Independent Scientific Advisory Board
for the Northwest Power and Conservation Council,
Columbia River Basin Indian Tribes,
and National Marine Fisheries Service
851 SW 6th Avenue, Suite 1100
Portland, Oregon 97204

Review of the Comparative Survival Study Draft 2017 Annual Report

Kurt Fausch
Stan Gregory
William Jaeger
Cynthia Jones
Alec Maule
Peter Moyle
Katherine Myers
Laurel Saito
Steve Schroder
Carl Schwarz
Tom Turner

ISAB 2017-2
October 18, 2017

Review of the Comparative Survival Study (CSS) Draft 2017 Annual Report

Contents

I. Background 1

II. Summary 1

III. Suggested Topics for Further Analysis 2

IV. Comments on New or Updated Analyses in the CSS draft 2017 Annual Report by Chapter 5

IV.1. Chapter 1. Introduction 5

IV.2. Chapter 2. Life cycle modeling of alternative spill and breach scenarios 6

IV.3. Chapter 3. Effects of the in-river environment on juvenile travel time, instantaneous mortality rates and survival 9

IV.4. Chapter 4. Patterns in annual overall SARs 10

Appendix B: Supporting tables on Chapter 4 - Overall SARs 10

IV.5. Chapter 5. SARs and productivity 11

IV.6. Chapter 6. Estimation of SARS, TIRS and *D* for Snake River Subyearling Fall Chinook..... 11

IV.7. Chapter 7. Patterns of variation in age-at-maturity for PIT-tagged spring/summer Chinook in the Columbia River Basin 11

IV.8. Chapter 8. CSS adult success: summer Chinook, Snake River sockeye and steelhead 12

IV.9. Appendix A: Survivals (SR), SAR, TIR, and *D* for Snake River Hatchery and Wild Spring/Summer Chinook Salmon, Steelhead, and Sockeye........ 17

IV.10. Appendices C through H 17

V. References 17

VI. Editorial Suggestions........ 18

Chapter 1........ 18

Chapter 2........ 18

Chapter 4........ 20

Chapter 5........ 20

Chapter 6........ 20

Chapter 7 20
Chapter 8 21

Review of the Comparative Survival Study (CSS) Draft 2017 Annual Report

I. Background

The Columbia River Basin Fish and Wildlife Program tasks the Fish Passage Center Oversight Board to work with the Fish Passage Center (FPC) and the ISAB to ensure independent and timely science review of FPC's analytical products. These reviews include evaluations of the Comparative Survival Study's (CSS) draft annual reports. The ISAB has reviewed these reports annually beginning seven years ago with the evaluation of the CSS's draft 2010 Annual Report and most recently the draft 2016 Annual Report (ISAB 2010-5, ISAB 2011-5, ISAB 2012-7, ISAB 2013-4, ISAB 2014-5, ISAB 2015-2, ISAB 2016-2). This review of the draft 2017 CSS Annual Report is the ISAB's eighth review of CSS annual reports.

The ISAB's review was aided by three presentations from CSS researchers at the ISAB's September 15, 2017 meeting: Treatment of total dissolved gas (TDG) in survival monitoring, SARs and Productivity, and Life cycle model evaluation of Snake River spring/summer chinook under alternative spill and breach scenarios.

II. Summary

This ISAB review begins with an overview of the latest report's key findings (this section). It then moves to suggesting topics for further CSS analysis (Section III), general comments on each chapter of the 2017 CSS Annual Report (Section IV), and ends with editorial suggestions (Section VI).

The annual CSS report is a mature product, typically including only updates with the latest year of data and expansion of analyses as more data are acquired. Many of the methods have been reviewed in previous ISAB reports, and so these methods now receive only a cursory examination. As more data are acquired, new patterns and questions arise on the interpretation of the results—this interpretation is now the primary focus of our reviews. The ISAB appreciates the CSS's detailed responses to suggestions provided in previous reviews, and we do not expect the CSS to necessarily respond immediately to new requests for further analyses by the next report—see the CSS response (Appendix J) to the ISRP's 2016 review.

Chapter 2 (Life-cycle modeling) has been updated with a revised fit of the life-cycle model using more data, and now separate smolt-to-adult ratios (SARs) are modeled for in-river and transported fish. As in the last report, the model examined 12 spill/flow scenarios. Similar to last year's results, more spill generally leads to higher in-river survival and improved SARs. A new component to this chapter is the modeled scenario exploring the impact of breaching four Snake River dams. The model predicts an increased in-river survival by about 10 percentage points and a doubling of SARs when dams are breached. The addition of the breach scenarios was a nice complement to the spill scenarios, producing interesting results. Further consideration of assumptions used in both sets of scenarios and recommendations for experiments (short of an actual breach) that could be done to test model results would be useful.

Chapter 3 (Effect of the in-river environment on juvenile travel time and survival) is updated with new data. In addition, a preliminary investigation of the impact of total dissolved gas (TDG) on the instantaneous mortality and survival probabilities parameters was made using a graphical approach. While this approach did not show any evidence of an impact of TDG on either instantaneous mortality or survival probabilities, a more comprehensive approach of including TDG directly in the modeling process would address concerns about the interrelationship between TDG, spill, and flow that may confound results.

