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 Comprehensive Passage (COMPASS) Model – Version 1.1

Richard Alldredge
Nancy Huntly
Stuart Hurlbert
Roland Lamberson
Colin Levings
Robert Naiman
William Pearcy
Dennis Scarnecchia
Peter Smouse
Chris Wood

ISAB 2008-3
June 2, 2008

ISAB Review of the COMPASS Model – Version 1.1

Contents

Executive Summary 1

Background and Charge from NOAA Fisheries 2

Response to NOAA Fisheries' questions 3

Response to Previous Critique 6

Detailed Commentary on Individual Sections 8

- Main Text 8
- Appendix 1: PIT-tag data 10
- Appendix 2: Model Diagnostics 10
- Appendix 3: Alternative Models 11
- Appendix 4: Dam Passage Algorithms 12
- Appendix 5: Dam Survival Estimates 13
- Appendix 6: Hydrological Processes 14
- Appendix 7: Uncertainty and Random Effects Modeling 15
- Appendix 8: Prospective Modeling 16
- Appendix 9: Sensitivity Analysis 16

ISAB Review of the COMPASS Model – Version 1.1

Executive Summary

This report is the fourth in a series of ISAB reports pertaining to the development of the Comprehensive Passage Model (COMPASS) created by NOAA Fisheries along with federal, state, and tribal agencies and the University of Washington. COMPASS is intended to predict the effects of alternative hydropower operations on salmon survival rates and provide ongoing evaluation for the new Federal Columbia River Power System Biological Opinion (BiOp). COMPASS is a welcome addition to the analytical tools available to both scientists and managers. These periodic ISAB critiques have been explicitly intended to provide constructive suggestions to facilitate continuing development of a valuable modeling tool. The specific questions for this round of review, and our responses, are the following:

(1) *Does the model successfully perform the desired capabilities, as listed below?*

(a) *realistically portray the hydro-system and variable river conditions* - The fit to available in-river and hydro-system data is quite good. With a few exceptions, the model has captured the impact of the variables considered. The question of how well the model will work for river conditions encountered in future years must await later data.

(b) *allow for the simulation of the effects of management actions* - COMPASS will permit evaluation of a reasonable range of management options, though the passage data are still insufficient to fine-tune the management choices. Full-blown management simulation is (mostly) a future challenge for COMPASS, but the possibilities are promising.

(c) *characterize uncertainty in prediction* - This version provides improved treatment of uncertainty, allowing for the correlation of estimates from sequential projects. The uncertainty is separated into components for stochastic sampling and for differences among time periods. How well that treatment serves the simulation effort must await a fully simulation-capable version.

(d) *represent hydro-system-related effects that occur outside the hydro-system* - The Bonneville Dam (BON) → Ocean → BON survival component of the model is still poorly characterized, in the absence of reliable data from below Bonneville Dam. The ISAB's sense is that continuing to elaborate latent mortality is somewhat pointless, given the lack of comparable data from the pre-hydro-system period. The ISAB also concluded that it is time to separate the detailed survival experience of fish transported from each collection point (Lower Granite (DAM), Little Goose Dam (LGS), Lower Monumental Dam (LMO), and McNary Dam (MCN) separately, because those seem to be different, and it is not possible to model the transportation alternatives if transported fish from all four projects are treated as a single cohort.

(2) *Is the model too complex or too simple?* - The ISAB's sense is that the model is now of about the complexity that will be useful, and it is manageable.

(3) *Does the model realistically represent the data and its variability?* - The model allows for variability in prediction, based on variability in the input parameters. And, at least where the requisite empirical data exist, the model does a credible job of reflecting a dynamic reality. The requisite data are sometimes in short supply, however, and both the COMPASS team and the ISAB recommend that more data of the necessary types be gathered.

(4) *Are the statistical methods sound?* - The COMPASS team's statistical methodology is generally sound, but questions remain about several of the methodology's finer points (the AIC criteria, log-linear vs. logit-linear regression and prediction, multinomial vs. normal error structures, and the inclusion or exclusion of a grand intercept term in the model). The effort is moving along nicely, but statistical methods are still evolving in this arena, and it is premature to view the methods embedded in COMPASS as firmly set.

(5) *Is the documentation adequate?* - The documentation is good, as far as it goes, though we offer some suggestions for additional improvements. The COMPASS team has decided to delay preparation of the User's Guide for a later effort. The ISAB's view is that deployment region-wide cannot realistically occur without that Guide. Strategic and management decisions are already being considered, and the BiOp is now reality, all of which argue for early availability.

In response to the ISAB's third-round critique of the COMPASS document, the COMPASS team provided a variety of responses. The ISAB has used this opportunity to provide additional feedback on those responses, by way of iterating the conversation.

Finally, we provide a detailed critique on each section of the current version of COMPASS 1.1. This critique is offered in the spirit of constructive suggestion to the COMPASS team, and we trust that our critique will be useful in continuing efforts to develop this valuable modeling tool for the region.

