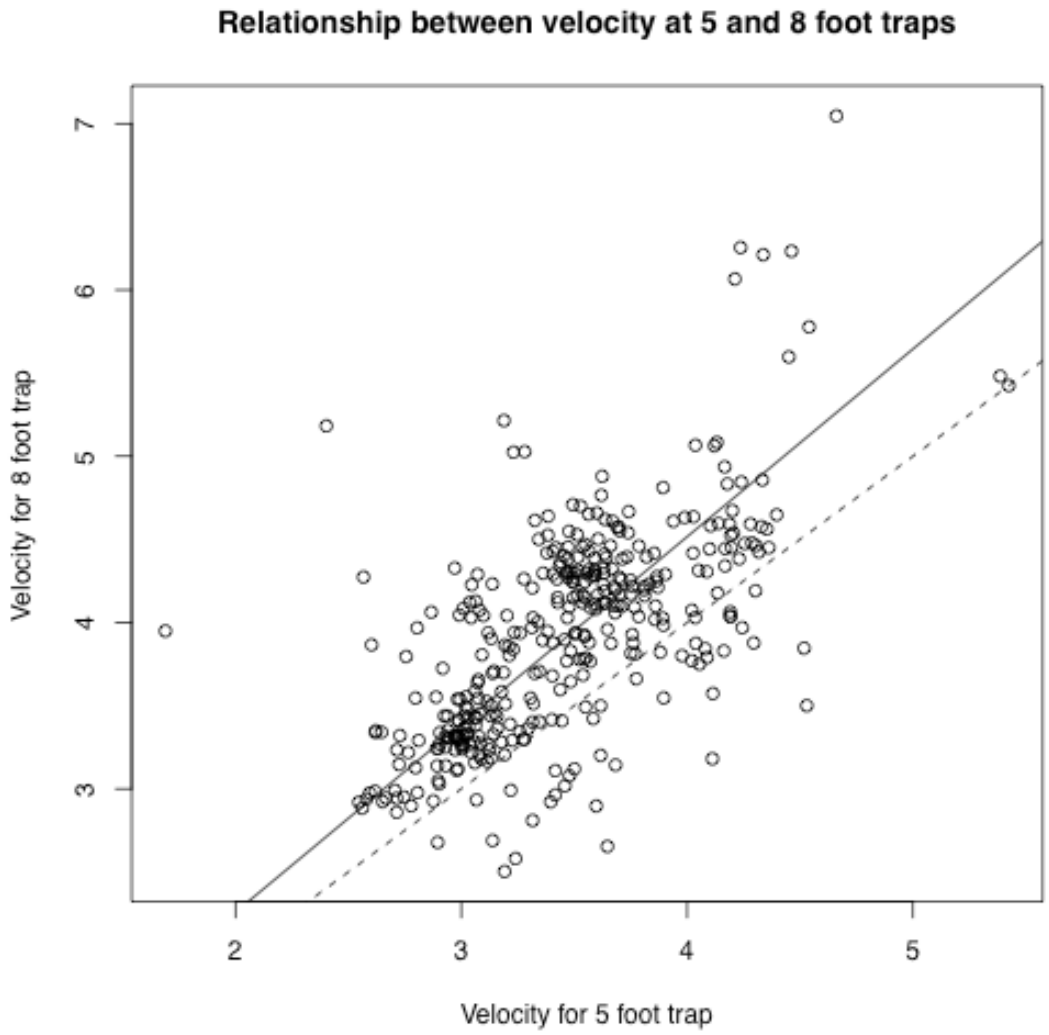


Figure 5.8. Relationship between stream velocity between the 8 ft and 5 ft traps at the Pear Tree site in 2007. Dashed line is the relationship Y=X; solid line is the relationship $Vel_{8\,ft} = 1.13 Vel_{5\,ft}$.

This relationship was used to interpolate in both directions when one of the readings was missing but the other was present. This imputation need to be done only at the Pear Tree site.

The volume of water sampled (cfs) by each trap *t* on date *d* at site *s* was estimated by:

$$DisSampled_{sdt} = \overline{Vel}_{sdt} \times .5\pi r_t^2$$

i.e. by assuming that the cross-section in the river was ½ of the area of the screw-trap's face. The total volume of discharge sampled at site *s* and date *d* was found by summing over the traps present and operating that date.

$$DailyDisSampled_{sd\bullet} = \sum_t \overline{Vel}_{sdt} \times .5\pi r_t^2$$

The number of traps operating may vary by site and date, and not all days in a year at a site have daily discharge samples.

The total river discharge (cfs) was obtained from the database of the official USGS Gauge Stations as outlined in
Table 2.4. This database is complete with readings from every site and date.

The database available did not have the number of hours each trap was operating each day, so it was implicitly assumed that all traps were operating 24 hours/day.

A plot of the daily discharge sampled vs. the daily total river flows are found in Figure 5.9.

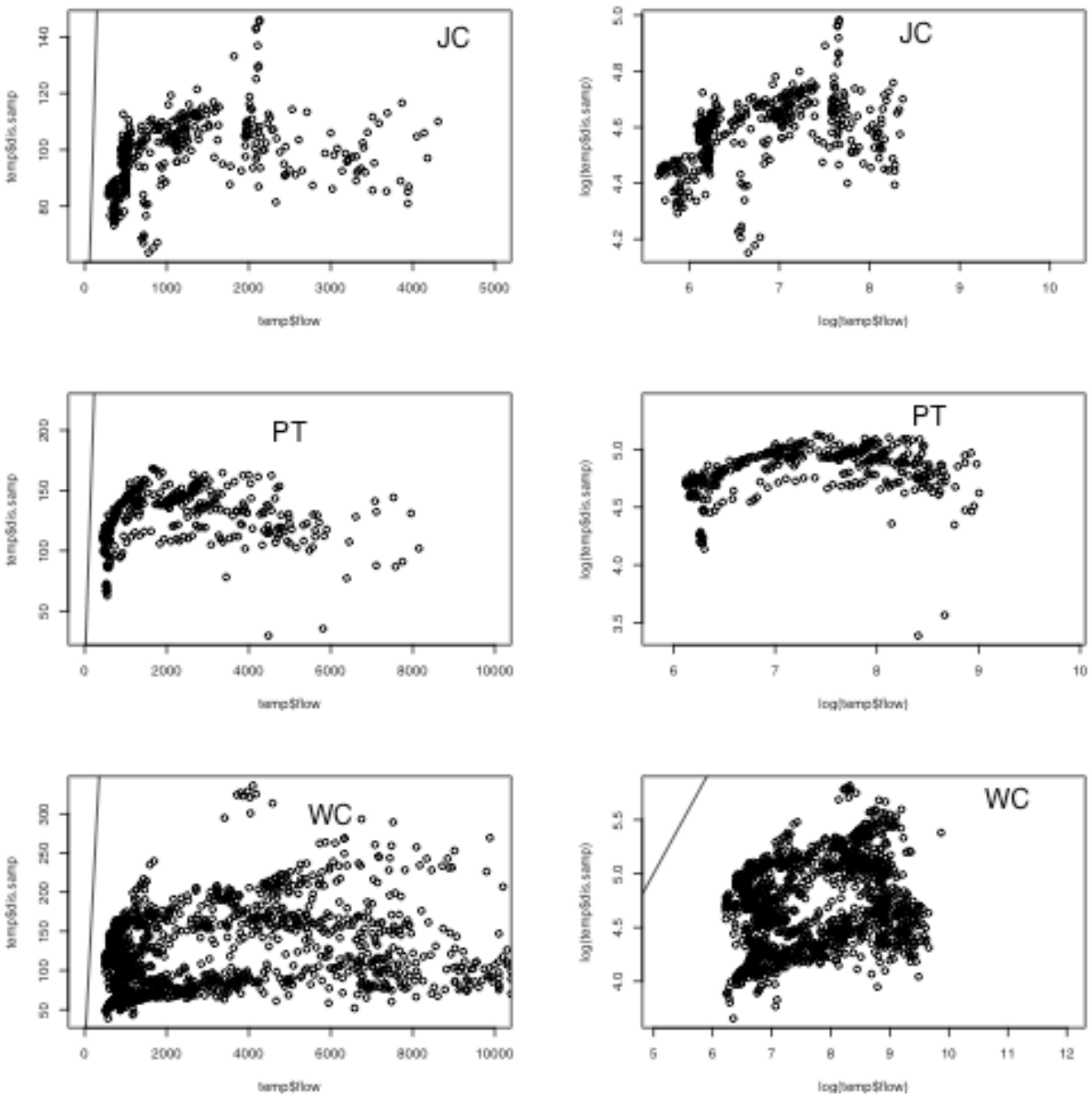


Figure 5.9. Plots of the daily discharge sampled (cfs) vs the daily total flow (cfs). The left column are in original units (cfs); the right column are on the log-scale. The top row is Junction City; the middle row is Pear Tree; the bottom row is Willow Creek. Pooled over all years and days.

The line Y=X can occasionally be seen. The values are expressed as cfs rather than volume over 24 hours.

Generally speaking, the discharge sampled by the traps is independent of flow, i.e. is consistent regardless of the total flow of the row. This is likely an artifact of the way the traps must be placed – it is too dangerous to place them in areas of heavy flow and they don't function well in areas of very low flow.

The total river volume passing the traps at site *s* in year *y* and Julian week *j* was found as:

$$TotalRiverVol_{syj} = \sum_{\text{days in julian week}} RiverVol_{syd} \ .$$

It is not necessary to scale up the cfs to total cubic feet by multiplying by the number of seconds in a week as this scaling factor will cancel when the estimated sampling fraction is found later. The total weekly discharge that is sampled was found as:

$$WeeklyDisSampled_{syj} = \sum_{\text{days in julian week with data}} TotalDisSampled_{sd}.$$

i.e. a simple sum over the trap operating on each day in that week. Again, no information was available in the database on the number of hours the traps were operating during a day, so it will be assumed that all traps were operating for 24 hours/day. This implicitly assumes that the average discharge sampled is consistent over the week which appears to be a reasonable assumption given Figure 5.9. A plot of the weekly measurement is found in Figure 5.10.

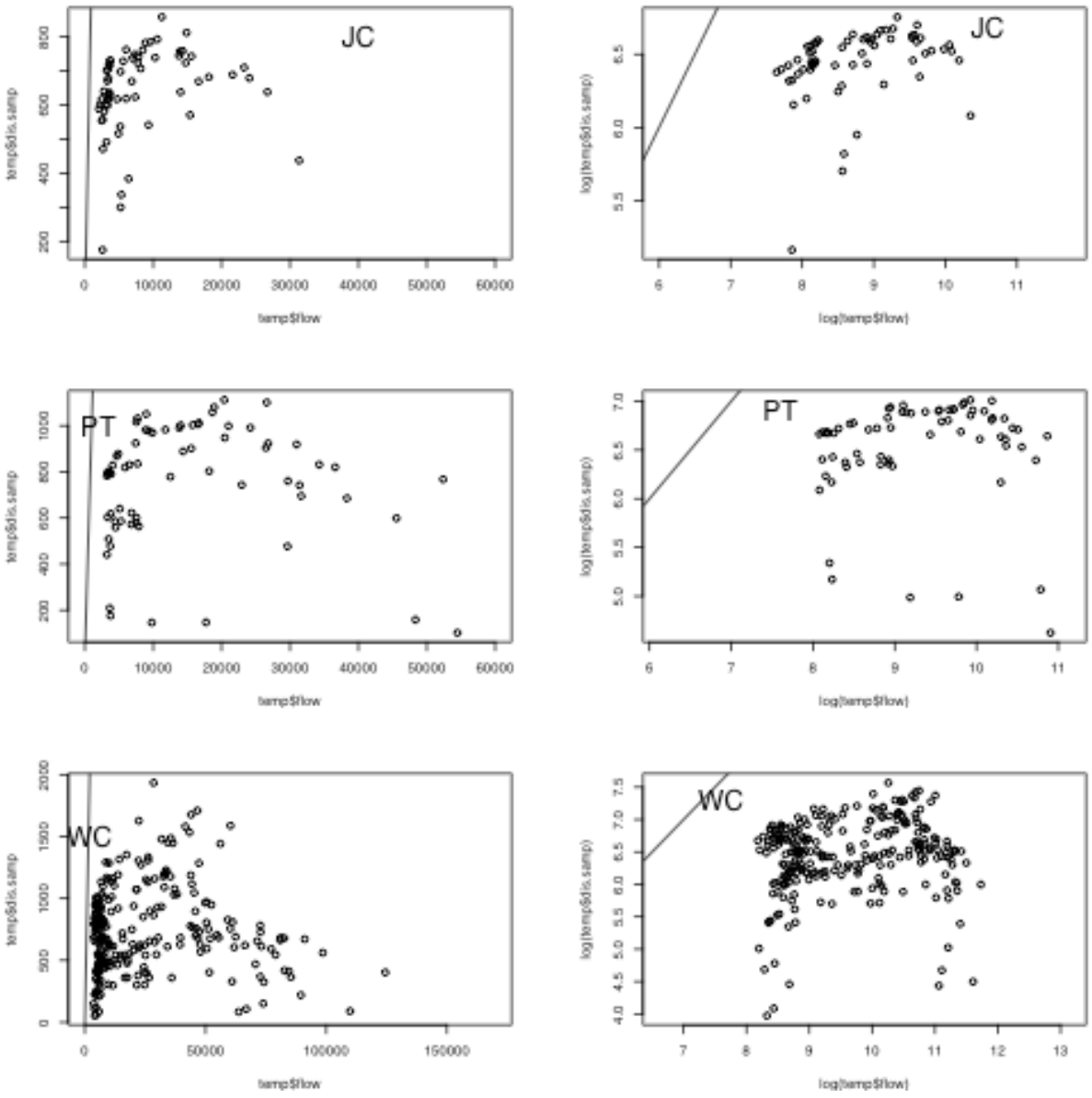


Figure 5.10. Plots of the weekly discharge sampled (cfs) vs the weekly total flow (cfs). The left column are in original units (cfs); the right column are on the log-scale. The top row is Junction City; the middle row is Pear Tree; the bottom row is Willow Creek. Pooled over all years and Julian weeks. The line Y=X can occasionally be seen.

A similar pattern is found, i.e. the discharge sampled is roughly constant over a wide range of flows.

Finally, the fraction of the river flow sampled at site *s* in year *y* for Julian week *j* was found as the ratio of the water flowing through the trap to the total river volume for that week.

$$\hat{p}_{syj}^{flow} = \frac{WeeklyDisSampled_{syj}}{TotalRiverVol_{syj}}$$

5.5.2 Comparison of the flow-based sampling rate and the mark-recapture estimates.

A plot of the empirical relationship between the flow-based sampling rate and the mark recapture estimates for each Julian week is found in Figure 5.11.

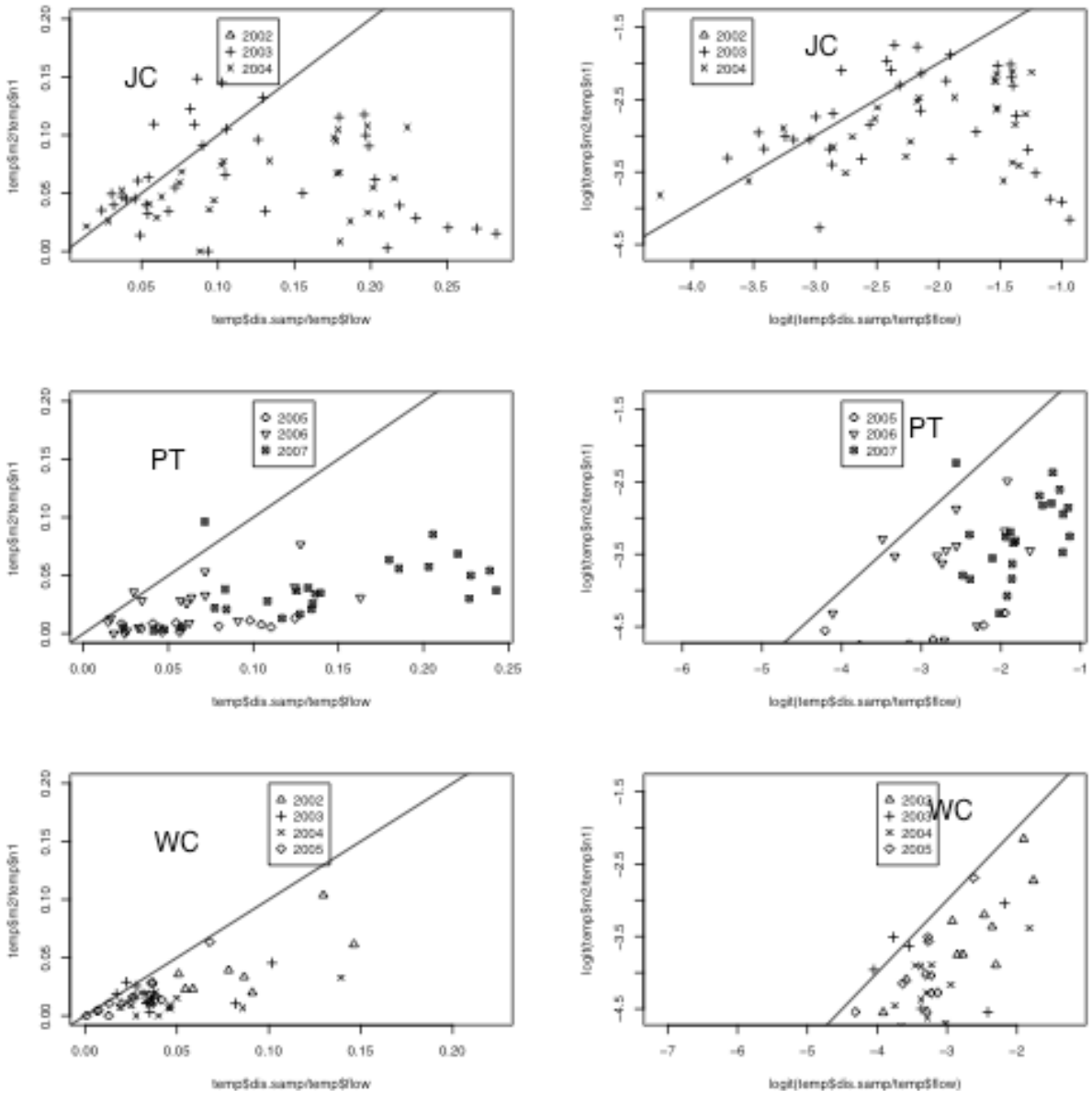


Figure 5.11. Plots of the empirical mark-recapture estimates of the capture-probability vs the flow-based estimates of the discharge sampled. Only those Julian weeks where at least 20 fish were released are used. The line Y=X is shown. The left column is the original scale

(probability); the right column is on the logit scale. The first row is for Junction City; the second row for Pear Tree; the third row for Willow Creek. Different plotting symbols are used for each year of the study (JC 2003-2004; PT 2005-2007; WC 2003-2006).

At the Junction City site, there appears to be strong relationship between the two estimates of capture efficiency of the traps until fraction of water sampled exceeds about 15%. This usually occurs are very low flow conditions. The relationship at Pear Tree is much tighter, but generally speaking the estimated capture rates measured by flow sampled is much smaller than that measured by the mark-recapture experiments. The relationship at Willow Creek appears to be consistent over time again with the flow based estimates of capture efficiency consistently lower than the mark-recapture based estimates.

5.5.3 Using the flow-based estimates of capture efficiency to estimate the number of outmigrating fish.

The flow-based estimates of capture-efficiency were then used to expand the number of unmarked fish captured in the trap for that Julian week to obtain an estimate of the total number of unmarked fish (wild+hatchery) passing the trap at site *s* in year *y* in Julian week *j*:

$$\hat{U}_{syj}^{W+H,\,flow} = \frac{u_{syj}}{\hat{p}_{syj}^{flow}}$$

Flow-based estimates of only the wild-population were obtained by expanding the number of unclipped fish recovered adjusted for the unclipped portion of the hatchery fish:

$$\hat{U}_{syj}^{W,\,flow} = \frac{\max\left(0, u_{syj}^{Non-clipped} - u_{syj}^{Ad-clipped}\frac{1-c_{sy}}{c_{sy}}\right)}{\hat{p}_{syj}^{flow}}$$

where c_{sy} is the fraction of the hatchery fish that are adipose-fin clipped in site *s* and year *y*.

5.5.4 Comparison of flow-based and spline estimates of outmigrant abundance and run timing

Plots of the estimated weekly-flow based estimates of the wild+hatchery combined and the wild only population size were compared to the corresponding estimates based on the spline methods in Figure 5.12, Figure 5.13 and Figure 5.14.

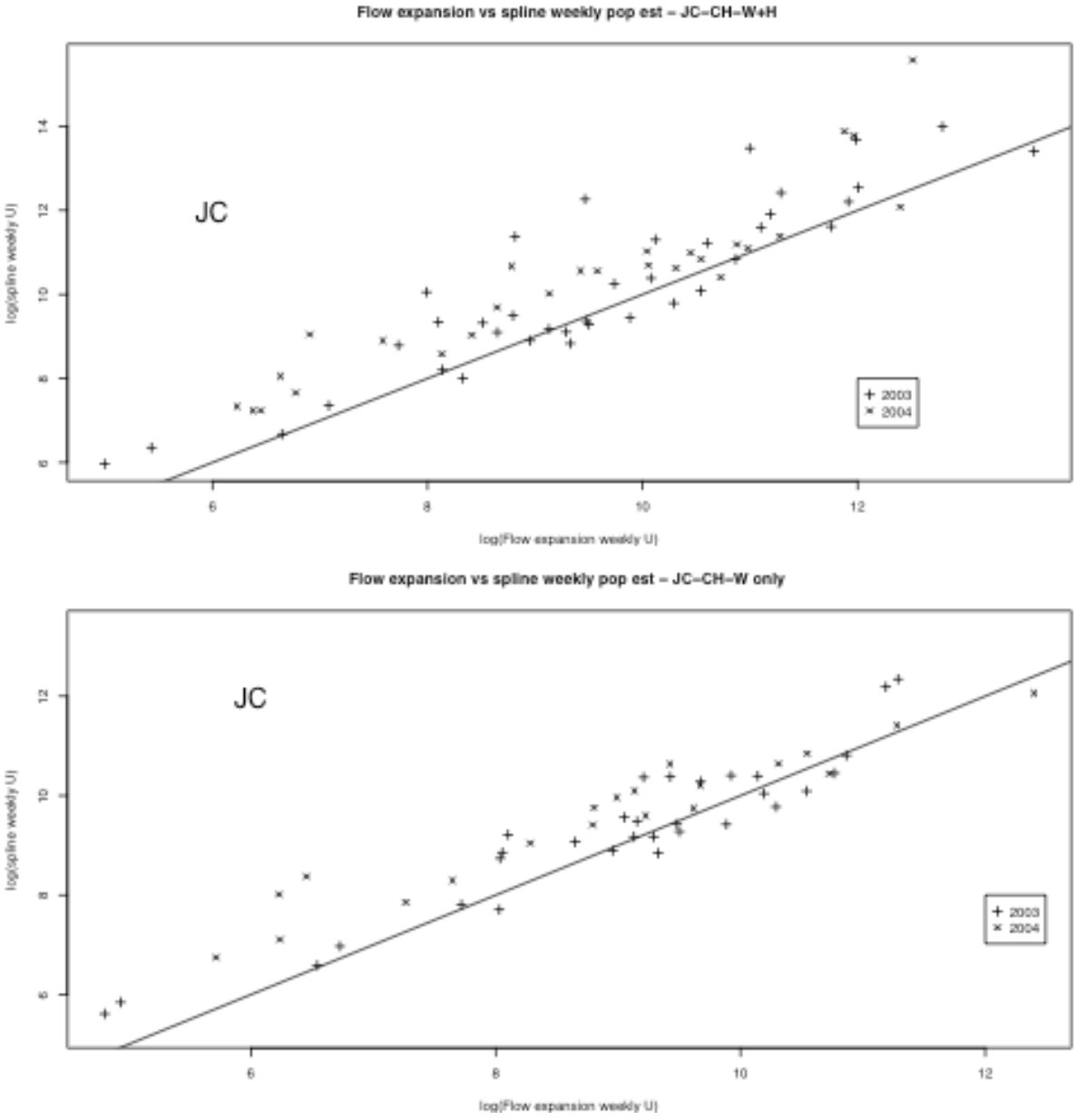


Figure 5.12. Junction City Site: Spline-based estimate of weekly numbers of unmarked fish vs. flow-based estimates for weeks when both estimates are available Top panel is wild+hatchery combined; bottom panel is wild only.

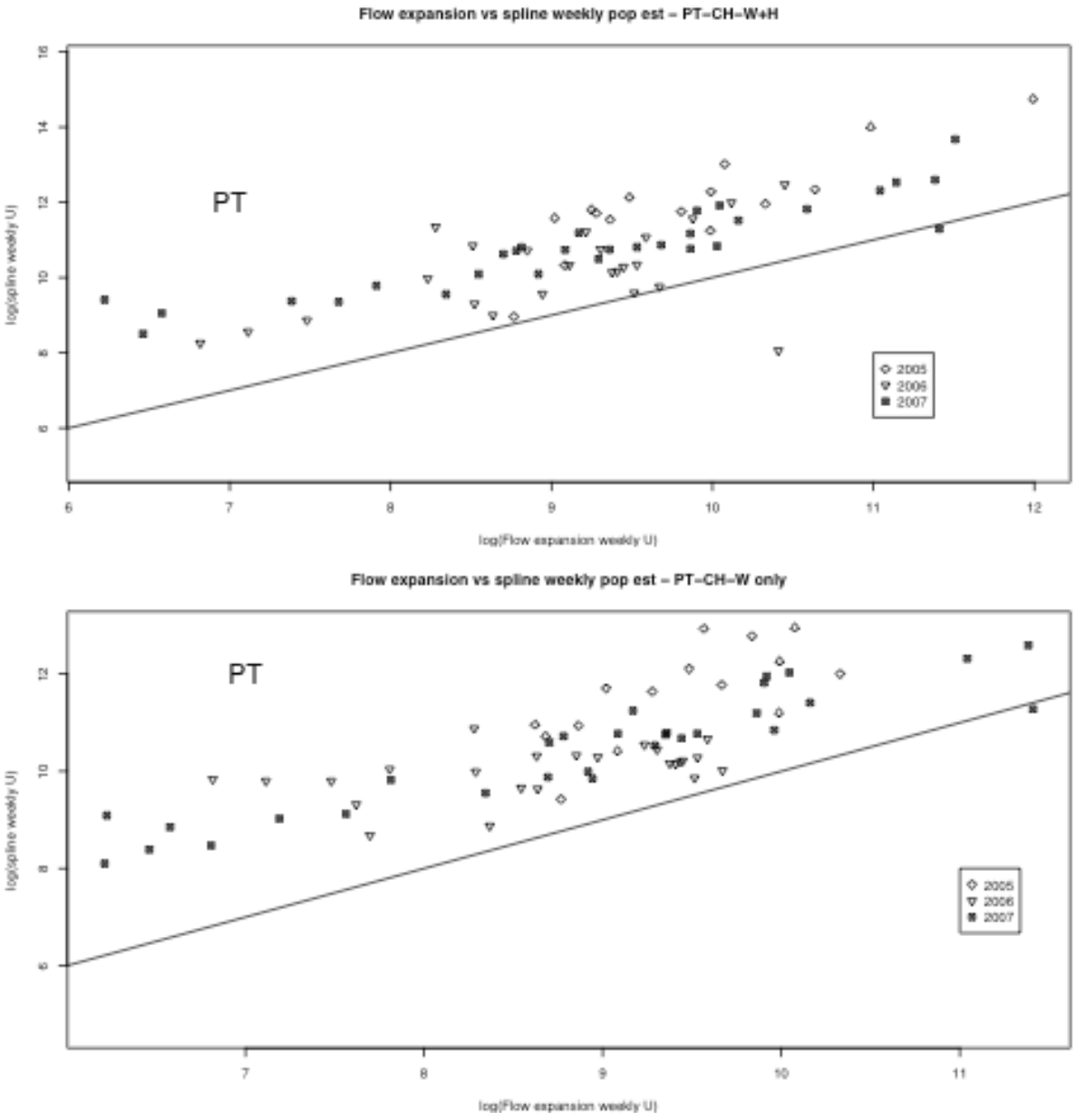


Figure 5.13. Pear Tree Site: Spline-based estimate of weekly numbers of unmarked fish vs. flow-based estimates for weeks when both estimates are available. Top panel is wild+hatchery combined; bottom panel is wild only.

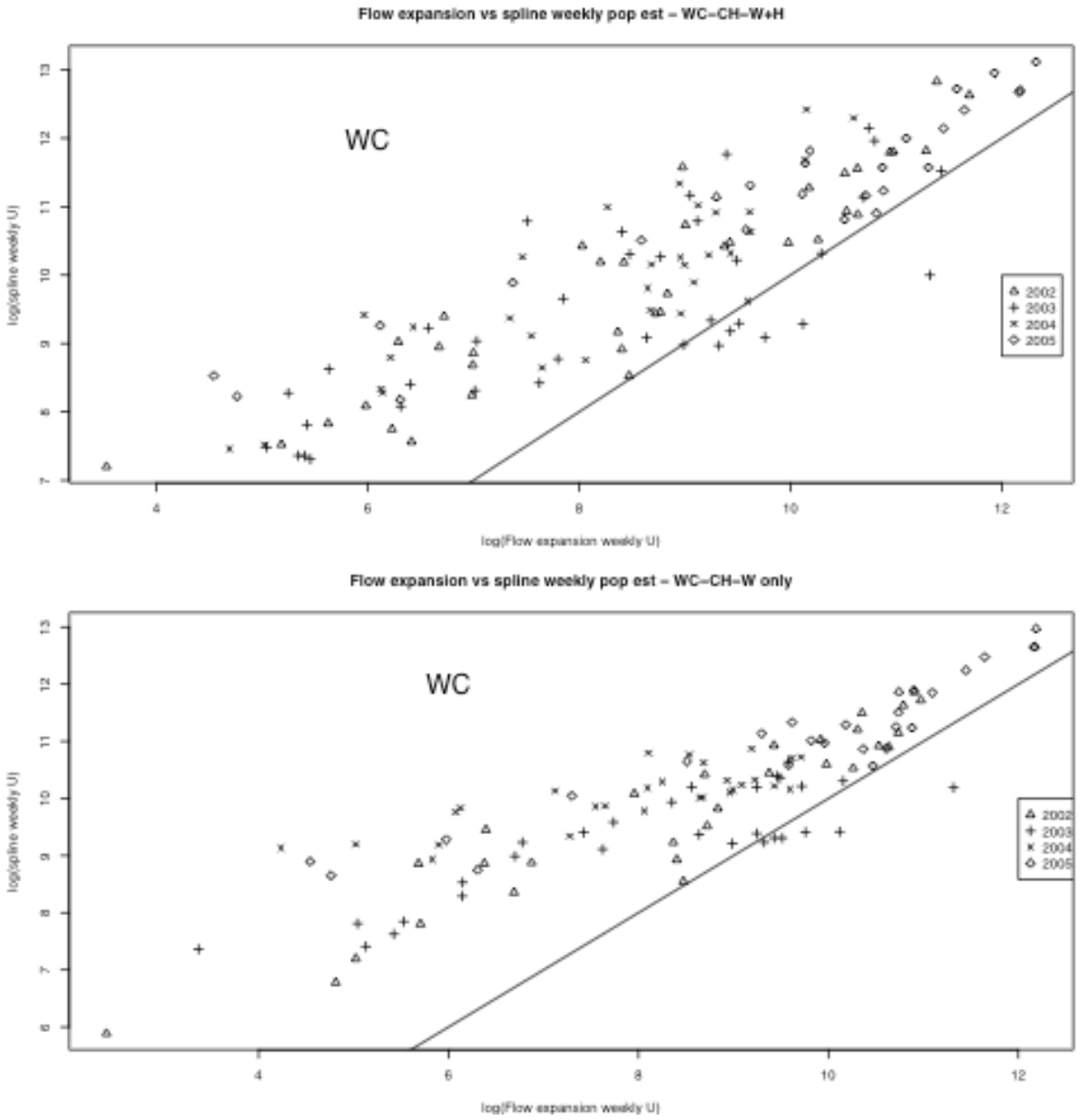
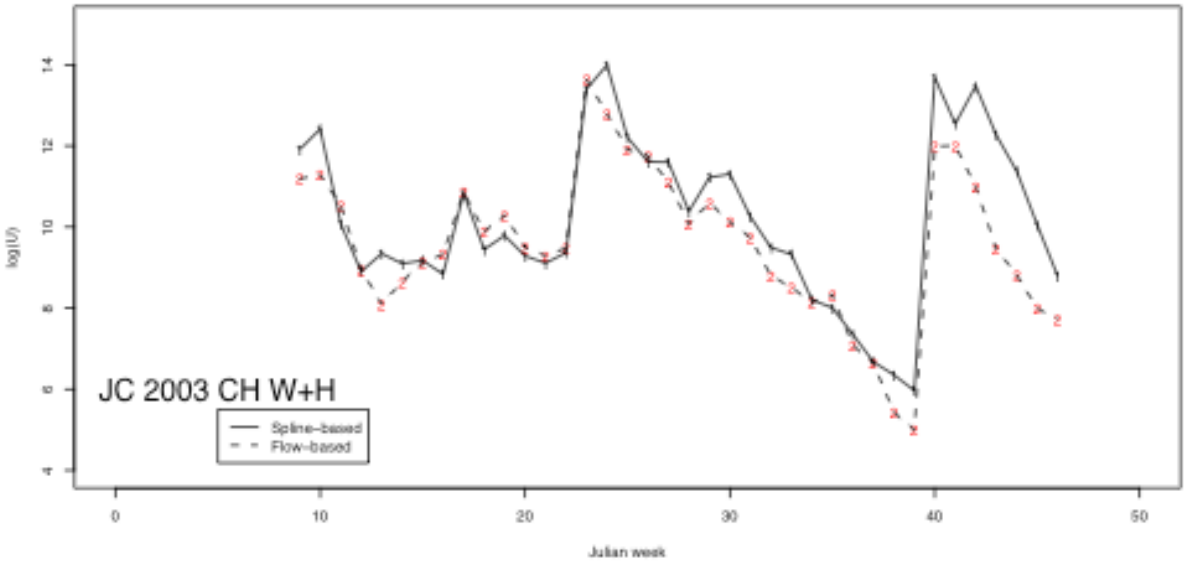
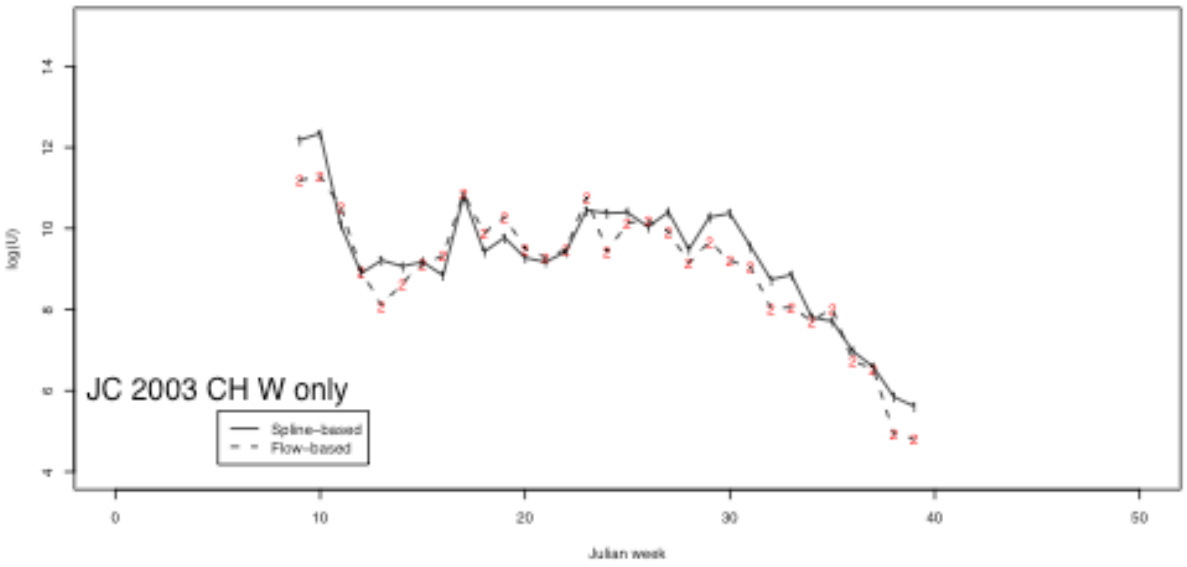


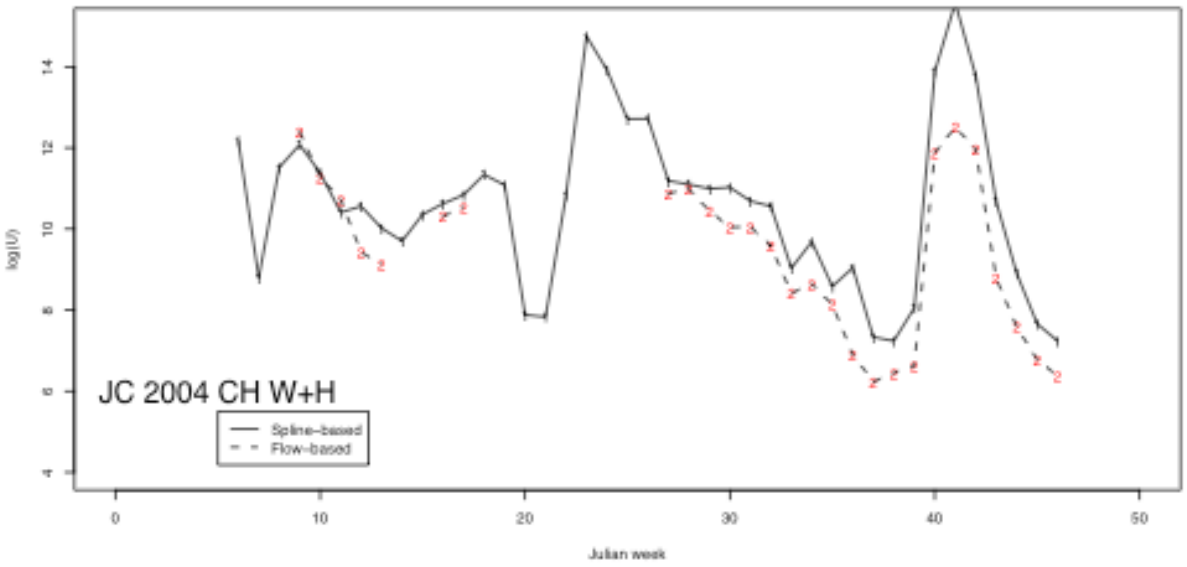
Figure 5.14. Willow Creek Site: Spline-based estimate of weekly numbers of unmarked fish vs. flow-based estimates for weeks when both estimates are available. Top panel is wild+hatchery combined; bottom panel is wild only.

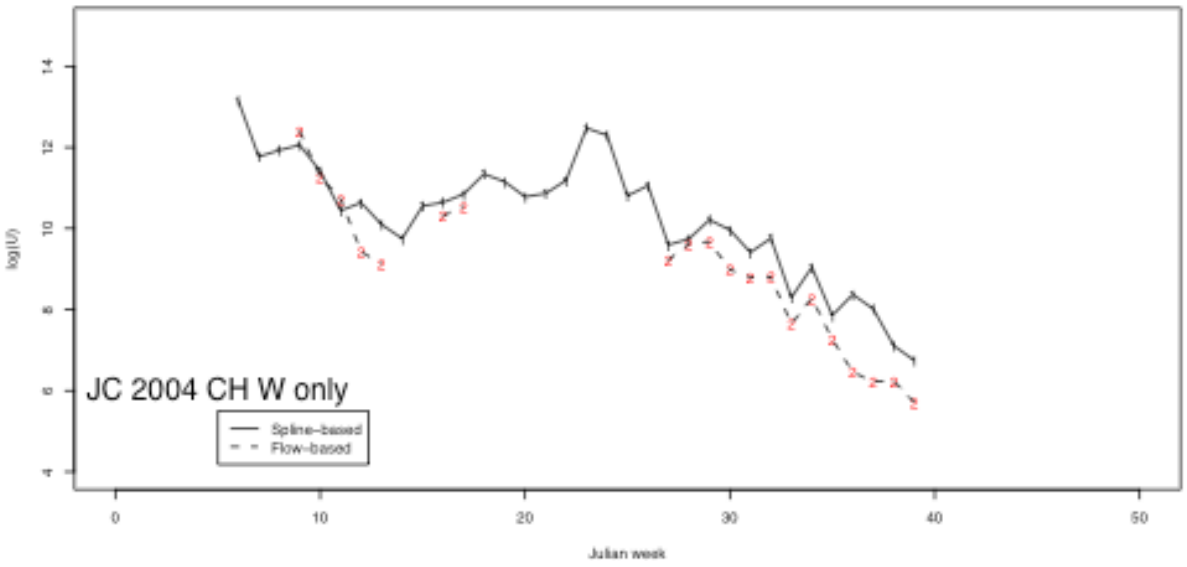
The Junction City estimates show a small but consistent underestimate of the total population size; the Pear Tree estimates show a larger, but consistent underestimate of the total population size; the Willow Creek estimates are not as consistent with some evidence that the underestimation become more severe at smaller population sizes.

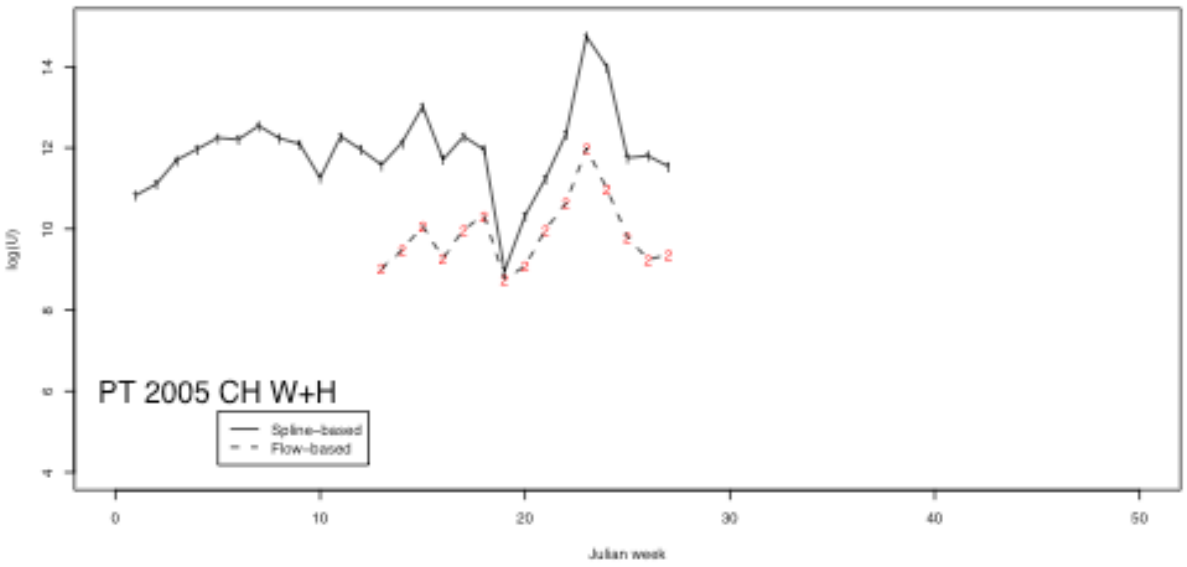
The series of panels in Figure 5.15 below show the same comparisons in time-series plots.