Chapters 4 (Patterns in SARs), Chapter 5 (SARs and productivity), and Chapter 6 (SARs for Snake River subyearling Chinook) are updated from previous years by including new population groups. As in past reports, pre-harvest SARs of 4%-6% are associated with pre-1970 levels of productivity for Snake River spring/summer Chinook. An unanswered question is that given the large amount of effort in the past to improve SARs through dam passage improvements, habitat improvements and other changes, to what extent might further improvements in hydrosystem management, predator control, and estuarine habitat lead to achieving SARs of 4%-6%?

Chapter 7 (Patterns of variation in age-at maturity for PIT tagged fish) is the same model as past years with recent data incorporated. The chapter is exploratory, and now sufficient data may be available to try to elucidate factors associated with observed changes.

Chapter 8 (CSS adult success) is a new chapter that looks at the relationship between survival of adults upstream of Bonneville and travel time, temperature, and arrival date. A complex modeling framework was used, but the ISAB is concerned that not enough assessment of the fit of the model to the data has been done to ensure that conclusions are appropriate.

III. Suggested Topics for Further Analysis

The latest CSS Report incorporates many of our past suggestions. For example, the current report has a substantial discussion of correlations among SARs from different regions or effects of transport on SARs (suggestion #1 in 2013; #1 in 2014). The life-cycle modeling now allows for variation in stream productivity and hydrosystem survival and simulates the correlative impacts of these changes on predicted future population abundances (#2 in 2013; #2 in 2014; #2, #3 in 2015). The ISAB appreciates this effort to respond to our past queries.

Some of the past recommendations from the ISAB appear to be beyond the current scope of the CSS (see several from 2016) but will become increasing important in the future. Some of our earlier and current recommendations may seem repetitive and unachievable to be accomplished within a year to inform the next report, but they deserve some forward planning as these issues will become much more pressing in the future. In particular, if there are data gaps, these gaps should be identified for potential new data collection procedures. When life-cycle models are modified, the modification should be flexible enough to incorporate these issues. This is reflected in our recommendations for future work below.

In 2017, we recommend the following topics for future reports:

1. Modeling flow, spill, and dam breach scenarios is very useful for policy makers. Consequently it is important that all assumptions be clearly stated and that the results are robust to these assumptions. For example, rather than using a single year to represent future flow conditions, variable flow conditions should be used to study the impact of flow/spill modifications under future climate change, and examine correlations between Pacific Decadal Oscillations (PDOs) and flows. What assumptions are being made about in-river predation under dam breach scenarios? What assumptions are being made about harvest under dam breach scenarios? Why are there discrepancies between the results of the Life-Cycle Model (Chapter 2) and the COMPASS model (refer to #4 in 2016)? The same scenarios should be run through both models and discrepancies resolved.

2. Include other important processes in the life-cycle models. In the current CSS analyses, each modeled population does not interact with other modeled populations as they migrate through the hydrosystem. Interactions among the various populations, including compensatory responses, are important and whenever possible should be folded into future modeling efforts, particularly if restoration actions increase the abundance of out-migrants.

   Similarly, there has been a great deal of interest in the impact of predator control programs on salmon returns, especially northern pikeminnow, birds and pinnipeds. Are these programs effective? Are there compensatory responses?

   Is there evidence in the existing data about either issue? What type of data would be needed to address these issues and include them in the life-cycle models? This recommendation builds upon our previous recommendation (#1 in 2015; #3 in 2016)

3. There appear to be sufficient data to try to elucidate reasons for shifts in the age distribution of returning spring/summer Chinook (# 5 in 2016). We suggest doing so.

4. The graphical analysis of the impact of TDG could be improved using direct modeling to deal with potential confounding effects of spill, flow, TDG, and temperature.

5. The (new) modeling of adult survival upstream of Bonneville should be continued and improved to identify the limiting factors to adult returns. Once these factors are identified, are there modifications to the hydrosystem operations that could be done to mitigate some of the factors?

6. The CSS report is a mature product and the authors are very familiar with the key assumptions made and the impact of violating the assumptions. These should be collected together in a table for each chapter to make it clearer to the readers of the report.

Below are the lists of topics we recommended in our reviews from 2013-2016. As noted above many of our recommendations were incorporated in subsequent CSS analyses and reports, and some recommendations require future planning and coordination with other entities.

In 2016, we recommended these topics (ISAB 2016-2, pages 5-6)

1. Use variable flow conditions to study the impact of flow/spill modifications under future climate change, and examine correlations between Pacific Decadal Oscillations (PDOs) and flows.
2. Examine impact of restricted sizes of fish tagged and describe limitations to studies related to types/sizes of fish tagged
3. Modify life-cycle model to evaluate compensatory response to predation.
4. Comparison of CSS and NOAA in-river survival estimates.
5. Examine factors leading to spring/summer Chinook declines of four and five-year olds and increases in three-year olds.

In 2015, we recommended these topics (ISAB 2015-2, pages 4-5):

1. Use SAR data to examine both intra- and interspecific density dependence during the smolt out migration and early marine periods
2. Propose actions to improve SARs to pre-1970s levels

3. Explore additional potential relations between SARs and climate and ocean conditions
4. Consider ways to explore the variability of inter-cohort response

In 2014, we recommended these topics (ISAB 2014-5, pages 2-3):

1. Hypotheses on mechanisms regulating smolt-to-adult return rates (SARs) [update from 2013 review]
2. Life-cycle modeling questions and Fish and Wildlife Program SAR objectives [update from 2014 review]
3. New PIT/CWT study

In 2013, we recommended these topics (ISAB 2013-4, Page 1):

1. Hypotheses on mechanisms regulating smolt-to-adult survivals (SARs)
2. Life-cycle modeling questions and Fish and Wildlife Program SAR objectives
3. Data gaps
4. Rationalization of CSS's Passive Integrated Transponder (PIT)-tagging, and
5. Publication of a synthesis and critical review of CSS results

IV. Comments on New or Updated Analyses in the CSS draft 2017 Annual Report by Chapter

IV.1. Chapter 1. Introduction

This chapter is similar to last year's report, and the ISAB has no major comments but has two suggestions for improving the chapter.