Background and Charge from NOAA Fisheries

At NOAA Fisheries' request of March 14, 2008, we reviewed the most recent Draft Version (1.1) of the COMPASS model. This report is the fourth in a series of ISAB reports pertaining to the development of this new comprehensive fish passage model, which was created by NOAA Fisheries along with federal, state, and tribal agencies and the University of Washington for use in informing the new Federal Columbia River Power System Biological Opinion (BiOp). In March 2006, the ISAB completed its first review of the then partially completed COMPASS model, specifically addressing several questions regarding model capabilities, complexity, data usage, statistical protocols, documentation, and graphical interface (ISAB 2006-2[1]). The ISAB concluded that the new COMPASS model should be a welcome addition to the analytical tools available to both scientists and managers. The ISAB's critique was explicitly intended to provide a series of strong but constructive suggestions to facilitate the continuing development of a valuable modeling tool for the region.

The ISAB's second review was a reply to the COMPASS team's responses to the ISAB's initial review (ISAB 2006-6[2]). The points at issue for both the ISAB's report and for NOAA's response were largely confined to statistical usage, over which there remained some differences of opinion that needed further discussion. The ISAB was encouraged by the efforts of the COMPASS team and provided comments to further the team's discussions and development of the model.

The third ISAB review assessed the COMPASS Model Version 1.0 (ISAB 2006-7[3]). ISAB commentary concentrated on issues surrounding dam passage algorithms and model uncertainty, and the resulting changes are presented in some detail in Appendices 4 and 7 of this new report, respectively. For this current report, the COMPASS team has also expanded its treatment of PIT-tag data (Appendix 1), diagnostics (Appendix 2), alternative survival models (Appendix 3), prospective modeling (Appendix 8), and sensitivity analysis (Appendix 9), respectively. The User's Manual has been deleted from Version 1.1, and will be dealt with at a later date.

The specific questions for this review round are the following:

(1) Does the model successfully perform the desired capabilities, as listed below?

- (a) realistically portray the hydro-system and variable river conditions
- (b) allow for the simulation of the effects of management actions
- (c) characterize uncertainty in prediction
- (d) represent hydro-system-related effects that occur outside the hydro-system

(2) Is the model too complex or too simple?

(3) Does the model realistically represent the data and its variability?

---

[1] www.nwcouncil.org/library/isab/isab2006-2.htm
[2] www.nwcouncil.org/library/isab/isab2006-6.htm
[3] www.nwcouncil.org/library/isab/isab2006-7.htm; Also see ISAB Latent Mortality Report 2007-1: www.nwcouncil.org/library/isab/isab2007-1.htm

(4) Are the statistical methods sound?

(5) Is the documentation adequate?

Response to NOAA Fisheries' questions

In general, Version 1.1 of the COMPASS Model is a marked improvement on Version 1.0. The documentation is more abundant, and the writing is much improved. As a consequence, the document is much easier to follow. The User's Manual is not at issue here, so the ISAB will refrain from commenting on it at this time, other than to make the obvious point that it deserves a major effort in its own right. We would urge timely attention to that Manual, because wide deployment of the COMPASS Model cannot occur until it becomes available to the many potential users. In general, we remain firm in our assessment that this product will be a useful tool for the management of salmonids in the Columbia River Basin.

Relative to the questions from NOAA, forwarded by Council, we have the following comments, separately for each section of the new report:

(1) *Does the model successfully perform the desired capabilities, as listed below?*

(a) *realistically portray the hydro-system and variable river conditions*

The statistical fit to the available in-river and hydro-system data is quite good. With a few exceptions, which we will address in our detailed commentary (see below), the model has captured the impact of the variables considered. The question of how well the model will work for river conditions encountered in future years awaits later data, but so far the results look good.

(b) *allow for the simulation of the effects of management actions*

Given the parameterization of the dam passage models and the data available to estimate those parameters (Appendix 5), the COMPASS platform should permit simulation of a reasonable range of available management options. Appendix 4 is devoted to evaluating the likely impact of different spillway and bypass options for particular projects, but, for routine runs, the "general configurations" on each project are constants for each period, which is an oversimplification. The passage data are not yet sufficiently complete to justify detailed fine-tuning of the model for actual practice, but COMPASS will clearly be able to accommodate detailed dam passage information for salmon as that becomes readily available. To simulate anything beyond the average configurations for each project, it will be necessary to incorporate the appropriate models into COMPASS and to have some calibration-caliber data on the alternatives, project by project, allowing further empiric testing of the model under variable management. COMPASS is not quite ready for routine simulation, and a final answer to this question must await simulation. Management simulation represents a future challenge for COMPASS, because the team will need better data support than exists to date, but COMPASS is now ready for the requisite data

(c) *characterize uncertainty in prediction*

Version 1.1 seems to allow for an adequate representation of uncertainty, and the models fit the data used to estimate the parameters well. The characterization of uncertainty allows for the correlation of estimates from sequential projects, as it should. The authors have modeled this as a multivariate normal error process and have separated the uncertainty into that due to stochastic sampling per se and that due to differences among time periods. The larger question is how well those measures of uncertainty serve the simulation aspects of the program, and we cannot evaluate that, pending a fully simulation-capable version.

(d) *represent hydro-system-related effects that occur outside the hydro-system*

Partitioning survival from Bonneville → Ocean → Bonneville is not easily done. There are still few reliable data on survival rates from the Bonneville → Estuary, Estuary → Ocean → Estuary, and Estuary → Bonneville phases of the life cycle. All one has are round-trip Bonneville → Bonneville survival rates, which still includes latent mortality components that are assumed to be the same for both in-river and transported fish, an assumption that is unnecessary (see the Latent Mortality review: http://www.nwcouncil.org/library/isab/isab_2007-1.htm). Given that D = ($SAR_{T,BON\rightarrow LGR}$) ÷ ($SAR_{I,BON\rightarrow LGR}$), the ratio of smolt to adult return rates for in-river vs. transported fish, is measurable and not dependent on any particular assumptions about in-river (L_I) and transported (L_T) latent mortality, there is no problem with providing an option for the user to "insert" arbitrary values. However, that option comes down to *assuming* something about latent mortality (relative to the pre-hydro-system period that is not supported by actual data).

