APRIL 2024

Peer Review of the Fish and Aquatic Effects Analysis for the Long-Term Operations of the Central Valley Project and State Water Project

A report to the Delta Science Program

Prepared by

Kenneth Rose, Ph.D. (Lead Author), University of Maryland Center for Environmental Science

Henriette Jager, Ph.D. (Panel Chair), Quantus

Nancy Monsen, Ph.D., Sole Proprietorship

Zhoajun Bai, Ph.D., University of California, Davis

Emily Howe, Ph.D., The Nature Conservancy

Delta Science Program
DELTA STEWARDSHIP COUNCIL

Table of Contents

Executive Summary

The United States Bureau of Reclamation (Reclamation) reinitiated Endangered Species Act (ESA) Section 7 consultation for the Long-term Operations (LTO) of the Central Valley Project (CVP) and the State Water Project (SWP). As part of this process, Reclamation is preparing an Environmental Impact Statement (EIS) and Biological Assessment (BA). Both the EIS and BA are informed by the Fish and Aquatic Effects Analysis (Draft Effects Analysis). This report is a review of the scientific basis for the methods used in the Draft Effects Analysis to date. The Draft Effects Analysis provided to the Panel is a collection of chapters, appendices, and attachments that document how various models were used to assess how alternative management actions related to the operations of the CVP and SWP would affect a suite of life stage-specific stressors of the six ESA-listed species and their critical habitats. These species are: Sacramento River Winter-Run Chinook, Central Valley Spring-Run Chinook, Central valley Steelhead, Green Sturgeon, Southern Resident Killer Whales, Delta smelt, and Longfin smelt. As expected for a work in progress, some of the analyses in the Draft Effects Analysis were incomplete or were not included in what was provided to the Panel. The Panel determined there was sufficient documentation on enough of the models and analyses for the Panel to provide a thorough review and either answer, or at least partially answer, the Charge questions.

The review is organized so that we first discuss three global comments that are highlighted (labeled "global") for their importance, we then provide answers to the Charge questions, followed by a list of overarching comments that span more than a single model or species, and finally offer general comments on the species chapters of the BA. Finally, we present a list of comments about each individual model. Because of the overlap of methods used for the EIS and BA, the Panel often did not distinguish comments based on EIS or BA. The Panel acknowledges the impressive effort to date by Reclamation and the Delta Science Program's assistance in navigating through the many documents.

The review highlights three global comments. The first is that the Panel acknowledges the impressive effort to date by Reclamation. Reclamation has amassed many datasets and models that can provide a sound basis for assessing the effects of the proposed alternatives on the status and habitats of the listed species. The conceptual and technical challenges in doing a comprehensive analysis at this scale are enormous and Reclamation has assembled many of the ingredients for completing such an assessment. The Draft Effects Analysis reflects a consideration of the multiple pathways for impacts and a thorough and thoughtful selection of data and models. The second global comment repeats a common theme in reviews of often complicated effects analyses. A complete analysis includes a transparent and rigorously applied integration and synthesis of stressor-by-life stage results. Reclamation has made significant progress towards this task, and the Panel urges Reclamation to complete this effort. The third global comment relates to Reclamation's explicit statement on their philosophy for interpreting model predictions. Reclamation's philosophy limits reporting of the results that rely on CALSIM-3 as input to monthly because of the time step of CALSIM-3 output. The Panel questions this stipulation as many biological responses are determined by sub-monthly (e.g., daily) variability. The Panel offers several suggestions on methods for adding sub-monthly variability to CALSIM-3 output that should be explored and assessed.

The Panel answered the five Charge questions as completely as possible. The questions focused on: (1) completeness of the explanation of the stressor effects on individuals, populations, and habitat, (2) whether the analyses constituted a scientifically defensible approach, (3) how well best available science was used, (4) whether data gaps and uncertainties were identified and gaps were filled in a reasonable way, and (5) whether models representing the effects of the different alternatives were adequate.

The Panel determined that the analyses covered (i.e., considered) the major stressors (Question 1). Question 2 was more difficult to answer because the Panel deliberations moved from a checklist of effects to evaluating the use of the specific models and data to predict responses to stressors and how the results lead to conclusions at the species-level (i.e., from qualitative to quantitative). The Panel considered the general approach of coupled models, all informed from a common source (CALSIM-3), that included climate change as a solid foundation. Specific issues were: (a) use of monthly output from CALSIM-3 which misses important daily variability critical to assessing some biological effects, (b) use of statistical correlations relating annual (i.e., highly aggregated) habitat values to year-class indices that often are either weak or fall apart, especially under climate change, (c) exclusion of more-recent historical data in analyses, (d) a lack of a rigorous scheme for integrating and synthesizing the results at the population level, and (e) inappropriate use of CALSIM-3 results to estimate entrainment as part of a "coarse level assessment".

The Panel answered the third question on best available science with a partial yes. Reclamation's leveraging of existing models is an excellent starting point but, by itself, is not sufficient to guarantee that it is appropriate to apply an existing model to a new situation. The Panel was comfortable with the way that many of the models were selected, updated, and calibrated by Reclamation.

However, the Panel is unable to give an unqualified endorsement because it was difficult to judge whether all model updates needed were identified and implemented. There are some remaining questions about calibration (especially in the life cycle models) as well as concerns that the results of the model updates were reported in a way that did not allow easy evaluation by the Panel.

The Panel had several significant issues related to questions 4 and 5 that prevented us from giving comprehensive answers. First, uncertainties were not comprehensively presented or discussed in the modeling or species chapters and they were unevenly described in the modeling appendices. Second, the use of CALSIM-3 output as a driver for the other models is sound but the lack of consistent statistical interpolation to finer time steps that maintain fine-scale variability and mass continuity is problematic. Finally, the practice of using parameter values from one species to represent another to fill information gaps requires better justification.

Our answer to Question 5 about the models distinguishing the effects among alternatives also had some important issues. For example, the fact that many of the alternatives produced similar responses may indicate that stressor differences among alternatives were averaged away or underestimated by the models, or that model responses were dominated by the common driver of climate change. Many of these issues can be addressed with documentation and some new targeted analyses. In many cases, the new information should be used to better understand uncertainties and therefore better interpret model results, especially when comparing alternatives relative to baselines.

The basis for the Panel’s answers to the Charge questions were the 13 overarching comments and the reviews of the 28 appendices/attachments that document the individual model analyses. The overarching comments cover the topics of communication of complicated results, use of multiple models, model sensitivity to alternatives and sources of uncertainty, use of better interpolation schemes with CALSIM-3 to ensure proper temporal variability is used for biological effects, avoiding averaging out of differences among alternatives, ensuring historical data used in models remains relevant to today’s Bay-Delta (“Delta”) and to projected plausible future Deltas, presenting results to enable assessment of tradeoffs among species, and expanding species responses that are currently abundance-centered.

The next section of this review focuses on nine additional comments relating to the species chapters specific to the BA. The Panel was encouraged by the species chapters and Reclamation’s attempt to-date to systemically evaluate effects by stressor and life stage for each species using conceptual models as guides. The organization of the BA is useful, and the supporting appendices are very informative. However, the analysis stops short of providing an integration of the stressor effects that limit understanding of population-level outcomes. Some components of Reclamation's scheme (e.g., lethality category, exposure to the stressor as percent of years) could be more transparently or rigorously estimated. The life cycle models were treated separately from the other models and 'weight-of-evidence' was invoked by Reclamation but evidence of such analysis across life stages and stressors was not described or implemented. The Panel recognizes the effort represented in the species chapters and suggests Reclamation use this as a start and consider using more rigorous frameworks for cumulative impacts analyses and ideas and concepts from Population Viability Analysis (PVA).

The final section of the review is an appendix with comments on each modeling analysis. The Panel reviewed 28 appendices/attachments. Comments were provided for each model and many of the overarching comments also apply to each model.

1. Introduction

1.1 Background

The United States Bureau of Reclamation (Reclamation) reinitiated Endangered Species Act (ESA) Section 7 consultation for the Long-term Operations (LTO) of the Central Valley Project (CVP) and the State Water Project (SWP). This reinitiation was based on anticipated modifications to the Proposed Action that may impact ESA-listed species or designated critical habitats not analyzed in the current U.S. Fish and Wildlife Service (USFWS) and National Marine Fisheries Service (NMFS) Biological Opinions. The Fish and Aquatic Effects Analysis (Draft Effects Analysis) is a portion of the Environmental Impact Statement, a report mandated by the National Environmental Policy Act of 1969 (NEPA), that is being developed by Reclamation for the LTO of the CVP and the SWP. The analyses inform a Biological Assessment (BA), which is necessary when a Federal Agency is proposing an action that may affect a listed species under the ESA. The USFWS and NMFS will then evaluate the Biological Assessment to determine whether the Proposed Action will jeopardize listed species.

The Draft Effects Analysis prepared by Reclamation is the subject of this review and includes numerous technical appendices describing the literature, models, and tools used to evaluate the effects of different project alternatives on fish species and their critical habitats. The purpose of the Draft Effects Analysis is to: (1) systematically evaluate the potential effects and outcomes of the LTO NEPA Alternatives on specific life stages of ESA-listed species, (2) assess the population-level consequences of LTO NEPA Alternatives on ESA-listed populations, and (3) support a biological assessment for consultation with the USFWS and the NMFS.

1.2 Purpose of the Review and the Panel

The intent of this review is to evaluate the scientific basis of the analytical approach taken by Reclamation in the Draft Effects Analysis. The term "analytical approach" includes how data and models were used to assess the exposure and effects of alternative project operations on individuals and populations, and how the data and models were used to assess the impact of project operations on critical habitats.

The Delta Science Program convened a Panel of experts to perform the review. The Panel included: Henriette Jager (Panel Chair), Kenneth Rose (Lead Author), Nancy Monsen, Zhaojun Bai, and Emily Howe[1]. Members were selected to ensure that the Panel, as a whole, had sufficient expertise in the needed technical areas. The members were screened for conflicts of interest.

2. Scope and Caveats

2.1 Documents

The review is limited to the documents provided by Reclamation, although at times, the Panel needed to refer back to the original source documents, especially when existing models were used. The documents for the review were provided in three categories: Background (relevant background information to support the Panel's review), Review (main focus of the Panel's review), and Supplemental (optional relevant material to the review, but not in scope to be reviewed)[2]. The Panel was provided a total of 6,880 pages, which included chapters for the Draft BA, appendices and attachments for the Draft EIS as primary documents (5,800 pages), and Supplemental (optional) information that included roughly 1,000 pages.

---

[1] Biographies are posted at https://deltacouncil.ca.gov/delta-science-program/long-term-operations-for-the-central-valley-project-and-state-water-project-fish-and-aquatic-effects-analysis-review-Panel

[2] https://deltacouncil.ca.gov/delta-science-program/long-term-operations-for-the-central-valley-project-and-state-water-project-fish-and-aquatic-effects-analysis-review-Panel

Reclamation used an extensive set of models to generate the results and very often used the same results presented in a form intended for the EIS and in a form intended for the BA. The methods used were very similar, although there were some differences in the baseline and alternatives evaluated. Due to this overlap and because this review is focused on methods, the Panel often did not distinguish comments for the EIS versus the BA. A notable exception was the species chapters that were intended for the BA. The Panel refers to the entire collection of documents provided as the "Draft Effects Analysis."

As expected for a work in progress, some of the analyses in the Draft Effects Analysis were incomplete or were not included in the Panel's materials. We use the term "incomplete" to refer to models whose methods were documented and presented to the Panel but lacked a complete interpretation of modeling results (e.g., only tables or figures of results were presented). We use the term "not included" when nothing about the specific modeling methods and results was included in the documents provided to the Panel; often, there were placeholders in the documents provided indicating that methods and/or results would be added later. Both the models included by Reclamation and not included in the documents provided (about 30% of the models) are listed in Table 1. The models whose analyses were not included in the documents were not reviewed by the Panel. In addition, many of the models and analyses provided by Reclamation (and therefore reviewed) either were not documented in detail in the documents provided (but referred to published papers) or were briefly documented because they had been described in other documents outside of the Draft Effects Analysis.

The Panel determined that there was sufficient documentation on enough of the models and analyses for the Panel to provide a thorough review of the available material and provide caveated answers to the Charge questions along with what we hope are useful suggestions. The Panel was confident in its review of the materials provided. The completely missing models were more problematic and more limiting to the review because the Panel was unable to properly review these materials. The Panel considered that some of the missing analyses (i.e., Particle Tracking Model for entrainment) are important to assess population responses. The Panel's review focused on methods, and the documentation provided enabled the Panel to provide useful responses to the Charge questions and identify important methodological issues. The review identifies issues related to the methodology and provides suggestions for Reclamation to consider. With analyses still ongoing, this review will be useful to clarify and improve the Draft Effects Analysis. Waiting for all analyses to be completed is not effective because then there is no possibility of improving the remaining analyses before they are finalized. Therefore, the Panel has conducted its review, and hopes that readers will recognize both the benefits and the challenges of a review conducted early in the process before the Draft Effects Analysis is finalized.

2.2 Reusing Existing Models

The Charge to the Panel suggested that the Panel focus on newer models or models being used for new purposes. Models, which have a long history and have been reviewed previously, being used without much modification can be reviewed but with less emphasis. Clearly, a model that has been used and reviewed before, especially for similar questions, has earned a higher level of confidence than a new model. However, it is still important to ensure that the historical data used in the original model development or previous applications remain relevant to the new questions being asked. It is imperative that the ecological state the model was calibrated to is still representative of the present state of the ecosystem. The Panel determined that most models, even existing peer-reviewed models with no structural modifications, required some level of review. Existing peer-reviewed models were tailored (calibrated and then used) for this specific application to ensure they remained realistic for the examination of new states of the system associated with the different alternatives. The Panel therefore did not simply accept a model analysis because the model had been previously reviewed.

2.3 Completeness of Review

The review is organized for the reader so that we first discuss three global comments that are highlighted (labeled “global”) for their importance, we then provide answers to the Charge questions, followed by a list of overarching comments that span more than a single model or species, and finally offer general comments on the species chapters of the BA. Finally, we present in the Appendix a listing of comments about each individual model. Individual model reviews were sufficient for us to know what was being done or likely to be done, minus the details. Similarly, when we moved up a level from individual model reviews to overarching comments, the Panel had sufficient information to identify major methodological and technical issues as overarching comments. Finally, in answering the Charge questions, the large number and diversity of models and analyses prevented the Panel from providing comprehensive answers to the Charge questions. We were able to provide partial answers and identify topics and issues to look for as the analyses are finalized (including suggestions based on our overarching comments and model reviews) which we hope will be useful in completing the analyses and documentation.

Finally, there were several important aspects of the Draft Effects Analysis that we did not review. We did not evaluate the alternatives for their feasibility or their performance. Often, model analysis results were presented for the Biological Assessment (BA) and then reformulated or reformatted for use in the Environmental Impact Statement (EIS). We accepted these as described and examined the methods used to assess the biological effects of project operations without considering the magnitude of effects. Our review focused on how well the models were positioned to predict the biological effects of the project alternatives. We also did not examine how the environmental baseline was formulated, nor the scientific basis for formulating versions of the Exploratory Simulations and other simulations used as baselines or reference conditions. Finally, we did not review Reclamation's interpretation of the conceptual life cycle models used in the Draft Effects Analysis.

3. The Panel's Charge

3.1 Overall Charge

The Charge to the Panel included this overall statement of the objective of the review:

> *The intent of the review is to evaluate the analytical approach taken by the United States Bureau of Reclamation (Reclamation) to assess how the long-term operations (LTO) of the Central Valley Project (CVP) and State Water Project (SWP) affect the aquatic environment and the exposure, response, and risk to select Endangered Species Act (ESA)-listed species (individuals and populations). In addition, the review will assess whether quantitative and qualitative methods and risk assessment tools are used appropriately.*

3.2 Specific Charge Questions

The Panel was asked to answer five questions as they reviewed the various models, datasets, and analyses:

1. *To what extent do the draft analyses explain the exposure, response, and risk from project operations for individuals and populations of the ESA-listed species, and physical and biological features of designated critical habitats under the approaches described by the alternatives?*

2. *To what extent do the draft analyses provide a scientifically defensible approach for evaluating effects on listed species and their designated critical habitats throughout the action area for different alternatives?*
3. *How well do the draft analyses use the best available scientific information in their analyses and findings?*
4. *How well do the draft analyses address data gaps and uncertainties? Are assumptions and methodologies suitable for addressing identified data gaps?*
5. *Of the key operations modeled, how adequate are the models for representing the effects of the different alternatives on aquatic listed species and their habitat?*

It is important that all Panel members use consistent interpretation and view of the questions, and that the Panel's interpretation of the questions is clear to readers. The Panel mutually agreed upon the following interpretation of the questions. An important aspect of this is ensuring the differences among the questions are clear.

Interpretation of Question 1: The question focuses on the term "**explain**." The Panel interpreted this question as a check on the completeness of the analyses in terms of whether the major routes of exposure and the resulting risks and responses of individuals and populations, as well as effects on key habitats, were addressed by the analyses. This question is conceptual and does not necessarily involve detailed reviews of specific models.

Interpretation of Question 2: The question focuses on the term "**scientifically defensible approach**." The Panel interpreted this question as whether the many datasets, tools, and models used, when viewed together, conform to scientific standards and are methods that are generally accepted by the scientific community as valid and reliable. In addition, this question (as opposed to Question 1 which focuses on individual stressor effects) includes how the effects can support statements at the species level.

Interpretation of Question 3: The question focuses on the term "**best available scientific information**[3]" with "analyses" and "findings." The Panel interpreted this question as involving a detailed examination of the methods, such as specific

---

[3] The Panel used guidance on Best Available Science from the Delta Science Plan, Appendix C. The Delta Plan (ca.gov)

datasets and specific models. Evaluating this, in some cases, required a fair amount of sleuthing to determine whether the latest models were being used.

Interpretation of Question 4: The question focuses on **data gaps and uncertainties**. The Panel interpreted this question as how well critical uncertainties were identified throughout each of the analyses (individual models) and how were they were combined across analyses to result in an understanding of the confidence appropriate with results, especially when comparing among alternatives. We also addressed whether data gaps were imputed or "filled in" using reasonable assumptions.

Interpretation of Question 5: The question focuses on whether the models are "**adequate**" for comparing alternatives. The Panel interpreted this question as: With what confidence can the differences in predictions of the Proposed Action and other alternatives (expressed as relative to a baseline) be interpreted as true differences (as expected to occur in nature) versus indistinguishable because of the uncertainties? Answering the adequate issue is the culmination of the other questions: Were all major effects included (question 1) using defensible (Question 2) and best available scientific information (Question 3)? Were uncertainties properly considered (Question 4)?

4. Review Process

The review began in November 2023, with the delivery of the documents provided by Reclamation. There was a kickoff meeting with the Delta Science Program on November 28, a meeting with Reclamation and the Delta Science Program on December 4, and regular Panel-only meetings roughly every two weeks during January and February. The initial reviews of each model were done by 1-2 Panel members. The reviews were then collated and reviewed by the entire Panel. Meetings were used for group discussion of overarching comments, answers to the Charge questions, and to collect feedback on specific issues and comments about individual models. This review report was unanimously approved for submission by all members of the Panel.

The Panel's comments and observations are organized into five sections:

a) Global comments,

b) Responses to charge questions,

c) Overarching comments,

d) General comments on species chapters, and
e) an Appendix of specific comments on individual models.

The Panel considered many of the documents provided by Reclamation in their deliberations, including those outside of the specific modeling documents (i.e., reports or papers that described earlier modeling analyses) to inform our reviews of the models, as the basis for answering the Charge questions, for formulating the global and overarching comments, and as part of reviewing the general approach of the species chapters. Throughout the review, the Panel uses “stressor” to represent the variables and factors (e.g., temperature, flows) altered by the alternatives, and “biological effects” as the responses of individual fish (e.g., mortality, growth, movement) and, in some cases, responses at the population level (e.g., population growth rate, abundance).

5. Global Comments

5.1 Global Comment 1: Impressive effort to date

The Panel wants to acknowledge the efforts of Reclamation to date. Reclamation has amassed many datasets and models that can provide a sound basis for assessing the effects of the proposed alternatives on the status and habitats of the listed species. The conceptual and technical challenges in completing a comprehensive analysis at this scale are enormous and Reclamation has assembled the ingredients for doing such an assessment. The Draft Effects Analysis reflects a consideration of the multiple pathways for impacts and a thorough and thoughtful selection of data and models. Because the Panel did not have access to the complete analysis used by Reclamation, especially the omitted models and a clear synthesis plan, the Panel cannot, at this time, fully endorse the approach and methods. However, the Panel determined that with continued diligence, incorporation of the comments in their modeling analyses, and a strong synthesis addition, Reclamation’s analyses will provide a sound scientific basis for assessing alternative impacts on the listed species. The Panel acknowledges the significant and thoughtful efforts of Reclamation in their progress to date.

5.2 Global Comment 2: The challenges of integration can and must be overcome

The documents provided to the Panel reflected many analyses using data and models providing a promising start to the synthesis of these results (i.e., species chapters). A major aspect of the synthesis (cumulative effects of multiple stressors) by Reclamation is the grouping of analysis results by species. This is a logical way to synthesize results since Biological Opinion decisions are based on each of the listed species.

The species chapters are an impressive attempt to organize the multi-stressor effects from a variety of analyses together to assess alternatives at the species level. We discuss the species chapters in section 8 of this review. The Panel notes that some of the steps used in the species chapters can be made more rigorous (e.g., methods for estimating exposure) and that the approach did not lead to an integration of modeling results to reveal species-level effects. The analyses stopped with a systematic collation of the results of stressors by life stage. There are multiple sources of uncertainty associated with each model and analysis that determine the level of confidence appropriate for the final predictions. When confidence is not specified for different analysis results, the unstated default is that all results are treated the same and thus all have the same level of confidence. This is very unlikely to be the case.

The Panel considers that when the different modeling efforts, including those not included in this review, are completed, the major stressor effects associated with Proposed Action or Alternatives will have been addressed. How these results will be rigorously, using quantitative and qualitative information, and transparently combined across analyses beyond the narrative and sequential listing of results in the species chapters was not clear to the Panel.

Reclamation repeatably mentions weight-of-evidence as a synthesis tool, which the Panel emphasizes is difficult to rigorously implement without some idea of the uncertainty of the results from the different analyses and a scheme for combining results (see Hope and Clarkson 2014 and Section 8.5 Comment 19). The use of life cycle models (LCMs) also provides a rigorous, scientific basis for synthesis, although none presented in the analyses addressed all effects for a species at relevant time scales. When available, the LCMs differed across species, and several species did not have LCMs. This complicates the effects analysis for a species because one must then combine effects covered by the models with effects estimated by using other approaches. The grounding of the analyses in conceptual models like the Salmon and Sturgeon Assessment of Indicators by Lifestage (SAIL) model and the Management, Analysis and Synthesis Team (MAST) model for Delta smelt is also a way of synthesizing information by species. The Panel suggests that Reclamation go further in their synthesis and continue advancing the use of LCMs across all listed species and expanding the effects that are represented.

5.3 Global Comment 3: Reclamation's philosophy on how to interpret model predictions

The Panel appreciates the clear general statement by Reclamation about how to interpret model results.

> *The models are not predictive models of project operations and results cannot be considered as absolute with a quantifiable confidence interval. The model results are only useful in a comparative analysis and can only serve as an indicator of conditions. Due to the assumptions involved in the input data sets and model logic, care must be taken to select the most appropriate time step for the reporting of model results. Sub-monthly (e.g., weekly, or daily) reporting of raw model results is not consistent with how the models were developed, and results should be presented on a monthly or more aggregated basis.*
>
> *Absolute differences computed at a point in time between model results from an alternative and a baseline to evaluate impacts is an inappropriate use of model results (e.g., computing differences between the results from a baseline and an alternative for a particular month and year within the period of record of simulation)* [Draft EIS, Appendix F Modeling, Modeling main report, document page F-8 (pdf page 12)].

The Panel considers this stipulation problematic because the temporal scale of some of the physical modeling is too coarse to properly evaluate some of the biological effects of stressors. The central challenge is to relate the effects of the alternatives on listed species, mediated by numerous stressors. The choice of physical model should reflect the purpose of the modeling (i.e., to assess biological effects). Because the reporting time step in some of the physical models is not appropriate for use in evaluating some of the key stressors for their biological effects on organisms, the entire modeling chain (from flow to stressor to population) is called into question. Furthermore, even if one accepts the philosophy expressed in italics above, it was not clear to the Panel that the philosophy was applied consistently across all models. There may be good reasons for this, but justification should be provided.

A major issue, and one that has been discussed in many other reviews, is the use of CALSIM-3 monthly output to assess biological effects on fish that operate on daily or finer time scales and depend on extremes, durations of exposure, and variation, rather than on a smoothed monthly average. The Panel recognizes that

Reclamation must and should use CALSIM-3, but there are accepted ways to add daily and finer variability to monthly averages that preserve the full distribution of sub-monthly variation (not simply by interpolating between monthly means or using the same value for all days in a month) that must be incorporated. Furthermore, it is well-known that variability in flow [and temperature] can have an important influence on cold-water fishes (Freeman et al. 2022; Steel et al. 2012). Otherwise, the models will not be able to represent variation at the temporal scale important to represent exposure and effects on fishes.

The Panel recommends conducting a separate analysis based on field data to evaluate the best statistical method for adding variability as part of the post-processing of CALSIM-3 output in a way that preserves the full distribution. We recommend that daily and finer time steps drivers produced in this way be used consistently by all modeling analyses. In addition, constraints, such as minimum flows and ramping rates, can then be incorporated into the modeling analyses. Another approach to providing daily and finer variability to certain variables (e.g., salinity, temperature), which was used in some analyses in the Draft Effects Analysis but summarized on a monthly scale, is to use the DSM2 and HEC5Q models that predict at daily and sub-daily time scales. This model-based option should be considered whenever possible and could be added to the empirical approach based on variability patterns in monitoring data to develop a generalized analysis for adding variability to drivers (from CALSIM-3 and other sources) used across models.

6. Answers to Charge Questions

Question 1: To what extent do the draft analyses explain the exposure, response, and risk from project operations for individuals and populations of the ESA-listed species, and physical and biological features of designated critical habitats under the approaches described by the alternatives?