First, much of the text in this chapter is identical to text from the 2016 CSS report, which at times makes it difficult to determine if new data or techniques were incorporated and used in the 2017 CSS report. Some sets of data have been updated to include later years, but others have not. At times, it even appears that less data might be included. For example, Figure 1.6, lowest panel, presents less data than in 2016 CSS report. Why was the last data point in 2016 CSS removed, and why wasn't an additional year of data included in this panel? As another example, on p. 14, second paragraph, the text says "Two other new traps are planned to begin operation by 2016." This same wording was in the 2016 CSS report, and now that this report is being prepared in 2017, the wording makes it unclear if these traps did begin operating. A careful editing is needed in this chapter and other chapters to ensure that it is clear that year ranges are consistent (please refer to Section VI of this report for some other places where apparent discrepancies exist). The ISAB suggests that future reports include a table that shows for each chapter, which sets of data have not changed, which have received additional data, and which are waiting for additional data (and why there is a delay). For example, during the presentation to the ISAB, it was clarified that only some updated data were received (e.g., updated spawner numbers were received but updated smolt numbers were not received).

Second, more and more groups are being marked (which is good), but it may be expensive to achieve large enough sample sizes for each group as each population group becomes more and more specific. Is there a "lower limit" to the size of the population group that can be monitored? For example, Table 1.1. shows that the Lyons Ferry group has only 6,000 fish marked compared to 20,000+ elsewhere. Similarly, Table 1.2 and Table 1.3 have groups with small numbers of tagged fish (e.g., Table 1.3 shows that the Yakima steelhead group had only 250 tags applied). Is anything useful learned from a marked group this small? Do these groups provide any useful information? Based on past experience, what is the "minimum" number of tags per year needed before useful findings emerge? Can a combination of methods (e.g., genetics) be used to augment the CSS methods and therefore increase the value of tagging information?

IV.2. Chapter 2. Life cycle modeling of alternative spill and breach scenarios

The life-cycle model from previous years is changed slightly to include more years of smolt data and is now also modeling the SAR of in-river versus transported fish as well the overall SAR (the C0SAR, TXSAR, and SAR terms on page 32, respectively). Each of these SARs now contributes to the model-fitting procedure (but see the ISAB comment below regarding page 34). Additional years of spawner data are now available to add to those used in previous model fits.

Density dependence was modeled only at the smolt production stage. However, if we are restoring populations and they are co-migrating in the river, do we need to account for factor limiting in-river or estuary survival or juveniles and the returning adults? Can the pattern of PIT tag detections from each group plus reasonable estimates of production and forecasted travel time be used to forecast the "aggregate" smolt population or adult returns moving through the hydrosystem? These forecasts can be used to evaluate if the aggregate number of fish is being affected by constraints in the system.

The fitted life-cycle model was used to predict the impact of 12 spill/flow scenarios prospectively as was done in the previous year's report. Three "representative" years were used as surrogates for years with high, average, and low flows. Transport was set at 20% for future scenarios, but, as noted in previous comments by the ISAB, this may ignore that transport changes with spill.

Simulating the impact of breaching the four Snake River dams was done by reducing the power house contact index and decreasing the water transit time (Table 2.2) and using these revised values in conjunction with the 12 spill/flow scenarios. However, if the dams are breached, there will be many changes to the migration environment besides faster migration times (such as water transit time (WTT)). For example, an important change might be reduced predation by smallmouth bass. Is this reduced predation implicitly assumed to be a consequence of faster migration? Will future versions of the life cycle model include modeling the impacts of predation? The addition of the breach scenarios was a nice complement to the spill scenarios with interesting results. Further consideration of assumptions used in both sets of scenarios and recommendations for experiments (short of an actual breach) that could be done to test model results would be useful.

Similar to last year's results, more spill generally leads to high in-river survival and improved SARs. However, slightly different parameter estimates were obtained in the revised model fitting (p. 61) and the revised model no longer predicts that in-river survival is higher for lower flows at a given spill level. Nevertheless, the overall conclusions about the scenarios are basically unchanged from last year's report.

Results show that under dam breaching scenarios, in-river survival increased by about 10 percentage points and SARs doubled. These results assume no compensatory responses or density-dependent responses later in the life cycle, and this assumption needs discussion and investigation. As noted in the discussion in the CSS draft report, the benefits from dam breaching may be overstated because other changes to the hydrosystem and the ecosystem as a result of dam breaching are not modeled.