Of greater practical concern is the treatment of all transported fish as interchangeable, regardless of the site (LGR, LGS, LMO, MCN) from which they were transported. The Comparative Survival Study's Ten Year Retrospective made it clear that smolt-to-adult survival rates (SAR-values) for smolts transported from sites lower on the river were higher than those transported from LGR (http://www.nwcouncil.org/library/isab/isabisrp_2007-6.htm). Whether the SAR_T improvements, measured from BON→ LGR, are offset by in-river survival losses from LGR → LGS, from LGR → LMO, and from LGR → MCN, respectively, for fish transported from those lower projects remains unclear. The next version of the COMPASS model should include provision for SAR values for transported fish, SAR(T_0 – from LGR), SAR(T_1 – from LGS), SAR(T_2 – from LMO), and SAR(T_3 – from MCN), but with due allowance for additional in-river losses from LGR to each of those projects. The question of whether to transport smolts from LGR, LGS, LMO or MCN may be a consequential management choice, so those options should be evaluated. One needs full survival accounting to evaluate those choices.

(2) *Is the model too complex or too simple?*

There is always a tradeoff between the need to provide enough detail to capture the nuances of the real situation and the need to keep the model as simple as possible. The ISAB's sense is that this latest version (1.1) of the COMPASS model strikes a healthy balance between simplicity and realism. This modeling tool has to serve myriad purposes, and we find ourselves calling for more detail at various points, while constantly reminding the team to "keep it as simple as possible."

(3) *Does the model realistically represent the data and its variability?*

The model allows for variability in prediction, based on variability in the input parameters. And, at least where the requisite empirical data exist, the model does a credible job of reflecting a dynamic reality. The requisite data are sometimes in short supply, however, and both the COMPASS team and the ISAB recommend that more data of the necessary types be gathered. The value of testing the performance of the model against the real world cannot be overestimated.

(4) *Are the statistical methods sound?*

The COMPASS team's statistical methodology is generally sound, but questions remain about several of the methodology's finer points (the precise usage of AIC criteria, log-linear vs. logit-linear regression and prediction, multinomial vs. normal error structures, and the inclusion or exclusion of a grand intercept term in the model). We comment in the context of the individual appendices, but statistical practice is fluid in this arena, and it is premature to consider the choices embedded in COMPASS 1.1 as firmly established.

(5) *Is the documentation adequate?*

The documentation is good, though we offer suggestions for additional improvements below. The decision has been made to delay preparation of the User's Guide for a later effort. The User's Guide represents unfinished business that will need timely attention. The ISAB's one comment on the matter is that any region-wide deployment of COMPASS cannot realistically occur without that Guide, and the urge to deploy is growing. Strategic and management decisions are already being contemplated (and argued about) that would profit from a COMPASS evaluation of alternatives.

Response to Previous Critique

The COMPASS modeling team has submitted a response to ISAB's most recent review (http://www.nwcouncil.org/library/isab/isab 2006-7.htm). Our reaction to that response is below.

(1) *Specificity of the modeling platform*

Response: The ISAB had voiced concerns about eventual expansion of COMPASS to *Evolutionarily Significant Units* (ESUs) other than Snake River spring/summer Chinook and steelhead. The response is that the platform can be expanded to other Columbia River Basin ESUs, provided that the appropriate calibration data exist. The necessary data do exist for the Upper Columbia River and many of its tributaries. The hope is to model spring-migrating ESUs (spring Chinook, steelhead, and sockeye) from the Upper Columbia River in the near future. At a later time, the hope is to model Snake River fall Chinook, though the life history is complex enough to represent an incremental challenge. Other ESUs in the Columbia River Basin (currently) do not seem to have sufficient data to support the enterprise. The COMPASS team will use data from similar ESUs as surrogates. The COMPASS platform is general enough that it could be adapted to other river systems in the future.

Reaction**:** We view these plans as commendable and would content ourselves with a trio of simple comments: (a) modeling from surrogate data is certainly better than not modeling at all, but the better (long term) solution is to obtain the data necessary to calibrate properly; (b) the complex life history of Snake River fall Chinook may well have some tricky management implications, and careful modeling will be important; and (c) while it is good to ensure that the COMPASS platform is general enough to allow adaptation to other drainage systems, to model the Columbia River Basin and its hydro-system accurately is a large enough challenge that it will keep this team productively engaged and quite busy for an extended period.

(2) *Survival probabilities > 1.0*

Response: The ISAB has raised several questions about the practice of allowing $\{S\}$ to be > 1 in COMPASS runs, because Cormack-Jolly-Seber (henceforth CJS) methods can sometimes yield $\{S\} > 1$. There has been an ongoing exchange between the COMPASS team and ISAB on what to do about this fact, both in terms of the best way to model survival, as it relates to river conditions and management scenarios, and in terms of setting survival probabilities for COMPASS simulations of downriver passage. The ISAB's suggestion that logit regression modeling might circumvent some of the limitations of CJS estimation has resulted in evaluation of logit modeling and a comparison with the results of classic log-linear modeling (see ISAB comments on **Appendix 3**). Our suggestion was to constrain $\{S\}$ to the [0,1] interval for deterministic and simulation runs of the model. The COMPASS team responds that they have constrained $\{S\}$ for deterministic runs, as requested, but that they have continued to let $\{S\} > 1$ for calibration.