In general, the analyses appear to cover the major effects anticipated from the alternatives. However, not all are explicitly included in analyses of biological effects and they are unevenly represented across species due to data and model limitations. For example, some potentially important risks to Green sturgeon (e.g., stranding of spawners) and effects of temperature and flows on migration stimuli are not represented and other effects (e.g., entrainment) could be modeled in a more species-specific manner. The Particle-tracking modeling methodology used to estimate entrainment at the South Delta pump facilities was not yet available and not included in our review documents. Similarly, although dewatering and stranding are listed as stressors in the American River, these were not modeled for spring-run Chinook salmon. The Central Valley Project Improvement Act (CVPIA) Decision Support/life cycle model used in the Critical Habitat analysis (Chapter 5) uses a constant egg-fry survival, and therefore does not consider the effects of Shasta operations on early life stages or spatiotemporal variation in entrainment. The CVPIA life cycle model used has other strengths as a tool for integrating across life stages and did show differences among alternatives at the population level. However, the juvenile rearing life stage during winter is listed as a significant concern, but it is not well-represented in the CVPIA model. The Panel noted other instances of potentially important missing stressors. For example, the potential exposure of larvae to high salinities within the Delta has not been quantified (Rodgers et al. 2019). If spawning is occurring farther downstream than it did historically, or if premature outmigration of salmon that have not completed smoltification occurs, this can cause mortality of early life stages of anadromous fishes (Giroux and Schleck 2020). This would then be further aggravated by sea-level rise.

The Panel suggests that Reclamation better explain the role of the Conservation Measures (and other operational actions) that are used to address (mitigate) stressor effects (Appendix H). These include ramping rates, minimum instream flows, spring pulse flows, and aspects of temperature management (e.g., targets, maximums). The Panel was unclear on the extent of mitigation by the Conservation Measures for these stressor effects and how they were simulated in the Draft Effects Analysis. We define it as partial inclusion if a stressor was included in a model but at a coarse level (e.g., monthly), whereas daily variation was assumed constrained by Conservation Measures. Thus, it seems that the effects analysis only included the coarse-level effects of the stressor (e.g., monthly resolution) and did not reflect the effects of Conservation Measures. Excluding mitigation constraints on stressors from the modeling analyses may also affect the models' capability to distinguish among alternatives (Question 5). It is likely the stressor exposure and effects realized in nature might further differ from the Draft Effects Analysis predictions because the Conservation Measures (not included in the modeling) will likely have resulted in differences in the stressors among alternatives.

The Panel identified a subset of stressor effects that could be addressed further. The use of conceptual models like SAIL and MAST are useful to consistently evaluate stressors and determine which subset of these to focus on when predicting biological effects on species. Based on interpretation of the question as a checklist, and assuming the missing models are completed as anticipated, we consider that most of the major effects have been considered. Taken together, the Panel determined that the major effects and stressors were at least mentioned by Reclamation and that the major effects were analyzed. There are some exceptions and the Panel identified some of these in the reviews of individual models.

Although the description of the differences in management actions among alternatives is clear, an analogous presentation based on the biological effects assessed across alternatives would be helpful. It may be that there is great overlap in effects across alternatives, but they differ in expected magnitude; or perhaps a few of the alternatives include effects that others do not.

The area of “risks” included in the Charge question may be lacking sufficient coverage. Effects (together lead to risks) were very much treated separately with a “one-at-a-time” approach, whereas the life-cycle models (LCMs) combined multiple risks but did not capture all of the stressors (e.g., egg-fry survival of salmonids). Risks are more often viewed on a species-level basis and are the culmination of all of the biological effects. The possibility of effects interacting with each other to generate non-linear effects (a multi-stressor framework) was not discussed. Also, there was much focus on population-level effects and little discussion of the Viable Salmonid Population (or similar) framework (McElhaney et al. 2000) that accounts, not only for abundance and population growth rate (the foci of this effects analysis) but also for population spatial structure and life-history diversity, including the degree of reliance on hatcheries. We appreciate that hatchery ‘sub-populations’ were added to the life cycle models, but little was done to explore sensitivity to these new additions. A discussion of risks may therefore be underestimated with a focus on abundance and population growth rates and without modeling populations in adjacent tributaries.

Question 2: To what extent do the draft analyses provide a scientifically defensible approach for evaluating effects on listed species and their designated critical habitats throughout the action area for different alternatives?

Question 2 is more difficult to answer than Question 1 because it moves the discussion from evaluating a checklist of effects (Question 1) to evaluating the use of specific models and data to predict the magnitude of the biological effects of stressors and how they lead to species-level conclusions. Two main areas of concern are: (a) integration of stressor effects to the population level, and (b) coupling physical and ecological models.

The analyses currently lack a rigorous scheme for integrating and synthesizing the stressor-by-life stage results. This is an issue of major importance and was discussed as a Global Comment above and is revisited in Section 8 which reviews the species chapters. Synthesis of results is important for us to evaluate the scientific defensibility of the Draft Effects Analysis.

The general approach of using coupled physical and ecological models is appropriate and scientifically sound. Further, requiring models to share the same time series of physical habitat output from CALSIM-3 and other physical models provides consistency across analyses. One challenge involves coupling physical models that operate on one temporal and spatial resolution with another model that operates on different scales. This is very common with coupled models and is solvable. The key is in the details.

Whereas many of the approaches used to couple physical and ecological effects models in the Draft Effects Analysis are reasonable, the Panel noted four areas where the details could create some problems in physical modeling that propagate through the modeling of biological effects, which could result in lowered confidence for predictions of biological effects. These areas were: (a) monthly output of CALSIM-3, (b) statistical correlations relating annual values of habitat to year-class indices, (c) inconsistent and/or outdated historical data in the analyses, and (d) inappropriate use of CALSIM-3 results as part of a “coarse level assessment” to represent Delta hydrodynamics for predicting entrainment at the South Delta export pump facilities.

The first area was how the monthly output from CALSIM-3 was used (as input or as boundary conditions) with models that operated on shorter time steps. Many models use CALSIM-3 output as input. Yet, many effects should be represented based on daily or finer time scales [e.g., temperature effects on egg survival as described in Martin et al. (2017) and Anderson et al. (2021)]. In this case, the model described in Attachment L.3 used daily values whereas the two life-cycle models (CVPIA Winter-run and Spring-run) and the temperature-dependent mortality (TDM) indicators used monthly values of temperature. Even when CALSIM-3 output was used for models, the Panel noted that interpolation was not done uniformly across models and the core interpolation that was done still had issues with mass-balance and unrealistic extrapolations. The output from CALSIM-3 is fundamental to the

Draft Effects Analysis as it is the driver for modeling a suite of biological effects among alternatives. This issue appears multiple times in this review, indicating its high importance to the Panel.

The second area was the use of statistical correlation models based on annual values that relate habitat to year-class indices. This correlation-based approach has a long history in fisheries. Such aggregated correlations have many well-known problems and limitations (e.g., Walters and Collie 1988, Bason 1999, Drinkwater and Myers 1987) and should either not be used or used with great caution. In some cases, such correlations may provide the best available models, but they may not be good enough to predict responses to new environmental and ecological conditions, such as those generated under the alternatives in the Draft Effects Analysis. Importantly, the aggregation of information to annual values does not allow for the resolution of stressors that occur in specific places at certain times. Therefore, the Panel questions the use of these models for comparison of alternatives.

The third area was the use of inconsistent or outdated historical environmental drivers and physical data in analyses. The use of existing models has advantages but also can limit the power of analyses because one wants to use the model in its original and previously reviewed form. Examples are the habitat analyses in the Draft Effects Analysis that used previously developed suitability functions to derive Weighted Usable Area (WUA) versus flow relationships. In some cases, the data used ended in 2009 and the issue that arises is whether the system has changed sufficiently in recent decades, which would question the relevance of the old suitability functions and relationships to the present-day or futures under climate change, and further, whether the new conditions created under the alternatives are within the domain of conditions where the relationships comfortably apply.

The fourth area was the Panel’s concern on whether the CALSIM-3 model represents the necessary details of Delta hydrodynamics to the level that entrainment at the South Delta pump facilities can be estimated, even a coarse level assessment. A hydrodynamic model of the Delta must be utilized to assess entrainment risk. The Draft Effects Analysis’ direct use of CALSIM-3 model results, rather than a more refined approach that uses a hydrodynamic model that includes Delta bathymetric features and is informed by CALSIM-3 for flow boundary conditions, creates high uncertainty and predictions with highly questionable use.

Question 3: How well do the draft analyses use the best available scientific information in their analyses and findings?

The advantage of using existing models and historical data to develop and test the models is that the analysis starts with a higher level of confidence. Most models reviewed here have accumulated a history of model evaluation and review, which is not true of newer models. The fact that these models exist already directly relates to the "available" part of "best available scientific information." Such leveraging of existing models is therefore an excellent starting point but, by itself, is not sufficient to guarantee that it is appropriate to apply an existing model to a new situation.

Models must be evaluated for their utility and appropriateness for each new question being answered. It appears that Reclamation did some of this, as some of the existing models were examined and updated. The Panel was concerned that Reclamation, in their evaluation, may have put too much weight on the simple fact that a model was previously reviewed. It is not the simple act of a review of a model that results in higher confidence, but also how the model is applied, how the review was conducted, for what purpose, what questions were asked at the time, what information was available, etc. The context of the review is very important to judge the generality of the model, and the likelihood the model can easily be used (and what updating is needed if any) for this specific Draft Effects Analysis.

It is not possible to give a broad assessment that applies to all of the models reviewed. The Panel was comfortable with the way that many of the models were selected, updated, and calibrated by Reclamation. In most cases, the Panel agrees that models chosen by Reclamation represent the "best available" options. However, the Panel is unable to give an unqualified endorsement because it was difficult to judge whether all of the necessary model updates were identified and implemented. This is because there are some questions about calibration (especially in life cycle models), and because the results of model updating were not consistently reported to allow easy evaluation by the Panel. In addition, the issues of incomplete documentation and missing models prevented a more definitive assessment by the Panel.

There are examples in the Panel's reviews of individual models (Appendix) that raise the question of whether the representation of processes in the individual models reflected the "best available" options. For example, the Panel noted that there seemed to be room for improvement in the representation of dewatering, predation risk, and thermal risks (the latter is addressed under Question 5). The Panel raised these to provide Reclamation with specific cases to further explore the available science.

However, models are approximations of the real world, and the job of the analyst is to decide which processes and stressors to include and at what level of detail in order to address the questions; in this case, quantify the biological effects on species associated with different project alternatives. The consequences of the results of the simplifications done as part of the modeling are typically not known. In some cases, separate analyses are done specifically to address how including a stressor or how more detailed representations of stressors would affect the predictions. It was not always clear to the Panel how Reclamation determined which models and what aspects needed to be updated. The rationale for updating the temperature modeling was an example of a model update with a well-documented line of reasoning.

Question 4: How well do the draft analyses address data gaps and uncertainties? Are assumptions and methodologies suitable for addressing identified data gaps?

The identification of data gaps is sprinkled throughout the individual model analyses making it difficult for the Panel to provide a single answer. It does appear that when needed, reasonable methods for addressing data gaps were used with several important exceptions that the Panel thought needed more attention (detailed below).

Another major issue related to data gaps and filling them is how the monthly output from CALSIM-3 (flows) was interpolated to estimate daily values that were used in multiple other analyses. In some cases, the interpolation was done specifically to a model using CALSIM-3 output so there are multiple versions of what should be the same interpolation, and there are other examples when models did not use CALSIM-3 output but used field data for interpolation (e.g., Delta smelt LCME model). How consistent are predicted values with those obtained from different methods and when these relationships are determined do they include climate change in the same way as others?

Could the same interpolated values or the values from DSM2 and HEC-5Q be used, when appropriate, throughout the Draft Effects Analysis? This downscaling is critical to many analyses and the Panel identified the benefits of using a single set of downscaled values across models, and the technical issues related to the interpolation method as a high priority for Reclamation to better document and quantify the level and types of errors involved.

Data gaps in species-specific parameters are potentially a significant problem for some of the analyses presented. Such gaps were filled by "borrowing" information from one species for use in modeling another species. When this is done, more explanation is needed, along with a description of the likely direction of error generated by borrowing and how the borrowing would affect the results. For the life cycle models, one important source of uncertainty is the assignment of Chinook salmon race in salvage data used in calibration. However, confidence is higher for the winter run than for the spring and fall runs. Appendix D and Heublein et al. (2017) identified some potential gaps that were filled by substituting information from other species: (a) lack of quantitative environmental relationships for Green sturgeon recruitment and (b) lack of sturgeon criteria for estimating entrainment at diversion intakes and screens (or do they exist but were not used?). The assessment of entrainment risk is based on parameters for salmon. However, a lab study found the risk of entrainment of green sturgeon to be much higher (4.2-22.3%) than that for juvenile Chinook salmon (0.3-2.3%) (Mussen et al. 2014). It is unclear whether screening criteria are equally protective of both species, but research has been done that seems sufficient to suggest refinement of the criteria for Green sturgeon is possible (see Poletto et al. 2014a, 2014b, 2018).

Using the parameters for winter-run Chinook salmon as surrogates for other salmon species is arguably more defensible than using salmon to represent sturgeon. However, there are concerns about using hatchery fall Chinook coded-wire tagged (CWT) data to assess winter-run travel and survival (e.g., in STARS, Appendix I - Attachment I.5). More justification is needed to establish the degree of overlap in migration timing between the two runs. In addition, hatchery fish tend to return to spawn and exit the system earlier than wild fish.

Analogous to species borrowing, transferring information among watersheds to fill data gaps and the use of historical time periods that do not extend to the present are two additional sources of uncertainty. It is inevitable that data is used from nearby and that the time period used in previous model applications be maintained; this is good practice in modeling if done carefully, i.e., after comparing the historical data used previously to the new conditions being examined. The use of earlier historical time periods that end prior to present-day is of some concern because of the changes in the ecosystem in recent years. The use of a common

CALSIM-3 set of outputs as inputs to models can reduce this uncertainty when alternatives are compared. However, it is still unknown how well a model developed and evaluated with historical data (e.g., WUA) reflects current relationships between biological effects and environmental variables. Furthermore, it is unclear how well these relationships will hold under future climate conditions.

The treatment of uncertainties (and certainties) was less well developed than the filling of data gaps. The Panel notes that the use of box plots to show the full distribution of predicted values was very helpful. The reporting of error bars should be done with caution because distributions of values are often skewed, and error bars only represent a subset of sources of variability. Any reporting of statistical significance should be accompanied by information on the effect size involved (Smith 2020) and statistical significance should be viewed in the context of effect size. For example, statistical regression models to link flows to zooplankton response were built between specific zooplankton species and life stages and the position of X2 (Appendix J, Attachment J.3 – Zooplankton-Delta Outflow Analysis). Reclamation then only considered species or life stages with statistically significant relationships to the position of X2. This resulted in many key prey species (or prey life stages) for Delta smelt being removed from the analysis, such that it is likely that the remaining collection of species was a questionable indicator of food availability for Delta smelt.

There was no consistent approach used for reporting uncertainties within each analysis and there was no overarching presentation of uncertainties to allow a “big picture” view across analyses. Attempting to quantitatively track uncertainties through the many coupled analyses is not feasible; however, some discussion of uncertainties is appropriate. The Panel recommends that Reclamation further address the issue of uncertainty as part of the Biological Assessment species chapters because it plays an important role in comparing alternatives and combining and interpreting results at the species level.

Equally important to uncertainties is a discussion of the certainties – what aspects of the analysis does Reclamation have relatively high confidence in? Partitioning years into water-year types is an excellent approach to address hydrological variability. Incorporating climate change into all CALSIM-3 simulations is a forward-looking way to deal with the uncertainty of the performance of alternatives under plausible futures. It is also well-aligned with the decision framework by which flows are managed. The use of multiple models (two LCMs for

Delta smelt and winter-run Chinook salmon) seems to be the set-up for dealing with uncertainty in representing growth, mortality, and recruitment of Delta smelt in order to predict population dynamics. However, there was no plan or strategy presented on how to combine the predictions from the two LCMs. Chapter 9 on Delta smelt stopped before the results from both models were combined. The use of multiple models can be an effective approach for dealing with structural uncertainty, but it requires careful set-up and strategic interpretation of predictions. It is not as simple as simply treating the two models as completely independent.

Reclamation used existing models and so opted to not provide complete documentation, including parameter values and their sources. This is reasonable if information about uncertainties is provided – perhaps in a separate document. At a minimum, there should be a listing for each model of where parameters came from in terms of species, sub-region, and time period. It is unclear whether these three ways of borrowing parameters could potentially result in a poor representation of the seasonal and spatial effects of project operations. The major instances of "borrowing" and "lack of model identifiability" (discussed next) in the Draft Effects Analysis should be summarized and justified by demonstrating that the uncertainty in model predictions is low enough to quantify important (ecologically meaningful) differences among the alternatives.

The Panel appreciates that the Draft Effects Analysis includes several examples where uncertainties were addressed within the modeling framework. Generally, the life cycle models were stochastic and thus produced a distribution of population-level predictions appropriate for making decisions about species conservation. First, the CVPIA life cycle model is calibrated against spawners below Keswick Dam, juveniles collected at Red Bluff diversion dam, juveniles collected in Knights Landing catches and rotary screw traps, and Chipps Island abundance. Nevertheless, the model has more parameters than can be uniquely specified (a model-identification issue). This can lead to selective choice of fitted parameter values that, for example, results in too much of the variability in survival being explained by stressors in the ocean or Delta. One approach used to address this in the analyses presented was the fixing or constraining of biological parameters with other information. A general approach is to evaluate the necessary complexity of a model, looking for the optimal balance between mechanistic detail and available data (Collie et al. 2016). Determining the optimal model structure is a long-standing issue in simulation modeling and best addressed on a site- and case-specific basis.

These are reasonable options. However, two potential issues are: (a) arbitrary parameter decisions can produce a good fit for the wrong reason, leading to a focus on the wrong stressors, and (b) both approaches can reduce the ability of the model to distinguish among alternatives when the differences among alternatives are due to temporal or spatial aspects of the stressors.

Alternatively, Bayesian multi-parameter modeling (see Piou et al. 2009 and Jager 2013) can address the problem of selective parameterization by avoiding the need to fix parameters or to arbitrarily choose one set that fits. Constraints can be included through the prior distribution, and the Bayesian approach weights parameter combinations based on fit to historical data and uses the resulting (posterior) combinations of parameters to project forward. Thus, multiple parameter combinations are incorporated into model predictions, which addresses parameter uncertainty. Representing variability is important, in part, because population viability depends on the variation in future projections of population trends, as well as the trends themselves.

Question 5: Of the key operations modeled, how adequate are the models for representing the effects of the different alternatives on aquatic listed species and their habitat?

At a conceptual level, the models appear adequate to distinguish among alternatives. The Panel was unable to affirm how well the models were capable of distinguishing the effects of the alternatives on each of the listed species at the qualitative level. We addressed this in our response to Question (1) about whether major effects were discussed and our response here to Question 5 rests on our caveated responses to Question (2) about a scientifically defensible approach and Question (3) about using best available science. We recognize that scaling up from the effects of local and intermittent stressors to population-level outcomes is a significant challenge, but an important one for the analyses to address.

Several general results of the predictions caught the attention of the Panel. Differences and similarities in how the models responded to water operations of the alternatives in releases (flows) and water temperatures were nicely presented for the Proposed Action prepared for the BA, but not as clearly presented for all of the alternatives used in the EIS in Appendix B. First, how the different alternatives generated different model drivers across all alternatives was scattered and unevenly presented in the individual modeling documents. We refer not simply to plots of CALSIM-3 outputs, but also how CALSIM-3 outputs (or other sources of input data to models) were actually used as input to the different models and at what resolution.

Second, the Panel noted that in many analyses there seemed to be small differences predicted among multiple alternatives. This was clearly illustrated with Appendix B for the Proposed Action alternative and also when comparing other alternatives for the EIS when the information was included in the modeling documents. Although this may be accurate, it also suggests that the models may not capture important differences among alternatives, perhaps because of the ways in which processes are represented or because the differences in model inputs (or outputs) were essentially averaged away by using a coarse time step for presenting model results.

Another example is the apparent insensitivity of the currently calibrated CVPIA model to flow alternatives, reflecting an insensitivity to freshwater processes. In one life cycle model, this is caused by calibration leading to very high juvenile survival during river migration. In another, it is caused by using a constant egg-to-fry survival. Therefore, it is not really possible for Draft Effects Analysis to assess future effects of Shasta operations that affect egg-to-fry survival of the Chinook salmon runs or any influences during migration. By extension, this makes it a less-than-optimal choice for evaluating the relative importance of river versus Delta versus ocean phases. Looking across species, it is possible that effects are missing or underestimated which would provide further contrast among alternatives. It is also possible that the small differences among alternatives are accurate.

Alternatives may also differ in how they represent extreme events (e.g., drought, flooding); how well these are included to represent plausible futures, including how their effects are represented, then becomes important for distinguishing among alternatives. The ability to represent future changes in the frequency of (and autocorrelation in) drought years is of high interest and could be highlighted by a separate analysis. Presenting model results by water-year type is helpful for assessing the impact of rare but influential year-types and is consistent with how water is managed. However, it does not address the risk of successive poor (presumably extreme) years. How well are the biological effects of extreme events represented, now and under future climate? Because the potential amplification of effects by successive drought years (and flood years), the Panel suggests that the relevant models be evaluated for their realism in how they represent extreme events and that such conditions be analyzed as part of the Draft Effects Analysis.

Third, the CVPIA SIT Decision Support Model is a stage-structured life cycle model. It is unclear that it is designed to assess effects of the Proposed Action relative to baseline operations due to the coarse resolution and the way the effects of management actions are represented (Attachment 2 of Appendix O). Although it uses CALSIM-3 output as input like many of the other models, it appears that the actual effects of diversions on survival are specified rather than simulated as an outcome of spatially explicit hydrodynamics and fish movements (Peterson and Duarte 2020). Thus, some of the differences among alternatives were specified rather than emerging from mechanistic representation of interactions between fish and local stressors. Thus, differences among alternatives depended on modeling decisions and did not emerge from processes represented in the model.

Fourth, both life cycle models for Delta smelt use a statistics-based approach based on life stages defined by the different surveys and transfer functions between survey-defined life stages fitted to environmental covariates. A great deal of correlation-based analyses was needed to convert CALSIM-3 output to appropriate covariates (e.g., flows to seasonal food index) in these statistical life cycle models. This likely caused averaging out of differences among alternatives, so many alternatives generated similar overall responses. This could be real or an artifact of the averaging process. The issue of monthly CALSIM-3 output enters the discussion since it can miss important differences in daily values of flow among alternatives.

Fifth, representation of stressors that might distinguish alternatives was inconsistent across the species. This is partly a reflection of using existing models that were developed for different purposes and so they reflect those decisions on what effects are explicitly included. For example, the modeling approach and representation of effects differed substantially among the salmon and Delta smelt life cycle models. Although some of these differences reflect biological differences, other differences were due to decisions of each model developer. Furthermore, no life cycle modeling was conducted for Green sturgeon, Steelhead trout, and Longfin smelt. With such uneven coverage, the representation of effects is inherently different across species. This is a manageable situation if the results are viewed and interpreted carefully and against this backdrop of uneven coverage. A listing of all biological effects by stressor and life-stage (the species chapters) is a good start, but additional synthesis across life stages to the population level and cross-species interpretation to understand trade-offs is also needed.

Finally, perhaps the imposition of climate change scenarios as a common driver of model predictions dominates the responses. It is possible this is realistic, and that this common driver is strong enough to mask the smaller differences among alternatives within a species and differences among species. This can be easily assessed with some simulations with and without climate change included.

7. Overarching comments

7.1 Comment 1: Consider ways to improve the communication of complicated results.

The documents reviewed by the Panel were considered "in progress" for some of the analyses and modeling results. As expected, there were some results presented as a series of many tables. Such results are necessary to include but are not sufficient; further graphical presentations can greatly help clear communication of the results to readers. The fact that some of the analyses did have a well-formatted and clear graphical presentation of results suggests that Reclamation is well aware of the importance of presenting results visually. The Panel reminds Reclamation of the importance of the graphical presentation of results (including error bars or showing each year, when appropriate) and the importance of consistent labeling graphics throughout the document. Careful consideration of the scaling of the axes, especially the y-axis (response variables) enables effective interpretation of results and ensures that differences among graphs are correctly compared. In addition, in at least one case (HEC-5Q for EXP1), results were presented despite stated (numerical) issues with the results.

The Panel also sees a need for an introduction that presents visual conceptual models (or overview tables) that show the effects and how they were assessed (perhaps with links to the materials), and graphically showing how analyses (stressors and biological effects) link together to produce population-level results relevant to assessing conservation status. For some species, the information is already there but it is in terms of extensive narrative text spread over many documents. The Panel would have found it helpful to have links between documents when appropriate. For example, in some cases, the methods were presented in one document and the results in another (TDM, WUA models; SAIL). The idea of presenting matrices of stressor effects by species should be considered, as well as a summary by stressor across species.

7.2 Comment 2: Specify how multiple models will be used to increase confidence.

Reclamation included two life cycle models for Delta smelt and winter-run Chinook salmon, whereas other species were not represented at all by LCMs. A multiple-model approach can effectively quantify the level of uncertainty of predicted responses to alternative operational scenarios (see ISAB 2023). In concept, each model represents a different (but plausible) view of the real system. Judging which model is best is often difficult because each is based on defensible assumptions and may perform similarly when compared to field data as part of calibration and validation. When predicted responses of multiple models agree, uncertainty is reduced, increasing confidence in the model results. Disagreements among models are also useful because they show a range of possible responses that are plausible.

The success and usefulness of multiple models in reducing uncertainty depends on how the models are selected and implemented and how results are interpreted in a specific situation. Key aspects of a multiple model analysis are: why the different models were selected; how much information they share (e.g., are they fitted against the same monitoring data?); the operating constraints for calibration, validation, and scenario analyses; and how the predictions are combined across models to inform decisions. What determines effectiveness and success is how the analyses are conducted and interpreted. For example, a common mistake is to use multiple models and then simply treat their predictions as completely independent. This can lead to overconfidence when predictions agree because the appropriate confidence level is less because the models are not independent.