One compensating response that is included in the model is the human response, harvest. It would be useful in future analyses of spill and dam breaching scenarios to include results when harvest is held at current levels. While there are good reasons to assume that harvest will go up with increased abundance since that is likely, there are also good arguments for holding harvest constant for purposes of interpretation of the model results (the clarity of interpreting a *ceteris paribus* response). Given the magnitudes of the changes in abundance due to spill and breaching scenarios, interpreting the benefits of those actions would be less ambiguous if harvest were held constant (or otherwise if the benefits to harvest were added to the benefits of return abundance, but that raises additional complications). If harvest response was limited, how would that affect the results in terms of the magnitude of increased abundance, the rate at which those increases occur, and the differences in improved abundances for different populations?

A companion review of other life-cycle models (ISAB 2017-1) showed that the COMPASS model was not as optimistic about the impacts of spill/flow modifications. Has the COMPASS model been used for the dam breaching scenario as well? How do the results compare? This chapter should include a discussion of the comparison of this life-cycle results with the results from COMPASS and why they may differ.

Specific comments

p. 27. Here and elsewhere, the treatment of the proportion transported (PTRANS) in the modeling approach could be presented more clearly. On this page referral is made to the CSS 2016 Final report, but a description of PTRANS is not provided here. While it makes the report longer, it is helpful to have a fully self-contained document.

p. 28. The potential effect of distinct population timing is not considered. The report needs a discussion of how this assumption could affect model results (see also comment on page 62 below).

p. 30. From where does the assumption of transportation survival probability of 98% come?

p. 34. Equation 2.19. This assumes that the in-river SAR (C0SAR), the transportation SAR (TXSAR), and overall SAR are independent in a given year. A poor ocean condition would tend to depress all three values simultaneously, so this assumption may not be tenable. Additional justifications are needed for this assumption, e.g., show bivariate plots of the three response measures. A multivariate likelihood term may be needed.

p. 35. It is mentioned that “Transport was set at 20% for all future years…,” but this requires the reader to understand that “Transport” = PTRANS and that this means that PTRANS is equal to 20% for all scenarios (this statement only mentions the future scenarios). The rationale for setting PTRANS at 20% should be clearly explained (reference to McCann et al. 2016 is not sufficient).

p. 37. Regarding the use of 50% improvement in survival of adults migrating through four fewer dams when Snake dams are breached (i.e., 4/8), have the studies of adult migrants in the Columbia and Snake rivers (e.g., Keefer et al. 2008) been reviewed to see if those studies and perhaps other adult detections at dams conclude whether or not mortality at each dam (or just Snake dams versus Columbia dams) is equivalent? The models of Chapter 8 may be suitable to get a better estimate of adult survival in a breached hydrosystem.

p. 38. In CSS 2016, the authors used 10,000 draws for the simulations, but in CSS 2017 they used 1,000 draws. What was the rationale for changing the number of draws? Are there any implications of the change in number of draws on the results?

p. 38. In the description of the Markov Chain Monte Carlo (MCMC) simulations, the wording was changed here from CSS 2016 and was difficult to follow. The description in CSS 2016 was clearer; does the change of wording mean the approach was changed? If so, why?

p. 38. Why did the initialized population projections of empirical spawners change from 2010-2014 in CSS 2016 to 2010-2013 in CSS 2017? Did this change affect results?

p. 44. Figure 2.4: Why was PTRANS removed from this figure compared to last year's report?

p. 45. Figure 2.5. The fits for transport SARs look better than the fit for in-river SARs. Is there an explanation for the lack of fit in the latter? Is there evidence that can be presented to support the hypothesis for this on page 41?

p. 62. The added discussion is a good addition, but only one assumption is discussed. It would be helpful to provide references on some of the findings from other studies noted here. The chapter could benefit from more discussion of other assumptions such as 1) that distinct population migration timing is not considered (p. 28); 2) PTRANS is being set at 20%; 3) PITPH, WTT, and flows are being fixed for future scenarios; 4) transportation survival probability is set at 98%; 5) harvest levels also changing. How might results differ if these assumptions were varied? How certain are the authors of the values they used?

IV.3. Chapter 3. Effects of the in-river environment on juvenile travel time, instantaneous mortality rates and survival

The analysis of juvenile travel time, instantaneous mortality and survival is updated with new data and the same methods as previous years. The ISAB has no concerns about this part of the chapter.

The chapter also investigated the impact of total dissolved gas (TDG) on instantaneous mortality and survival probabilities. This was done by plotting the regression RESIDUALS from the models from past years versus TDG and looking for a pattern (Figure 3.9 and similar). This method is equivalent to investigating the impact of TDG after adjusting for the other covariates. The results are interesting. However, this method cannot account for the correlation between TDG and the other environmental variables.

A preferred method would be to include TDG directly into the regression models of previous years and again use an information theoretic framework. If TDG is a preferred covariate over the other covariates, this will be demonstrated by a high model weight on TDG. The correlation between TDG and the other environmental variables should also be explored (e.g., show plots of TDG versus the other environmental covariates) and this may help explain the lack of observed effect. The combined analysis may also indicate that the effects of variables highly correlated with TDG should have a quadratic component in the models.

Specific comments

p. 70. Equation 3.6 uses a log transformed instantaneous mortality (Z) rather than a square root equation, so results are not directly comparable but will be similar. In general, a log() transformation is preferable to the sqrt() transformation on biological grounds.

p. 79. Some of the graphs show NA for the variable importance. Does this indicate that the variable was not considered at all or that models with the variable were included but had model weights that were 0?

IV.4. Chapter 4. Patterns in annual overall SARs

This chapter was updated by the use of newly acquired data and the inclusion of new major population groups (MPG).

