Independent Review of the Coordinated Long-Term Operation of the Central Valley Project and State Water Project

Prepared for:
National Marine Fisheries Service

By:
Kenneth A. Rose
France-Merrick Professor in Sustainable Ecosystem Restoration
University of Maryland Center for Environmental Science
Horn Point Laboratory
PO Box 775
Cambridge, MD USA 21613
(225) 773-5625 (cell)

Executive Summary

This an independent scientific peer review of the Biological Opinion (BiOp) on the Reinitiation of Consultation on the Coordinated Long-Term Operations of the Central Valley Project and State Water Project (CVP/SWP). The purpose of the review is to comment on the incorporation and application of best available scientific information in the assessment of effects of the proposed CVP/SWP operations on aquatic species. I was provided with background material, which include the Biological Assessment (BA), and draft versions of selected sections of the BiOp (and various appendices) related to the effects methodology and that reported the results of the analyses for the effects on individuals. Because only sections of the BiOp was available for this review, a comprehensive evaluation was not possible. I therefore provide answers the charge questions that are restricted in scope due to not knowing the details of the entire BiOp (i.e., I could not provide globally answers to all of the charge questions). However, my answers to the questions apply to my restricted evaluation and I was able to do a general evaluation of the effects methods used. The reader should consider what I comment on, but know that a lack of comments is not an endorsement of methods or analyses by me as a reviewer. The overarching comments of my review are:

- The overall framework for assessing the effects of the Proposed Action (PA) is scientifically-sound and can lead to well-structured analyses that provide critical information on the effects of the PA. The approach is well described (albeit in general terms) for the effects on individuals, and sufficiently described for populations and diversity groups. (Q1.4)
- How the individual-level effects will be folded-up to population and to diversity groups is more difficult to review because it depends on the implementation (not just general description but details leading to the results) that would be documented in the synthesis and integration (S&I), which was not part of my review. (Q1.4)
- The analytical approach for the BFE's (physical and biological features, habitat) is described less clearly and results are apparently mostly qualitative statements. The approach for habitat would benefit from explanations about what deconstructed actions are considered as effects on habitat versus effects on individuals of the species, as many of the deconstructed actions affect individuals through changes in their habitat. (Q1.4)
- Baseline should be defined very carefully and clearly, including whether, and if so how, recent droughts, climate change, ongoing Reasonable and Prudent Actions (RPAs), and proposed conservation measures (CMs) are incorporated. The terms Baseline, COS (current operations), no-action, and others should be clearly defined and explained which, if any, is the official baseline. (Q1.4)
- The analysis of effects (results) on individuals (which was regional) is thorough and uses the available information in a reasonable and effective manner. (Q1.5)
- The application of the effects analysis going to population and diversity groups is more difficult to review because much of this would occur in the cumulative effects and S&I, which were not available. A major re-occurring theme for S&I is the need for consistency across deconstructed actions and effects on individuals within a region (likely pretty consistent) and across regions (this is where it gets more challenging). (Q1.5)