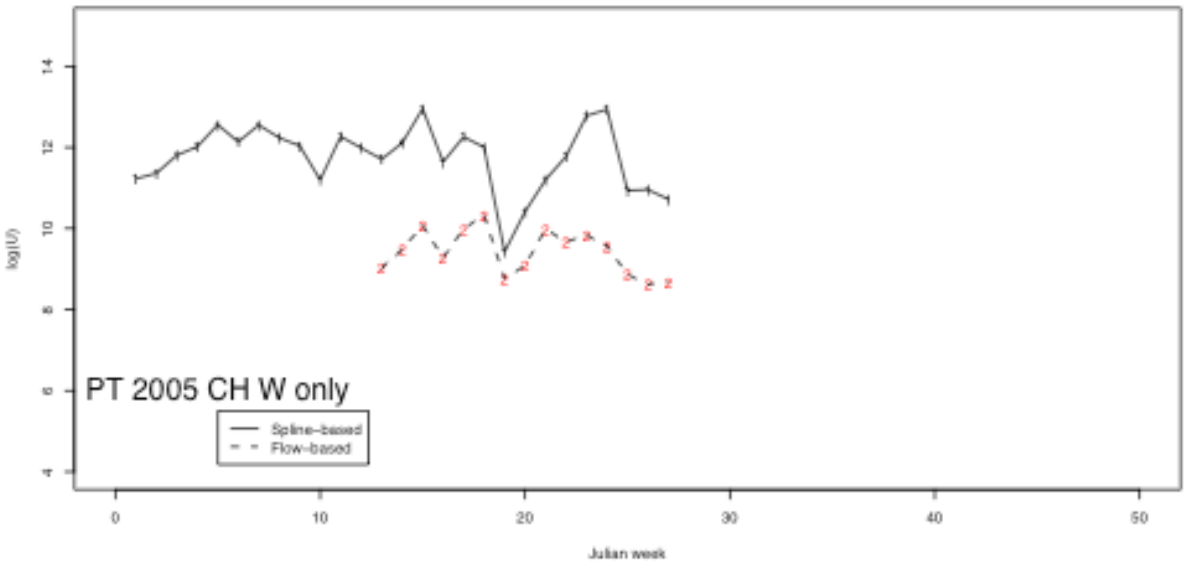

JC 2003 CH W+H
Spline-based
Flow-based
log(U)
Julian week

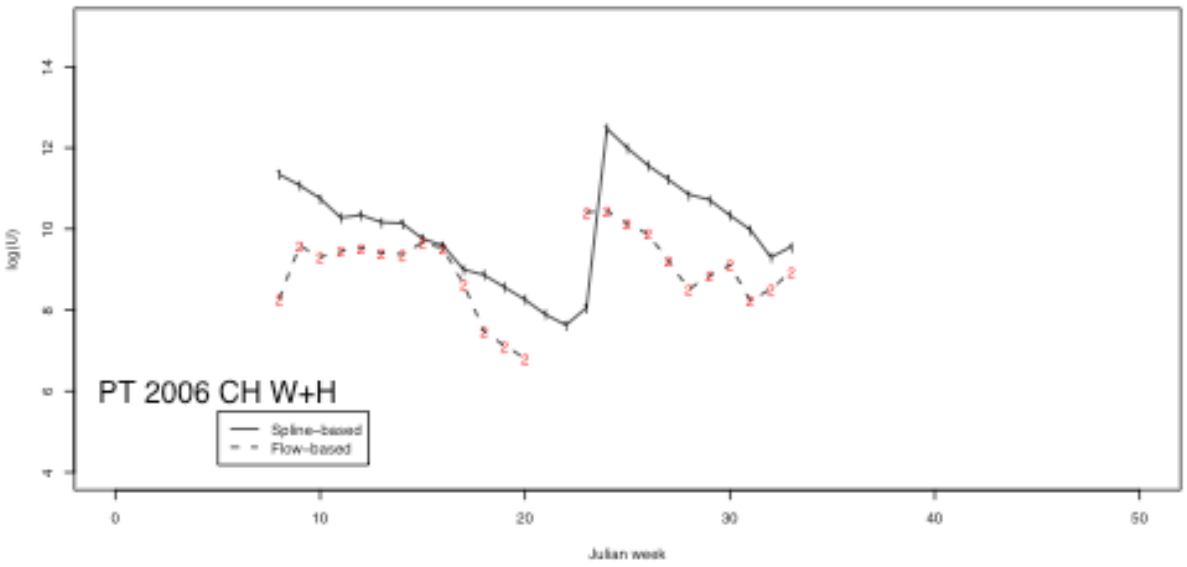

JC 2003 CH W only
Spline-based
Flow-based
log(U)
Julian week

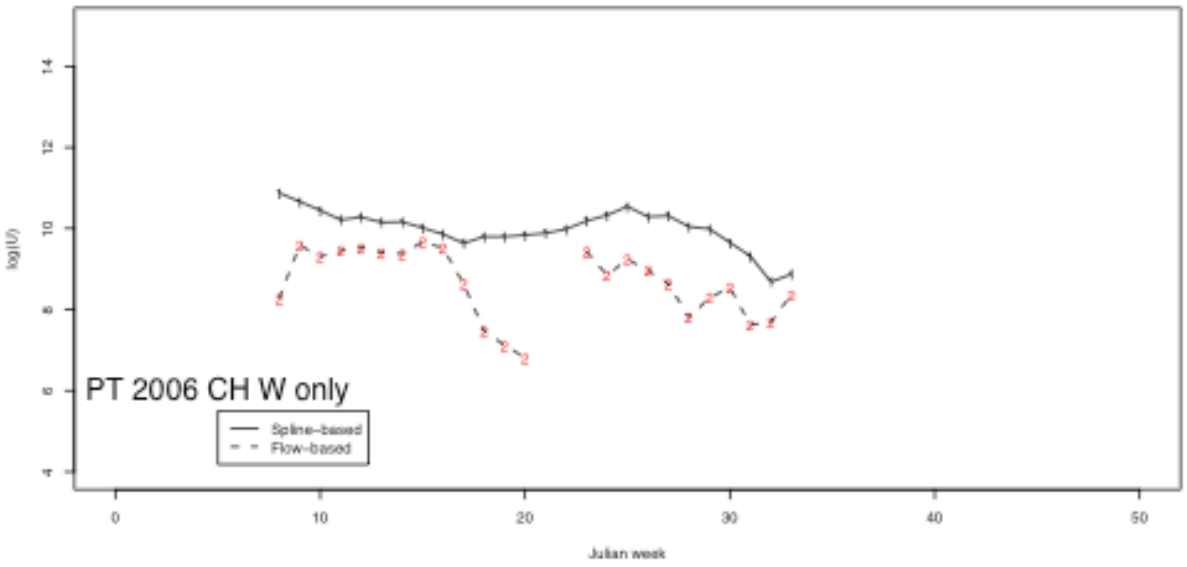

JC 2004 CH W+H
Spline-based
Flow-based
log(U)
Julian week

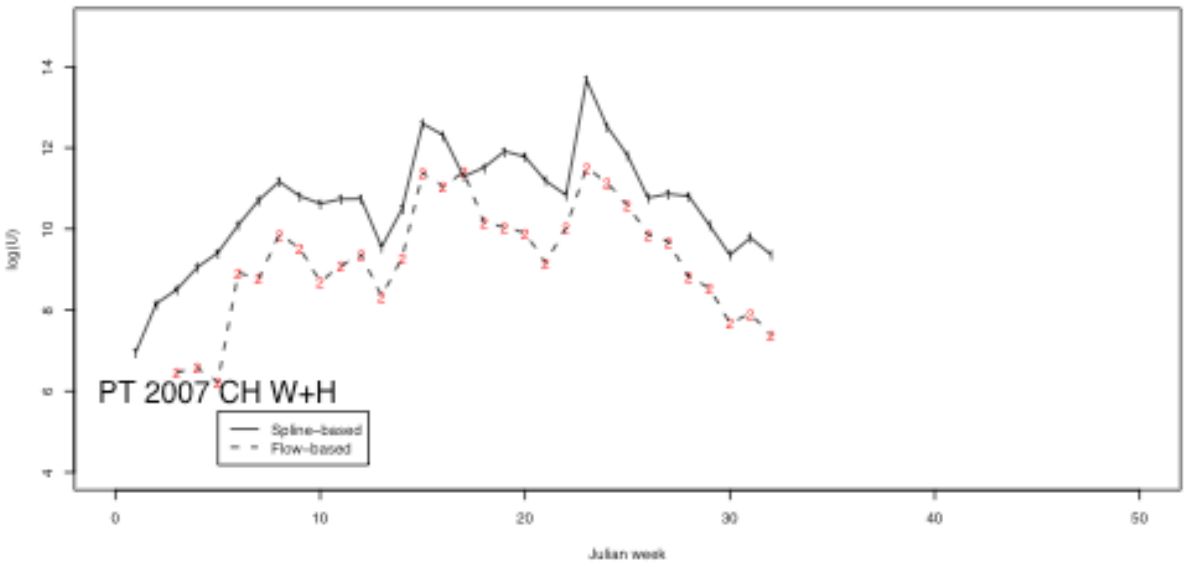

JC 2004 CH W only
Spline-based
Flow-based
log(U)
Julian week

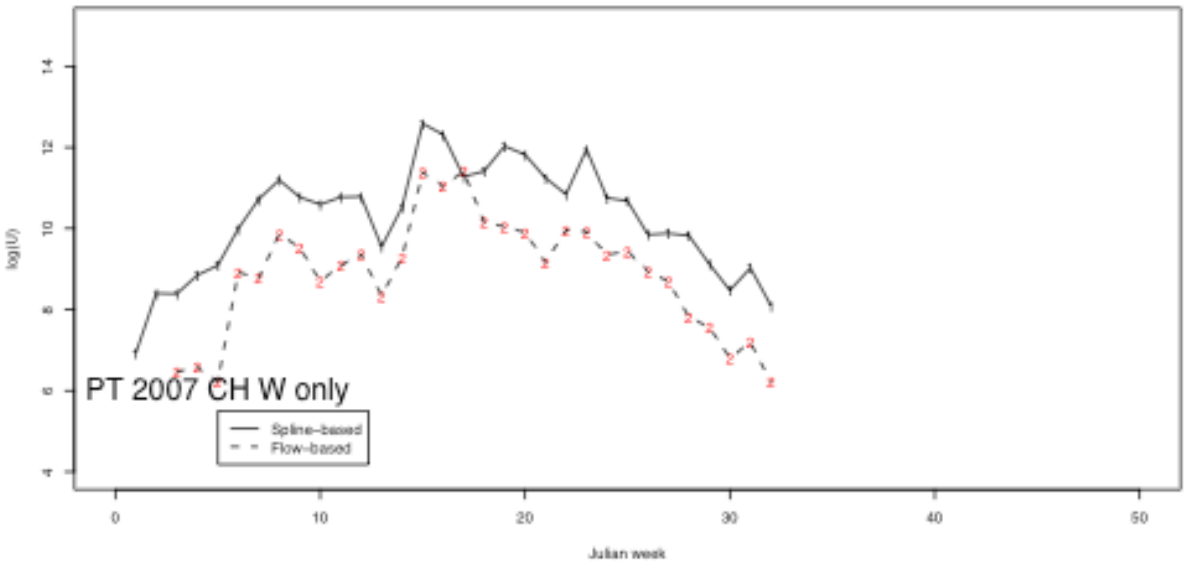

log(U)
PT 2005 CH W+H
Spline-based
Flow-based
Julian week

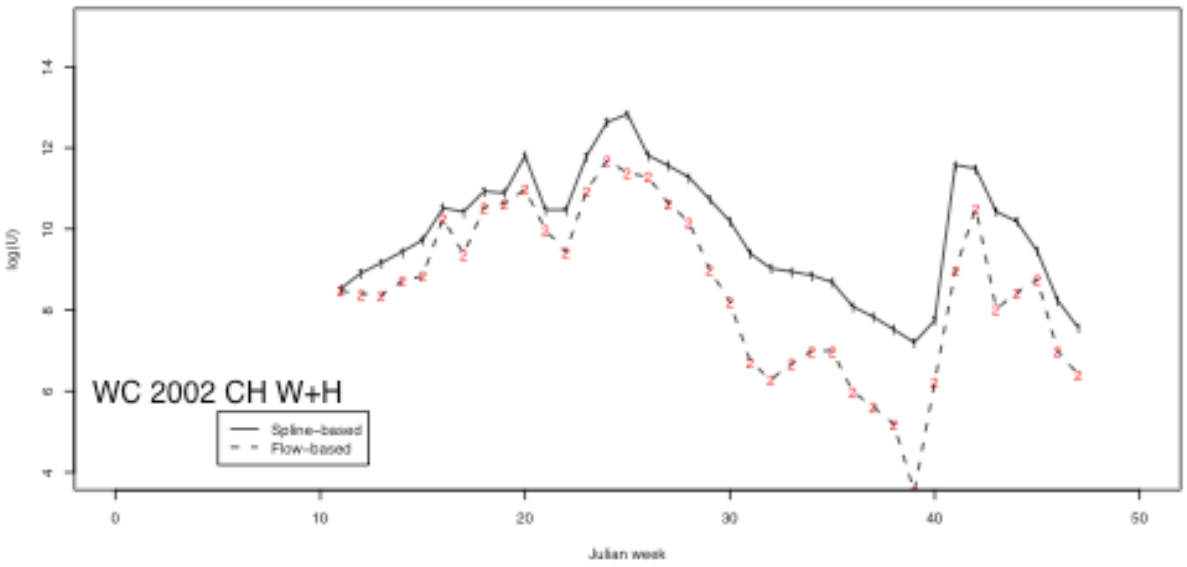

log(U)
PT 2005 CH W only
Spline-based
Flow-based
Julian week

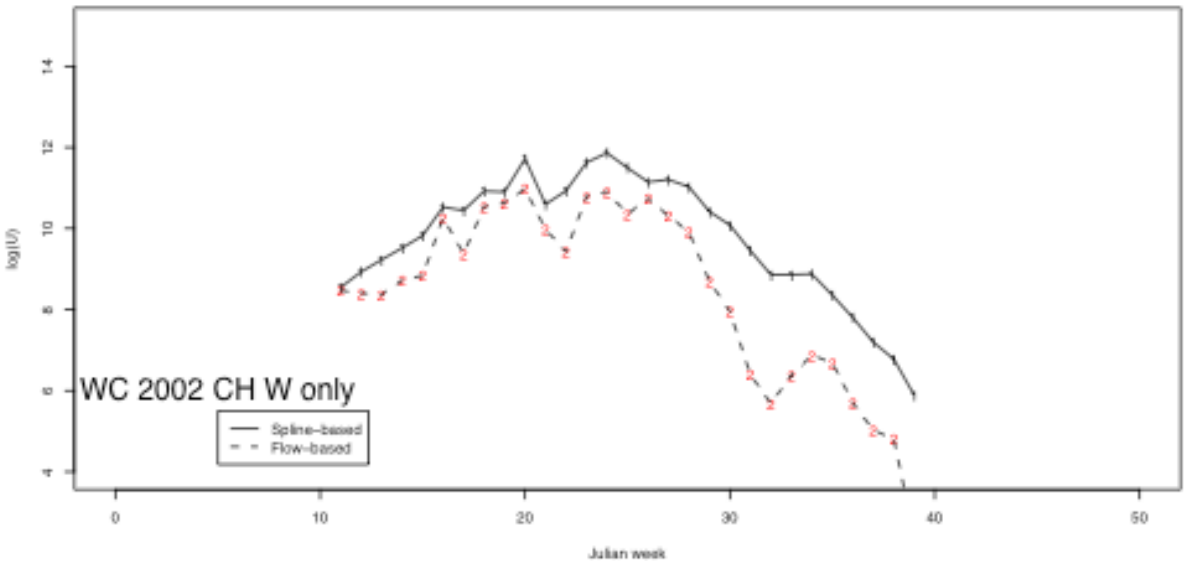

log(U)
14
12
10
8
6
4
0
10
20
30
40
50
Julian week
PT 2006 CH W+H
Spline-based
Flow-based

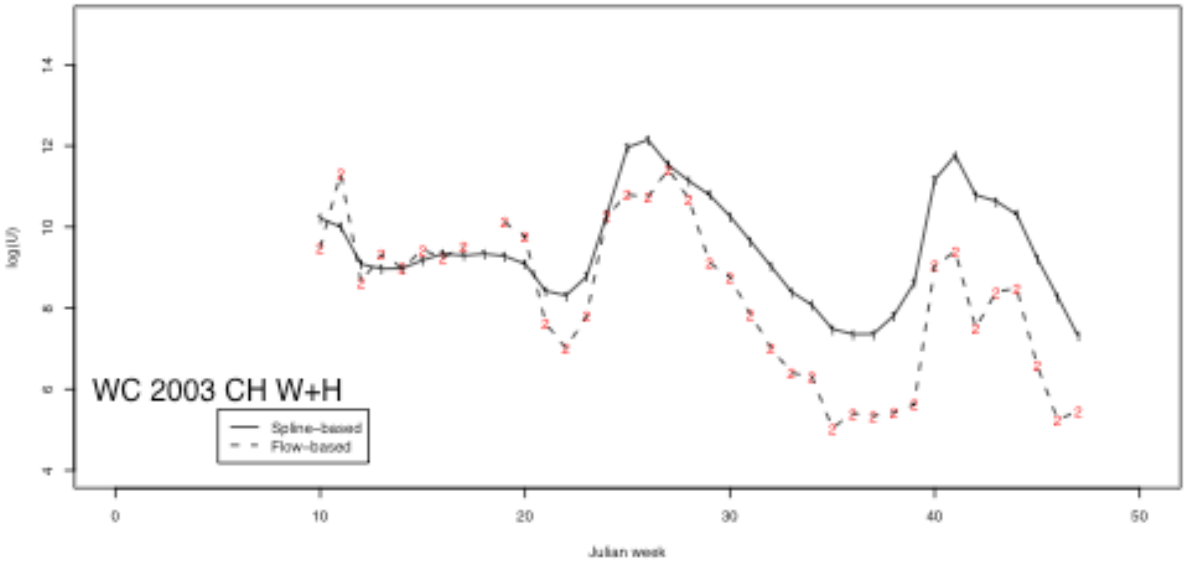

log(U)
14
12
10
8
6
4
0
10
20
30
40
50
Julian week
PT 2006 CH W only
Spline-based
Flow-based

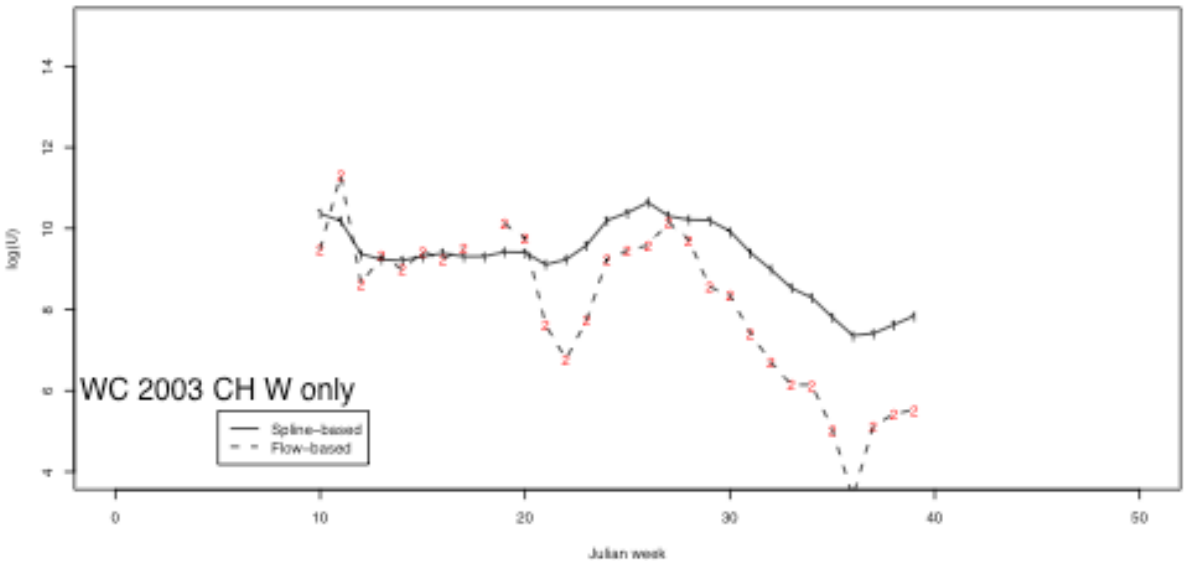

PT 2007 CH W+H
Spline-based
Flow-based
log(U)
Julian week

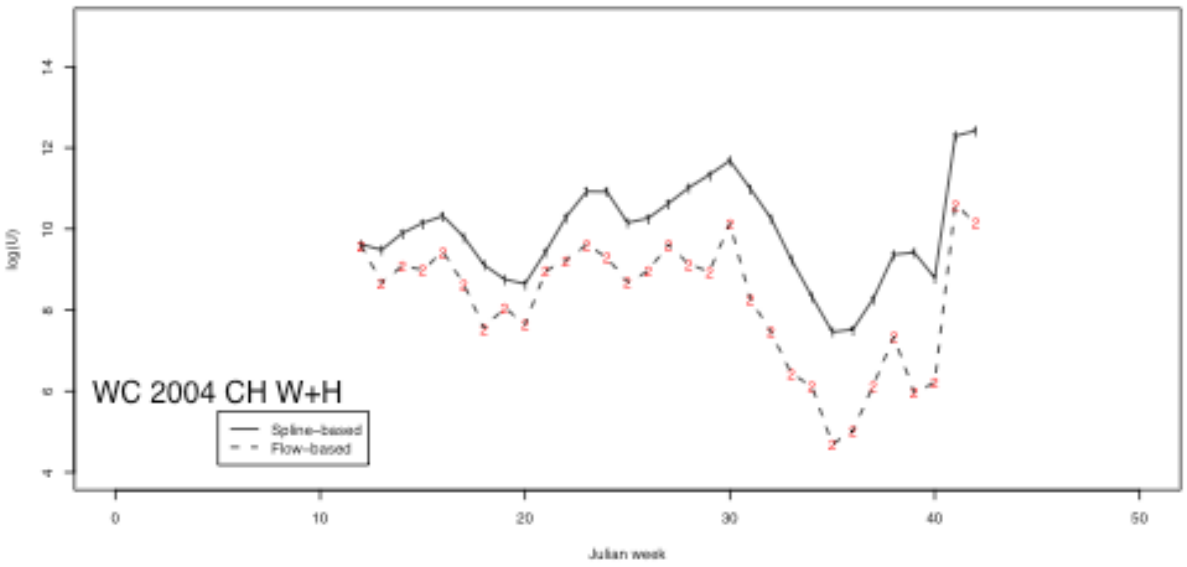

PT 2007 CH W only
Spline-based
Flow-based
log(U)
Julian week

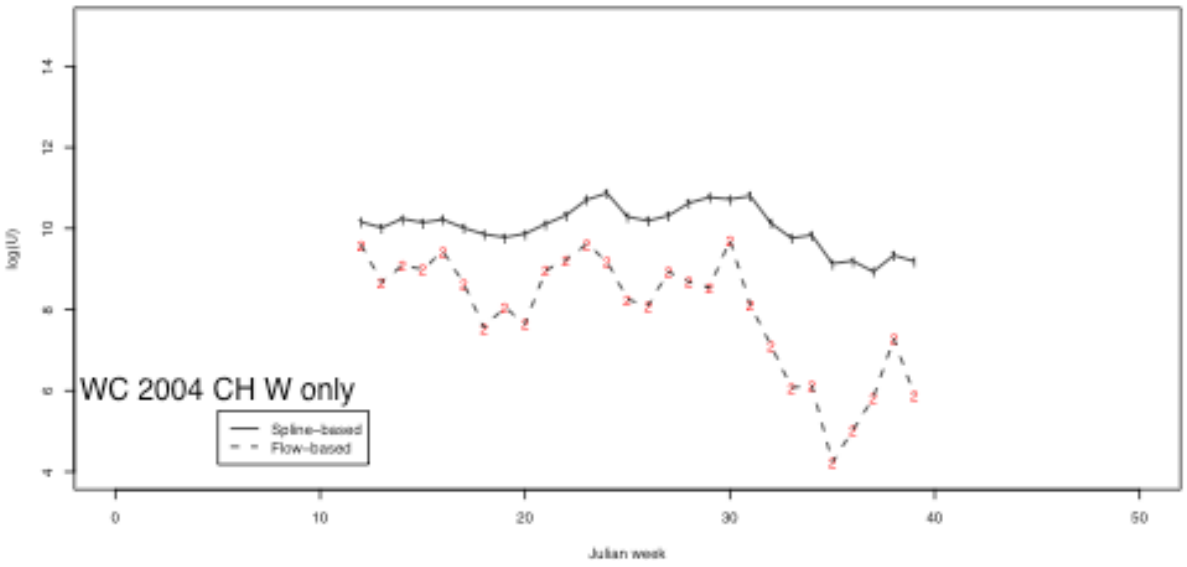

WC 2002 CH W+H
Spline-based
Flow-based
log(U)
Julian week


WC 2002 CH W only
Spline-based
Flow-based
log(U)
Julian week


WC 2003 CH W+H
Spline-based
Flow-based
log(U)
Julian week


WC 2003 CH W only
Spline-based
Flow-based
log(U)
Julian week


WC 2004 CH W+H
Spline-based
Flow-based
log(U)
Julian week


WC 2004 CH W only
Spline-based
Flow-based
log(U)
Julian week

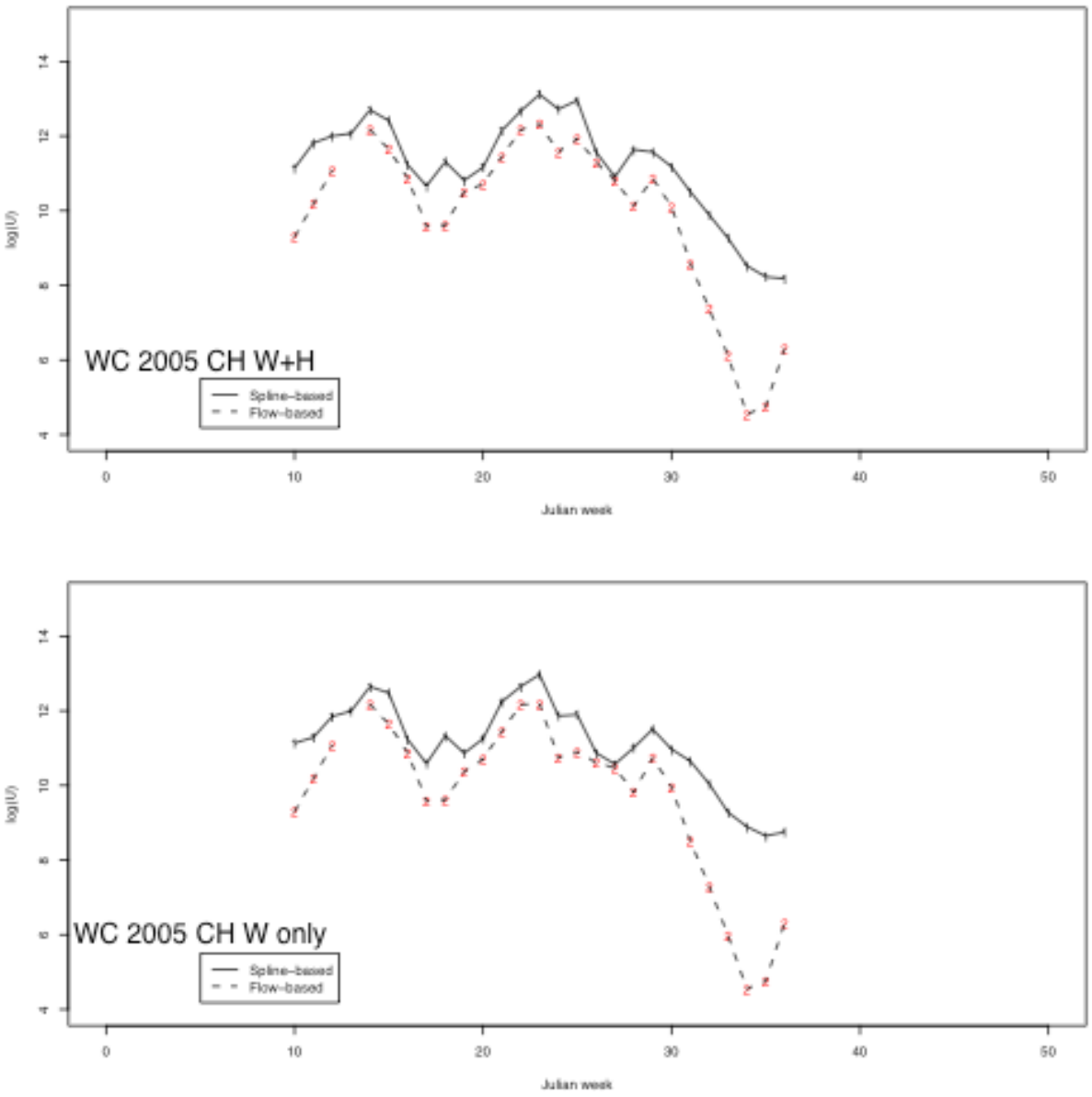


Figure 5.15. Time series plots of the spline-based estimates (solid line) of total outgoing fish vs. the flow-based estimates (dashed line). Not all weeks had discharge sampled and flow-based estimates not computed for those weeks. Top panel of each pair is for wild+hatchery fish combined; bottom panel of each pair is for wild fish only. Site and year labeled on each plot. No flow based estimates could be computed for Junction City 2002, Pear Tree 2003 and Pear Tree 2004 and these plots omitted.

The flow-based estimates of the outmigrant population appear to have the same basic shape as the spline-based estimates. At the Junction City site, the two estimates are very comparable until the second hatchery release arrives at the trap. At Pear Tree and Willow Creek, the flow-based estimates consistently under-estimate the outmigrant population size compared to the spline-based methods.

Finally, the total number of estimated outgoing fish over Julian weeks when estimates are available from both sources (flow-based and spline-based) are compared in Table 5.4 and Table 5.5.

Table 5.4. Comparison of flow-based and spline-based estimates of outgoing Chinook salmon wild + hatchery population for Julian weeks when both estimates are available.

Site	Year	Flow-based Pinnix et al. 2007[1] (millions)	Mark-recapture Pinnix et al.2007[2] (millions)	Ratio of flow-based to mark-recapture	Spline-estimates using same reporting period as Pinnix et al.	Flow-based from this report (millions)	Spline-based covering same periods as flow based (millions)	Ratio flow-based/spline-based
Junction City	2003	-	-	-	-	2.41	5.30	.45
Junction City	2004	-	-	-	-	1.16	8.74	.13
Pear Tree	2005	-	-	-	-	.44	5.68	.08
Pear Tree	2006	-	-	-	-	.26	1.14	.23
Pear Tree	2007	-	-	-	-	.59	3.00	.20
Willow Creek	2002	.75	1.33	.56	1.40	.73	1.97	.37
Willow Creek	2003	.36	.67	.54	.62	.39	1.14	.34
Willow Creek	2004	.28	.43	.65	.63	.26	1.26	.20
Willow Creek	2005	1.80	2.29	.79	2.74	1.48	3.55	.42

[1] Table 17 of Pinnix et al. (2007)

[2] Table 26 of Pinnix et al. (2007). They indicate that these estimates are for the weeks when mark-recapture data are available which may not be consistent with weeks in spline-estimates are available. For example, in Willow Creek 2002, mark-recapture estimates available in Julian weeks 17-27, but Figure 5.15 shows that spline estimates available for Julian weeks 10 to 47. The sum of the spline based estimates for the same reporting period as used in Pinnix et al. are also reported in the table.

Table 5.5. Comparison of flow-based and spline-based estimates of outgoing Chinook salmon wild only population for Julian weeks when both estimates are available.

Site	Year	Flow-based Pinnix et al. 2007[1] (millions)	Mark-recapture Pinnix et al.2007[2] (millions)	Ratio of flow-based to mark-recapture	Spline-estimates using same reporting period as Pinnix et al.	Flow-based from this report (millions)	Spline-based covering same periods as flow based (millions)	Ratio flow-based/spline-based
Junction City	2003	-	-	-	-	.53	.87	.61
Junction City	2004	-	-	-	-	.43	.59	.73
Pear Tree	2005	-	-	-	-	.21	.2.38	.09
Pear Tree	2006	-	-	-	-	.17	.59	.29
Pear Tree	2007	-	-	-	-	.42	1.84	.23
Willow Creek	2002	.54	-	-	.86	.48	1.11	.43
Willow Creek	2003	.28	-	-	.21	.26	.43	.60
Willow Creek	2004	.20	-	-	.53	.15	.74	.20
Willow Creek	2005	1.49	-	-	2.14	1.23	2.74	.45

[1] Table 17 of Pinnix et al. (2007)

The flow based estimates computed in this report are very similar to those reported by Pinnix et al. (2007) with a larger difference for the 2005 study. The differences may be attributable to different way in which anomalous data points are removed and differences in how missing days within a Julian week are accounted for. The estimates based on the mark-recapture methods are not directly comparable between the Pinnix et al. (2007) and this report because they cover different reporting periods.

The estimated ratio between the discharged-sample estimates and spline-based estimates is fairly consistently over time within a study site except for a few anomalous points. For example, the ratio for the Pear Tree site in 2005 is very low, but this apparently is caused by some data anomalies later in the season which give a vary large-spline based estimate of abundance (it actually exceeded the number of hatchery fish known to be released by a factor of two!). The actual data looks consistent but something may have gone wrong in the data base – further investigation is needed.

Similarly the very low and very high ratio at Willow Creek in 2003 and 2004 also are highly influenced by potential data anomalies. In the case of 2003, the discharge-based estimates appear to be inflated by a few weeks where an unusually low fraction of the flow was sampled compared to surrounding weeks. This may simply be a case of missing information because it is currently assumed that if no flow

information is recorded, then no trap was running. In the case of 2004, the results depend highly upon the estimates at the end of the study when the population estimate was very high. Again, further investigation is needed.

The estimates of the wild and wild and hatchery population using the flow-based methods at Willow Creek is summarized in Table 5.6. No estimates of precision are presented. These estimates will be used in the across-year evaluation of the program in Section 8.

Table 5.6 Estimated number of fish from discharge-sampled approach at Willow Creek.

		Wild+Hatchery (millions)	**Wild Only (millions)**
Willow Creek	1998	1.22	0.64
Willow Creek	1999	0.52	0.41
Willow Creek	2000	0.47	0.37
Willow Creek	2001	0.87	0.71
Willow Creek	2002	0.73	0.49
Willow Creek	2003	0.52	0.30
Willow Creek	2004	0.26	0.17
Willow Creek	2005	1.59	1.32
Willow Creek	2006	0.26	0.07

A comparison of the run timing estimates computed from the two methods is presented in Table 5.7 and the yearly run curves are summarized in Figure 5.16. The flow based estimates generally capture the shape of the outgoing migration, but because of cases where the flow-based estimates of the population are much lower than the corresponding spline-based estimate, the estimates of run timing tend to be shifter earlier compared to the comparable run timing estimates from the spline methods.

Table 5.7 Comparison of run timings from flow-based estimates and spline based estimates. Only those site-year combinations where the flow-based estimates were contiguous are presented.

			Quantile					
Site	**Year**	**Method**	**0%**	**10%**	**30%**	**50%**	**70%**	**90%**
Wild and hatchery fish combined								
Junction City	2003	spline	9	19.2	24.3	26.2	40.7	42.7
Junction City	2003	flow	9	17.8	23.5	24.1	26.0	41.0
Pear Tree	2007	spline	1	11.9	16.9	22.9	23.7	25.4
Pear Tree	2007	flow	3	12.7	16.6	18.9	23.5	25.8
Willow Creek	2002	spline	11	19.4	24.1	25.3	27.0	41.8
Willow Creek	2002	flow	11	18.1	21.6	24.6	26.1	29.7
Willow Creek	2003	spline	10	18.8	26.0	27.5	35.9	42.4
Willow Creek	2003	flow	10	11.3	14.3	20.7	26.5	31.2
Willow Creek	2004	spline	12	18.5	27.7	30.8	41.4	42.5
Willow Creek	2004	flow	12	14.8	22.8	28.0	41.1	42.3
Willow Creek	2004	spline	12	18.5	27.7	30.8	41.4	42.5
Willow Creek	2004	flow	12	14.8	22.8	28.0	41.1	42.3
Wild fish only								
Junction City	2003	spline	9	9.4	10.3	11.4	23.1	29.3
Junction City	2003	flow	9	9.7	11.8	18.5	23.6	28.2
Pear Tree	2007	spline	1	9.4	15.5	17.1	20.3	23.9
Pear Tree	2007	flow	3	10.9	15.8	17.1	19.0	23.7
Willow Creek	2002	spline	11	17.6	20.8	23.9	25.7	28.8
Willow Creek	2002	flow	11	16.7	19.9	23.0	24.9	27.5
Willow Creek	2003	spline	10	11.5	18.7	25.0	27.4	30.8
Willow Creek	2003	flow	10	11.2	11.8	15.5	19.9	27.3
Willow Creek	2004	spline	12	14.9	21.6	25.3	29.6	33.6
Willow Creek	2004	flow	12	13.4	16.8	22.6	25.2	30.6
Willow Creek	2004	spline	12	14.9	21.6	25.3	29.6	33.6
Willow Creek	2004	flow	12	13.4	16.8	22.6	25.2	30.6

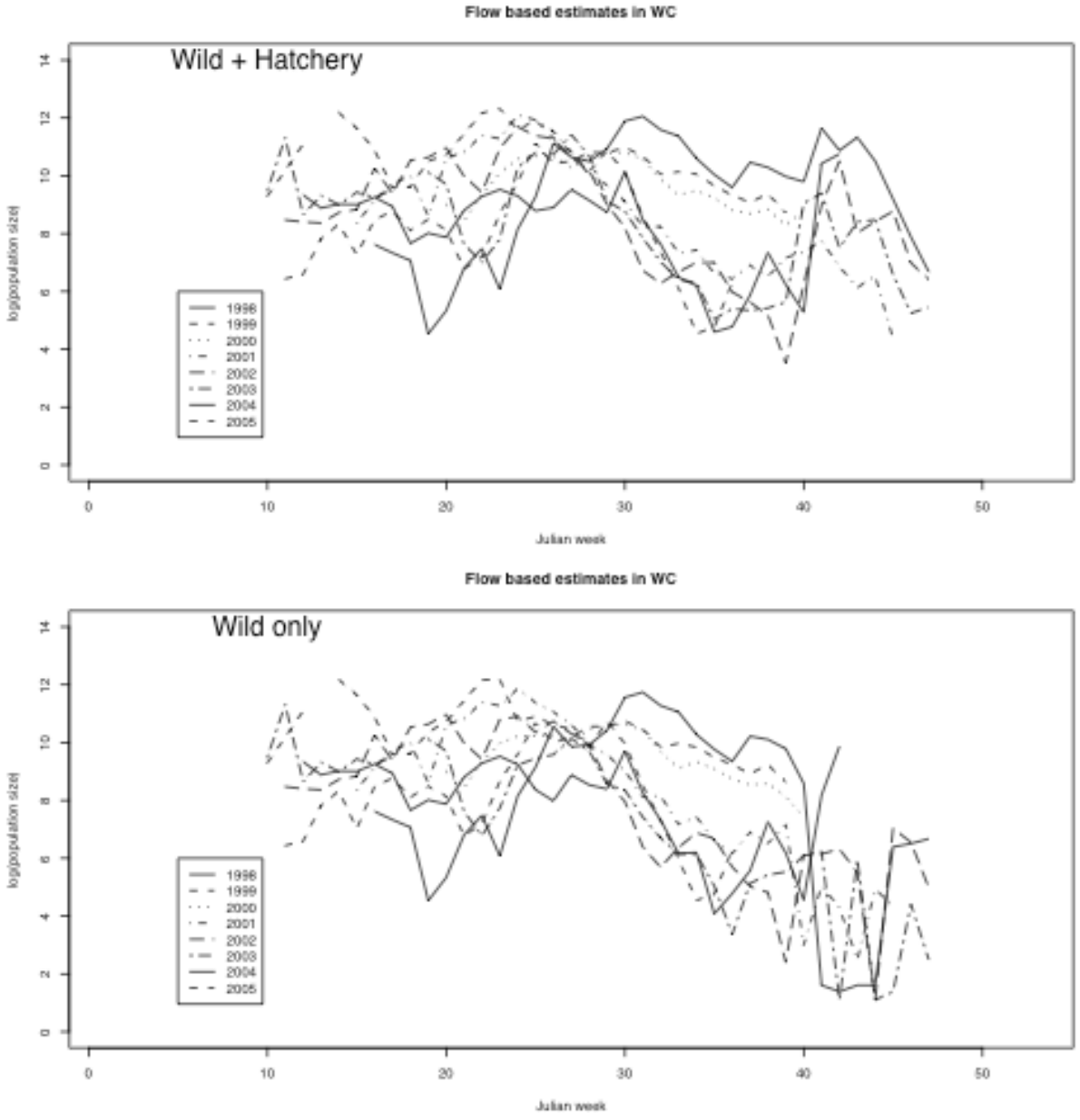


Figure 5.16 A comparison of the flow-based estimates for all years at the Willow Creek site.

The run-time estimates are also obtained for all years at Willow Creek and are available on the web site.

5.6 Discussion and Guidance

Both the spline-model with the log(flow) as a covariate (Section 5.3) and the analysis described in Section 5.4 showed that the relationship between catchability and flow as measured at the HVT/USGS gauges is weak and changeable over years within a site. As noted earlier, this may be an artifact of fish behavior where fish migrate along the margins of the river at all flows where the screw-traps are situated. This

implies that using flow as an estimate of the efficiency of the screw-trap will not lead to large gains in precision. Because the relationship between flow and catchability may change among years, this also adds an additional level of uncertainty when trying to use a common relationship across all years.

The methods using the fraction of discharge sampled appear to capture the general shape of the outgoing migration pattern quite well but the estimated capture-efficiency based on the amount of flow sampled underestimates the actual efficiency of the traps as measured by the mark-recapture methods. This implies that the index of outgoing migration underestimates the actual number of outgoing migrants. Unfortunately, the constant of proportionality may vary considerably across years even within the same study and further work is needed to try and identify the reasons for such wide variation before applying these flow-based estimates to years without a mark-recapture study. Based on the results in Table 5.4 and

Table 5.5, the estimated flow-based population sizes could vary by more than a factor of 2 because of unknown causes. Estimates of run timing are not as severely impacted, but there is evidence that the flow based estimates tend to produce run timings that are earlier than the corresponding spline-based estimate. This is likely due to the presence of one or two weeks where the flow-based estimate falls considerably below the estimate from spline-based method (see previous plots) which then tends to distort all future run timings.

No standard errors have been presented for the flow-based estimates. It is not entirely clear how to compute such standard errors that properly account for all of the sources of variation. There is some information available in the raw data. For example, the variation in revolutions across the locations measured at the trap and the variability in the discharge sampled across days within a Julian week could be propagated forward to give a measure of uncertainty for $\hat{p}_{syj}^{flow}$. This uncertainty, and the variability in u_i (the number of unmarked fish captured in a week) could then be combined together to give an overall uncertainty for estimated population size for the week which are then combined to give an overall measure of uncertainty for the season. The algebra is tedious and a bootstrap approach may be more realistic.

The flow-based method has some obvious shortcomings. For example, no estimates are available when measurements are not taken. The method could be extended in much the same way as the mark-recapture methods by using a spline-based approach to "interpolate" weeks when no flow measurements are available.

The results of this investigation suggest a hybrid approach that may be suitable for future years. The figures indicate that the constant of proportionality is fairly consistent across weeks within a year. Consequently, a study where flow measurements taken over the entire season are supplemented with mark-recapture experiments in a few weeks to calibrate the traps and establish the constant of proportionality between the flow-based and mark-recapture efficiencies may work quite well. This is particularly true in cases where electronic monitoring of the flow can be done.

6. Run timing

6.1 Methods

Once the model has been fit using MCMC methods, a sample from the posterior distribution for the weekly number of outgoing fish $\{U_i; i = 1,\mathsf{K}\ , s\}$ is available. The estimated run timings (e.g. at what time has 75% of the population passed the traps) are readily computed.

The quantiles are found using the methods for grouped data. For example, suppose that in weeks 1 to 5, the following are the estimated counts of fish passing the trap {100, 200, 100, 300, 100} for a total passage of 800 fish. The 20th percentile would correspond to the time when 20%x800=160 fish would have passed the trap. This must have occurred in the second week. In this example, 100 fish passed the trap by time 2.0, and 300 fish passed by time 3.0. Simple linear interpolation would estimate that the 160th fish would have passed at time 2.3. The estimated time for the 20th percentile of the run timing is then 2.3 weeks.

Such an estimate is made for each sampled set of $\{U_i; i = 1,\mathsf{K}\ , s\}$ from the posterior. The run times can be estimated for each point in the posterior. The mean and SD over the estimated run times are then summary measures for the run timing in the same way as any other parameter.

6.2 Results

Estimates of run timing can only be computed when estimates of the outmigrant population size are available at a weekly level. Consequently, as noted earlier, estimates of run timing can be obtained in all 8 studies for Chinook salmon (Table 6.1), in selected studies for steelhead (Table 6.4), but not for coho. Because the results for run timing are similar with and without using flow as a covariate, only the results for the case of the spline model with no flow covariate is presented.

6.2.1 Chinook salmon run timing

Estimates of run timing for Chinook salmon (all ages, wild and hatchery pooled) are presented in Table 6.1.

Table 6.1. Estimate of run timing (Julian week) for Chinook salmon (all ages, wild and hatchery combined).

		Percentile										
Site		**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Junction City 2002	Mean	9.0	23.9	25.1	25.5	26.0	26.4	26.8	27.5	29.2	41.6	46.0
	SD	0.0	0.2	0.1	0.1	0.1	0.1	0.1	0.2	0.3	0.1	0.0
Junction City 2003	Mean	9.0	19.4	23.7	24.3	24.7	26.3	38.6	40.7	41.9	42.7	47.0
	SD	0.0	4.0	0.1	0.1	0.1	0.7	3.2	0.1	0.2	0.1	0.0
Junction City 2004	Mean	6.0	23.0	23.7	24.6	32.6	40.7	41.2	41.5	41.7	41.9	47.0
	SD	0.0	0.7	0.1	0.4	5.8	1.4	0.1	0.0	0.0	0.0	0.0
Pear Tree	Mean	2.0	16.2	17.1	18.0	22.9	32.5	34.1	34.4	34.7	35.0	39.0

2003	SD	0.0	1.0	0.4	0.7	4.5	2.9	0.4	0.2	0.1	0.0	0.0
Pear Tree 2004	Mean	12.0	23.1	24.2	25.3	30.2	37.6	40.2	40.5	40.7	40.9	47.0
	SD	0.0	1.3	0.3	2.4	6.2	5.1	0.9	0.1	0.0	0.1	0.0
Pear Tree 2005	Mean	1.0	6.8	10.6	14.9	18.9	22.7	23.4	23.7	24.0	24.7	28.0
	SD	0.0	0.7	1.5	1.5	2.6	1.1	0.2	0.1	0.2	0.1	0.0
Pear Tree 2006	Mean	8.0	9.9	12.9	18.4	23.0	24.7	25.4	26.2	27.3	29.2	34.0
	SD	0.0	1.2	3.3	4.9	3.1	1.0	0.4	0.5	0.6	0.6	0.0
Pear Tree 2007	Mean	1.0	11.9	15.7	17.0	19.7	22.5	23.3	23.7	24.1	25.4	33.0
	SD	0.0	0.7	0.1	0.4	0.4	0.7	0.1	0.1	0.1	0.2	0.0
Willow Creek 2002	Mean	11.0	19.4	22.0	24.0	24.7	25.3	25.9	27.2	31.1	40.2	48.0
	SD	0.0	0.6	1.0	0.3	0.2	0.2	0.4	1.4	5.0	3.2	0.0
Willow Creek 2003	Mean	10.0	19.7	25.3	26.0	26.7	27.6	29.4	35.1	40.8	42.4	48.0
	SD	0.0	3.0	0.5	0.2	0.3	0.6	2.3	4.9	1.7	0.5	0.0
Willow Creek 2004	Mean	12.0	19.3	24.4	27.2	29.4	31.5	35.1	40.3	41.9	42.4	43.0
	SD	0.0	2.4	1.6	1.7	1.3	3.1	4.2	2.8	0.4	0.3	0.0
Willow Creek 2005	Mean	10.0	13.2	14.7	17.0	20.8	22.7	23.6	24.5	25.5	27.6	37.0
	SD	0.0	0.6	0.6	1.9	1.7	0.7	0.4	0.4	0.5	0.9	0.0

The estimates of run timing in the Junction City 2004 study occur much later in the year for higher percentiles because of the very large estimate of the run in later Julian weeks where the numbers of recaptures were low but not implausible. Because the size of the hatchery-released population is large relative to the natural population, the run timing is heavily influenced by the timing of the hatchery release. Data anomalies may be present. For example, the flow of the river would suggest that a certain percentile of the run should increase in time as fish flow down the river, (e.g. Junction City < Pear Tree < Willow Creek) but this was not always reflected in the run timing percentiles.

Estimates of run timing for Chinook salmon (YOY, wild and hatchery separated) are presented in Table 6.2, and Table 6.3.

Table 6.2. Estimate of run timing (Julian week) for Chinook salmon Wild YOY.

		Percentile										
Site		**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Junction City 2002	Mean	9.0	16.9	21.7	25.4	25.9	26.3	26.7	27.0	27.7	28.8	41.0
	SD	0.0	0.6	1.9	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.0
Junction City 2003	Mean	9.0	9.5	10.0	10.3	10.7	13.0	17.3	22.5	25.8	29.2	40.0
	SD	0.0	0.2	0.2	0.3	0.3	2.1	2.6	2.0	0.8	0.6	0.0
Junction City 2004	Mean	6.0	6.6	7.5	8.5	10.6	14.9	19.4	22.7	24.0	25.4	40.0
	SD	0.0	0.2	0.8	1.2	2.0	3.3	2.3	1.1	0.3	0.5	0.0
Pear Tree 2003	Mean	10.0	13.0	15.2	17.1	19.1	21.3	23.6	25.4	27.2	30.0	40.0
	SD	0.0	0.5	0.9	1.1	1.5	1.6	1.5	1.0	0.8	0.9	0.0
Pear Tree 2004	Mean	12.0	13.0	14.3	15.5	17.1	19.4	22.0	23.6	24.7	26.1	40.0
	SD	0.0	0.5	0.9	1.1	1.6	2.3	1.8	0.6	0.5	0.4	0.0
Pear Tree 2005	Mean	1.0	5.0	6.8	8.7	11.5	14.2	15.9	18.4	22.9	24.3	28.0
	SD	0.0	0.8	0.9	1.1	1.3	1.1	0.7	1.4	0.9	0.2	0.0
Pear Tree	Mean	8.0	9.5	11.1	13.2	15.8	18.9	22.1	24.7	26.6	28.9	34.0

2006	SD	0.0	0.6	1.2	1.5	1.9	2.2	1.7	1.0	0.7	0.5	0.0
Pear Tree 2007	Mean	1.0	9.4	13.9	15.5	16.2	17.2	19.0	20.3	22.2	23.9	33.0
	SD	0.0	0.5	0.9	0.1	0.2	0.5	0.4	0.3	0.5	0.1	0.0
Willow Creek 2002	Mean	11.0	17.6	19.7	20.9	22.7	23.9	24.7	25.7	27.2	28.8	40.0
	SD	0.0	0.6	0.5	0.5	0.6	0.3	0.2	0.3	0.4	0.5	0.0
Willow Creek 2003	Mean	10.0	11.9	14.6	17.7	20.8	23.3	25.2	26.7	28.7	31.2	40.0
	SD	0.0	1.0	1.8	2.2	2.5	2.2	1.7	1.2	1.0	0.9	0.0
Willow Creek 2004	Mean	12.0	15.1	18.0	21.4	23.6	25.4	27.6	29.5	31.1	33.3	40.0
	SD	0.0	1.0	1.6	1.6	1.1	1.2	1.3	0.9	0.9	1.3	0.0
Willow Creek 2005	Mean	10.0	13.2	14.4	15.5	18.3	21.0	22.5	23.3	24.3	26.8	37.0
	SD	0.0	0.8	0.6	1.0	2.1	1.5	0.7	0.4	0.6	1.5	0.0

Table 6.3. Estimate of run timing (Julian week) for Chinook salmon Hatchery YOY.

		Percentile										
Site		**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Junction City 2002	Mean	23.0	24.2	24.9	25.3	25.6	26.0	26.3	26.6	26.9	28.0	41.0
	SD	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0
Junction City 2003	Mean	23.0	23.4	23.7	24.1	24.2	24.4	24.6	24.8	25.1	26.9	40.0
	SD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0
Junction City 2004	Mean	23.0	23.2	23.4	23.5	23.7	23.9	24.1	24.6	25.1	26.5	40.0
	SD	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.1	0.0
Pear Tree 2003	Mean	24.0	24.3	24.6	24.8	25.1	25.6	26.0	26.8	27.9	29.8	40.0
	SD	0.0	0.0	0.1	0.1	0.1	0.2	0.2	0.3	0.5	0.6	0.0
Pear Tree 2004	Mean	23.0	24.0	24.2	24.3	24.5	24.6	24.8	24.9	25.3	26.2	40.0
	SD	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0
Pear Tree 2005	Mean	22.0	23.1	23.2	23.4	23.6	23.7	23.9	24.1	24.4	24.8	28.0
	SD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0
Pear Tree 2006	Mean	24.0	24.3	24.5	24.8	25.1	25.5	26.0	26.8	28.0	29.8	34.0
	SD	0.0	0.1	0.1	0.2	0.3	0.5	0.6	0.8	0.9	0.8	0.0
Pear Tree 2007	Mean	23.0	23.2	23.4	23.5	23.7	23.9	24.1	24.6	25.1	26.8	33.0
	SD	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.3	0.0
Willow Creek 2002	Mean	24.0	24.3	24.6	24.9	25.1	25.4	25.7	26.0	26.7	28.3	40.0
	SD	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.2	0.4	0.6	0.0
Willow Creek 2003	Mean	25.0	25.3	25.7	26.0	26.3	26.7	27.0	27.6	28.3	29.7	40.0
	SD	0.0	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.3	0.5	0.0
Willow Creek 2004	Mean	24.0	25.8	26.5	27.2	27.9	28.6	29.4	30.1	31.0	32.6	40.0
	SD	0.0	0.3	0.2	0.3	0.4	0.5	0.5	0.5	0.7	1.0	0.0
Willow Creek 2005	Mean	23.0	23.9	24.3	24.6	25.0	25.2	25.5	25.7	26.0	27.0	37.0
	SD	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.5	0.0

A comparison of the run timing between the hatchery and wild components shows that the wild YOY component is not as concentrated as the hatchery YOY component. The hatchery YOY component passes the screw-trap in about 5 weeks while the wild YOY component is spread over more than 10 weeks. [This

may be an artifact of the overlap in spring- and fall-run outmigration life histories and of hatchery fish release practices and patterns.]

6.2.2 Steelhead run timing

Estimates of run timing for steelhead (all ages, wild and hatchery pooled) are presented in Table 6.4 through Table 6.7. Many of the studies at the Pear Tree site had insufficient data, and mark-recapture studies were not done at the Willow Creek sites to compute estimates of run timing.

Table 6.4. Estimate of run timing (Julian week) for steelhead (all ages, wild and hatchery combined).

		Percentile										
Site		**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Junction City 2002	Mean	1.0	5.1	6.5	7.8	9.0	9.8	12.3	16.2	20.0	24.2	38.0
	SD	0.0	0.2	0.4	0.4	0.4	0.7	2.0	1.1	1.0	0.6	0.0
Junction City 2003	Mean	1.0	4.2	4.9	5.6	6.5	7.5	8.5	11.0	16.3	23.8	39.0
	SD	0.0	0.2	0.3	0.4	0.7	0.7	1.3	3.8	5.1	3.0	0.0
Junction City 2004	Mean	6.0	10.8	12.3	13.3	14.3	15.5	16.8	18.8	22.2	29.0	47.0
	SD	0.0	1.7	1.5	1.3	1.3	1.4	1.9	3.0	4.5	4.6	0.0
Pear Tree 2003	Mean	Insufficient data										
	SD											
Pear Tree 2004	Mean	Insufficient data										
	SD											
Pear Tree 2005	Mean	Insufficient data										
	SD											
Pear Tree 2006	Mean	Insufficient data										
	SD											
Pear Tree 2007	Mean	1.0	12.5	13.7	14.8	15.9	17.0	18.5	21.1	23.8	27.2	33.0
	SD	0.0	0.2	0.4	0.5	0.5	0.6	1.1	1.2	0.9	1.0	0.0
Willow Creek 2002-2005		Mark-recapture not done.										

Table 6.5. Estimate of run timing (Julian week) for steelhead Wild YOY.

		Percentile										
Site		**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Junction City 2002	Mean	9.0	24.3	25.9	27.5	28.4	29.3	31.1	32.3	33.5	36.3	46.0
	SD	0.0	0.4	0.8	0.6	0.3	0.7	0.8	0.4	0.6	0.8	0.0
Junction City 2003	Mean	9.3	26.2	28.3	29.5	30.3	31.1	32.2	33.8	35.4	37.8	47.0
	SD	0.6	0.7	0.8	0.5	0.5	0.6	1.0	1.1	0.8	2.7	0.0
Junction City 2004	Mean	6.0	27.6	29.1	30.3	31.6	32.9	34.0	35.2	36.8	39.9	47.0
	SD	0.2	0.7	0.7	1.0	1.2	1.2	1.1	1.1	1.6	2.1	0.0
Pear Tree 2003	Mean	Insufficient data										
	SD											

Pear Tree 2004	Mean	Insufficient data									
	SD										
Pear Tree 2005	Mean	Insufficient data									
	SD										
Pear Tree 2006	Mean	Insufficient data									
	SD										
Pear Tree 2007	Mean	4.7	19.6	20.6	22.0	23.4	24.5	25.7	27.2	28.6	30.7	33.0
	SD	1.9	0.3	0.5	1.1	0.8	0.7	0.9	0.8	0.8	1.0	0.0
Willow Creek 2002-2005		Mark-recapture not done.										

Table 6.6. Estimate of run timing (Julian week) for steelhead Wild 1+.