The Panel strongly recommends that, as part of the Effects Analysis, Reclamation develop a plan on how the multiple model predictions will be used and interpreted. A key aspect will be clear statements of the assumptions underlying how the model predictions were used to increase confidence in conclusions.

7.3 Comment 3: Confirm and discuss the lack of model sensitivity to many alternatives.

Model predictions of small biological effects for an alternative relative to baseline are as important as predictions of large biological effects. Small effects are important because of the issue of “false negatives”, that is, the model wrongly predicts a small effect to a stressor for an alternative when in fact, the actual effect is large. The Panel noted that for many models, multiple alternatives generated similar biological effects, and in some cases, these common responses showed little differences among alternatives. Often, one alternative differed from the others.

Given the complicated nature of the alternatives and their differences when each has a collection of management actions, the similarity of predicted biological effects (both when small effects and large effects) across alternatives suggests that key effects that would have distinguished alternatives were averaged away or key effects were missing. The models seemed too resistant to expressing differences among alternatives. This may be reasonable and realistic but needs to be confirmed. The species with LCMs are the best way to examine the realism of this pattern of similarity of predicted responses to stressors across alternatives. LCMs allow for easy examination of multiple stressors that would differentiate between alternatives.

7.4 Comment 4: Provide more explicit presentation of uncertainty in predictions from different sources.

The importance of uncertainties (and associated risks) is indicated by having a Charge question devoted to the topic. Dealing with uncertainties (and certainties) is always a challenge with a complicated coupled modeling approach because sources of uncertainty are propagated along the chain of coupled physical and ecological models. All of the model analyses in the Draft Effects Analysis involve uncertainties, and the models differed in how the collection of uncertainties manifested themselves by the predicted biological effects of the stressors associated with alternatives. While strict, explicit treatment of uncertainties may not be feasible, additional discussion of key uncertainties is warranted. Given the similarity of predicted responses to stressors among alternatives, a clear understanding of the ability to distinguish small differences as ecologically meaningful and management-relevant in the context of uncertainties becomes critical and efforts to reduce uncertainty are important.

7.5 Comment 5: Clarify and standardize the baseline for comparison with alternatives.

Reclamation went to great effort to explain and document the different baseline conditions for the EIS and the BA. The next step is to make certain all analyses use these baselines the same way. The more standardization that is done, the easier it is to present results across species and to synthesize results in a defensible and transparent manner. A thorough sweep through the documents to ensure the terms “Baseline”, “NAA” (No Action Alternative), “Alternative”, “Proposed action", etc. are used consistently and labeled identically within and across the EIS and BA would add clarity for readers.

7.6 Comment 6: Including climate change is a sound approach.

The Panel supports using drivers that include the effects of climate change in the simulation of baseline and alternatives. Climate change seems to be included in a reasonable way, although the Panel did note some cautions and possible issues (see CALSIM-3 review and climate change review). What is missing is a clear demonstration of the role assumptions about climate change played in affecting the predictions and as a potentially dominant driver of response to the different alternatives. Presuming the representation of climate change is reasonable, how well does the representation of alternatives accommodate climate change-induced changes in the ecosystem? Further, how do the effects of climate change interact with the different biological effects among alternatives? This would be important information to know when interpreting differences (or lack thereof) among the alternatives. This may cause a pivot on how the alternatives are viewed. Rather than focusing on the differences among alternatives, one might increase the focus on how the effects of climate change are represented (including drivers, individual stressors, and population-level responses) and the robustness of the alternatives to other scenarios about plausible climate futures. To date, realized climate change has tended to follow what was considered near-worst-case assumptions (Schwalm et al. 2020) about air temperature, sea level rise, and other environmental variables. Selected simulations of baseline and some alternatives, with and without climate change, would provide a direct way to assess the role of climate change and if any further work is needed.

7.7 Comment 7. Always use DSM2 and HEC-5Q salinity and temperatures, when appropriate.

The Panel noted that DMS2 and HEC-5Q outputs were not always used when possible. For example, CALSIM-3 output monthly flows were converted to changes in aggregate food for Delta smelt by using flow-to-salinity and salinity-to-food relationships based on field data. Why was salinity from DSM2 not used? The apparent use of a mix of salinity and temperature values arising from different methods used across models, rather than a common mechanistic-based source, likely adds avoidable uncertainties.

7.8 Comment 8. Avoid averaging out effects.

The Panel noted that multiple alternatives generated similar biological effects; a few alternatives consistently generated biological effects that differed dramatically from the others. If this result is realistic, then it is very important. However, the issue is whether such a pattern could be, at least partially, a result of the way various models represented effects that differ among alternatives. Convergence of model predictions can result from a lack of resolution on how CALSIM-3 output is used in models (e.g., monthly, no extremes), the many steps needed to convert differences in alternatives from CALSIM-3, DSM2, and HEC5Q into inputs for the models (average away differences among alternatives), and by the representation of effects within the models being artificially insensitive to the differences generated among alternatives.

7.9 Comment 9. Establish consistent protocols for interpreting results.

The Panel suggests that Reclamation develop a procedure that provides the modeling groups with guidance for presenting and interpreting model predictions among alternatives. Another viable approach is to form an Integration Team (if not already done) that has access to the results as processed by each team and also to the “raw” outputs, if needed.   This will greatly help in the integration of the species level and cross-species comparisons. The synthesis is a critical part of the assessment and the more the individual effects analyses are integrated, taking readers to the end, the easier it is for others to understand the results.

7.10 Comment 10. Consider presenting species tradeoffs under alternatives.

Tradeoffs among the species’ responses to alternatives may be important to understand. Important information is knowing whether tradeoffs (or complementarities) differ across alternatives. Some examples of tradeoffs were briefly described in the documents that we reviewed. For example, species sharing habitat below Keswick Dam and tradeoffs were discussed in Peterson and Duarte (2020). Some graphical presentations would be sufficient to summarize any tradeoffs.

7.11 Comment 11: Further reconcile time and space scales among models.

CALSIM-3 served as a critical source for baseline and the alternatives for many of the analyses. The Panel has noted the challenges of using monthly outputs for certain effects, such as temperature effects on egg survival, in multiple places throughout this report.

There is also a conceptual issue and a technical issue in how CALSIM-3 outputs were used when the situation demanded less than monthly resolution. The conceptual issue was that interpolated values from monthly output does not capture the maximums or minimums and other moments (e.g., variance, possibly autocorrelation) of the within-month distribution, but such variability is important to accurately predict many biological effects. The technical issue is that the interpolation scheme itself, while modernized, did not address the computational issues of mass balance and unrealistic values. Thus, the modernization was mostly about the coding of the interpolation scheme and not the resulting realism. The magnitude of the errors in the interpolation can be examined to ensure that when they are used, the introduced error is known and manageable.

The use of CALSIM-3 monthly outputs as input to effects, even when interpolated to daily, likely misses potentially important exposures and effects. Certain effects are more a function of extremes or variability than highly smoothed daily values determined from monthly values. The Panel suggests that, for practical reasons, Reclamation examines this issue as a special analysis to gauge its potential influence rather than as a comprehensive use of daily values across in all of the many models. There are field data and output from DSM2 and HEC-5Q that do provide more realistic daily variability. Effects analyses that use those outputs can easily be repeated with smooth daily values to mimic the interpolated value from CALSIM-3.

7.12 Comment 12. Reconcile relevance of historical data used by models to present-day conditions.

The various models used different historical time periods in their development and testing. The ecosystem has undergone, and continues to undergo, changes in its productivity and structure, both from proximate stressors and from climate change. How well historical time periods (that do not include the most recent years) represent baseline and maintain their integrity and realism under climate change and alternatives into the future is unknown. It is possible that a model developed from an earlier time period was built upon relationships that no longer apply as strongly as they once did. For example, using a model developed for the early 2000s that performed well for those conditions may not perform similarly for present-day conditions. Reclamation discussed this issue with most of the individual modeling analyses and some models were appropriately updated. An example where this does not occur was WUA analyses that used well-developed relationships, but from data often ending in the early 2000s.

Another layer of the issue occurs when alternatives are simulated. They can involve novel combinations of conditions that have not been observed in the calibration and validation datasets. The same problem arises when considering whether the relationships estimated historically can be used for novel conditions. The Panel emphasizes that this issue cannot be easily solved, and that Reclamation should consider this carefully as results are interpreted.

7.13 Comment 13: Consider critical species' responses to stressors other than population.

Most of the biological effects used to assess the performance of the alternatives were relevant to two of four recovery goals (see McElhaney et al. 2000), namely population abundance and trends. The main abundance-related response was survival, and the main trend-related response was finite population growth rate. The Panel noted that although there was mention of habitat restoration efforts, the Alternatives modeled did not focus on spatial diversity (i.e., metapopulation-level responses) or life-history diversity as modeling endpoints, nor was the VSP recovery framework discussed. In addition, not all Central Valley rivers supporting spawning were modeled in some cases.

8. Comments on Species Chapters

The Panel reviewed the species chapters (designed for the BA) and determined that a single set of general comments was appropriate rather than comments on each of the species chapters. A single set of general comments was possible because Reclamation used a standard format for all of the species chapters. The chapters were not finalized, and some model results were not available yet, so comments specific to each species are potentially incomplete.

A method or process for accumulating the effects of the stressors by life stage is a critical element to a scientifically defensible approach. The method must be comprehensive and transparent. Whether required or not, the method should also provide the basis for not just listing the effects but also for integrating ("rolling up") the effects to enable statements about the effects of the Proposed Action alternative at the species level. This is often referred to as Cumulative Impacts or Cumulative Effects Analysis and is part of most assessments, such as EISs and

Biological Opinions. There is extensive literature on methods for performing cumulative impacts or effects analysis (Blakley and Franks 2021).

8.1. Comment 14. The method used to date in the species chapters for integrating effects is useful.

The Panel was encouraged by the species chapters and Reclamation's attempt to date to evaluate the effects by stressor and life stage for each species. Reclamation's approach starts with a statement of the status of the species (distribution, abundance, temporal and spatial domains) and habitat, a description of the limiting factors, threats, and stressors by life stage, and a listing of management activities. This is followed by a short summary of monitoring data and the current incidental take statement.

The analysis of the Proposed Action consists of a systematic evaluation by stressor and life stage. Each stressor effect by life stage that cannot be deemed discountable or insignificant (i.e., stressors with potential effects) is categorized by its **severity** (e.g., sub-lethal, lethal, beneficial, or minor), the **proportion** of the population affected (e.g., small (≤2%), Medium (>2% and <70%) or large (≥70%), sometimes prefaced by "likely"), and the **frequency** of occurrence of the increased (or decreased) stressor (e.g., high (≥75%), Medium (25-75%), or low (<25%), sometimes with a "likely").

A series of appendices prepared by Reclamation enabled the linking of conceptual models, field data, and model predictions (i.e., indicators) of individual stressor effects to inform the proportion of the population affected and the frequency of occurrence of the stressor: Appendix D analyzed potential stressors for the seasonal operation of the CVP and SWP, Appendix C summarized when fish may be present in different locations based on historical monitoring, Appendix G analyzed potential stressors due to facility-specific operations, and Appendix H analyzed conservation measures to minimize or compensate for adverse effects. These appendices were valuable documentation of stressor-by-stressor impacts on the life stages of each species. The Panel recognizes the effort involved in assembling and clearly presenting the many sources of information used in these appendices and the species chapters.

Reclamation then stated that they used a weight-of-evidence approach to infer the proportion of the population that will be affected and the frequency of an increase in the stressor; however, the Panel did not see any actual use of a weight-of-evidence approach. In the section entitled "Effects Analysis" in each species chapter, the sources for the evidence were systematically listed for each stressor and life stage, followed by a list of Conservation Measures that are designed to reduce the impact.

8.2 Comment 15. The analysis stops short of providing an integration of the stressor effects that limits understanding of population-level outcomes.

Although the approach used by Reclamation is logical, the analysis presented stops too soon and does not state any population-level conclusions. Whereas some of the strongest modeling to detect the local effects of Reclamation activities were the stressor models (e.g., TDM models), they were not integrated into a life cycle modeling framework (expanded in Comment 17). Even if such conclusions are not required, the information in the chapters on stressor-by-stressor effects should be presented so that such determinations are easy and transparent. The analysis lists the sources of the evidence under the weight of evidence statement and then stops. Appendix D has some of the information needed but it is not easy to combine this with the species chapters. Thus, the Panel recommends that Reclamation consider modifying or adding documentation to the species chapters to enable an easy extension to assessing species-level responses. The Panel also notes that the organization of the information in the species chapters would be very useful to the evaluation of all alternatives (i.e., the EIS), including the addition of a graphical presentation of the results.

8.3 Comment 16. Some components of the Draft Effects Analysis destined for the BA should be better estimated.

The severity determination seems too coarse, and it does not seem informative to a population (or Evolutionarily Significant Unit (ESU) or Distinct Population Segment (DPS))-level assessment. The severity terms come from situations focused on individual organisms. It was not clear how lethality was determined for some stressors because the processes (vital rates) affecting populations (growth, mortality, reproduction, movement) are interrelated. For example, mortality is often size-dependent and so it depends on growth. Would the reduced reproduction be lethal or sub-lethal? Importantly, how was the classification used (or will be used) to inform the severity of the effect and then used with the other terms (i.e., proportion and frequency)?

In some cases, the estimation of the exposure (proportion of the population affected) seems weak and relies too much on the percent occurrence of conditions across years rather than a more relevant measure that quantifies the degree of exposure over ecologically relevant time scales. For example, a typical statement (this is from the Delta smelt chapter) is: *"In 21 out of 27 (~78%) years, spring outflow was low (Figure 9-8)."* There are many data sources listed that should provide more informative exposure information and the Panel suggests this be explored.

8.4 Comment 17. The life cycle models are treated separately from the effects predicted by other models.

This may not be possible in the short term, but the absence of life cycle models for some species is both surprising and disappointing given their listing status. When a life cycle model is available for a species, its results are treated separately from the effects estimated from models focused on individual stressors. The estimation of effects with the life cycle models involves the simultaneous effects of multiple stressors across life stages, but the models do not represent all of the stressors. Integrating the models that looked at a single or a subset of stressors within the framework of a life cycle model is challenging but necessary. One approach is to incorporate mechanistic stressor-effects sub-models or modules into the life cycle models so they better represent and cover the important local and fine-resolution stressors. This is doable, although it likely would involve recalibration and validation.

8.5 Comment 18. There is no weight of evidence analysis presented.

The analysis of each stressor and life stage in the Draft Effects Analysis always stops with a form of the general statement:

> *To evaluate the weight of evidence for the X stressor, multiple location- and species-specific datasets and studies have been evaluated to infer the proportion of the population that will be affected and the frequency of an increase in the stressor.*

The Panel recommends caution in labeling the analyses as using "weight of evidence." Weight of evidence is not a simple listing of the sources of evidence or even a listing of the actual evidence. Weight of evidence is a formal methodology (Hope and Clarkson 2014; Linkov et al. 2009). The present sequential listing of stressor effects provided to the Panel did not include a clear and transparent weight-of-evidence analysis. The Panel found the qualitative summaries using a standard typology (e.g., severity-abundance-frequency) very useful and we recommend using a more formal weight-of-evidence approach, especially explaining how the different sources were used and how they were considered.

Using such an approach could also make better use of the life-cycle modeling and would ensure that multiple stressor effects on the same life stage and the effects over different life stages and processes (e.g., reproduction, mortality, migration) were appropriately 'weighted' to produce population-level outcomes (see Comment 18).

8.6 Comment 19. Clarify how stressor effects were determined to be insignificant or discountable.

The determination that a stressor was insignificant or discountable requires additional explanation beyond that provided. Reclamation states:

> *Based on best judgment, a person would not be able to meaningfully measure, detect, or evaluate insignificant effects. Discountable effects are those extremely unlikely to occur. Based on best judgment, a person would not be able to expect discountable effects to occur.*

It was not clear to the Panel how these criteria were applied to each stressor by life stage. The Panel notes that, in some cases, insignificant effects that were based on modeling results may be a consequence of modeling that underestimates differences among alternatives. Also, the Panel considered that Reclamation appeared to use modeling results in some places as absolute values – does this violate their philosophy about how modeling results should be used to make relative predictions? Given the recent weather patterns (drought, heavy rains), what does Reclamation consider to be "extremely unlikely to occur?" Is this referring to seismic activity causing levee failures?

8.7 Comment 20. Consider other related frameworks to improve and expand the approach.

In the field of Conservation Biology, there are frameworks to consider that are specifically designed to focus on threats to listed species. When available, the best available science includes the use of a Population Viability Analysis (PVA) to systematically add and remove multiple stressors (Murray et al. 2021). In fact, the best use of stochastic PVA models (i.e., stochastic life-cycle models) is to evaluate threats posed to species at risk, taking uncertainty into account (Caughley 1994). The Panel suggests Reclamation consider the ideas and concepts of PVA, if not an actual PVA, going forward.

In addition, the literature on how to perform weight-of-evidence (Linkov et al. 2009, Suter et al. 2017) and cumulative impacts analysis (Blakley and Franks 2021) is extensive. Of particular relevance here is how many cumulative-type analyses use a variation of the concept of the Drivers-Pressures-State Change-Impact-Response (DPSIR) framework (Patrício et al. 2016) and when quantitatively combined use a form of the index proposed by Halpern et al. (2008, 2015). Cumulative impacts have been analyzed for many ecosystems and projects (Foley et al. 2017); examples of large-scale ecosystem restoration include the California Delta (Diefenderfer et al. 2021). Whereas many of the example applications focus on spatial (often GIS-based) information (Halpern and Fujita 2013; Hammar et al. 2020), the approaches and examples also apply to non-spatial situations; in the case of the Draft Effects Analysis, the multiple species, multiple stressors, and life stages form the dimensions of the analysis matrix.

The approach by Reclamation could be clarified and strengthened by explaining how the analyses were applied and relating the approach used to other approaches used in similar situations. Reclamation may consider modifying their approach in light of these other analyses. For example, how multiple stressors affect each other (additive, synergistic, antagonistic - Crain et al. 2008) is often part of assessing effects when multiple stressors are present. Reclamation does not discuss this and so seemingly assumes all effects are additive, when in fact, it is likely some stressors show synergistic or antagonistic effects with other stressors. If stressors act synergistically, treating them separately would underestimate the total effect and could influence determinations of “discountable.”

8.8 Comment 21. Some of the studies cited are possibly outdated.

The Panel encourages the use of all studies with the caveat that some studies may have less relevance because of new methods becoming available and the changing ecosystem. For example, the Kimmerer (2008) analysis of entrainment of Delta smelt is relied on heavily by Reclamation. This analysis used data from 1995 to 2005, which is now more than 20 years old. Smith et al. (2021) is a more recent analysis of entrainment and is also presented. How these two studies, both of which are considered excellent analyses, can be combined is not clear from the present cumulative analysis. The Panel uses Delta smelt here because much has changed with Delta smelt in the system; similar examples can be cited for the other species.

8.9 Comment 22. A graphical presentation of the results of species-by-stressor effects is needed.

As part of the species chapters (and also perhaps for the EIS), the Panel suggests that Reclamation develop graphical visuals to communicate the results of the analysis by species and stressor. There are many examples in the literature of ways to visualize the results so that readers can see the patterns and properly interpret the results described in the text. The systematic approach is good, but the bulleted narrative form is difficult for the reader to comprehend because it is impossible to keep track of a long list of results when presented individually as text. This is closely related to Overarching Comment 1 that applied to the entire Draft Effects Analysis.

9. Tables

Table 1. The models listed in the documentation of the Draft Effects Analysis provided to the Panel. The models are categorized by type or purpose, whether they were included in the documentation provided to the Panel (versus not included and noted by Reclamation with a placeholder), and whether they had been previously peer-reviewed and updated [from the briefing presentation given by Reclamation on December 4, 2023]. The seven specific chapters for the BA were also reviewed as part of model reviews and treated as a collection about how they integrated modeling results on stressor effects. The Panel opted to not review Appendix K – Summer and Fall Delta Outflow and Habitat that included models. The appendix was not on the list of models to focus on provided by Reclamation to the Panel and there are two ongoing reviews that include Delta outflow and habitat.[4,5]

Type or Purpose	Sources used to evaluate models	Included	Previously reviewed
Numeric and Life Cycle Models	Appendix F Sections 1-3 and all Attachments related labeled 1-X and 2-X	Yes	No
	Appendix F Attachment 2-5 DSM2 Salinity	Yes	Yes
	Appendix F Attachment 2-11 HEC5Q	Yes	Yes
	Attachment F.3 CVPIA Winter-Run Life Cycle Model; Attachment F.2 CVPIA Winter and Spring-run Life Cycle Model	Yes	Yes, but updated
	Attachment F.4 CVPIA Spring-Run Life Cycle Model	Yes	Yes, but updated
	Attachment F.1 Maunder and Deriso in R Model	Yes	No

---

[4] National Academy of Sciences' Review of the Long-Term Operations of the Central Valley Project and the State Water Project https://www.nationalacademies.org/our-work/review-of-the-long-term-operations-of-the-central-valley-project

[5] Delta Science Program review Panel on Summer-Fall Habitat Action Monitoring and Science Plans and Structured Decision Making Approach Peer Review https://deltacouncil.ca.gov/delta-science-program/summer-fall-habitat-action-monitoring-and-science-plans-and-structured-decision-making-approach-peer-review

Type or Purpose	Sources used to evaluate models	Included	Previously reviewed
	Attachment F.5 Delta Life Cycle Model with Entrainment (LCME)	Yes	Yes
Old and Middle River	Attachment I.6 Volumetric Influence Analysis	Yes	No
	Attachment I.5 Survival, Travel Time, and Routing Simulation Model (STAR)	Yes	Yes
	Delta Passage model	No	Yes
	Particle tracking/fate modeling	No	No
	Eco-PTM	No	No
	Attachment I.3 Delta Export Zone of Influence Analysis	Yes	No
	Attachment I.1 Negative Binomial Salvage Model	Yes	No
	Attachment I.2 Old Middle River Salvage-Density Model Loss	Yes	No
	Winter run CWT proportional loss	No	Yes, but updated
	Attachment I.4 Longfin Smelt Salvage Old Middle River Relationship	Yes	Yes, but updated
	Flow into Junctions	No	Yes, but updated
Spring Delta Outflow	Attachment J.3 Zooplankton-Delta Outflow Analysis	Yes	No
	XT model	No	Yes
	Flow threshold salmon survival	No	Yes
	Appendix J.2 Sturgeon Year Class Index and Delta Outflow	Yes	No
	Appendix J.1 Longfin Smelt Outflow	Yes	No
Summer and Fall X2	SCHISM Habitat suitability modeling	No	Yes
	Winter run juvenile production index model	No	No

Type or Purpose	Sources used to evaluate models	Included	Previously reviewed
Shasta Coldwater Pool	Attachment O.3 Sacramento River Weighted Useable Area Analysis	Yes	No
	Sacramento Dewatering analysis	No	No
	Sacramento Juvenile stranding analysis	No	No
	Attachment L.1 Coldwater Pool Storage and Coldwater Pool Exceedance Analysis	Yes	No
	SacSalMort & Reclamation egg mortality modeling	No	Yes
	Attachment L.3 Egg-to-fry Survival and Temperature-Dependent Mortality	Yes	Yes, but updated
	Attachment L.2 Sacramento River Water Temperature Analysis	Yes	No
Folsom Flow and Temperature	Attachment M.3 American River Weighted Useable Area Analysis	Yes	No
	Attachment M.2 American River Water Temperature Analysis	Yes	No
	Attachment M.1 American River Redd Dewatering Analysis	Yes	No
Stanislaus Stepped Release Plan	Attachment N.1 Stanislaus River Water Temperature Analysis	Yes	No
	Appendix O- Tributary Habitat Restoration – only documentation for Stanislaus WUA	Yes	No
Tributary Habitat	Attachment O.2 Science Integration Team Life Cycle Model Habitat Estimates	Yes	No
	Attachment O.1 Coldwater Pool Clear Creek Weighted Useable Area Analysis	Yes	No

Table 2. Specific models reviewed in the Appendix (Section 10) showing their source document and the alphanumeric label assigned to their review. Several models were not evaluated with separate reviews and these are noted in the table with information about where in the report the reader can find comments relevant to these models.

Model	Review Label or Location
CASLIM-3, using multiple documents	A
Climate change, including CALSIM-3	C
Appendix F - Attachment 2-5 DSM2 Salinity	N
Appendix F - Attachment 2-11 HEC5Q	O
Attachment F.1 Maunder and Deriso in R Model	H
Attachment F.2 CVPIA Winter and Spring-run Life Cycle Model	Q
Attachment F.3 CVPIA Winter-Run Life Cycle Model	W
Attachment F.5 Delta Life Cycle Model with Entrainment (LCME)	L
Attachment I.1 Negative Binomial Salvage Model	E
Attachment I.2 Old Middle River Salvage-Density Model Loss	F
Attachment I.3 Delta Export Zone of Influence Analysis	X
Attachment I.4 Longfin Smelt Salvage Old Middle River Relationship	M
Attachment I.5 Survival, Travel Time, and Routing Simulation Model (STAR)	G

Attachment I.6 Volumetric Influence Analysis	V
Attachment J.1 Longfin Smelt Outflow	I
Attachment J.2 Sturgeon Year Class Index and Delta Outflow	J
Attachment J.3 – Zooplankton -Delta Outflow Analysis	Y
Appendix K – Summer and Fall Delta Outflow and Habitat	Not reviewed because there are ongoing in-depth reviews by the National Academy of Sciences and the Delta Science Program.
Attachment L.1 Coldwater Pool Storage and Coldwater Pool Exceedance Analysis	B
Attachment L.3 Egg-to-fry Survival and Temperature-Dependent Mortality	S
Attachment M.1 American River Redd Dewatering Mortality	Z
Attachment M.2 American River Water Temperature Analysis	S
Attachment M.3 American River Weighted Useable Area Analysis	D
Attachment N.1 Stanislaus River Water Temperature Analysis	S
Appendix O - Tributary Habitat Restoration – only place the Panel found documentation for Stanislaus WUA	R
Attachment O.1 Coldwater Pool Clear Creek Weighted Useable Area Analysis	K
Attachment O.2 Science Integration Team Life Cycle Model Habitat Estimates	T

Attachment O.3 Sacramento River Weighted Useable Area Analysis	U
Ch. 11 Killer Whale (zooplankton only)	P
Ch. 05 Winter-run Chinook Salmon	See species chapter comments (Section 9)
Ch. 06 Spring-run Chinook Salmon	See species chapter comments (Section 9)
Ch. 07 Steelhead	See species chapter comments (Section 9)
Ch. 08 Green Sturgeon	See species chapter comments (Section 9)
Ch. 09 Delta Smelt	See species chapter comments (Section 9)
Ch. 10 Longfin Smelt	See species chapter comments (Section 9)
Ch. 11 Killer Whale	See species chapter comments (Section 9) and review comments in this Appendix related to zooplankton effects – labeled P.