- The effects analysis of individuals uses an impressive suite of datasets and models. Much of the analyses were from the BA, which further confirms that either the BA used the best available information or that there is was insufficient time to do a lot of new analyses. My cursory look at the BA modeling suggests it is the former: the BA modeling seems to be done well. (Q1.7)
- I was expecting more use of the monitoring data to show status and trends of the populations. This is data rich system, especially for salmon. (Q1.7)
- I suggest there is an opportunity for NMFS to gain additional confidence in the effects analysis beyond my review. The opportunity is to make use of previous reviews of the 2009 BiOp and the WaterFix EIS, which used similar methods for analyzing effects. (Q1.7)
- It is not possible to list all assumptions for every deconstructed effect analyzed but a list of the major assumptions that underlie the effects analysis would be helpful. The assumptions should be presented with a focus across regions, and a table with the assumptions about baseline and the period of record used for analyses by effect and across regions would help the reader keep track of major assumptions. (Q1.8.1)
- Presuming a similar approach to S&I as in 2009 BiOp suggests uncertainties will be dealt with qualitatively and not step-by-step of the analyses. This may be sufficient, but I cannot judge. One suggestion is to keep track of bias versus precision, which together can be thought of as comprising uncertainty. The implications of these are quite different. Second, with any categorization (e.g., low versus high uncertainty), express a scale and use an example to illustrate and anchor the categories. Finally, combining effects by summing the mean values (qualitatively or quantitatively) can lose possible responses related to non-linearity and thresholds and also lose responses to extreme events. (Q1.8.1)
- The issue of hatchery fish versus natural-run fish may not be raised to sufficient level in the BiOp. This discussion may be in other sections than those I reviewed. (Q1.8.2)
- Based on my partial review and without the S&I but with a good idea of how the S&I will be done, I consider that the effects analysis (not the entire BiOp since I did not have it) adequately covered the effects of the proposed action on spawning, rearing, and out-migration. I am less clear on the how the analyses included ongoing RPAs and the benefits of ongoing and proposed CMs. I did not review the green sturgeon and killer whale analyses for their completeness. (Q1.9.1)
- The use of the available conceptual models by NMFS is very good. Those models were developed with an extensive process and can greatly help ensure that effects are "thoroughly" analyzed and risks are not overlooked." However, I think that more can (and should) be done, as they can be used to view multiple effects within a region, trace major effects across regions (and hence through the life cycle), and be especially useful in the S&I. (Q1.9.2)
- There were many effects considered and that can hinder transparency when a reader gets to the S&I and can also "numb" the reader so that the major effects get diluted by the presentation. The S&I should have a format that allows for easy identification of the major effects across regions. (Q1.9.2)
- Much of the analysis effects on individuals use the 82 year run of CALSIM2 that captures hydrological variation. Does this time period and the use of CALSIM2 output sufficiently capture rare but important extreme situations such as prolonged droughts? (Q1.9.2)
- Climate change seems unevenly treated or the unevenness appears because of how its documented. Climate change was sometimes explicitly included (e.g., via the BA modeling to 2030) and sometimes seemingly ignored. (Q1.9.2)

- In general, the appropriate models were used for the effects analysis at the individual level. The collection and linkages among the models, assuming the passing of outputs to become inputs and the processing of the outputs to get summary measures of outputs was done correctly, then reflects very good use of the available models. (Q.1.9.3)
- The heavy (almost complete) reliance by NMFS on the modeling done in BA means that one is assuming the BA methods are most appropriate and done correctly. The test of this is whether if NMFS had done their own modeling from scratch, how closely would it have mimicked the modeling reported in the BA? (Q1.9.3)
- Some additional clarification about how the effects are reported would help increase clarity. For example, in the Delta regional analysis of individual effects, the presentation of DPM results themselves is confusing. The text refers to percentages in several ways, and seems to mix fractions and changes (PA minus COR) using various expressions of percentages. (Q.1.9.3)
- The results of the IOS and WRLCM need more explanation and context. Use of such life cycle models is an excellent addition to the effects analysis. The preponderance of life cycle modeling results of no difference between the COS and PA, despite the extensive lists of effects on individuals in the regional analyses, needs explanation. Perhaps showing the outputs of each model as they were inputted into the next model would help. (Q1.9.3)

In conclusion, the effects on individuals appears to be scientifically-sound. The effects of the PA on diversity (variability) are less well enumerated. How the results will be accumulated to the populations and diversity groups is feasible if sufficient consistency was maintained among regional analyses. Clarification of baseline, how well it includes events like drought and how climate change is factored in throughout the many analyses would greatly improve transparency.

TABLE OF CONTENTS

1 Introduction

1.1 Background [the text below is from the final charge from NMFS to the reviewers]

Reclamation is consulting with the U.S Fish and Wildlife Service (FWS) and the National Marine Fisheries Service (NMFS) pursuant to Section 7(a)(2) of the Endangered Species Act (ESA) on the coordinated operation of the Central Valley Project and State Water Project (CVP/SWP). As a part of these consultations, Reclamation has written a Biological Assessment (BA) that summarizes the effects of the Reinitiation of Consultation on the Coordinated Long-Term Operations of the CVP/SWP (ROC on LTO) on ESA-listed species and their designated critical habitats. NMFS will complete its assessment of effect and jeopardy determination in a biological opinion (BiOp), expected to be completed by July 1, 2019, as directed by the October 19, 2018, White House memorandum Promoting the Reliable Supply and Delivery of Water in the West. The purpose of this independent scientific peer review is to obtain the views of experts not involved in the ROC on LTO ESA consultation on the incorporation and application of best available scientific information and information and assessment of effects on aquatic species of the proposed CVP/SWP operations.