		Percentile										
Site		**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Junction City 2002	Mean	9.0	12.5	13.8	14.8	15.7	16.5	17.2	17.7	18.6	21.4	46.0
	SD	0.0	0.4	0.4	0.4	0.5	0.4	0.3	0.4	0.7	1.0	0.0
Junction City 2003	Mean	9.0	11.0	11.8	12.4	13.2	14.4	15.4	16.1	17.1	19.9	47.0
	SD	0.0	0.4	0.4	0.4	0.7	0.9	0.5	0.4	0.6	1.5	0.0
Junction City 2004	Mean	6.0	11.0	12.2	13.1	14.1	15.0	15.8	16.8	17.9	19.9	47.0
	SD	0.0	0.7	0.6	0.7	0.6	0.5	0.6	0.6	0.7	1.6	0.0
Pear Tree 2003	Mean	Insufficient data										
	SD											
Pear Tree 2004	Mean	Insufficient data										
	SD											
Pear Tree 2005	Mean	Insufficient data										
	SD											
Pear Tree 2006	Mean	Insufficient data										
	SD											
Pear Tree 2007	Mean	1.0	9.2	11.6	12.6	13.5	14.5	15.3	16.3	17.5	20.1	33.0
	SD	0.0	0.8	0.6	0.4	0.6	0.5	0.4	0.5	0.5	1.2	0.0
Willow Creek 2002-2005		Mark-recapture not done.										

Table 6.7. Estimate of run timing (Julian week) for steelhead Hatchery 1+.

		Percentile										
Site		**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Junction City 2002	Mean	12.0	12.7	13.4	14.2	15.1	15.7	16.3	16.9	17.4	17.9	46.0
	SD	0.0	0.1	0.3	0.4	0.5	0.4	0.3	0.2	0.2	0.2	0.0
Junction City 2003	Mean	12.0	12.4	12.7	13.1	13.5	14.0	14.8	15.6	16.2	17.0	47.0
	SD	0.0	0.1	0.2	0.3	0.4	0.6	0.6	0.4	0.4	0.3	0.0
Junction City	Mean	12.0	12.9	13.4	13.7	14.1	14.5	15.0	15.6	16.3	17.2	47.0

2004	SD	0.0	0.2	0.2	0.2	0.2	0.3	0.2	0.3	0.3	0.4	0.0
Pear Tree 2003	Mean	Insufficient data										
	SD											
Pear Tree 2004	Mean	Insufficient data										
	SD											
Pear Tree 2005	Mean	Insufficient data										
	SD											
Pear Tree 2006	Mean	Insufficient data										
	SD											
Pear Tree 2007	Mean	12.0	12.6	13.1	13.5	14.0	14.6	15.2	15.7	16.3	17.0	33.0
	SD	0.0	0.2	0.2	0.3	0.3	0.4	0.3	0.3	0.3	0.2	0.0
Willow Creek 2002-2005		Mark-recapture not done.										

The run timing estimates indicate that the wild age 1+ fish tend to move about 15 weeks earlier than the wild YOY fish but the spread in the run appears to be comparable at about 9-10 weeks.

6.3 Discussion and Guidance

It is relatively simple to obtain estimates of run timing from the spline-based methods as an estimate of the population passing the screw-trap in each week is obtained. It is implicitly assumed that passage is uniform within the week which is clearly not the case, but the error introduced by this assumption is small relative to sampling errors.

A key requirement to ensure that estimates of run times are sensible, is that the study covers the entire migration period (or at least that the number of fish moving outside the study window is negligible). It is possible to use the spline-based methods to interpolate outside the study boundaries. Given that the tail of the study usually has few fish, interpolation after the study period is unlikely to be problematic. However, in several of the datasets, the study began when a large number of fish were passing the traps and interpolating before the study begins is highly problematic.

Section 8 will discuss the cross-year comparison of the run timing percentiles.

7. Evaluation of Condition

One component of the IAP Objective 3.2.2 is to assess the improvement in growth, physical condition and health from baseline conditions in the mainstem Trinity River. This section of the report evaluates the existing condition related data. What data are available? What metrics of growth and physical condition can be generated from the existing data and how do we generate them? For species and years where sufficient data are available, we report annual estimates for a sample of condition related metrics. This subject is currently being investigated by an IAP technical working group and so we simply provide a few examples of how to generate annual estimates of condition from the available data. Finally we summarize the feedback provided to us by the TRRP and Program partners regarding potential metrics of interest and related hypotheses. We hope that this information will prove useful to the IAP technical working group as they move forward with their investigations.

7.1 Methods

7.1.1 Data availability

The recently developed USFWS database can store fork length, weight, and the incidence of a variety of health-related indicators for individual fish (e.g., bloody vent, open wounds, fin rot). Fork length data have been collected throughout the dataset for all sites and all species (Table 7.1). Weight data were available only for the most recent years at each site except Junction City. There were insufficient health observation data to complete any analyses.

Table 7.1. Summary of the availability of fork length and weight data.

Site	Fork length	Weight	Health observations
Willow Creek	1993-2006*	2004-2006	-
Pear Tree	2003-2007*	2006-2007	-
Junction City	1997-2004	-	-

* 2006 is the latest year for which data were made available to the review team for Willow Creek. 2007 is the latest year for which data were made available for Pear Tree.

7.1.2 Fork length

Fork length is relatively easy to measure and has the longest time-series for condition-related data available on the Trinity River (Table 7.1). The length of this time-series makes it worth examining metrics directly related to fish length. Size of outmigrating juveniles has been positively related to survival for some stocks of Chinook salmon and other salmon species (Wedemeyer et al. 1980; Kjelson et al. 1981; Healey 1991). So, from this point of view, length can be used directly as an index of condition and tendency to survive the transition to estuarine and oceanic life stages. However, this should be done with caution as variability in fish length can be large and obtaining a random sub-sample of fish can be difficult (Hilborn and Walters 1992), and this relationship can vary depending on stock and on natural versus hatchery-origin fish. But keeping these qualifications in mind, we want to produce an annual metric of condition based on fork length that can then be evaluated across many years providing an

assessment tool for the TRRP (Section 8). Many different annual metrics are possible. We suggest several options:

- *Simple Julian date:* average length on a given Julian day or week
- *Run timing*: average length at a given point in the run (e.g., when 50% of the fish have passed the trap)
- *Degree date*: average length on a given degree day (e.g., when the cumulative degree days attain some threshold)

Additionally, we asked the TRRP and Program partners for feedback and suggestions for metrics related to fork length, these ideas are summarized in Table 7.2. These comments raise a number of important ideas and will be helpful to the IAP Technical working groups as they move forward.

Table 7.2. Summary of feedback from TRRP and Program partners for length related indices.

Question	**Feedback from TRRP and partners**
How useful are the proposed fork length metrics?	I think they have some promise. However, hatchery fish can blur the picture as not all hatchery fish are marked. Also, especially at the lower trap, there are lower river populations that contribute to the catch and biosample which during certain periods of the season can greatly influence FL data. (USFWS)
How should the fork length data be summarized: using the daily means, weekly means, moving average, or a modeled value?	I think daily means is best compared to some modeled (i.e. ideal) value. (USFWS) It seems logical that if pop. est. is stratified by weekly estimates, maybe most of the other data should be as well. (Yurok)
Would degree date be a useful way to report fork length or other condition metrics?	I think this also has some promise. (USFWS)
How should we select specific metrics?	This is really something that should be worked on through the workgroup process so that all program partners have the opportunity to provide input. (USFWS) I like table 6.2 [table illustrating several metrics simultaneously across years, now 7.3] in being able to compare the different metrics…….maybe combine the values of all 3 to develop an index…. (USFWS)
What specific dates would be of interest?	1)Pre- vs. post- TRH releases, 2) vs. various temperature objectives @ Willow Creek. (Yurok)
Are there any other triggers that might be of interest to monitor across years (e.g., besides run timing, Julian date, or degree days)?	No feedback was received at this time.

In order to calculate each of these metrics we first need to identify the date of interest. This date may be the same each year (e.g., Simple Julian Day), or it might change annually depending on run timing or degree days or some other trigger of interest (Table 7.3).

Table 7.3. This table shows an example of how different metrics could be translated into Julian days for a series of different years.

Year	**Simple Julian Day**	**Run timing**	**Degree Day**
Metric of interest	**July 19**	**50% of run**	**X degrees C**
1993	200	200	100
1994	200	260	170
…	…	…	…
2007	200	250	140

Once the Julian date for each year is identified, we simply use the available data to estimate the mean fork length on that day. There are a number of possible methods for estimating the mean fork length for a given day (Table 7.4). As mentioned earlier, variation in fork lengths can be large and there is substantial day to day variation in the observed data. The methods described here differ in the amount of smoothing applied to the raw data. In each case we are using the value for a single day to represent the performance for the entire year and the method selected to estimate that day is critical. Figure 7.1 illustrates a range of possible methods for summarizing a single year's fork length data.

Table 7.4. This table describes several different methods for estimating the mean fork length on a given day.

Method	**Reference**
Use the raw daily data and interpolate as necessary for days with no available data.	Figure 7.1a
Split the data into weeks and produce weekly estimates.	Figure 7.1b
Use a weekly moving average, i.e., a value for each day would be estimated by taking the mean of the previous 3 days and next 3 days as well as the current day.	Not shown
Use a smoothing spline to interpolate between points (Appendix A). The amount of smoothing can be varied within this method.	Not shown
Use locally weighted regression to fit a smooth curve to the points. The amount of smoothing can be varied within this method.	Figure 7.1c
Fit a straight line to the data.	Figure 7.1d

Using the raw data could result in very large differences in the annual index by simply choosing one day earlier or later, Figure 7.1 a). Aggregating weekly is a simple method in which this variability could be reduced Figure 7.1 b). A LOWESS fit could reduce the day to day variability and yet allow for local effects to be captured (i.e., local in a temporal sense) Figure 7.1 c). A simple straight line fit to the fork length data within a year would be simple to compare across years, even allowing comparison of the slope

rather than having to choose a single day. A straight line model fits this particular example quite well (Figure 7.1 d), but in many cases (i.e., other years, species, and locations) there were plateaus observed in the fork length data which would not be captured effectively by a straight line model.

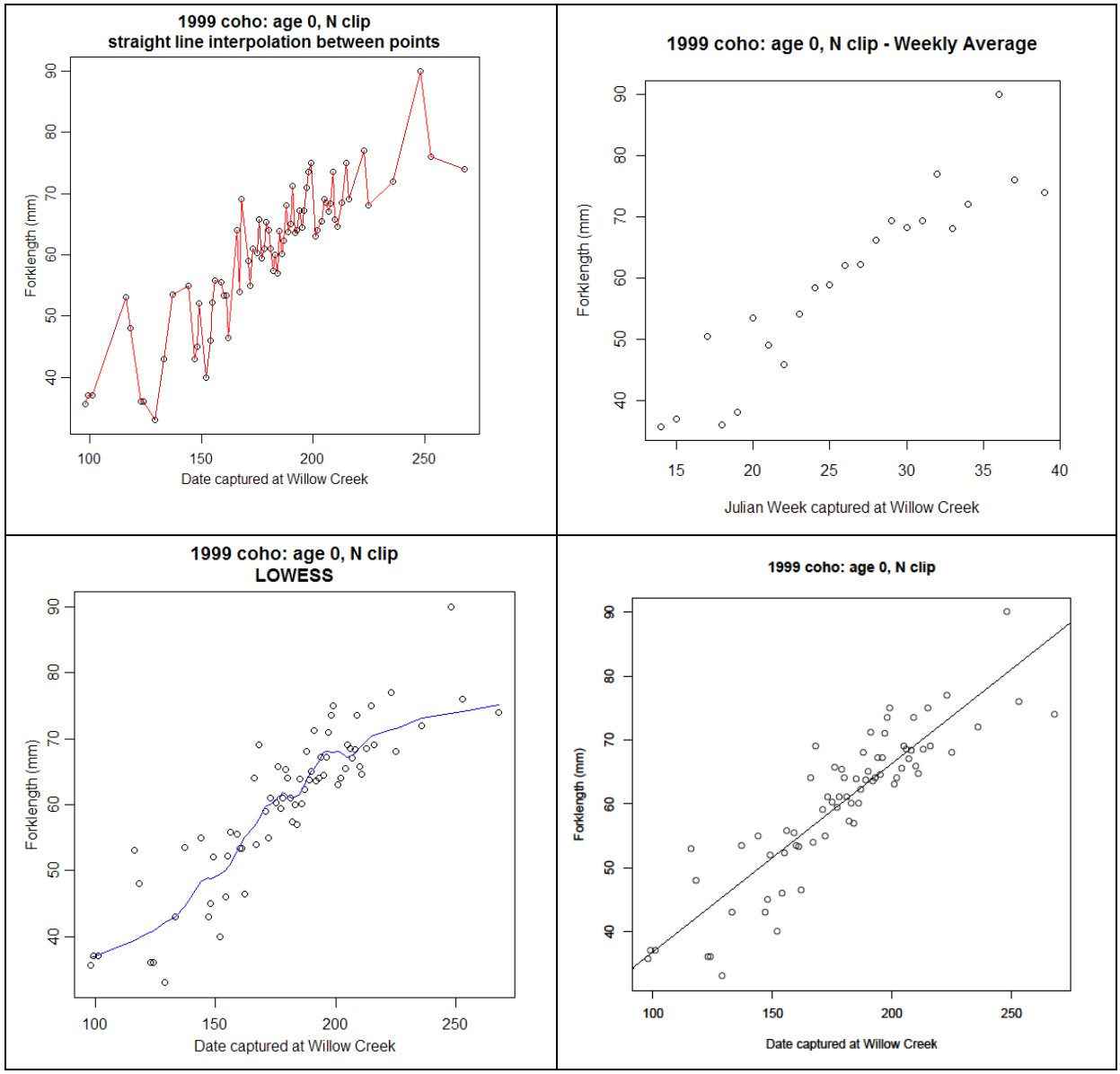


Figure 7.1. a) This figure shows an example of the relationship between daily mean fork length and Julian day for natural origin young of year coho in 1999 at Willow Creek, with four examples of how the data may be summarized: a) using raw data with straight line interpolation between points (to allow estimation for missing days); b) weekly averages; c) fitting a locally weighted regression model d) fitting a straight line regression model across the entire sampling window. These are just a few examples that illustrate the range of smoothing available, but one can see that the choice will affect the resulting index as there is significant within year variation in fork lengths on any given day.

It is important to remember that the goal is to find a metric that allows us to assess whether or not the restoration program is changing the characteristics of the juvenile salmonids across years (Section 8). Some level of smoothing is going to improve the ability to detect changes in fork length related indices across years. For example, if the sample mean from July 18th is 48 mm, and from July 19th is 40 mm, it may be more informative to know what the average fork length was in the week surrounding July 19th, or some other coarser measure that minimizes the within year day to day variability. What metric will have the most meaning across years? The IAP Technical working group can evaluate how well these different methods work under different conditions. It isn't necessary to simply choose one method. Multiple indices may be reported from the same dataset and compared. Another important consideration is separating out natural and hatchery origin stocks (Section 7.1.5). On average this method is effective, but substantial noise is added to the daily fork length estimates for Chinook and so it is likely not a very good idea to the use the daily averages rather than some smoothed value (e.g., weekly average or LOWESS).

Weight

Relative to fork length there are few weight data available (Table 7.1). In recent years where sufficient data were made available to us, we estimated a traditional Fulton condition factor for individual fish as follows: K = weight / length^3 (Ricker 1975). The IAP technical working group is actively evaluating potential metrics for assessing condition. This material was most recently presented at the 2007 Trinity River Science Symposium (Hayden and Pinnix 2007).

We plot the Fulton condition factor against the Julian day, similar to the length analyses. However, the mean value remains fairly consistent across the trapping season. In this case the variance in condition factors may provide more useful information. In order to use the Fulton condition factor to evaluate the program we suggest two metrics which can be compared across years:

1) *A simple annual mean condition factor index*

   We plot this index against the Julian day, similar to the length analyses. However, we found that the mean value remains fairly consistent across the trapping season.

2) *Date of change in variance.*

   Due to the consistency in the mean values across the year, we explored the idea of using the variance in condition factor rather than the mean as an indicator. We tried to identify the within year date where a reduction in variance is observed, under the hypothesis that this may be able to be used as an indicator of the date by which most of the juveniles have smolted.

At this time we simply report these two metrics where sufficient data exist and leave it to the IAP technical working group to develop these further. A summary of the feedback received from the TRRP and Program partners regarding weight related indices of condition is provided in Table 7.5.

Table 7.5. Summary of feedback from TRRP and Program partners for weight related indices of condition.

Question	Feedback
Would you expect the condition index to remain constant across the year? If so then a simple annual mean is probably sufficient, but if not it might be necessary to report the condition index at a given date as with the fork length.	I think we would like to see, hopefully, a change in condition factor as restoration actions are implemented (i.e. more rearing habitat availability might result in higher condition factor). This is really only applicable to Chinook. (Yurok) For Chinook salmon, this may be constant through the sampling season as we are far enough from the estuary that true 'smolting' is not occurring. However, with steelhead and coho smolts, the condition factor tends to decrease through the season as Smoltification progresses. (USFWS)
Does investigating the variance in condition factor across the season make sense?	Again, I don't think we are close enough to the ocean to detect Smoltification in Chinook salmon….maybe with steelhead and coho. (USFWS)
Are there any other annual metrics related to weight that people would like to see?	No feedback received at this time, but the IAP Technical working group is actively addressing this question.

7.1.3 Proportion of fry to smolts

The proportion of juveniles emigrating as fry may be expected to change in response to TRRP restoration activities. Metrics associated with this proportion might be useful for evaluating the impact of the program on the juvenile fish. For example:

- *Annual proportion of fry vs. smolts at each trap.*
- *Proportion of fry vs. smolts related to run timing*

In order to assess this for the historical data we could use a simple size cut-off to identify fry vs. smolts. This metric was not evaluated in more detail in this report, but we include a summary of the feedback we received from the TRRP and Program partners (Table 7.6).

Table 7.6. Summary of feedback from TRRP and Program partners for indices related to life-stage at emigration.

Question	Feedback
Should we consider tracking the proportion of fry:smolt at the rotary screw traps?	This is also one of the issues that are to be addressed in the IAP, specifically in the definition of what is production and how the estimates are partitioned into these classes if it is deemed necessary. (USFWS)

If yes, could we use length to obtain historical estimates from fork length? Is there any concern with confounding with age which is often based on length as well?	Using a standardized length-based cut-off seems problematic if one of the goals of the TRRP is to increase fry-rearing. Could there be an increase in size before the onset of smoltification that would be ignored if using a "length cut-off". Is there any other method that would be appropriate. (Yurok) I don't think it's an issue with age…..the 1+ fish are obvious in their size at date……we age larger steelhead from scales to get the age distribution. (USFWS)
What would the appropriate length cut-off be? 50mm for Chinook?	This seems appropriate, but could be a moving target. (USFWS)

7.1.4 Growth

In order to estimate the growth of fish as they move between the upper and lower traps, we either need some estimate of average travel time between traps or we need to compare individual or batches of marked fish released from the upper trap and recovered at the lower trap. Again, this metric was not evaluated in more detail in this report, but we include a summary of the feedback we received from the TRRP and Program partners (Table 7.7).

Table 7.7. Summary of feedback from TRRP and Program partners for indices related to growth of juvenile salmonids.

Question	**Feedback**
Would metrics that assess the growth between the two rotary screw traps be useful?	All TRRP restoration efforts (other than ROD flows) are solely implemented above the Pear Tree site, so growth in between the 2 sites might be considered inconsequential or anecdotal. (Yurok)
Is it reasonable to assume that fish marked at Pear Tree and recaptured at Willow Creek came from the most recent batch or not, since batch marks are re-used over the season?	This problem might currently being remedied by the use of freeze-branded hatchery fish that have unique weekly marks. (Yurok) We are addressing this very issue in 2009. We are using unique batch marks at both trap sites to definitively identify a recapture to a specific release date. (USFWS)
Could we use information about hatchery release dates and the date that hatchery marks first show up at each trap to estimate the travel time?	This is problematic because the hatchery releases fish over a 2 week window. (USFWS) I don't think it is reasonable because you have to use the first release date and the first capture so you will only be 'measuring' the fastest fish. (USFWS)
Could you use best estimate of travel time based on experience as a starting place and then test that in	No feedback received at this time.

upcoming seasons?	
Would it be useful to use simple assumptions to try to get a historical estimate of growth between the traps?	I don't think this will work, but might be worth some effort. The other issue is that there are introductions to the population between the trap sites (north fork, south fork, new river). (USFWS)
Do you think that travel time between traps for hatchery fish is markedly different from natural origin fish?	Yes (USFWS)
Would it be worth exploring the use of PIT tags to estimate growth of individual fish and also to provide information about travel time?	We have begun to talk about this, and are testing some trap modifications in 2009 (PIT tag readers in the traps). I think the big drawback is the number of tags released to get useable information. What would be great is an analysis of how many PIT tags need to be released to get a valid sample size at the Willow Creek trap site. (USFWS)

7.1.5 Separating Hatchery and Natural Stocks

We are primarily interested in the condition of the naturally spawned juveniles, as the restoration efforts are targeted at these fish. Natural coho and steelhead are easy to distinguish from hatchery fish, as there is a 100% hatchery marking rate for these species. However, there is only a 25% hatchery marking rate for Chinook. In order to assess the average characteristics (length and weight) of natural Chinook, we need to separate the unmarked fish into hatchery and natural fish. The method here is similar but slightly more complicated than that traditionally used to estimate the number of natural origin fish in a sample, as we need to track the length and weight of the different groups as well as the number of individuals in each group.

Initially we tried using the assumption that the sub-sampled fish were selected completely at random and therefore the 25% marking rate is maintained in the sub-sample. However in discussions with Hoopa Valley Tribe we found that the sub-sampling protocol is to take a sample of a fixed size from each group of interest (e.g., 30 fish from each of the following groups: clipped & not-clipped). To account for this we instead used the daily catch data and 25% marking rate to estimate the proportion of the not-clipped fish that were of natural origin. Then we assumed that the sample of not-clipped fish was taken at random from within that group (e.g., no size bias due to sampling protocol). We used this proportion as an estimate of the proportion of natural origin fish from the not-clipped sub-sample.

We know how many fish were sub-sampled (N_T), and how many of these were marked (N_M) or not marked. We know that all of the marked fish were hatchery fish. If we assume the mean length from the sample of known hatchery fish represents the average behavior of all hatchery fish, then we can determine what the average length of the natural fish must have been in order to obtain the overall mean length observed for unmarked fish.

P_N	Estimated proportion of natural fish in the not-clipped fish from the total daily catch
N_T	Total number of fish in the sub-sample
N_M	Number of hatchery marked fish in the sub-sample

$N_U = N_T - N_M$	Number of unmarked fish in the sub-sample
S_T	Sum of the lengths of all fish in the sub-sample
S_M	Sum of the lengths of all hatchery marked fish in the sub-sample
$L_H = S_M / N_M$	Estimated mean length of hatchery fish
$S_H = N_H * L_H$	Estimated sum of the lengths of all hatchery fish in the sub-sample
$L_N = (S_T - S_H) / N_N$	Estimated mean length of natural fish
$N_H = N_T - N_N$	Estimated number of hatchery fish in the sub-sample
$N_N = N_U * P_N$	Estimated number of natural fish in the sub-sample

There is a fair bit of noise during the portion of the dataset where we are estimating means based on the proportion of natural not-clipped fish estimated from the total catch. This isn't surprising, because although on average we would expect the proportion to hold in the sub-sample, even from a completely random sample, there will be many cases where by chance alone the 30 fish selected for the sub-sample do not maintain this proportion. If there are more hatchery fish than we predict then the natural mean will be underestimated to compensate. If there are fewer hatchery fish than we predict the natural mean will be overestimated to compensate.

This method works well during many years (Figure 7.2), however there are several years where the noise is substantial. In some cases there is a single value which is very small or large, almost certainly reflecting an unmarked sample that substantially differed from the assumed proportion. It may be sensible to use some kind of criteria for dropping a day from the sample when the values are improbable based on the biology. If a year is particularly bad, it may be better to simply report the actual not-clipped means rather than trying to separate them out. Alternatively it may be useful to select at least one metric of fork length that is early enough in the year to typically occur prior to the hatchery release. Some form of smoothing will likely be necessary to mitigate for the additional noise resulting from imperfect knowledge of the origin of the un-clipped fish (Table 7.4).

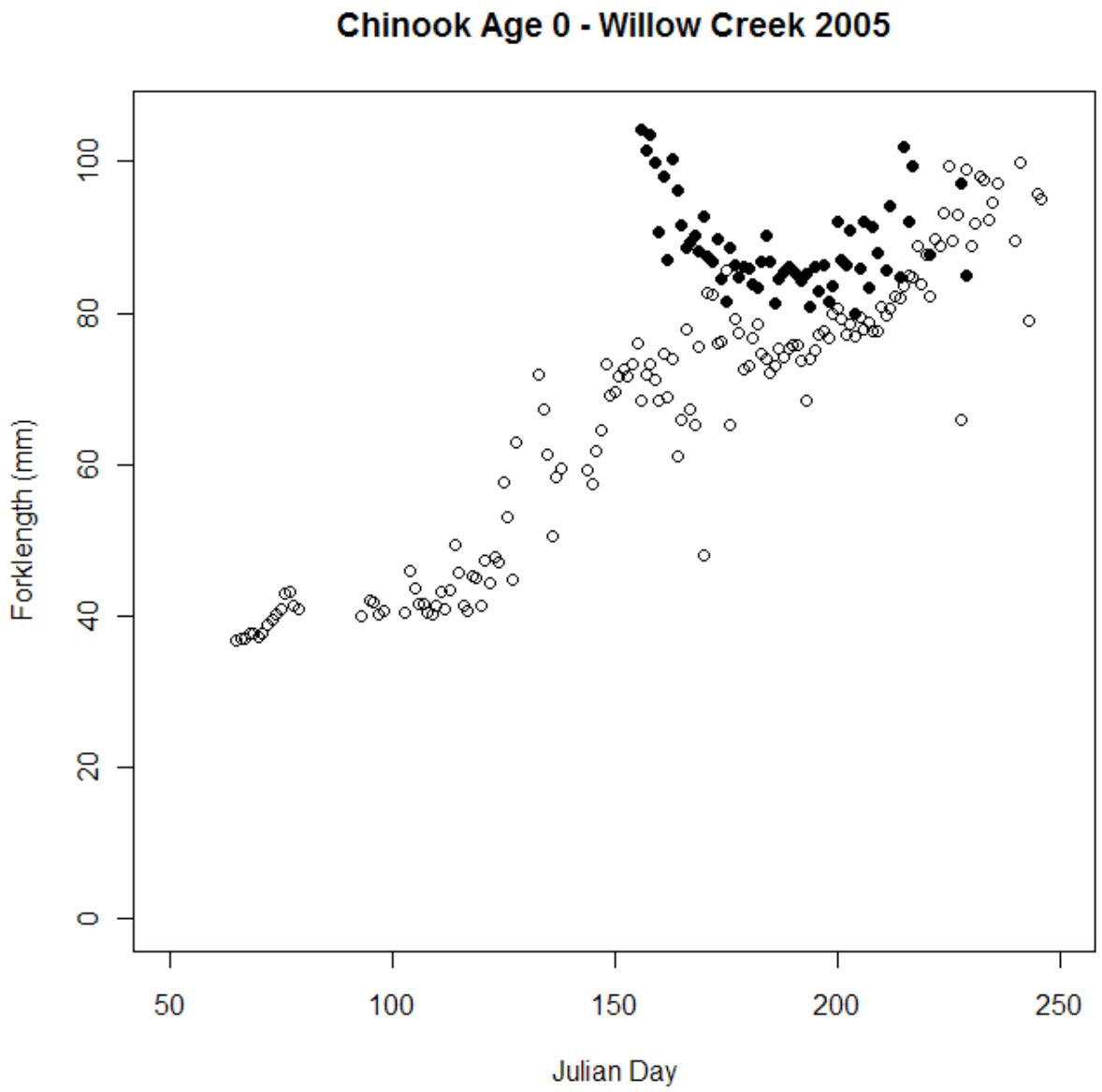


Figure 7.2. An example of the mean daily fork length estimated separately for hatchery (solid dots) and natural (open dots) Chinook, for Willow Creek 2005.

7.2 Results

We report annual estimates for a sample of condition related metrics where sufficient data exist. Where possible, we report results separately for each species, origin (hatchery vs. natural), and age-class. We focused on the analysis of historical fork length data as this dataset is very rich. We estimated condition factors for Chinook in recent years. We did not report estimates for the other proposed metrics: fry:smolt ratio, growth, as there is insufficient information to generate historical /baseline estimates and the IAP technical working group is actively evaluating ideas for future monitoring. The methods described above can be applied to generate a great variety of indices of condition. Annual estimates of condition discussed

here or developed by the IAP technical working group can then be used to evaluate the program using methods described in Section 8.

7.2.1 Fork length

We show an example of the type of information that can be generated from the historical data (Table 7.8 and Table 7.9). This example is for a fixed date across the years. This could be generated for any date of interest or for a date that depends on some sort of trigger such as run timing or degree days. We show results for Julian days 100 and 200. These annual estimates can then be used to evaluate hypotheses of interest to the TRRP (Section 8). In practice the IAP working groups will need to consider the most relevant metric (e.g., Julian Day) for each of their hypotheses of interest. The metric may change by hypothesis, species, and age-class. For example, there are many years where Julian day 100 is either outside the range of steelhead observations or is at the very beginning where the model fits are less accurate.

Unusual observations obtained from these methods should be investigated. For example, the 2005 Willow Creek steelhead estimate of 95.9 mm is likely a result of incorrectly aging the fish (Table 2.6 and Figure 2.1).

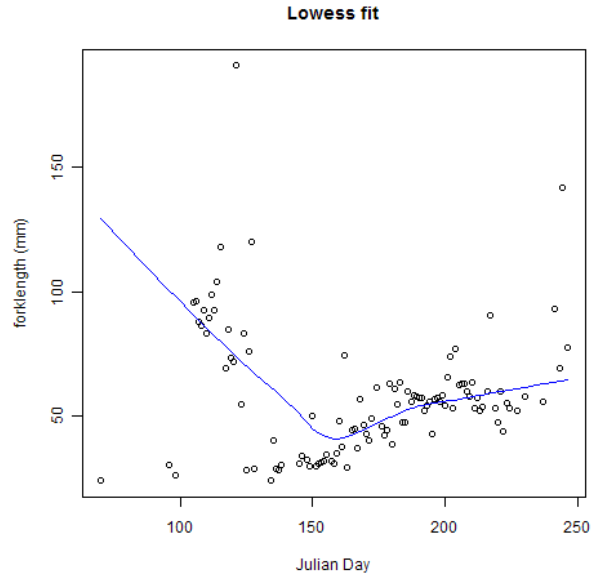


Figure 7.3 A closer look at the Willow Creek 2005 steelhead data suggests that there may have been a problem aging the fish in the early part of the year.

Table 7.8. Average fork length for natural origin young of year salmonids on April 10 (Julian day 100) across all years where data are available. This example was produced using estimates from a lowess model (with smoothing parameter, f=2/3) of fork length x Julian day. (JC=Junction City, PT=Pear Tree, WC=Willow Creek).

	Chinook, age 0				Coho, age 0				Steelhead, age 0		
Year	**JC**	**PT**	**WC**		**JC**	**PT**	**WC**		**JC**	**PT**	**WC**
1993	-	-	36.6		-	-	34.2		-	-	NA
1994	-	-	46.9		-	-	35.4		-	-	NA
1995	-	-	40.8		-	-	NA		-	-	NA
1996	-	-	41.0		-	-	40.0		-	-	NA
1997	*	-	43.7		*	-	NA		*	-	NA
1998	*	-	NA		*	-	NA		*	-	NA
1999	*	-	44.3		*	-	35.7		*	-	NA
2000	*	-	NA		*	-	NA		*	-	NA
2001	*	-	42.9		*	-	NA		*	-	NA
2002	49.7	-	49.4		46.1	-	40.6		26.2	-	30.9
2003	NA	49.0	47.8		NA	39.2	38.7		NA	15.3	18.2
2004	49.9	46.2	41.2		32.0	34.6	37.9		23.3	12.3	29.4
2005	-	54.9	44.7		-	37.8	37.8		-	26.6	[9]95.9
2006	-	44.9	NA		-	36.4	NA		-	29.0	NA
2007	-	54.8	-		-	38.6	-		-	NA	-

*The 1997–2001 Junction City data can be included once the naming convention has been resolved (Appendix E).
(-) represent years with no data
(NA) represent years with data, but where the Julian date selected occurred outside the range of observed data.

[9] Unusual observations should be examined in more detail

Table 7.9 Average fork length for natural origin young of year salmonids on July 19 (Julian day 200) across all years where data are available. This example was produced using estimates from a lowess model (with smoothing parameter, f=2/3) of fork length x Julian day. (JC=Junction City, PT=Pear Tree, WC=Willow Creek).

	Chinook, age 0				Coho, age 0				Steelhead, age 0		
Year	**JC**	**PT**	**WC**		**JC**	**PT**	**WC**		**JC**	**PT**	**WC**
1993	-	-	79.8		-	-	65.3		-	-	58.9
1994	-	-	81.9		-	-	66.9		-	-	65.2
1995	-	-	90.2		-	-	NA		-	-	58.7
1996	-	-	83.2		-	-	81.9		-	-	62.9
1997	*	-	88.5		*	-	70.9		*	-	63.0
1998	*	-	87.2		*	-	71.5		*	-	51.4
1999	*	-	86.0		*	-	66.8		*	-	58.1
2000	*	-	88.7		*	-	76.9		*	-	61.9
2001	*	-	81.3		*	-	NA		*	-	57.7
2002	83.9	-	83.3		71.8	-	66.2		58.8	-	58.9
2003	82.1	82.1	87.1		68.9	71.6	71.0		58.2	52.7	58.5
2004	70.9	70.7	79.1		69.1	66.5	67.8		55.0	48.0	59.1
2005	-	NA	80.1		-	NA	70.7		-	NA	55.9
2006	-	72.5	80.8		-	74.2	75.5		-	49.4	57.2
2007	-	74.9	-		-	75.8	-		-	55.9	-

*The 1997–2001 Junction City data can be included once the naming convention has been resolved (Appendix E).
(-) represent years with no data
(NA) represent years with data, but where the Julian date selected occurred outside the range of observed data.

7.2.2 Condition:

Annual average condition factor

The annual average condition factor is shown for all species, sites, and years where sufficient data exist (Table 7.10). In the case of Chinook we did not report a wild only estimate of condition. This is because the estimates of condition factor were generated for an individual fish and then averaged. It isn't possible to obtain fish specific estimates of wild only weight or fork length data, instead a daily average of each metric would have to be generated (Section 7.1.5) and it wasn't clear whether this would produce useful information.

Table 7.10. Annual average condition factor for natural origin coho and steelhead, and for not-clipped Chinook. (JC=Junction City, PT=Pear Tree, WC=Willow Creek).

	Chinook, age 0			Coho, age 0			Steelhead, age 0	
Year	**PT**	**WC**		**PT**	**WC**		**PT**	**WC**
1993 - 2003	-	-		-	-		-	-
2004	-	0.974E-05		-	1.75E-05		-	1.21E-05
2005	-	1.09E-05		-	1.13E-05		-	1.12E-05
2006	1.05E-05	1.12E-05		1.07E-05	1.1E-05		1.09E-05	1.15E-05
2007	1.1E-05	-		1.11E-05	-		1.15E-05	-

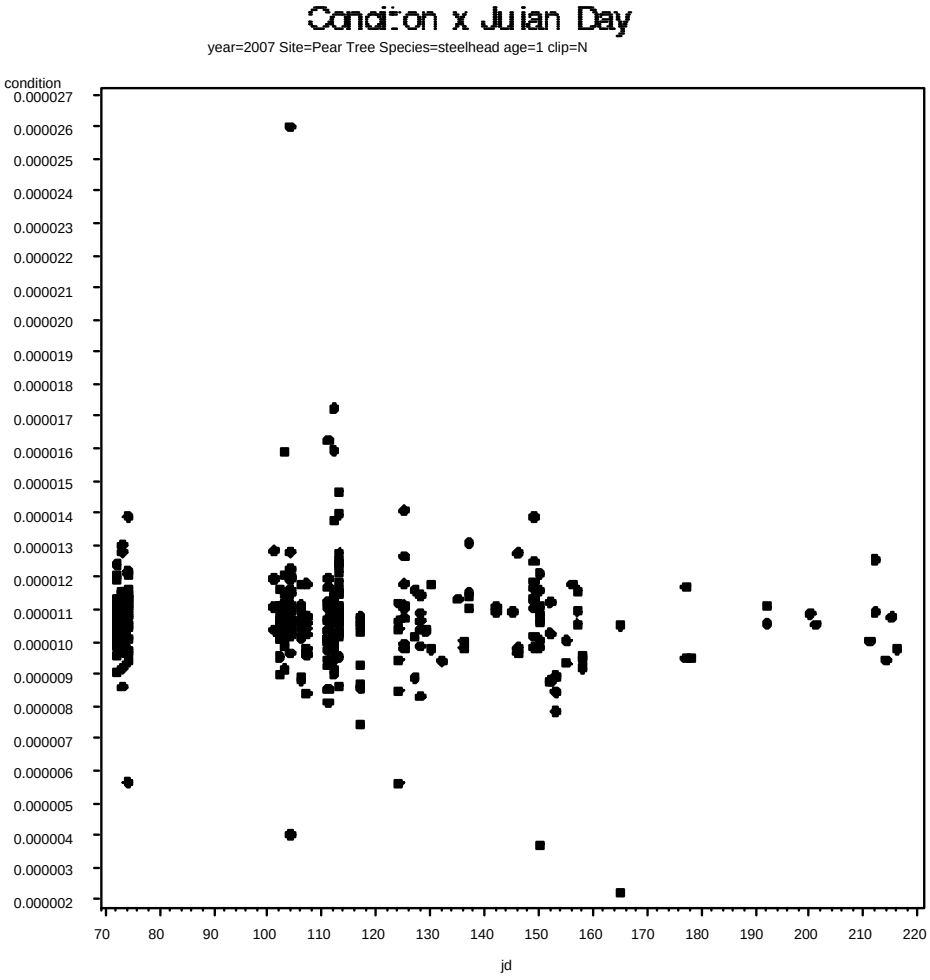


Figure 7.3. An example of the raw observed condition factor, for Pear Tree, age 1, natural steelhead, 2007.

Date of change in variance

Using the existing data we are unable to assess the value of the proposed variance based metric. There is no obvious change point in the variability (Figure 7.4). Figure 7.3 is somewhat misleading because there is very little data later in the year so it appears to have reduced variance, but with only a single data point it is impossible to evaluate. If this metric is believed to have merit and has been useful elsewhere then it would be helpful to collect more weight data during the period where the change-point is expected to occur.

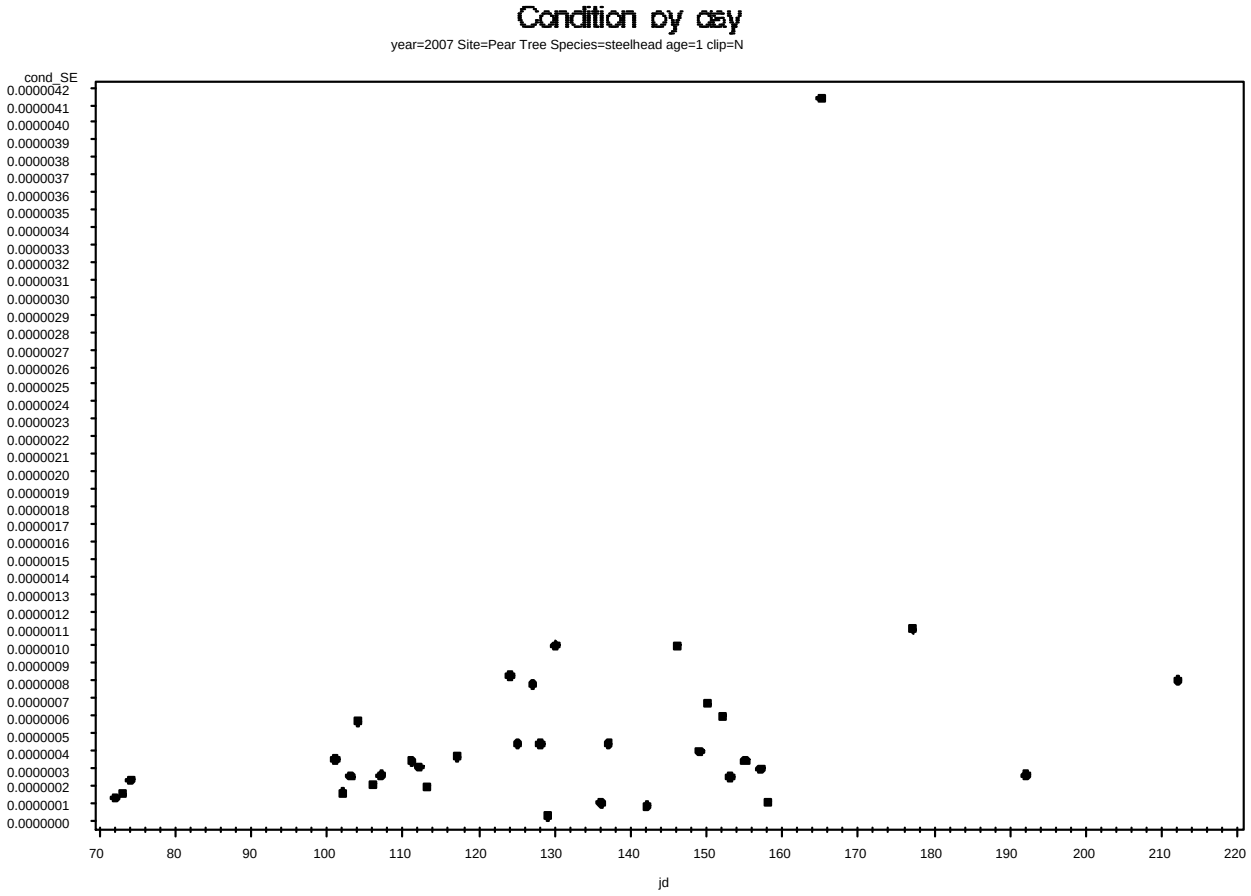


Figure 7.4. The daily standard error of the mean for condition factor plotted against Julian day. Note that there was insufficient data to estimate the variance on many days.

7.3 Discussion and Guidance

As described above the IAP working groups are currently investigating possible indicators of condition. We hope that this chapter will provide a good summary of the available data as well as some ideas about the derivation and computation of possible metrics. The techniques used to separate wild and hatchery Chinook can theoretically be used for any metric, but are limited in that they will only provide a daily estimate, not a fish specific estimate, and as is seen in Figure 7.2 the variability in daily estimates increases when both hatchery and natural fish are present. The methods described in the fork length section are applicable to other measures of fish condition and illustrate the variety of strategies that can be used to generate such metrics. The examples simply provide illustration of the methodologies. We expect

that the IAP working groups will improve upon our examples and define detailed hypotheses to test and derive specific metrics. A general recommendation is to remember that it is more important to describe the average behavior for a given time period (e.g. Julian day or week) rather than the exact value on a given day and year when conducting multi-year analyses that test Program hypotheses. This means that some method of smoothing should be incorporated, especially to account for the additional variation that results from separating the hatchery and wild Chinook salmon.

8. Outmigrant Data as a TRRP program assessment tool

8.1 General Overview

The majority of this report examined methods to obtain within-year estimates of fish condition (e.g. fork length at a specific Julian date), abundance (e.g. wild and hatchery outmigrant populations) and run timing (e.g. date when 50% of the outmigrant population passed a screw-trap). In this section, methodology will be illustrated on how to use the collected time series to evaluate trends across years. These methods can be used by the TRRP and Program partners to assess program hypotheses that are described in the Integrated Assessment Plan (TRRP & ESSA 2008).

The across-year trends that can be evaluated are of several types:

- Step changes, e.g. is the mean level of the response variable the same across groups of years classified by flow regime (high/low flow years) or type of hydrograph (with and without a bench);
- Gradual changes (usually linear) where the mean level of the response variable gradually changes over time;
- Combination of step and gradual changes (e.g. is there evidence of a difference in the mean response over time after adjusting for different flow conditions).

The standard statistical tool to evaluate these trends is the linear model. This model is very flexible and able to incorporate both step, linear, and combinations of step and linear changes. Statistical software for the linear model is readily available (e.g. the *lm()* function in *R*) and details on the fitting process (e.g. least squares) will not be provided here.

For example, a linear model to test for a time trend is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where Y_i is the response variable; β_0 the intercept; β_1 the long-term slope; X_i a covariate (such as year); and ε_i is random variation. The ε_i are assumed to be independent of each other (i.e. observations in different years are independent of each other) with a constant variance (spread about the fitted line) over time.

The variance of the response variable consists of two parts – sampling and process error (Figure 8.1). The sampling error arises in each year because only a portion of the outgoing fish can be sampled and measured. This sampling error can be quantified within each year by the standard error of the estimate. However, for long-term evaluations, the process error is often as large (or larger) than sampling error. Process error is the variation of the response variable about the trend line even if perfect information were available in each year, i.e. if perfect information (no sampling error) were available, the response variable still would not lie perfectly on the trend line. Process error cannot be estimated from within-year studies. The combination of process and sampling error is what determines the precision of the estimated trend line and controls the power to detect trends. For example, if process error is large, even perfect information in each year may not be sufficient to detect trends over short time series.

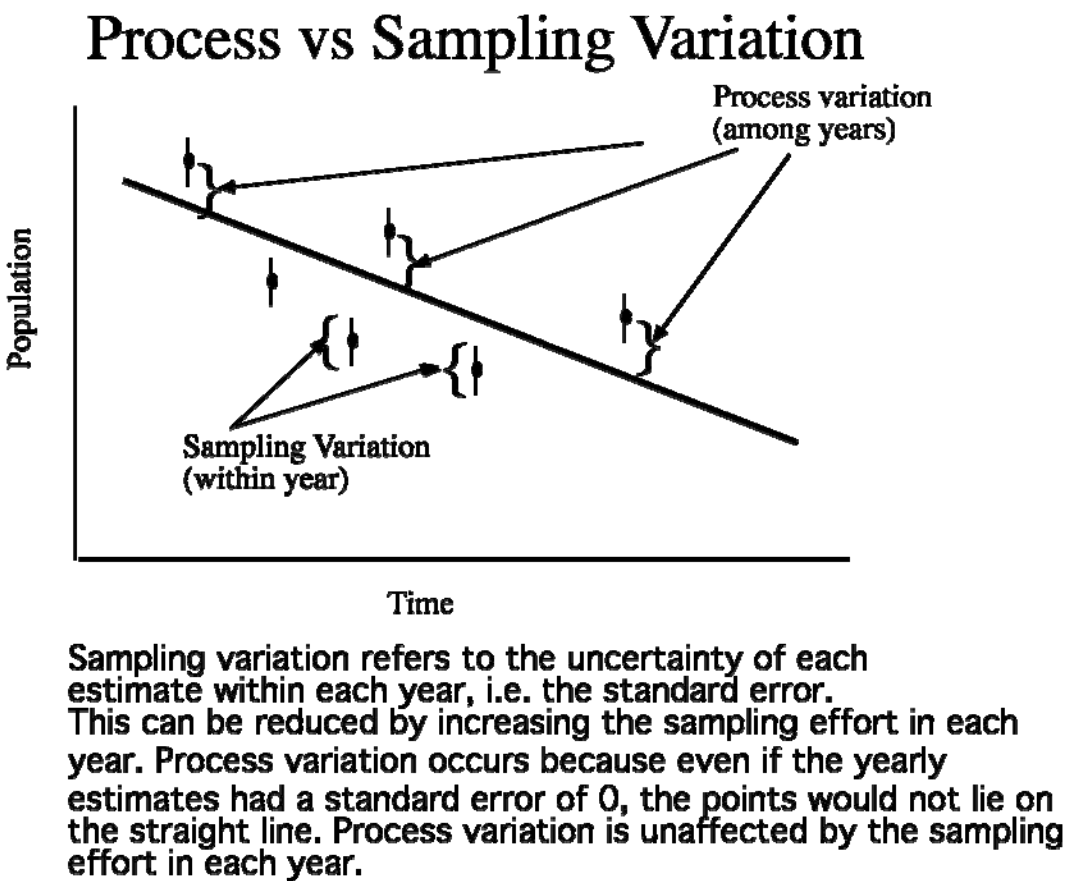


Figure 8.1 Separation of sampling and process error.

Because different fish are handled each year, it is plausible that sampling errors are independent across time. Process error is typically generated by uncontrollable year-specific factors. The process error may not be independent across time. For example, an El Nino event may depress the response variable (i.e. below the trend line) for several years in a row; if the time series is long enough, the response variable from the progeny of brood years may be related (e.g. stock-recruitment dynamics). The combined sampling and process error may not then be independent across years. Dealing with non-independence is not covered in this report because generally the time-series considered are short and the studies will have little power to detect this non-independence.

Another assumption made is that the process + sampling variation is the same across years. In many cases, sampling variation is small relative to process variation so changes in effort in individual years (which affect the standard error of the individual estimates) will have little effect. This assumption is assessed in the usual way (e.g. looking at residual plots). If this assumption is violated (i.e. process +sampling variation is not equal across years), then estimates of trend remain unbiased, but there is a loss in efficiency (i.e. the power to detect trends is reduced) if ordinary least-squares are used. A more proper analysis would use weighted least squares where the weights are proportion to the inverse of the variation (i.e. years with smaller process + sampling error would be given more weight). Unless the differences in variation are extreme, there is usually little benefit to using a weighted regression.

A related question in the analysis of across-year data is power. Power is defined as the probability that the study will detect an effect of a certain size. There are several aspects of studies that will affect power.

First is the size of the effect with larger effects being easier to detect than smaller effects. Second is the variability of the responses where studies with larger variation in the data having lower power than studies with smaller variation in the data. This variation is a combination of process and sampling error. Third is the sample size with longer studies (i.e. a larger sample size) having greater power to detect differences than smaller studies. Lastly, power also depends on the significance level (the α level) adopted with smaller α levels (e.g. α=.01 vs α=.05) making it harder to detect differences and resulting in a lower power. There is no absolute standard for an adequate power for a study, but two common goals are to achieve a power of about 80% at α=.05; or a power of 90% with α=.10.

Power analyses can also be conducted prospectively and retrospectively. In prospective power analyses, estimates of required sample sizes are obtained for future studies. In retrospective power analyses, the power for an existing study is obtained. As noted by Gerard et al (1998) and Steidl et al (1997), retrospective power analyses are not recommended because there is actually no new information obtained. It can be shown mathematically that if the result of a statistical test was negative, a retrospective power analysis will report low power and if the result of a statistical test was positive, the retrospective power analysis will report high power. However, prospective power analyses are very useful to provide the researcher with guidance on how long the study must continue to detect effects of interest.

The theory on power analysis in linear models is outlined in Morrison (1983). In the case of detecting trends over time, the computations simplify considerably as outlined in Gerrodette (1987). Let $\{T_i; i=1,\ldots n\}$ be the times at which the study is to be conducted. For example, a study that is conducted in 10 consecutive years would have $T_1 = 1$, $T_2 = 2$, …, $T_{10} = 10$. [The actual values are not important as long as they are consecutive.] If a study is conducted in alternate years, then one possible set of values could be $T_1 = 1$, $T_2 = 3$, etc. Let σ^2_{resid} be the residual variance, i.e. the variance of data points about the trend line, which should include both process and sampling error. [This is often estimated from an existing study.] Finally, let β_1^* be the effect size of interest, i.e. the slope that should be detected in the experiment.