Reaction: The basic dependence on CJS estimation of survival probabilities, while it has major attractions, has led the COMPASS team to survival estimates of $\{S\} > 1$ for a number of data points. This problem of survival probability estimates greater than 100% is particularly acute for small data sets, where precision, accuracy, and the ability to model productively are all poor. The COMPASS team's position is that CJS-generated data points with $\{S\} > 1$ are unbiased and that they provide allowance for substantial estimation error. One consequence of $\{S\} > 1$ is that collective likelihood values (L-values) for small data sets are also sometimes "out of bounds" (i.e., $L > 1$). That may be part of the cause of the AIC confusion (see comments on **Main Text**). Quite apart from the choice of model predictors, the "noise" in the survival values is also unavoidably large with small sample sizes, and, for modeling purposes, there is a natural urge to capture that variation in the modeling exercise, translating estimation uncertainty into an honest portrayal of the resulting predictive uncertainty. For simulation purposes, the question becomes whether to use the "out of bounds" estimates or to constrain the $\{S\}$-values to the [0,1] interval. If the COMPASS team wishes to ignore estimation error for modeling purposes – as suggested in **Appendix 7**, concentrating solely on the "process (survival sampling) error", then one could make the case that accounting for estimation error should not be a consideration in simulation modeling. However, if all the uncertainty is to be accounted for in simulation modeling, then one needs to account for both the estimation and sampling variation.

The COMPASS team should be consistent and explicit on its choice of estimation, regression-evaluation, and simulation methods. The choices matter, particularly where the available data sets are small, and that is precisely where the difficulties (and in all likelihood subsequent quibbles with the results) will emerge. Such data-poor cases are those that will be difficult to model well, and subtleties of technique might be consequential.

(3) *Modeling choices*

Response: The ISAB raised concern about using relative variance for weighting, as well as the log-linear form of the model used to describe the impact of several environmental variables on survival. The COMPASS team responded that $\mathrm{Var}\{S\}/\{S\}^2$ is the appropriate weight for $\log\{S\}$, so they have retained it. They concede that $\mathrm{logit}\{S\}$ performs well (**Appendix 3**, see below), but they prefer the log-linear form, and plan to continue using it, on the basis of theoretical and practical considerations.

Reaction: If log-linear modeling is to be continued, then we agree that relative variance weighting should suffice. On the question of whether it is better to use log-linear or logit-linear modeling, we have more to say in response to **Appendix 3**. However, we note here that logit-linear modeling naturally constrains $\{S\}$ to the [0,1] interval, and it seems to be working well in those cases tested, though we suggest that Figures A3 1 and A3 2 be plotted with alternative constructs (see comments on **Appendix 3**).

Response: The ISAB had suggested the generic use of a grand intercept, and suggested that it would be appropriate to allow for the negative correlations of CJS-derived $\{S\}$-“observations” from adjacent projects. The authors now use a covariance matrix for analysis and show that a grand intercept works particularly well for steelhead. They plan to explore that further in future work.

Reaction: We view the use of a formal covariance matrix for weighting as an improvement and view the inclusion of a generic grand intercept as a no-cost (and sometimes high-value) addition to the modeling strategy. We view the grand intercept model as the generic null hypothesis.

(4) *Stochasticity* (or more accurately, *model uncertainty*)

Response: The ISAB suggested in its previous review that the methods of characterizing the impact of model uncertainty should be reworked, and that it would be better to draw randomly from parameter space to seed any particular model run. The COMPASS team is now doing that.

ISAB Reaction: Sounds reasonable.

(5) *Partitioning survival variation*

Response: The ISAB had raised several concerns about partitioning survival variation across reaches (river segments). The authors have responded in detail in **Appendix 7**. There were also minor ISAB comments on documentation, biases, and diagnostics, all now addressed in

Appendix 2. Concerns were also expressed about the dam passage algorithms, now addressed in **Appendix 4**.

ISAB Reaction: We will comment under the sections for **Appendices 7**, **2** and **4**, respectively.

Detailed Commentary on Individual Sections

As is the ISAB's standard practice, we include below a number of particular comments on the individual sections of the report, intended to aid the COMPASS team in its continuing efforts to improve the performance of the model, as well as the documentation.

Main Text

Precision of the PIT-tag data depends on sample sizes, and wherever the sample sizes are small, precision is poor. Sample sizes for hatchery fish are much larger than those for wild fish, which has obvious consequences for precision. The practice in the past has been to pool both hatchery and wild fish to increase overall sample size, thus increasing precision, but as described in our comments on **Appendix 1** (see below), the average survival performance of hatchery and wild fish is different enough that it is better to keep them separate.

Sample sizes for spring/summer Chinook are larger than those for steelhead, again with the obvious consequences. The authors also note that the precision of the available PIT-tag data for the Snake River segment of the journey is greater than that for the Columbia River segment, and that precision for the lower Columbia River, based on limited detection capability below Bonneville, is poor enough that it compromises the ability to model effectively. They note the value of future studies that will provide more precise estimates for the lower Columbia, and the ISAB concurs.