Each reviewer will review NMFS' analytical approach, status of the species and critical habitats, environmental baseline, and effects analysis sections of the draft BiOp. Each reviewer will also receive relevant background information and supplemental materials to consider in their review. NMFS will be available for a conference call during the review period to provide answers to questions or address clarification needs during the review. Reviewers are expected to convene at least one conference call among themselves to discuss major findings and identify and attempt to rectify any conflicting recommendations. The review is expected to culminate with individual reports from each reviewer, according to the format provided by the hiring contractor. This report is one such individual review.

1.2 General Observations

[The reader will find general observations in the Executive Summary and in my responses to the charge questions].

1.3 Review Activities

I was provided with background information on April 30 and then with the draft version of key sections of the BiOp that are the subject of this review on June 3. The sections of the draft BiOp that were provided were: Sections 2.1 through 2.6 and Appendices H, I, J, K, and L. I also used the BA, including Appendices a-h, as reference material (did not review the BA). The reviewers had two conference calls with NMFS: Kick-off (June 3) and Interim (June 7). On the interim conference call, the

reviewers described to NMFS how they would approach the review, given the time constraints and that an incomplete BiOp was provided to the reviewers.

The reviewers decided we would be most constructive by focusing on key areas and topics, rather than details, in order to provide as broad coverage as possible. The synthesis and integration (S&I) chapter was not available for the review so we (the reviewers) used how the synthesis and integration was done for the 2009 BiOp as being a reasonable template (not exactly the same and the 2009 results are not relevant) for the new synthesis and integration. As part of my review, I offer suggestions and ideas for the synthesis and integration that would ensure transparency and scientific soundness to the very challenging task of synthesizing the many individual analyses of individual-level effects together.

I therefore provide answers the charge questions that are restricted in scope due to time constraints and not knowing the details of the entire BiOp (i.e., I could not provide global answers to all of the charge questions). However, my answers to the questions apply to my restricted evaluation and I was able to do a general evaluation of the effects methods used. It is important to distinguish why there is a lack of comments about topics or details. It is not because I thought those analyses were done perfectly or that I had no comments; rather, it is because they were likely not reviewed. The reader should consider what I comment on to be complete and thorough, but that a lack of comments is not an endorsement by me as a reviewer.

Some comments may be simply a reflection of the partial BiOp being reviewed. If so, these comments are easy for NOAA to confirm that the comments are satisfied by ensuring where the information does appear now, it is made more clear and easy to find and access.

Responses to Questions

1.4 How well does the analytical approach explain how the exposure, response, and risk from project operations will be assessed for: individuals, populations, and diversity groups of the listed species? physical and biological features of designated critical habitats?

[The key phrase here is "explain."]

The overall framework for assessing the effects of the PA is scientifically-sound and can lead to well-structured analyses that provide critical information on the effects of the PA. The use of the process (deconstruct the action → exposure → responses → risk), with the Viable Salmonid Populations (VSP) focus on abundance, productivity, spatial, and diversity, is an excellent way to organize the complicated effects of the PA. The use of life cycle models as additional tools is also noteworthy and a very positive step in preparing these BiOps. The approach is well described (albeit

in general terms) for the effects on individuals, and sufficiently described for populations and diversity groups. How the individual-level effects will be folded-up to population to diversity groups is more difficult to review because it depends on the implementation (not just general description but details leading to the S&I results) that would be documented in the S&I, which was not part of my review. The concepts that form the basis of the approach (deconstruct, exposure, response, risk; VSP) are an excellent foundation for assessing the impacts on individuals, and if implemented well, would provide a solid basis for quantifying individual effects that can be carried over to the higher levels of population and diversity groups. If time to address these comments is pressing, my statement suggests that effort should focus on the degree of consistency of the individual-level effects analyses (i.e., look across regions) to ensure effective synthesis and integration.

The analytical approach for the PBFs (physical and biological features, habitat) is less clear and would benefit from explanations about what deconstructed actions are considered as effects on habitat versus effects on individuals of the species, as many of the deconstructed actions affect individuals through changes in their habitat. The description of how effects on habitat would be assessed should be expanded. The results are mostly general statements with a lot of "may cause this" without much semi-quantitative or quantitative information in support, seems to be a reflection of this lack of formalization of the analytical approach for habitat. Perhaps the habitat results are clearly folded into the S&I with the species effects?