The power of the statistical test for the slope depends upon the non-centrality parameter:

$$\lambda = \frac{\left(\beta_1^*\right)^2 \text{var}(\{T_i\})(n-1)}{\sigma^2_{resid}}$$

and the power is found as

$$Power = P(F > F_{1-\alpha,1,n-2}; 1, n-2, \lambda)$$

i.e. the probability of exceeding the critical value of the *F*-distribution from a non-central F-distribution with non-centrality parameter λ. [*R*-code to compute power is included with the report.]

Power increases with increasing values of λ. From the expression above, it is seen that power increases with increasing effect size ($|\beta_1^*|$), an increasing spread in the time points sampled which increases the $\text{var}(\{T_i\})$ term, or a reduction in the residual variance. The effect of the alpha level is found in the increase in the critical F-value as alpha declines.

This expression gives the power for a given set of parameters. It can be solved in reverse to find the number of years required to obtain a specified power.

The above expressions can also be used for a power analysis of step changes, e.g. is the mean response different between high/low water years or between years where the hydrograph has a bench or doesn't have a bench. The non-centrality parameter and power are computed in the identical ways except that the

T values are now coded a 0 or 1 corresponding to the two groups. For example, in a 10 year study where even years had a bench in the hydrograph and odd years did not, then $T_1 = 1$, $T_2 = 0$, $T_3 = 1$, etc. The sample size of the two groups is implicitly given in the number of 1's and 0's. Because of the assumption of independence across years, it does not matter in which order the two classes occur over time.

An online power calculator is also available at Lenth (2009).

R-code and full details on the analyses are available at the web site:
http://www.stat.sfu.ca/~cschwarz/Consulting/Trinity/Phase2

In this chapter we present several examples of how the Trinity River outmigrant data may be used to assess the restoration program. A variety of analyses are illustrated including: trend analysis, incorporating covariates to explain year to year variability, comparing between different year types, and associated power analyses. Working with the TRRP we described several hypotheses of interest which could be tested using the existing data. Many more hypotheses have been or will be described by the IAP and the associated working groups.

8.2 Across-year analysis of fork length.

As outlined in Chapter 7, the mean daily fork length was computed within each year at a site for a species by fitting a lowess (a non-parametric smoother) to the mean daily wild YOY sampled fish (e.g., Figure 8.2).

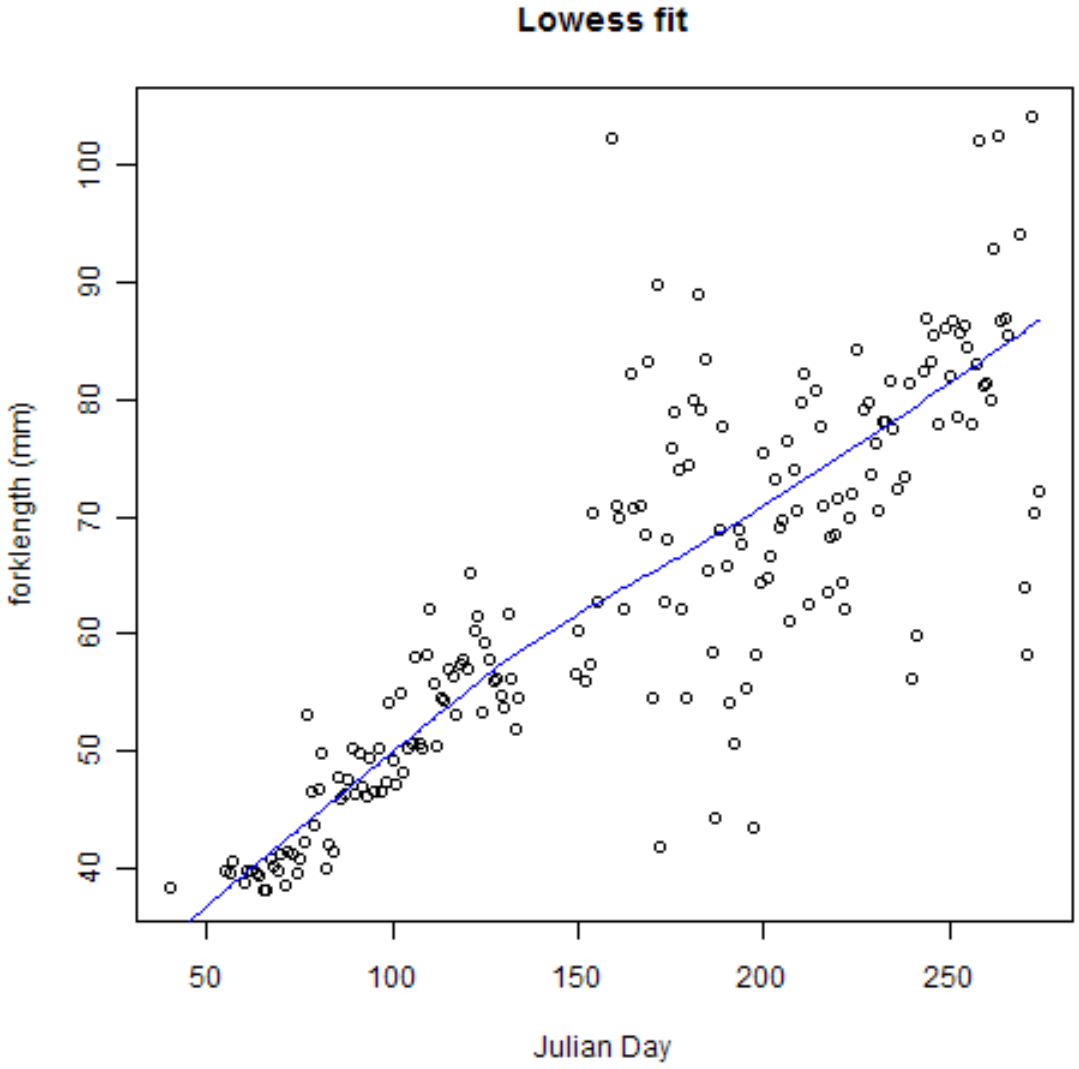


Figure 8.2 An illustration of a lowess fit to the mean daily fork length of wild YOY fish. This data represents Chinook in Junction City in 2004.

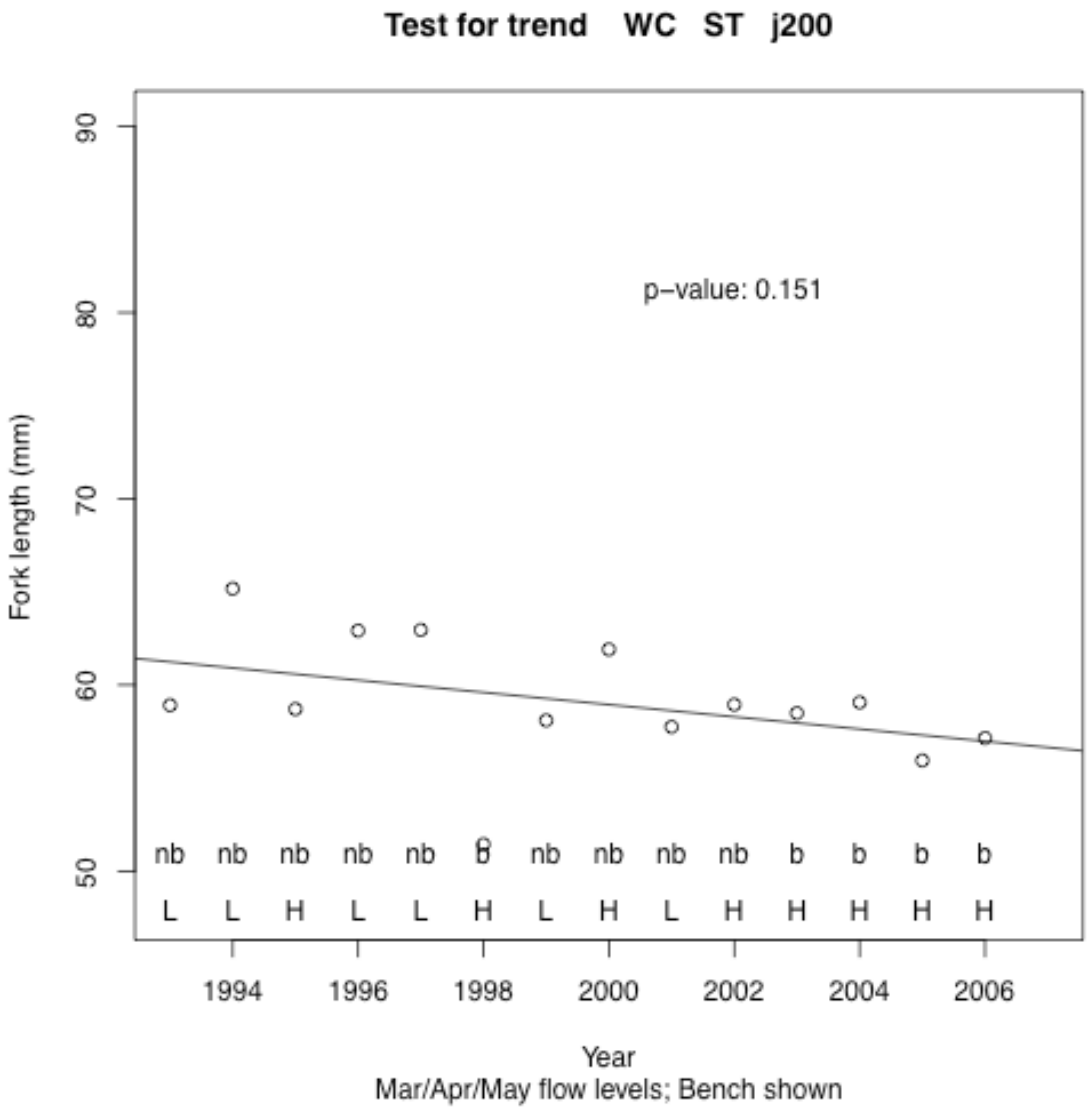
The lowess line was used to estimate the mean fork length at Julian day 100 (early in the season) and Julian day 200 (late in the season) at the study sites over the years where data were available.

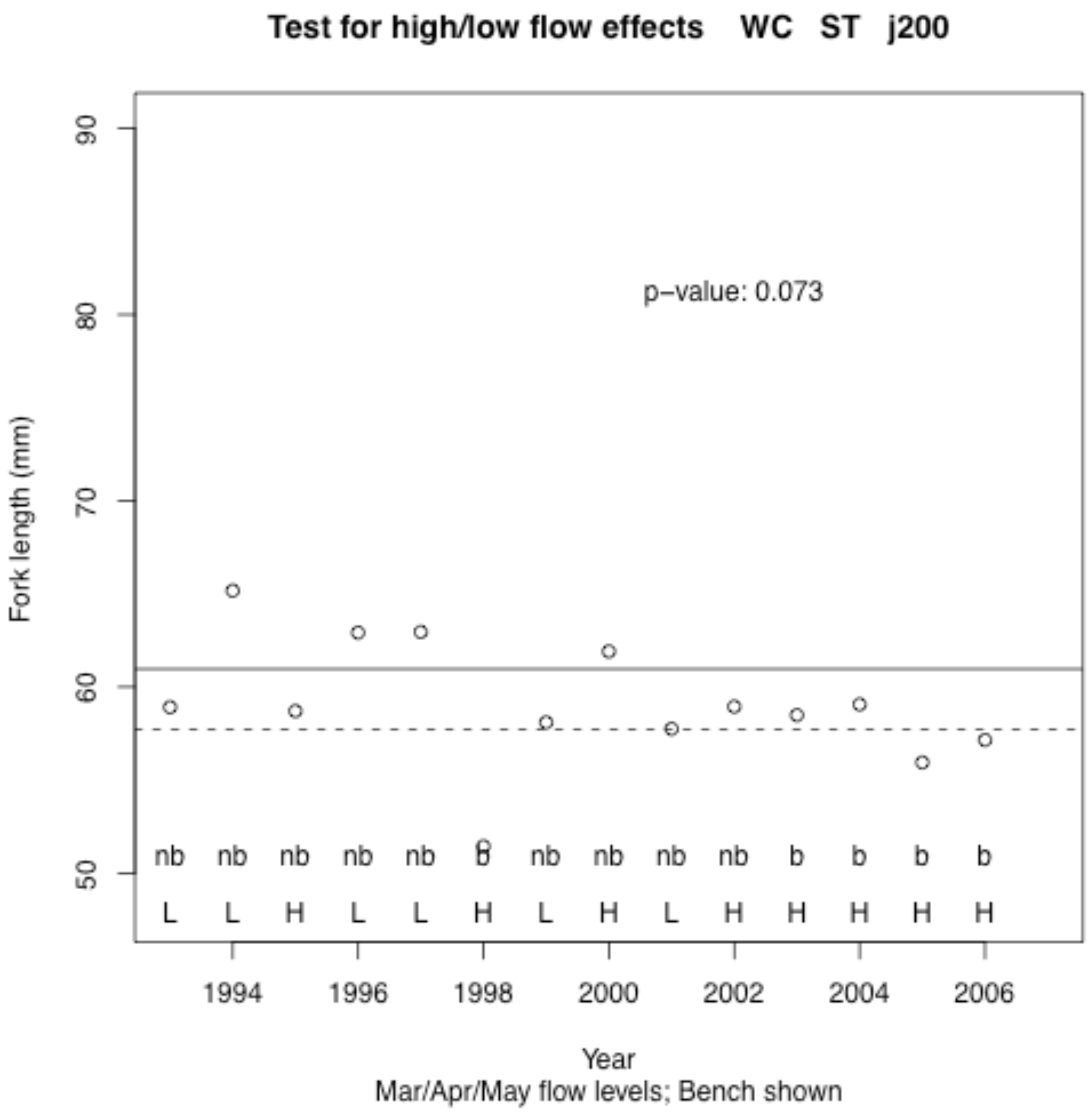
Three across-year tests for trend were conducted for changes in the mean fork length at each of the two Julian days for each site-species combination:

- A linear trend;
- A comparison of the mean fork length in years with high and lower water levels in the March-May portion of the year;
- A comparison of the mean fork length in years when the hydrograph did and did not have a bench.

An illustration of the fitting and testing process is found in Figure 8.3(a), (b), and (c). Notice that the variability seen in the plots about the trend line is a reflection of process and sampling error and appears

to be roughly constant over time. A complete set of graphs is available at the web-site. A summary of the results for all sites species and two time steps is found in Table 8.1.


Test for trend WC ST j200
p−value: 0.151
Fork length (mm)
90
80
70
60
50
nb nb nb nb nb b nb nb nb nb b b b b
L L H L L H L H L H H H H H
1994
1996
1998
2000
2002
2004
2006
Year
Mar/Apr/May flow levels; Bench shown


Test for high/low flow effects WC ST j200
p−value: 0.073
Fork length (mm)
90
80
70
60
50
nb nb nb nb nb b nb nb nb nb b b b b
L L H L L H L H L H H H H H
1994
1996
1998
2000
2002
2004
2006
Year
Mar/Apr/May flow levels; Bench shown

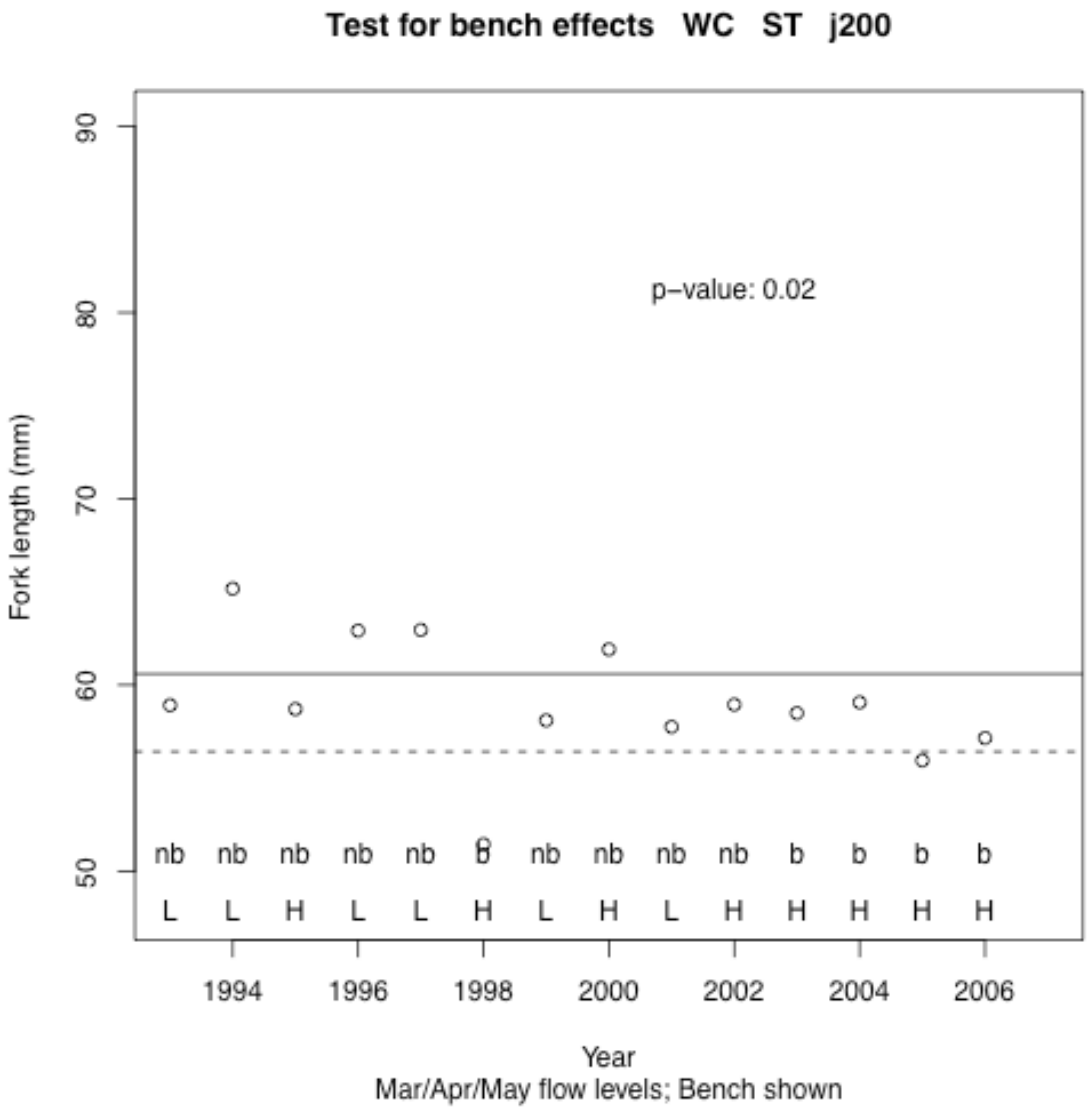


Figure 8.3 (a), (b) and (c). Plots of the estimated mean fork length at Julian day 100 for Willow Creek for Chinook. In each plot, the linear trend or the step function model was fit, and p-value for a test of no linear trend, or no difference in mean fork length between the two classes is displayed.

Table 8.1 Summary of test for changes in mean fork length across years at two different Julian dates. Analyses with statistically significant (or close to statistical significance) are in bold.

		Julian day 100			Julian day 200		
Site	**Species**	**Linear trend n Slope[1] (SE) p-value**	**High/Low Flow n_H Diff[2] (SE) p-value**	**Bench/Non-bench n_B Diff[3] (SE) p-value**	**Linear trend n Slope (SE)[1] p-value**	**High/Low Flow n_H Diff (SE)[2] p-value**	**BenchNon-bench n_B Diff (SE)[3] p-value**
Junction City	Coho	2	2	1	3	3	**2**
		-7.08	-	-14.17	-1.37	-	**-2.81**
		-	-	-	0.88	-	**0.13**
		-	-	-	0.36	-	**0.03**
Pear Tree	Coho	5	5	4	4	4	3
		0.07	-	-1.63	1.6	-	-5
		0.68	-	2.2	1.09	-	4.5
		0.93	-	0.51	0.28	-	0.38
Willow Creek	Coho	8	4	3	12	7	5
		0.26	2.4	0.94	0.13	1	0.6
		0.17	1.43	1.74	0.36	3.06	3.07
		0.17	0.14	0.61	0.73	0.75	0.85
Junction City	Chinook salmon	2	2	1	3	3	2
		0.06	-	0.13	-6.48	-	-7.37
		-	-	-	2.71	-	9.68
		-	-	-	0.25	-	0.59
Pear Tree	Chinook salmon	5	5	4	4	4	3
		1.03	-	-6.05	-1.26	-	0.24
		1.61	-	4.95	1.71	-	7.04
		0.57	-	0.31	0.54	-	0.98
Willow Creek	Chinook salmon	11	5	3	14	8	5
		0.4	2.23	1.38	-0.27	1.09	-1.92
		0.25	2.22	2.57	0.25	2.09	2.11
		0.15	0.34	0.60	0.31	0.61	0.38
Junction City	Steelhead	2	2	1	3	3	2
		-1.45	-	-2.9	-1.88	-	-2.17
		-	-	-	0.74	-	2.74
		-	-	-	0.24	-	0.57
Pear Tree	Steelhead	4	4	4	4	4	3
		5.55	-	-	0.79	-	-5.85

		Julian day 100			Julian day 200		
Site	**Species**	**Linear trend n Slope[1] (SE) p-value**	**High/Low Flow n_H Diff[2] (SE) p-value**	**Bench/Non-bench n_B Diff[3] (SE) p-value**	**Linear trend n Slope (SE)[1] p-value**	**High/Low Flow n_H Diff (SE)[2] p-value**	**BenchNon-bench n_B Diff (SE)[3] p-value**
		2.23	-	-	1.25	-	2.77
		0.13	-	-	0.59	-	0.17
Willow Creek	Steelhead	4	4	3	14	**8**	**5**
		20.62	-	16.95	-0.33	**-3.26**	**-4.18**
		12.72	-	48.49	0.21	**1.66**	**1.56**
		0.25	-	0.76	0.15	**0.07**	**0.02**

[1] If there are only 2 points for a linear trend, no estimate of the standard error or p-value is available.
[2] If all or none of the years were high flow, then there is no contrast in the data and no estimate of the flow effect is available. If there are no replicates, estimates of the standard error or of the p-value are not available.
[3] If all or none of the hydrographs had a bench, then there is no contrast in the data and no estimate of the bench effect is available. If there are no replicates, estimates of the standard error or of the p-value are not available.

There are two major challenges in the cross-year assessment. First is the small number of years for which data are available especially at the Junction City site. In many cases, no sensible estimates can be obtained. Second, is the lack of contrast, especially for testing the effect of high vs. flow regimes where at Junction City and Pear Tree sites, all the data are from a single flow regime. No estimates of the effect of flow regime are then possible.

The Willow Creek site had a sufficiently long time series at Julian day 200. There was evidence of a negative effect of the high flow regime and the bench upon the mean fork length. These are illustrated in Figure 8.3(a) (b) and (c). Notice that because of the high association between the flow level (H or L) and hydrograph (Bench or non-bench), it will be very difficult to disentangle the individual effects of these two factors.

Because of very small sample sizes, a power analysis was conducted based on the results of the analysis only from the Willow Creek site with measurements taken at Julian day 200 (

Table 8.2) For example, the power to detect a 10% change over 10 years (i.e. a 1% change/year) is only 43% for Chinook salmon. With about 15 years of data at Willow Creek, power is reasonably high (exceeds 80%) to detect a 1% change/year for Chinook salmon and steelhead but appreciably lower (around 50%) for coho. The reduction in power for coho is caused by the much larger process + sampling error (5.2 mm vs. about 4 mm for Chinook salmon and steelhead).

Table 8.2 Estimated power to detect trend in mean fork length over time based on Willow Creek Julian day 200 results.

	5% over study	**10% over study**	**15% over study**	**20% over study**
Coho (base mean 71 mm; process + sampling std dev 5.2 mm)				
5 years	0.06	0.10	0.15	0.23
10 years	0.08	0.19	0.37	0.58
15 years	0.11	0.29	0.56	0.80
20 years	0.13	0.38	0.70	0.91
Chinook salmon (base mean 84 mm; process + sampling std dev 3.7 mm)				
5 years	0.08	0.17	0.32	0.49
10 years	0.15	0.43	0.76	0.94
15 years	0.21	0.64	0.93	1.00
20 years	0.28	0.78	0.98	1.00
Steelhead (base mean 59 mm; process + sampling std dev 3.2 mm)				
5 years	0.07	0.13	0.23	0.36
10 years	0.11	0.31	0.59	0.83
15 years	0.16	0.47	0.81	0.96
20 years	0.2	0.61	0.92	0.99

A similar power analysis can be conducted for a simple two group comparison (e.g. high vs. low flow years; bench vs. non-bench hydrographs) (Table 8.3). For example, the power to detect a 10% difference in mean fork length in a 10 year study with 5 years in one group and the other 5 years in the other group for Chinook salmon is 87%. In a 15 year study, the power is quite high (exceeds 90%) to detect a 10% difference in mean fork length (under an equal allocation) for Chinook salmon and steelhead; power is again round 50% for Coho. The reduction in power for Coho is caused by the much larger process.+.sampling error (5.2 mm vs. about 4 mm for Chinook salmon and steelhead).

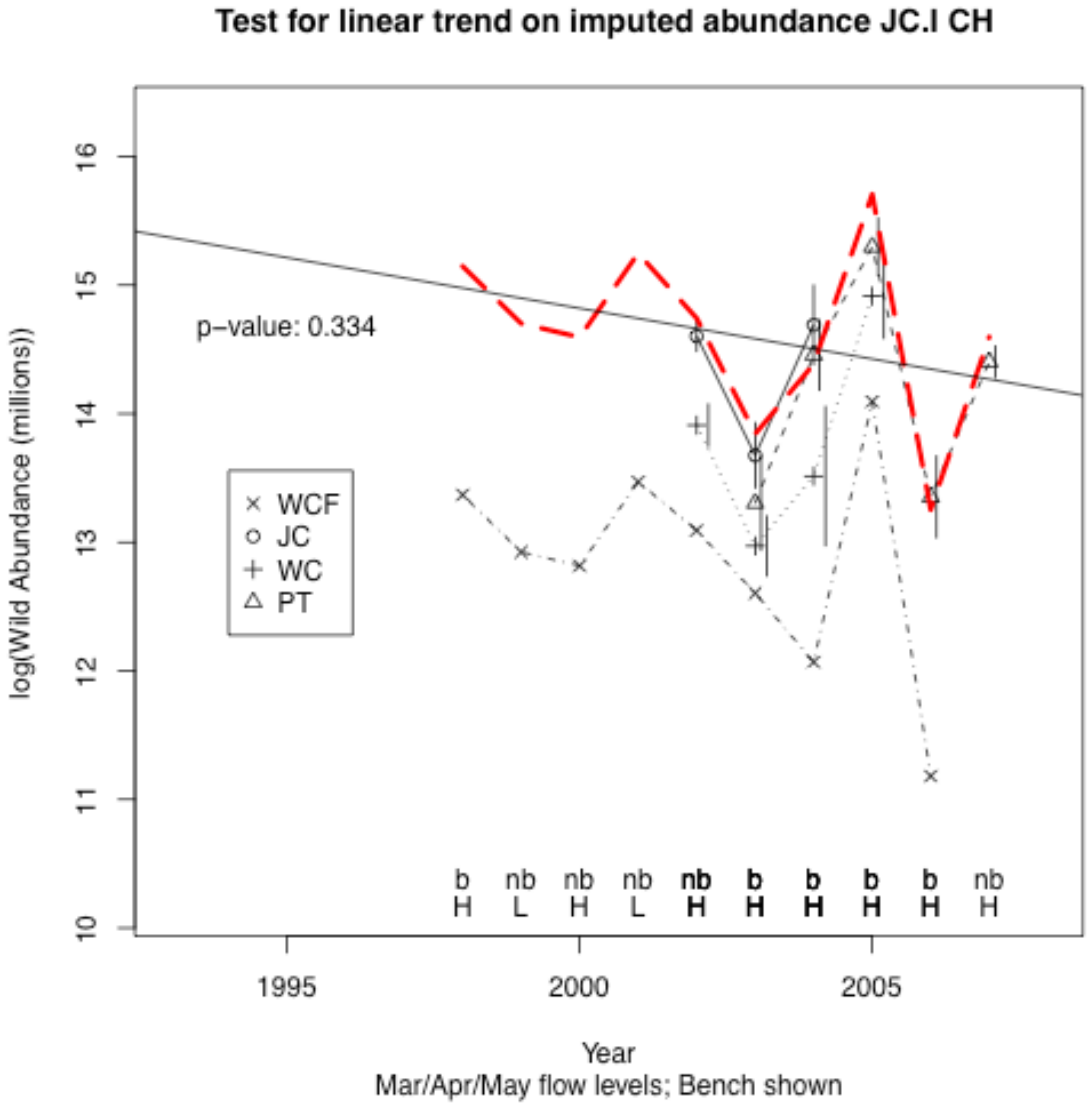
Table 8.3 Estimated power to detect change in mean fork length over time between two groups based on Willow Creek Julian day 200 results. Sample years were allocated equally as possible over the years.

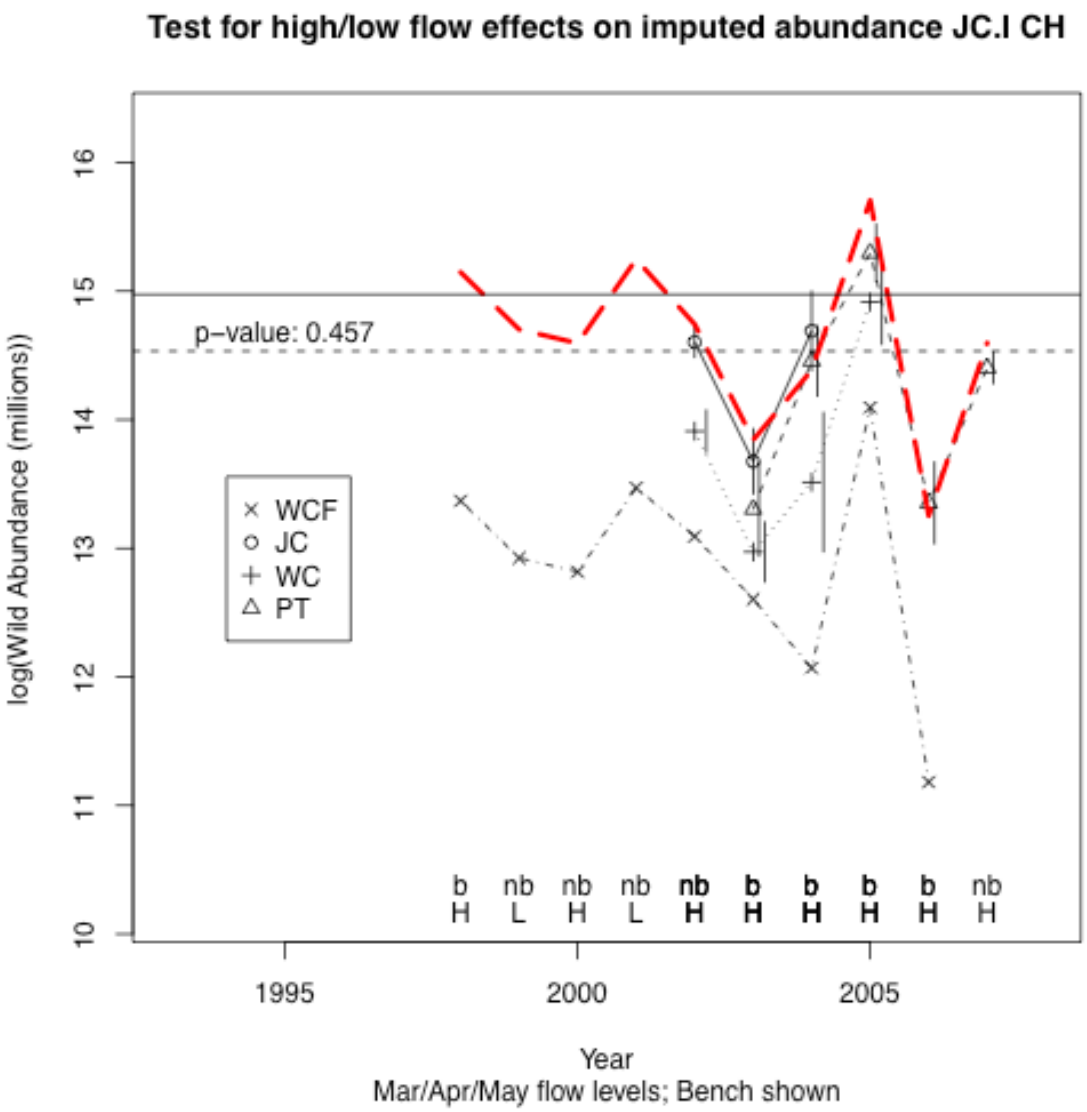
	5% difference	**10% difference**	**15% difference**	**20% difference**
Coho (base mean 71 mm; process + sampling std dev 5.2 mm)				
5 years	0.05	0.10	0.15	0.2
10 years	0.08	0.19	0.34	0.53
15 years	0.16	0.47	0.81	0.96
20 years	0.23	0.68	0.95	1.00
Chinook salmon (base mean 84 mm; process + sampling std dev 3.7 mm)				
5 years	0.14	0.4	0.69	0.89
10 years	0.35	0.87	1.00	1.00
15 years	0.52	0.98	1.00	1.00
20 years	0.66	1.00	1.00	1.00
Steelhead (base mean 59 mm; process + sampling std dev 3.2 mm)				
5 years	0.11	0.29	0.53	0.76

10 years	0.25	0.72	0.97	1.00
15 years	0.37	0.90	1.00	1.00
20 years	0.49	0.97	1.00	1.00

8.3 Across-year evaluation of abundance

Figure 8.4(a), (b) and (c) presents a plot of the estimates of log(abundance) of the wild YOY Chinook salmon population based on the spline model and the sampled-discharge flow models. There were only a handful of years where estimates of steelhead abundance were available and no years where abundances of coho were available; only the Chinook salmon abundance will be considered further.




Test for high/low flow effects on imputed abundance JC.I CH
log(Wild Abundance (millions))
p−value: 0.457
WCF
JC
WC
PT
b H
nb L
nb H
nb L
nb H
b H
b H
b H
b H
nb H
1995
2000
2005
Year
Mar/Apr/May flow levels; Bench shown

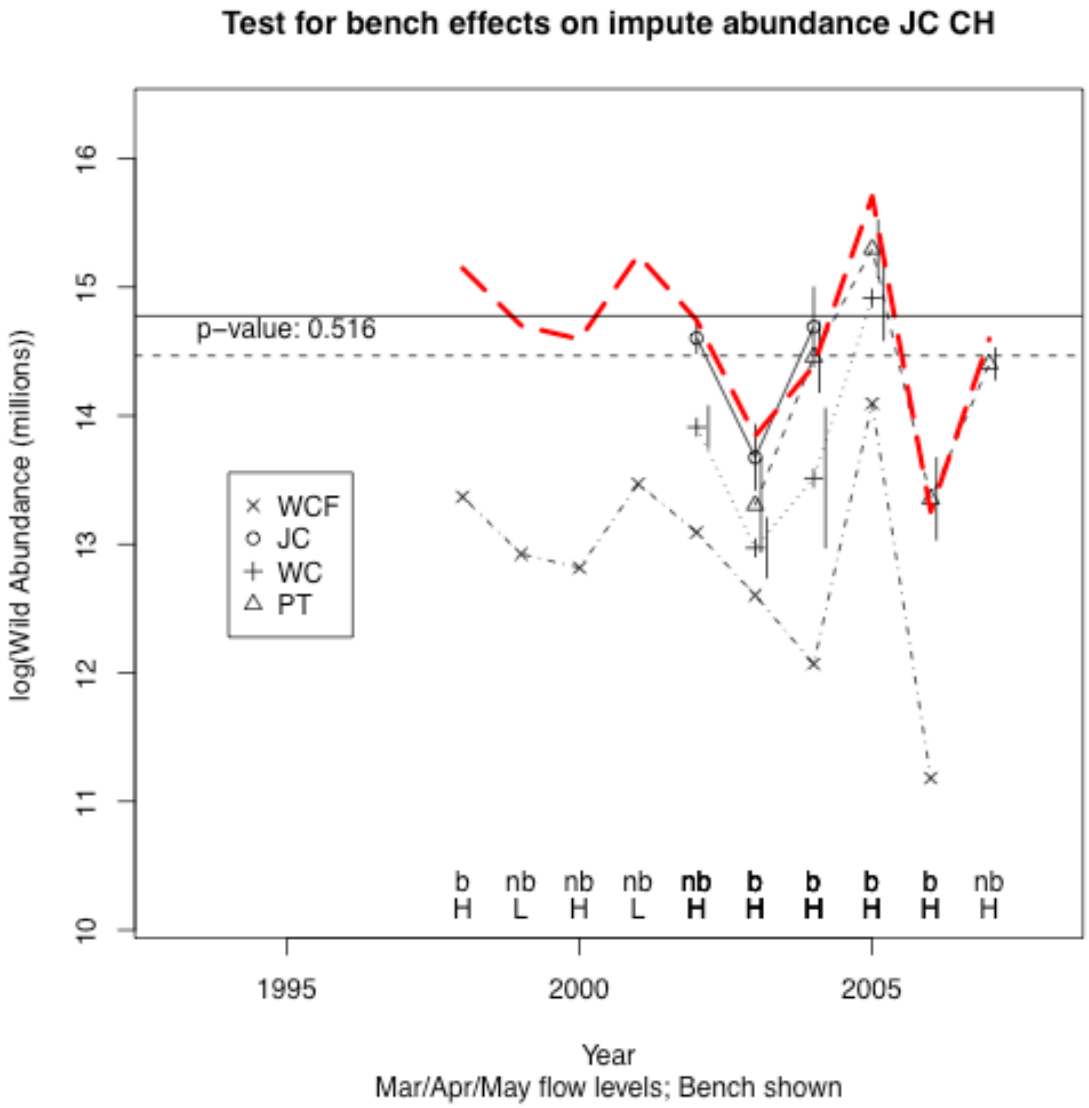


Figure 8.4 Summary of test for (a) linear trend in log(wild YOY abundance), (b) changes in mean log(wild YOY abundance) between high and low flow years, and (c) changes in mean log(wild YOY abundance) between years with a bench in the hydrograph and with out bench. WCF indicates the sampled-discharge flow-based estimates from Willow Creek. The heavy (red) line is the imputed log(abundance) at Junction City based on a simple block model. Error bars are 95% confidence intervals.

The pattern of the spline-based estimates, show a high degree of parallelism which is not surprising given that they all measure mostly the same population (albeit at different parts of the river). The sampled-discharge flow-based estimates from Willow Creek are not as consistent but still show the same general pattern.

Trend analyses for each individual source using such short time series are not very informative because they are unlikely to detect anything but gross changes. The use of Analysis of Covariance (ANCOVA) models to combine the multiple sources with a common trend is not recommended because measurements

within the same year from different sources are likely to be highly correlated which violates a key assumption of the ANCOVA model.

In order to combine all sources of data, an imputed curve for Junction City both before 2002 and after 2004 is computed. This was done by fitting an anova-block model:

$$\log W_{ij} = \mu + year_i + source_j + e_{ij}$$

where $\log W_{ij}$ is the log(wild abundance) in year *i* from source *j* (the three sites and the Willow Creek flow-based estimate); μ is the overall mean; $year_i$ is the (average) effect of year *i* , i.e. the average increase or decrease over the sources; and $source_j$ is the (average) effect of source *j* upon the abundance, i.e. the average separation of the line. This was fit using least-squares ignoring the non-independence of the errors and the predicted abundance at Junction City (the red solid line) is also displayed in Figure 8.4. This imputed population abundance closely follows the Junction City estimates when data are available and extrapolates earlier and later in the series based on the trends in the other sources. A key assumption is that the flow-based estimate from Willow Creek has a strong correlation with actual abundance.

This imputed abundance was then used to test for a linear trend in the log(abundance), to test for a difference in the mean(log abundance) at high vs. low flow conditions, and to test for a difference in the mean(log abundance) in years with and without a bench in the hydrograph as summarized in Figure 8.4 and Table 8.4. None of the three hypotheses of no changes in the mean(log abundance) were statistically significant.

Table 8.4 Summary of tests for trend using the imputed log(abundance) at Junction City.

Site	**Species**	**Linear trend n Slope (SE) p-value**	**High/Low Flow n_H Diff (SE) p-value**	**Bench/Non-bench n_B Diff (SE) p-value**
Junction City	Chinook salmon	10	8	5
		-.08	-.44	.-31
		.03	.57	.46
		.33	.46	.52

It is fairly clear from Figure 8.4 that there are tremendous fluctuations in log(abundance) around the trend line or the high/low flow or bench/non-bench means. The standard errors of the estimates of abundance from the spline-based methods are also shown in Figure 8.4 and show that sampling error is small relative to the sampling + process error. As a rough approximation, the estimated sampling + process error was obtained from the linear fit and is estimated from the variance of the residuals as. 49; the average sampling variance as on the order of .02; process error comprises over 90% of the total variation seen in the log(abundances) over time. If this process error cannot be reduced (e.g. perhaps some of the variation can be explained by other factors not considered in this report), then spending considerable effort to obtain precise estimates of precision within each year has little impact on the ability to detect longer-term trends.

A power analysis on the ability to detect various rates of change based on the estimated process + sampling error from the linear fit is presented in Table 8.5. For example, a 10 year study would have only a 21% power to detect a 10% change per year (i.e. essentially a doubling of abundance over the 10

years!). The reason for this poor power is that the process + sampling variation is very large – a value of .7 on the log(abundance) scale implies that in absence of change, abundances can fluctuate by 2(.7)=1.4 on the log(abundance) scale or by a factor of exp(1.4)=4 on the raw abundance scale. Consequently it is difficult to detect event a doubling of abundance over 10 years.

Table 8.5 Estimated power (alpha=.05) to detect linear changes in mean log(abundance) based on imputed Junction City Chinook salmon values. Process + sampling standard deviation is .7.

	Percent change in abundance PER year.[1]									
	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%
5 yrs	0.05	0.05	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.10
10 yrs	0.06	0.07	0.11	0.15	0.21	0.28	0.36	0.45	0.54	0.63
15 yrs	0.07	0.14	0.27	0.43	0.60	0.76	0.87	0.94	0.98	0.99
20 yrs	0.11	0.29	0.56	0.80	0.94	0.99	1.00	1.00	1.00	1.00

[1] This was converted to an equivalent change on the log-scale when power computations were done.

Similarly, a power analysis to detect a simple change in the mean log(abundance) between two types of years (e.g. high vs. low flow; bench vs. non-bench) is presented in Table 8.6. For example, the power to detect a 30% difference in abundance between the classes based on 10 years of sampling with equal numbers of years in each class is only 9%. Again the dismal power is a result of the very large process + sampling variation which makes it hard to detect effects.

Table 8.6 Estimated power (alpha=.05) to detect simple changes in mean log(abundance) between two classes of year based on imputed Junction City Chinook salmon values. Process + sampling standard deviation is .7. Years are divided equally between the two classes.

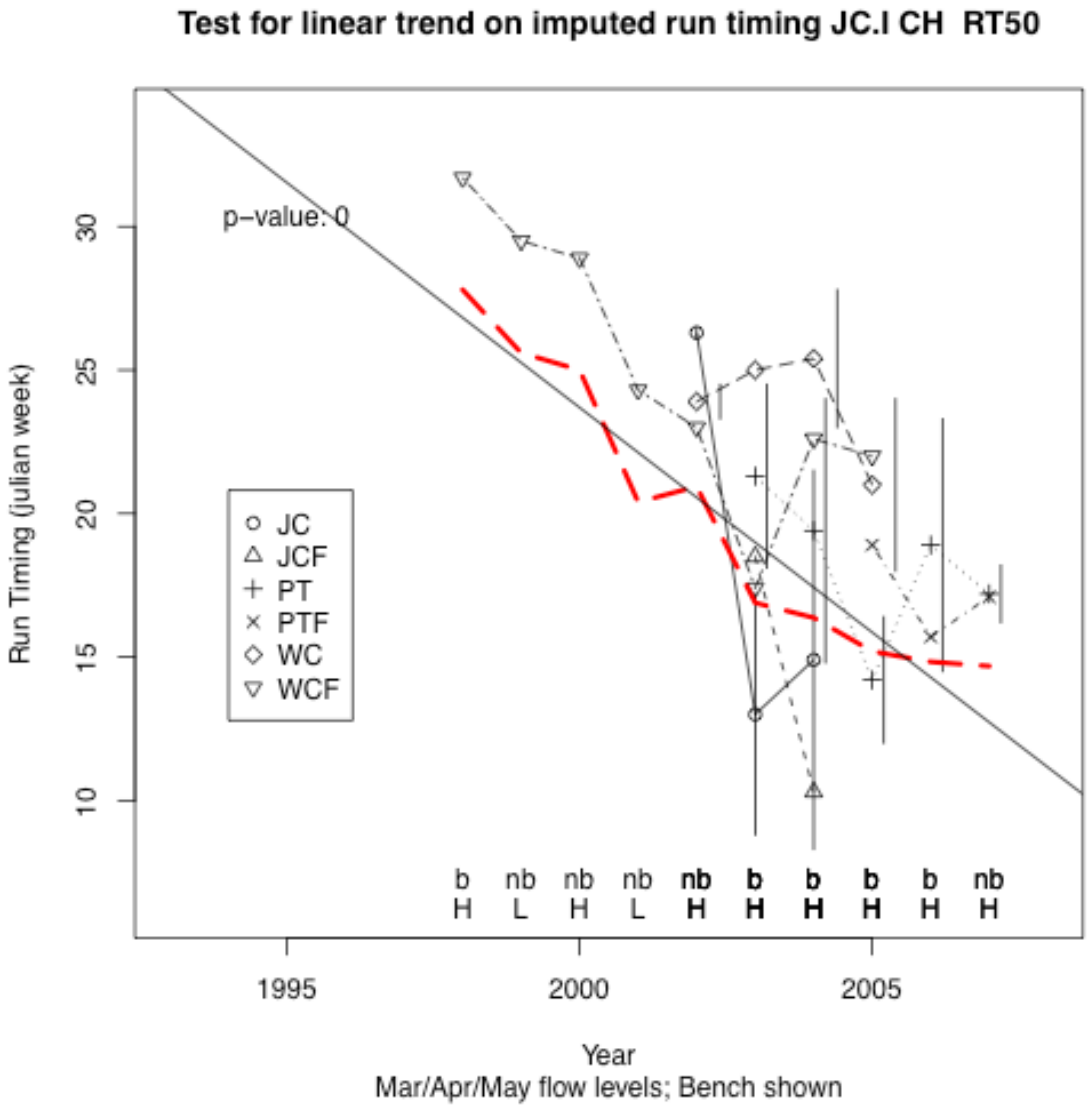
	Percent difference in abundance between two classes.[1]									
	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
5 yrs	0.05	0.05	0.05	0.06	0.06	0.06	0.07	0.07	0.08	0.09
10 yrs	0.05	0.05	0.06	0.07	0.08	0.09	0.11	0.13	0.15	0.17
15 yrs	0.05	0.06	0.07	0.08	0.10	0.12	0.15	0.18	0.21	0.25
20 yrs	0.05	0.06	0.07	0.09	0.12	0.15	0.19	0.23	0.28	0.33

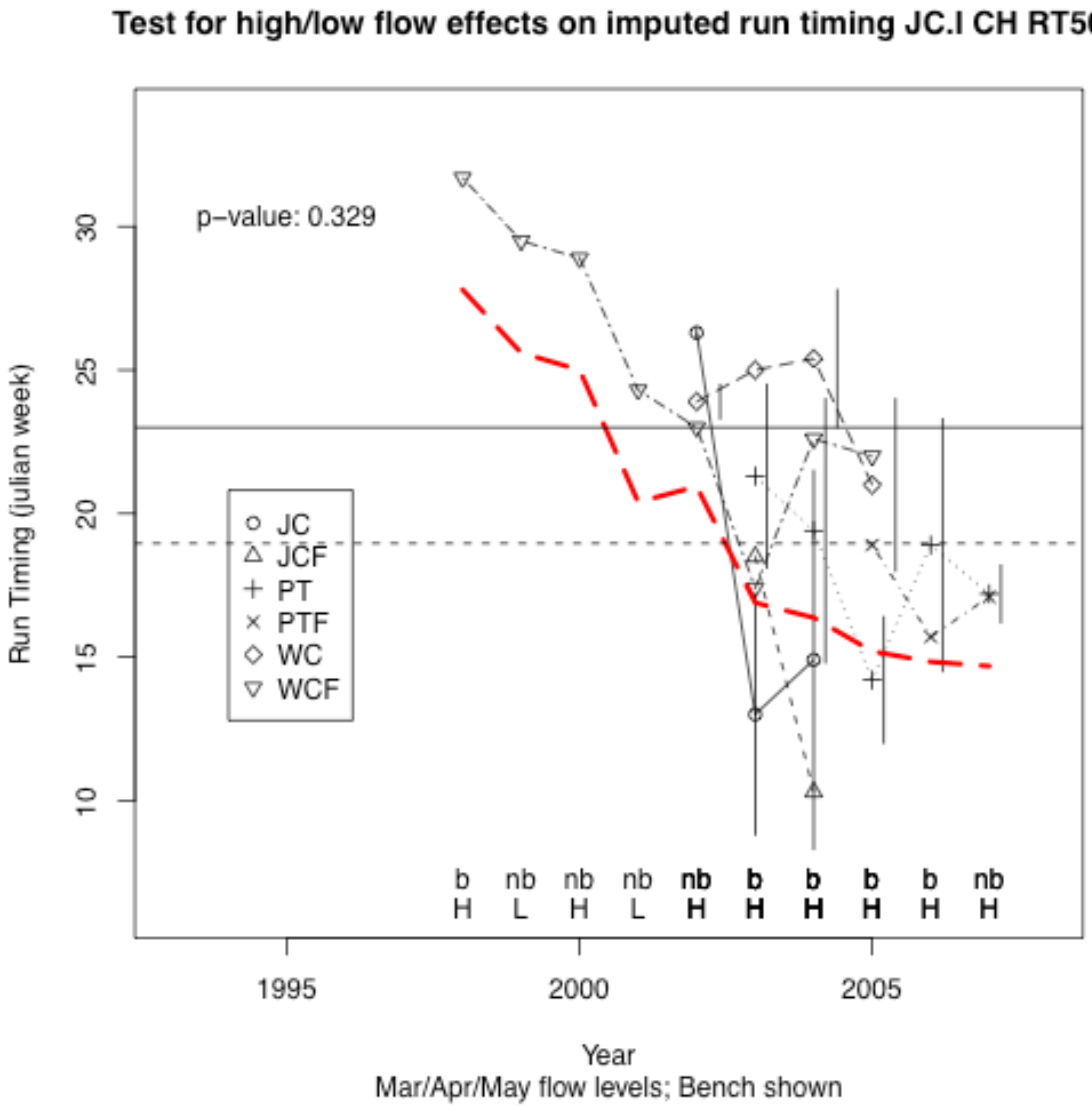
[1] This was converted to an equivalent change on the log-scale when power computations were done.