First full paragraph on Page 13, line 3 – should read "**. . .** is run in scenario **. . .**"

Top of Page 14 – define FGE and SPE here. They are not defined until later, and premature use of acronyms does not aid comprehension.

R^2 – equation on Page 17 – Given that the modeling is done in $\log\{S\}$, rather than $\{S\}$, and given that the error structure (in $\log\{S\}$) is assumed to be multivariate normal, it would make more sense to evaluate performance (coefficient of determination) in terms of $\log\{S\}$, rather than in terms of S. The authors also try logit modeling later, where the appropriate scale is $\text{logit}\{S\} = \log\{S/(1 - S)\}$ instead, and there, it would seem to make more sense to use logit scaling for R^2. The comparison of R^2 across scales is problematic, of course, but the choice is usually dictated by the scaling appropriate for the error distribution, for any particular regression model.

AIC-weights on Page 17 – The weights $\{w_i\}$ for this equation are not the same as the weights for the R^2-equation above it, but that is the implication of using the same symbol for both. Change one or the other symbol.

Table 3, Page 18 – The AICc values for both the Chinook and steelhead models for the upper river (above MCN) are both negative, while those for both Chinook and steelhead for the lower river (below MCN) are both positive. The standard formula is AIC = $\{-2*\log L(\theta \mid \text{data}) + 2p\}$, where θ is the parameter set and p the number of parameters estimated for the model being evaluated. That renders AIC strictly positive, under normal circumstances, and the strategy is to minimize it, so something is odd here. In a subsequent discussion, Rich Zabel made the point that when $\{S\} > 1$, $\log\{S\} > 0$ and $-\log\{S\} < 0$, which might explain the odd results. The fact that all of the AIC values are large and negative for the upper river, however, suggests that many of the model-estimated $\{S\}$-values are > 1, and this is the sample set for which the data are quite ample. It may be that these strange results derive from the use of log-linear models, which are prone to this problem, rather than logit-linear models, which are not, but the only way to be sure is to track these odd results to their source, and we would suggest that the COMPASS team do that. There are different "correction factors" used to convert AIC → AICc; a subsequent memo from Rich Zabel indicates that the form being used here is AICc = AIC + $2p(p + 1)/(n - p - 1)$, which is a strictly positive correction. Define the AICc criterion explicitly in the text.

Figure 6, Page 19 – The top two panels indicate that the model fits data from the upper river reasonably well (and is relatively co-log-linear) for both species, but the lower two panels show that the model fits data from the lower river poorly. In fact, the models are poor enough from the lower river that they are almost irrelevant, R^2-values notwithstanding. The authors have pointed out the greater imprecision (variance) of the lower river data, so the scatter (and the limited ability to model in Table 3) are not a surprise, but we note that none of these models has an overall intercept (γ for the final equation on Page 12), and we have to wonder whether it might not be a good idea to insert such an intercept in the models. Considerations of parsimony suggests that such a model forms the natural reference (null hypothesis) condition in virtually every case, and that it should be evaluated as a default trial. The text is also a little unclear on the question of how the weighted models will be used for projection. The text states that the weights will be used for simulation, but will they use weighted combinations of the predictive regression coefficients or weighted combinations of the survival predictions? It would be good to clear that up by being algebraically explicit. Also, why not use the weights in a similar way to evaluate the data one has for adequacy of fit? Obviously, the fit will not be as close as when using the best model, but how much degradation of performance is there from using a weighted prediction?

Page 20, 2nd paragraph below ***Passage Efficiency Relationships*** – It would be better to word this as "**. . .** the points (0,0) and (1,1). **. . .**" and similarly, elsewhere in the text.

Appendix 1: PIT-tag data

Bottom of page 2 of A1-1 – Correlation between *observation* and *prediction*? In $\log\{S\}$ or in $\{S\}$? Be a little more explicit here. Also, for clarity, it would be better to use "r" than "ρ" for the estimates. The same comment applies to the tabular presentation.

Top line of ***Results*** – One too many repetitions of "are expressed"

Page 4 of A1-1, last line before ***References*** – misstatement, "**. . .** smaller fish **. . .** than smaller ones." Cannot be correct.

Tables on A1-1 page 5 – too many ruler lines in the tables.

Plots in A2-5-6 – What is "Wild Wild" on the axis for migration rate supposed to indicate?

Appendix A1-2, page 2: First line mentions having more precise survival estimates without suggesting how. A presentation of some particular alternative would be useful.

Appendix A1-2, page 2: last paragraph – should be "combining fish" rather than "combing fish".

The authors conclude that pooling wild and hatchery fish is not prudent, given their performance differences, even though the estimates for wild fish are less precise because of smaller sample sizes. The ISAB concurs. The COMPASS team attempted to increase sample sizes by using two-week cohorts, but the practice did not help enough to make the effort worthwhile. It will be important to face this reality systematically in future work. The obvious solution would be to tag larger numbers of wild smolts, but whether that is possible or desirable is an open question.

Appendix 2: Model Diagnostics

Page 1, Section 1, 1st paragraph – Plotting residuals versus observed values is not useful, because the residuals and observed values are correlated, whereas the residuals and the predicted values are not (Draper and Smith, Applied Regression Analysis, 2nd ed.). Replot.