I conclude that the description for effects on species (individual to population to diversity groups) is well-founded and on a solid scientific basis, but that the analytical approach (or acknowledgment there is not one and why) should be better described for the habitat.

Explanation of the effects analysis includes describing the Environmental Baseline, upon which the individual-level effects of the PA will accumulate for the population and then added to this baseline. Section 2.4 has a lengthy discussion of the baseline and yet did not provide me with some of the information I am looking for. Baseline should be defined very carefully and clearly. It may be in a section I was not provided. What is in baseline and how is it extrapolated to the account for future actions and climate change? The field-based information and data that were used in the regional analyses and modeling were all historical, so are all activities that are supposed to be in the baseline fully reflected in the data? Examples of things that should be clearly described as part of baseline or not, and why, are the recent droughts, climate change, ongoing RPAs (reasonable and prudent alternative), and proposed conservation measures (CMs). For example, the treatment of the CMs as part of baseline is confusing. The material provided to me included an assessment any negative effects from the CMs. However, I did not see how the benefits of the CMs were assessed as part of baseline? Often, NMFS stated that due to lack of details of a CM, it was therefore considered baseline or for programmatic consultation (page 107 in Section 2.5.5 of the BiOp). Does the baseline include the recent extensive drought? It does not seem to, as most of the analyses do not include

such recent years. There are lags in ecological responses to actions, including management actions, and droughts and other stressors, and (a) are these included in baseline, and (b) if these are included, does baseline reflect their fully realized effects? The BA used 2030 for their modeling analysis, and the BA modeling results were also used by NMFS. Does this mean that some of the analyses used a baseline that was adjusted to 2030 and some analyses of effects did not include climate change?

Another critical component that needs clarity is the time periods covered by field data to assess effects and I think needs to be accommodated into defining baseline. For example, excellent use is made of the field data to establish exposure based on how much their occurrence in a region overlaps with when a deconstructed action would impact the environment. This for the species that itself would be affected by climate change with or without the PA, and so is climate change a consideration for baseline as well as for assessing PA impacts relative to baseline. Of course, one cannot change the field data and so when does the adjustment occur or not occur? Should aspects, like exposure, be adjusted from the historical data to account for climate change? I suspect not, but this needs to be clearly stated and then how climate change will be used to account for these climate change effects in the S&I.

There are various terms used for the conditions of baseline and PA that are confusing and should be rectified across regions and modeling. These terms include: Baseline, No-action, COS (current operations), PA, parts of the PA (flow and temperature) and no-action, among others. Is COS the same as Baseline? It seems to include a climate change effect in some cases and not in others. The Winter-run Life Cycle Model (WRLCM) refers to Appendix H, which is very good. But the text in Appendix H stopped before I could possibly see if the differences in inputs was presented.

1.5 How effectively is the analytical approach applied in the effects analysis on the listed species and designated critical habitats?

[Note: The key phrase here is "applied."]

The key to how well the effects of the PA can be assessed depends on how the general framework (see comments to Q1.4) is actually implemented or applied. It is in those details, how the analyses are parsed into manageable units (e.g., regionally or by species), and the use of imperfect data and knowledge and models that can have difficult-to-quantify uncertainties, where unevenness and inconsistencies can arise in the implementation of the effects analysis. I considered the deconstruction of the PA, identification of the **effects on individuals**, determination of exposure, and the resulting risk statements to be well founded and, for the most part, complete to the extent the data and models permit.

The analysis of effects on individuals (which was regional) is thorough and uses the available information in a reasonable and effective manner. Each regional analysis appears to cover the major

effects and, at least qualitatively, uses similar enough approaches, data, modeling results, and organization and presentation to enable a reader to follow the PA across regions from the deconstruction of the PA to the resulting effects on individuals.

The application of the effects analysis going to population and diversity groups is more difficult to review because much of this would occur in the cumulative effects and S&I, which was not available. A major re-occurring theme for S&I (here and in many other situations) is the need for consistency. In this BiOp, it is consistency across deconstructed actions and effects on individuals within a region (likely pretty consistent) and across regions (this is where it gets more challenging). This may sound obvious, but consistency in how effects were quantified and uncertainties documented, plus a transparent way to combine the effects, are the keys to a strong and scientifically-defensible folding up of effects to the population level in the S&I.