10. Appendix. Specific Comments on Individual Models

Guidance for Using the Panel Reviews of Individual Models

Major Comments on Certain Models

Many of the individual model attachments and documents were incomplete and the Panel has commented on these to the extent possible using the provided documentation. Many analyses also used existing models and thus, while the Panel had comments, we erred on the side of assuming that these models, based on their previous use, were reasonable approaches. Exceptions, where the Panel had significant issues with the models in their current form, were:

a) Appendix I OMR, Attachment I.6, Volumetric Influence Analysis – Labeled review V,

b) LTO Appendix J – Spring Delta Outflow, Attachment J.1 Longfin Smelt Outflow – Labeled review Eye,

c) LTO Appendix J – Spring Delta Outflow, Attachment J.2 Sturgeon Year Class Index and Delta Outflow – Labeled review J,

d) LTO Appendix J, Attachment J.3 – Zooplankton Delta Outflow Analysis – Labeled review Y, and

e) Ch. 11 Killer Whale – Labeled review P.

Panel Comments on Species Chapters

The Panel reviewed the species chapters destined for the BA, but written comments are not included here. These were: Ch. 05 Winter-run Chinook Salmon, Ch. 06 Spring-run Chinook Salmon, Ch. 07 Steelhead, Ch. 08 Green Sturgeon, Ch. 09 Delta Smelt, and Ch. 10 Longfin Smelt. The exception was Ch. 11 Killer Whale, which we included a written review of because the analyses provided on food effects greatly differed from the other species and the effects analysis was also quite different.

The general format for the species chapters included a summary of the status of the species followed by a summary of the Draft Effects Analysis results, and this is an excellent start. Presently, the chapters mostly consist of a listing of effects results, generally by life stage and stressor (i.e., a long listing of stressor effects by life stage). These are the raw ingredients for a synthesis of these effects and responses to risks at the species level. Our major comments about the species chapters are covered in the Global and Overarching Comments and apply to all of the species chapters. Other relevant comments can be found in the individual model reviews below.

In particular, proper synthesis of stressor responses for each species remains a major challenge that must be overcome with a consistent and transparent “roll-up” of the life stage species results. The life cycle models, when available, provide a framework for the synthesis but are not always sufficiently comprehensive in representing stressor responses across all life stages and effects. Weight-of-evidence is cited by Reclamation but that requires a level of rigor and systematic methodology (not just a listing of effects) which was undefined in the documents provided to the Panel, and also not used to synthesize across life stages and stressors.

(A) CALSIM-3 (Multiple documents)

Material reviewed:

a) LTO – Draft Biological Assessment, Chapter 1 to Chapter 4,
b) Appendix F – Modeling, main report,
c) Appendix F – Sections_1-1_to_1-3_Attachments_1-1-to-2_11, and
d) CALSIM-3 Report (August 2022) – not part of review material provided by Reclamation

The CALSIM-3 model is a standard operational hydrology model used for the operations of the Central Valley Project. As such, this is the model that Reclamation knows the best and they provide very detailed information about the settings and assumptions used for each of the simulation runs.

Topic #1: Application of climate change perturbations to drive CALSIM-3 simulations

Data uncertainty is a great challenge to any quantitative modeling, which is particularly true for climate change. The LTO team developed model simulations to support the analysis of CVP and SWP LTO under climate change, including 15cm of assumed sea level rise. For rim boundary conditions, “*historical and perturbed meteorological data were used for simulating projected surface runoff, base flow, surface water evaporation, and potential evapotranspiration variables for future periods using the Variable Infiltration Capacity (VIC) model*' (Appendix F, Attachment 1-1). In another example, CALSIM-3 projected hydroclimate input data under different change scenarios using methods such as the perturbation of total flows of major watersheds (Appendix F, Attachment 1-1). Overall, the perturbation method seems reasonable but the provided documents did not describe in sufficient detail how the perturbation method was applied to simulate climate change in CALSIM-3. If it is a uniform perturbation of historical and meteorological data, then it may not be the best practice for this study. Better simulation and regression methods should be considered (e.g., Kolstad and Moore 2020).

Topic #2: Filling data gaps for CALSIM-3 simulations

The draft analysis numerously reported data gaps and/or missing data. Examples in Appendix F, Attachment 1-3 include: "*streamflow data (for CALSIM-3) exist at the watershed outflow point for only a limited period between water years 1922 and 2021". "(As) evaporation data is incomplete it is necessary to develop a standard method of estimating reservoir evaporation rates beginning October 1921."* Techniques were reported in the draft BA report to address this data gap issue. For example, since no gage data exists for the watershed, "*it is assumed that runoff is proportional to the product of drainage area and average annual precipitation depth over the watershed. Outflow was determined through association of the watershed with a similar but gaged watershed and the use of multiplicative factors representing the ratio of watershed areas and the ratio of precipitation depths*" (Appendix F, Attachment 1-3). The technique of filling gaps with existing data is also known as imputation. The imputation techniques described in the Draft Effects Analysis seem very limited in scope. There are various missing data imputation techniques (see Jager et al. 2021).

Topic #3: Interpolation techniques to preserve volume continuity when linking CALSIM-3 model output (monthly time step) to connecting models (e.g. HEC5Q, DMS2) that require daily or hourly time steps to drive the model

As an overarching planning model to simulate operations of the CVP and SWP over a range of hydrologic conditions, CALSIM-3 simulates system operations for a multi-year period using a monthly time step. The model assumes that facilities, land use, water supply contracts, and regulatory (e.g., water quality, instream flows) requirements are constant over this period, representing a fixed level of development. For input of CALSIM-3 output into other models, the preprocessor aggregates various CALSIM-3 time series as well as interpolates the time series, as needed, from monthly to daily values (Appendix F, Attachment 1-3).

For the HEC5Q temperature model, several time series within each basin model require disaggregation from monthly CALSIM-3 inputs to a daily time series. The CAL2DOM utility program translates data from CALSIM-3 to USRDOM, including conversions from monthly to daily operations and the disaggregation and consolidation of flow data.

Reclamation undertook a modernization of the temperature preprocessor to improve their code transparency, understandability, and maintainability. The revised preprocessor utilizes the PchipInterpolator from the Python Scipy library to perform the spline interpolation (instead of using legacy Fortran code) for interpolating the time series from monthly to daily time steps. It is indeed a modernization. However, as noticed by the draft report,*"PchipInterpolator does not preserve the monthly volumes which is necessary to prevent an unphysically realistic trough"* (Appendix F, Attachment 1-3).

To address the volume conservation issue, Reclamation explains *"volume was enforced through a preconditioning operation that incrementally adjusts the maximum monthly magnitude until the average value of the spline matches the CALSIM monthly value."* Meanwhile, *"to prevent an unphysically realistic trough prior to large increases in magnitude, the code shifts the date of the maximum monthly magnitude backwards in time if the months differ in magnitude by more than a factor of two. This results [in] a continuous time series that is more smooth and representative of the CALSIM monthly time series than would otherwise be produced by PchipInterpolator with the maximum flow occurring mid-month. The maximum monthly flow is limited to occurring five days before the end of the month."*

Unfortunately, the methodologies used to address these two critical issues seem ad hoc. It is unclear whether these methodologies have been thoroughly validated and mathematically proven to be effective. The following plot (Figure 1) illustrates the negative flow phenomenon for the disaggregation of the monthly flow to daily flow by the PchipInterpolator for monthly volume (August 2014 to March 2015) of Cow Creek, a Sacramento River tributary in Shasta County.

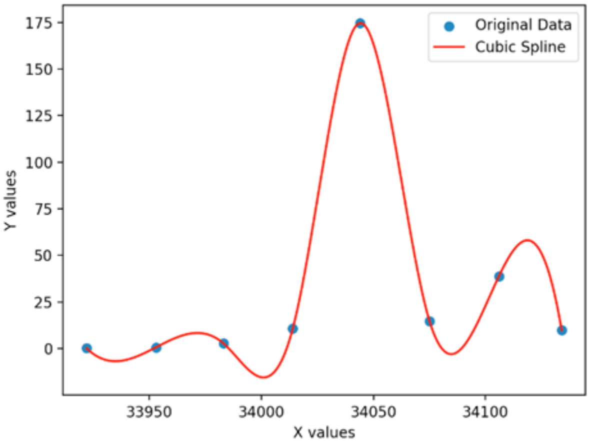

Figure 1. Disaggregation of monthly flow to daily flows generated by the PchipInterpolator for monthly volume from August 2014 to March 2015 at Cow Creek, a Sacramento River tributary in Shasta County. The x-axis is the count of days, and the y-axis is daily flows (cfs). The plot illustrates the "negative flow" phenomenon and the peaks and troughs pattern versus the observed data.

There are various studies on interpolation techniques in mathematics and engineering for mass conservation that suggest maintaining non-negativity and continuity (e.g., Hittmeir et al. 2018).

Topic #4: Optimization guidelines applied to the CALSIM-3 operations model to represent ESA-listed species effects

There are two key changes to the operation of the CVP that are being proposed and modeled for the Draft Effects Analysis: (a) water management for Shasta, based on water-year type to preserve cold water pool volumes under drought conditions and (b) Sacramento River pulse flows.

The water management philosophy for the CVP changed to Victorian objectives for the Shasta Reservoir: *"In order to recognize and adapt to these significant changes to the system, as a whole, Reclamation is proposing a new approach to managing Shasta which changes the balance between risks of flood control releases (aka spills) and maintaining water in storage for future drought protection and temperature management. This approach, described below, places a higher priority on maintaining storage for drought protection for all project purposes while limiting the frequency of spilling water due to flood control limitations"* (BA, Chapter 03 Proposed Action, p. 3-13; pdf p. 22).

Of particular importance for this review are operations for drought conditions (Bin 3-Protect Shasta operations): *"Under Bin 3, critically dry conditions exist, the system is stressed, and water resources are not available to meet all demands. There is low* *confidence to meet sufficient temperatures at the Clear Creek gage and future drought protection is at risk. The main biological objective is to protect winter-run Chinook salmon against decline. This Bin includes the widest array of potential water supply and fishery management actions to protect winter-run Chinook salmon from significant impacts and to protect against future drought risks"* (BA, Chapter 03 Proposed Action; p. 3-20 – 3-21; pdf p. 29-30).

To increase the outmigration survival of Chinook salmon, Reclamation would release up to 150 thousand-acre-feet (TAF) in pulse flow(s) for the Sacramento River each water year, typically in the spring, to benefit Chinook salmon in the Sacramento River watershed when the pulse does not interfere with the ability to meet temperature objectives or other anticipated operations of the reservoir. Reclamation plans to schedule this pulse after coordination through the Sacramento River Group (SRG) and Shasta Operations Team (SHOT) and may include coordinating timing with natural flow events, potential storage management operations, and/or pulse flows in tributaries (BA, Chapter 03 Proposed Action pdf p. 20; 3-11).

Most of the documentation about the simulations focuses on how Reclamation incorporates these two operational changes (Shasta management and Sacramento River pulse flows) into the simulations. These descriptions also include the other regulatory conditions that must be met and the Voluntary Agreements with stakeholders. **The Panel will not comment on the current regulatory implementation details or voluntary agreements.**

CALSIM-3 output is used as boundary conditions for HEC5Q and DSM2 models that then drive biological effects models, as well as directly in biological effects models. The Panel had some comments from the perspective of the end-use of the biological effects models.

The first comment is about the use of model outputs to estimate species-specific cumulative salvage or loss threshold. Reclamation states "*Since salvage or loss cannot be directly modeled in CALSIM, historic salvage data at the fish facilities at Banks and Jones Pumping Plants and other triggers for these actions were analyzed for the 2010 – 2022 period. Based on this historic data and water year type, the modeling used an OMR index of negative 3,500 CFS in a portion of each January through June*" (Appendix F Section 1-1 Modeling Methodology, p. 39). The Panel requests additional evidence be provided that this use of monthly output is a reasonable way to estimate the salvage-related effects.

A second comment is about turbidity bridge avoidance. Reclamation explains *"January through March in any Sacramento (40-30-30) Index Water Year type, if first flush has already occurred and if the turbidity trigger is reached (SACRI greater than or equal to 20,000 CFS), Projects operate to OMR Index of negative 2,000 CFS for ten days*" (Appendix F Section 1-1 Modeling Methodology, p. 39). Again here, the Panel questions if the use of modeling output treated in a way that enables statistically robust estimation of first flush and the trigger.

Thirdly, Reclamation states "*not all salinity requirements are included, as CALSIM-3 is not capable of predicting salinities in the Delta. Instead, empirically based equations and models are used to relate interior salinity conditions with the flow conditions*" (Appendix F Section 1-1 Modeling Methodology, p. 11). This linked model uses empirical equations without evidence of its accuracy. It was unclear to the Panel from the provided text whether the Department of Water Resources' artificial neural network was modified for the Draft Effects Analysis to reflect the new empirical relationships between salinity and flow that incorporated sea level rise.

(B) Attachment L.1 Coldwater Pool Storage and Coldwater Pool Exceedance Analysis

This appendix analyzes alternatives for the management of Shasta Reservoir for water temperatures downstream of Keswick Dam. The last remaining population of winter-run Chinook salmon is below Keswick Dam and relies upon the operation of the CVP to provide cold water for spawning and incubation over the summer months.

An initial alternative report (LTO 2021) developed potential options for the LTO of the CVP and SWP. Reclamation identifies 8 management questions to inform the formulation of alternatives, such as *"does real-time onset and shaping of temperatures improve winter-run Chinook salmon production or does a fixed schedule based on historical observations protect fish with limited water supply impact"?*

Reclamation solicited input from agencies and interest parties for their "knowledge base paper", and conducted full 82-year CALSIM II simulations for alternatives; followed by HEC-5Q and temperature-dependent mortality (TDM) models.

The initial findings provided answers to five of the 8 questions, including partial answers, while answers to the other three questions are still "under development".

Among these answered questions, some answers are clear, even quantitatively. However, some findings are less formative, such as those presented for “limited effect”.

A comprehensive set of literature, datasets, and models for the development of “the knowledge base paper” are detailed in section 5, with a total of 26 pages, consisting of the main body of the appendix. Some datasets include data gaps or shorter sampling efforts than others, but overall, a large body of data is available. These datasets, in conjunction with the modeled data (i.e., CALSIM-3, DSM2, USRDOM), serve as input for models that can be used to understand and predict the effects of CVP and SWP operations on environmental conditions and fish distribution and loss. Overall, the contents are solid and informative.

Rationales behind different concepts and approaches to coldwater pool management strategies were documented during the development of the alternatives. These concepts are described here as lines of evidence in section 6. Reclamation stated that the analyses will be done to assess how the storage and cold-water pool criteria are met across alternatives.

In section 7 on Uncertainty, Reclamation developed a special study plan to answer a couple of key questions. The new models may be used alongside or combined with existing TDM models to evaluate the effects of operations and the Panel supports such efforts. At the time of this review, such information is forthcoming and not included in the document provided to the Panel.

(C) Climate change, including CALSIM-3 (multiple documents)

Material reviewed:

a) Appendix F, main report,
b) Appendix F, Section 1-1, CALSIM-3, DSM2 and HEC5Q Modeling Simulations and Assumptions,
c) Appendix F, Section 1-2 Callout Tables, and
d) Attachment 1-1, Climate Change

Based on section F.3.1 of the main report of Appendix F, climate change impacts representing 2022±15 median climate conditions and 15cm of assumed sea level rise were analyzed by updating CALSIM-3 meteorological and hydrologic boundary conditions for LTO, including the No Action Alternative, EXP1, and EXP3. In addition, from the callout table of DSM2 in Section 2.1 of Appendix F, 2022±15 median climate conditions are also included in HEC-5Q.

In addition to updating meteorological and hydrologic boundary conditions, Reclamation developed further climate conditions and a set of different scenarios to review the range of uncertainty. Historical and perturbed meteorological data were used for simulating projected surface runoff, baseflow, surface water evaporation, and potential evapotranspiration variables for future periods. Inputs of CALSIM-3 for the 2022±15 median condition are described in detail in sections 2.5 and 2.6. Three sensitivity scenarios: 2022±15 hot-dry, 2022±15 warm-wet, and 2040±15 median climate conditions are described in sections 2.7 and 2.8.

One key issue is how to represent future climate conditions. It appears that an overarching technique is perturbation. For example, for applying the 2022±15 median climate condition, rim inflows in the upper San Joaquin of CALSIM-3 were impaired before the perturbation process. The rim inflows were re-impaired after perturbing the unimpaired inflows to present future climate conditions. Missing data for some specific local project operations (impairment) was calculated as the difference between the unimpaired historical flow and the CALSIM-3 inflow time series, under the assumption that the local project operations will be the same in future climate conditions, not accounting for potential adaptation in local project operations. While all these descriptions are sound, they lack substantial techniques that are needed for the Panel to fully endorse that this approach uses best available scientific information. The approach used by Reclamation seems reasonable and may be the best available, but the Panel would need further details and some additional analysis to conclude definitively.

As recognized by Reclamation, climate change represents the most significant and least understood threat to Reclamation's operations in California. Section 2.10 states "limitations and appropriate use of results", essentially the same statements as used in Section 4 of the main report of Appendix F, which should be applied to all numerical models developed and applied for the LTO of a complex water resources system.

(D) Attachment M.3 American River Weighted Useable Area Analysis

The document provided presented relatively complete methods and an initial presentation of the results. However, the results section was two paragraphs, one of which simply stated what results are found in which tables. There was little comparison among alternatives and no further explanation of why such results occurred. Several sections were identical to the text in Attachment O.1 – Coldwater Pool Clear Creek Weighted Useable Area Analysis. Both analyses were similar so some comments here (Attachment M) are repeated in Attachment O for completeness.

Data used in the analysis are from Bratovich (2017). That analysis was well described and provides a solid foundation for the use in this analysis. There are only WUAs for spawning, as no reliable data on rearing WUAs exist- except for a 1980s report that was deemed unreliable. The Panel notes there are potentially usable relationships from other rivers in the region. Composite WUA curves exist for eight stretches of river and include simulations from a 93-year period of record. WUAs were estimated for spawning months for each species under each water-year type, all water-year types combined, and for each spawning month.

Flows used in the analysis were obtained from CALSIM-3 below Nimbus Dam. The analysis assumes those flows represent flows in the downstream 10 miles where WUA was applied. An explanation that is reasonable (e.g., no major tributaries) would strengthen the justification for using these flows. Also, Bratovich created a composite WUA derived from eight different river stretches, however, it is unclear if Reclamation uses this composite WUA curve to compare across scenarios. If Reclamation does, that could be potentially problematic. While it may be defensible to roll up WUAs to create a composite WUA, it may not be defensible to work in the other direction. The Panel suggests providing a rationale for whether it would be preferable to model the flow for each of the eight sections of river under each management scenario and then rolling those results into a composite curve. A better explanation of the approach is warranted.

In addition, is there an assumption that the changes in flow under the alternatives must be realized and further must adhere to how the components (velocities and depth) are each assumed to affect habitat suitability? That is, if predicted values of velocity and depth were available for each alternative, and these were plugged into their habitat suitability curves (rather than flow into the composite relationship), would you get the same responses in WUA?

CALSIM-3 produces a monthly time step, and WUAs should be interpreted as monthly averages, which "*faithfully represent the average conditions affecting fish*". Therefore, Reclamation states that using monthly averages is justified as "acceptable". Acceptable for what? It is acceptable for ensuring that the data and interpretations match in terms of their time step, and that we do not assume more specificity than the models can produce. However, it is not acceptable to assume that monthly average conditions describe spawning success because stochastic events and/or daily variables, such as scour from flood events, drops in water surface elevations, etc., all operate on smaller time step than monthly, and influence spawning success. The Panel cautions against assuming monthly time steps are adequate for an effects analysis on spawning success. However, there is substantial literature on what scales are relevant to fish (e.g., Witman et al. 2023, Thompson et al. 2013, plus many others).

There is very little interpretation or description of the results provided, other than to show the plots and make mention of the following notable findings: (a) for Steelhead, WUA increases from wetter to drier years, for all scenarios, (b) for Chinook, the maximum WUA occurs in the above normal water-year types, and reaches a near minimum WUA in wet water-year types, and (c) there appears to be very minimal differences among alternatives and no statistical or ecological comparisons are provided.

While the plots provided are easy enough to compare, the Panel suggests that tables be created to help with interpretation. It would be helpful to provide tables that outline percent decreases or increases from the No Action Alternative, or whichever baseline comparison is required. It is unclear how Reclamation plans to handle tradeoffs in model output among species or river segments, etc., or how Reclamation plans to integrate WUA metrics into the assessments of impact at the population scale. This may be discussed elsewhere, but in general, the connections among separate analyses were not very clear.

There is no discussion section in this document. There should be an explanation that makes the connection between the main spawning season for the species of interest so that one can see how alternatives not only compare among year types but also across months and how that maps onto the peak or tails of the spawning season. From there, the Panel (and Reclamation) could understand whether the alternatives promote the tails or center of the spawn time distribution which has important implications for genetics and population dynamics.

The results and discussion sections should both be expanded upon to provide an explanation for why the presented responses occurred across the different scenarios. As the methods currently stand, it would be impossible to know which of the habitat characteristics used in the WUA are responsible for increases or decreases in WUA under each scenario. For example, there are a lot of outliers in the Chinook figures, but no discussion as to what is going on there compared to the

Steelhead plots. Why are the Chinook so much more variable? Is it a model convergence issue? Something about the Chinook input data?

Lastly, it would be helpful if the scenario flows used as inputs to the WUA analysis could be placed upon the composite WUA curves so we can visualize how they differ from one another. There needs to be an explicit paragraph linking the WUA responses to the Proposed action alternatives with respect to flow, not just name.

(E) Attachment I.1 Negative Binomial Salvage Model

The document provided was a draft version that was incomplete.

Why was a different statistical method used here as compared to LTO Appendix I – Attachment I.2, OMR salvage-density Model Loss? This approach seems (at least appears) to be better. Results for both facilities are combined here but kept separate in Attachment I.2.

Where was catch per unit effort (CPUE) measured? This relates to how well CPUE would index abundance that is vulnerable to the pumps.

The analysis uses a clever treatment of explanatory variables to try to get more precise estimates of salvage for the alternatives. Another positive aspect of the analysis is the cross-validation using a standard sub-setting of data.

The results section consists of narrative text describing the results in the tables in great detail. The plots shown are very helpful and a similar format should be used in other analyses (e.g., Attachment I.2). While this is a reasonable first step and provides the basis for plotting and comparisons, synthesis of the results, especially to compare among alternatives, is needed. A strategy for synthesizing the results across analyses (i.e., operationally becomes across attachments) is also needed to effectively communicate the tabular results and compare alternatives. For example, this analysis combines the facilities while they are kept separate in Attachment I.2. There may be a logic to this but without any explanation, it adds unnecessary differences when comparing alternatives across analyses and can be viewed as a lack of coordination within the analysis team. Standardizing the results to use similar plotting across analyses would be the simplest way to increase consistency. This can be done by adding such summarizations (text and plots) to the end of the appropriate Attachments, and then using those in the synthesis.

The plots nicely show which months have high salvage, but this also tends to reduce the differences among alternatives, which is the purpose of the analysis. For example, are the differences among alternatives, which look small for a given month, important? When Alt1 is added, a different pattern is seen in Figure 2. A similar scaling of the y-axis is appropriate, and it seems another layer of presentation is needed to guide the reader. The Panel had difficulty comparing alternatives except for very dramatic differences in some plots. Reclamation should explore graphical ways to display this type of results. It would be relatively easy to select a month (March or April or whichever is highest) and make plots that highlight differences among alternatives – the same plots used in other analyses when annual values are shown (here also as a single value, but monthly rather than annually) would help with the interpretation of these results and add to the synthesis. For example, Figure 5 is scaled by the very high values in the wet year, which then compresses the values for all other water-year types. Scaling the y-axes to show patterns while also striving for as much consistency as possible across plots when so many plots are involved is difficult. Sometimes the best solution is to simply provide different versions of the same plot. Note this was done already in some attachments to display results for EIS versus the BA. Duplicate plots can be separated by putting the plots needed to compare alternatives into the synthesis section of each attachment.

There should be some thought and narrative given to what is the smallest difference that is meaningful. Variability is shown but does not seem to be used in the draft text in the results.

(F) Attachment I.2 Old Middle River Salvage-Density Model Loss

The document provided was a draft version that was incomplete. The discussion of results was limited to describing the results that were presented in the tables; no figures or plots were presented. The results consisted of 256 pages of tables that were poorly formatted (e.g., an entry of “29%” was spread over three lines). However, the Panel can offer some observations and suggestions as the information provided was sufficient for the Panel to decipher how the analyses were done and how they will likely be used. However, endorsement and definitive comments are not possible without seeing the interpretation.

The analysis method is straightforward. Daily salvage numbers (2009-2022) are used as number/TAF exported for a month and multiplied by the monthly export from CALSIM-3, separately for the CVP and the SWP. This resulted in monthly salvage estimates for 1922 to 2021 by species, by pump, and by alternative.

The results section consists of narrative text describing the results in the tables in great detail in very long paragraphs. While this is a reasonable first step and provides the basis for plotting and comparisons, there was no synthesis or plots of the results. The text mimics the tables just in text form. The Panel presumes this type of presentation is being prepared and will be added later. It was not practical for the Panel to examine the results for reasonable interpretation because of difficulties in determining how the alternatives differed. The Panel did note there were many zeroes for year-month combinations. A strategy for synthesizing the results is needed to effectively communicate the tabular results and compare alternatives. Likely, an additional variable that combines the two pumps will be helpful. Summarizing the results using plotting as in other analyses would be the simplest way to improve consistency.