Tables 4.1 shows that model weights are relatively diffuse, yet the results use only the top model. Model averaging should be used for all analyses, predictions, figures, and discussion.

CSS reported on the relatively large absolute difference in SARs based on PIT-tags versus run reconstruction (the values are highly correlated, however). As in previous reports, this chapter listed various hypotheses. An email from Michele Dehart to the ISAB indicated that a study is underway to further evaluate PIT-tag effects on salmon survival, but results will not be ready for analysis until after summer 2017 when tagged age-5 Chinook have returned, and the report should be available in 2018. Potential bias in survival caused by tagging methodology (or in the run reconstruction methodology) is an important issue to resolve.

Specific comments

p. 91. The authors have removed “from outmigration to the estuary and ocean environments” at the end of the description of this chapter. The ISAB notes that this change was made between the Draft 2016 CSS report we reviewed and the Final 2016 CSS report. Is the rationale for this change explained in Chapter 4, and is it relevant? Is this in response to an ISAB comment?

Appendix B: Supporting tables on Chapter 4 - Overall SARs

The ISAB has no concerns with this appendix, which is an update from the one presented in previous reports. Information on additional MPGs and new data has been added.

IV.5. Chapter 5. SARs and productivity

Chapter 5 continues the examination of the relationships between life cycle productivity and SARs, including the level of SARs needed to reach or exceed population replacement. This year's report is an update from previous years and includes an addition of a new major population group (Pahsimeroi). The ISAB has no major concerns with this chapter.

How were hatchery origin Chinook adults on the spawning grounds identified and excluded from the productivity estimates of the natural-origin population? There is reference to the CSS estimates of natural origin returns (p. 137), but no details are provided. A sentence or two here would be useful to describe how hatchery Chinook returns are excluded to make this chapter self-contained. For steelhead, the text indicates a weir was used to identify natural origin adult returns and was used to exclude hatchery fish.

The findings suggest that pre-harvest SARs of 4%-6% are associated with pre-1970 levels of productivity for Snake River spring/summer Chinook. How do these SARs compare with SARs from viable wild Chinook populations in other regions? To what extent might improvements in hydrosystem management, predator control, and estuarine habitat lead to SARs of 4%-6%? From the results in Chapter 4, it seems major improvements would be needed to reach the desired SAR range given recent poor ocean conditions which led to low SARs. This was a query by the ISAB in previous years, but it does not seem to be addressed.

IV.6. Chapter 6. Estimation of SARS, TIRS and *D* for Snake River Subyearling Fall Chinook

This chapter contains updates from previous years on SARs by route of passage and TIR for 2006 to 2013 with the addition of new MPGs. The ISAB has no major comments.

The 2016 CSS report included a statement regarding estimating bias in SARs due to detections of overwintering and holdover detections and late season migrants through simulations. This was removed from the 2017 CSS report. Is this issue resolved?

IV.7. Chapter 7. Patterns of variation in age-at-maturity for PIT-tagged spring/summer Chinook in the Columbia River Basin

This chapter examined the mean age-at-maturity among different stocks, years, and fish type (wild or hatchery) using regression methods and is an update from last year with new data. The ISAB has no major concerns about this chapter but hopes that analyses to explain the year effects is ongoing.

A two-factor ANOVA was used to analyze the mean age-at-maturity. The analysis of the proportion of age-3, age-4, and age-5 fish was done using a binomial logit model in a Bayesian context with overdispersion.

There was evidence of both stock and year effects with 2007, 2008, and 2010 resulting in the lowest mean age of maturity and highest proportion of age-3 returns. Is there anything obvious about these years either in the river system or ocean conditions that could serve as a hypothesis for future years? As noted on page 174, one consideration for the future work is looking for factors associated with the year effects such as ocean factors associated with growth and climate change, differences in hatchery practices, or freshwater environments (tributary temps, or annual differences in migration corridor). Some preliminary investigations should be made, e.g., using plots. The authors should continue to coordinate with the Council's Ocean and Plume Science and Management Forum.

Specific comments

Chapter 7 in the 2016 CSS report previously examined delayed mortality, but this topic was not covered in this year's Chapter 7. Was this topic covered elsewhere in the report? If it was removed, a reference/appendix should be added to include this information, if still relevant.

p. 170. The Bayesian model should be described here to make the CSS document self-contained as many readers will not be familiar with fitting overdispersed models using Bayesian methods. This could go into an appendix for this chapter.

p. 170. A simple regression model applied to the empirical logit (Wharton, 2011) would automatically account for overdispersion in the residual variance term and is an alternative to the Bayesian model. No change is necessary, but just mention this in the text.

p. 176-178. The proportions of age-3, age-4 and age-5 must add to 100%. Does a ternary plot (https://en.wikipedia.org/wiki/Ternary_plot) with the points jointed over time provide any insight?

IV.8. Chapter 8. CSS adult success: summer Chinook, Snake River sockeye and steelhead

In this chapter, the authors look at the impacts of water temperature (as recorded at Bonneville) and transportation (as juveniles) on travel time and survival between Bonneville and McNary dams for summer Chinook, sockeye, and steelhead that were PIT-tagged from 2003 to 2016.

For travel time, the authors fit a Generalized Additive model (GAM) between the log(travel time) and the temperature and arrival date. A comparison of the proportion of fish that had a long travel times (e.g., more than 10 days for Chinook) was made.