8.4 Across-year evaluation of run timing

A multi-year analysis of the run timing (Julian week by which 50% and 90% of fish have passed) was done for wild YOY Chinook salmon. Again data was sparse for the Coho and steelhead and an analysis was not done.

Figure 8.5 (a), (b), (c) and Figure 8.6 (a), (b), (c) present summary plots of the analyses for the two percentiles.


Test for linear trend on imputed run timing JC.I CH RT50
p−value: 0
JC
JCF
PT
PTF
WC
WCF
b nb nb nb nb b b b b nb
H L H L H H H H H H
Run Timing (julian week)
10
15
20
25
30
1995
2000
2005
Year
Mar/Apr/May flow levels; Bench shown


Test for high/low flow effects on imputed run timing JC.I CH RT50
p−value: 0.329
JC
JCF
PT
PTF
WC
WCF
b H
nb L
nb H
nb L
nb H
b H
b H
b H
b H
nb H
Run Timing (julian week)
10
15
20
25
30
1995
2000
2005
Year
Mar/Apr/May flow levels; Bench shown

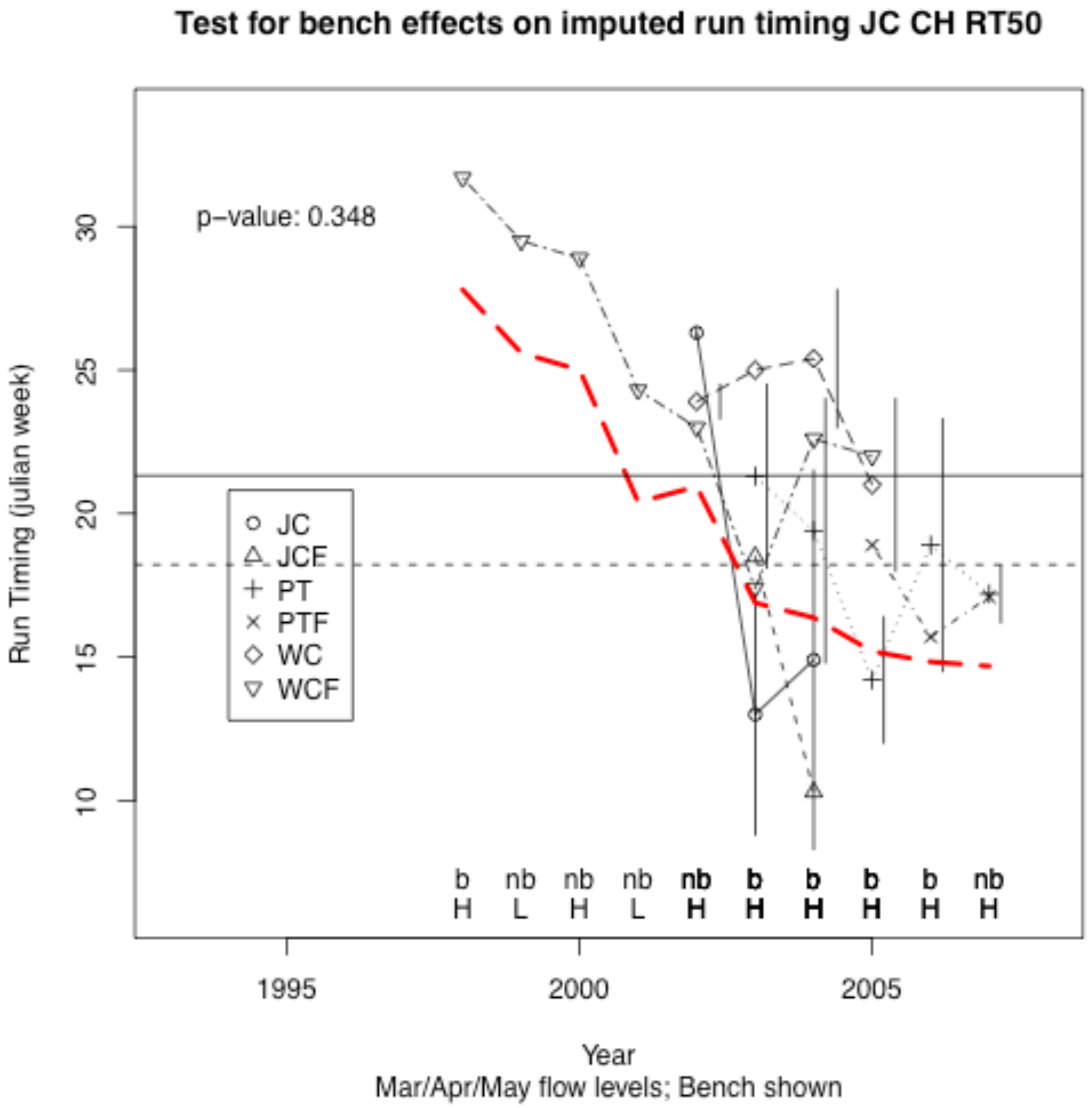
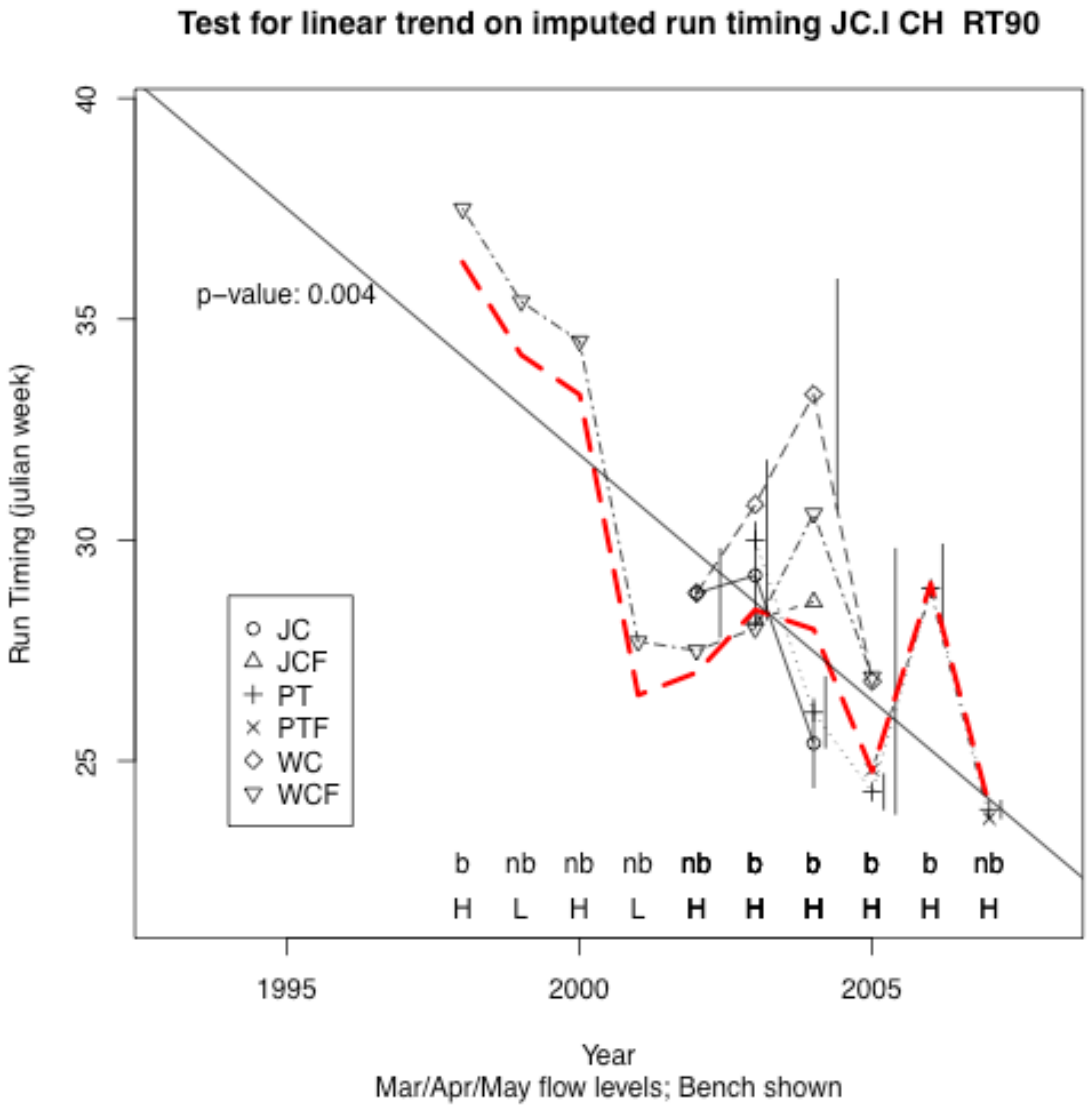
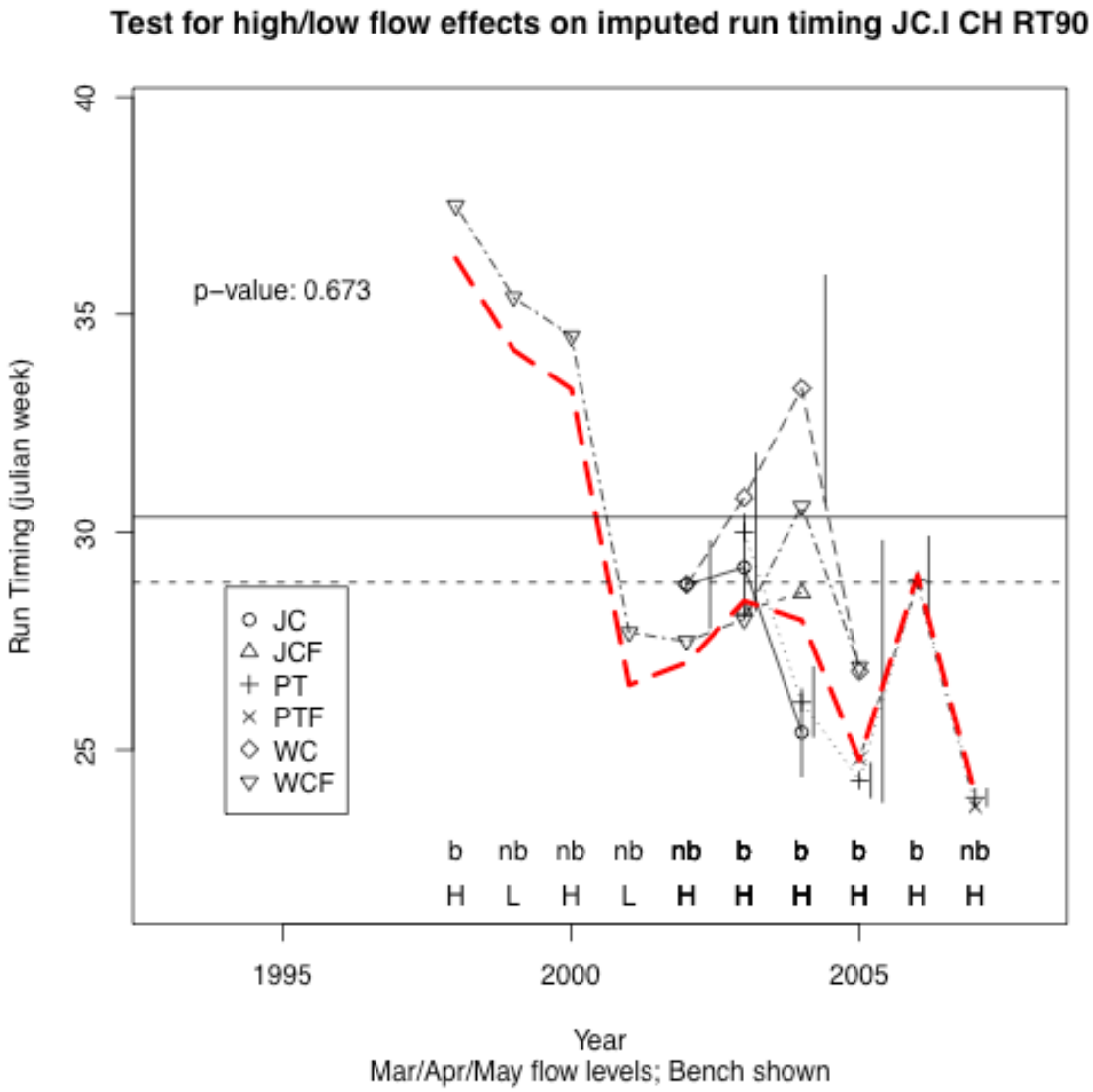


Figure 8.5 Summary of test for (a) linear trend in wild YOY 50th percentile of run timing, (b) changes in mean wild YOY 50th percentile of run timing between high and low flow years, and (c) changes in wild YOY 50th percentile of run timing between years with a bench in the hydrograph and with out bench. The heavy (red) line is the imputed wild YOY 50th percentile of run timing at the Junction City site based on a simple block model. Error bars are 95% confidence intervals. JCF, PTF, and WCF indicate the sampled-discharge flow-based estimates from the Junction City, Pear Tree, and Willow Creek sites respectively.


Test for linear trend on imputed run timing JC.I CH RT90
p−value: 0.004
Run Timing (julian week)
40
35
30
25
JC
JCF
PT
PTF
WC
WCF
b nb nb nb nb b b b b nb
H L H L H H H H H H
1995
2000
2005
Year
Mar/Apr/May flow levels; Bench shown


Test for high/low flow effects on imputed run timing JC.I CH RT90
p−value: 0.673
JC
JCF
PT
PTF
WC
WCF
b nb nb nb nb b b b b nb
H L H L H H H H H H
Run Timing (julian week)
25
30
35
40
1995
2000
2005
Year
Mar/Apr/May flow levels; Bench shown

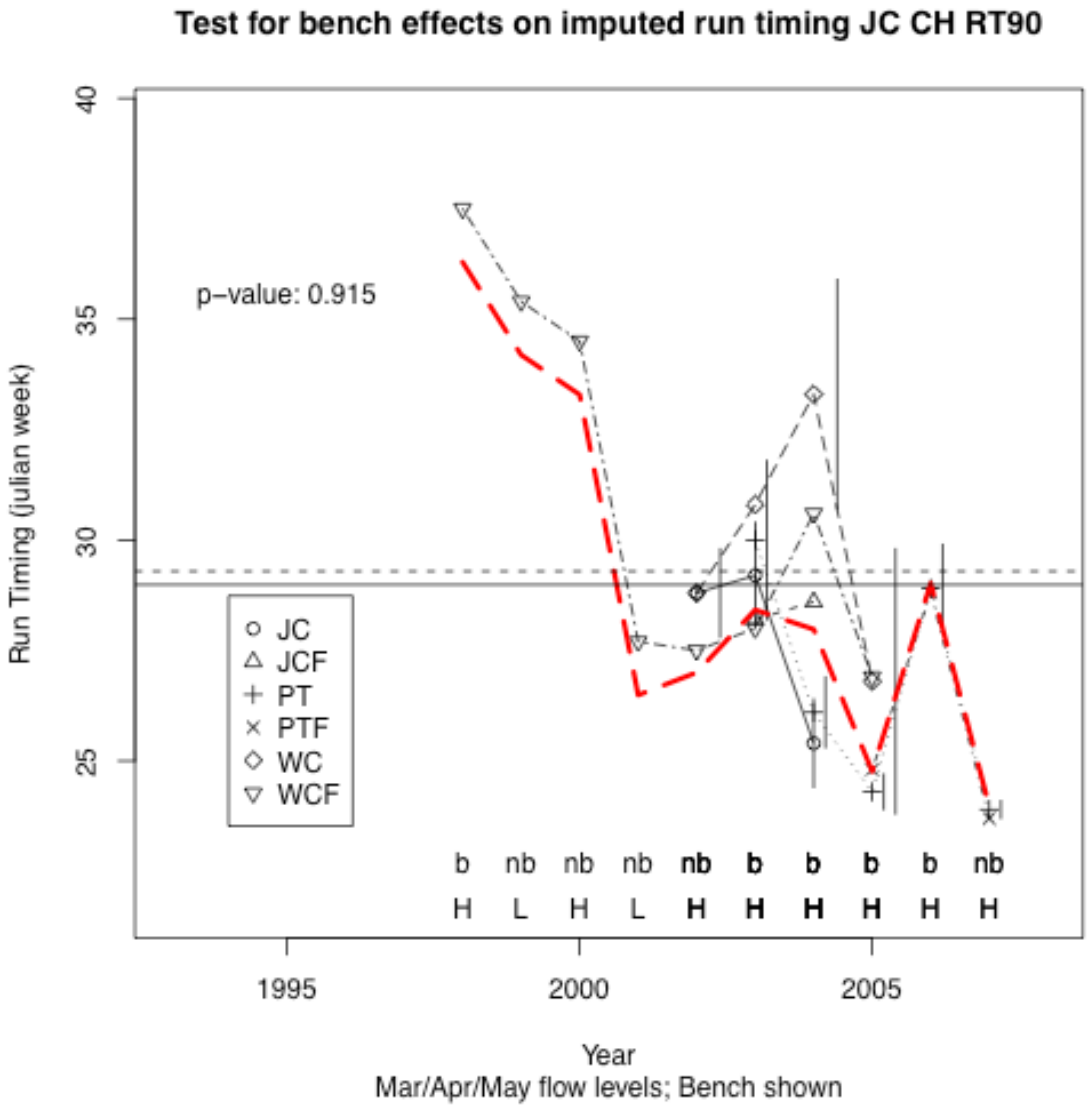


Figure 8.6 Summary of test for (a) linear trend in wild YOY 90th percentile of run timing, (b) changes in mean wild YOY 90th percentile of run timing between high and low flow years, and (c) changes in wild YOY 90th percentile of run timing between years with a bench in the hydrograph and with out bench. The heavy (red) line is the imputed wild YOY 90th percentile of run timing at the Junction City site based on a simple block model. Error bars are 95% confidence intervals. JCF, PTF, and WCF indicate the sampled-discharge flow-based estimates from Junction City, Pear Tree, and Willow Creek respectively.

The degree of parallelism is weaker for the 50th percentile than for the 90th percentile, but it is still evident that the various measurements are tracking the same trend. As with the trend analysis on abundance, individual analyses on each source with short time series are not very informative because they are unlikely to detect anything but gross changes. The use of Analysis of Covariance (ANCOVA) models is again not recommended because measurements within the same year from different sources are likely to be highly correlated which violates a key assumption of the ANCOVA model.

Similar to the analysis of the log(abundance) series, an imputed run timing that uses all of the sources of information was obtained by using a similar anova-block model (shown in red in the plots). The imputed run timing was then used to test for a linear trend, to test for a difference in the mean Julian week for each percentile between years with low and high flow conditions and between years where the hydrograph had or did not have a bench. The results are summarized in Table 8.7. There was strong evidence of (declining) linear trend in both percentiles of run timing, but no evidence of a difference in mean Julian week for either percentiles among either of the two classes of years. The estimated decline is fairly large (over a week/year!) which obvious cannot continue indefinitely.

Table 8.7 Summary of test for changes in mean Julian week corresponding to 50th and 90th percentile of run timing across years.

		50th Percentile			90th Percentile		
Site	Species	Linear trend n Slope (SE) p-value	High/Low Flow n_H Diff (SE) p-value	Bench/Non-bench n_B Diff (SE) p-value	Linear trend n Slope (SE) p-value	High/Low Flow n_H Diff (SE) p-value	BenchNon-bench n_B Diff (SE) p-value
Junction City I[1]	Chinook salmon	**10**	8	5	**10**	8	5
		-1.57	-4.04	-3.11	**-1.11**	-1.50	.31
		.16	3.88	3.12	**.28**	3.42	2.77
		<.001	.33	.35	**<.001**	.67	.91

[1] Imputed values at the Junction City site from an anova-block model.

A power analysis for detecting a linear trend is moot (a trend has been detected) but for completeness as been included in Table 8.8. For example, the power to detect a .21 Julian week/year (corresponding to a 1.5 Julian day/year) over a 10 year study is 79% for the 50th percentile of run timing. Power is high to detect moderate changes (more than 2 days/year) with at least 10 years of data.

Table 8.8 Estimated power (alpha=.05) to detect linear changes in run timing based on imputed Junction City Chinook salmon values.

	Julian week changes/year. [E.g., 07 Julian weeks/year = .5 day/year]									
	50th Percentile Process + sampling variation=1.4 weeks					90th Percentile Process + sampling variation=2.5 weeks				
	0.07	0.14	0.21	0.28	0.35	0.07	0.14	0.21	0.28	0.35
5 yrs	0.06	0.07	0.12	0.15	0.20	0.05	0.06	0.07	0.08	0.10
10 yrs	0.22	0.35	0.79	0.88	0.97	0.10	0.14	0.34	0.43	0.60
15 yrs	0.63	0.86	1.00	1.00	1.00	0.25	0.41	0.85	0.93	0.99
20 yrs	0.95	1.00	1.00	1.00	1.00	0.53	0.77	1.00	1.00	1.00

A power analysis to detect changes in the mean Julian week for the 50th and 90th percentile of the run timing is presented in Table 8.9. For example, the power (alpha=.05) to detect a 2.5 week change in the mean Julian week for the 50th percentile in a 10 year study with 5 years in each class is 69%. Power is generally high to detect a 3 week or larger difference in mean run timing in studies with 10 or more years

at the 50th percentile, and high to detect a 5 weeks or larger difference in a 10 years study at the 90th percentile.

Table 8.9 Estimated power (alpha=.05) to detect simple change in mean Julian week of run timing between two classes of year based on imputed Junction City Chinook salmon values. Years are divided equally among the two classes.

	Julian week difference in mean run timing between two classes									
	0.50	1.00	1.50	2.00	2.50	3.00	3.50	4.00	4.50	5.00
	50th percentile of run timing. Process + sampling standard deviation is 1.4 weeks.									
5 yrs	0.06	0.09	0.13	0.20	0.27	0.36	0.46	0.56	0.65	0.73
10 yrs	0.08	0.17	0.32	0.50	0.69	0.83	0.93	0.97	0.99	1.00
15 yrs	0.10	0.24	0.48	0.71	0.88	0.97	0.99	1.00	1.00	1.00
20 yrs	0.12	0.32	0.61	0.85	0.96	0.99	1.00	1.00	1.00	1.00
	90th percentile of run timing. Process + sampling standard deviation is 2.5 weeks.									
5 yrs	0.05	0.06	0.08	0.10	0.12	0.15	0.19	0.23	0.28	0.33
10 yrs	0.06	0.09	0.13	0.20	0.28	0.38	0.49	0.60	0.70	0.78
15 yrs	0.06	0.11	0.19	0.30	0.43	0.57	0.70	0.81	0.89	0.94
20 yrs	0.07	0.13	0.24	0.39	0.56	0.71	0.84	0.92	0.96	0.99

8.5 Discussion

The key challenge in evaluation across-year trends is the small time series available for most measures. This section demonstrated a way to combine information from different sources to extend the time series, but this method makes a critical assumption that the response measures from all sources moves in parallel (allowing for some random noise). This seems reasonable for the latest data where measurements at the Junction City, Pear Tree, and Willow Creek sites are taken on basically the same population, but there is little information (if any) available to calibrate the Willow Creek data prior to 2002.

The limiting factor in detecting longer term trends is the process error. The sampling error can be controlled by adjusting effort within each year, but in many cases, sampling error is small relative to process error. Unless additional covariates can be obtained to remove some of the process error, this large variation apparent in the response measures over time lead to studies with low power unless the number of years in the study is 10 or more.

Some of the variability across years especially in terms of log(abundance) and run timing is the survey window within each year. It is implicitly assumed that the measurements on the wild-population are comparable, e.g. that the individual studies started early enough and terminated late enough to sample the entire run of wild YOY fish. For this reason, it is important that a consistent study design be used across years to reduce some of the process error that arises because of year-to-year changes in study design.

Certain metrics are easier to measure and less sensitive to study design. For example, fish condition (e.g. fork length, weight) is relatively easy and inexpensive to measure over the course of a year. Although it is currently hypothesized that fish condition metrics should reflect an increase in health and survivability, it is uncertain how individual measures of condition of outmigrating smolts respond to improvements in fish rearing due to habitat restoration. Abundance would seem to be a direct measure of rearing habitat improvement, but has high process variation and is expensive to measure.

It is important to recognize that the implementation of restoration activities has implications for the ability to interpret the resulting data and to truly understand causal relationships. So many changes have occurred in the Trinity River without sufficient contrast or replication that it is difficult to attribute responses to any one type of action. As we have shown in the above examples a good starting place is to group responses by repeated features that are hypothesized to be important to various TRRP objectives (e.g., high flow years; years with a bench). We recommend that a timeline of restoration activities and important environmental data (e.g., water year type) be documented to date, and then be maintained as part of the TRRP. This timeline will improve the ability to interpret the monitoring data and understand the success of the program.

9. Final Summary

9.1 The spline-based methodology

The spline-based methodology in this report was developed to provide defensible estimates of outgoing migration based on a single-trap mark-recapture system. The stratified-Petersen estimator is the main competitor to this method. The key advantages of the spline-based method are:

- The hierarchical model for catchability improves estimates in cases where the sample size is small in a week; it also provides a mechanism for imputing an estimate of catchability for weeks where no mark-recapture is done.
- The spline model for abundance improves estimates based on the pattern of neighboring weeks in cases with small sample size; it also provides a mechanism for imputing an estimate when information is missing on the number of unmarked fish captured.
- It is self-calibrating in the sense of fitting a more complex model when data are rich and fitting simpler models when data are sparse.
- It can be used to separate the wild and hatchery components of the run and provide estimates of uncertainty for both.
- It readily deals with common problems in the data such as missing weeks, unrealistic numbers of marked fish recovered, etc.
- It readily provides estimates of run timing.
- It automatically incorporates all sources of variation in the final estimates of uncertainty,

The key disadvantage of the method is its complexity compared to a stratified-Petersen estimator but a suite of software has been developed that makes fitting these models as simple as possible. The complexity implies that a good understanding of the role of the hierarchical model and the spline are needed to ensure that the final model is sensible. For example, sharing information across weeks is not sensible if the sampling protocol across weeks is not consistent; or using a spline to interpolate is not sensible if the underlying pattern of the run does not have a "smooth" shape.

9.2 Use of discharge-sampled estimates

The discharge-sampled methods appear to track population trends within years fairly well, but are typically biased low. This may provide a mechanism for improving within year estimates if the proportion of the flow sampled is used a covariate for the capture-efficiency. Because of the consistency in the

relationship within a year, estimates of runtime based on the discharge-sampled methods should be comparable to those from the spline-based methods.

The estimate of total abundance based on the discharge-sampled methods underestimates the abundance (as measured by the spline based methods) with a correction factor that appears to be consistent across years except for a few instances. These latter cases need to be investigated further to determine the reasons for the variation from the other years. Additional years of analysis would be helpful to determine if the relationship continues past the scope of this report.

No estimates of uncertainty have been provided for the discharge-sampled methods. It is unclear how to propagate measures of uncertainty forward through the chain of computations and some sort of bootstrap method will likely be required.

Further use of these methods will require careful attention to the collection and storage of such data. For example, the current data summaries do not provide information on how many hours the traps were actually in operation for each day of the season.

9.3 Fish condition factors

Several potential fish condition factors were considered for use in long-term evaluations of the program. Fork length has the longest time series available and was the only metric evaluated in this report. There are several key concerns about using fork length. First, random samples must be taken from the fish available each day at the trap. Second, it is important to estimate the mean-fork length at each day for both wild and hatchery fish. With Chinook salmon, only a portion of hatchery-fish are adipose-fin clipped, so an estimate of the mean fork length for wild and hatchery fish must be imputed from the mean fork length for ad-clipped and non-clipped fish which adds an extra level of variability into the process. A smoother should be used to estimate the mean fork length for a particular Julian day to remove some of the variation in the daily means resulting from the small samples measured each day. Lastly, we arbitrarily choose two Julian dates (day 100 and day 200), and some thought needs to be given to appropriate ways to standardize interannual comparison of lengths, such as degree days since hatch.

9.4 Across-year program evaluation

The major limitation in the program assessment using multiple years of data are the short time-series available and large process error evident in the results. Some of the process error may be a result of changes in the study design over the years so consistency in study design across years is important.

It is possible to combine several short series from different sources if the data shows some evidence of parallelism in response – an example of the long-term evaluation of abundance showed how this could be done. It is again important to try and minimize process error that is attributable to changes in sampling protocols over years. For example, this report used the total run of wild fish as the response variable, but the studies started at different Julian weeks in different years, so some of the process variation seen may be a result of this sampling artifact.

9.5 Planning within year studies

It is not surprising that the estimated run size and associated precision are often similar to those obtained from a Stratified Petersen estimator. Indeed, if sample sizes were large in each Julian week with no missing data, the estimates would be virtually identical. The major reason that the more complicated

spline-based model is required is to deal with weeks with missing or odd data, and weeks with very small numbers of fish marked and recaptured.

Using this analogy, several remarks can be made about how to plan the within-year studies.

It is important to mark and release sufficient numbers of fish when the number of fish passing the traps is large, e.g. at the peaks of the wild and/or hatchery runs. Sparse or missing mark-recapture data in these weeks implies that the spline-based model must impute plausible values of the trap-efficiency based on the other weeks and variability in the trap-efficiency in other weeks leads to large variability in the estimate of the run for this particular week. Conversely, in weeks when few fish are expected to pass the trap, it is not important to mark many fish as even a poor estimate of trap efficiency had negligible effect on the precision of the overall run.

Because of the ability of the spline-based methods to interpolate for missing weeks, there is some flexibility in data collection across weeks. Again, it is very important to ensure that the traps are monitored during the peak of the run, but during periods of low migration (e.g. the tails), monitoring could be reduced to every second week to reduce costs.

In several of the datasets analyzed, it was found that a very large number of fish migrated in the first week of sampling with little mark-recapture effort. This is problematic in two ways. First, the imputation of the capture-efficiency for this week results in estimates with poor precisions. Second, there is no information available on how many fish passed through the system prior to the first week! Sampling should begin before large numbers of fish exit the system.

As part of the previous report and this report, a small simulator is available to help in planning the allocation of effort across the year. Please contact the senior author for more details.

Literature cited

Bjorkstedt, E.P. 2000. DARR (Darroch Analysis with Rank-Reduction): A method for analysis of stratified mark-recapture data from small populations, with application to estimating abundance of smolts from outmigrant trap data. Administrative Report SC-00-02, National Marine Fisheries Service. Available at: http://swfsc.noaa.gov/publications/FED/00116.pdf, Accessed 2009-07-28.

Bonner, S.J. 2008. Heterogeneity in capture-recapture: Bayesian methods to balance realism and model complexity, Ph.D. Thesis, Simon Fraser University. Available at: http://www.stat.sfu.ca/people/alumni/Theses/Bonner-2008.pdf

Brooks, S.P. and A. Gelman. 1998. Assessing convergence of Markov Chain Monte Carlo algorithms. Statistics and Computing 8(4): 319–335.

Brooks, S.P., E.A. Catchpole and B.J.T. Morgan. 2000. Bayesian animal survival estimation. Statistical Science 15: 357-376.

Darroch, J. N. 1961. The two-sample capture-recapture census when tagging and sampling are stratified. Biometrika 48: 241-260.

Eilers, P.H.C. and B.D. Marx. 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11: 89-121.

Freeman, M. F. and J.W. Tukey. 1950. Transformations related to the angular and square root. Annals of Mathematical Statistics 21: 607-611.

Gelman, A., J.B. Carlin, H. Stern, and D.R. Rubin. 2004. Bayesian data analysis, 2nd Edition. Chapman and Hall, New York.

Gerard, P.D., D.R. Smith and G. Weerakkody. 1998. Limits of Retrospective Power Analysis. Journal of Wildlife Management 62: 801-807.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68: 1364-1372.

Green, J., E. Logan, D. Zajanc, P. Petros. 2007. Emigrant surveys and rotary trapping. Estimation of abundance and outmigrating juvenile salmonids in Trinity River using rotary screw traps in 2002 and 2003.

Hayden, T. and B. Pinnix. 2007. Condition Factor Indices of Outmigrating Juvenile Trinity River Anadromous Fish at Willow Creek, CA. Presentation at the 1st Trinity River Science Symposium. 2007. Weaverville, CA.

Hastie, T.J. 1992. Generalized additive models. Chapter 7 of Statistical Models in S.J.M. Chambers and T. J. Hastie (eds.). Wadsworth & Brooks/Cole, Pacific Grove, California.

Healey, M.C. 1991. Life history of Chinook salmon. Pages 313-393 in C. Groot and L. Margolis (eds.). Pacific Salmon Life Histories. University of British Columbia Press, Vancouver, British Columbia, Canada.

Hilborn, R. and C.J. Walters. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics & Uncertainty. Kluwer Academic Publishers, Boston, MA.

Kjelson, M.A., P.F. Raquel, and F.W. Fisher. 1981. Influences of freshwater inflow on Chinook salmon (*Oncorhynchus tshawytscha*) in the Sacramento-San Joaquin estuary. Pages 88-102 in R.D. Cross and D.L Williams (eds.). Proceedings of the National Symposium on Freshwater Inflows to Estuaries. U.S. Fish and Wildlife Srevice, Biological Services Program, FWS/OBS-81/04(2).

Lang, S. and Brezger, A. 2004. Bayesian P-splines. Journal of Computational and Graphical Statistics 13: 183–212.

Lenth, R.V. 2009. Java Applets for Power and Sample Size [Computer software]. Accessed 2009-10-05, from http://www.stat.uiowa.edu/~rlenth/Power.

Lunn, D.J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10:325-337.

Mantyniemi, S. and A. Romakkaniemi. 2002. Bayesian mark-recapture estimation with an application to a salmonid smolt population. Canadian Journal of Fisheries and Aquatic Science 59: 1748–58.

Morrison, D.F. 1983. Applied Linear Statistical Models. Prentice-Hall, New York.

North State Resources, Inc., ESSA Technologies Ltd., and Simon Fraser University, Dept of Statistics and Actuarial Science. 2008. Trinity River Restoration Program: Juvenile Salmonid Outmigrant Monitoring Evaluation, Final Technical Memorandum. Prepared for the Trinity River Restoration Program, Weaverville, CA.

Pinnix, B., J. Polos, A. Scheiff, S. Quinn, and T. Heyden. 2007. Juvenile salmonid monitoring on the mainstream Trinity River at Willow Creek, California, 201-2005. U.S. Fish and Wildlife Service. Arcata Fisheries Data Series Report DS 2007-09.

Read, C.B. 1993. FreemanTukey chi-squared goodness-of-fit statistics. Statistics & Probability Letters 18: 271-278.

Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada 191:1-382.

Schwarz, C.J. and Taylor C.G. 1998. The use of the stratified-Petersen estimator in fisheries management with an illustration of estimating the number of pink salmon (*Oncorhynchus gorbuscha*) that return to spawn in the Fraser River. Canadian Journal of Fisheries and Aquatic Sciences 55: 281-296.

Seber, G.A.F. 1982. Estimation of animal abundance, 2nd edition. Griffen, London.

Seber, G.A.F. and Felton R. 1981. Tag loss and the Petersen mark-recapture experiment. Biometrika 68: 211-219.

Steidl, R.L., J.P. Hayes and E. Schauber. 1997. Statistical power analysis in wildlife research. Journal of Wildlife Management 61: 270-279.

Steinhorst, K., Y. Wu, B. Dennis and P. Kline. 2004. Confidence intervals for fish out-migration estimates using stratified trap efficiency methods. Journal of Agricultural, Biological, and Environmental Statistics 9: 284-299.

Thomas, A., B. O'Hara, U. Ligges, and S. Stutz. 2006. Making BUGS open. R News 6: 12-17.

Trinity River Restoration Program, ESSA Technologies Ltd. 2008. Integrated Assessment Plan, Version 0.99 – December 2008. Draft report prepared for the Trinity River Restoration Program, Weaverville, CA. 269 pp.

United States Fish and Wildlife Service (USFWS) and Hoopa Valley Tribe (HVT). 1999. Trinity River Flow Evaluation Study - Final Report. A report to the Secretary, US Department of the Interior, Washington, D.C.

Wedemeyer, G.A., R.L. Saunders, and W.C. Clarke. 1980. Environmental factors affecting smoltification and early marine survival of anadromous salmonids. Marine Fisheries Review 42(6): 1-14.

Appendix A: Introduction to Splines

A.1 What is a spline?

A common method of curve fitting to data is a form of regression analysis. The simplest regression model attempts to fit a single straight line to a set of data. For example, a straight line can be parameterized by the intercept (β_0) and the slope (β_1) and has the form $Y_i = \beta_0 + \beta_1 X_i$.

However, in real life, it is very rare for the data points to fall exactly on the straight line and the goal of regression analysis is to find the best fitting line to the data. Under a least squares model, the parameters of the model (e.g. the slope and intercept) are chosen to minimize the sum of the squared vertical deviations, i.e. find the values of the slope and intercept such that the total vertical sum of squared deviations is minimized[10]:

$$\underset{\beta_0,\beta_1}{\arg\min} \sum \left[Y_i - (\beta_0 + \beta_1 X_i) \right]^2$$

While it is possible to write out explicit solutions for the simplest case above, this is not feasible in more complicated models.

In many cases, the relationship between *Y* and *X* is not a simple straight line. One solution is to use a more flexible curve than a straight line such as a quadratic relationship between *Y* and *X*:
$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$. A least squares approach can also be used to estimate the parameters:

$$\underset{\beta_0,\beta_1,\beta_2}{\arg\min} \sum \left[Y_i - (\beta_0 + \beta_1 X_i + \beta_2 X_i^2) \right]^2$$

But, these simple functions may be too inflexible to fit the relationship over the full range of the data. A cubic (or rarely higher order) polynomials could also be fit, but these are rarely successful because the entire shape of the curve over the whole range of the data is determined by a few parameters and the curve is rarely flexible enough.

A more flexible approach is through the use of *splines.* A spline is composition of simpler curves on smaller intervals that span the range of the data. The simpler curves are called *basis* functions. The range of the data is partitioned into smaller intervals at points called *knots*. For example, consider the basis functions:

$$\{1, X, (X-4)_+, (X-8)_+\}$$

where the basis function

$$(X-K)_+ = \begin{cases} (X-K) & \text{if } X \geq K \\ 0 & \text{if } X < K \end{cases}$$

Consider the behavior of the curve (the spline):

[10] The ArgMin function says find the values of the parameters (listed underneath) that minimize the function to the right.

$$Y = \beta_0(1) + \beta_1 X + \beta_2(X-4)_+ + \beta_2(X-8)_+$$

evaluated at the following points:

Simple Spline evaluated at selective values of X.	
X	Y
0	β_0
1	$\beta_0 + \beta_1$
2	$\beta_0 + \beta_1(2)$
3	$\beta_0 + \beta_1(3)$
4	$\beta_0 + \beta_1(4)$
5	$\beta_0 + \beta_1(5) + \beta_2(1)$
6	$\beta_0 + \beta_1(6) + \beta_2(2)$
7	$\beta_0 + \beta_1(7) + \beta_2(3)$
8	$\beta_0 + \beta_1(8) + \beta_2(4)$
9	$\beta_0 + \beta_1(9) + \beta_2(5) + \beta_3(1)$
10	$\beta_0 + \beta_1(10) + \beta_2(6) + \beta_3(2)$

Notice in the range of 0 to 4, the curve is straight line with slope β_1; in the interval 4 to 8, the curve is a straight line with slope ($\beta_1 + \beta_2$) as for every increase in *X* by one unit (e.g. from 4 to 5), the value of *Y* increases by $\beta_1 + \beta_2$; in the interval 8 to 10, the curve is a straight line with slope ($\beta_1 + \beta_2 + \beta_3$). This is shown graphically (Figure A-1) for particular values of the parameters as:

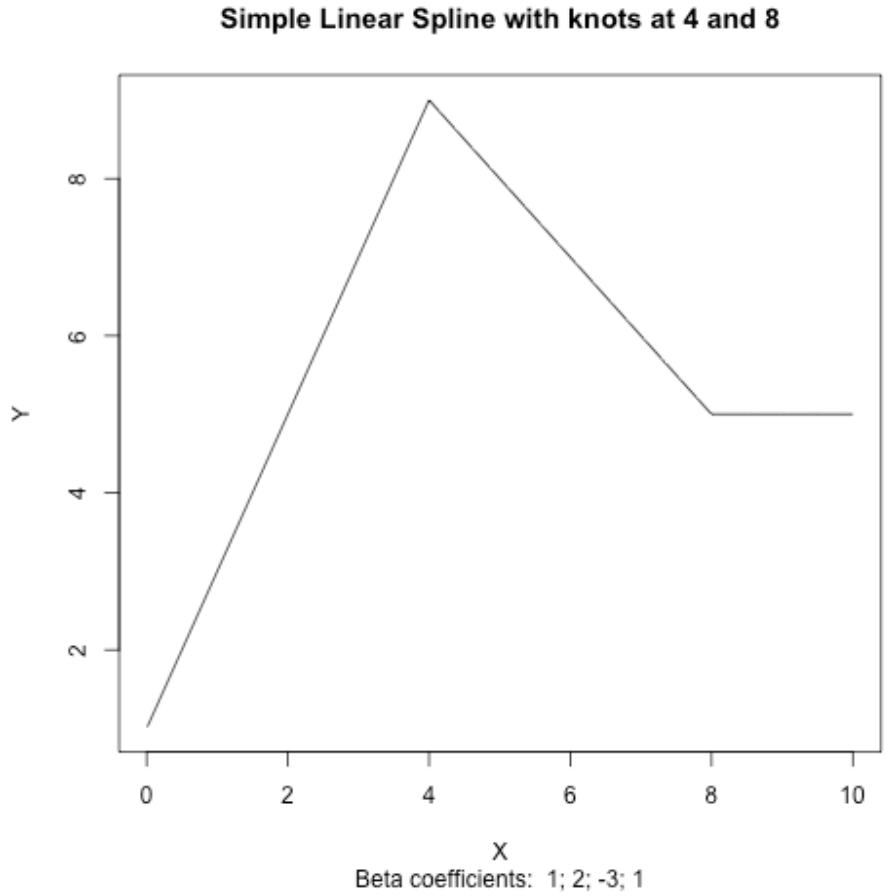


Figure A-1. Example of simple linear spline with knots at 4 and 8.

Linear splines can have "sharp" changes at the knots.[11] A more flexible spline can be enabled by increasing the degree of the basis functions or the number of knots. For example, to use quadratic basis functions between the knots, the basis functions are now:

$$\left\{1, X, X^2, (X-4)_+^2, (X-8)_+^2\right\}$$

[11] More formally, the first derivative does not exist at the knot points.

A plot of this curve for a set of specific parameters is found in Figure A-2:

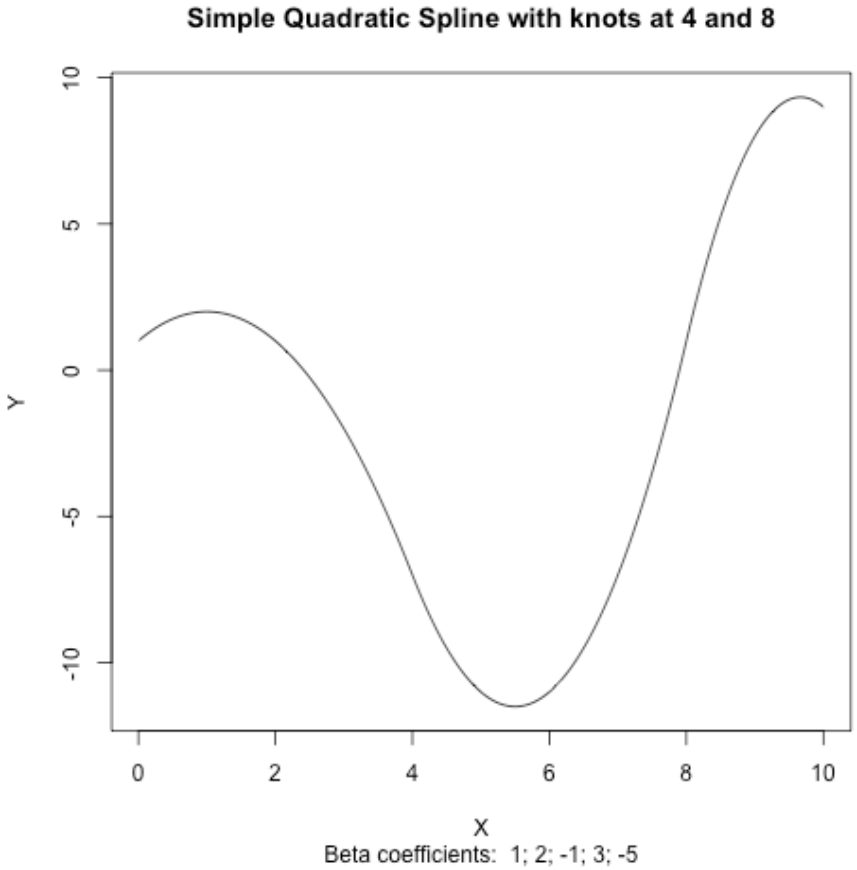


Figure A-2. Example of a simple quadratic spline with knots at 4 and 8.

If the number of knots is increased, the curve is also "smoother" as there are more individual segments to be fit. For example, a set of linear basis functions with knots at 2, 4, 6, and 8 could give the plot in Figure A-3:

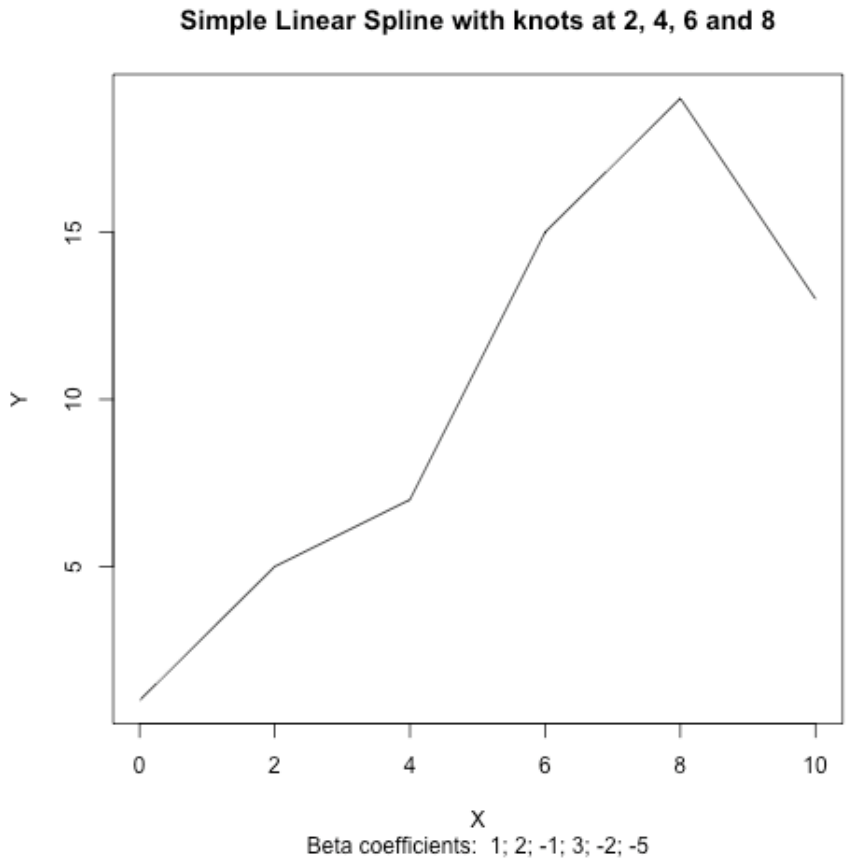


Figure A-3. Simple linear spline with more knots..

A reasonable compromise between smoothness and number of knots is to use cubic basis functions with knots spaced approximately every fourth observation (cite Rupert's paper). One advantage of cubic splines over the linear splines, is that the curve based on cubic splines is continuous in the first and second derivatives at each knot, i.e. the curve is "smooth" as it traverses the knots.

A.2 Fitting the spline using least squares

Fitting a spline to a set of data is relatively straightforward. For example, consider the following set of 40 data points in the range of 1 to 40 shown in Figure A-4:

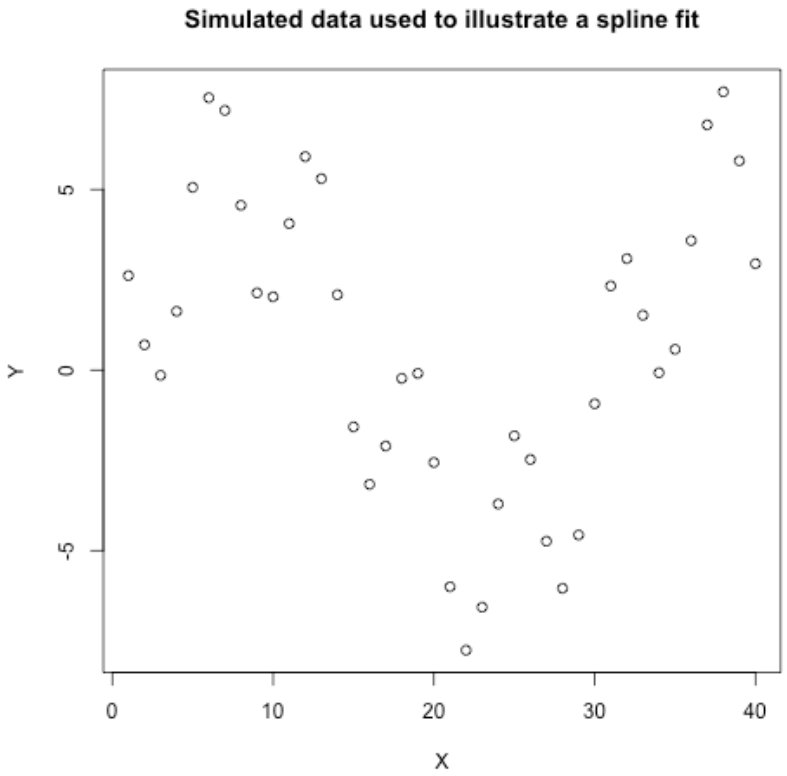


Figure A-4. Simulated data to illustrate a spline fit.

First, the number and position of the knots must be chosen. As noted earlier, setting the number of knots equal to about ¼ of the data points and spacing them equally through the range works well. In this case, 9 knot points will be chosen, spaced at positions 4, 8, 12, …, 36.

Second, the design matrix is constructed consisting of the basis functions evaluated at each knot. The basis functions chosen will be:

$$\left\{1, X, X^2 (X-4)_+^2, (X-8)_+^2, \text{K}\ , (X-36)_+^2\right\}$$

The design matrix will have a separate row for each data point (40 rows in total), and $1+q+k$ columns where q is the degree of the highest order polynomial term ($q=2$) and k is the number of knot points ($k=9$) for a total of 12 columns. Part of the design matrix is shown in Table A-1.