There should also be some thought given to what is the smallest difference that is meaningful. There were no variance estimates included anywhere in the analysis. Statistical variance is available but not reported and, further, statistical variance is only one component of the variance that can tell you meaningful differences.

Multiple paragraphs start with the over-generalized sentence that the salvage density model, calculated across all water-year types for each month and all alternatives, has a wide range. In a few cases, it appeared to the Panel that the range shown was actually a "small" range rather than a "wide" range. The useful statement of results, when present, then occurs somewhere buried in the paragraph.

(G) Attachment I.5 Survival, Travel Time, and Routing Simulation Model (STAR)

The STARS model (Survival, Travel time, and Routing Simulation) is an individual-based simulation that predicts the survival, travel time, and entrainment of juvenile salmonids migrating through the Delta. Estimates are based on acoustically tagged, hatchery-origin late-fall Chinook salmon from 2007-2011[6].

Daily flow inputs are obtained from monthly CALSIM-3 outputs (set constant across days). We were not provided with documentation for the model (refers to Perry et al. 2018, which we did not review), but we understand that the spatial structure of the model is coarse, with juveniles moved between eight polygons. One source of variation in entrainment and survival stems from the scheduling of open vs. closed Delta Cross-Channel (DCC) gates.

---

[6] https://oceanview.pfeg.noaa.gov/shiny/FED/CalFishTrack/

The results reported show that routing of more juveniles into the interior Delta leads to lower survival. STARS seemed to distinguish different flow alternatives when used in previous analyses, showing that alternative monthly flow regimes specified by the 2019 Biological Opinion elevated survival compared to the baseline by routing fewer juveniles into the interior Delta. These regimes corresponded to winter months with greater Delta inflows, especially from Sacramento River inflows. Gate configurations also differed among alternatives on a sub-monthly basis. How gate-closure schedules differed among alternatives may be described elsewhere.

No interpretation is provided, so we are unsure to what extent differences in flow regimes versus gate configurations were responsible for simulated differences among alternatives. It is also possible that the timing of gate closures was responsible for pushing entry into the interior Delta away from February and toward shoulder months (December or April).

Comment 1. As pointed out in the Appendix, unless daily variation is reflected in the inputs (i.e., not monthly average flows), the STARS model will not simulate the effects of pulse flows to stimulate migration timing or any effects of sub-monthly Reclamation operations. Thus, the effects of flow on migration and survival may not be simulated at an adequate temporal resolution, a concern with how the STARS model was implemented for the applications.

Comment 2. The Panel recommends updating this analysis to use more recent calibration data (see for example Hearn et al. 2014). The short time-period used in developing posterior distributions is less than one ENSO cycle and likely does not include a wide range of water-year types or climate-related shifts from more recent decades.

Comment 3. The Panel recommends updating this analysis using data from the right ESU (winter-run versus hatchery late-fall). At the very least, information regarding differences in timing and in the sizes of juveniles of the two races at the time of migration should be reported in any analysis that requires borrowing parameters from another race or species. Winter-run emigrate through the Delta from September through June and, according to Williams (2012) travel slowly, appearing at the pumps mostly in February-March. By contrast, evidence suggests that late-fall run juveniles travel in the fall (peaking in October). Williams (2012) describes great life-history variation among Chinook races in migration timing, size, development (fry vs. smolt) at the time of migration, and length of estuary residence. In addition, they claim that hatchery releases are generally larger and therefore would travel faster and (presumably) experience lower predation mortality, although they also report that winter-run juveniles at the pumps averaged 120 mm, which is large. We note that late-fall Chinook were not covered in Appendix C – Species Spatial and Temporal Domains, and perhaps should be. Clearly, modeling contingent decisions leading to variations in migration behavior related to estuary rearing and size at migration can quickly become a very complex modeling exercise, and as modelers, we appreciate the need to keep it as simple as possible, but some justification for borrowing information from hatchery late-fall run juveniles should be given.

(H) Attachment F.1 Maunder and Deriso in R Model

The use of life cycle models is an excellent approach for combining multiple effects from the alternatives within the life cycle of Delta smelt in order to express the effects at the population level.

The document provided was a draft version that was informative but still incomplete. There was a short discussion of results comparing alternatives, but this was limited to one paragraph that only referred to Table 3 (never refers to Tables 4-6) and did not refer to any of the figures showing results (no reference to Figures 3-6 in the text). There seems to be two sections labeled "Results" with the latter one having two bullet points that are conclusions.

The text states *"important differences between the original M&D model and the application of Polansky et al. nevertheless remain, and include model structure, surveys used, inference method, covariates tested and consideration of density dependence; these differences are summarized in Table 1."* This is critical information needed to understand how to interpret the results of this model with the results of the other Delta smelt life cycle models. However, Table 1 is only a list of candidate covariates included in model selection. A key table seems to be missing.

The text is inconsistent as to whether density dependence was included or not in the transitions. At the end of the first paragraph in the Model Development section, the document states that *"all transitions were assumed to be density independent."* But then results from density-dependent versions are discussed as if density-dependent relationships were developed and included. For example, the text later states: *“The stochastic approach involved random selection of two covariates per transition from the complete set of candidates (Table 1) and random selection of which, if any, life stages were subject to density dependence (options for density dependence were weighted such that there was equal probability of no density dependence and any density dependence).”* Also later: *"The overall “best” model identified after application of the hybrid stochastic-stepwise model selection process included South Delta Secchi depth and Beverton-Holt density dependence for the sub-adult survival transition.”* Additionally, Table 2 shows the best models with density dependence. Perhaps the fitting and selecting of the best model included density-dependence but then a density-independent form was used for comparing alternatives? The Panel recommends the use of the density-independent version as more appropriate for this analysis because of the high uncertainty about the strength and life stage when density-dependent mortality occurs and because the density-independent version will more likely overestimate than under-estimate stressor effects.

Figure 2 which shows OMR values and outflow values used as inputs is very helpful. More plots like this are needed to fully understand the differences among the alternatives.

What is shown in the box plot (Figure 6) needs careful explanation. Is the geometric mean shown as the mean? What are the different values shown by the box, line, and points?

An important result is that the geometric mean of the population growth rates (λs) was below one for all ALT-2 alternatives except EXP1 and EXP3. Only Alternative-3 generated a λ value close to one. The presentation of the results is well done, and the plots are clearly labeled and logical. Estimates of variability about the predicted λs would further add to a solid analysis. Table 3 is a nice summary.

The second bullet in the second Results section proposes an explanation for why EXP-1, EXP-3, and Alt-3 perform better than the others. The proposed explanation is that more positive OMR values and relatively high June-August outflows occurred during dry years. This should be explored further to confirm the reasons why these generated higher geometric mean growth rates. This would be valuable information for understanding these results and for when the modeling results are synthesized with other analyses.

General Comments that Apply to the Maunder and Deriso model and the LCME model

The sharing of input data (e.g., covariates) with the other Delta smelt life cycle model (LCME) is appropriate. This does create issues with how to compare the predictions of the population growth rate of the alternatives between the two models. This issue of unclear use of multiple models is discussed as an overarching issue.

The heavy reliance on the choice of model parameters for how well they fit multiple survey indices has advantages, but also limitations. This approach enables clear statistical fitting of the model and the opportunity to judge which version fits the best and how well it fits. This empirical basis for model building and fitting can provide confidence in predictions. The limitations arise from the assumption that the indices relate to each other in a consistent way over time that is well-captured by the transfer function and covariates. One can argue that the Maunder and Deriso model is a large and complicated (and advanced) correlation analysis rather than a life cycle model. While the model does simulate the entire life cycle, it does not use the standard framework of defining ecologically meaningful life stages (egg, larvae, juvenile, adults) and uses growth, mortality, and reproduction to transfer from one life stage to the next. Indices can include multiple life stages and thus can have a fuzzy relationship with classical life stages. Nevertheless, the Panel sees a valuable contribution from the Maunder and Deriso model, as long as the results are properly interpreted.

Population growth rate (λ) is used to compare alternatives. This is a useful model prediction for comparing alternatives because it shows population trends, but it is not sufficient alone as an index of future persistence or extinction risk. First, a tenet of conservation modeling (PVA) is that the variability in population growth rates is also very important when assessing extinction risk in addition to the trend itself. This is an important feature that should be considered in all the life cycle modeling efforts, and this is why stochastic models should be used.

Lambda alone does not tell the reader the size of the population. In this model, λ is basically the ratio of population abundance in successive years so high or low λ can occur with high or low abundances. In addition, differences between λ values for alternatives have different implications for the population, depending on the status of the population. For example, a difference in λs of 0.2 would be different if the actual growth rates were about 0.2 (i.e., a doubling) versus if the population growth rates were close to one. Presentation of model predictions of predicted population status (i.e., indices) that accompany the λ values, as well as a risk of extinction based on variability around λ, would help with interpretation. The model was fit to these population indices (not to observed population growth rate). Although the Panel appreciates the caution in using predicted abundance indices, they do provide important context for interpreting the population growth rates. For example, it is possible to predict the same high growth rate at very low and high population sizes. Furthermore, high variability around growth rate can lead to a high extinction risk when the population is small (Staples et al. 2004).

Despite the great similarity between this modeling and the LCME, it is very difficult to compare the results between the models. The plots are reasonable for each analysis but are not coordinated across the models. It is very difficult, and unnecessarily challenging, to determine the similarities and differences between results from the two models. This creates easily avoidable problems in synthesizing the results of both models.

A question that will arise is how robust model predictions are to different covariates included. The authors discuss this issue in terms of collinearity among the covariates. Adding predictions from the lower-ranked versions that use different combinations of covariates but still have reasonable fits would be helpful, for example using multi-model inference (Burnham and Anderson 2002). This could be for a subset of results (those influential in comparing alternatives) and not needed for all results. This would build confidence in the primary modeling results that use the best-fit model. Given the high degree of overlapping methods and data between the LCME and Maunder and Deriso models, the Panel does not see an advantage in using both models unless additional work is done to determine how to properly interpret the models when used together. The two models are far from independent yet may have important differences. Their predictions cannot be treated as two independent predictions, but they deserve more confidence than the predictions from a single model. The use of multiple models is discussed further as an overarching comment.

(l) Attachment J.1 Longfin Smelt Outflow

The document provided was a draft version that was incomplete. There was no discussion of results and there was no discussion about differences and similarities in salvage among alternatives. The text of the results starts with a placeholder to insert "*Key Take Aways Here*" indicating more interpretation is coming, and only presents figures and tables. The Panel can offer some observations and suggestions going forward as the information provided was sufficient for the Panel to decipher how the model was fit and how it will likely be used.

The approach of the analysis seems statistically sound and attempts to quantify uncertainty using several techniques. The authors use advanced and appropriate statistical methods, including a Bayesian framework and model averaging. The overall fit is encouraging as a description of historical trends, and differences from earlier analyses are well documented. Table 3 is included for complete documentation. The figures are good, but an explanation of the violin plots is needed, as readers will not know the details that are being shown.

Why is outflow labeled as “Sacramento River” water-year type (e.g., Figure 2_EIS)? Also, there is no text with the results. The Panel realized that the figure names indicated the similar plots were for EIS, the BA, and combined.

The pattern of the predicted Fall Midwater Trawl (FMWT) index shows little differences among alternatives and across water-year types, partially due the log scale used for the y-axis (e.g., Figure 2_EIS). Yet, there seems to be a much larger variation shown in the time series plots (e.g., Figure 3_BA). The time series of the two outflow explanatory variables should be shown. Is the lack of resolution among alternatives because outflows are so similar or because the statistical model is too uncertain? One conclusion could be that the model is not useful for evaluating the alternatives. This needs to be addressed in the results and discussion (when added). Perhaps illustrate the “power” of the model by showing how the predicted FMWT index would differ with specific values of outflow representative of the alternatives, including the credible intervals. This is different than showing the predicted index value for alternatives (made even more difficult to interpret by not presenting the outflow values by alternative).

Acknowledging the log scale on the y-axis, the predictions of FMTE index are mostly between 100 and 1000, while the historical values are above that for the pre-*Potamocorbula* period. Is the explanation that the predictions for the alternatives used post-*Potamocorbula* conditions?

The stratification of years into water-year types here and through other analyses is a good framework for analyses and presenting results.

How many years are in each water-year type for the model fitting? This could be deduced from Table 3 but should be made clear to the reader. Also, showing the observed versus predicted water-year type is needed, as that is how predictions are used later. The time series presentation is helpful but not sufficient.

All high FMWT values are for before 1987 and so fall completely into the pre-*Potamocorbula* regime. Does this mean the model used to compare alternatives basically predicts a near-constant (and low) index value no matter the outflow? This would be the case if the model provided a good fit of the data.

The inclusion of a variable indicating the Pelagic Organism Decline (POD) as a covariate is interesting. The assumption is POD conditions will continue into the future as they exist now. Does that seem inconsistent with the state of the system now? The predictions are what would have happened under historical hydrological conditions of 1922 to 2022, given the present configuration of the Delta and that the regimes occurred. Each year Reclamation uses outflow and looks up the regime value and predicts the index. This use of regime needs to be discussed because it seems inconsistent with other similar regression analyses applied to salvage or indices. The repeating of hydrology with the current Delta is consistent with other analyses, but the regimes may not be. Are the predicted indices for the alternative projections looking forward as they seem to be presented as?

This is where interpretation of the results and a discussion would help and, when added, needs to address this.

Coarse relationships between highly aggregated variables, such as the Longfin smelt index and Delta outflow, based on correlations using annual values have many potential problems. These have been repeatedly documented (e.g., Keyl and Wolff 2008, Tyler 1992, Gargett et al. 2011, Walters and Collie 1988) and provide little predictive power for new conditions not within the range of the data. Correlations fall apart with the addition of new data and as empirical models; they fail to provide a mechanistic understanding (aggregated variables result in spurious relationships). The Panel greatly cautions against the use of the LFS index and outflow correlations for comparing alternatives without careful development and a level of mechanistic understanding of the reasons for the observed relationships; useful information can be obtained but not with simple development and application.

(J) Attachment J.2 Sturgeon Year Class Index and Delta Outflow

A sturgeon Year Class Index (YCI) is a relationship with Delta outflow, based on annual historical data, that uses White sturgeon as a surrogate species. However, the Green and White sturgeon have different life histories and little justification is given for using this information. Annual Delta outflows used in the YCI are too coarse to represent the effects of pulse flows on either the adults moving upstream or juveniles moving downstream or any seasonal upstream effects due to operations or diversions.

Please see our comments above at the end of the review of Attachment J.1 regarding caution in using this type of correlation model based on annual values. Nevertheless, we provide some comments on the details of the sturgeon YCI model below.

The document provided was a draft version that was incomplete. The discussion of results was limited to describing the results that were presented in the tables, and there were only a few sentences that synthesized the results (labeled for the EIS); no figures or plots were presented. The Panel can offer some observations and suggestions going forward as the information provided was sufficient for the Panel to decipher how the model was fit and how it will likely be used. However, endorsement and definitive comments are not possible without seeing the interpretation.

The index is used without any variance estimates, which may be a limitation of what information is available. The index, however, shows wide fluctuations in its annual values and also includes many zeroes. This suggests other statistical methods (e.g., Poisson regression) may be more appropriate. There are standard methods for data with zeroes, and methods that accommodate rare but large values should be investigated. There were no plots showing the model fits (the Fish (2010) reference had some predicted vs observed plots but for different outflow months that suggest there is a relationship). The reference to ICF International (2016) is not very helpful for the reader in finding the details (~1,300 pages). The Panel examined the plots, which reinforced our impression that the statistical model is weak for predicting the index among alternatives within water-year types. Basically, the model predicts a decreasing index with increasing dryness and no differences among alternatives (except higher values for Alternative 3). Also, the higher values for Alternative 3 start off much different under wet conditions and, with increasing dryness, converge to the predicted index values for the other alternatives.

The cross-validation analysis is a good addition and can greatly help in interpreting alternatives, although in this case there are no real differences except for clearly higher values for Alternative 3 under wet years.

As presented, the use of regression modeling to assess alternatives is questionable. With better fitting that accounts for zeroes and ensures models work well for average index values and also predict extremely high values, the use of this approach can be informative. How different are the outflow explanatory variables among the alternatives?

The choice of explanatory variables (i.e., how outflow is averaged) should be documented. Table 3 in Fish (2010) shows that many other months provided very similar evidence for the strength of the relationship with the index. Since the purpose is to compare alternatives, it may be important if the alternatives differ in outflow in other months than the April-May and March-July used in the regression equation.

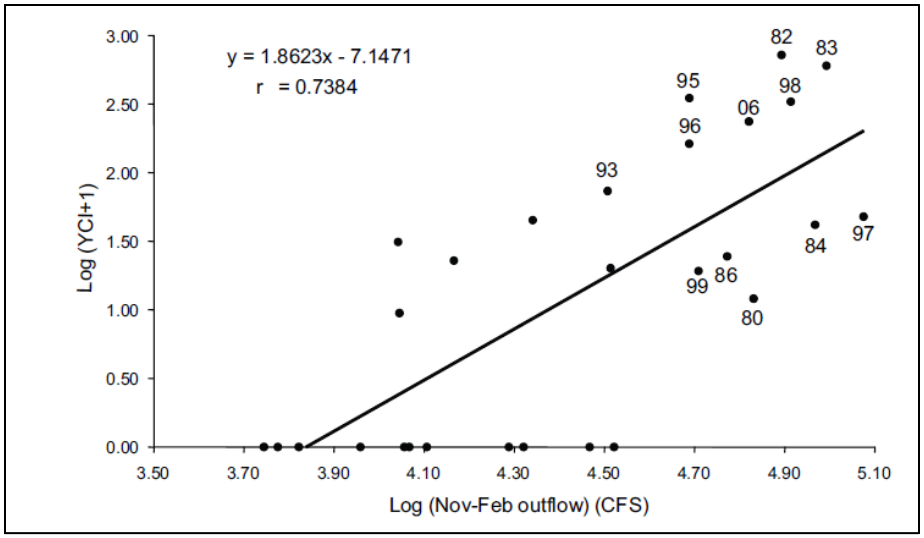

Figure 2. Plot of the log of year-class index of White sturgeon versus log of the mean daily outflow (cfs) calculated from November to February. The numbered points show select year-classes. This is Figure 2 from Fish (2010).

This figure is from Fish (2010) cited in this attachment. Different months were used in the Draft Effects Analysis but these data generally show the data used in the attachment.

(K) Attachment O.1 Coldwater Pool Clear Creek Weighted Usable Area Analysis

The general approach to habitat analysis, here and elsewhere, appears solid and based on well-described studies from USFWS.

Data used for the analysis is from the USFWS report for field studies conducted between 2004-2009. Spawning and rearing were evaluated using CALSIM-3 outflow data for each month of the 100-year period of record. RIVER2D was the primary hydraulic model used. Spawning WUA was assumed to be a function of water depth, flow velocity, and substrate particle size. Rearing WUA was assumed to be a function of water depth, flow velocity, adjacent velocity, and availability of cover. WUAs were developed by the USFWS for 3 sections of the creek for each species/run and life stage. Figures of these curves are provided.

CALSIM-3 flow data for Whiskeytown Lake releases to Clear Creek were used to estimate WUA for each species/run and life stage across the management scenarios. Total weighted means for spawning and rearing WUA that combine monthly WUA results from all three stream segments were computed. This seems to match the Bratovich (2017) approach on the American River, but because of the differences in presentation, the Panel was not certain if the same method for deriving the composite relationship, and then applying the same with alternative flows, was used. Given the apparent similarities across the multiple WUA analyses, it would be helpful to see a flow chart of the process to ensure the Panel is interpreting the order of operations correctly, and for Reclamation to identify whether different methodologies are being used across different river systems.

The use of monthly information again raises the question of what level of aggregation is reasonable to use. While the Panel understands that is what is generated from CALSIM-3, perhaps such a coarse scale is simply too limiting? There is substantial literature on what scales are relevant to fish (e.g., Witman et al. 2023, Thompson et al. 2013, plus many others).

Is there an assumption that the changes in flow under the alternatives must be realized and adhere to how the components (velocities and depth) are each assumed to affect habitat suitability? That is, if predicted values of velocity and depth were available for each alternative, and these were plugged into their habitat suitability curves (rather than flow into the composite relationship), would you get the same responses in WUA?

A major issue that almost always arises with WUA yet goes unmentioned is problems in interpretation. WUA is an index of habitat capacity not realized spawning.

The Panel suggests that more explanation be provided about the role and source for the weighting factors applied to each river segment for each scenario for Chinook. The document refers to Appendix C, Figures 35 and 36 for the weighting factors. These figures appear to be for *O. mykiss*, not Chinook, and are the proportion of redds by date and river mile. The Panel recommends Reclamation carefully review weighting factors for each WUA to ensure that all inputs to the analyses are sufficiently documented.

There is an enormous amount of data that is being pulled/available in Appendix C, and the paragraph describing which data was used is incomplete.

CALSIM-3 flow data at Whiskeytown Reservoir outflow is described as appropriate for application to Clear Creek's entire length because there are only minor tributaries, which only influence Clear Creek under high runoff conditions. The citation for this is weak, as it is listed as "USFWS?". Also, it is unclear what the definition for "high runoff conditions" is, and whether that should apply or affect Wet Year scenarios.

The uncertainties are the same for the American River across all WUA analyses, there is no explanation of how Reclamation plans to handle the uncertainty beyond identifying it, or how they plan to interpret the uncertainties with respect to the results. This should be included in discussion sections.

A new uncertainty listed for Clear Creek is fixed spawning periods were used to determine the effects of changes in flow on spawning. However, the time of spawning varies among years depending on flows. This feedback mechanism is not accounted for in the model, however, because spawn timing is a conservative, genetically controlled trait (Quinn 2005), the impact of this uncertainty is likely to be small. The Panel suggests a check on this rationale.

The discussion section regarding the validity of WUA analysis is the same across WUA models, with a minor, one-sentence modification explaining that improvements made to WUA models exist but are not currently available for the river of interest. This is repeated for each WUA, but there is no attempt to explain how one should take these limitations into account with respect to the specific scenarios being examined. Given that new models exist, what are the biases they correct? And how should managers then think about those biases with respect to the model comparisons at hand? Does the model consistently over or underestimate WUA under high or low flows? Are the WUA models fitted in the early 2000s representative of a small subset of environmental conditions (i.e. low flow years or high flow years?), or do they capture a wide distribution of flow conditions to which fish can then respond?

Results, figures, and tables are presented. The results section was a series of tables and figures only. There is no explanation or narrative to go along with the tables and figures. Results, however, are much less clear for Clear Creek compared to the American River. Box and Whisker plots are dominated by outliers, with some plots reflecting no boxes, just points, which is confusing for the reader. Interpretation of these results and of what Reclamation plans to do with these plots is needed. It appears the models for Clear Creek are unfinished, so evaluation is difficult.

The plots are a good idea, but they need further adjustments and supporting explanations. What is being shown? Key points on the boxes, vertical lines, and points need to be clearly defined. For example, Figure 1 seems to have the blue boxes cut off, and the different alternatives by month are difficult to see with the small similar-looking symbols. Why do Figure 1 and Figure 2 use different formats? Were all points considered "outliers" and shown as individual points in Figure 2? The only difference is which alternatives are shown, which is displayed in the legend. The Panel thought there should be 100 points shown since there were 100 years of CALSIM-3 used. Where are the missing points for Figure 2?

There are no plots for fry-rearing or juvenile-rearing. It appears that there is no effect of water-year type nor alternatives on the two rearing WUAs for both species. The Panel could not interpret the plots for spawning WUA due to uncertainties about what is plotted.

Results need to be explained in relation to projected flows, and there needs to be a discussion that incorporates uncertainties and explains why there are greater effects in some conditions and lesser effects under other conditions. There also needs to be an explanation of how these results build off previous models (i.e. CALSIM-3), how uncertainties are carried forward from previous models, and how the results and uncertainty from the WUA models will be carried forward to the next model they inform (presumably life cycle models).

(L) Attachment F.5 Delta Life Cycle Model with Entrainment (LCME)

The use of life cycle models is an excellent approach for combining the multiple stressor effects from the alternatives with the life cycle of Delta smelt and expressing the overall population-level response.

The document provided was complete and well-written with an excellent presentation of modeling results. The modeling reflects the long history of development and refinement of the LCME, and the diligent attention to detail by the authors. The inclusion of the memos to the Collaborative Science and Adaptive

Management Program (CSMAP) on key topics related to the modeling is very helpful.

The conversion of alternatives to their food effects is quite complicated and can be followed with careful reading. A concern is that the multiple layers of averaging may act to eliminate important differences in food among alternatives. The authors used a bottom-up approach starting with many prey species in spatial boxes and collapsing until a single metric per year was obtained. The two final metrics used as covariates were food in January-February for late-sub adults and in March for early adults. Many steps are involved from CALSIM-3 predicted salinity changes to changes in these yearly food metrics. Another approach might be to develop relationships using aggregated zooplankton data. That is, work at the level of the LCME rather than at a more resolved level that is then collapsed into a yearly value. Other modeling has shown the importance of food to Delta smelt population dynamics and λ (Kimmerer and Rose 2018) so the apparent insensitivity of the LCME to food differences from the alternatives warrants further investigation. It may be the correct result or may be due to the need to summarize the information with so many steps and decisions that differences in food metrics among alternatives important to Delta smelt growth were greatly reduced.

Why are the food covariates not included in Figures 5, 6, and 7? Showing all covariates is needed.

The general pattern of results seems to be that most alternatives generate similar growth rates and the few that differ result in higher growth rates in already good years. Figure 9 summarizes the overall result and the time series plots of growth rate show the bump in the few cases they occur, in good years.

The Panel had several high-level cautions about the LCME. These were already listed in the Panel comments on the other life cycle model (Maunder and Deriso) that uses a very similar approach to the LCME. The reader is referred to the comments under General Comments that Apply to the Maunder and Deriso model and the LCME model in the review labeled H.