A multi-step procedure was used to investigate the relationship between survival and temperature, arrival time, and transportation. Because detection probabilities are very high at McNary Dam (99%+ for Chinook and steelhead and 97%+ for sockeye), a detection/non detection at McNary is tantamount to survival/not survival between Bonneville and McNary Dam. A GAM was first fit to the detect/non-detect at McNary using a smoothed function of temperature and linear function of arrival date. Plots of the fitted relationship with temperature and arrival date were used to assess if the relationship of these covariates with survival was linear, piece-wise linear, quadratic, or cubic. These relationship forms were then used in a mixed-effects logistic regression (again using detection/non-detection at McNary as a surrogate for survival) to explore the impact of transportation, interactions between the covariates and transportation, and random effects of years and individuals. Backwards stepwise selection and AIC were used to select the most appropriate model. Finally, the most appropriate model was then fit using a Cormack-Jolly-Seber (CJS) model to the detection history at McNary and upstream of McNary using a Bayesian state-space model.

General comments

The chapter should indicate why travel time and survival between Bonneville and McNary is of primary interest, but travel time and survival upstream of McNary is not. Are data upstream of McNary lacking?

This analysis does not include spring Chinook. Why not? This may have been mentioned in earlier reports, but it is not clear in this chapter. The authors should consider linking this type of analysis to estimate the change in adult survival if the Snake River dams are breached, as in Chapter 2.

This chapter needs more model assessment (absolute goodness-of-fit assessment) for all models. A Bayesian *p*-value check was done for the final Cormack-Jolly-Sever (CJS) model, but basic plots of observed data versus predicted values are missing. For example, when fitting models for survival to McNary Dam, no plots were shown of the empirical survival probabilities (e.g., in degree intervals) for transported/non-transported fit to assess if Figure 8.4 is realistic or not. Similarly, no plots were shown of the empirical mean travel time (again in degree intervals) versus temperature to compare to the fit shown in Figure 8.2. Without such plots it is difficult to determine if some predictions (e.g., dramatically different predicted survival or travel times in the extremes) are simple artefacts of the model fitted or are actually supported by data.

There is some potential bias inherent in using travel times only of fish that survive between Bonneville and McNary dams. This is recognized by the authors (page 204) and adequately discussed.

The choice of linear, piece-wise linear, quadratic, or cubic relationships using AIC or backwards selection is an attempt to fit an appropriate functional form. The method used to choose the breakpoint for the piece-wise linear function is *ad hoc* (depends on spotting a break in a plot), is not recommended, and further analysis does not account for the uncertainty in the choice of the breakpoint (Toms and Lesperance 2003). Because only the functional form of the relationship is of interest, the authors should investigate the use of smoothing splines rather than using *ad hoc* methods to select an appropriate functional form. Smoothing splines will "automatically" adjust for non-linear relationships and will not suffer from the (observed) problems of fitting a cubic, which has unrealistic predictions in the extremes because of the degree of the polynomial.

Fitting models in this chapter with random effects of individuals is problematic. No fish was observed more than once at a dam, and no fish was observed in more than one year. So the individual effect is completely confounded with residual error and reflects overdispersion in the logistic mixed-effect analysis or CJS analysis. Unfortunately, with individual data, it is extremely difficult to estimate overdispersion using random effects. Standard diagnostics, such as deviance/df, perform poorly. To investigate overdispersion, the authors will have to combine data into groups (e.g., release groups when juvenile) and use these groupings to estimate the degree of overdispersion. Also, fitting individual random effects is numerically problematic for both the mixed-logistic model and the CJS model because of the very large number of parameters introduced. It would not be surprising if the mixed-logistic model has convergence issues and the fit was not reliable. It would be more profitable to try to estimate the degree of overdispersion in the data and, if needed, use the standard methods to adjust for overdispersion (such as variance-inflation factors).

Specific comments

p. 185. "… water temperature (with smoothing function)." The authors need to explain in more detail what smoothing function is used in the GAM or point to the appropriate *R* software reference.

p. 185. "We decided on a model that was the most biologically plausible to us and/or the lowest AIC value (Akaike 1973). Then we selected explanatory variables using a backward elimination stepwise process. During which, we started with all candidate variables in the model, and deleted them one by one until no more could be eliminated without losing significant model fit. We evaluated the model fit using a sequence of likelihood ratio tests and comparisons of AIC values." This is a mixture of AIC and hypothesis testing after a stepwise elimination procedure which is extremely *ad hoc* in nature. We suggest either adopting AIC for the entire process, or as noted in the general comments, switching to splines that will automatically "adjust" for non-linear relationships.

p. 185. "(b) a quadratic form." The usual practice is to fit a centered quadratic form to reduce correlation between Temp and Temp2 terms. Was this done? [It appears to have been done in the Bayesian analysis, but it is unclear if this was done for the non-Bayesian analyses.]

p. 186. "(a) a third degree polynomial." Third (and higher) polynomials are seldom appropriate given the severe restrictions placed on where the curve increases/decreases with the explanatory variable. If more than a quadratic is to be considered, a fit using splines with a small number of knots is likely more interpretable and not that much more difficult to fit.

p. 186. A paired-test appeared to be used to compare the proportion of fish with > 10 days travel time. It is not clear what the pairing variable is (year, temperature group, both?). Because proportions are being compared in a blocked design, a Manzel-Cochran-Hanzel test is the appropriate version of the chi-square test to use. Alternatively, many authors suggest a t-test on the empirical logit of the proportions (Warton 2011) rather than the raw proportions.