Table A-1. Part of the design matrix used to fit a spline using least squares.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	1	1	1	0	0	0	0	0	0	0	0	0
[2,]	1	2	4	0	0	0	0	0	0	0	0	0
[3,]	1	3	9	0	0	0	0	0	0	0	0	0
[4,]	1	4	16	0	0	0	0	0	0	0	0	0
[5,]	1	5	25	1	0	0	0	0	0	0	0	0
[6,]	1	6	36	4	0	0	0	0	0	0	0	0
[7,]	1	7	49	9	0	0	0	0	0	0	0	0
[8,]	1	8	64	16	0	0	0	0	0	0	0	0
[9,]	1	9	81	25	1	0	0	0	0	0	0	0
[10,]	1	10	100	36	4	0	0	0	0	0	0	0

Third, the spline is fit using standard least squares (multiple regression). The fitted coefficients (and associated standard error) are found in Table A-2:

Table A-2. Fitted coefficients and standard error from the fitted spline found using least squares.

	est_beta	se(est_beta)
SplineDesign1	7.36	4.80
SplineDesign2	-5.72	3.58
SplineDesign3	1.11	0.57
SplineDesign4	-1.66	0.76
SplineDesign5	0.71	0.41
SplineDesign6	-0.26	0.37
SplineDesign7	0.12	0.36
SplineDesign8	0.18	0.36
SplineDesign9	-0.13	0.36
SplineDesign10	-0.11	0.36
SplineDesign11	0.25	0.38
SplineDesign12	-0.77	0.51

The estimated coefficients are usually not very interesting and have no easy interpretation. However, the fitted values on the spline can be estimated as in regular regression by substituting in values of X and evaluating the spline. A plot of the fitted spline function is shown in Figure A-5:

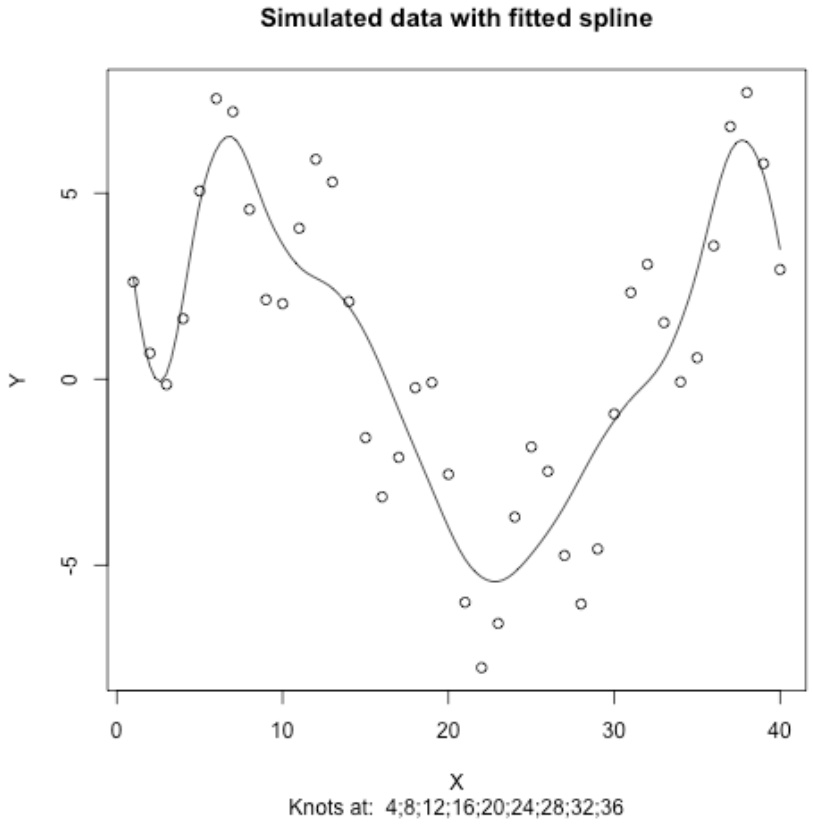


Figure A-5. Fitted spline curve using least squares.

A.3 B-Splines

The $(X-K)_+^q$ may not be the best choice of basis function for many models. The problems are that the resulting design matrix has columns that are highly "correlated" because of the way they are constructed and the least squares fit may be numerically unstable because of the potential for large values. A more numerically stable set of basis functions are the *B-spline* basis functions (Hastie 1992). There are more cumbersome to set up but many computer packages have functions to do this (e.g. the *bs()* function in the *splines* library of *R*). For example, the first rows of the design matrix constructed using the *B-spline* basis are shown in Table A-3.

Table A-3. First few rows of design matrix for B-splines.

	1	2	3	4	5	6	7	8	9	10	11	12
[1,]	0.5625	0.40625	0.03125	0.00000	0.00000	0	0	0	0	0	0	0
[2,]	0.2500	0.62500	0.12500	0.00000	0.00000	0	0	0	0	0	0	0
[3,]	0.0625	0.65625	0.28125	0.00000	0.00000	0	0	0	0	0	0	0
[4,]	0.0000	0.50000	0.50000	0.00000	0.00000	0	0	0	0	0	0	0
[5,]	0.0000	0.28125	0.68750	0.03125	0.00000	0	0	0	0	0	0	0
[6,]	0.0000	0.12500	0.75000	0.12500	0.00000	0	0	0	0	0	0	0
[7,]	0.0000	0.03125	0.68750	0.28125	0.00000	0	0	0	0	0	0	0
[8,]	0.0000	0.00000	0.50000	0.50000	0.00000	0	0	0	0	0	0	0
[9,]	0.0000	0.00000	0.28125	0.68750	0.03125	0	0	0	0	0	0	0
[10,]	0.0000	0.00000	0.12500	0.75000	0.12500	0	0	0	0	0	0	0

The B-Spline design matrix has entries that are small, and has a banded-diagonal structure making it more stable for numerical estimation. The estimated coefficients are also found using least squares and shown in Table A-4:

Table A-4. Estimated coefficients from B-spline fitted using least squares.

	est_beta	se(est_beta)
bs.SplineDesign1	7.36	4.80
bs.SplineDesign2	-4.08	3.02
bs.SplineDesign3	8.41	2.10
bs.SplineDesign4	3.03	1.93
bs.SplineDesign5	2.41	1.89
bs.SplineDesign6	-1.92	1.89
bs.SplineDesign7	-6.06	1.89
bs.SplineDesign8	-4.33	1.89
bs.SplineDesign9	-0.89	1.91
bs.SplineDesign10	0.74	1.99
bs.SplineDesign11	8.67	2.43
bs.SplineDesign12	3.48	2.03

These coefficients have no easy interpretation. Fortunately, the fitted curve is identical under any basis function as shown in Figure A-6:

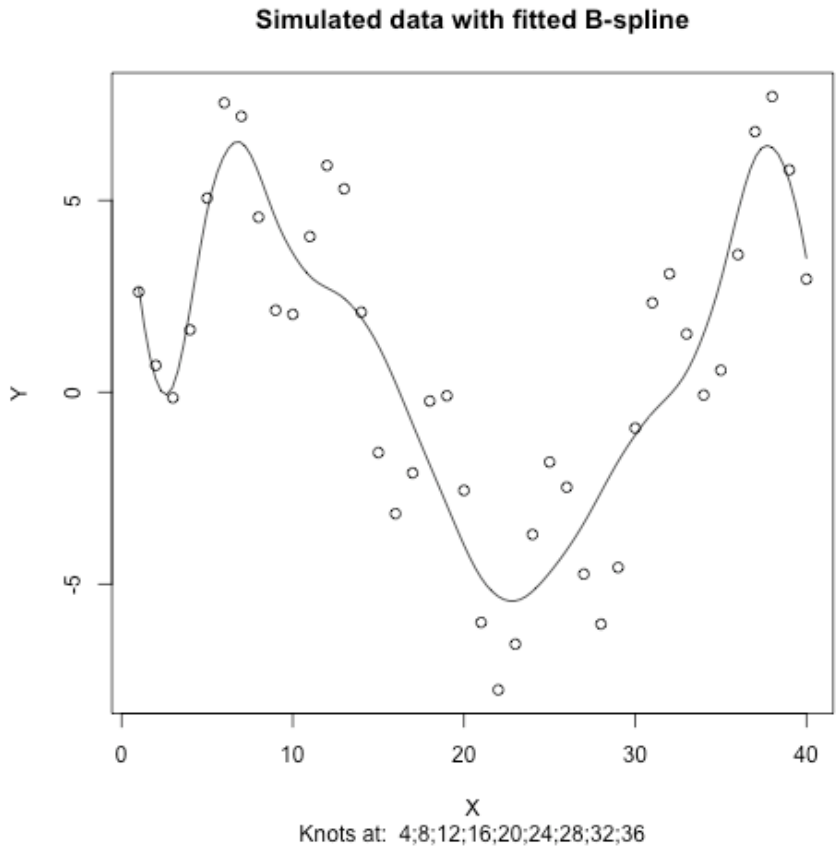


Figure A-6. Fitted B-spline curve to simulated data. The curve is identical to Figure A-5.

An Introduction to Bayesian Inference and MCMC Methods for Capture-Recapture

Contents

1 Bayesian Inference for the Binomial Model

Bayesian inference is a branch of statistics that offers an alternative to the frequentist or classical methods that most are familiar with. At its core, Bayesian inference is based on an alternative understanding of probability which many have argued is more intuitive and meaningful than the classical long-run-averages interpretation. In practice, Bayesian methods present alternatives that often allow for more intricate models to be fit to complex data. Growth in Bayesian inference has been particularly strong in recent years due to advances in computing – both in the development of new algorithms and new hardware. We will begin with a review of maximum likelihood estimation for the binomial model and then introduce the fundamental concepts of Bayesian inference.

1.1 Maximum Likelihood Estimation for the Binomial Distribution

The Binomial Distribution The binomial distribution plays a key role in capture-recapture analysis because it models the distribution of the size of a sample drawn at random from a fixed population. Suppose that a population contains n marked individuals and that a sample is drawn in such a way that: 1) every marked individual has the same probability of being captured, denoted p and 2) whether or not one individual is captured does not affect the probability that any other individual is captured. In this case, the number of individuals captured, which we

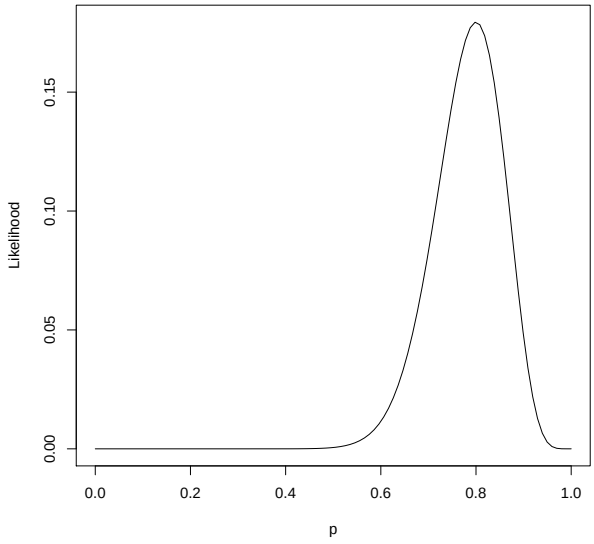


Figure 1: Probability mass function for the binomial distribution with population size $n = 30$ and capture probability $p = .8$.

will call m, will have a binomial distribution. We commonly denote this by writing $m \sim$ Binomial(n, p) and the exact probability that m individuals are captured is:

$$P(m|p) = \binom{n}{m} p^m (1-p)^{n-m}.$$

This function is called the probability mass function, and the notation $m|p$ emphasizes that we are considering how the probability changes with m while p remains fixed. Figure 1 plots the probability mass function for a binomial model with population size $n = 30$ and capture probability $p = .8$.

The Likelihood Function The probability mass function for a model tells us how the probabilities of the different possible outcomes for an experiment vary as functions of fixed values of the parameters. After the experiment is conducted the outcome is fixed and instead we want to know how the probability of the specific outcome, the observed data, varies if the parameters are changed. This is the role of the likelihood function. The likelihood function is in fact exactly equal to probability mass function, treated as a function of the parameters for fixed data. For the binomial model, the likelihood function is:

$$L(p|m) = \binom{n}{m} p^m (1-p)^{n-m}.$$

The notation $p|m$ emphasizes that we are now considering the probability changes with p while m remains fixed. Figure 2 illustrates the likelihood function for a specific binomial experiment in which $m = 24$ individuals were captured from a population of $n = 30$.

Maximum Likelihood Estimates Simply put, the maximum likelihood estimator of a parameter is the value which maximizes the likelihood function given the observed data. Intuitively, this is the value of the parameter for which the observed data was most probable to occur. It is simple to show that likelihood function for the binomial model is maximized when p is equal to the proportion of individuals captured in the sample. Hence, the maximum likelihood estimator of p is $\hat{p} = n/N$.

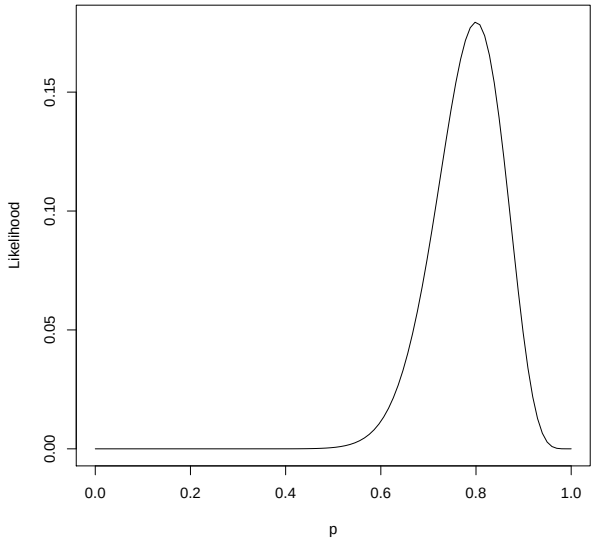


Figure 2: Likelihood function for a binomial experiment in which $m = 24$ of $n = 30$ individuals were captured.

Measures of Uncertainty Although the probability of the observed data is maximized by the this estimator, there may be other parameter values for which the probability of the observed data is almost as high. These values are also plausible guesses for the parameter and providing the maximum likelihood estimate alone may be misleading. Figure 2 shows clearly that the likelihood function for the binomial experiment with $n = 30$ and $m = 24$ is maximized when $p = .8$ and is almost equal to 0 when p is less than .5 or greater than .95. However, there is a broad range of values around .8 where the likelihood is quite high and these values can't be ignored.

To describe our uncertainty, we need to provide some measure of how far the true value of the parameter could possibly be from our estimate. Common measures are the standard error and 95% confidence interval. Imagine that the same experiment could be repeated many, many times without changing the value of the parameter, and that we could computed maximum likelihood estimates for each of the resulting data sets. The standard error is the standard deviation of these estimates. The 95% confidence interval is an interval defined by a pair of values which are computed in the same way for each of the data sets and would bound the true parameter value for at least 95% of the replications. The standard error of $\hat{p}$ for the binomial experiment is $SE_p = \sqrt{\hat{p}(1-\hat{p})/m}$ and the limits of a 95% confidence interval are given by:

$$\hat{p} - 1.96SE_p \text{ and } \hat{p} + 1.96SE_p.$$

In practice, we consider any value inside this interval as a plausible guess for the true parameter. If the interval is narrow, then there are few plausible values and we can be certain that the true parameter value is close to our estimate. If the interval is wide, then we have a lot of uncertainty about the true parameter value. The standard error for the specific data $n = 30$ and $m = 24$ is .07 and the 95% confidence interval is (.66,.94).

1.2 Bayesian Inference and the Posterior Density

Including Prior Information Classical statistical methods, including maximum likelihood, may be considered objective because the computation of estimates depends only on the observed data. Given a model of the system, we solve for the values of the parameters which maximize the probability of the observed data. There is only one solution and any researcher who used the same model and observed the same data would arrive at the same estimates. (There are deeper philosophical issues concerning the subjectivity in defining a model, but we won't get into these.) Suppose, however, that we had prior knowledge about the behaviour of the system. In the case of capture-recapture, we might have data from prior experiments with similar populations which suggests

that some values of the capture probability are more likely than others. Is there a way that this information can be incorporated into the analysis?

The answer is yes – this is exactly what Bayesian statistics does. Rather than using the likelihood function, Bayesian parameter estimates are computed from a new function which incorporates both the information from the data and any information about the parameters that is available *before starting the experiment*.

Imagine that in the case of the binomial experiment a pilot study had been conducted in which $m_1 = 10$ individuals were captured out of a population of $n_1 = 20$. If we can assume that the capture probability in the pilot study is the same (or very similar to) the capture probability in the full study then we may wish to combine the data to improve the precision of our estimates. The likelihood function for the capture probability in the pilot study is:

$$L(p|m_1) = \binom{20}{10} p^{10}(1-p)^{10}$$

and the combined likelihood from both studies would be:

$$L(p|m, m_1) = \binom{20}{10} p^{10}(1-p)^{10} \cdot \binom{30}{24} p^{24}(1-p)^{6} = \binom{20}{10}\binom{30}{24} p^{34}(1-p)^{16}.$$

The new maximum likelihood estimate would be $\hat{p} = 34/50 = .68$.

Even if no pilot study had been conducted, there may be other reasons to that some values of the parameters are more plausible than other values, and we may wish to make use of this information in our analysis. A Bayesian analysis begins with the definition of a function called the prior density which mimics the likelihood function which may have been obtained from a pilot study if one had been conducted. If, for example, we believed strongly that the capture probability were close to .5 then we may define the probability density to be:

$$\binom{20}{10} p^{10}(1-p)^{10}$$

which exactly mimics the likelihood from the pilot study above. Alternatively, if we believed this less strongly then we might define the prior density to be:

$$\binom{2}{1} p(1-p).$$

Figure 3 compares these two choices for the prior density. Notice that the first is more concentrated around .5 while the second spreads more evenly over the entire interval from 0 to 1.

After the experiment is conducted, the posterior density is constructed by multiplying the selected prior density with the likelihood function defined from the observed data. A Bayesian analysis then uses posterior density, instead of the likelihood function alone, to compute parameter estimates and associated measures of uncertainty. Figure 4 compares the shapes of the prior density, the likelihood and the posterior density given both the strong and weak prior densities defined above. In both cases, the posterior density forms a compromise between the other two curves, putting the highest credibility at values of p that are supported by both the prior density and the likelihood. However, the posterior density given the weak prior information is more similar to the likelihood function whereas the posterior density given the strong prior information is more similar to the prior density.

This simple example also illustrates one of the most important facts of Bayesian analysis – as we collect more data the posterior distribution becomes closer to the likelihood, and parameter estimates become closer to the objective maximum likelihood estimates. As the population gets bigger in the binomial experiment (n increases) the value of $\hat{p}$ will move toward m/n. If $m = 240$ and $n = 300$ and the same prior density is selected then $\hat{p} = .78$. If $n = 2400$ and $N = 3000$ then $\hat{p} = .80$, exactly the same as the maximum likelihood estimate to 2 decimals of accuracy. Heuristically, we say that the large amount of data overwhelms the prior beliefs about the system.

1.3 Summarizing the Posterior Distribution

The Posterior Density is a True Probability Density One very important fact of Bayesian inference is that the posterior density is a true probability density (Actually, we have to normalize the density so that it integrates to 1 over the entire parameter space, but this is a technicality). This is in fact the key difference between classical and Bayesian inference. In classical inference, we summarize our uncertainty in terms of how parameter estimates would vary if an experiment were repeated many times yielding many different data sets. In Bayesian inference, we summarize uncertainty in terms of direct probability statements about the parameters.

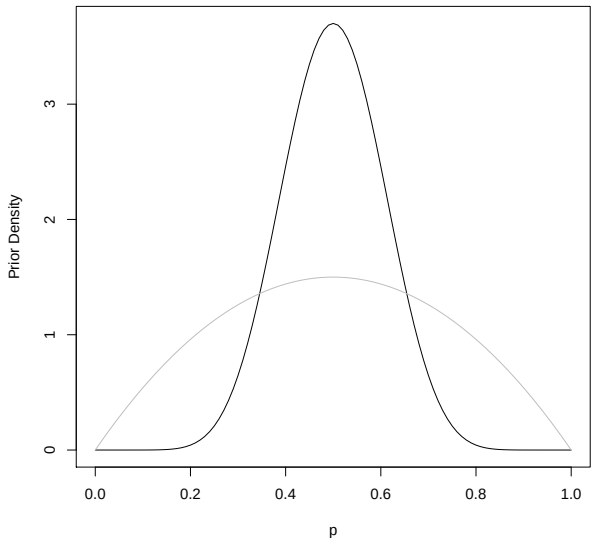


Figure 3: Comparison of the prior densities $\binom{20}{10}p^{10}(1-p)^{10}$ (black) and $\binom{2}{1}p(1-p)$ (grey).

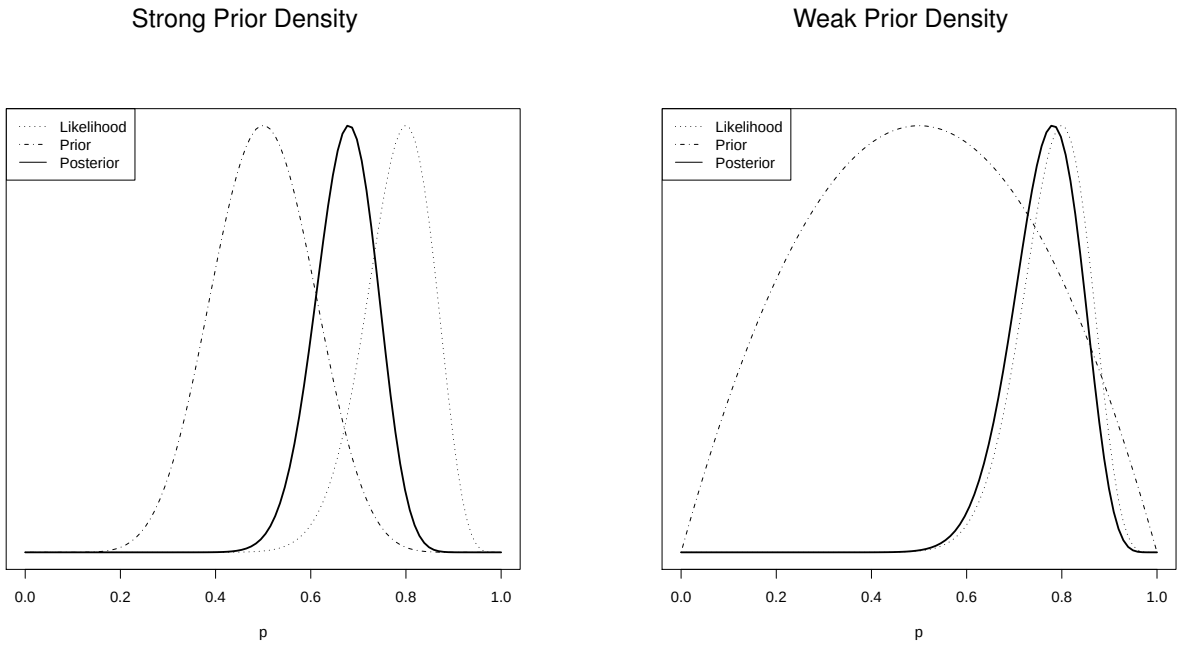


Figure 4: Comparison of the prior density, the likelihood and the posterior density for the binomial model with $n = 30$ and $m = 24$. The left hand panel compares these quantities given the strong prior density associated with the hypothetical data $n = 20$ and $m = 10$. The left hand panel compares these quantities given the weak prior density associated with the hypothetical data $n = 2$ and $m = 1$.

Bayesian Point Estimates When computing a Bayesian point estimate we want to find a single value that conveys information about the centre of the posterior distribution – what value of the parameter is most likely in some sense. The estimates we computed for the binomial problem above maximize the posterior density, and for this reason are called the posterior modes. The posterior mode will be a good measure of the centre of the posterior density in many problems but always. For example, if the distribution is very skewed or has several modes then the posterior distribution may be far from what most people interpret as the centre. Two alternatives that are commonly used instead are the posterior mean and the posterior median – the mean and median of the posterior density. As we will see, these values are also easier to compute in more complicated problems. For the binomial model, the posterior mode, mean, and median are all very close because the posterior density is unimodal and almost symmetric about the mode.

Posterior Standard Deviation The posterior standard deviation is the Bayesian equivalent of the standard error. As the name implies, it is the standard deviation of the posterior density. For the binomial experiment, the posterior standard deviation of p is .06 with the strong prior and .07 with the weak prior, which are both very close to the standard error for the maximum likelihood estimator given above. Note, however, that the interpretation of these two values is quite different. The posterior standard deviation is a direct statement about the uncertainty in the true value of the parameter p. The standard error is a statement about the uncertainty of $\hat{p}$, how much would the estimate vary in repeated trials of the experiment with the same underlying value of p. One needs to know that Bayesian inference depends on a different interpretation of probability to fully understand this difference, but this is beyond the scope of this introduction.

Confidence Intervals versus Credible Intervals Consider the 95% credible interval we computed for p in the previous section. The interpretation of this quantity is that if the true value of the parameter were exactly equal to the estimate, $.8$, and the experiment were repeated many times (thousands or millions) then 95% of the intervals constructed in this way would cover the true parameter value. What can we say after collecting only one data set and computing one interval? Very little can be said because we have no way of knowing if we have the interval we have computed is one of the 95% that cover the true value or one of the 5% that doesn't. The chances are good that the interval does cover the true parameter values, but it is not a certainty.

The Bayesian equivalent of the confidence interval is the credible interval. A 95% credible interval is any interval which contains a 95% percent of the posterior probability. Because the posterior density is a true probability density, we can compute quantiles and percentiles of the parameter. The simplest 95% credible interval is bounded by the 2.5^{th} and 97.5^{th} percentiles. This interval is called a symmetric credible interval because it removes equal probability (2.5%) from both tails of the distribution. Other 95% credible intervals can be formed by removing $\alpha < 5$ percent from the left tail and $5 - \alpha$ percent from the right tail, and these intervals may be better in that they may be shorter but still cover 95% of the posterior distribution. However, they are often much harder to compute and so we will only consider the symmetric intervals.

Exercises

1. Bayesian inference for the binomial
   The file `Intro_to_splines\Exercises\binomial_1.R` contains code for plotting the prior density, likelihood function, and posterior density for the binomial model described in this section. Vary the values of `N`, `n`, and `alpha` to see how the shapes of these functions and the corresponding posterior summaries are affected.

2 MCMC Methods for the Binomial Model

2.1 An Introduction to MCMC

Sampling from the Posterior Distribution In the case of the simple binomial model we can compute the posterior summaries (the posterior mean or mode, standard deviation, and approximate credible intervals) by hand. However, this is rarely the case. The posterior distribution can be very complicated for models that include

large numbers of parameters and this may make it impossible to compute summary statistics analytically. How then can we make inference about the parameter values?

Because the posterior density is a true probability density, one way to do this is to sample values from the distribution and then to compute the sample statistics. Imagine that we have a box full of numbered pieces of paper and we wish to compute the mean. The exact approach would require taking all of the pieces of paper out of the box, recording the values, and then computing their average. This would work if the box was small, but if there were many pieces of paper in the box then it would take a very long time. Alternatively, we could collect a sample of values by drawing a small number of pieces of paper from the box and use the mean of these values to estimate the true mean. Provided we draw enough values and the numbers are fairly regular, our approximation should be very close to the true mean. The same procedure can also be used to approximate any other characteristic of the distribution of the numbers in the box – like the standard deviation or the quantiles.

When confronted with complicated Bayesian problems we use exactly the same concept to approximate the posterior summary statistics. Rather than trying to compute the posterior summaries exactly, we generate a sample of values from the posterior density and then use these values to approximate posterior means and standard deviations, credible intervals, or any other quantity we are interested in.

The Very Basics of Markov chain Monte Carlo Sampling from an arbitrary distribution can be very complicated, but there is one method that can be applied quite widely and that has opened the doors of Bayesian statistics in recent years. This is the method of Markov chain Monte Carlo (MCMC for short).

In probability, a Markov chain is a sequence of events whose distribution depends only on the outcome of the previous event. If the probability of being fog-bound at Arcata airport on one day depends only on whether or not the airport was fog-bound the day before, then this sequence of events forms a Markov chain. One of the cornerstones of Markov chain theory says that if the probabilities associated with different events are constructed in the correct way, and the chain is run long enough, then the distribution of the events can be made equal to any arbitrary distribution – including the posterior distribution. The theory of how to construct these chains to achieve the proper distribution can be quite complicated, but suffice it to say that there are some general methods that can be used in most problems and that are implemented in available software packages. Here we will focus on how to use one of the most developed and widely used packages, `WinBUGS` (or its open source partner `OpenBUGS`).

2.2 Introduction to WinBUGS

The file `Intro_to_splines\Exercises\binomial_model_winbugs.txt` contains code to implement the Bayesian binomial model in `WinBUGS` . This is a simple text file, so you can view and modify it using any text editor (like `notepad` or `wordpad` on `Windows`). To specify a model in `WinBUGS` we need to provide three pieces of information: 1) a model which defines the structure of the likelihood and the prior distributions for each parameter, 2) a listing of the data, and 3) a listing of initial values to start the chain (remember that one step in the Markov chain depends on the previous step so we need to give the software some values to get started). These pieces of information can be provided in a single file or in separate files, as you prefer. All of the information in a single file for the binomial model because the code is short.

To run the model in `WinBUGS` or `OpenBUGS`, open the file using the menu option `File > Open...` (you may have to select `Text (*.txt)` in the file type selector in the bottom right corner of the file selector to see the appropriate file). The `BUGS` language is fairly simple, but we won't go into the details here. However, there are a few things that you should now:

1. Comments begin with the # symbol and text on a line following this symbol is ignored.
2. The definition of the model must begin with the keyword `model`.
3. The listings of data and initial values must begin with the keyword `list` and be enclosed in parentheses.

You can read about the `BUGS` language in the manuals available under the `Help` menu if you are interested in writing your own models.

To initialize `WinBUGS` we must load the three pieces of the specification in the correct order. Open the model specification tool via `Model > Specification...` and follow these steps:

1. Check the model:
   Highlight the keyword `model` at the start of the model definition and click on the `check model` button in the `Specification Tool` window. `WinBUGS` will check the syntax of the model and, if correct, will print "model is syntactically correct" in the notification bar at the bottom of the main window. The buttons labelled `load data` and `compile` will also become active.

2. Load the data:
   Highlight the keyword `list` at the start of the data definition and then click `load data`. The message "data loaded" will appear in the notification bar if the data is loaded successfully.

3. Compile the model:
   Click on the `compile` button. `WinBUGS` will check that the appropriate data has been loaded and then construct the Markov chain. The message "model compiled" will appear in the notification bar if this is successful.

4. Load the initial values:
   Finally, highlight the keyword `list` at the start of the initial values and click on `load inits`. If the initial values are read successfully then the message "model is initialized" will appear in the notification bar. (Note: it is not necessary to provide initial values. `WinBUGS` can generate a set of initial values by itself, but this can sometimes lead to troubles and so it is best to supply your own initial values if you can.)

As `WinBUGS` says, the model is now initialized and ready to run.

Before we actually run the chain,however, we need to tell the software what variables to keep track of as the chain iterates. In this simple model there is only one parameter, p, but in complex models we may only want to keep track of a subset of parameters of interest. Open the `Sample Monitor Tool` via `Inference > Samples...` In the `node` dialogue enter `p` and click on `set`. This tells `WinBUGS` to keep track of the values of p as the chain iterates so that we can later compute the posterior summary statistics.

We are now ready to run the Markov chain. Open the `Update Tool` via `Model > Update`. Change the number of updates from 1000 to 10,000 – we'll discuss later how to choose this number – and click on the `update` button. The message "model is updating" will appear in the notification bar and the iteration counter in the `Update Tool` window will increment. When the updater is finished, a message will be displayed in the notification bar telling you how long it took to complete the iterations. That's it; the Markov chain has been run producing a sample of 10,000 values from the posterior distribution of p.

To compute posterior summary statistics, return to the `Sample Monitor Tool`, and type or select `p` in the `node` dialogue. Now click on the `stats` button. This will open a new window containing summary statistics from the sample of p including the posterior mean and standard deviation, the MC errror (which we is discussed below) and the values of the 2.5^{th}, 50^{th}, and 97.5^{th} percentiles. If all has gone well, the mean should be very close to .68 and the standard error close to .06 (the theoretical values we computed above). The 2.5^{th} and 97.5^{th} percentiles should also be close the bounds of the 95% credible interval we computed. Clicking on `density` button will produce a plot of the posterior density of p. This should be very similar to the density we plotted in `R` , but will be somewhat rough because of the error in sampling.

3 Bayesian Inference for Two-Stage Capture Recapture Models

3.1 The Simple-Petersen Model

The Two-Stage Capture-Recapture Model In modelling the data from a two-stage capture-recapture model, we have to make use of the binomial model twice – once to model the sample obtained from the population of marked individuals and once to model the sample obtained from the population of unmarked individuals. In essence, the model for the marked individuals provides information to estimate the capture probability which is then used to draw inference about the size of the unmarked population.

Let n and U denote the total numbers of marked and unmarked individuals in the population and m and u the numbers captured. Denoting the capture probability by p, the model components are:

1. $m \sim$ Binomial(n, p), and

2. $u \sim$ Binomial(U, p).

The likelihood function is then given by the product of the probabilities for the observed m and u:

$$L(p, U|n, m, u) = \binom{n}{m} p^n (1-p)^{n-m} \cdot \binom{U}{u} p^u (1-p)^{U-u}.$$

A maximum likelihood estimate can be computed numerically for specific data, but we will turn instead to the Bayesian solution.

Bayesian Formulation To complete the Bayesian formulation of the model we need to define prior distributions for the two unknown quantities, p and U. There are several different ways that the priors for these parameters can be specified, but these issues are beyond the level of this workshop. For now, we will define the same prior density for the capture probability as we used in the previous section – called the beta density. For the size of the unmarked population, we will use a slightly more complex prior density which is constructed by assuming that any value of $log(U)$ is equally likely (i.e., the prior density of $\log(U)$ is proportional to 1). This is called Jeffrey's prior for U and represents one concept of complete uncertainty about the value of U, though the theoretical argument is somewhat complex. Interested readers can find more information about Jeffrey's priors and the definition of prior densities in general in any introductory textbook on Bayesian statistics.

Implementation in `WinBUGS` The file `Intro_to_splines\Examples\cr_winbugs.txt` contains code which implements the two-stage capture-recapture model in `WinBUGS` . The file comprises three sections as in the code for the binomial model: the model definition, the data, and the initial parameter values. The model is very similar to that of the binomial model, except that it now contains two statements for the likelihood, one specifying the binomial distribution for m and one for u. The data contains the same values for m and n and the parameters of the prior distribution for p, but also includes a value for u. Run the model in `WinBUGS` by following the steps above to obtain posterior summaries for both p and U (Note: you will have to enter both parameters separately in the `Sample Monitor Tool`).

If the code runs successfully, then the posterior mean of p should be similar to what we obtained in the previous section, and the posterior mean of U should be near 1470. Note that this is actually quite a lot larger than the Lincoln-Petersen estimate for U given by $un/m = 1250$. The reason for this is that we have included strong prior information about p which has led us to believe that the capture-probability is lower than $24/30 = .8$ and so hence that the size of the unmarked population is greater than $1000/.8 = 1250$.

Exercises

1. Bayesian inference for the simple Petersen model
   Use the code provided in the file `Intro_to_splines\Examples\cr_winbugs.txt` to implement the Bayesian formulation of the simple Petersen model. Change the parameters of the beta prior density for p to both be equal to 10 and recompute the posterior summary statistics.

3.2 The Stratified-Petersen Model

The Stratified Model The simple model is appropriate when it can be assumed that the capture probability is the same over the entire experiment. However, this is rarely the case. To account for differences over time, we need to allow the capture probabilities to vary over time. The way to do this is by stratifying the data – essentially, by fitting separate models to each stratum (day or week) of the experiment. The new data set will contain separate records for the numbers of marked fish released and recaptured in each stratum as well as the number of unmarked fish captured each stratum, denoted by n_i, m_i and u_i respectively for stratum i. The new model components are:

1. $n_i \sim$ Binomial(N_i, p_i), and
2. $u_i \sim$ Binomial(U_i, p_i)

where the capture probability, p_i, is assumed to be the same for all fish passing the traps in one period but is allowed to change from one period to the next. The number of periods in the experiment will be denoted by s.

Implementation in WinBUGS The file Intro_to_splines\Examples\cr_stratified_winbugs.txt contains code for implementing the stratified-Petersen model in WinBUGS along with simulated data and initial values. If we run the model in WinBUGS then we can generate posterior summaries for all of the values of U_i using the stats button as we did previously. However, the numerical summaries are hard to interpret for all values of U_i simultaneously. To help visualize the results, WinBUGS has some limited, built-in graphical capabilities. After running the model, open the Comparison Tool via Inference > Compare..., enter U in the node dialogue, and click box plot. This will produce a boxplot of the values of U_i sampled for each stratum. The values displayed for each box are the extents of the 95% credible interval (indicated by the whiskers extending from the top and bottom of each box), the extents of the 50% credible interval (indicated by the bounds of the green box), and the posterior mean (indicated by the horizontal line through the box). For the simulated data $U_i = 10,000$ for every i, and the posterior means should all be roughly equal to this value.

Exercises

1. Bayesian inference for the stratified-Petersen model
   Implement the stratified-Petersen model for the simulated data set and produce a boxplot for the values of p (if you didn't specify p in the sample monitor than you will need to do so and re-run the chain). Notice that the 95% credible intervals are much wider for some values of p_i than for others. Why is this?

3.3 Hierarchical Modelling – A Compromise

Hierarchical Models The values of p_i which have larger credible intervals in the example are those for which we only marked and released 10 individuals instead of 1000. We obtain less information about the capture-probability in these strata because fewer individuals were marked, and so the estimates are less precise. If we believe that the capture probabilities from one week to the next vary widely and are completely unrelated then this is the best we can do. However, it seems reasonable to believe that the capture probabilities in different weeks are similar to each other, and that information from the weeks where many individuals were marked be used to improve the estimates in the weeks where there is little data. The hierarchical model provides a way of doing this without having to make the assumption that the capture probability is exactly the same in every strata.

In essence, the hierarchical model builds on the stratified model by incorporating further structure in the prior distribution of the capture probabilities. To construct a hierarchical prior distribution, we model the values of p_i as a random sample from some distribution whose parameters are unknown. The periods for which we have lots of data then provide information about these parameters, which in turn provide information about the values of p_i in the periods with little data. Suppose, for example, that the values m_i were very all very close to 500 in all of the strata in which 1000 marked fish released. Unless we believe that the number of marked fish released is somehow related to the capture probability, it seems reasonable to believe that the capture probability is also very close to .5 in the strata in which only 10 marked fish were released. Using this type of logic, the hierarchical model shares information between the strata and provides precise estimates even in the weeks with little data.

Implementation in WinBUGS Code to implement the hierarchical model is provided in the file Intro_to_splines\Examples\cr_hierarchical_winbugs.txt using the same simulated data set as provided for the stratified model. In this code, the capture probabilities are modelled by a normal distribution with unknown mean and variance on the logistic scale (i.e., $\log(p_i/(1-p_i)) \sim N(\mu, \tau^2)$. Run this model in WinBUGS and compute the posterior summaries for the capture probabilities. The estimates should be more precise than for the stratified model (i.e., smaller posterior standard deviations and narrower credible intervals).

Exercises

1. Bayesian inference for the hierarchical Petersen model
   The hierarchical model can be used even in the more extreme case in which no marked fish are released in one period or the number of recoveries is missing, so that there is no direct information about the capture probability in that period. The file Intro_to_splines\Examples\cr_hierarchical_2_winbugs.txt contains the code for fitting the hierarchical model to the simulated data, except that some of the values of n_i have been replaced by the value NA, WinBUGS notation for missing data. Run the model and produce

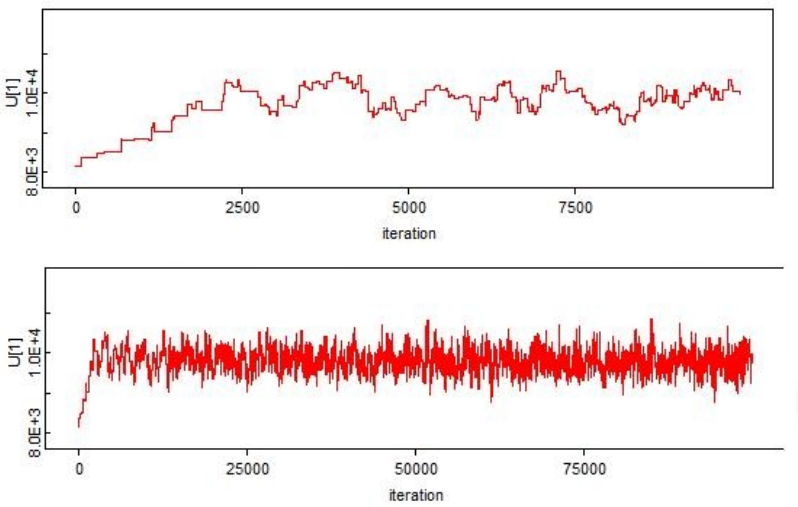


Figure 5: Traceplots for the parameter U_1 in the hierarchical model after 10,000 iterations (top) and 100,000 iterations (bottom).

boxplots for U and p. Note that you will have to use the `gen inits` button in the `Specification Tool` window to generate initial values for the missing data after loading the initial values for p and U.

4 Further Issues in Bayesian Statistics and MCMC

4.1 Monitoring MCMC Convergence

Convergence and Mixing One important question that must be asked when sampling from any distribution via MCMC simulation is: how many samples are needed to accurately approximate the characteristics of the posterior distribution? What makes this questions difficult to answer is the samples generated on successive iterations are not independent of one another. Frequently, the values from one iteration and the next will be highly correlated, and a very large number of iterations will be necessary to make sure that the sample covers the entire range of the distribution.

As an analogy, suppose that you were to estimate the mean height above sea level of California by measuring the elevation at every step of a random walk starting in Eureka. You would never reach the mountains, unless you walked for a very long time, and you would greatly underestimate the average elevation. You would like to make much larger moves between each of the measurements, ideally, choosing each location completely at random.

The same is true of Markov chain sampling. We would like our Markov chain to move about the space covered by the distribution freely. When this is the case, and the outcome of one iteration has little effect on the next iteration, we say that the chain is mixing quickly. If the outcomes on successive iterations are highly linked then we say that the chain is mixing slowly. If the chain is mixing slowly then it will have to be run for a long time until we can be sure that our sample properly represents the posterior distribution.

Trace Plots The simplest tool for visualizing the convergence of a Markov chain is the trace plot: the plot of the values generated from the Markov chain versus the iteration number. A traceplot of the most recent values obtained from the chain can be produced in `WinBUGS` by pressing the `trace` button in the `Sample Monitor Tool`. A traceplot of the entire chain is produced by `history`. The top panel of Figure 5 shows the traceplot for the parameter U_1 after running the Markov chain for the hierarchical-Petersen model for 10,000 iterations. The plot shows that the value of U_1 is mixing slowly, meaning that it changes only a little on each iteration. It is difficult to know if the sample of 10,000 values covers the entire distribution or if there could be other plausible regions that the chain has simply not reached yet. The figure in the bottom panel shows the traceplot for U_1 after running the chain for 100,000 iterations in total. This plot shows that the chain is mixing well, moving back and

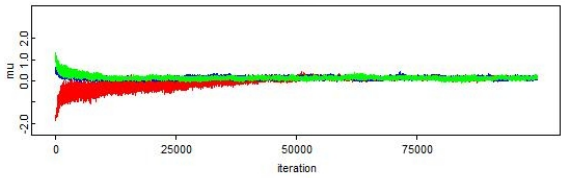


Figure 6: Traceplot for the parameter μ in the hierarchical model obtained by running three chains starting at widely spaced initial values for 100,000 iterations.

forth over the space, and suggests that the sample of 100,000 values is more than enough to produce accurate approximations of the posterior summaries.

MC Error Another tool for judging whether enough iterations have been generated is the MC error printed along with the posterior summary statistics. This value represents the amount of uncertainty in the posterior summaries that is due to the approximation by a finite sample. It will always decrease as the sample size increases, and we would like the MC error to be much smaller than the posterior standard deviation, 5% or less, to ensure that the posterior summary statistics are accurate. In one run of the hierarchical model, the MC error for U_1 was approximately 10% of the posterior standard deviation after 10,000 iterations and 4% after 100,000 iterations. This again suggests that 10,000 iterations are insufficient, but 100,000 are adequate. The same check should be applied to all other parameters in the model.

Thinning If the Markov chain is mixing slowly then it may not be necessary to retain all of the values from the chain. Although there is no harm in doing so, it can take a lot of disk space to store the outcomes from many thousand or million iterations and if the values are highly correlated then the same information can be conveyed by a much smaller sample. In this case, one can retain only a subset of the iterations – perhaps every 10^{th} or 50^{th} iteration depending on how correlated the values are. The chain can be thinned in `WinBUGS` by entering a number greater than 1 in the `thin` dialogue in the `Update Tool` window before running the chain. Alternatively, statistics can be computed with a thinned sample, speeding the calculations but not saving disk space, by changing the value in the `thin` dialogue fo the `Sample Monitor` window.

Burn-in, Multiple Chains, and the Brooks-Gelman-Rubin Diagnostic Although the chain in the bottom panel of Figure 5 seems to be mixing adequately the figure illustrates another important concern of MCMC sampling. Notice that the values from the very beginning of the chain, approximately the first 2500, are quite different from the remaining values. This is called the burn-in period of the chain and arises because the initial values don't represent a proper sample from the distribution. We need to remove the burn-in period from the sample in order to compute proper estimates of the posterior summary statistics. However, how many iterations to remove is not always an easy question to answer.

One way to assess the length of the burn-in period is to start several chains from widely spaced initial values and then compare their behaviour. The chains will have converged when they begin to produce similar values. In `WinBUGS` , we can run multiple chains by setting the `num of chains` value in the `Specification Tool` window *after loading the data and before compiling the model* and loading separate lists of initial values for each of the chains.

Figure 6 shows the traceplot for the parameter μ from three chains run in `WinBUGS` starting from widely spaced initial values for all parameters. The values from the three chains are represented by the different colours. The plot shows that the chains initially produce very different – the values of the red series are all below 0 at first while those in the blue and green series are all above 0. However, the values become closer as the chain proceeds and appear to coalesce somewhere near iteration 50,000. This suggests that the first 50,000 iterations for any chain are unreliable as they belong to the burn-in period and should be discarded before computing the posterior summaries.

Judging when the multiple chains become similar provides a quick guess at the length of the burn-in period but is still quite subjective. To avoid the subjectivity in interpreting the traceplot, we can use diagnostic measures

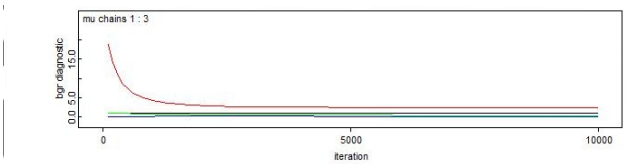


Figure 7: Brooks-Gelman-Rubin diagnostic plot for the parameter μ after 10,000 iterations.

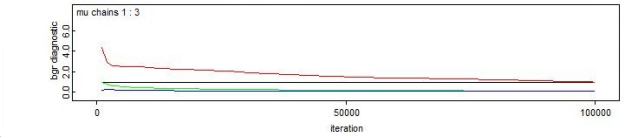


Figure 8: Brooks-Gelman-Rubin diagnostic plot for the parameter μ after 100,000 iterations.

to infer exactly when the chains converge.

The most common diagnostic measure, and the one that is built into `WinBUGS` , is the Brooks-Gelman-Rubin diagnostic. This diagnostic is compares the variance of the values sampled by each of the three chains separately with the variance of the values in the pooled sample from all of the chains. If the chains have converged then the variances from the separate samples and the combined sample should be very similar. The specific implementation in `WinBUGS` compares the average width of the symmetric 80% credible interval in each chain with the width of the interval computed by pooling all three chains together. Clicking the `brg diag` button in the `Sample Monitor Tool` will generate a plot illustrating the evolution of this diagnostic value over the iterations of the parallel chains.

Figure 7 shows the diagnostic plot generated by `WinBUGS` for the first 10,000 values of μ sampled from the 3 chains. The blue line represents the average width of the 80% credible intervals computed from the 3 separate chains, the green line the width of the 80% interval computed from the pooled data, and the red line the ratio of these two values. At convergence, the ratio will be 1. However, the plot shows that the ratio is initially greater than 15 and still as high as 5 after 10,000 iterations. This suggests that many more iterations are required to reach convergence. Figure 8 shows the same plot for all 100,000 iterations. This plot shows that the ratio is 2.0 after 50,000 iterations but very close to 1.0 by the 100,000th iteration. This suggests that the burn-in period is in fact much longer than expected based on visual inspection of the multi-chain traceplot. Note that the same diagnostics should be computed for all other parameters in the model to ensure convergence of all values.

Exercises

1. Bayesian inference for the hierarchical Petersen model: convergence diagnostics
   The file `Intro_to_splines\Examples\cr_hierarchical_bgr_winbugs.txt` contains the code to run three parallel chains for the hierarchical capture-recapture model. Use this file to plot the traceplots and compute the Brooks-Gelman-Rubin diagnostics. To initialize the model you will need to enter 3 in the `num of chains` dialogue and then load the three sets of initial values one at a time.

4.2 Model Selection and the DIC

The Priniple of Parsimony An important consideration in any complex analysis is the need to choose between competing models for the same data. Here we will focus on one method called the Deviance Information Criterion or the DIC for short. One way to compare the fit of two different models is to compare the values of the likelihood evaluated at the parameter estimates. The model with the larger likelihood fits the data better. However, the likelihood always increases (or stays the same) when parameters are added to a model, and so considering the likelihood alone will always select the model with the most parameters. Instead, we would like to select the model which provides the best fit to the data with the fewest number of parameters. This is the principle of parsimony.

The Deviance Information Criterion The DIC attempts to identify the most parsimonious model by creating a balance between the likelihood and the number of parameters in the model. Rather than using the likelihood directly, the DIC is computed from minus twice the value of the log of the likelihood, a quantity called the deviance which plays a very important role in classical statistics. To this is then added an estimate of the number of unique and estimable parameters in the model, p_D, called the effective number of parameters. The reason that this is an estimate is that the exact number of parameters is not well defined for complex, multi-level models. For example, the variation between the capture probabilities in the hierarchical-Petersen model depends on the parameter τ. If τ is zero then the capture probabilities will all be the same and so there is really only one parameter for the capture probabilities, regardless of the number of strata. If τ is very large then the capture probabilities will all be different and the number of parameters will be equal to the number of strata. However, if τ is moderate in size, and there is some sharing of information between the strata, then the number of parameters will lie somewhere in between. Moreover, some parameters appearing in the model may not actually be estimable and these should be removed from the count of parameters. The p_D value is designed to account for such issues in computing the number of parameters for a model.

Because the deviance is opposite sign to the likelihood, the best fitting model will be the one with the lowest DIC (i.e., the highest penalized likelihood). In practice, the value of the DIC will vary somewhat for a given model because of error in the MCMC sampling and noise in the data. Differences of 5 or more in the DIC values for two models present clear evidence in favour of one model over another. Differences of 2 or 3 provide weaker evidence. Smaller differences may very well be the result of random variation and should not be considered as support for one model over another.