Figure A2-1 2, upper left panel – same comment as for Figure 6 of the main text. The residuals are symmetric, which is good, but there is almost nothing one can say here.

Figure A2-1 4, upper left panel – virtually the same comment, but here the distributions, one mode for each reach. Both of these graphs show that we have almost no useful ability to model survival for the lower river, a point also made in the text by the authors.

Figure A2-1 5 through A2-1 8 – residuals become progressively asymmetric, probably a tipoff that a general intercept (γ) is in order for the model (there are non-trivial but unexplained delays that are not being accommodated by the models). The points do not lie along the 45^{o} line.

Survival diagnostics – The plots are far too small to read properly, much less to evaluate, though the results look reasonable, on quick visual scan. Perhaps place one set of six on a page, oriented in portrait fashion, which doubles the number of pages but renders them readable.

Passage distributions – Look good.

The larger punchline from this exercise is that given the limited data from the lower reaches, we can predict very little about survival from MCN → BON. That is a serious problem that only better data can solve, as the COMPASS team points out. Once again, the ISAB concurs.

Appendix 3: Alternative Models

The tables could be simplified by explicitly defining ΔAICc in the text and then suppressing AICc in the tables. For model weighting, it is ΔAICc that matters, and that presentation is clear.

Page 4, top paragraph – The meaning of "backed out" is not coming through. Do it algebraically to clear up any confusion.

Next paragraph – The difficulty of comparing logit{*S*} and log{*S*} is noted, but there might be another way to do that without using the method described here. One can compute the log-likelihood under both treatments, and log-likelihood "currency" is universal. The adjustments for numbers of parameters fit and sample sizes are identical, given the same predictor variables, so a difference in log-likelihoods between the two transforms of {*S*} is tantamount to ΔAICc, either positive or negative for any given sample (reach/year/project). Back-transformation into {*S*} and computation of an inverse-variance weighted sum of squared deviations of observed *S*-values from predicted *S*-values seems a bit off-target, given the basic log-linearity of both treatments. The object is <u>not</u> to minimize $SS\{S_{observed} - S_{predicted}\}$, but rather to minimize AICc. If plots are needed, as in Figures A3.1 and A3.2 (currently mis-numbered), plot log*L*(for logit{*S*}) – log*L*(for log{*S*}) against some (probably) logarithmic function of observed {*S*}. Any systematic departure of the difference from 0.0 will probably emerge in relation to that function of decreasing {*S*}, and a choice of modeling strategy should emerge from the results.

The comparison of models containing a grand intercept (γ) with those lacking such an intercept, plotted in Figures A3.3 and A3.4 (currently mis-numbered), seems to suggest that it would not hurt and would sometimes help to include that grand intercept. As pointed out under **Main Text**, we could probably view the grand intercept model as a generic null (reference condition).

Appendix 4: Dam Passage Algorithms

Previous treatment of the dam passage algorithms has employed wide averaging of parameter-values over dams and weekly cohorts, in the absence of sufficient data to explore differences. The COMPASS team has now begun to examine the finer points, dam by dam and period by period, and – where data permit – ESU by ESU and stock by stock. Not surprisingly, averages hide more than they reveal. The Appendix describes in some detail how estimates were obtained for each of the projects, using whatever data were available and the methods deemed most reasonable, given the particular (and variable) circumstances. The delivery provides a sense of the extent to which the current numbers ensure precision and accuracy, which varies quite a lot. Spill passage efficiency (SPE) and fish guidance efficiency (FGE) are considerably more variable than had previously been assumed, and the relationships with project configuration are complex. Within the spotty limits of available data, several prospective runs were attempted, some of them using project configurations at particular dams that are not (yet) possible, just to see what one might expect.

Page 2, 3rd paragraph, 2nd sentence: Replace "proposal" with "proposed." Also state what criterion was used to conclude that predictions were improved at some sites.

Page 5, 1st paragraph, last sentence: What are the consequences of extrapolating that single estimate to multiple years? Will the predictive ability of the models be overestimated?

Page 11, second paragraph, last line: In the absence of data at the high or low end of the range, why must the curves be extrapolated?

Page 14, first paragraph, last sentence: What is the justification for assuming equality of slopes across species and projects? Do data exist to support this assumption?

Page 20, first paragraph, first sentence: What is the justification for assuming that SPE and FGE are linear functions of explanatory variables on the logit scale? Do data exist to support this assumption?

Page 22, top of page and page 24, bottom of page: Is it possible that coding RSW as "on" if any of the fish in the cohort passed the dam while the RSW was on could create a bias? The RSW results seem odd. Could the RSW effect be collinear with another term in the model, thus influencing the sign of the estimated RSW effect?

Figures A4 1 through A4 5 – Valuable, but much too small to be useful. Consider placing one project (both species), plots + residuals, on one page (4 panels). The distinction between the bold line and the dotted line is neither clear nor indicated in the captions. We suspect that the dotted lines are the regression models, but then what are the solid lines and why do we need them? The residual patterns are revealing, and it would be good to see them in readable form.

The results suggest two conclusions: (a) Fine-tuning individual project configurations, in an attempt to optimize dam passage survival, should prove effective. The use of COMPASS to model alternative configurations shows a lot of promise. (b) The data available are still too spotty and incomplete, however, to support systematic deployment throughout the basin. It seems clear that augmentation of the current data set by additional and systematic experimentation with individual project configurations would constitute a high-return investment for the region.