Conclusions about the use the Delta smelt life cycle models

As stated in the review of the Maunder & Deriso model, despite the great similarity between the two modeling approaches (e.g., same years, shared covariate data), it is very difficult to compare the results between the models. The plots are reasonable for each analysis but are not coordinated across the models. It is very (and unnecessarily) challenging to determine the similarities and differences between the two models. This lack of coordination in presenting the results creates easily avoidable problems in synthesizing the results of both models.

Given the high degree of overlap in methods and data between the LCME and Maunder and Deriso models, the Panel does not see an advantage in using both models unless additional work is done to determine how to properly interpret the models when used together. The two models are far from independent yet may have important differences. Their predictions cannot be treated as two independent predictions, but they deserve more confidence than the predictions from a single model. The use of multiple models is discussed further as an overarching comment.

(M) Attachment I.4 Longfin Smelt Salvage Old Middle River Relationship

The document provided was a draft version that was incomplete. There was no discussion of results nor about differences and similarities in salvage among alternatives. The text ended with the presentation of figures and tables. The Panel can offer some observations and suggestions going forward as the information provided was sufficient for the Panel to decipher how the model was fit and how it will likely be used.

A rationale should be provided as to why years with high salvage were selected. Usually, the widest spread in the data is best for fitting regression models. There is no need for contiguous years to be used, so are there years from the past that can be added to the dataset? The gap between 1,000 to 5,000 cfs is particularly critical to fill as best as possible. That is where the alternatives are operating yet there is no data for the model fitting. While there is an attraction to using the same data as Grimaldi, that paper was published in 2009. Updating the analysis is needed, even if it confirms the earlier model because, at minimum, there will be more confidence in the predictions allowing for a better comparison among alternatives. Can years since 2008 be added with a covariate since they are post-Biological Opinion (2009)? Can any earlier years be added that have high or intermediate salvage?

The stratification of years into water-year types here and through other analyses is a good framework for analysis and presenting results.

How many years are in water-year types for the fitting? Is it 13 years spread among 5 types? The Pane;ll suggest Reclamation show the years for fitting by type.

What is meant by *"annual estimates were made for the mean upper and lower 95% prediction limits of salvage estimates"*? The Panel did not know what this meant and results using this were never shown.

What is shown in the box plots? Annual values grouped into water-year types from 1922-2021? And what is shown on the boxes?

There was little difference among the alternatives in OMR for the No Action Alternative, versions of Alternative 2, and Alternative 4. Yet, there seem to be larger differences in predicted salvage across some alternatives with similar OMR flows. How can this occur if the OMR is the only explanatory variable in the regression equation? While there is a monotonically decreasing relationship between salvage and OMR, there seems to be an important interaction between water-year type and OMR. Is it the variability in OMR? This is where the interpretation of the results and a discussion would help and when added, needs to address this.

While split sample testing is not feasible, it is possible to complete a jackknife process in which you delete 1 year at a time, however, some level of validation is needed.

Why were age-0 and adult life stages combined into one regression equation? This may be a good decision, but there needs to be an explanation. Whether age-0 or adults are entrained (indexed by salvage) can have important implications on population-level impacts.

How can the fit be improved as the regression is really between two clusters of points and may underestimate salvage at negative OMR flows?

Where is the tabular summary and graphic display of the fit of the new equation?

Are there other variables that would be affected by the Proposed Action that are not captured with OMR? For example, turbidity is mentioned. The analysis is really about how OMR, under the alternatives, will affect salvage and should not be over-extended to how the alternatives will affect entrainment. In addition, salvage occurs throughout most months of the year, so the predictions provided include only April-May salvage and not annual salvage.

An explanation is needed for how absolute salvage can be lower in dry years when OMR flows also seem more negative.

(N) Appendix F Attachment 2-5 DSM2 Salinity

Additional material reviewed:

a) Main Report,
b) LTO Appendix F: Attachment 1-3 Model Updates

Main Report - The use of the DSM2 Delta Hydrodynamic model is fairly limited in scope for this Draft Effects Analysis. Only basic salinity modeling has been reported in the documents provided to the Panel and therefore included in this review. No major changes were made to the DSM2 model for this application. The model is appropriate for representing Delta hydrodynamics and salinity intrusion. The application of the DSM2 results for the Zone of Influence analysis is discussed in a separate review (labeled “X”). The DSM2 particle tracking simulations were not available at the time of this review.

To account for sea level rise and the associated salinity intrusion, the western boundary of the DSM2 model (at Martinez) is driven by a tidal stage boundary and salinity. Salinity measurements were based on a flow-salinity correlation developed from simulations with a multi-dimensional hydrodynamic model of San Francisco Bay and the Delta that incorporated the assumed level of sea level rise. That multi-dimensional model has a tidal boundary on the ocean side of the Golden Gate Bridge with a constant ocean salinity value applied at that boundary. This approach to account for sea level rise and associated salinity at the DSM2 western boundary is well known and has been used in other studies.

For the Draft Effects Analysis, *“the DSM2 models assume a 15 cm increase in sea level rise. The Martinez electrical conductivity (EC) boundary condition is modified to account for the salinity changes related to the sea level rise using the regression equation derived based on the three-dimensional (SCHISM) modeling of the Bay-Delta under the future conditions with 15 centimeters (0.5 feet) sea level rise. The hydrodynamics and salinity changes in the Delta due to sea level rise were determined from the SCHISM three-dimensional Bay-Delta model simulations based on 2009 through 2010 historical hydrology. SCHISM results for changes of stage at Martinez were dominated by a scalar shift of about 0.5 feet. Given that the magnitude of the phase shift is very small relative to the DSM2 time step, it was assumed that 0.5 feet sea level rise would lead to 0.5 feet incremental change at Martinez with no phase shift”* (Appendix F Modeling Attachment 1-3, Model Updates, p. 13).

DSM2 flow boundaries are based on the **monthly** flow time series from CALSIM-3 (Appendix F Modeling Section 1-1, CALSIM-3, DSM2 and HEC5Q Modeling Simulations and Assumptions, p. 14). The timing operation of the DCC gates matters for salinity intrusion. Specifying that the DCC is closed for a certain number of days within a month is rather vague. DSM2 is a hydrodynamic model for an estuarine system. Therefore, tides matter and results are important on a tidal timescale. Yes, the flow inputs will be monthly values specified by the CALSIM-3 simulations with climate change adjustments. But the valuable information is at the tidal timescale. The anticipated particle tracking simulations need to be reported on a timescale of less than 1 month.

Issues to consider related to DSM2 modeling are: (a) appropriate use of the model in the Zone of Influence modeling (separate review), (b) reporting model results on a subtidal time step is appropriate for particle tracking and used to estimate entrainment (PTM not include in review material), and (c) binning results based on the volume of water entering at the boundary of the Sacramento and San Joaquin rivers (low, medium, high) is appropriate and recognizes that the Sacramento and San Joaquin rivers are distinct parts of the Delta system with unique water quality characteristics. For example, see the Zone of Influence modeling where high/high, medium/high, medium/medium binning was applied (Appendix I, Old and Middle River flow management, p. 3).

(O) Appendix F Attachment 2-11 HEC5Q

Additional material reviewed:

a) Appendix F Modeling, Section 1-1, CALSIM-3, DSM2 and HEC5Q Modeling Simulations and Assumptions

Although HEC5Q has been widely used for these California reservoirs, this model was significantly changed to operate the model in forecasting mode for this Draft Effects Analysis.

To date, an impressive amount of programming effort has been made to convert the HEC5Q model into a tool that can be applied to this application, which is very different from the standard operational use of this model. This effort included extending input data, filling in data gaps, creating a more realistic representation of Shasta Temperature Control Device operations, and changing techniques for iterative processes that were previously done manually. The old manual approach involved making adjustments that were dependent on the modeler’s judgment. While making these changes, the modelers have also found calibration errors that prompted a major review of the calibration of the model. Because of this major calibration, a calibration review document should be cited and available for reference.

The temperature modeling is a critical component of the overall analysis for both the BA and the EIS. Simulating Shasta Coldwater pool volume in addition to downstream temperatures in key spawning habitats is critical information. HEC5Q is a powerful modeling tool that produces an enormous amount of valuable output data at a 6-hour time step. It is vital that this information is synthesized in a format that best supports the needs of the ESU evaluation. As an example, the model output available to support evaluating key spawning habitat includes the maximum temperature experienced and the exposure duration exceeding a maximum temperature criterion.

To follow the "Appropriate Use of Model Results" guidelines (Appendix F Modeling, p. F-8; pdf p. 12), the HEC5Q model results are presented as mean monthly temperatures with the probability of percent exceedance at sampling stations. Other approaches to statistically present the data need to be considered. In addition to the mean temperature, please report the range of the temperature output (10% and 90% range) as well as the number of successive days of exceedance per month.

The modelers state that temperature modeling for the EXP1 temperature simulations is faulty for significant numerical modeling reasons. This scenario created numerical challenges due to the very low storage utilized by the HEC5Q basin models outside of their intended range of inputs. The Panel recommends that Reclamation not plot the EXP1 temperature simulation output data alongside all the other alternatives. The EXP1 temperature output is not in the same range as the other alternatives and the differences have already been identified as errors from the model representation of this unique configuration.

The Panel notes the following quote: ***"****The only definitive method to fully resolve the HEC5Q numerical issues under the EXP1 operations logic would be to re-architect the HEC5Q model engine itself to correct the problematic algorithms. Such an undertaking is not within the scope of the 2021 LTO and would require full revalidation/recalibration of the HEC5Q basin models. It was therefore decided to utilize an approach to minimize the numerical issues within the current HEC5Q model engine"* (Appendix F Modeling, Attachment 1-3 Model Updates, p. 25).

Much more synthesis of results is needed. For example, when a critically dry year fails to meet the set criteria, what are the underlying mechanisms? For example, was there a unique feature of the input hydrology? Was this year part of a multi-year critically dry event? Was there a carry-over in stratification from the previous year? Was this a simulated year where there was missing data and a default value was applied? The Panel recommends looking for themes to identify how to improve both the temperature model and the operations guiding criteria.

The preprocessing of CALSIM boundary data needed to be changed with the switch over to CALSIM-3 format. The time series interpolation was changed and temporal downscaling using spline interpolation was applied to the monthly time series. The Panel noted that the spline interpolation did not preserve monthly volumes and had to be done with a preconditioning operation that adjusts the maximum monthly magnitude until the average value of the spline matches the CALSIM monthly value. In addition, to prevent a physically unrealistic trough, the code shifted the date of the maximum magnitude backward if the magnitude changed more than a factor of two. Giving an example showing how the data was processed would greatly help the reader. How did these changes affect the original simulation results?

A second issue was the Meteorological Data Extension for these simulations. The Panel noted the extension of the HEC5Q to 2022, which covers recent conditions. Because California Irrigation Management Information System (CIMIS) does not provide coverage back to 1921, the period CIMIS data has **been augmented to is based on water-year types to backfill** for the full CALSIM period. If you are doing the simulation by individual year and not looking at year-over-year carryover, what is the purpose of going back to 1921 when there is no driving meteorological data available? Is it so you have a statistical range of temperature values to use in your wet/dry/critically dry classification? Do the water-year classifications change as climate change shifts the hydrology or do you use the historical water-year classification for binning data? How is climate change incorporated into these simulations? It has been suggested that water-year types be updated regularly to account for non-stationary future conditions (Rheinheimer et al. 2016).

A third issue was the use of Gerber, Nicolaus, and Modesto stations. *"Solar radiation, the primary variable used to calculate equilibrium temperature and the heat transfer coefficient, and the wind speeds were markedly different in both trend and magnitude between the CIMIS values and the existing HEC5Q meteorology*" (Appendix F Modeling, Attachment 1-3 Model Updates p. 23). This triggered a review and the primary finding of the Reclamation review was that "*the* *total solar radiation as measured at the CIMIS station was not being utilized in favor of top of atmosphere short wave radiation …. The differences between the CIMIS station information and the existing HEC5Q meteorology were significant enough to warrant additional consideration during the present extension. The difference due to geometric factors and wind speed could not be satisfactorily resolved*” (Appendix F Modeling, Attachment 1-3 Model Updates p. 23). The solution was “*revised geometric correction factors were applied to the top of atmosphere solar radiation estimates and the reduction factors were eliminated*” (Appendix F Modeling, Attachment 1-3 Model Updates p. 24). The Panel understands the necessity of these adjustments and suggests Reclamation better document how they influenced predictions.

The current presentation of the data makes extensive use of exceedance probability charts. Only one temperature per month is reported for each simulation; multiple years are combined by month. How is that temperature determined? Was it the maximum or median temperature of the time series? How many times during the month was that temperature exceeded? Importantly for fish, for how many days in a row are the criteria exceeded? A discussion of the exceedance charts in locations where there is exceedance should be discussed. What was the mechanism that caused relatively long simulated exceedances?

The time series of simulated temperatures for an example year should be displayed and the data processing steps used to determine the representative temperature for the month should be explained. In addition, a map of key stations (similar to BA 4-11) would also be useful. Starting with the time series, the aggregation into the final figures can be described. Even if the results do not match the observed temperature exactly, the model should still represent basic trends. Are predictions staying within a realistic range over time or are there concerning, consistent deviations? We know that the time series was broken into year-long sequences to reduce any accumulation of errors. However, what are the circumstances under which the model becomes numerically unstable? Can the model simulate 2-3 years of drought in a row when there is a carryover stratification?

HEC-5Q modeling analysis enumerates the frequency at which mean monthly simulated water temperatures exceed water temperature criteria for winter-run Chinook salmon obtained from the scientific literature. Reliance on exceedance based on monthly values seems highly questionable. **While this is in keeping with**

the guidance on the proper use of model outputs, this analysis ignores valuable model data by not reporting daily values.

Because biological effects of temperature are non-linear, the average of the function applied to daily values does not equal the function of the average, especially when compared to thresholds, such as maximum temperature tolerance. In our view, modeling thermal risks at a monthly resolution is not adequate for making decisions based on the potentially lethal effects of short-term extreme temperature exposures. The Anderson et al. model seems to be a reasonable approach because it uses a daily time step (i.e., realistic high-resolution variability in water temperatures).

(P) Zooplankton – Killer Whale (Chapter 11 Killer Whale, Appendix D Seasonal Operations Deconstruction)

In general, the Draft Effects Analysis for Killer Whales relied on qualitative information more than the other species. Because it differed from the other species, we review the assumptions here.

Some key assumptions used in the analysis:

a) Central Valley Chinook is 22% of sampled Chinook off Oregon coast, and 50% of Chinook of California Coast (both areas within designated critical habitat),
b) 40% of Southern Resident Killer Whale (SRKW) diet when whales are off the coast of California, and 18% of their diet when whales are off the Oregon coast,
c) Coded wire tags indicate 21% of returning Chinook to the Central Valley are natural origin fall-run Chinook salmon, and
d) 21% is used to represent the percentage of natural-origin Chinook in the ocean from the Central Valley that can be available for SRKW diet.

Assumptions about prey availability as a stressor included:

a) Prefer Chinook. Central Valley Chinook identified in prey of SRKW; 19% of SRKW prey collected in outer coastal waters and 5% or prey/diet items collected in Puget Sound,
b) The relationship between SRKW and Chinook is getting weaker. SRKW demographics decrease regardless of varying Chinook levels (likely due to multiple interacting stressors that need to be taken into account); whales can exhibit stress when they have less access to food and multiple stressors accumulate,

c) High uncertainty for winter, as data are biased towards summer and early fall months when whales are in inland waters and boat-based research activities can take place.

Assumptions about other stressors included:

a) Whales exposed to persistent organic pollutants (POPs) through prey, which they then pass on to offspring or release/metabolize when hungry,
b) Vessel effects caused sound interference with hunting success, energy expenditure, social cohesion, communication, foraging efficiency, etc., and
c) Oil spills would result in exposure to petroleum hydrocarbons, with resulting serious health impacts to SRKW. The Proposed Action is stated as not anticipated to change these stressors: Pollution and Contaminants, Vessel Effects, Oil Spill, or Acoustic.

The stressor of prey availability was anticipated to change *at insignificant or discountable levels.* Prey availability would be impacted by about 10% for 5 months of the year by the Proposed Action. The Panel notes that a 10% shift in the diet of an organism needs to be considered in the context of whether or not prey availability is near a starvation threshold or not. Simply saying only 10% of their diet may be affected, and therefore the effect on the whales is likely insignificant needs further confirmation. However, the analysis suggests that the Proposed Action will not affect the production of natural origin Chinook (which make up that 10% of the diet). The Panel advises that if there is any uncertainty as to the Proposed Action's impact on the natural origin Chinook, then this is a place where caution and further analyses may be needed. SRKW are surviving on near starvation thresholds, which can lead to malnutrition that causes them to metabolize fats wherein POPs are bioaccumulated. The release of these POPs reduces fecundity and survivorship of their young, leading to declining population rates. So, while the proposed action may not directly increase pollutant exposure, if it limits food resources, it can indirectly expose SRKWs to increased pollutant levels.

Under Critical Habitat Area the Draft Effects Analysis states that the *"final rule maintains the previously designated, but not in the action area, critical habitat in inland waters of Washington, and expands it to include certain coastal waters off Washington, Oregon, and California".* While it is true that SKRWs do not inhabit the Delta or rivers upstream, this missed the point. It is a connected hydrological system, and the key food source, Chinook Salmon, crosses between both realms as part of their life cycle. The range of Chinook consumed by SRKW includes the designated waters off the coasts of California and Oregon and is impacted within the action area. This connection should not be discounted.

The Draft Effects Analysis also makes the following conclusions about impacts on habitat: (a) no impacts to water quality because essential physical and biological features (i.e. ocean water quality) are not affected by the Proposed Action, and (b) no impacts are expected from Proposed Action on passage conditions to allow for migration, resting, and foraging because these actions occur in the coastal ocean.

(Q) Attachment F.2 CVPIA Winter and Spring-run Life Cycle Model

Additional material reviewed:

a) 01_LTO Appendix F Modeling,
b) 01_LTO Appendix F Modeling, Section F.4, and
c) Appendix O – Tributary Habitat Restoration.

Life cycle models offer the most suitable tool for scaling from operations to population-level effects. Two LCMs were reported for winter-run Chinook salmon, the CVPIA Winter-run and Spring–run life-cycle models. The CVPIA LCM is a spatially discretized aggregated state-space model with a monthly time step. The use of a stochastic PVA/life cycle model is necessary when estimating population growth and persistence, and we support the decision to use the peer-reviewed CVPIA LCM. It is impressive that such a large area has been represented in a way that incorporates available data and with potential for higher-resolution linkages to relevant drivers in the Proposed Action. The model is stochastic at the level of hydrologic year types, which makes it more suitable for representing population trends and low spawner abundances, i.e., 100, 1,000 females for Winter-run Chinook salmon.

The model also considers straying or diversity of spawning habitats, but only for Spring-run Chinook salmon. However, evidence suggests spawning may be occurring elsewhere and this may increase in the future with access to new spawning areas. In addition, limited information was provided about hatchery operations and how they were represented in the model and used to distinguish trends in hatchery-origin fish, wild fish, and combined populations.

However, the ability of the CVPIA model to examine impacts for a listed species is only as good as its ability to represent and discriminate among different alternatives. An important limitation of the CVPIA LCM is that the monthly time-step does not represent linkages between flow and temperature controls and survival and growth of early life stages well enough to capture the effects of ramping, pulse flows, or short-term temperature excursions. The calibrated model also produced migration survival in the main-stem Sacramento that was high across scenarios (pinned against 1), so that processes in the Delta, by default, control estimates of population growth (Appendix F, Attachment F.2). Because neither egg-fry survival nor migration survival respond to flow and temperature, the calibrated model would not appear to be designed to evaluate the effects of seasonal changes in flow releases (i.e., from spring to summer) or changes in ramping rates.

We note that there is a discrepancy in how early-life stage survival in the CVPIA Decision Support Model (DSM) is described in Appendix F and O, possibly because different versions are used. In Appendix O, section 4.7.3, CVPIA SIT DSM is used to model redd dewatering using WUA and monthly flow drivers, but in the version used for Appendix F, egg-fry survival, juvenile growth, and the temporal distribution of spawners are not allowed to vary (page 5).

The Panel notes that some of the short-term variability is proposed to be controlled by Conservation Measures; however, it is important to note the limitations of the models for representing these important effects from short-term variation because it questions the accuracy of their predictions that ignore this variation. It is also unclear whether the beneficial aspects of variable flows (e.g., pulses that attract up-migrating spawners or push juveniles downstream) were included in the modeling.

Reclamation also produced a modified version of the Peterson & Duarte (2020) model[7]. The model was designed to address questions related to passage and screening of diversions and habitat restoration. It is highly discretized at a coarse temporal resolution. The published version was used by Reclamation with some modifications that required recalibration. The LTO decision model by Peterson and Duarte was not described in the information that we reviewed; however, the paper provided valuable information and we considered the paper in our review. We infer that hatchery populations may have been added and that the Chinook runs may have been run separately. In the Peterson & Duarte published analysis, simultaneous fitting of the three Chinook runs was reported. This seems like a good idea. However, the analyses reported seem to treat them separately and they have separate models on GitHub[8].

---

[7] https://onlinelibrary.wiley.com/doi/full/10.1111/rec.13244

[8] https://cvpia.scienceintegrationteam.com/cvpia-sit/resources/dsm-r-packages

It is unclear how and whether minimum viable population size is treated by the life cycle models; is there an extinction threshold? Responses of spawning and rearing habitat to flow are represented either through WUA relationships or through more-detailed floodplain hydraulic models between Keswick Dam and Battle Creek on the Sacramento River (Appendix O Tributary Habitat Restoration, section 4.7.3). However, the fact that, in the modeling, migration survival is high and egg-smolt survival did not vary among alternatives reduces any possibility of differences among alternatives. Additional explanation is needed to understand and confirm this representation because it can force similarity among the alternatives.

An optimization framework (dynamic programming to select restoration options) limits the number of restoration actions possible at each time interval. Notes from the CVPIA Integration Team indicate that restoration actions in the Upper Sacramento were optimal at every time step. A sensitivity analysis was reported by Peterson & Duarte (2020). The most influential inputs for all Chinook runs were existing habitat, median discharge, and temperature. In addition, winter-run predictions were sensitive to initial abundance and total water diverted for the winter run.

Some uncertainties include how gaps in parameters and among species are filled from similar tributaries and runs, respectively. It is unclear whether this borrowing of parameters could potentially result in a poor representation of the seasonal and spatial effects of operations. Uncertainty in the environmental drivers is another concern. The temperature interpolation decisions for the CVPIA LCM seem reasonable as long as the effects of changes in operations are not impacted. This involved imputing equilibrium tailwater temperatures based on air temperature and distance downstream (note: this should also include flow because the distance traveled is higher at high flows), and basin matching. This is an important thing to check. Because temperature monitoring is widely available, empirical data should be able to be used for model testing and improvement.

The Panel questions the ability of the currently calibrated CVPIA LCM to represent different alternatives because of its insensitivity to freshwater processes. This is caused by very high juvenile survival during river migration and a constant egg-to-fry survival rate. Therefore, we question whether it is really possible for Reclamation to assess future effects of Shasta operations that affect egg-to-fry survival of the Chinook salmon runs or any influences during migration. By extension, this makes the use of the model questionable for evaluating the relative importance of river versus Delta versus ocean phases. We note that these issues can be easily fixed by calibrating against monitoring data in the river to produce more reasonable migration survival estimates and by linking egg-fry survival to the TMD indicator in SAIL. We encourage using an existing approach that can represent the effects of daily temperature on egg-fry survival, the effects of pulse flows, and moderation by minimum flows and ramping rates that influence stranding and dewatering mortality of freshwater life stages.

(R) Stanislaus River WUA (Appendix O Tributary Habitat Restoration)

Draft EIS Stanislaus models appear in Appendix N; however, there is no WUA analysis presented in either document in Appendix N. One analysis is on water temperature. The other focuses on the stepped release plan, The New Melones Stepped Release Plan.

It is mentioned that pulse flows are needed to support different life stages of steelhead. This work does not appear to be linked to the WUA analyses. Performance metrics for fish appear to be linked to migration cues and temperature.

In Appendix O Tributary Habitat Restoration, however, there is a section on the Stanislaus River describing that WUA curves exist for fall-run Chinook, for 23 miles of river below Goodwin Dam. Data are from 1993 (USFWS).

The curves are presented for the alternatives and baselines being considered for this review as tables in Appendix O Tributary Habitat Restoration, but would be better presented as figures.

(S) Attachment L.3 Egg-to-fry Survival and Temperature-Dependent Mortality

Additional material reviewed:

a) Appendix L Shasta Cold Water Pool, Attachment L.2 Sacramento River Water Temperature Analysis,
b) Appendix L Shasta Cold Water Pool, Attachment L.3 Egg-to-fry Survival and Temperature-Dependent Mortality,
c) LTO Appendix F, Attachment F.2 CVPIA Winter-run and Spring-run Life Cycle Models,
d) Appendix M Folsom Reservoir Flow and Temperature Management, and
e) Appendix N Stanislaus Stepped Release Plan

Temperature modeling below the Shasta and Folsom Dams may be important to evaluate alternatives. Two aspects include physical modeling and modeling of biological effects. Simulated temperatures below Keswick Dam and the Upper Sacramento River are used as inputs to models of biological effects produced using various approaches. Likewise, temperatures are modeled below Folsom Reservoir on the American River and below New Melones on the Stanislaus River. The approach is similar for the two tributaries and below Shasta/Keswick, so the rest of the Panel review focuses on the Sacramento River.

The results presented in the Draft Effects Analysis used HEC-5Q, a 2D model (width averaged) that represents leakage zones associated with the temperature control device (TCD) in Shasta.

The modeling, done in 2003, is described online[9].

A previous review of Temperature modeling recommended some changes[10].

Another joint agency temperature modeling effort not included in our review is the Central Valley Temperature Mapping and Prediction (CVTEMP) model[11].

Integration of this temperature model with CALSIM-3 has not yet been done, but simulated operation by ResSim-TCD was successful in demonstrating that meeting a downstream temperature target would be possible, provided sufficient cold water was available in Shasta Lake.