p. 190. It is not clear how the piece-wise linear function of temperature was fit using the notation given in the table. Usually, a single new variable is created which is 0 prior to the breakpoint and then (x-breakpoint) following the breakpoint.

p. 190. "… mean survival rate." What is a "mean survival"? The authors interpret the (anti-logit) of the intercept as the "mean survival," but the intercept does not have this interpretation. These are not rates as there is no time element to the survival probabilities.

p. 190. Not clear exactly what model is being fit. The authors created a 3-element detection history with the first element representing the "release" at Bonneville, the second element representing detection at McNary, and the third element representing detection upstream of McNary. There are two survival probabilities in the CJS model. The first is from Bonneville to McNary and the second is upstream of McNary. The latter is confounded with detection upstream of McNary. Why doesn't Table 8.4 have this second survival parameter? The appendices appear to show that the same survival probability was assumed between Bonneville and McNary and upstream of McNary, but this was never explicitly stated. The model with equal survival "breaks" the confounding seen in the last interval of the CJS model, but there is no information to assess if this assumption is justified.

p. 192. Figure 8.4. It is not clear how the temperature-specific credible intervals for the survival probability were computed. These appear to be based on some sort of envelope of the fitted curves, but how were the fitted curves selected to represent the credible bounds. The easiest way to compute the temperature-specific credible intervals would be to predict the response at selected temperatures and use the 2.5 and 97.5 percentiles of the predictions.

p. 193. Figure 8.5. The X-axis is labeled as "Last" detection at Bonneville. What does the "last" refer to? Are these Julian dates? It is hard to infer the arrival distributions from the plots. We think you need to plot a smoothed density curve for the arrival distribution for the two groups of fish. There appear to be missing temperature data for some fish (e.g., in 2004). Were the temperature data then interpolated for these fish?

We make similar comments for the analyses of the sockeye and steelhead data.

p. 195. Rather than using a cubic polynomial, we suggest using a spline with a few knots. Indeed, this would likely alleviate the problem where the authors choose the quadratic model because the results were more plausible.

p. 196. The report indicates the individual random effects were included, but there are no variance terms for individuals in Table 8.8.

p. 206. The state space model indicates that there are two survival probabilities (as noted earlier). Only by reading the explicit equations, is it possible to deduce that the authors are forcing the two survival probabilities to be equal. Please explain explicitly in the text.

It is not clear why the observation process has the detection probabilities having a Uniform(0,1) distribution. Perhaps this is the Bayesian prior? If so, this is an odd choice for a prior because there is very good prior knowledge that the detection probabilities are 90%+. Fortunately, the choice of prior for the detection probabilities is somewhat moot given the very large sample sizes.

The model for the survival probabilities indicates that the two survival probabilities are forced to be equal, which was never indicated in the text describing the model.

p. 209. Goodness of fit test should not include $t=1$ because the model only uses fish released alive at $t=1$ and there is no stochasticity.

p. 213. Why did the authors switch to N() priors for this species compared to Cauchy priors for the previous species? Given the large sample sizes, the choice of priors is somewhat moot.

IV.9. Appendix A: Survivals (SR), SAR, TIR, and *D* for Snake River Hatchery and Wild Spring/Summer Chinook Salmon, Steelhead, and Sockeye

This is an update from the previous year's report. New figures are included that now display confidence limits for many parameters and are broken out by population group.

IV.10. Appendices C through H

The draft that we reviewed did not include these appendices, but they will be included in the final CSS report and will likely contain updates for 2017. For Appendix C, the authors should refer to the ISAB's comments on Chapter 1 about the small number of tags in some groups.

V. References

Keefer, M.L., C.A. Peery, and C.C. Caudill. 2008. Migration timing of Columbia River spring Chinook salmon: effects of temperature, river discharge and ocean environment. Transactions of the American Fisheries Society 137, 1120–1133.

Toms, J.D. and M.L. Lesperance. 2003. Piecewise regression: A tool for identifying ecological thresholds. Ecology 84, 2034-2041.

Warton, D.H. 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology 92, 3-10.

VI. Editorial Suggestions

Chapter 1.

There are some places where the exact same years of data or exact same text are used as in CSS 2016 without being updated for 2017. These should be checked:

p. 14. Refer to second paragraph noted above ("Two other new traps…")

p. 14. The last two sentences have the same years as CSS 2016.

p. 15, middle of third paragraph. "To maintain a time-series of PIT-tagged…and this is expected to continue through 2015." The next sentence seem to indicate that things changed in 2015, but the wording of "expected to continue through 2015" makes interpretation of what happened confusing.

p. 15, first sentence of last paragraph. This uses the same years as CSS 2016.

p. 17, last sentence of first paragraph. Wording is identical as CSS 2016 and states that "…2016 CSS report will include SARS for upper Columbia wild summer Chinook (RRE-BOA, MCN-BOA) for MY 2011-2013 (and possibly 2014)." Now we are reviewing 2017 CSS report, so did these things get included?

p. 23. Appendices I and J should refer to the 2017 CSS meeting and report.

Figure 1.1. Define C0, C1 in the caption, and perhaps use T instead of transported (once defined) to be consistent with using C0 and C1. These terms did not get defined until p. 11 and so the caption to this figure using these terms was confusing.