Appendix 5: Dam Survival Estimates

The tabular presentation of these voluminous data would be well served by suppressing most of the ruler lines in the tables. It seems clear that once reliable data become available from more projects and configurations, it will be possible to mount telling evaluations of project survival, under various water management scenarios. It should even be possible to "optimize" survival through any particular project by "fine tuning" that model, but the current limitations of these dam passage survival estimates (see comments on **Appendix 4**) are substantial.

The COMPASS team has done what it can with the data available to extract some central patterns, but only better and more complete data will permit calibration of the required precision and accuracy. Each of the dams has its own characteristics that influence project survival through the season and for varying configurations. COMPASS 1.1 uses specific project survival values for each dam, for each of the weekly cohorts, and we are still a long way from being able to evaluate finely-tuned management alternatives with any precision, but we can obviously anticipate some substantive impact of water management alteration on dam passage survival.

Appendix 6: Hydrological Processes

The approach in COMPASS to determining water velocity in reservoirs from river flow and reservoir geometry arises from a simple, classic model. The development is clear, though it would be very helpful to have an illustration of the volume $V_2(E)$. The reservoirs in the Columbia River Basin have a relatively simple structure, so they should be well suited to this model structure. There appear to be sufficient data to calibrate the model. The model output provides a good fit to observations of pool volume over a range of pool elevations for Lower Granite and Wanapum reservoirs (Figure 3). More importantly, for Lower Granite, the water particle travel time predicted from flow (via COMPASS) seems to compare well with particle travel time computed by the US Army Corp of Engineers (Figure 4). The authors have not shown how well the reservoir pool volume (or pool elevation) predicts flow (or particle travel time), and that would be an informative addition to the Appendix. The exposition is clear, but there are a few places where further clarification would help.

Page 2 – Define “thalweg” volume in words or by reference to the diagrams.

Page 5 – Why is the John Day River value of 4.5×10^{-3} kfs used as a “default” value?

Page 6 **–** θ is defined in *arctan* terms. It would probably be better to use tan^{-1} than *atan*.

Appendix 7: Uncertainty and Random Effects Modeling

The authors describe an approach to estimating uncertainty in COMPASS projections of reservoir survival based on a variance component analysis. The variance component estimates are used in conjunction with a regression analysis that relates survival probability to distance, flow, spill, travel time, and temperature. The approach described is a step in the right direction for dealing with a model that contains both fixed and random effects. The ISAB encourages the timely implementation of an integrated mixed-effects (random and fixed) modeling approach the authors mention as their intention.

Although data are available for weekly cohorts of both wild and hatchery Snake River Spring/ Summer Chinook and Snake River steelhead, analysis is only presented for wild fish. A comparison of results for wild and hatchery fish is a desirable as a future activity.

Page 1, last paragraph, line 5: should read "Because they contain survival both at the dams *and* in the reservoirs …"

Page 2, paragraph 2: The observational unit is defined to be a single reach for a single cohort, but Table A7 1 refers to the number of observations for each segment. The distinction between reach and segment should be clarified.

Page 3, paragraph below Table A7 1: Some justification is necessary for using the same reservoir survival model for the LGR-LMN and LMN-MCN reaches and also for the MCN-JDA and JDA-BON reaches.

Page 3, last sentence: Is it reasonable to consider the tagged cohorts as a random sample? That is, justify why selection bias is unlikely.

Page 5, 6th line from bottom: Why use a method of moments estimator rather than say, REML?

Page 6, Eq. [4] – This model would probably be better for having a grand intercept, as consistent with comments elsewhere in this report and some of the COMPASS team's own findings.

Page 6, middle of the page: Why use the estimated variance-covariance matrix of the log-transformed CJS survival probability estimates, rather than the Fisher information matrix approach recommended by Franklin et al (2002)?

Page 6, 1st sentence in section on Model Fitting for Random Effects Models: The meaning of "reasonably favored" is not clear. More details on how models were selected would be illuminating.

Page 8: Providing Tables A7 2 and A7 3 (page 13) was useful for illustrating calculation details.

Page 11, Section B: Should read "Project reservoir survival from Lower Monumental to McNary Dam"

Page 14, first full paragraph: The ISAB looks forward to resolution of the choice of weights for computing the predicted mean project survival probabilities.

Pages 15-18: Presentation of Figures A7 1-4 provided a useful summary of the process and results of the uncertainty modeling presented.

Table A7 1 – It is unclear whether an "observation" is meant to indicate a single fish or a cohort, though only a cohort would make sense in this context. Perhaps "observed cohort" would be a better choice of term, assuming that is what is meant.

Appendix 8: Prospective Modeling

Appendix 8-1 – The object was to compare the likely survival of current BiOp recommendations with the base (2004) survival rates for both spring/summer Chinook and steelhead. The final version of the BiOp may be different, but the point is to illustrate how COMPASS can be used to evaluate changing management regimes. What the analyses show is that with greater flow and later transportation, Chinook returns increase but steelhead returns decrease, reflecting the fact that early transportation augments steelhead SARs but hurts Chinook SARs. Starting transportation later in the spring helps Chinook but hurts steelhead. The delivery is quite clear in general, and it is a nice example of how COMPASS can be used.

The plots in Figures A8-1 2 and A8-1 3 indicate that the model fits the data reasonably well on average, but apparently not as well as has been the case shown in earlier Appendices. Is there any particular reason for that?