A thought experiment to consider is whether the same results at the population level would have been obtained if the needed effects analyses on individuals had been divided into work units by species rather than by region. The approach taken in the BiOp for S&I (assuming it similar to that used in the 2009 BiOp) can work, using pieces from each region to assemble them into the effects on the population. This is quite feasible and makes sense from the point of view of deconstructing the PA that is very place-based and forming effective work teams, but also means that the level of consistency across regions must be high to enable the effects on a species, estimated by different people, to be accurately combined in a transparent manner. Is the magnitude of effects judged small or large the same across regions? Was the PA deconstructed the same way? Exposure is very likely consistently done across regions because it is based on when individuals are present in the region. Weight of evidence (confidence level) had a scale to work with based on the amount of evidence. The weight of evidence rating scale (Section 2, page 20) has a nice set of criteria for ensuring sufficient consistency across regions and species. Were the time periods analyzed the same across regions? Did the regional analyses view baseline the same in terms the productivity and status of the individuals and the effects of drought, climate change, and CMs? How similarly were the CMs in the deconstruction of the PA treated across regions? Are there any effects in one region that would affect the effects quantified in the "downstream" or "upstream" regions? A second thought experiment would be to think whether if the people from one region did the analysis for a different region, how would the two analyses (local and transplanted scientists) concur?

As the analysis proceeds beyond the individual-level to population and diversity group, the analyses become less rigorous and the critical information shifts from how the deconstructed PA would affect the environment important to individual salmon and sturgeon, to how to gather up and fold together the various pieces done at the specific action and individual levels within different regions and then across regions. I relied on the S&I approach used in the 2009 BiOp for some idea of how S&I would be done here.

I suspect that the fold-up from individual effects by region to the population level is quite doable, provided steps were taken to ensure consistency and therefore additivity (collation) of the effects across regions. Diversity as part of the VPN (distinct from diversity as in diversity groups) will remain a challenge. This not because of the lack of analytical methods used but rather due to data limitations and because diversity in this context focuses on effects on variability (rather than the average effect) and thus is always more challenging to quantify. In my opinion, the combination of individuals-level effects to obtain population responses should be semi-quantitative (also because there are modeling tools available), while a qualitative argument for diversity within populations (and effects at the diversity group level) is reasonable. Qualitative does not mean that numbers, data, and models are not used; it means that quantitative information can, and should, be used but the conclusions are best expressed in qualitative terms. Saying that effect X MAY or CAN result in lower Y is too weak. I point this out to be clear that qualitative statements are still definitive, but they focus on direction and rough magnitudes (high or low) rather than reporting effects on a continuous scale (e.g., The population will have 1,212 less individuals or decrease by 12.3%).

The likely use of tables as a basis for the S&I, as described in the effects analysis, is reasonable, but their entries must be extensively documented. Not just what region the effects and associated information came from with short-hand bullets, but rather some documentation of the thought process and logic and rationale used to construct the entries in the tables would greatly help the S&I. Without enough documentation, the S&I can appear to be weaker than it really is and more difficult to justify.

1.6 To what extent does the approach for assessing effects provide a scientifically defensible approach for evaluating effects to listed species and their designated critical habitats throughout the action area?

[Note: The key phrases here are "scientifically defensible approach" and "throughout the … area."]

The analysis for adverse effects on individuals is reasonable and defensible. While difficult to review without the details and results, I expect the roll-up to populations to be reasonable, although it will depend on the consistency of the estimation of the individual-level effects that were done across deconstructed actions within a region, and even more critically, across regions. The synthesis of effects across regions is of particular concern to me because the estimated effects involves different staff specific to a region, as there are many effects and most are seemingly small but a few are large. With just so many effects and different analyses by region, it will be a challenge to track the effects and their uncertainties (and certainties) through the accumulation of individual effects to the population level.

In my view, the analysis of effects on habitat is less strong than the effects on species, relying mostly on qualitative statements about the general state of the habitats within each region and how the PA may affect them. I discussed this more in my responses and comments to other questions.

1.7 How well does the draft biological opinion use best available scientific and commercial information in the effects analysis and findings?

[Note: The key phrase here is "best available… information."]