Table 1.3. USFS Grand Total is "#######."

Chapter 2

General proof reading and editing required.

p. 32. Equation 2.12. Why isn't the overall SAR normalized by the number of smolts initially produced for that brood year?

p. 32. Equation 2.13. What is the phi term in equation 2.13 and 2.14? It does not appear in Table 2.1. These SARs are normalized by the number of smolts that are transported.

p. 31 and onward. Consistency of capital S versus lower case s is needed. According to Equations 2.1 and 2.8 and Table 2.1, capital S refers to number of spawners, whereas lower case s refers to survival (e.g., Equation 2.2, Table 2.1, Equations 2.9-2.11). However, the results tend to use capital S as survival, which is very confusing (see for example, pages 32 below Equation 2.8, 40, 41, Figure 2.5, and other places). Capital S is used for spawners in Figure 2.7 (which is correct usage according to the equations).

p. 33. Equation 2.17: What is “n” in the denominator?

p. 37. Should “and breach conditions” be added to the end of the second sentence (i.e., “In the hydrosystem, those are the four spill alternatives evaluated at three flow scenarios and breach conditions.”)?

p. 38, 8th sentence, “conditions can be the same as historical conditions, i.e., same environmental.” Should “same” be “similar”? Historical conditions are not exactly being replicated.

p. 38, end of 8th sentence. There appears to be a typo on “to mimic a of interest”

p. 38, 10th step of first logic description. Should “breach” be added to “each spill scenario j”?

p. 41, second full sentence. Change “with the joint posterior” to “when the joint posterior”

p. 48, second full sentence. Should Figure 2.10 be Figure 2.9?

p. 48, second line. This should this read “The SAR can be viewed as more THAN a smolt to spawner ratio.”

p. 49. Figure 2.8 is a common way to present three variables using a three-dimension plot, but some readers may have difficulty reading the plot. Other representations may be easier to read. For example, plot PITPH versus WTT in two-dimensions, with points for the 12 scenarios joined to show how breaching changes these two variables and a “circle” representing in-river survival with the size of the circle or the shading of the circle increasing as survival increases?

p. 51. Put Figure 2.10 and 2.11 on same graph for breach versus non-breach comparisons. Same comments for other graph pairs as well.

p. 53. Figures 2.12, 2.13 versus Figures 2.14 and 2.15: Why is it necessary to have both sets of figures? The paragraph about these on Page 48 is confusing, and it is not clear why it is necessary to show both sets of results.

p. 54. Isn’t Figure 2.12 contained in Figure 2.13 with the bottom set of lines in each panel?

p. 58. Place Figure 2.16 and 2.17 on the same plot.

p. 61. The last sentence of 2nd paragraph (starts with “PITPH for John Day Chinook…”) is worded awkwardly and is hard to follow. Can it be written more clearly?

p. 61, second to last sentence. What are the two competing perspectives referred to here?

p. 62, fifth sentence of conclusions. Does “empirically” mean “experimentally”? If so, suggest using “experimentally” because “empirically” has an ambiguous meaning in a chapter about modeling.

Chapter 4

p. 99. Keep figure legends on same page as figure.

p. 99. The phrase “decreased four-fold” is an unclear way to say something decreased by three-quarters.

p. 109. This chapter uses log() to represent natural logarithmic transformation while other chapters use ln(). Please be consistent across chapters. Table 4.1 uses both notations.

p. 110. Table 4.1 Many readers will not be familiar with the R convention that a colon (:) represents purely an interaction while an asterisk (*) represents both main effects and interactions. Use a computer language independent convention.

Chapter 5

p. 138. Some of the terms in the model were not displayed properly in the paragraphs under the equation.

p. 145. Table 5.3. The column for the slope is labeled as “-B” because the model used “–beta” in the model. This is confusing. Why not fit the model using “beta” directly and avoid the negation in the presentation? Similarly, the hypothesis of beta < 0 would be changed to beta > 0 because the model equation uses “-beta.”

p. 148. Figure 5.5 (and others). Add the average SRI to the graphs, i.e. add a line segment at -2.02 for the top panel.

p. 149. Figure 5.6. Try to avoid the same symbol with different colors for different years so that black and white copies can be read properly or readers who have difficulty distinguishing color are not confused.

Chapter 6

p. 155. What are “BDA” and “IPC” used in several tables (Tables 6.1, 6.2 and others)?

Chapter 7

p. 175. Figure 7.2 (and others). Y-axis and X-axis tick-mark values not aligned properly.

p. 179. Figure 7.6. Y-axis legend seems to be truncated and not centered on axes. X-axis legend should be centered on axis.

Chapter 8

p. 187. Figure 1. The figure could be improved by plotting the midpoints of the temperature range on the bars.

p. 188. Figure 2. Plot the raw data along with the fitted curve to provide some assessment of the fit.

p. 189. It is usual in AIC tables to show the number of parameters and the likelihood value as well. Same comment applies to all AIC tables in this chapter.

p. 197. Figure 8.9. "… focuses on the top 95th percentile." Authors likely meant the temperatures above the 5th percentile.

p. 209. Figure 8.16. Legend uses "eta" rather than "beta" and uses "$" (similar to LaTeX).

p. 214. "… some parameters has effect sizes smaller than desired." What does this mean?

p. 217 (and elsewhere). Reporting *p*-values that are very small is rarely useful or sensible. Just state that p<.0001 for very small *p*-values*.*