In Figures A8-1 6 and A8-1 7, the "relative return rates" seem to be defined differently than in the text. They appear to be defined as the quotient (base or proposed return rate)/(base return rate). In the figures, the dashed line seem to be (base rate)/(base rate) = 1 and the solid lines the (proposed rate)/(base rate). If this is the case, then the captions for both figures have the designations of the lines reversed. Relative return rates need to be consistent. The definition should not change from a difference to a quotient in a matter of a few pages.

Appendix 8-2 – The idea was to model how survival depends on arrival times just below Bonneville, while ignoring the differences between the projects from which transported fish were barged. In full COMPASS runs, the intent is to "keep the books" in full detail, so as to be able to evaluate the efficacy of different transport points and dates. Here, the point is to illustrate.

Modeling is logistic for this Appendix, with a survival error term described as binomial, but with variation in the parameter set treated as multivariate normal for modeling purposes, which seems odd. In-river and transported fish were modeled from the same date of cohort arrival, to obtain a comparison of their SARs to LGR. Models were chosen randomly, based on ΔAICc weights.

The results are voluminous and show that year and a quadratic function of date are high-ΔAICc-weight components of the model in all cases. The models are also informative as to the likely effects of transportation. Different hydro-system scenarios are modeled, and the results suggest that COMPASS will be very useful for this sort of scenario evaluation.

It remains a bit unclear as to whether the grand mean (μ) is included in the year-effects models or not. Some clarification on this point would be good. That is to say, are they comparing

$$g_y(d) = \mu \quad \text{with} \quad g_y(d) = \mu + \psi_y \quad \text{with} \quad g_y(d) = \mu + \psi_y + \phi_y \bullet d \quad \text{etc.},$$

or are they comparing

$$g_y(d) = \mu \quad \text{with} \quad g_y(d) = \mu + \psi_y$$

and

$$g_y(d) = \mu + \phi \bullet d \quad \text{with} \quad g_y(d) = (\mu + \psi_y) + (\phi + \phi_y) \bullet d \; ,$$

or what? They can compare whichever models they desire, of course, but clarity would be served by being algebraically explicit in both the text and in the tables that follow.

Page 1, third and fourth paragraphs: - The statements "**. . .** survival of fish during barging (typically assumed to be 0.98)." and "**. . .** assumed to have 100% survival in the barge", appear to be inconsistent.

Figure A8-2 7 – What is "two-tier" meant to represent?

Section 8-3 – What are scroll case temperatures and WQM case temperatures?

In the text, the tables and figures are designated XX.1. They probably need renumbering.

Appendix 9: Sensitivity Analysis

The point here is to investigate two different sorts of model sensitivity: (a) the effects of varying levels of flow, temperature, and spill on dam survival, inriver survival, and travel time; and (b) the effects of varying transportation start date and levels of spill on SARs and the proportions of fish transported. Both sets of input variables are under management control and they have impact on the performance of salmonids, so it is important to have a clear picture of "how it all plays out" in some detail. Over 300 scenarios of each type were run for both yearling Chinook and steelhead. The results are presented graphically and are nicely summarized in the text.

In general, COMPASS runs show that hydro-system management has non-trivial impact on salmonid performance. The model is living up to its advance billing, and it will obviously prove to be a valuable management tool. We have just a few small comments.

(a) The graphics would profit from being a little larger. There is space on the page for a judicious expansion. Especially for the graphs showing that there is not much separation between scenarios, the extra size would aid visualization.

(b) All of the runs to date have utilized "generic configurations" for either all the dams or separately for the Snake and Columbia River sections of the hydro-system. That is helpful, as a means of showing "how much difference" hydro-system operation can make, but it is (understandably) oversimplified for routine practice. We found ourselves wondering how well the salmonids would do if we were to continue "doing what we do now, day by day"?

(c) What is done routinely in practice is to manage each dam separately, but presumably in an orchestrated way, attempting to optimize a plethora of competing demands for flood control, irrigation and potable water, power generation, salmonid passage, etc. We cannot help but wonder how well "routine practice," designed at least in part to provide those other deliverables, impacts the salmon, and how well salmonid survival stacks up against the best and worst scenarios shown in Appendix 9. It might be instructive to run COMPASS with the real settings for a few years.

(d) We presume, though the documentation lacks specificity on this point, that the configuration of each project can be set separately. In addition to modeling a real year, with actual settings, it should be possible to optimize salmonid performance through the hydro-system, by setting the configuration of each project separately (but in orchestrated fashion). That would require "iterative tweaking" of the settings, beginning at LGR and working progressively downriver, optimizing each project and reach in succession. The ISAB recommends that as a priority.

(e) The point here is to assess the impact of hydro-system operation on the salmonids, and to be able to do that can only be viewed as a major step forward. The COMPASS team is to be commended for their considerable achievement to date and encouraged to explore further.

(f) Sooner or later, someone is going to have to model the hydro-system for all of its outputs simultaneously, not just salmonid survival. That is an Operations Research problem on a vast scale, and the tradeoffs are going to be complicated and almost surely sobering. That one is for the future, of course, but sooner would be better than later. There are going to be some tough choices ahead for the region, and we will need additional modeling capability.

---

w:\em\ww\isab projects and reports\1 isab final reports\isab 2008-3 compassversion1-1 2june08.doc