I evaluated the effects analysis based on effects on individuals. I can only offer some general scattered comments on the analyses beyond individuals. The effects analysis of individuals uses an impressive suite of datasets and models. Much of the analyses were from the BA, which further confirms that either the BA used the best available information or that there is was insufficient time to do a lot of new analyses. My cursory look at the BA modeling suggests it is the former: the BA modeling seems to be done well. One new life cycle model was added by NMFS, which is also good. As best as I could, I evaluated the modeling and, despite some missing information on inputs and lack of clarity about baseline, the modeling (BA and NMFS's use and own life cycle modeling) seemed solid.

I was expecting more use of the monitoring data to show status and trends of the populations. This is data rich system, especially for salmon. I was looking for trends from the various monitoring data sources for the status of the species discussions in Section 2. An example is Figure 2.4.4-1 for Spring-run Chinook.

I suggest there is an opportunity for NMFS to gain additional confidence in the effects analysis beyond my review. The opportunity is to make use of previous reviews. The new BiOp seems to follow, to some extent, the 2009 BiOp and so reviews of the 2009 BiOp remain mostly relevant. Also, similar effects analyses were done for the WaterFix EIS, which was also reviewed for its basis in best available science. The BiOp uses excerpts from the WaterFix EIS to document the methods in this new BiOp being reviewed here. This would be a way to leverage the peer review process. There are few instances in the BiOp that refer to the earlier reviews of the 2009 BiOp, but these are scattered and full appreciation of responses to old reviews would help. Given the similarities in the effects analysis approach between the 2009 and 2019 BiOp, documenting which of the comments on the 2009 BiOp have been addressed and how, and which ones were not (which is fine – reviewer comments are suggestions not mandates). The same argument applies to the review of the effects analysis used for the WaterFix EIS. I emphasize that not all comments must or even should be addressed but seeing these comments in one place and seeing if they are considered valid by NMFS and which ones were partially or fully addressed. A compendium of major comments and responses would help this review process. It would be a way to leverage this review.

1.8 Does the draft biological opinion adequately address data gaps and uncertainties? Specifically:

1.8.1 Are uncertainties and assumptions in the effects analysis clearly stated and reasonable based on current scientific knowledge?

It is not possible to list all assumptions for every deconstructed effects analyzed but a list of the major assumptions that underlie the effects analysis would be helpful. The assumptions should be more presented across regions, and a table with the assumptions about baseline, the period of record used for analyses by effect (any with deconstruction, exposure, risks) and across regions, would help the reader keep track of the major assumptions. For example, were the comparisons of PA to COS for egg mortality based on the same time period as other effects in the same region and for effects analyzed in other regions? Some deviations within and among regions will occur due to different data, information, and models used. Presently, the degree of coordination and therefore consistency within and among regions, is challenging to piece together from the regional analyses. This can be done as part of S&I but probably better done as part of the effects analysis methods so the reader knows these as they read the regional analyses. If such information is put together, or already exists in other sections of BiOp, then the key to making it useful is its specificity in the documentation (not just a list of datasets or models, but also time periods, etc.).

To the cursory extent I could evaluate the BA modeling, which the BiOp relies on, thee BA modeling appears to be sound. Assumptions are well stated and, with any complicated analysis based on linked models, it is a challenge to document the modeling while still making the key information easily accessible (e.g., 1000's of pages may be complete but the cost is reduced transparency).

Three specific areas that would benefit by much more clarity and statement of assumptions are: (1) how the different field data were used across analyses and across regions, especially the time periods covered and how they affect interpretation. This will come up with S&I to get to the population level, (2) what was assumed to be baseline and what was not and, in both cases, why, and (3) how was climate change incorporated into baseline and the effects analyses (data, assumptions, modeling). Presently, it is a lot of effort for the reader to reconstruct this information.

Dealing with uncertainties is always an issue in these types of analyses. While we are very good at analyzing uncertainties in a single model or single analysis of data, we continue to be challenged by formal analysis of uncertainty when multiple analyses and models and data sources are used in complicated and highly linked schemes. The linked scheme used here is appropriate for this BiOp, but a disadvantage is tracking how uncertainty propagates through the analyses. Much of how uncertainties are treated in this BiOp will likely appear in the S&I. Presuming a similar approach

as in 2009 suggests uncertainties will be dealt with qualitatively and not step-by-step of the analyses. This may be sufficient, but I cannot judge.