Below, we review approaches to modeling temperature-dependent mortality (TDM) using three different modeling approaches:

a) monthly multi-species TDM indices available for all species and life stages (described in Appendix L Shasta Cold Water Pool Attachment L.2 Sacramento River Water Temperature Analysis),
b) daily TDM models for the early life stages of Chinook runs (described in Appendix L Shasta Cold Water Pool: Attachment L.3 Egg-to-fry Survival and Temperature-Dependent Mortality (based on Anderson et al. 2020; Martin et al.), and

---

[9] https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/california_waterfix/exhibits/docs/petitioners_exhibit/dwr/part2/DWR-1084%20RMA%202003%20SRWQM.pdf

[10] https://deltacouncil.ca.gov/pdf/science-program/review-materials/2022-07-14-bureau-of-reclamation-cvp-project-operations-and-temperature-management.pdf

[11] https://oceanview.pfeg.noaa.gov/CVTEMP/river/survival.

c) monthly TDM modeling in LCMs for all life stages of Winter- and Spring-run Chinook salmon (described in LTO Appendix F – Attachment F.2 based on Peterson & Duarte 2020).

The modeling of monthly mean temperatures for the CVPIA model is described[12]. The HEC-5Q drivers are runs with a cryptic labeling scheme (e.g., AR_5QCS.dat, AR_5Q-CL.OUT). The Panel was unsure how climate change that was included in CALSIM-3 runs was included in these HEC-5Q runs. An important uncertainty could be the availability of tailwater temperature gages and models of reservoir release temperatures, but it is difficult for the Panel to evaluate this. The provided documents did not describe performance metrics measuring how well HEC-5Q represented reservoir stratification and cold-block storage nor an evaluation of the predictions from the linked reservoir–tailwater simulations.

For the purpose of estimating TDM, the Panel favors the Martin et al. (2017) and Anderson et al. (2022) models of egg-fry survival based on a daily time step. These were used to predict egg-to-fry survival for winter-run Chinook salmon as a function of temperature-dependent egg mortality, background mortality, and density-dependent mortality (but not redd-dewatering mortality, which is estimated separately). Temperature-dependent mortality (TDM) is largely understood to be caused by dissolved oxygen depletion. A stage-dependent model (Anderson et al. 2022) accounts for accumulated thermal units (ATU) to represent ATU-dependent development and to identify a critical period for thermal mortality for winter-run Chinook salmon. The Anderson et al. model was used with modifications to estimate TDM, but it's not clear that these results are fed into a life cycle model or what they were used for. The model was originally applied for all three Chinook runs together, assuming identical parameters across runs. A problem identified in the report is the covariance among parameters that makes it hard to distinguish the effects of temperature conditions from habitat limitation (e.g., superimposition) and baseline survival. It is unclear whether baseline mortality was allowed to vary among Chinook salmon runs. The Bayesian estimate of critical temperature in the Martin et al. model, $T_{crit} = 11.8^{o}$ C seems quite low and the published value to which it is compared does not have a citation. It would also be useful to see the equation.

Daily TDMs were used to estimate early life history survival for the three Chinook salmon races (described in Appendix L Attachment L.3). Advantages of these models include that the analyses are conducted using daily-resolution temperature

---

[12] https://cvpia-osc.github.io/DSMtemperature/reference/stream_temperature.html#watershed-modeling-details

drivers and that they provide stochastic results. The disadvantage is that TDMs may not represent all competing risks.

Life cycle models can represent competing risks and incorporate monitoring data, but at the current time resolution, they are not the best option for representing temperature-dependent mortality. In our view, modeling thermal risks at a monthly resolution is not adequate for making decisions based on the potentially lethal effects of short-term extreme temperature exposures. The CVPIA LCM acknowledges that it does not represent differences in egg-fry survival due to operations. It does not appear that survival effects of daily-scale variations in flow and temperature from Shasta releases were integrated into the life cycle models to inform state transitions. Therefore, it is not really possible for LTO to assess the future effects of Shasta operations that affect egg-to-fry survival.

However, because egg-fry survival involves competing risks such as redd dewatering, scouring, and superimposition, the Panel favors the use of a life cycle model to integrate risks associated with temperature and other factors and to allow comparison to available monitoring data. Ideally, the strengths of both approaches would be combined.

The best available models estimate risk at a daily resolution and by estimating physiological risk as a function of duration (see approaches by Martin et al. 2017, Anderson et al. 2020, Bowman et al. 2020, and Troia et al. 2023), represent competing risks, and make predictions that can be compared to monitoring data. Because biological effects of temperature are non-linear, the average of the survival function applied to daily values does not equal the function of the average, especially when compared to thresholds, such as maximum temperature tolerance. The Panel recommends estimating daily egg-fry survival from the highest resolution of thermal exposure data available (e.g., daily maxima), and thresholds based on laboratory studies. Daily TCM models could be refined by modeling risk using magnitude-duration relationships based on laboratory studies (see Troia et al. 2023).

Another important consideration is the role of dissolved oxygen. Because both temperature-dependent mortality and redd-dewatering mortality are thought to be driven by dissolved oxygen (DO), it would be helpful to have some idea of how Shasta/Keswick operations (including the use of the temperature control device and spill) influence tailwater DO. Whereas the focus on temperature is important, it is possible for DO and temperature to be decoupled in hypolimnetic releases. If the proximate driver for TDM and redd dewatering mortality is low DO, then further justification is needed for assuming that lower temperatures mean higher DO (due to saturation) in a tailwater setting and we are not sure whether Keswick provides aeration. In short, the Panel is asking Reclamation to justify modeling biological effects in response to long-term equilibrium temperature rather than as responses to transient dynamics in water quality based on the ability to discern differences in lethality among alternatives.

(T) Attachment O.2 Science Integration Team Life Cycle Model Habitat Estimates

The analysis attempts to answer a series of management questions. The overall question is where habitat is a primary factor influencing survival. The Appendix is not complete and points to two attachments (‘knowledge base papers’) for the methods, which are standard Instream Flow Incremental Methodology (IFIM)/WUA analyses. However, the CVPIA salmon life cycle model description also refers to the use of hydraulic models. These are not described here but rather summarized in Appendix O Tributary Habitat Restoration.

Table 1 provides a nice summary of the locations where habitat suitability criteria (HSC) (depth, velocity, substrate preference, or habitat suitability curves) were estimated for each species. We recommend a) collecting the habitat modeling into one place (or cross-referencing) and b) adding a column or map to show where the models were applied for each species.

This approach, while well-established, could be refined by representing the temporal distribution of spawners. It has also been pointed out by the Panel that modeled hydraulic (e.g., velocity and depth) information is directly available for use, which might improve the transferability of the IFIM modeling.

The Panel notes the following: Responses of spawning and rearing habitat to flow are represented either through WUA relationships or through more-detailed floodplain hydraulic models, where available. Based on looking at the code, it appears that modeling is simply based on the number of weeks that the floodplain is inundated. Additional explanation of the roles of hydraulic models and WUA relationships within the life cycle modeling is warranted.

(U) Attachment O.3 Sacramento River Weighted Useable Area Analysis

In general, the Weighted Usable Area analysis completed for the Upper Sacramento River winter-run Chinook salmon, fall-run Chinook salmon, and Central California Steelhead is based on a strong history of use and methods, and relies on data obtained from three USFWS reports, dated from 2003-2006. The spawning and rearing WUA estimates presented for the BA and EIS modeled scenarios are based on CALSIM-3 flow data for each month of the 93-year period of record.

Models were developed separately for the different races and species of salmon and assume that habitat suitability for spawning is based on substrate particle size, depth, and flow velocity. The habitat suitability characteristics for spawning were developed by taking observational data for these three metrics at active redds. Hydraulic modeling was then used to quantify the amount of suitable habitat available at different river flows, for different Habitat Suitability Criteria (HSC) levels. These results were combined to develop the WUA curves and tables, which are used to look up the amount of habitat available at different flows during the spawning periods.

Generally, the reports from which the WUA curves were developed are thorough, robust, and well-documented. Reclamation has pulled an extraordinary amount of data together and appears to have a strong, systematic approach to connecting WUA curves to the alternatives and baselines. However, without a stronger narrative that synthesizes all the WUA analyses and the methodology behind them, the Panel found it difficult to track methodologies across river systems and species to understand where, if any, differences occur and how that should be handled in assessments of effects on ESA-listed species.

The Panel recommends that a table that outlines whether the WUA curves for a specific river/species/life stage and reach were developed using PHABSIM vs. RIVER2D would be helpful, as well as any substitutions of data from one species or river system to another. Limitations, uncertainties, and tradeoffs between the underlying models should be discussed (i.e. for systems using PHABSIM, an increased level of uncertainty should be identified since this model has multiple identified issues).

No WUA curves were developed for spring-run Chinook spawning, so fall-run Chinook salmon WUA curves were used to quantify spring-run spawning habitat. This gives the Panel some cause for concern, as there is no discussion or evidence presented that discusses the degree to which this surrogacy is appropriate. Reclamation cites a personal communication to Mark Gard via email that USFWS staff endorsed this practice, and it has since been adopted into subsequent studies. Without further information, the Panel cannot comment as to the appropriateness of this endorsement. Interestingly, it seems spawning peaked at 700-900 cfs for spring-run but at 5,000 cfs for fall-run Chinook salmon. Does this reflect the preference of fall-run Chinook for deeper mainstem habitat and of spring-run Chinook for tributary habitat?

The HSC data (responses to depth and velocity) used to develop WUA for Steelhead trout were obtained from the American River. Local HSC data for Sacramento Steelhead were not available. The Panel suggests that Reclamation provide evidence that this space substitution is appropriate.

There is a section that states that the upper and lower limits of the range in WUA values are determined by the ranges of the fry-rearing WUA curves from which they are estimated (i.e. the curves differ across sections of the river for the same flow). However, there are no results that show the upper or lower ranges – only one WUA value is shown, and there is no explanation of how the upper and lower limits were incorporated into the modeling framework or represented in the output tables and figures for each alternative scenario.

The Panel notes that the analyses tend to focus on the peak of the spawning season; care should be taken to ensure the tails of the spawn timing distribution are carefully considered, as phenotypic variation in run-time is a specific objective of Recovery Planning. When effects analyses and resulting management actions focus on the spawning peak, it may have the effect of reducing genetic or phenotypic diversity.

The data used to inform spawning and rearing WUAs is from the late 1990s and early 2000s. The Panel questions whether the years in which these data were collected are representative of the range of environmental conditions experienced in the river systems presently, or whether the conditions are representative of wet, dry, etc. year types, or whether the river-scape is similar enough now. A source of uncertainty is whether or not the placement of redds and the biological effects of stressors on early life stages are similar across water-year types.

Attachment O.3 provides a map of the section of the Sacramento River wherein WUA curves were developed. However, in this attachment, and all other WUA attachments, it is difficult to discern the spatial extent to which the WUA curves are being applied. Are they extended beyond the original stream reaches for comparison among scenarios? If so, is that appropriate in terms of similar channel shape and flow magnitude? If not, can the HSC relationships be applied instead to ensure that the effects of alternatives at the population (or sub-population) level are fully understood?

Differences in mean spawning WUA curves across the management scenarios are presented by month, water-year type, and for all water-year types combined (for each species/race), and account for differences in density across different river segments. This is a careful, strong effort that ensures temporal and spatial differences are considered. The Panel suggests an explanation of how the Reclamation plans to integrate this degree of complexity into the life cycle models.

For months when the Anderson-Cottonwood Irrigation District (ACID) Dam boards (barriers to movement) are installed, how is WUA adjusted other than by flow?

Reclamation acknowledges that many fry and steelhead rear downstream of Red Bluff Diversion Dam, where no WUA curves have been developed.

The report repeats a section (Sacramento River Rearing) starting on page 11.

A limitation identified for all the WUA curves is that all of the habitat-based studies assume the channel characteristics of the river during the time of field data collection are in dynamic equilibrium. This report confusingly states the field data collection is by USFWS from 1995-1999, but also discusses redd surveys through 2021. This statement should be clarified as to whether it refers only to the rearing data, or also to the redd data.

The report also identifies that WUA curves were applied as far downstream as Red Bluff Diversion Dam, although known rearing occurs below this point. It is unclear how this data gap is handled in the integration of the effects for Chinook and Steelhead, but this is a significant portion of the Sacramento River wherein rearing habitat appears to be unevaluated with respect to the alternatives. This may also affect how the benefits of floodplain habitat are represented, which growth, food web, and bioenergetics studies suggest is critical to early marine survival. The Panel suggests that this omission be addressed in the text.

Estimated WUA values for the Draft Effects Analysis are similar for May, June, and August, but are lower and more variable in July, when CALSIM-3 flows are higher. The higher July flows are described as substantially higher, and possibly contributing to lower and more variable spawning WUA results. The Panel is unable to evaluate this rationale without a better explanation in the document. The July flows should be apparent in the document so that a reader could go back to the WUA curve and identify whether this statement is reasonable, and also determine which sections of the river are likely driving the reduction in spawning WUA.

Similar to spawning, the highest WUA for rearing habitat (fry and juveniles separately) were generally higher under critically dry years and lower in above-normal or wet water years. The largest difference between the No Action Alternative and the scenarios is a 0.7% increase for Alternative 4 in critical water years. The largest reduction is 0.5% for Alternative 2 without UCP Systemwide VA in below-normal water years. Across months, estimated fry-rearing WUA curves are similar (August through December), except extreme WUA results, particularly high values, which are more prevalent for December, and somewhat prevalent in November. This is presumably due to more frequent high flows, which correspond to higher WUAs on the winter-run Chinook WUA curve. Winter-run Chinook is the only Sacramento River salmon race to show this pattern.

There is a section that states that the upper and lower limits of the range in WUA values are determined by the ranges of the fry-rearing WUA curves from which they are estimated (i.e. the curves differ across sections of the river for the same flow). However, there are no results that show the upper or lower ranges, only one WUA value.

Perhaps there is a module within the RIVER2D model that accounts for this range in WUA values, but it would seem to be better to apply the appropriate HSC derived for each river section to the river sections under each scenario, and then roll those up into a composite value. This section is confusing to understand, and clarity is needed about how different sections of the river were handled in the modeling process.

(V) Attachment I.6 Volumetric Influence Analysis

This analysis is too coarse and based on a simplistic, flawed assumption. The Panel's primary concern is that decision-makers will look at these results and conclude that there are no significant differences between the alternatives. As an example, one of the stated EIS Key Takeaways is: *"Among the other alternatives there is great overlap in the distribution and variation among where the peaks in the distribution among the water types are observed"* (p. I.6-3).

Output from CALSIM-3 is not the correct source of data for this analysis. More appropriate models of water circulation within the Delta exist and should be used when analyzing issues related to Delta hydrodynamics.

The Volumetric Influence Analysis assumes that the full inflow from the Sacramento and San Joaquin Rivers is available for export. The basic assumption of the analysis that the Delta can be represented as a giant Continuously Stirred Mixed Reactor is hydrodynamically wrong. In reality, the majority of Sacramento water stays in channels of the Sacramento River. The amount of Sacramento sourced water available for export is limited to the amount that is transferred to the San Joaquin River and the South Delta via the San Joaquin/Mokelumne junction, the Three Mile Slough junction, and the Sacramento/San Joaquin confluence.

To estimate the percent of Sacramento-sourced river that is transferred to the San Joaquin and South Delta and available for export, one would take any Delta hydrodynamics model (DSM2, etc.) and calculate how much Sacramento-sourced water is going through the three above-mentioned junctions. As an extremely rough estimate, calculate the Sacramento flow just below the Sacramento/Georgina Slough junction as the amount that is not transferred to the San Joaquin and South Delta region.

For this analysis, the total volume of inflow water available for export is not the full Sacramento and San Joaquin flow. It is the San Joaquin inflow plus a percentage of the Sacramento flow based on Delta hydrodynamics and the Delta Cross Channel operations.

Many conclusions were stated at the end of the Volumetric Influence Chapter. We will discuss here the validity of several key conclusions and how the assumptions underlying the analysis contributed to the conclusions.

First, the statement is made: "*the frequency of specific observations can be used to qualitatively assess which alternatives have the most observations of low percent Delta inflow*" (p. I-6-3). The Panel asks: Is it a low percent of the actual inflow that is available to be exported at the pumps? A large amount of water can be headed through the Sacramento system, creating a large total Delta inflow. However, that large Sacramento flow is not hydrodynamically available at the export pumps.

Second, the statement: "*Among the other alternatives there is great overlap in the distribution and variation among where the peaks in the distribution among the water types are observed*" (p. I.6-3). The Panel reads this to mean that it was hard to tease apart differences in these different simulations. That is because the basic premise of the analysis is incorrect. Only a limited percentage of the Sacramento-sourced water crosses over to the South Delta and is available for export.

Third, the statement is made: "*Inflow groups with high Sacramento River flows have a large amount of overlap in their distribution among the alternatives*" (p. I.6-19). The

Panel notes that this is because high Sacramento flow is the case where the Total Sacramento plus Total San Joaquin inflow assumption works the least.

Fourth is the statement: "*A similar pattern occurs in the lolo and lomed groupings, however medmed (medium-medium), and medhi (medium-high) have distinctly high frequencies between 20% and 30% in Alt2woTUCPwoVA*" (p. I.6-19). The Panel notes that the medhi category means that Sacramento is a medium-level flow, and the San Joaquin is a high-level flow. So, the San Joaquin-sourced water has a high influence on export in these cases. Therefore, the basic assumption of the analysis is not swayed as much compared to when Sacramento flow is high. In the case where San Joaquin water dominates inflow volume, signals that distinguish between the alternatives show up.

Fifth is the statement: "*The lohi group is unlike any other group by having its highest peak in Alt2woTUCPAllVA, but this is likely driven by the sample size. The NA group is introduced because of some of the observed values falling outside of the delta inflow group definitions*" (p. I.6-19). The Panel thinks that this finding could also be related to the assumptions of the analysis rather than sample size. This is the only case where the dominant water source is the San Joaquin. Therefore, the results are not diluted by the wrong assumptions related to Sacramento water volume. Here, a signal that distinguishes between different alternatives shows up.

The Panel also had comments on some of the conclusions (EIS, p. 3). One conclusion was that the lowest percent diverted occurred in an above-normal year. The Panel interprets this as simply that if you have more water in the system, the percentage of the Delta inflow exported would be lower. Another conclusion was that the maximum value of 65% Delta inflow exported in multiple years. The likely explanation is that this is because the regulations that limit the CALSIM-3 export operations kicked in. Finally, the observed lowest (non-zero) mean percent Delta inflow exported was in Alt 3 and observed in the hihi inflow group at 6.7%. This made sense to the Panel because if both rivers have a high volume, you would expect that the percent diverted would be the lowest.

Hydrodynamic models (and associated PTMs) need to be used to tease apart the results. Even the basic DSM2 (1-D channel assumption) puts in necessary detail. The CALSIM-3 hydrology model can be used to drive the flow boundaries of Delta hydrodynamic models. However, CALSIM-3 should not be used to make conclusions about the influence of South Delta export facilities.

(W) Attachment F.3 CVPIA Winter-Run Life Cycle Model

Additional material reviewed:

a) Appendix I – Old and Middle River Flow Management, and
b) Model Description for the Sacramento River Winter-run Chinook Salmon Life Cycle Model[13]
c) The Winter-run Chinook Salmon Life Cycle Model (WRLCM) is a spatially and temporally explicit stage-structured, stochastic simulation model that estimates the number of winter-run Chinook salmon at each geographic area and time step for all stages of their life cycle. Hatcheries have been added (although we did not see information about the assumptions used, which could be important). The model uses the Newman (2003) Bayesian state-space model to estimate juvenile Delta survival based on biological and environmental covariates. The WRLCM has been reformulated into a state-space model. Many of the stage transition equations describing the salmon life cycle are direct or indirect functions of water quality, depth, or velocity from DSM2 outputs, thereby linking management actions to the salmon life cycle. The approach to modeling early survival in response to temperature effects is more sophisticated than that in the CVPIA model. Although the model represents linkages between the survival of each life stage and flow-temperature drivers, in some cases the temporal resolution is too coarse (e.g., egg-to-fry survival is based on 3-month temperature averages).

The WRLCM now uses the enhanced particle tracking model (ePTM) to estimate the survival of out-migrating smolts originating from Lower Sacramento River, Delta, and floodplain habitats. ePTM represents two stressors: predation and diversions. Calibration of ePTM is based on the survival of released coded-wire-tagged (CWT) juvenile hatchery fall-run Chinook salmon and their recovery at Chipps Island, which do not leave the system at the same time as winter-run juveniles. The WRLCM is calibrated against spawners below Keswick Dam, juveniles collected at Red Bluff Diversion Dam, juveniles collected in Knights Landing catches and rotary screw traps, and Chipps Island abundance. The Panel would like to see more justification for using hatchery fall-run Chinook as a surrogate.

How well does the model represent project operations? An earlier review[14] concluded that WRLCM incorporated the needed linkages between project operations and population dynamics, as determined by the distribution, survival, and movement of salmon within the river system. However, the review also

---

[13] Model Description for the Sacramento River Winter-run Chinook Salmon Life Cycle Model (noaa.gov)
[14] Winter-run Chinook Salmon Lifecycle model (noaa.gov)

questioned the sensitivity of the model to flow alternatives, and covariation between outflow, temperature, and spawner returns, which was a concern of the earlier Newman model, a predecessor to this model. One concern is that historical correlations may be broken in the future, changing the ability of the model to predict population-level effects of CVP decisions. The WRLCM model was used to compare scenarios releasing colder water in spring. The analysis found an important seasonal tradeoff between early pulse flows and maintaining the cold-water pool at Shasta until September. Elevated September temperatures resulted in poorer simulated outcomes for winter-run Chinook salmon (Appendix L Shasta Cold Water Pool Management). Hendrix suggests that, because of its design, the model may be overly sensitive to operations.

The WRLCM model is not uniquely specified in its parameters, reflecting the reality that multiple combinations can produce the same outcomes. It may therefore be unable to distinguish between different combinations of management decisions leading to the same fit against downstream monitoring data. One solution to this is to use Bayesian multi-parameter modeling (see Piou et al. 2009; Jager 2013). To avoid over-specification, we also agree with the decision to minimize the complexity of spatial representation, for example, in ePTM. The best solution to the model-identification issue is to find ways (data) to distinguish hypotheses about the response of salmon to specific operations higher up in the system, so this is an important data gap that should be addressed.

(X) Attachment I.3 Delta Export Zone of Influence Analysis

Additional material reviewed: Chapter 5 Winter-Run Chinook Salmon (p. 5-77 to 5-89)

The Panel identified three issues for Reclamation to consider:

The first issue is the approach for the channel length analysis dilutes the results to the point that the differences between alternatives are obscured. The base channel length for the calculation of the percentage should be the length of the channels in the south Delta and the San Joaquin River. Including the length of the Sacramento River and associated tributaries north of the San Joaquin River unnecessarily dilutes the results.

Second, the Panel agrees that future analysis should anchor the results to export volume rather than OMR values. "*Future work may visualize results by exports instead of OMR to better understand if exports are a more direct driver of the* *spatial extent of the zone of influence*" (LTO Appendix I Old and Middle River Flow Management, Attachment I.3, Delta Export Zone of Influence Analysis, p. I.3-36).

The third issue is that there are multiple placeholders in this section for particle tracking results. It is unfortunate that the Panel did not have an opportunity to review those sections. Placeholder Sections include: Flow into Junctions, Particle Tracking Models, ECO-PTM (BA Chapter 05, p. 5-90).

The introduction of the topic and approach is much better in Appendix I than in the BA – Winter Run Salmon section. Much of this introduction is needed in the BA Winter Run Salmon section to explain the results. To enhance understanding of the concept, provide a map of the stations for Figures 20-22. Also, make sure the x-axis in the figures is consistent – the direction for Figure 22 is flipped. (BA-Winter Run Salmon p. 5-79 through 5-81).

Appendix I did not have definitions of what Alt2wTUCPwoVA, Alt2woTUCPAllVA, Alt2woTUCPDeltaVA stand for. This information may be buried in other documentation. However, since these simulations are trying to compare the results of combinations of alternatives and climate change, it would be helpful for the reader to get a one-paragraph summary with a reference to the expanded simulation documentation at the start of this discussion.

The interpretation of the contour plots raised some questions for the Panel. The panel agrees that the Sacramento and San Joaquin inflow categories (hi, med, lo) should be used to divide the results into flow groups. The information in the contour plots (Figures 6-12) is presented well and makes intuitive sense. All the figures show the same general patterns and provide a basic lesson on how the export pumps influence the South Delta region.

Note that the region of influence is always on the export pump side of the San Joaquin River. This is important evidence supporting the Panel's comments regarding the Volumetric Influence analysis and the Channel Length analysis.

Gaussian Kernel Density Estimation (KDE) Proportional Overlap Contour Maps: Since the Delta is a tidal system, for fish entrainment, the range of velocity magnitude the fish experience over a tidal cycle matters. The changes throughout the tidal cycle influence how fish navigate through the connected labyrinth of channels in the South Delta. For example, Figure 4 (p. 12) shows a very distinct difference in the velocity fields over the tidal cycle at Station Old River and Middle River in the 'with pumping' and 'no pumping' scenarios. Note that the average median velocity over a tidal cycle may change very little (e.g. Figure 4, Velocity Differential = 0.14 fps) between the pumps on vs. pumps off scenario. But, fish entrainment is a tidal process, not a tidally-averaged process.

Sacramento River channel lengths of the Delta should not be included in a calculation of the zone of influence. The “Channel Length of Delta” figures of the Proportion of total channel length in the Delta (e.g. Figures 13-21) are diluted significantly from the actual results. There is no physical basis for including channels all the way up to the I Street Bridge on the Sacramento, Mokelumne System, Sutter Slough, Georgiana Slough, Yolo Bypass, and the Deep Water Ship Channel in a “Zone of Export Pump Influence” calculation. Likewise, it is not clear whether other channels in the Suisun Bay region were also incorporated into this calculation.

The results of all the zone of influence charts clearly demonstrate that the zone of influence of pumping rates is associated with all the channels on the export pump side of the San Joaquin River. The Panel would argue that there is also an influence of pump operations at channel junctions along the San Joaquin River. However, these results clearly show that Sacramento channels (all tributary channels north of the San Joaquin River) are not part of the zone of influence of the export pumps.

The modeling group responsible for the Gaussian KDE analysis needs to do much more synthesis of results. A map is necessary to show where the stations are for Figures 20-22. Figure 22 (p. 5-80) has the x-axis scale flipped compared to the other two figures. This axis should be consistent to promote understanding of the underlying mechanisms causing the velocity shifts.