Computing the Deviance Information Criterion in `WinBUGS` In `WinBUGS` , the `DIC Tool` is accessed via `Model > DIC`. Set the monitor after the burn-in period is complete and then run the chain for more iterations. Clicking the `stats` button will then provide the value of the DIC and p_D, along with some other information. The output for the hierarchical-Petersen model applied to the simulated data (with some variation for sampling error) is:

	Dbar	Dhat	DIC	pD
m	195.2	177.4	213.0	17.85
u	318.1	289.6	346.6	28.47
total	513.3	467.0	559.6	46.32

The bottom row of this table provides the DIC and p_D values for the entire model. The DIC is 559.6 and the estimated number of parameters is 46.3. In the rows above this, `WinBUGS` attempts to decompose the DIC according to the different components in the data. The two-stage capture recapture model comprises two data components: the numbers of marked fish captured each day (m_i) and numbers of unmarked fish captured each day (u_i). The first line of the table describes the contribution to the DIC and p_D for the model of the marked fish. The model depends only on the capture probabilities and so we might expect the value of p_D to be close to the number of strata, which is 30 for this data set. However, there are 11 strata in the data set with very low numbers of marked fish and so the effective number of parameters is estimated to be around 18. The second row describes the contributions for the unmarked fish. Because the data for the unmarked fish provides no information about the capture probabilities these parameters are not counted. Only the U_i are counted and the effective number of parameters is 28, very close to the number of strata.

Exercises

1. Bayesian inference for two-stage capture-recapture: model selection
   The file `Intro_to_splines\Examples\cr_stratified_dic_winbugs.txt` contains code for a slightly modified version of the Bayesian stratified-Petersen model. Run this model for 100,000 iterations, set the DIC monitor, and then run the model for a further 100,000 iterations. Compare the DIC for this model with the hierarchical model. Which model is selected?

4.3 Goodness-of-Fit and Bayesian p-values

The DIC provides a method for ranking different models of a given data set, but there is no guarantee that even the highest ranked model actually fits the data well. To assess this, we need some way to check that the selected model adequately describes the data. In Bayesian statistics, the fit of a model can be tested by generating new data from the model and then comparing it to the observed data. If the model is good then the simulated data and observed data should be similar. If the model is poor then there will be systematic differences that arise between the observed and simulated data sets.

To compare the observed and simulated data we define some mathematical criterion, called a discrepancy measure, that is a function of the both data and the model parameters. For example, the observed and expected counts of unmarked individuals captured each in each strata can be compared using the discrepancy measure:

$$\sum_{i=1}^{s}(\sqrt{U_i p_i} - \sqrt{u_i})^2$$

based on the Tukey difference.To assess the fit of the model, first sample a set of parameters from the posterior distribution (i.e., extract the parameter values for one MCMC iteration) and simulate a new data set using these parameters. Then compute the value of the chosen discrepancy measure for both the observed data and the simulated data. If the discrepancy is larger for the observed data, then this is evidence that the model does not fit well. Of course, there is noise in the process of simulation and so it is possible that the simulated data may have a higher or lower discrepancy simply by chance. To guard against this, repeat the process many times and compute the proportion of times that the discrepancy for the simulated data is higher than that of the observed data. This proportion is called the Bayesian p-value. If the proportion is low then the model consistently fits the simulated data better than the observed data, which suggests that the model does not describe the observed data adequately. If the proportion is close to .5 then the p-value provides no evidence that the fit of the model is inadequate.

The discrepancy measures above is specifically designed assess the component modelling the capture of the unmarked individuals. Other discrepancy measures can be defined to test that the model accurately describes other components of the data, like the capture of marked individuals or the system as a whole. Different measures of discrepancy that assess different parts of the model will isolate the components of the model that are inadequate and guide the construction new models that better describe the data.

5 Bayesian Penalized Splines

Recall that the smoothness of a spline constructed from the B-spline basis depends on the differences between the spline coefficients. The spline is smooth if the differences are small and may have sharp changes if the differences are large. In the classical setting, the spline is fit by minimizing the penalized least squares criterion and the smoothness of the spline is determined by the value of the smoothing parameter, λ. However, different values of λ can produce vastly different splines for the same data.

The Bayesian implementation of penalized splines removes the decision about λ by assigning this value a prior density. Moreover, inferences will account for the uncertainty in the value of λ which is more honest than trying several values of λ and choosing the best by visual inspection. Ideally, the prior density should favour smoothness but allow for sharp changes in the fit if this is warranted. Such a prior can be defined by the gamma density:

$$\lambda \sim \Gamma(\alpha, \beta)$$

where the parameters α and β are chosen so that the density is concentrated near 0 but is highly skewed and allows for some very large values. The full Bayesian penalized spline model defines a hierarchical model for the coefficients of the spline. The first level of the hierarchy assigns a normal prior to the individual differences between coefficients:

$$(b_k - b_{k-1}) \sim N(b_{k-1} - b_{k-2}, (1/\lambda)^2)$$

and the second assigns the gamma prior to λ. Although the interpretation of λ is more strict in the Bayesian formulation, it is now the inverse variance of the differences in the coefficients, it plays exactly the same role as the smoothing parameter in the classical setting. The spline is smooth if λ is large and the variance of the coefficients is small.

Appendix C: How to use the R-wrapper for the spline models

C.1 Fitting the spline model for the combined wild and hatchery fish.

All of the routines used in the population estimation are available in an R-package called BTSPAS (Bayesian Time Stratified Petersen Analysis System) that can be downloaded from the CRAN and R-forge libraries.

You can join the BTSPAS mailing list to be kept up-to-date on revisions to the program by visiting the website:

https://lists.r-forge.r-project.org/cgi-bin/mailman/listinfo/btspas-discussion

These spline models are fit using Bayesian methods and the WinBugs/OpenBugs program when called using the R statistical program. This Appendix will illustrate how to format the data and fit the spline models.

1. Download and install WinBugs. Visit

http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml

to download the program. The webpage also has instructions on how to install it on your system. By default the program is installed at:

c:/Program Files/WinBUGS14/

If WinBugs is installed in a different directory, please record the location.

2. Download and install OpenBugs. Visit:

http://mathstat.helsinki.fi/openbugs/SoftwareFrames.html

to download the program. The webpage also has instructions on how to install it on your system. By default the program is installed at:

c:/Program Files/OpenBugs/

or c:/Program Files/Bugs/

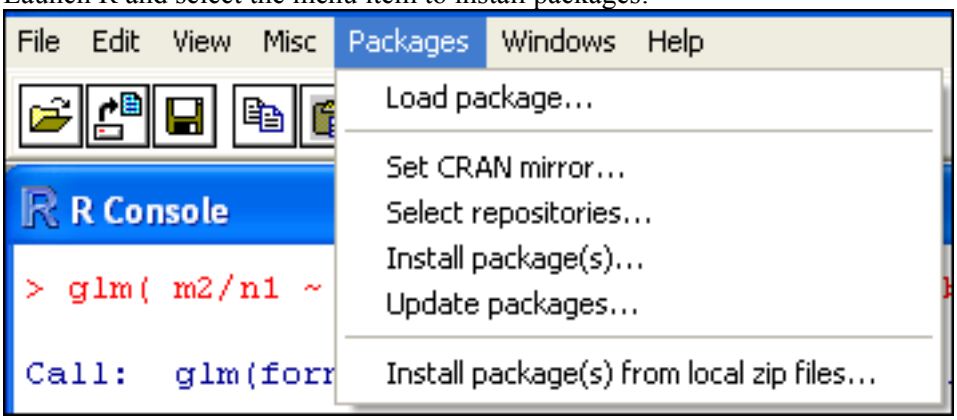
If OpenBugs is installed in a different directory, please record the location.

3. Download and install R. Visit

http://cran.r-project.org/

The directory where R is installed is not crucial

4. The software of this document makes use of several "packages" that need to be installed into your R system. Launch R and select the menu item to install packages:



The R-system will prompt you to select a local "mirror" where the packages are available. Select a location close to you, the choice is not critical.

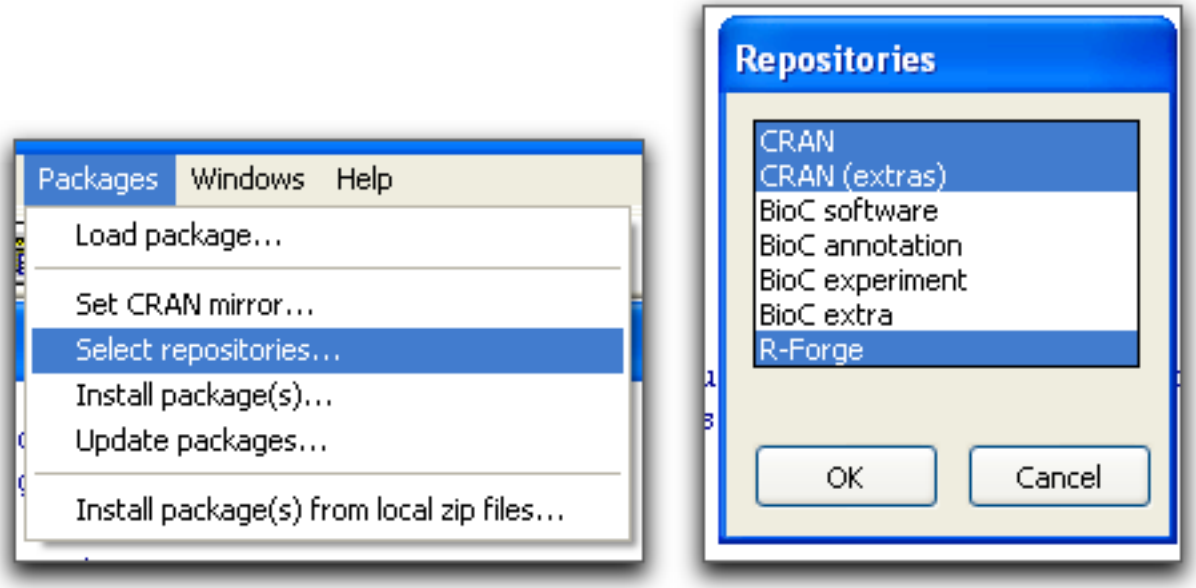
The R-system will then present an (alphabetic) list of packages. Scroll down the list and select the packages: BTSPAS, R2WinBUGS, coda, actuar, and BRugs (in any order). This may require multiple separate steps. After each package is selected, the R-system will go the mirror via the internet, download, and install the package. You should see a message on the R console if the installation is successful similar to the one shown below for each package installed.

```
> utils:::menuInstallPkgs()
--- Please select a CRAN mirror for use in this session ---
trying URL 'http://cran.stat.sfu.ca/bin/windows/contrib/2.9/actuar_1.0-2.zip'
Content type 'application/zip' length 1230196 bytes (1.2 Mb)
opened URL
downloaded 1.2 Mb

package 'actuar' successfully unpacked and MD5 sums checked

The downloaded packages are in
        C:\Documents and Settings\cschwarz\Local Settings\Temp\RtmpQhUhJa\downl$
updating HTML package descriptions
```

If the package is not available at CRAN, you may need to switch repositories:



and then search for the package. The development directory of BTSPAS is stored in the R-Forge repository and downloading from this will give the development version, but the CRAN repository has the latest production version.

The spline-based program also requires the use of the *BRugs* package, but this has been recently removed from the usual package libraries for some modifications. Fortunately, it can still be located using the

install.packages("BRugs")

command.

5. Create a directory to hold the R-wrappers and data. For example, create directory called *TrinityData on your desktop.*

Download a sample R-wrapper from
http://www.stat.sfu.ca/~cschwarz/Consulting/Trinty/Phase2.
Several examples are available in the *TrinityWrappersAndData* subdirectory.

For example, download the *JC_2003_CH.csv file* (containing the raw data for the Junction City 2003 Chinook example of this report), and the *JC_2003_CH_TSPDE.r* file containing the R-wrapper to read in the raw data and to call the program. Place this in the same directory as your data.

6. The TSPDE program has several inputs that are compulsory and several that are options for advanced users. The compulsory arguments are

Argument (case important)	Description
title	Character string that serves as the title for the analysis. For example title <- "JC 2002 Chinook"
prefix	Character string used as the prefix for created output files. For example prefix <- "JC-2002-CH-TSPDE" will create files of the form "JC-2002-CH-TSPDE-xxxx". Ensure that the prefix will result in valid file names.
time	A numeric vector of stratum numbers (e.g. Julian weeks). This is usually read in with the data (see below). Values do not have to start at 1 (i.e. the first Julian week can be 10), nor does it have to be contiguous. Missing stratum numbers are assumed to have $n_i = m_{ii} = u_i = 0$. The program will use a spline to interpolate population numbers for these missing strata.
n1, m2, u2	Numeric vectors of the number of fish marked and released (n1), the number recaptured in the stratum of release (m2) and the number of unmarked fish captured (u2). The vectors n1, m2, u2 should be same length as the *time* vector.
sampfrac	The sampling fraction used to inflate/deflate the u2 to account for sampling that was not complete during the stratum. For example, the strata are Julian weeks but the traps were only operating for 5 of the 7 days, then the corresponding sampling fraction should be specified as 5/7=.714. This vector should be the same length as the *time* vector.
jump.after	The spline model may not be able to react quickly to sudden increases in population sizes across strata. For example, hatchery fish tend to arrive in a large pulse. The jump.after vector list the stratum numbers (corresponding to elements in the *time* vector) **AFTER** which the population is allowed to jump. For example, *jump.after <- NULL* implies a single spline will be fit. The code: *jump.after <- c(12,22)* implies that the spline is allowed to jump AFTER the strata numbered 12 and 22, i.e. three separate splines will be bit. The values of jump.after should be taken from the elements of the *time* vector.

bad.m2	A vector of stratum numbers (possibly null) where the number of recaptures is suspect. For example, in the 2003 Chinook data, the number of recaptures in Julian week 41 was suspect. The program will set the m2 values for these strata to missing and will interpolate as needed. For example *bad.m2 <- NULL* implies that all the m2 values should be used. The code *bad.m2 <- c(14,33)* implies that the m2 values in strata numbers 14 and 33 will be ignored.
logitP.cov	(Optional). This is the matrix of covariates to model the logit(p) values using linear regression. This matrix would normally include a column of 1's for the intercept. If omitted, the TSPDE will fit a mean model for the logitPs. The matrix should have as many rows as the number of strata. CAUTION. If the *time* vector has missing entries, this may cause unexpected results!
Prior parameters	There are many parameters for the default priors for the spline model. Normally the default values can be used. Contact cschwarz@stat.sfu.ca for further information on modifying these parameters.
run.prob	(Optional). A numeric vector specifying the quantiles for which run timings should be computed. The default value is : run.prob=seq(0,1,.1) which corresponds to the 0, .1, .2, .3, …, 1.0 quantile (i.e. the deciles).
debug	(Optional, default=FALSE). If debug=TRUE, then a shortened call to the spline program with a limited number of iterations will be done. This is useful when running the program for the first few times to ensure that everything is working properly.
openbugs	(Optional, default=TRUE) There are two related programs to fit MCMC models. The OpenBugs program provides more control to the programmer and is the default method. If openbugs=FALSE is set, then the WinBugs program will be called to fit the model. The WinBugs program also provides a better debugging environment if things go wrong. Please contact cschwarz@stat.sfu.ca for more details.
WINBugs.directory	The directory where the WINBugs program is stored. Defaults to C:Program Files/WinBugs14. This is the standard installation location.
OPENBugs.directory	The directory where the OPENBugs program is stored. Defaults to C:Program Files/OPENBugs This is the standard installation location.
InitialSeed	(Optional, default= random integer between 1 and 1 billion (inclusive)). The initial seed used for the random number generators in OpenBugs.

The easiest way to bring the data into R in the proper format is to create a spreadsheet. The first row of the spreadsheet should have the labels *time, n1, m2, u2, sampfrac.* [The case of the labels is important.] If

covariates are to be used (such as log(flow)), enter these in subsequent columns labeled *covar1, covar2,* etc. The remaining rows of the spreadsheet should have the data values as numeric values.

Save the spreadsheet as a *.csv file in your data directory.

The function *read.csv(file="your filename", header=TRUE)* can be used to read the data into a data frame. The usual R commands can be used to extract the various data vectors, to set the *title, prefix, jump.after, and bad.m2* values. Here is a sample input file with annotations in **bold**,:

The *.csv file is JC-2002-CH.csv and the header names for the data columns were *TotalMarks*, *TotalRecaps*, *CH.TOT.0,* and *DaysOperating*.

```
Site       <- "Junction City"
SiteShort  <- "JC"
Year       <- 2003
FishType   <- 'CH'

jump.after <- c(22,39)  # Julian weeks after which jump occurs
input_file <- "JC-2003-CH.csv"
bad.m2     <- c(41)   # list Julian week with bad m2 values

# read in the data file for all the years/sites/species

Fish <- read.csv(input_file, header=TRUE)  # reads in file
Fish[1:5,]  # list the first few records

# Now to extract the subset of data, do any fancy adjustments, and fit
the data.

#create the prefix and title.
prefix <- paste(SiteShort,"-",Year,"-",FishType,"-TSPDE",sep
title <- paste(Site," ",Year," Species ",FishType,sep="")


# extract the data
n1 <- Fish$TotalMarks
m2 <- Fish$TotalRecaps
u2 <- Fish$CH.Tot.0
sampfrac <- Fish$DaysOperating/7
jweek <- Fish$SampleWeek + 8  # convert from sample week to Julian week
```

The TSPDE is then called using the following code segment. The arguments can be in any order.

```
library("BTSPAS")  # makes the BTSPAS functions available

results <- TimeStratPetersenDiagError_fit(
                title=title,
                prefix=prefix,
                time=jweek,
                n1=n1,
                m2=m2,
                u2=u2,
                sampfrac=sampfrac,
                jump.after=jump.after,
```

```
                bad.m2=bad.m2,
                debug=TRUE)
```

The TSPDE program will run in debug mode (a small number of MCMC iterations) and will return the results in an MCMC-object called *results*. This can be used for further analyses – contact cschwarz@stat.sfu.ca for details.

The sample code can be run against the sample data using the *source* facility of R. Under *File->Source* point R to the R-wrapper *JC.2003.CH.TSPDE.*

The main part of the output is a series of output files and graphical files created in the directory with the R-wrapper with the prefix passed to it. For example, the above code will generate the following files in the data directory.

File	Contents.
JC-2003-CH-TSPDE-results.txt	Text listing with the raw data, various Pooled-Petersen, Stratified-Petersen estimators, summary statistics on the MCMC posterior distributions (e.g. mean and SD on the total population estimate, the spline coefficients, etc), run timing quantiles etc.
JC-2003-CH-TSPDE-initialU.pdf JC-2003-CH-TSPDE-logU.pdf	Plots of the initial and final population estimates over time.
JC-2003-CH-TSPDE-logitP.pdf	Plot of the fitted logit(p)'s over time
JC-2003-CH-TSPDE-Utot-acf.pdf JC-2003-CH-TSPDE-Utot-posterior.pdf JC-2003-CH-TSPDE-Ntot-posterior.pdf	Plot of the posterior distribution of U-total and the autocorrelation plot of successive estimates of U-total form the MCMC chain. The N-total posterior plot would be used in two trap situations where the total population size is the sum of unmarked and marked fish.
JC-2003-CH-TSPDE-GOF.pdf	Plots of the posterior predictive distributions used for goodness of fit.
JC-2003-CH-TSPDE.data.txt JC-2003-CH-TSPDE.inits1.txt JC-2003-CH-TSPDE.inits2.txt JC-2003-CH-TSPDE.inits3.txt	Intermediate files used when calling OpenBugs/WinBugs from R. Normally not useful to the user. These contain the raw data and the initial values for 3 chains.
JC-2003-CH-TSPDE.CODAchain1.txt JC-2003-CH-TSPDE.CODAchain2.txt JC-2003-CH-TSPDE.CODAchain3.txt JC-2003-CH-TSPDE.CODAindex.txt	Intermediate files called when calling OpenBugs/WinBugs from R. Normally not useful for the user. These contain the results from the MCMC chains are automatically processed and summaries produced in other files.

JC.2003.CH.TSPDE-saved.Rdata	Save results in R data dump with the results form the spline-fit. This can be subsequently loaded with the *load()* command and further processed.

7. Interpretation of sample output. The following is selected portions of the output from the analysis of the JC 2003 Chinook data in the file *JC.2003.CH.TSPDE-results.txt*. The text not in Courier Font is the added explanation:

```
Time Stratified Petersen with Diagonal recaptures and error in
smoothed U -  Mon Jul 27 11:13:46 2009

 Junction City 2003 Species CH Results

*** Raw data ***
      time   n1  m2    u2 SampFrac logitPcov[1]
 [1,]    9    0   0  4135     0.43            1
 [2,]   10 1465  51 10452     1.14            1
 [3,]   11 1106 121  2199     0.86            1
 … … (some output omitted)
[32,]   40 4757 188 35118     1.00            1
[33,]   41 2876   8 34534     1.00            1
[34,]   42 3989  81 14960     1.00            1
```

The sample data are listed as read in. If there are no covariates, the mean logit (capture probability) is estimated which is equivalent to fitting a model for covariates with only the intercept (a column of 1's) on the right portion of the data listing.

```
Jump point are after strata:  22 39
```

Separate splines are fit when large changes in population numbers are expected. In this example, large jumps in the run size occur AFTER Julian weeks 22 and 39, i.e. in Julian weeks 23 and 40.

```
*** Pooled Petersen Estimate prior to fixing bad m2 values  ***

The following strata are excluded because n1=0 or NA values in m2 or
u2 : 9

Total n1= 50489 ;  m2= 2486 ;  u2= 205860
Est U(total)  4,179,300   (SE  81,714 )
```

The pooled-Petersen using ALL of the data are presented.

```
*** Pooled Petersen Estimate after fixing bad m2 values  ***

The following strata had m2 set to missing:  41
The following strata are excluded because n1=0 or NA values in m2 or
u2: 9 41
```

```
Total n1= 47613 ;  m2= 2478 ;  u2= 171326
Est U(total)  3,290,666   (SE  64,348 )
```

The pooled-Petersen after excluding strata with no releases ($n_i = 0$ in Julian week 9) or deliberately excluded because the number of recaptures appears to be odd (Julian week 41).

```
*** Stratified Petersen Estimator for each stratum PRIOR to removing
bad m2 values ***

      time   n1  m2    u2     U[i] SE(U[i])
 [1,]    9    0   0  4135     4135        0
 [2,]   10 1465  51 10452   294693    40133
 [3,]   11 1106 121  2199    19961     1704
… (some output omitted)
[37,]   45  485  14   679    22031     5596
[38,]   46  115   4   154     3595     1568

Est U(total)  16,018,079   (SE  3,680,070 )
```

The simple stratified-Petersen estimator is computed using data for each strata and summing over all strata.

```
*** Stratified Petersen Estimator for each stratum AFTER removing bad
m2 values ***

      time   n1  m2    u2    U[i] SE(U[i])
 [1,]    9    0   0  4135    4135        0
 [2,]   10 1465  51 10452  294693    40133
 [3,]   11 1106 121  2199   19961     1704
… (some output omitted)
[32,]   40 4757 188 35118  884106    63018
[33,]   41 2876  NA 34534      NA       NA
[34,]   42 3989  81 14960  727979    79559
[35,]   43 1755  27  3643  228530    42837
[36,]   44 1527  30  1811   89313    15873
[37,]   45  485  14   679   22031     5596
[38,]   46  115   4   154    3595     1568

Est U(total)  4,978,392   (SE  209,845 )
```

The simple stratified-Petersen estimator is computed using data for all strata except those deliberately excluded because of bad m2 values (in this case Julian week 41)

```
*** Test if pooled Petersen is allowable. [Check if marked fractions
are equal] ***

(Large Sample) Chi-square test statistic  986.1172  has p-value
2.411562e-184
```

```
      time n1-m2  m2 E[n1-m2] E[m2] X2[n1-m2] X2[m2]
 [1,]   10  1414  51   1388.8  76.2       0.5    8.4
 [2,]   11   985 121   1048.4  57.6       3.8   69.9
… (some output omitted)
[35,]   45   471  14    459.8  25.2       0.3    5.0
[36,]   46   111   4    109.0   6.0       0.0    0.7

 Be cautious of using this test in cases of small expected values.
```

The chi-square test for pooling over all strata is presented. Check the X2 values that are large to ensure that these are not a consequence of very small counts.

```
*** Revised data ***
   time   n1  m2    u2 sampfrac new.logitP.cov jump.indicator
1     9    1  NA  9648     0.43              1
2    10 1465  51  9146     1.14              1
… (some output omitted)
32   40 4757 188 35118     1.00              1
33   41 2876  NA 34534     1.00              1
34   42 3989  81 14960     1.00              1
35   43 1755  27  3643     1.00              1
36   44 1527  30  1811     1.00              1
37   45  485  14   679     1.00              1
38   46  115   4   216     0.71              1
```

The revised data are presented. Notice that in Julian week 9 when there were no releases and Julian week 41 where the number of recaptures was odd have been modified.

```
*** Information on priors ***
… (some output omitted)
```

Information on the priors used for the analysis are presented.

```
*** Summary of MCMC results ***

Inference for Bugs model at "model.txt", fit using OpenBUGS,
 3 chains, each with 2e+05 iterations (first 1e+05 discarded), n.thin
= 50; n.sims = 6000 iterations saved
```

Summary of the MCMC results. There were 3 chains run with a 100,000 iteration burn-in, and a 100,000 post-burn-in sample. Every 50th results form the MCMC sampling is saved for a total of 6000 = 3 x 2000 iterations saved.

The output below are the summary statistics from the MCMC chains. For each parameter, the mean and SD over the MCMC iterations is presented along with selected percentiles. In some cases, the output is broken into two chunk with the larger percentiles on the second chunk (see below).

```
                    mean         sd        2.5%         25%      50%
U[1]           148470.795  74594.032   70586.075   97364.000  122537.000
U[2]           247035.349  32894.558  184933.000  224031.000  243808.000
U[3]            23954.843   2057.245   20238.725   22507.750   23834.500
… (some output omitted)
U[37]           23233.763   5565.252   14750.875   19323.750   22554.500
U[38]            6553.574   2765.356    3054.850    4673.000    5973.500
Utot          5302252.274 184268.185 4992346.449 5168412.249 5288462.000
```

Estimates of the run size for each stratum and the total run size.

```
bU[1]              12.165      1.141      10.390      11.352      12.001
bU[2]              10.782      0.432       9.951      10.489      10.776
… (some output omitted)
```

Estimates of the spline coefficients (likely not of interest)

```
beta.logitP[1]     -2.805      0.117      -3.038      -2.883      -2.803
… (some output omitted)
```

Estimates of the coefficient of the covariate for logit(P). In this case, the covariate was a single column of 1's (representing the intercept), so the estimate is presented is for the mean logit(p).

```
deviance          607.029     11.950     585.690     598.813     606.275
… (some output omitted)
```

The deviance used for the DIC and goodness-of-fit computations.

```
logUne[1]          11.425      0.690      10.205      10.939      11.378
logUne[2]          10.848      0.419      10.017      10.558      10.849
logUne[3]          10.410      0.318       9.774      10.199      10.413
… (some output omitted)
```

Estimates of $\log(U_i)$

```
logitP[1]          -2.550      0.461      -3.528      -2.749      -2.458
logitP[2]          -3.249      0.138      -3.508      -3.345      -3.245
logitP[3]          -2.116      0.094      -2.299      -2.180      -2.115
… (some output omitted)
```

Estimates of logit(p) for each stratum.

```
m2[1]               0.081      0.272       0.000       0.000       0.000
m2[33]            365.069     65.723     263.000     319.750     349.000
… (some output omitted)
```

If there are bad data values, estimates of how many marks should have been recaptured given the number of releases.

```
                     75%       97.5%  Rhat n.eff
U[1]          160256.000  337371.948 1.509     8

For each parameter, n.eff is a crude measure of effective sample size,
```

```
and Rhat is the potential scale reduction factor (at convergence,
Rhat=1).
```

The second part of the summary output with higher percentiles and the Brook-Rubin-Gelman statistic for assessing convergence. The *Rhat* values should be close to 1. The *n.eff* is the effective number of independent MCMC iterations for this parameter. These should be close to the number of saved values seen earlier. The *n.eff* will be quite small if the successive MCMC steps are highly correlated. In this case, the estimate of the run size at time 1 is not well determined, but plays a minor role in the overall estimate of population size.

```
DIC info (using the rule, pD = Dbar-Dhat)
pD = 66.6 and DIC = 673.6
DIC is an estimate of expected predictive error (lower deviance is
better).
```

The DIC is presented as measure of model fit along with the effective number of parameters.

```
*** Summary of Unmarked Population Size ***
   mean      sd    2.5%     25%     50%     75%   97.5%    Rhat   n.eff
5302252  184268 4992346 5168412 5288462 5426633 5684909       1      50
```

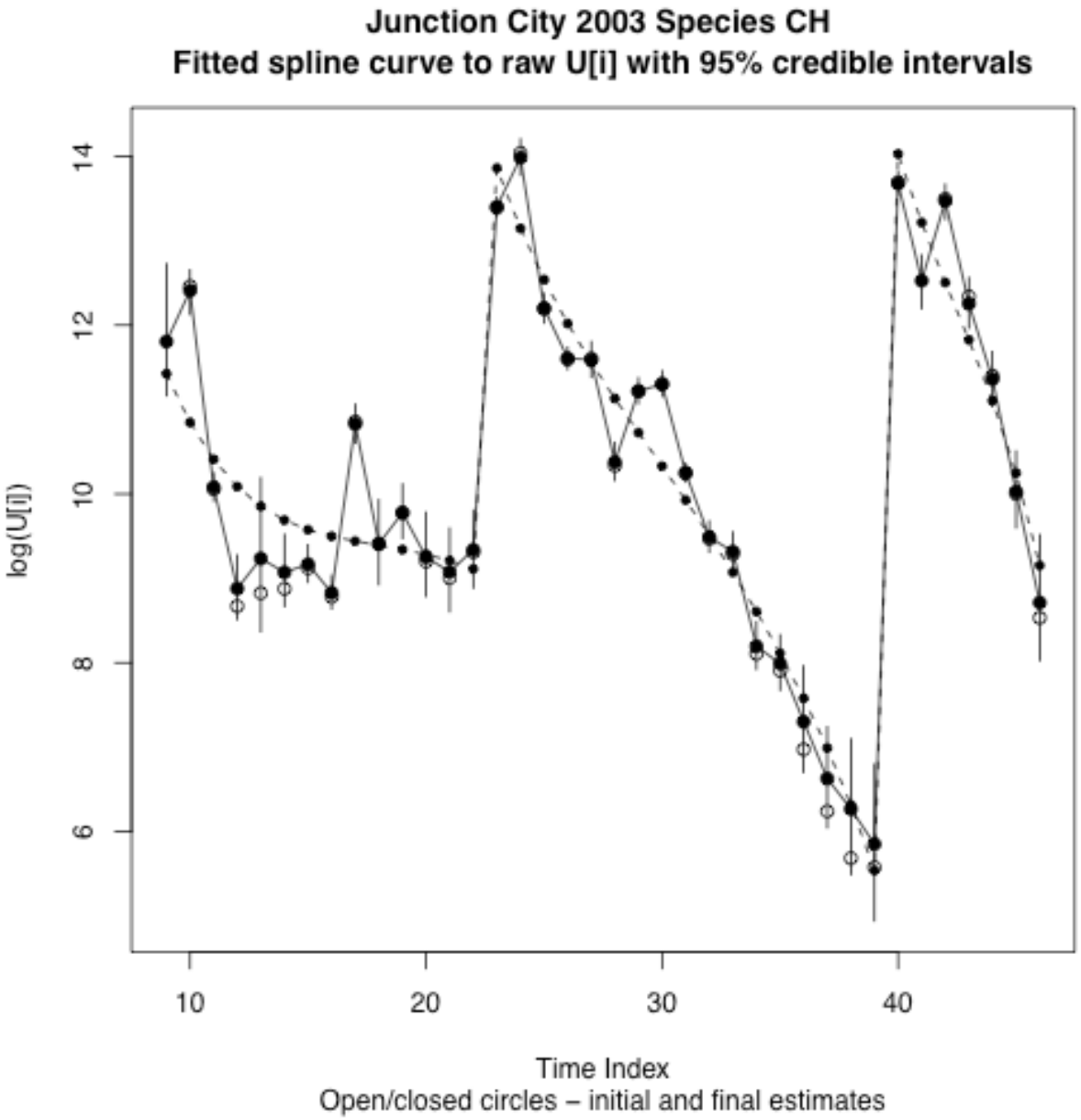
Estimate of the total run size.

```
*** Summary of Quantiles of Run Timing ***
    This is based on the sample weeks provided and the U[i] values
       0%   10%   20%   30%   40%   50%   60%   70%   80%   90% 100%
Mean    9 19.37 23.73 24.29 24.74 26.29 38.59 40.69 41.89 42.72   47
Sd      0  4.00  0.13  0.07  0.07  0.70  3.16  0.09  0.21  0.06    0
```

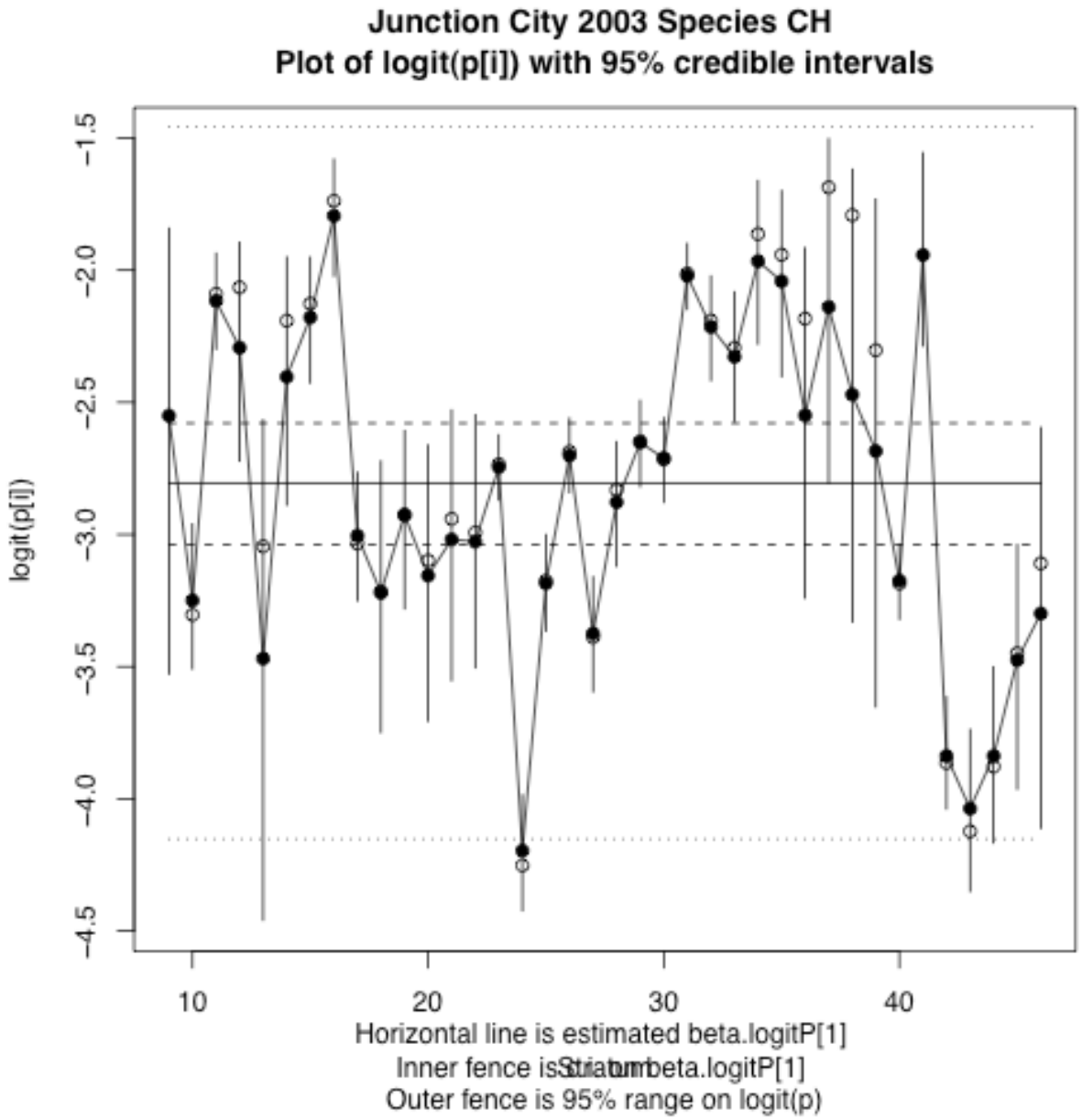
Estimates of the run timing.

```
*** end of fit ***  Mon Jul 27 11:23:17 2009
```

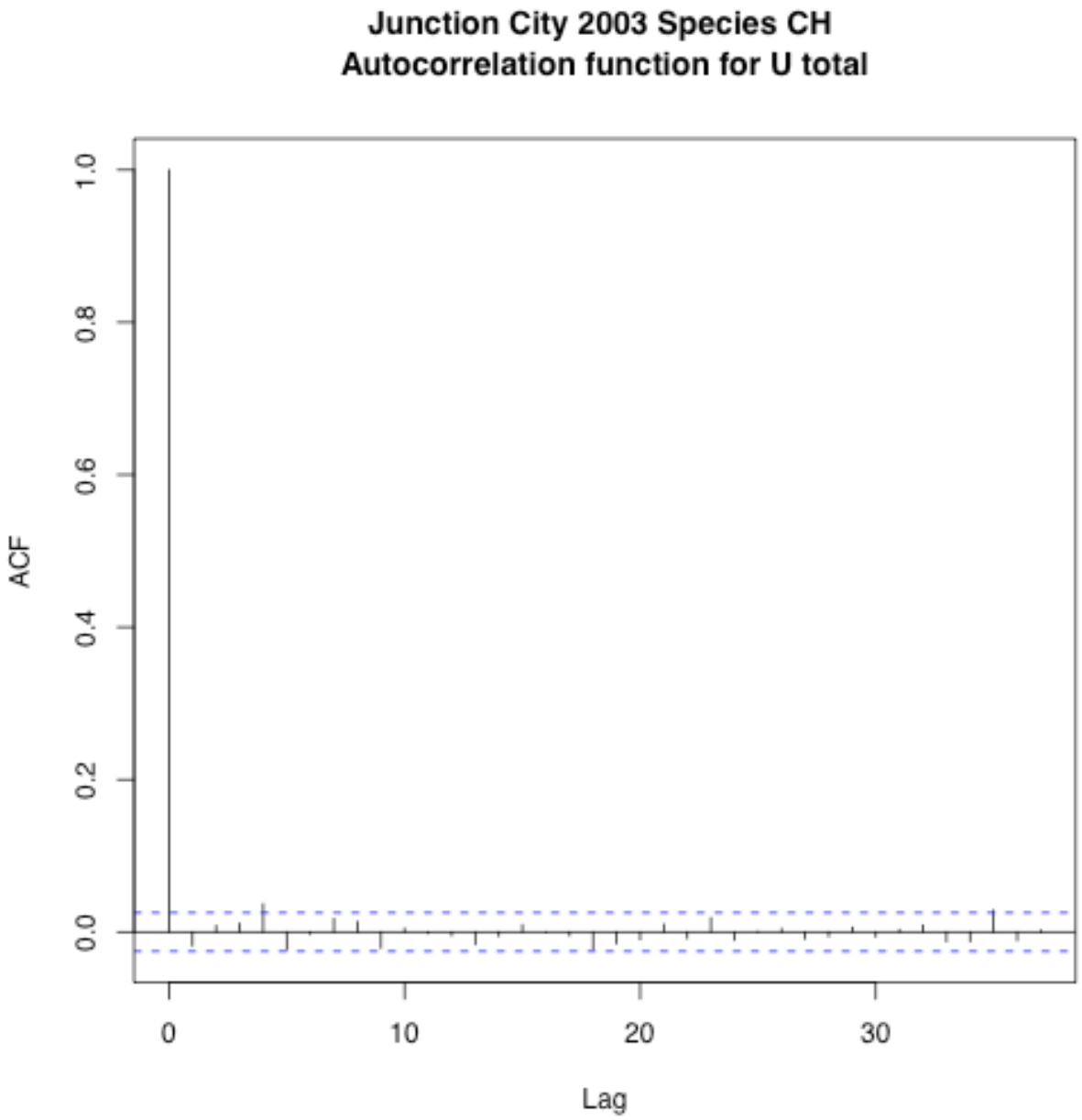
The *JC.2003.CH.TSPDE-logU.pdf* (below) shows the fitted spline curve (dashed lines) and the estimated run sizes (solid line(). Note the estimates of run size at Julian week 9 and 39 have much wider credible intervals because of the poor data. The spline was allowed to jump in two locations as noted earlier. This data set is quite rich so provides good estimates for most Julian weeks.



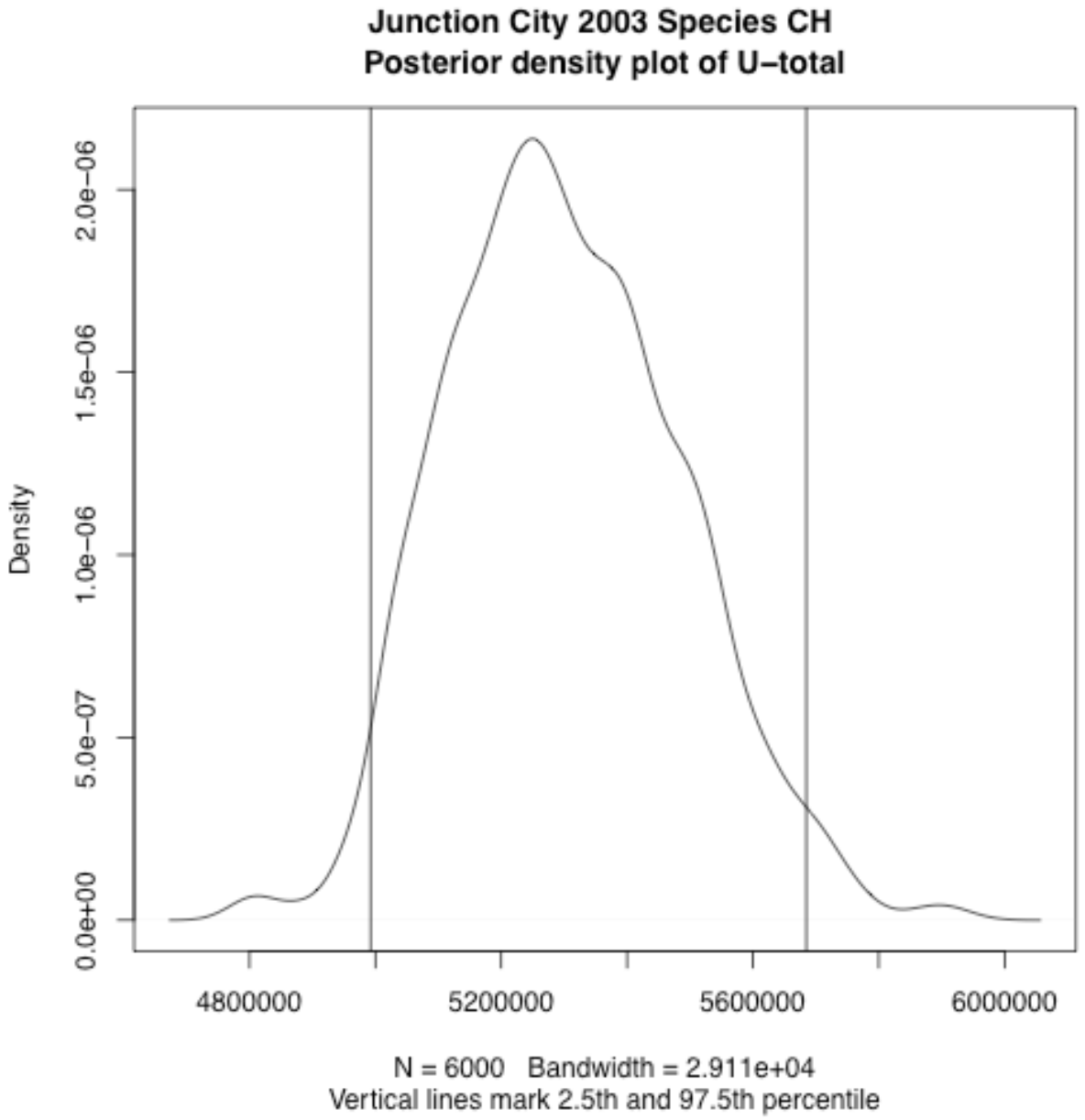
The *JC.2003.CH.TSPDE-logitP.pdf* (below) shows the estimated logit(P) over time. The solid horizontal line is the estimate mean logit(P) over the entire experiment and the inner dashed lines are the 95% credible interval for the mean logit(P). The outer dashed lines are 95% intervals for the individual values of p over the strata. Notice that the credible intervals in Julian weeks 9 and 39 are much wider than the other weeks because of the poor data for these weeks. The open circles are the "raw" estimates of logit(P) based on the actual mark-recapture data.



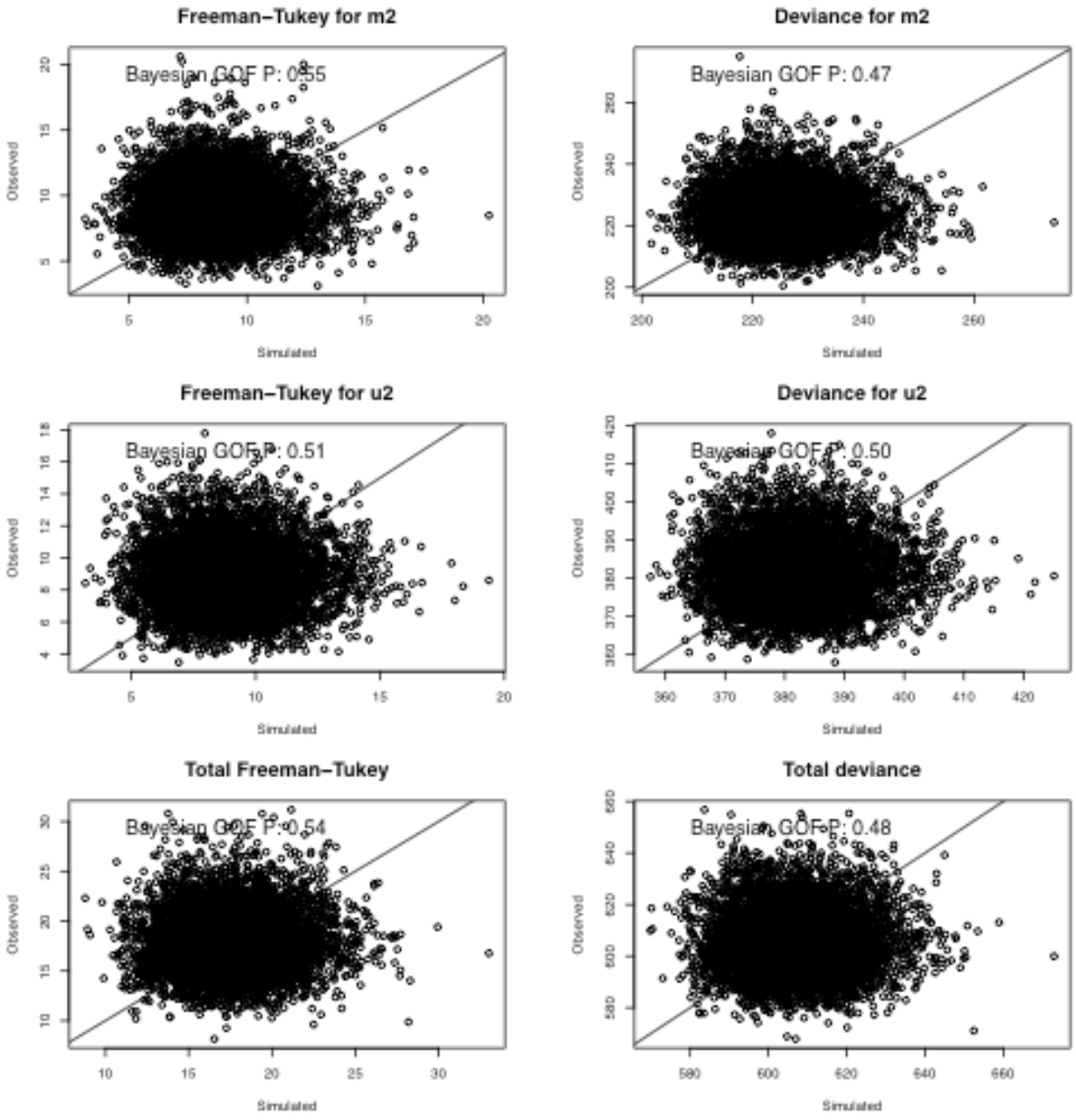
The *JC-2003-CH-TSPDE-Utot-acf.pdf* plot (below) plots the autocorrelation of the estimates of the total runsize over successive (saved) iterations of the MCMC chain. The acf at lag 0 must (by definition) be 1 (the correlation of each value with itself). The autocorrelation at lag *i* is the correlation of each observation with the same observation *I* iterations previously. With a thinning rate of 50, the ACF plots shows little evidence of autocorrelaton over time. These should be close to 0. The dashed lines on the plot indicate the approximate range of sample autocorrelations that could be expected if the true autocorrelation were 0. Most autocorrelations should fall between these limits. If the MCMC samples exhibit high autocorrelation, this is not "bad", but the results should be interpreted with care.



The *JC-2003-CH-TSPDE-Utot-posterior.pdf* plot (below) shows the posterior distribution of the total run size along with the 95% credible intervals. The plot should look smooth with no bimodality etc. Minor bumps in the density plot are OK and are artifacts of only saving 6000 iterations from the MCMC process. In this case, the posterior distribution shows a slight right skewness, but is fairly symmetric.



The *JC-2003-CH-TSPDE-GOF.pdf* plot (below) represents the goodness-of-fit assessment for this model. There are two (related) measures of goodness-of-fit based on the Freeman-Tukey (Freeman and Tukey 1950; Read 1993) and the Deviance statistics (left and right columns) and goodness of fit is assessed against the number of marked fish recaptured (top row), the number unmarked fish captured (middle row) or the combination of the two. A good fitting model should be centered around the line X=Y (drawn on the plots) and the Bayesian p-values should be close to .50.



The remaining files generated by this example would not ordinarily be of interest and are provided for more advanced analyses.

The output from the R-wrapper also leaves an object in the R-workspace called *JC.2003.CH.TSPDE.* This is a list, with much of the information from the analysis collected into one container.

The elements of the list are:

```
> names(JC.2003.CH.TSPDE)
 [1] "n.chains"        "n.iter"          "n.burnin"        "n.thin"          "n.keep"
 [6] "n.sims"          "sims.array"      "sims.list"       "sims.matrix"     "summary"
[11] "mean"            "sd"              "median"          "root.short"
"long.short"
[16] "dimension.short" "indexes.short"   "last.values"     "program"
"model.file"
```

```
[21] "isDIC"          "DICbyR"          "pD"              "DIC"             "data"
```

The more interesting elements in the list are:

- `JC.2003.CH.TSPDE$summary` which contains the summary of the MCMC results (the long table printed in the *-results.txt file.
- `JC.2003.CH.TSPDE$sims.array` which contains the individual values from each iteration of the MCMC chain. This can be used to derive new estimates of parameters not defined in the analysis, .e.g. the 53rd percentile of run timing.
- `JC.2003.CH.TSPDE$data` which contains the raw data used in the fit

These can be used to obtain customized estimates – please contact C. Schwarz for details.

C.2 Fitting the spline model to separate wild and hatchery YOY for Chinook.

The steps in running the R-wrapper to estimate the wild and hatchery YOY for Chinook are similar to those for estimating the total out-migration population. The following are the key differences. In many cases, file names now include the character WH to indicate the wild vs hatchery separation.

Steps 1..4 are the same as in C.1

5. Create a directory to hold the R-wrappers and data. For example, create directory called *TrinityData on your desktop.*

Download a sample R-wrapper from
http://www.stat.sfu.ca/~cschwarz/Consulting/Trinty/Phase2.
Several examples are available in the *TrinityWrappersAndData* subdirectory.

For example, download the *JC_2003_CH_.csv file* (containing the raw data for the Junction City 2003 Chinook example of this report), and the
JC_2003_CH_TSPDE_WH.r
file containing the R-wrapper to read in the raw data and to call the program. Place this in the same directory as your data.

Use a text-editor to open the *JC_2003_CH_TSPDE_WH.r* file and ensure that the line near the top of the file points to the location for the spline-fit program from (5).. For example, the current line:

```
source(paste(dirname(getwd()),
      "/TrinityCode/TimeStratPetersenDiagErrorWHChinook.r", sep=""))
```

will look for the code by looking at the parent of the file location for the R-wrapper and then looking for the directory *TrinityCode* to find the program.

6. The TSPDE_WH_Chinook program has several inputs that are compulsory and several that are options for advanced users. The compulsory arguments are similar to those in C.1.

Argument (case important)	Description
title	Character string that serves as the title for the analysis. For example title <- "JC 2002 Chinook Wild vs Hatchery"
prefix	Character string used as the prefix for created output files. For example

	prefix <- "JC-2002-CH-TSPDE-WH" will create files of the form "JC-2002-CH-TSPDE-WH-xxxx". Ensure that the prefix will result in valid file names.
time	A numeric vector of stratum numbers (e.g. Julian weeks). This is usually read in with the data (see below). Values do not have to start at 1 (i.e. the first Julian week can be 10), nor does it have to be contiguous. Missing stratum numbers are assumed to have $n_i = m_{ii} = u_i = 0$. The program will use a spline to interpolate population numbers for these missing strata.
n1, m2, u2.A u2.N	Numeric vectors of the number of fish marked and released (n1), the number recaptured in the stratum of release (m2), the number of adipose-clipped fish capture (u2.A), and the number of unclipped fish captured (u2.N). The vectors should be same length as the *time* vector.
sampfrac	What fraction of the week was sampled. For example, if the traps were running for 6 days, the corresponding element of the vector would be 6/7=.8514.
clipfrac	The fraction of the Hatchery fish that are adipose-fin clipped. In these studies this is usually set to .25;
hatch.after	The spline model uses a single spline for both the wild and hatchery populations. The hatchery populations usually arrive at the screw-traps in mid-study and so the number of hatchery fish is 0 before they arrive. This argument gives the Julian week AFTER which the hatchery fish arrive – this can usually be determined by looking directly at the raw data.
bad.m2 bad.u2.A bad.u2.N	A vector of stratum numbers (possibly null) where the number of recaptures or captured fish is suspect. For example, in the 2003 Chinook data, the number of recaptures in Julian week 41 was suspect. The program will set the m2 values for these strata to missing and will interpolate as needed. For example *bad.m2 <- NULL* implies that all the m2 values should be used. The code *bad.m2 <- c(14,33)* implies that the m2 values in strata numbers 14 and 33 will be ignored.
logitP.cov	(Optional). This is the matrix of covariates to model the logit(p) values using linear regression. This matrix would normally include a column of 1's for the intercept. If omitted, the TSPDE will fit a mean model for the logitPs. The matrix should have as many rows as the number of strata. CAUTION. If the *time* vector has missing entries, this may cause unexpected results!
Prior parameters	(Optional) There are many parameters for the default priors for the spline model. Normally the default values can be used. Contact cschwarz@stat.sfu.ca for further information on modifying these

	parameters.
run.prob	(Optional). A numeric vector specifying the quantiles for which run timings should be computed. The default value is : run.prob=seq(0,1,.1) which corresponds to the 0, .1, .2, .3, …, 1.0 quantile (i.e. the deciles).
debug	(Optional, default=FALSE). If debug=TRUE, then a shortened call to the spline program with a limited number of iterations will be done. This is useful when running the program for the first few times to ensure that everything is working properly.
openbugs	(Optional, default=TRUE) There are two related programs to fit MCMC models. The OpenBugs program provides more control to the programmer and is the default method. If openbugs=FALSE is set, then the WinBugs program will be called to fit the model. The WinBugs program also provides a better debugging environment if things go wrong. Please contact cschwarz@stat.sfu.ca for more details.
WINBugs.directory	The directory where the WINBugs program is stored. Defaults to C:Program Files/WinBugs14. This is the standard installation location.
OPENBugs.directory	The directory where the OPENBugs program is stored. Defaults to C:Program Files/OPENBugs This is the standard installation location.
InitialSeed	(Optional, default= random integer between 1 and 1 billion (inclusive)). The initial seed used for the random number generators in OpenBugs.

The easiest way to bring the data into R in the proper format is to create a spreadsheet. The first row of the spreadsheet should have the labels *time, n1, m2, u2.A, u2.N, sampfrac.* [The case of the labels is important.] If covariates are to be used (such as log(flow)), enter these in subsequent columns labeled *covar1, covar2,* etc. The remaining rows of the spreadsheet should have the data values as numeric values.

Save the spreadsheet as a *.csv file in your data directory.

The function *read.csv(file="your filename", header=TRUE)* can be used to read the data into a data frame. The usual R commands can be used to extract the various data vectors, to set the *title, prefix, hatch.after, bad.m2, bad.u2.A,* and *bad.u2.N* values. Here is a sample input file with annotations in **bold**,:

The *.csv file is JC-2003-CH.csv and the header names for the data columns were *TotalMarks*, *TotalRecaps*, *CH.YOY.AD.0, CH.YOY.NC.0,* and *DaysOperating*.

```
Site       <- "Junction City"
SiteShort  <- "JC"
Year       <- 2003
FishType   <- 'CH'

hatch.after <- c(22)  # Julian week after which hat fish arrive
input_file <- "JC-2003-CH.csv"

bad.m2     <- c()     # list Julian weeks with bad m2 values
bad.u2.A   <- c()     # list Julian weeks with bad u2.A value
bad.u2.N   <- c()     # list Julian weeks with bad u2.N values
```

```
clip.frac.H <- .25    # what fraction of the hatchery fish are adipose
fin clipped?


# read in the data file for all the years/sites/species

Fish <- read.csv(input_file, header=TRUE)  # reads in file
Fish[1:5,]  # list the first few records

# Now to extract the subset of data, do any fancy adjustments, and fit
the data.

#create the prefix and title.
prefix <- paste(SiteShort,"-",Year,"-",FishType,
      "-TSPDE-WH",sep="")  # indicate files to save information
title <- paste(Site," ",Year," Species ",FishType,
      " Wild vs Hatchery", sep="")



# extract the data
n1 <- Fish$TotalMarks
m2 <- Fish$TotalRecaps

u2.A <- Fish$CH.YOY.AD.0 + Fish$CH.1..AD.0  # YOY Ad-clipped fish. See
previous comment about 1+ fish recorded in earlier weeks

u2.N <- Fish$CH.YOY.NC.0 + Fish$CH.1..NC.0  # YOY (Julian weeks 9-> 39)
and 1+ (Julian weeks 40 onwards) Non-ad-clipped fish


sampfrac <- Fish$DaysOperating/7
jweek <- Fish$SampleWeek + 8  # convert from sample week to Julian week

YoY.select <- jweek < 40         # only Julian weeks 9-> 39 are YOY
fish. Julian week 40 onwards is 1+ hatchery fish only
```

The TSPDE is then called using the following code segment. The arguments can be in any order.

```
library("BTSPAS") # make the functions available

resi;ts <- TimeStratPetersenDiagErrorWHChinook_fit(
                  title=title,
                  prefix=prefix,
                  time=jweek[YoY.select],
                  n1=n1     [YoY.select],
                  m2=m2     [YoY.select],
                  u2.A=u2.A [YoY.select],
                  u2.N=u2.N [YoY.select],
                  clip.frac.H=clip.frac.H,
                  sampfrac=sampfrac[YoY.select],
                  hatch.after=hatch.after,
                  bad.m2=bad.m2,
```

```
bad.u2.A=bad.u2.A,
bad.u2.N=bad.u2.N,
debug=TRUE
)
```

The [YoY.select] on the data arguments restricted the data to the Julian weeks consisting only of YOY fish (typically Julian week 39 or earlier) and was set using the

```
YoY.select <- jweek < 40
```

command.

The TSPDE program will run in debug mode (a small number of MCMC iterations) and will return the results in an MCMC-object called *results*. This can be used for further analyses – contact cschwarz@stat.sfu.ca for details.

The sample code can be run against the sample data using the *source* facility of R. Under *File->Source* point R to the R-wrapper *JC_2003_CH_TSPDE_WH.r*

The main part of the output is a series of output files and graphical files created in the directory with the R-wrapper with the prefix passed to it. For example, the above code will generate the following files in the data directory.

File	Contents.
JC-2003-CH-TSPDE-WH-results.txt	Text listing with the raw data, various Pooled-Petersen, Stratified-Petersen estimators, summary statistics on the MCMC posterior distributions (e.g. mean and SD on the total population estimate, the spline coefficients, etc), run timing quantiles etc.
JC-2003-CH-TSPDE-WH-initialU.pdf JC-2003-CH-TSPDE-WH-logU.pdf	Plots of the initial and final population estimates over time.
JC-2003-CH-TSPDE-WH-logitP.pdf	Plot of the fitted logit(p)'s over time
JC-2003-CH-TSPDE-WH-UtotH-acf.pdf JC-2003-CH-TSPDE-WH-UtotW-acf.pdf JC-2003-CH-TSPDE-WH-UtotH-posterior.pdf JC-2003-CH-TSPDE-WH-UtotW-posterior.pdf	Plot of the posterior distribution of U-total (UtotW=Wild and UtotH=Hatchery) and the autocorrelation plot of successive estimates of U-total from the MCMC chain. .
JC-2003-CH-TSPDE-WH-GOF.pdf	Plots of the posterior predictive distributions used for goodness of fit.
JC-2003-CH-TSPDE-WH.data.txt JC-2003-CH-TSPDE-WH.inits1.txt JC-2003-CH-TSPDE-WH.inits2.txt JC-2003-CH-TSPDE-WH.inits3.txt	Intermediate files used when calling OpenBugs/WinBugs from R. Normally not useful to the user. These contain the raw data and the initial values for 3 chains.

JC-2003-CH-TSPDE-WH.CODAchain1.txt JC-2003-CH-TSPDE-WH.CODAchain2.txt JC-2003-CH-TSPDE-WH.CODAchain3.txt JC-2003-CH-TSPDE-WH.CODAindex.txt	Intermediate files called when calling OpenBugs/WinBugs from R. Normally not useful for the user. These contain the results from the MCMC chains are automatically processed and summaries produced in other files.
JC.2003.CH.TSPDE.WH-saved.Rdata	Save results in R data dump with the results form the spline-fit. This can be subsequently loaded with the *load()* command and further processed.

8. Interpretation of sample output. The following is selected portions of the output from the analysis of the JC 2003 Chinook data in the file *JC.2003.CH.TSPDE-WH-results.txt*. The text not in Courier Font is the added explanation:

```
Time Stratified Petersen with Diagonal recaptures, error in smoothed
U, separating wild and hatchery fish -  Sun Sep 06 23:10:56 2009

 Junction City 2002 Species CH Wild vs Hatchery Results

*** Raw data ***
      time   n1  m2 u2.A  u2.N SampFrac logitPcov[1]
 [1,]    9   73   3    0    88     0.43            1
 [2,]   10 1773 150    0  1362     0.86            1
 [3,]   11 3543 207    1  2164     1.00            1
 … … (some output omitted)
[30,]   38   76  23    6    92     1.00            1
[31,]   39    0   0    5   167     1.00            1
[32,]   40   78  11    0    98     1.00            1
```

The sample data are listed as read in. If there are no covariates, the mean logit (capture probability) is estimated which is equivalent to fitting a model for covariates with only the intercept (a column of 1's) on the right portion of the data listing.

```
Hatchery fish are released AFTER strata:  22

Hatchery fish are clipped at a rate of : 0.25
```

The number of hatchery YOY fish is assumed to be 0 until stratum 23 when the hatchery (partially) clipped fish arrive.

```
… Output omitted on pooled Petersen and stratified-Petersen estimator
…

*** Revised data ***
   time   n1  m2 u2.A  u2.N sampfrac    logitP.cov hatch.ind
```

```
1     9   73   3    0   205     0.43              1
2    10 1773 150    0  1589     0.86              1
3    11 3543 207    1  2164     1.00              1
… (some output omitted)
```

The revised data are presented. Notice that in Julian week 9 when there were no releases, the data have been modified.

```
*** Information on priors ***
… (some output omitted)
```

Information on the priors used for the analysis are presented.

```
*** Summary of MCMC results ***

Inference for Bugs model at "model.txt", fit using OpenBUGS,
 3 chains, each with 2e+05 iterations (first 1e+05 discarded), n.thin
= 50; n.sims = 6000 iterations saved
```

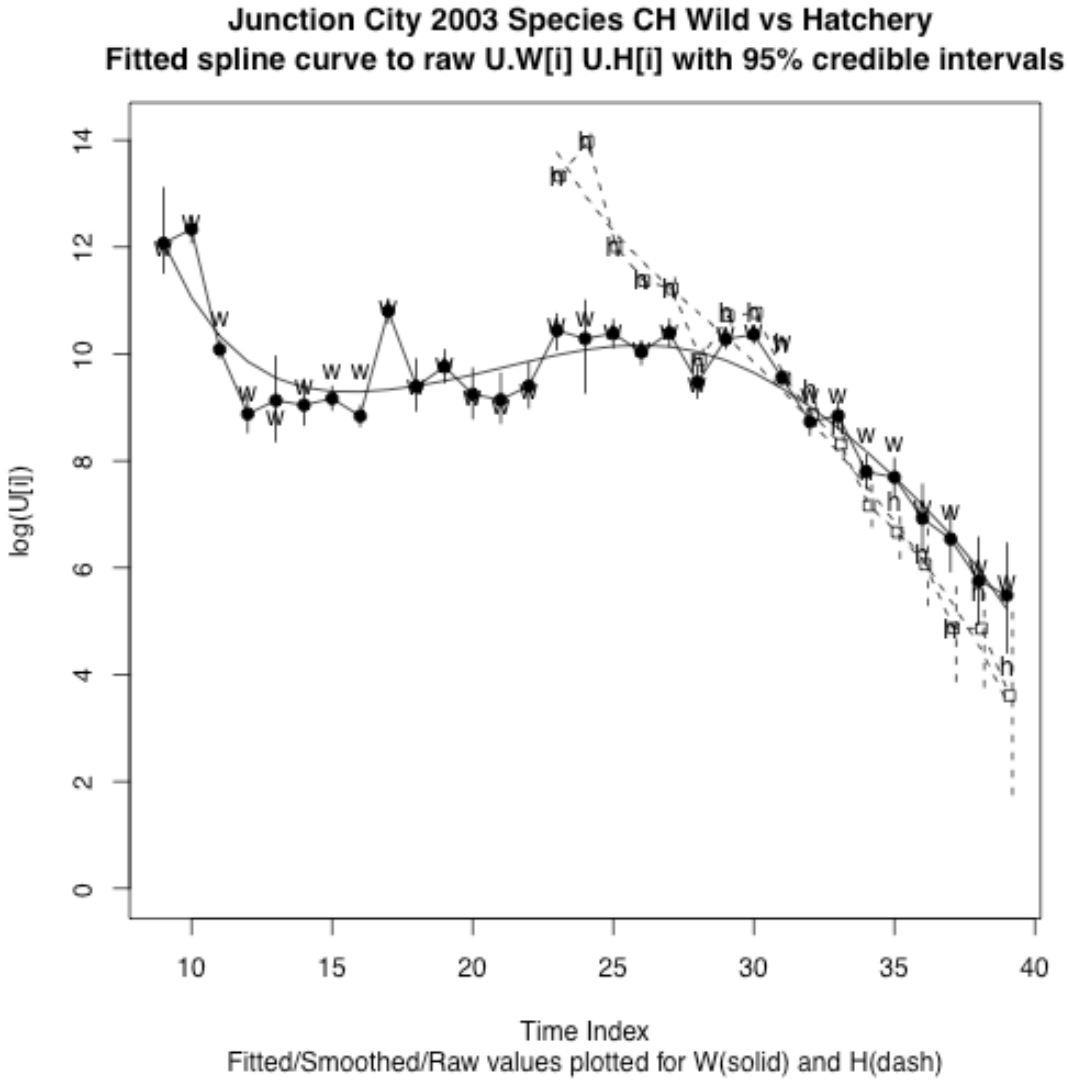
Summary of the MCMC results. There were 3 chains run with a 100,000 iteration burn-in, and a 100,000 post-burn-in sample. Every 50th results form the MCMC sampling is saved for a total of 6000 = 3 x 2000 iterations saved.

The output (not shown) the presents the summary statistics from the MCMC chains. For each parameter, the mean and SD over the MCMC iterations is presented along with selected percentiles. The output is similar to that shown in C.1.

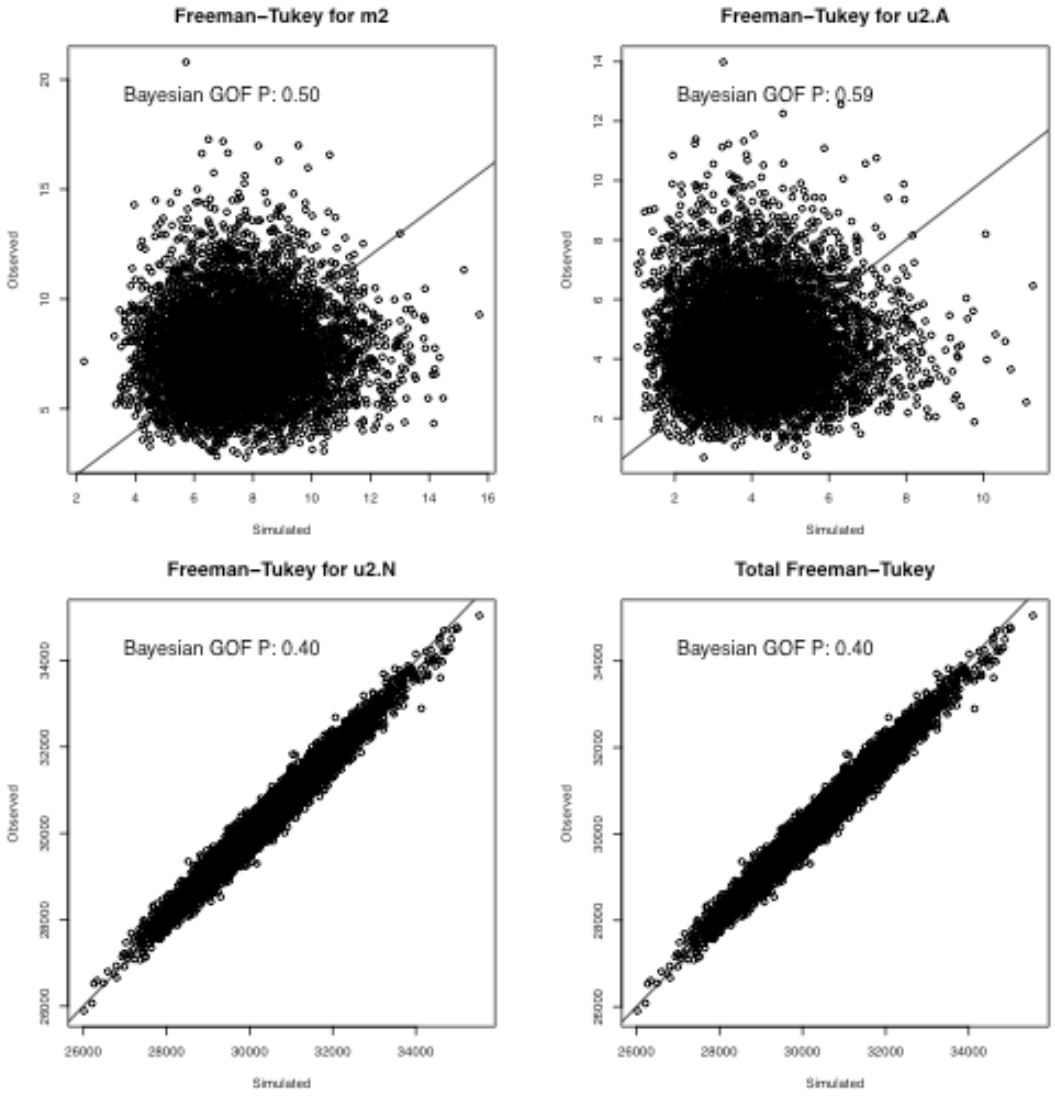
The lines labeled U.H[i] and U.W[i] are the hatchery and wild YOY populations respectively. Estimate of the total hatchery and wild run size are presented along with quantiles of the run timing in much the same way as in C.1.

The plots are similar to those in C.1. The notable changes are as follows.

The *JC.2003.CH.TSPDE-WH-logU.pdf* (below) shows the spline and fitted curve for the wild (solid lines) and hatchery fish (dashed lines). Note the estimates of run size at Julian week 9 has a much wider credible intervals because of the poor data. This data set is quite rich so provides good estimates for most Julian weeks.



The *JC-2003-CH-TSPDE-WH-GOF.pdf* plot (below) represents the goodness-of-fit assessment for this model. The Freeman-Tukey statistic is computed for the u2.A and u2.N data. A good fitting model should be centered around the line X=Y (drawn on the plots) and the Bayesian p-values should be close to .50. The deviance is not easily computed for this model, and so GOF tests based on the deviance are not provided.



The remaining files generated by this example would not ordinarily be of interest and are provided for more advanced analyses.

The output from the R-wrapper also leaves an object in the R-workspace called *JC.2003.CH.TSPDE.* This is a list, with much of the information from the analysis collected into one container. It as the same format as in Section C.1.

C.3 Fitting the spline model to separate wild and hatchery components for steelhead

The steps in running the R-wrapper to estimate the wild and hatchery components for Steelhead are similar to those for estimating the total out-migration population. The following are the key differences. In many cases, file names now include the character WH to indicate the wild vs hatchery separation.

Steps 1..4 are the same as in C.1

5. Create a directory to hold the R-wrappers and data. For example, create directory called *TrinityData on your desktop.*

Download a sample R-wrapper from
http://www.stat.sfu.ca/~cschwarz/Consulting/Trinty/Phase2.
Several examples are available in the *TrinityWrappterAndData* subdirectory.

For example, download the *JC_2003_ST_.csv file* (containing the raw data for the Junction City 2003 Steelhead example of this report), and the
JC_2003_ST_TSPDE_WH.r
file containing the R-wrapper to read in the raw data and to call the program. Place this in the same directory as your data.

Use a text-editor to open the *JC_2003_ST_TSPDE_WH.r* file and ensure that the line near the top of the file points to the location for the spline-fit program from (5).. For example, the current line:

```
source(paste(dirname(getwd()),
    "/TrinityCode/TimeStratPetersenDiagErrorWHSteel.r", sep=""))
```

will look for the code by looking at the parent of the file location for the R-wrapper and then looking for the directory *TrinityCode* to find the program.

6. The TSPDE_WH_Steel program has several inputs that are compulsory and several that are options for advanced users. The compulsory arguments are similar to those in C.1.

Argument (case important)	Description
title	Character string that serves as the title for the analysis. For example title <- "JC 2003 Steelhead Wild vs Hatchery"
prefix	Character string used as the prefix for created output files. For example prefix <- "JC-2003-ST-TSPDE-WH" will create files of the form "JC-2003-ST-TSPDE-WH-xxxx". Ensure that the prefix will result in valid file names.
time	A numeric vector of stratum numbers (e.g. Julian weeks). This is usually read in with the data (see below). Values do not have to start at 1 (i.e. the first Julian week can be 10), nor does it have to be contiguous. Missing stratum numbers are assumed to have $n_i = m_{ii} = u_i = 0$. The program will use a spline to interpolate population numbers for these missing strata.
n1, m2, u2.W.YoY u2.W.1 u2.H.1	Numeric vectors of the number of fish marked and released (n1), the number recaptured in the stratum of release (m2), the number of wild YOY fish (u2.W.YoY), wild age 1+ fish (u2.W.1), and the number of hatchery age 1+ fish (u2.H.1). The vectors should be same length as the *time* vector.
sampfrac	The proportion of the week that was sampled. For example, if the traps ran for 6 days, the corresponding element of this vector would be 6/7=.8514. Same length as the data.
hatch.after	The spline model uses a single spline for both the wild and hatchery populations. The hatchery populations usually arrive at the screw-traps in mid-study and so the number of hatchery fish is 0 before they arrive.

	This argument gives the Julian week AFTER which the hatchery fish arrive – this can usually be determined by looking directly at the raw data.
bad.m2 bad.u2.W.YoY bad.u2.W.1 bad.u2.H.1	A vector of stratum numbers (possibly null) where the number of recaptures or captured fish is suspect. The program will set the data values for these strata to missing and will interpolate as needed. For example *bad.m2 <- NULL* implies that all the m2 values should be used. The code *bad.m2 <- c(14,33)* implies that the m2 values in strata numbers 14 and 33 will be ignored.
logitP.cov	(Optional). This is the matrix of covariates to model the logit(p) values using linear regression. This matrix would normally include a column of 1's for the intercept. If omitted, the TSPDE will fit a mean model for the logitPs. The matrix should have as many rows as the number of strata. CAUTION. If the *time* vector has missing entries, this may cause unexpected results!
Prior parameters	(Optional) There are many parameters for the default priors for the spline model. Normally the default values can be used. Contact cschwarz@stat.sfu.ca for further information on modifying these parameters.
run.prob	(Optional). A numeric vector specifying the quantiles for which run timings should be computed. The default value is : run.prob=seq(0,1,.1) which corresponds to the 0, .1, .2, .3, …, 1.0 quantile (i.e. the deciles).
debug	(Optional, default=FALSE). If debug=TRUE, then a shortened call to the spline program with a limited number of iterations will be done. This is useful when running the program for the first few times to ensure that everything is working properly.
openbugs	(Optional, default=TRUE) There are two related programs to fit MCMC models. The OpenBugs program provides more control to the programmer and is the default method. If openbugs=FALSE is set, then the WinBugs program will be called to fit the model. The WinBugs program also provides a better debugging environment if things go wrong. Please contact cschwarz@stat.sfu.ca for more details.
WINBugs.directory	The directory where the WINBugs program is stored. Defaults to C:Program Files/WinBugs14. This is the standard installation location.
OPENBugs.directory	The directory where the OPENBugs program is stored. Defaults to C:Program Files/OPENBugs This is the standard installation location.
InitialSeed	(Optional, default= random integer between 1 and 1 billion (inclusive)). The initial seed used for the random number generators in OpenBugs.

The easiest way to bring the data into R in the proper format is to create a spreadsheet. The first row of the spreadsheet should have the labels *time, n1, m2, u2.W.YoY, u2.W.1, u2.H.1, sampfrac.* [The case of the labels is important.] If covariates are to be used (such as log(flow)), enter these in subsequent columns labeled *covar1, covar2,* etc. The remaining rows of the spreadsheet should have the data values as numeric values.

Save the spreadsheet as a *.csv file in your data directory.

The function *read.csv(file="your filename", header=TRUE)* can be used to read the data into a data frame. The usual R commands can be used to extract the various data vectors, to set the *title, prefix, hatch.after, bad.m2, bad.u2.W.YoY, bad.u2.W.1,* and *bad.u2.H.1* values. Here is a sample input file with annotations in **bold**,:

The *.csv file is JC-2003-ST.csv and the header names for the data columns were *TotalMarks*, *TotalRecaps*, *CH.YOY.AD.0, CH.YOY.NC.0,* and *DaysOperating*.

```
Site       <- "Junction City"
SiteShort  <- "JC"
Year       <- 2003
FishType   <- 'ST'

hatch.after <- c(11)  # Julian week after which hat fish arrive
input_file <- "JC-2003-ST.csv"

bad.m2     <- c()     # list Julian weeks with bad m2 values
bad.u2.W.YoY <- c()
bad.u2.W.1   <- c()
bad.u2.H.1   <- c()


# read in the data file for all the years/sites/species

Fish <- read.csv(input_file, header=TRUE)  # reads in file
Fish[1:5,]  # list the first few records

# Now to extract the subset of data, do any fancy adjustments, and fit
the data.

#create the prefix and title.
prefix <- paste(SiteShort,"-",Year,"-",FishType,
      "-TSPDE-WH",sep="")  # indicate files to save information
title <- paste(Site," ",Year," Species ",FishType,
      " Wild vs Hatchery", sep="")



# extract the data
n1 <- Fish$TotalMarks
m2 <- Fish$TotalRecaps

u2.W.YoY <- Fish$ST.YOY.NC.0
u2.W.1   <- Fish$ST.1..NC.0 + Fish$ST.2..NC.0
u2.H.1   <- Fish$ST.YOY.AD.0+ Fish$ST.1..AD.0 + Fish$ST.2..AD.0
```

```
sampfrac <- Fish$DaysOperating/7
jweek <- Fish$SampleWeek + 8  # convert from sample week to Julian week
```

The TSPDE is then called using the following code segment. The arguments can be in any order.

```
library("BTSPAS")  # make the functions available

results <- TimeStratPetersenDiagErrorWHSteel_fit(
                  title=title,
                  prefix=prefix,
                  time=jweek,
                  n1=n1,
                  m2=m2,
                  u2.W.YoY=u2.W.YoY,
                  u2.W.1=u2.W.1,
                  u2.H.1=u2.H.1,
                  sampfrac=sampfrac,
                  hatch.after=hatch.after,
                  bad.m2=bad.m2,
                  bad.u2.W.YoY=bad.u2.W.YoY,
                  bad.u2.W.1  =bad.u2.W,1,
                  bad.u2.H.1  =bad.u2.H.1,
                  debug=TRUE
                  )
```

The TSPDE_WH_Steel program will run in debug mode (a small number of MCMC iterations) and will return the results in an MCMC-object called *results*. This can be used for further analyses – contact cschwarz@stat.sfu.ca for details.

The sample code can be run against the sample data using the *source* facility of R. Under *File->Source* point R to the R-wrapper *JC_2003_ST_TSPDE_WH.r*

The main part of the output is a series of output files and graphical files created in the directory with the R-wrapper with the prefix passed to it. For example, the above code will generate the following files in the data directory.

File	Contents.
JC-2003-ST-TSPDE-WH-results.txt	Text listing with the raw data, various Pooled-Petersen, Stratified-Petersen estimators, summary statistics on the MCMC posterior distributions (e.g. mean and SD on the total population estimate, the spline coefficients, etc), run timing quantiles etc.
JC-2003-ST-TSPDE-WH-initialU.pdf JC-2003-ST-TSPDE-WH-logU.pdf	Plots of the initial and final population estimates over time.
JC-2003-ST-TSPDE-WH-logitP.pdf	Plot of the fitted logit(p)'s over time
JC-2003-ST-TSPDE-WH-UtotH.1-acf.pdf	Plot of the posterior distribution of U-total

JC-2003-ST-TSPDE-WH-UtotW.1-acf.pdf JC-2003-ST-TSPDE-WH-UtotW.YoY-acf.pdf JC-2003-ST-TSPDE-WH-UtotH.1-posterior.pdf JC-2003-ST-TSPDE-WH-UtotW.1-posterior.pdf JC-2003-ST-TSPDE-WH-UtotW.YoY-posterior.pdf	(UtotW.YoY=Wild YoY, UtotW.1=Wild age 1+, and UtotH.1=Hatchery age 1+) and the autocorrelation plot of successive estimates of U-total from the MCMC chain. .
JC-2003-ST-TSPDE-WH-GOF.pdf	Plots of the posterior predictive distributions used for goodness of fit.
JC-2003-ST-TSPDE-WH.data.txt JC-2003-ST-TSPDE-WH.inits1.txt JC-2003-ST-TSPDE-WH.inits2.txt JC-2003-ST-TSPDE-WH.inits3.txt	Intermediate files used when calling OpenBugs/WinBugs from R. Normally not useful to the user. These contain the raw data and the initial values for 3 chains.
JC-2003-ST-TSPDE-WH.CODAchain1.txt JC-2003-ST-TSPDE-WH.CODAchain2.txt JC-2003-ST-TSPDE-WH.CODAchain3.txt JC-2003-ST-TSPDE-WH.CODAindex.txt	Intermediate files called when calling OpenBugs/WinBugs from R. Normally not useful for the user. These contain the results from the MCMC chains are automatically processed and summaries produced in other files.
JC.2003.ST.TSPDE.WH-saved.Rdata	Save results in R data dump with the results form the spline-fit. This can be subsequently loaded with the *load()* command and further processed.

8. Interpretation of sample output. The following is selected portions of the output from the analysis of the JC 2003 Steelhead data in the file *JC.2003.ST.TSPDE-WH-results.txt*. The text not in Courier Font is the added explanation:

```
Time Stratified Petersen with Diagonal recaptures, error in smoothed
U, separating wild and hatchery fish, STEELHEAD ONLY -  Mon Sep 07
17:57:13 2009

 Junction City 2003 Species ST Results

*** Raw data ***
      time   n1 m2 u2.W.YoY u2.W.1 u2.H.1 SampFrac logitPcov[1]
 [1,]    9    0  0        0     58      0     0.43            1
 [2,]   10    0  0        0    357      2     1.14            1
 [3,]   11    0  0        0    720      0     0.86            1
 … … (some output omitted)
[36,]   44    0  0       19      7      0     1.00            1
[37,]   45    0  0       46      4      0     1.00            1
[38,]   46    0  0      229      7      0     0.71            1
```

The sample data are listed as read in. If there are no covariates, the mean logit (capture probability) is estimated which is equivalent to fitting a model for covariates with only the intercept (a column of 1's) on the right portion of the data listing.

```
Hatchery fish are released AFTER strata:  11

The following strata had m2       set to missing:   NONE
The following strata had u2.W.YoY set to missing:   NONE
The following strata had u2.W.1   set to missing:   NONE
The following strata had u2.H.1   set to missing:   NONE
```

The number of hatchery YOY fish is assumed to be 0 until stratum 11 when the hatchery fully clipped fish arrive. There are no data points which are appear to be anomalous.

```
… Output omitted on pooled Petersen and stratified-Petersen estimator
…

*** Revised data ***
   time   n1 m2 u2.W.YoY u2.W.1 u2.H.1 sampfrac new.logitP.cov
hatch.indicator
1     9    1 NA        0    135      0     0.43             1
2    10    1 NA        0    312      2     1.14             1
3    11    1 NA        0    840      0     0.86             1
… (some output omitted)


*** Information on priors ***
… (some output omitted)
```

Information on the priors used for the analysis are presented.

```
*** Summary of MCMC results ***

Inference for Bugs model at "model.txt", fit using OpenBUGS,
 3 chains, each with 2e+05 iterations (first 1e+05 discarded), n.thin
= 50; n.sims = 6000 iterations saved
```

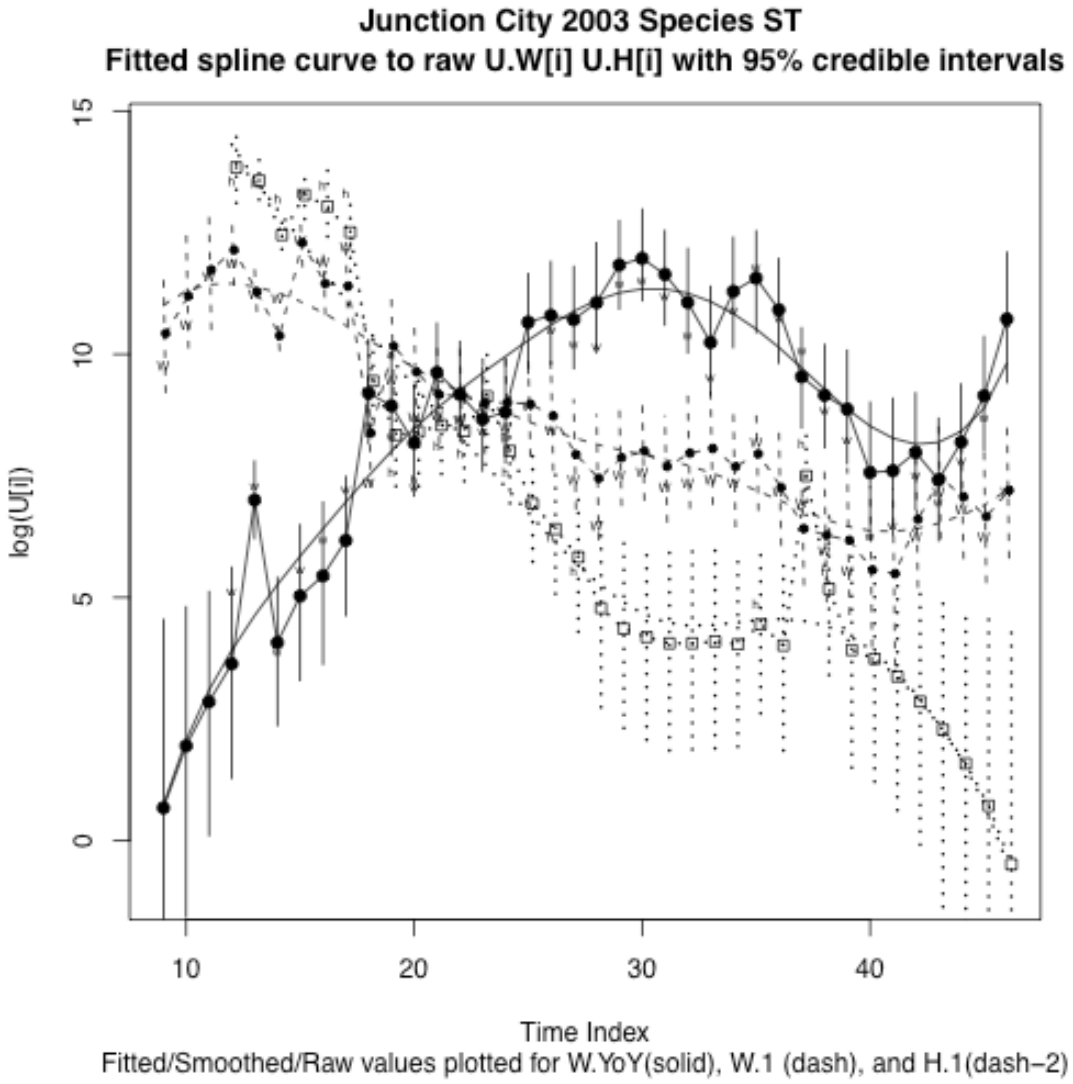
Summary of the MCMC results. There were 3 chains run with a 100,000 iteration burn-in, and a 100,000 post-burn-in sample. Every 50th results form the MCMC sampling is saved for a total of 6000 = 3 x 2000 iterations saved.

The output (not shown) the presents the summary statistics from the MCMC chains. For each parameter, the mean and SD over the MCMC iterations is presented along with selected percentiles. The output is similar to that shown in C.1.

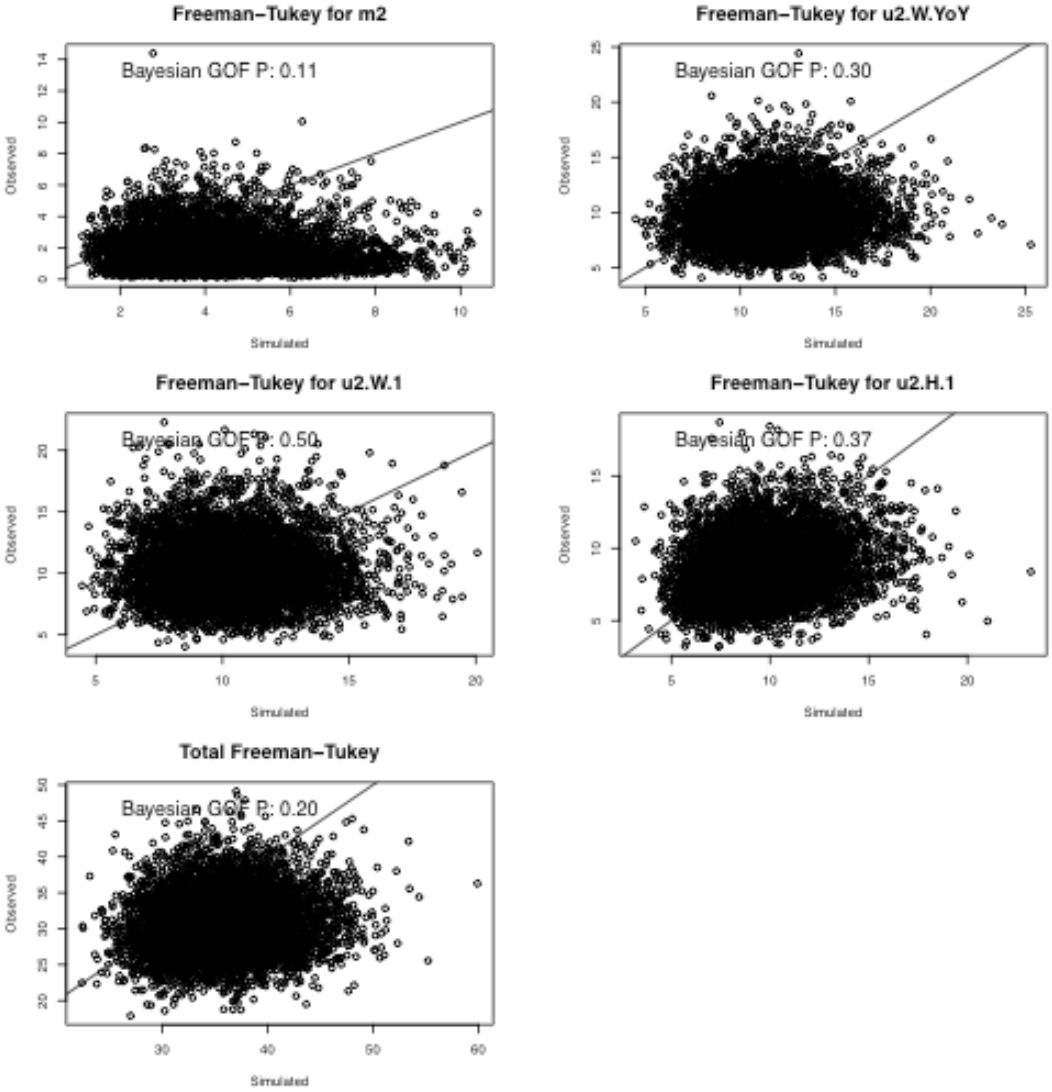
The lines labeled U.H.1[i], U.W.1, and U.W.YoY[i] are the hatchery aged 1+, wild age 1+, and wild YOY populations respectively. Estimate of the total run size over the hatchery and wild components are presented along with quantiles of the run timing in much the same way as in C.1.

The plots are similar to those in C.1. The notable changes are as follows.

The *JC.2003.ST.TSPDE-WH-logU.pdf* (below) shows the spline and fitted curve for the wild YOY (solid lines) wild age 1+ (dashed lines) and hatchery age 1+ fish (dotted lines). Note the estimates of run size at are very poor because of the limited number of strata in which marking occurred.



The *JC-2003-ST-TSPDE-WH-GOF.pdf* plot (below) represents the goodness-of-fit assessment for this model. The Freeman-Tukey statistic is computed for the m2, u2.W.YoY, u2.W.1, u2.H.1 and the combined data. A good fitting model should be centered around the line X=Y (drawn on the plots) and the Bayesian p-values should be close to .50.



The remaining files generated by this example would not ordinarily be of interest and are provided for more advanced analyses.

The output from the R-wrapper also leaves an object in the R-workspace called *JC.2003.ST.TSPDE.* This is a list, with much of the information from the analysis collected into one container. It as the same format as in Section C.1.

Appendix D

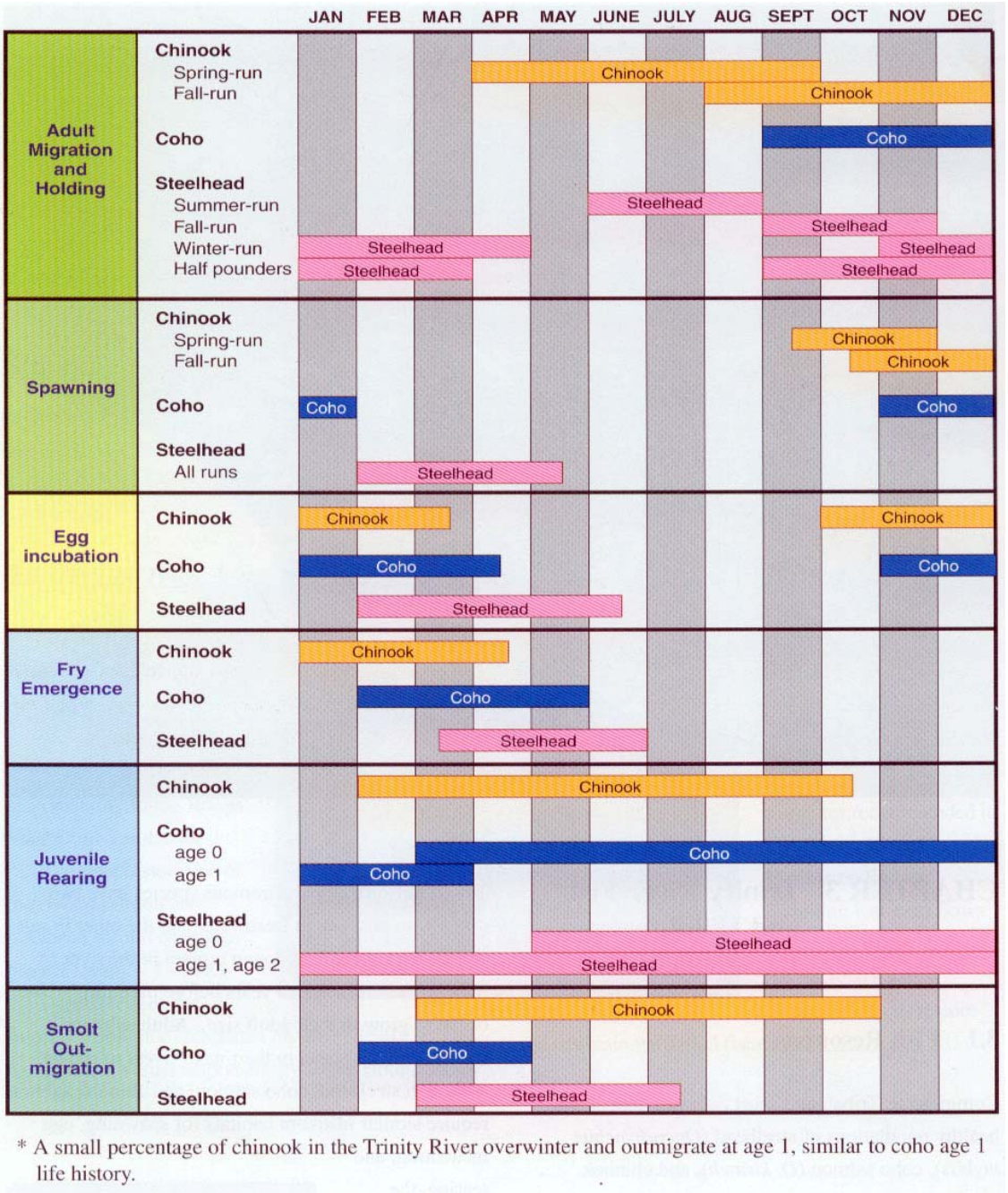


Figure 1. Diagram of the timing and duration of various life-history events for Chinook salmon, coho salmon, and steelhead in the Trinity River. (Reproduced with permission from USFWS and HVT 1999)

Appendix E

The early Junction City data did not use the same naming conventions as the more recent data. This table describes our interpretation of the old data.

1997

Species	Age?	Marked	Assumptions
Chinook salmon	0	True, False, & blanks	?
Coho	?	a few false, mostly blanks	assume blanks mean true (i.e. hatchery?)
STHD 0+	0	mostly false, quite a few blanks, no true	assume blanks mean true (i.e. hatchery?)
STHD Parr	1	mostly false, quite a few blanks, no true. NOTE: the blanks are generally bigger than the FALSE	assume blanks mean true (i.e. hatchery?)
STHD Smolt	2	mostly FALSE, a few blanks, no TRUE. Blanks were bigger, and note that a few of the FALSE were quite small.	assume blanks mean true (i.e. hatchery?)

1998

Species	Age?	Marked	Assumptions
Chinook salmon	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
Coho	?	TRUE, FALSE, no blanks	True = hatchery, False = natural
Coho 1+	1	TRUE, FALSE, no blanks	True = hatchery, False = natural
Coho YOY	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
RBT < 70 mm	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
RBT 70 to 130 mm	1	TRUE, FALSE, no blanks	True = hatchery, False = natural
RBT > 130 mm	2	TRUE, FALSE, no blanks	True = hatchery, False = natural

1999

Species	Age?	Marked	Assumptions
Chinook salmon	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
Coho 1+	1	TRUE, FALSE, no blanks	True = hatchery, False = natural
Coho YOY	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
RBT < 70 mm	0	all FALSE	False = natural
RBT 70 to 130 mm	1	TRUE, FALSE, no blanks	True = hatchery, False = natural
RBT > 130 mm	2	TRUE, FALSE, no blanks	True = hatchery, False = natural

2000

Species	Age?	Marked	Assumptions
Chinook salmon	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
Coho 1+	1	only 3 fish, all FALSE	False = natural
Coho YOY	0	mostly FALSE, 2 TRUE, no blanks	True = hatchery, False = natural
RBT < 70 mm	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
RBT 70 to 130 mm	1	mostly FALSE, 2 TRUE, no blanks	True = hatchery, False = natural
RBT > 130 mm	2	TRUE, FALSE, no blanks	True = hatchery, False = natural

2001

Species	Age?	Marked	Assumptions
Chinook salmon	0	TRUE, FALSE, no blanks	True = hatchery, False = natural
Coho	0 & 1, Should I try and separate these? If so what cutoffs to use?	TRUE, FALSE, no blanks	True = hatchery, False = natural
RBT	0, 1, & 2. Should I try and separate these? I could use the 70 & 130 cutoffs like above?	TRUE, FALSE, no blanks	True = hatchery, False = natural