Proportion of total channel length analysis: Fix Figure 24 (which channels should be included) before doing the analysis in Figure 29. The signal has been too diluted to give meaningful results.

(Y) Attachment J.3 – Zooplankton-Delta Outflow Analysis

Additional material reviewed:

a) BA Chapter 9 Delta Smelt (9-20 to 9-95), and
b) LTO Appendix J – Winter and Spring Pulses and Delta Outflow (Section 6.7, p. 55-57; pdf p. 63-65)

The Panel identified multiple issues regarding the Zooplankton-Delta Outflow analysis. There are issues both with hydrodynamics and biological aspects of the analysis.

While Reclamation included appropriate caveats on how to interpret these results in the texts, the Panel wants to emphasize these are major uncertainties and not just the usual cautions. They go to the foundation of the analysis. Some examples are:

> *"Yet the mechanism for why CPUE increases in the low salinity zone during higher outflow has not been clearly and definitively established. Kimmerer (2002) found lower tropic level taxa (zooplankton) responded inconsistently with flow across seasons and historical periods. Kimmerer also found that chlorophyll showed little response to flow, suggesting a bottom up, "agricultural model" explanation for increased CPUE with higher flows is unlikely" (LTO Appendix J, Attachment J.3, p. J.3-5).*
>
> *"Another possible mechanism is that increased flow also increase subsidies of zooplankton from higher abundance freshwater regions into the LSZ (Hassrick et al. 2023, Kimmerer et al. 2019)" (LTO Appendix J, Attachment J.3, p. J.3-6).*
>
> *"A historical regression of zooplankton CPUE with flow may be too simple and including other factors such as salinity, temperature, chlorophyll-a, residence time, etc. may have more explanatory power" (LTO Appendix J, Attachment J.3, p. J.3-3 – J.3-4).*

The Panel wants to be sure that these caveats and others permeate throughout the entire document, especially in the interpretation of results. The Panel suggests that an important question to answer as part of the analysis is: What magnitude of Delta outflow is needed before a significant shift in zooplankton response can be observed?

The variability in the simulation flow data from CALSIM-3 is averaged out of the analysis. This analysis should use a seasonal averaging approach rather than an annual averaging approach. As a result, the differences between the different Baseline Conditions and Alternatives are muted in results.

To create the regression curves, Delta outflow historic data from 2000-2021 from the DAYFLOW database was used. Delta outflow is available on a daily time step in this database. For this analysis, "for each taxon, mean annual log-transformed catch per unit effort + 1 was regressed against **mean annual log$_e$-transformed Delta outflow** for each season period" (LTO Appendix J, Attachment J.3, p. J.3-2, bold added).

Instead of regressing the catch per unit effort +1 values against the Delta outflow representative of the months of the season of interest (Spring, Summer, Fall), Reclamation used the average Delta outflow over the entire year. This means annual Delta outflow may or may not be representative of the Delta outflow during the season when the sampling took place. A better alternative would be to find a seasonal mean Delta outflow for Spring (March-May), Summer (June-August), and Fall (September-November).

The analysis states that the regressions "*were then applied to the 1922-2021 CALSIM-3-modeled data for Baseline Conditions and Proposed Action scenarios, with predictions back-transformed to the original measurement scale for summary of results*" (LTO Appendix J, Attachment J.3, p. J.3-2). The CALSIM-3 produces Delta outflow results on a monthly time step but the regression analysis does not incorporate this flow variability available in the dataset. Instead, for each year in the simulation, the CALSIM-3 Delta outflow results are averaged over the entire year to produce a mean annual Delta outflow. If the regressions in the pre-processing step used seasonal mean Delta outflow instead of mean annual Delta outflow, the flow variability in the CALSIM-3 results could be represented.

For the figures in the document (LTO Appendix J, Attachment J.3, p. J.3-17 – J.3-34), the 1922-2021 data was binned by water-year type. There is only one CPUE value per year of the water-year type represented. The range bars represent the variability over the entire simulation of 1922-2021. Here again, there is more variability in the flow dataset that is not being incorporated in the synthesis charts.

Chapter 9 discusses how the Proposed Action *may increase food availability stressors*. The mechanism is that the proposed storage and diversion of water associated with the Proposed Action will reduce Delta inflows and outflows. Delta smelt feed on calanoid copepods (*E. affinis* and *Sinocalanus doerrii*), which exhibit a positive correlation with Delta outflow in spring. Pulse spring flows in dry water years can increase copepod biomass near Suisun Bay. Thus, if the Proposed Action reduces pulse spring flows or Delta outflows, there will be a subsequent reduction in zooplankton prey. This needs to be discussed in some detail. For example, the frequency of an increase in the stressor (low food availability/use Delta outflow as a proxy) is likely high (78% of years were low spring outflow; 81% of years were low winter outflow).

The two studies cited as "multiple studies" in Chapter 9 are Merz et al. (2011) in the white literature/non-peer reviewed and the other is a non-species specific zooplankton flow analysis model. Given that Delta smelt are specific in their prey selection, the analysis should try to be as species-specific as possible to accurately assess food as a stressor.

The food availability stressor links to Delta outflow-zooplankton regressions. The Panel notes that all life stages in the Delta Smelt Chapter 9 are treated the same: same rationale, data, and explanation is given linking Delta smelt to zooplankton. The Panel suggests that more description and rationale be provided about the datasets used for each life stage, as some datasets are more complete and relevant than others. It is important to indicate why certain data sets and zooplankton species were considered. The documentation should be sufficient so that the reader can determine that the species and life stages of the zooplankton are appropriately applied to the life stages of the Delta smelt.

A table is needed to show Delta outflow and % change in Delta outflow with corresponding species-specific zooplankton CPUE and % change in zooplankton CPUE. The Panel was unable to gauge the sensitivity of the zooplankton response among scenarios without a table of this type. Reclamation provides zooplankton output tables, but not in the context of the Delta outflow – just the water-year type and the scenario type. The reader must assume there are differences in the amount of Delta outflow, but this information is not explicitly available in the Zooplankton-Delta Outflow Appendix J Attachment J.3 methodology, or in Ch. 9 (Delta Smelt).

For simulated Delta outflow from CALSIM-3, is there one Delta outflow metric per season that holds steady (Appendix J, Attachment J.3)? What are the 95% CIs of that flow per season, and how sensitive are the zooplankton responses across the range of CALSIM-3 Delta outflow results?

The methodology appears to only use ONE parameter (Delta outflow) to explain Delta zooplankton. Multiple papers (mentioned in the materials) identify other factors that are predictive of Delta zooplankton production. As a result, many prevalent prey items are excluded from the analysis because they do not have a significant relationship with Delta outflow. The Panel would like to see how Delta outflow relates to these factors (e.g. turbidity, spatial location, temperature, actual salinity residence time, etc.). While Delta outflow is likely correlated with these factors, including each parameter and being able to test the sensitivity of the zooplankton response to those parameters may be valuable. Choosing the parameters to include in the best-fit model would require running an Akaike Information Criterion (AIC) analysis.

Zooplankton export from productive ecosystems near Liberty Island, Cache Slough Complex, Barker Slough, and the Deepwater Ship Channel has been shown to influence zooplankton densities in the Delta and is positively correlated with Delta outflow. This is mentioned in the text and seems like it could be incorporated into the modeling effort herein to understand how upstream zooplankton production might be conveyed to the Delta low salinity zone (LSZ) habitat for consumption, and how that production and conveyance might change across the Proposed Action and scenarios. This linkage is missing and would require linking an upstream/freshwater habitat zooplankton/flow model to Delta outflow.

Plankton production is often linked with shallow shoals and marsh edge habitats. How do Delta outflows for each alternative map the preferred salinity and turbidity ranges of Delta smelt onto the preferred habitats of smelt at different life stages? What percentage of the shallow water and fringing marsh habitats available in the Delta, which produce zooplankton, are captured in the low/preferred salinity zone?

The Panel was confused by the tables (Tables 1-17, Appendix J, Attachment J.3). For each set of paired tables per species, there needs to be a measure of Delta outflow.

No measures of uncertainty are reported in these Tables (i.e. % change in CPUE in *E. affinis*). Each prey species is reported individually, but fish can consume a range of prey types. It would also be good to understand the total zooplankton CPUE response. Overall, if some are increasing and others decreasing, is the general population holding steady across scenarios? Decreasing? Increasing?

Figures show a high degree of uncertainty with error bars. There is no explanation provided on whether they are 95% CIs or something else. Given the error bars, it would appear that most scenarios are not significantly different from one another. That said, the summary suggests that in Critical Years, alternative action is better for zooplankton action than No Action. This should be clarified, as Alt2woTUCPwoVA does not improve food web conditions in the spring (but you can only really see that for “other calanoid copepods”, and this size class is actually the one to pay attention to for larval-juvenile Delta smelt).

Modeling zooplankton species only represents a proportion of Delta smelt diets (i.e. only those with a significant relationship to Delta outflow). It would be good to have a table showing what % of Delta smelt diets are represented by the prey items in the models. For example, a 15% decrease in a species rarely consumed has a very different impact than a 15% difference in a preferred prey species.

Note on Table 13 (p. J.3-14 of Appendix J, Attachment J.3) - "*Provide biological, ecological, and operational explanation for the observation".* That has not happened yet in the table explanation of the results but the Panel concurs that these comments should be included.

The document (Appendix J, Attachment J.3) discusses regression equations from Hennessy and Burris 2017, which predict *E. affinis* and *N. mercedis*, and *P. forbesi* to mean June-September Delta outflow. Reclamation decided these regressions were geographically too simplistic and too temporally broad for applying to the effects of operations (Appendix J, Section 5.3.2.2, p. 36). However, the Panel did not see how the approach used by Reclamation was a major advancement for being more geographically complex. And further, with the approach that was used, *P. forbesi* was dropped from the equations because there was not a statistically significant relationship with the Delta outflow.

The Panel suggests that Reclamation consider adding a spatial component to the analysis. The results, as presented, provide an analysis of potential changes to the food web for the Proposed Action and each of the alternatives binned by water year. Further dividing responses spatially might be helpful, and that could complement if a seasonal reporting of Spring, Summer, Fall, and Winter is also done. (This is because Delta smelt use different areas of the estuary in different percentages, and it would be important to know if the areas where zooplankton are most likely to change (+ or -) overlap with the areas with the most Delta smelt.) What spatial areas might become more limited? What areas might become less limited?

The following statement is made in Appendix J: "*In the spring CPUE is LESS under alternatives compared to NAA for all but critical water years. Thus, alternatives appear to provide benefits over NAA to smelt in the spring of critical water years only.*" This is an important result. However, the Panel does not see how the figures in Appendix J support this conclusion. Spring CPUE for the NAA vs Alternatives is variable by the Alternative, water year, and species. Mostly there is strong overlap, and it would be difficult to statistically and ecologically distinguish differences. The Panel does note that Alt2woTUCPAIIVA appears better than the NAA for spring *E. Affinis* across multiple water-year types.

(Z) Attachment M.1 American River Redd Dewatering Mortality

Additional material reviewed:

a) Appendix M Folsom Reservoir Flow and Temperature Management, and
b) Appendix O Tributary Habitat Restoration

Modeling dewatering risk applies to Steelhead and fall-run Chinook salmon below Nimbus Dam on the American River. Historic Steelhead redd dewatering is mentioned (Appendix M, section 6.7), along with a comment that it has not been reported in recent years. The modeling approach described for the American River is reported in Appendix O Tributary Habitat Restoration as part of the Redd Dewatering Analysis (section 4.7.2). The analysis for the Stanislaus River is missing (a placeholder) in Appendix O and there does not seem to be a corresponding description for the Upper Sacramento below Keswick Dam, although dewatering is referred to in the LCMs.

The approach described in Appendix M for Steelhead trout assumes that dewatering mortality occurs when monthly flows decrease from those at the time (month) of spawning. The approach is very coarse in some respects. For example, to represent a 2-month period of potential risk, only the maximum flow reduction based on monthly flow is used to estimate dewatering mortality. If the Panel understands this correctly, this would give only two numbers to compare, the value in the first month and the value in the second month. In addition, temperature, which influences the rate of egg and larval development (but is also moderated by groundwater), is neglected.

In other respects, the analysis is detailed and site-specific, considering the distribution of spawning times based on redd counts and estimates of redd depths for each reach. It is unclear whether this spawning time "distribution" is also across just two months. The distribution of redd depths also influences the estimate of mortality (Table M.1-1). A minor suggestion is that Table M.1-1 would be more concisely presented as a graph of cumulative curves. In addition, an equation is needed to communicate exactly how the estimate (and its variance) is produced. The approach estimated that the proportion of redds dewatered was fairly high, especially in wetter years.

The Panel recommends:

a) tracking daily changes and the duration of exposure of redds, giving credit for rewetting,

b) consider using HEC-5Q daily output to allow simulation of development and differences between eggs and alevins,
c) providing information about model validation, and
d) presenting an equation for mortality.

Ultimately, it will be important to combine dewatering with competing risks, such as temperature-dependent mortality, superimposition, and redd scouring. The Panel appreciates that the results show the full distribution of percentages to allow a comparison among alternatives.

11. Literature Cited (including papers cited in the Appendix)

Anderson, J.J., Beer, W.N., Israel, J.A., and Greene, S. 2022. Targeting river operations to the critical thermal window of fish incubation: Model and case study on Sacramento River winter-run Chinook salmon. *River Research and Applications* 38: 895-905.

Basson, M. 1999. The importance of environmental factors in the design of management procedures. *ICES Journal of Marine Science* 56: 933-942.

Blakley, J.A., and Franks, D.M. (eds.) 2021. *Handbook of Cumulative Impact Assessment*. Edward Elgar Publishing.

Burnham, K.P., and Anderson, D.R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd Edition. Springer-Verlag, NY.

Caughley, G. 1994. Directions in conservation biology. *Journal of Animal Ecology* 63: 215-244.

Collie, J.S., Botsford, L.W., Hastings, A., Kaplan, I.C., Largier, J.L., Livingston, P.A., Plagányi, É., Rose, K.A., Wells, B.K., and Werner, F.E. 2016. Ecosystem models for fisheries management: finding the sweet spot. *Fish and Fisheries* 17: 101-125.

Crain, C.M., Kroeker, K. and Halpern, B.S., 2008. Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters* 11: 1304-1315.

Diefenderfer, H.L., Steyer, G.D., Harwell, M.C., LoSchiavo, A.J., Neckles, H.A., Burdick, D.M., Johnson, G.E., Buenau, K.E., Trujillo, E., Callaway, J.C., and Thom, R.M. 2021. Applying cumulative effects to strategically advance large-scale ecosystem restoration. *Frontiers in Ecology and the Environment* 19: 108-117.

Drinkwater, K.F., and Myers, R.A. 1987. Testing predictions of marine fish and shellfish landings from environmental variables. *Canadian Journal of Fisheries and Aquatic Sciences* 44: 1568-1573.

Fish, M. 2010. A white sturgeon year-class index for the San Francisco Estuary and its relation to Delta Outflow. *Interagency Ecological Program Newsletter* 23(2):80–84.

Foley, M.M., Mease, L.A., Martone, R.G., Prahler, E.E., Morrison, T.H., Murray, C.C., and Wojcik, D. 2017. The challenges and opportunities in cumulative effects assessment. *Environmental Impact Assessment Review* 62: 122-134.

Freeman, M.C., Bestgen, K.R., Carlisle, D., Frimpong, E.A., Franssen, N.R., Gido, K.B., Irwin, E., Kanno, Y., Luce, C., Kyle McKay, S., Mims, M.C., Olden, J.D., LeRoy Poff, N., Propst, D.L., Rack, L., Roy, A.H., Stowe, E.S., Walters, A., and Wenger, S.J. 2022. Toward improved understanding of streamflow effects on freshwater fishes. *Fisheries* 47: 290-298. https://doi.org/10.1002/fsh.10731

Gargett, A.E., Li, M., and Brown, R. 2001. Testing mechanistic explanations of observed correlations between environmental factors and marine fisheries. *Canadian Journal of Fisheries and Aquatic Sciences* 58: 208-219.

Giroux, M., and Schleck, D. 2021. The effects of temperature and salinity on the endocrinology in two life stages of juvenile rainbow/steelhead trout (*Oncorhynchus mykiss*). *Journal of Fish Biology* 99: 513–523.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D’Agrosa, C., Brunoa, J.F., et al., 2008. A global map of human impact on marine ecosystems. *Science* 319: 948–952.

Halpern, B.S., and Fujita, R. 2013. Assumptions, challenges, and future directions in cumulative impact analysis. *Ecosphere* 4: 1–11.

Halpern, B.S., Frazier, M., Potapenko, J., Casey, K.S., Koenig, K., Longo, C., Lowndes, J.S., Rockwood, R.C., Selig, E.R., Selkoe, K.A., and Walbridge, S. 2015. Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nature Communications* 6.

Hammar, L., Molander, S., Pålsson, J., Crona, J.S., Carneiro, G., Johansson, T., Hume, D., Kågesten, G., Mattsson, D., Törnqvist, O., and Zillén, L. 2020. Cumulative impact assessment for ecosystem-based marine spatial planning. *Science of the Total Environment* 734, p.139024.

Hearn, A.R., Chapman, E., Singer, G., Brostoff, W., LaCivita, P., and Klimley, A.P. 2014. Movements of out-migrating late-fall run Chinook salmon (*Oncorhynchus tshawytscha*) smolts through the San Francisco Bay Estuary. *Environmental Biology of Fishes* 97: 851-863.

Heublin, J., Bellmer, R., Chase, R., Doukakis, P., Gringas, M., Hampton, D., Israel, J., Jackson, Z., Johnson, R., Langness, O., Luis, S., Mora, E., Moser, M., Rohrback, L., Seesholtz, A., Sommer, T., and Stuart, J. 2017. Life history and current monitoring inventory of San Francisco Estuary sturgeon. NOAA-TM-NMFS-SWFSC-589.

Hittmeir, S., Philipp, A., and Seibert, P. 2018. A conservative reconstruction scheme for the interpolation of extensive quantities in the Lagrangian particle dispersion model FLEXPART. *Geosciences Model Development* 11: 2503–2523.

Hope, B.K., and Clarkson, J.R. 2014. A strategy for using weight-of-evidence methods in ecological risk assessments. *Human and Ecological Risk Assessment* 20: 290-315.

IEP MAST. 2015. An updated conceptual model of Delta Smelt biology: our evolving understanding of an estuarine fish. Interagency Ecological Program, Management, Analysis, and Synthesis Team. Technical Report 90.

ISAB (2023) Review of spring chinook salmon and steelhead models for assessing responses to EIS alternatives for the Willamette Valley system. Independent Scientific Advisory Board, SAB 2023-1, March 3, 2023.

Jager, H., 2013. Documentation of a linked-population viability model (LPVA) to evaluate recovery options for Snake River Fall Chinook Salmon. ORNL/TM-2013/183. https://info.ornl.gov/sites/publications/files/Pub42661.pdf

Jäger, S., Allhorn, A., and Bießmann, F. 2021. A benchmark for data imputation methods. *Frontiers in Big Data* 4:693674. doi: 10.3389/fdata.2021.693674

Kimmerer, W.J., and Rose, K.A. 2018. Individual-based modeling of delta smelt population dynamics in the Upper San Francisco Estuary III. Effects of entrainment mortality and changes in prey. *Transactions of the American Fisheries Society* 147: 223-243.

Keyl, F., and Wolff, M. 2008. Environmental variability and fisheries: what can models do? *Reviews in Fish Biology and Fisheries* 18: 273-299.

Kimmerer, W.J. 2008. Losses of Sacramento River Chinook salmon and delta smelt to entrainment in water diversions in the Sacramento–San Joaquin Delta. *San Francisco Estuary and Watershed Science* 6(2).

Kolstad, C.D., and Moore, F.C. 2020. Estimating the economic impacts of climate change using weather observations. *Review of Environmental Economics and Policy* 14: 1-24.

Linkov, I., Loney, D., Cormier, S., Satterstrom, F.K., and Bridges, T. 2009. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. *Science of the Total Environment* 407: 5199-5205.

Martin, B. T., Pike, A., John, S. N., Hamda, N., Roberts, J., Lindley, S. T., and Danner, E. M. 2017. Phenomenological vs. biophysical models of thermal stress in aquatic eggs. *Ecology Letters* 20: 50-59.

McElhany, P., Ruckelshaus, M.H., Ford, M.J., Wainwright, T.C., and Bjorkstedt, E.P. 2000. Viable salmonid populations and the recovery of evolutionarily significant units. NOAA Technical Memorandum NMFS-NWFSC-42,156 pp.

Murray, C.C., Hannah, L.C., Doniol-Valcroze, T., Wright, B.M., Stredulinsky, E.H., Nelson, J.C., Locke, A., and Lacy, R.C. 2021. A cumulative effects model for population trajectories of resident killer whales in the Northeast Pacific. *Biological Conservation* 257: 109124.

Mussen, T.D., Cocherell, D., Poletto, J.B., Reardon, J.S., Hockett, Z., Ercan, A., Bandeh, H., Kavvas, M.L., Cech Jr, J.J., and Fangue, N.A. 2014. Unscreened water-diversion pipes pose an entrainment risk to the threatened green sturgeon, *Acipenser medirostris*. *PloS one*, 9(1), p.e86321.

Patrício, J., Elliott, M., Mazik, K., Papadopoulou, K.-N., and Smith, C.J. 2016 DPSIR—Two Decades of Trying to Develop a Unifying Framework for Marine Environmental Management? Frontiers in Marine *Science* 3:177. doi: 10.3389/fmars.2016.00177

Peterson, J.T., and Duarte, A. 2020. Decision analysis for greater insights into the development and evaluation of Chinook salmon restoration strategies in California's Central Valley. *Restoration Ecology* 28: 1596-1609.

Piou, C., U. Berger, and V. Grimm. 2009. Proposing an information criterion for individual-based models developed in a pattern-oriented modeling framework. *Ecological Modelling* 220:1957-1967

Poletto, J.B., Cocherell, D.E., Mussen, T.D., Ercan, A., Bandeh, H., Levent Kavvas, M., Cech Jr, J.J., and Fangue, N.A. 2014a. Efficacy of a sensory deterrent and pipe modifications in decreasing entrainment of juvenile green sturgeon (*Acipenser medirostris*) at unscreened water diversions. *Conservation Physiology* 2(1), cou056.

Poletto, J.B., Cocherell, D.E., Ho, N., Cech Jr, J.J., Klimley, A.P., and Fangue, N.A. 2014b. Juvenile green sturgeon (*Acipenser medirostris*) and white sturgeon (*Acipenser transmontanus*) behavior near water-diversion fish screens: experiments in a laboratory swimming flume. *Canadian Journal of Fisheries and Aquatic Sciences* 71: 1030-1038.

Poletto, J.B., Cocherell, D.E., Ho, N., Cech, J.J., Klimley, A.P., and Fangue, N.A. 2018. The effect of size on juvenile green sturgeon (Acipenser medirostris) behavior near water-diversion fish screens. *Environmental Biology of Fishes* 101: 67-77.

Quinn, T. P. 2005. The Behavior and Ecology of Pacific Salmon and Trout. University of Washington Press, Seattle. 378 pp.

Rheinheimer, DE., Null, S.E., and Viers, J. 2016. Climate-adaptive water year typing for instream flow requirements in California's Sierra Nevada. *Journal of Water Resources Planning and Management* 142:04016049.

Rodgers, E.M., Poletto, J.B., Gomez Isaza, D.F., Van Eenennaam, J.P., Connon, R.E., Todgham, A.E., Seesholtz, A., Heublein, J.C., Cech, Jr, J.J., Kelly, J.T., and Fangue, N.A. 2019. Integrating physiological data with the conservation and management of fishes: a meta-analytical review using the threatened green sturgeon (Acipenser medirostris). *Conservation Physiology* 7: 1-21.

Schwalm, C.R., Glendon, S., Duffy, P.B., 2020. RCP8.5 tracks cumulative CO2 emissions. Proceedings of the National Academy of Sciences 117, 19656-19657.

Smith, W.E., Newman, K.B., and Mitchell, L. 2020. A Bayesian hierarchical model of postlarval delta smelt entrainment: integrating transport, length composition, and sampling efficiency in estimates of loss. *Canadian Journal of Fisheries and Aquatic Sciences* 77: 789-813.

Smith, E.P. 2020. Ending reliance on statistical significance will improve environmental inference and communication. *Estuaries and Coasts* 43: 1-6.

Staples, D.F., Taper, M.L., and Dennis, B. 2004. Estimating population trend and process variation for PVA in the presence of sampling error. *Ecology* 85: 923-929.

Steel, E. A., Tillotson, A., Larsen, D.A., Fullerton, A.H., Denton, K.P., and Beckman, B.R. 2012. Beyond the mean: The role of variability in predicting ecological effects of stream temperature on salmon. Ecosphere 3(11): art104.

Suter, G., Cormier, S., and Barron, M. 2017. A weight of evidence framework for environmental assessments: Inferring qualities. *Integrated Environmental Assessment and Management* 13: 1038-1044.

Thompson, R.M., Beardall, J., Beringer, J., Grace, M. and Sardina, P., 2013. Means and extremes: building variability into community-level climate change experiments. *Ecology Letters* 16: 799-806.

Troia, M.J. 2023. Magnitude–duration relationships of physiological sensitivity and environmental exposure improve climate change vulnerability assessments. *Ecography* 2023(1).

Tyler, A.V. 1992. A context for recruitment correlations: why marine fisheries biologists should still look for them. *Fisheries Oceanography* 1: 97-107.

Walters, C.J., and Collie, J.S. 1988. Is research on environmental factors useful to fisheries management? *Canadian Journal of Fisheries and Aquatic Sciences* 45: 1848-1854.

Williams, J. 2012. Juvenile Chinook salmon *(Oncorhynchus tshawytscha)* in and around the San Francisco estuary. San Francisco Estuary & Watershed Science 10(3) https://escholarship.org/uc/item/96f2t9xw

Witman, J.D., Pershing, A.J. and Bruno, J.F. 2023. Smooth and spiky: the importance of variability in marine climate change ecology. *Annual Review of Ecology, Evolution, and Systematics* 54: 129-149.