I can offer some suggestions that I would look for in how uncertainties could be dealt with in the S&I. One is to try to keep track of bias versus precision, which together can be thought of as comprising uncertainty. The implications of these are quite different. Bias is how close the value is to truth, while precision is the variability around the value. We tend to simply discuss uncertainty and not to differentiate between bias and precision. While you may not be able to quantify these, considering them as the S&I is done may help the interpretation. Bias is about how wrong can you be, while precision can be considered related to risk. To give the benefit of doubt to the species (institutionalized caution), requires a certain level of knowledge about the biases in the deconstruction, exposure, and effects, and then especially in the cumulative effects and S&I. It is easy to say one is adhering to this principle, it is much more difficult to point out places this principle was used and how much adjustment was made. Also, do not use the term risk loosely; risk of what? This discussion of bias, precision, and risk is an over-simplification but illustrates the idea of considering bias and precision and not just a lumped uncertainty.

Second, with any categorization (e.g., low versus high uncertainty), one should express a scale and use an example to illustrate and anchor the categories. There should be agreement across regions and effects about how to rate the effects, uncertainties, and risks. What is a meaningful reduction? Was the same data and time period used to define exposure across regions for the same species?

Finally, combining effects by summing the mean values (qualitatively or quantitatively) can lose possible responses related to non-linearity and thresholds and also lose the effects of extreme events. These arise when the uncertainties of the effects are considered and you get extreme values of multiple effects occurring together (i.e., "the perfect storm"). I saw many analyses of effects on individuals that divided the CALSIM2 simulation in water-year types (this is good), but it was not clear if critically-dry covered multiple years in a row of drought.

1.8.2 How extensively are gaps in aquatic species life history information considered and appropriately addressed?

I noted some instances where results of effects on individuals from one species were used as an approximation for the effects on another species. The occurrences I noticed included some discussion of the rationale, but such substitutions of results have typically large and unknown uncertainty, and great care must be used when they are combined with other effects. I think that, in general, the information for the salmon is reasonably complete and the information is limited, but sufficient, for sturgeon. Using surrogates for sturgeon is, of course, a challenge because of their life history.

The issue of hatchery fish versus natural-run fish may not be raised to a sufficient level in the BiOp. This discussion may be in other sections than those I reviewed. I put the issue of hatchery fish here because it relates, to some degree, to gaps in life history. How are hatchery fish being considered in terms of the effects analysis? Are there life history differences between hatchery and natural fish that should be considered in the effects analysis, and especially in the S&I when effects are scaled to the population and diversity group levels? Given the high attention paid to the issue of hatchery fish in the past, I suspect the information is somewhere in the BiOp. If not, then the issue of what percent of returns are hatchery and how would hatchery fish affect effects and diversity may need to be further considered.

I was underwhelmed by the treatment of habitat, but I do not think that was because of gaps in life history information related to habitat quantity and quality by life stage. I think it was more the level of analyses of effects on habitat that were done, which seemed limited relative to the fairly extensive amount of information that is available about life stage-specific needs of habitat and their linkages of stages into their life cycles. Perhaps this was covered in the species analyses due to overlap between effects on individuals (often through habitat) and effects on habitat?

1.9 How adequately does the draft biological opinion address the key operational effects of the proposed action? Specifically:

1.9.1 Do the analyses provide sound information and analyses to adequately characterize the effects of operations on spawning, incubating, rearing, and outmigrating salmonids and sturgeon?

Based on my partial review and without the S&I but with some idea of how the S&I will be done, I consider that the effects analysis (not the entire BiOp since I did not have it) adequately covered the effects of the proposed action on spawning, rearing, and out-migration of salmon. Due to time constraints, I am not sure how complete the analyses were for effects on sturgeon. I am also less clear on the how the analyses for salmon included ongoing RPAs and ongoing and proposed CMs. It seems the negative effects of the CMs were analyzed well. It is not clear to me how well the benefits of the CMs were included as part of the PA. There is always some vagueness in the specificity and future effects of CMs; I am looking for clarity in how the positive effects of the CMs were or were not included as part of baseline and the PA. Perhaps this is done in other sections of the BiOp.

1.9.2 *How thoroughly do the data, analyses, and findings presented in the biological opinion capture the risks to individuals and populations, and to critical habitat, from the proposed action? Are there significant risks that have been overlooked or other scientific information that should be considered?*

[Note: The key phrases here are "thoroughly … capture the risks" and "risks … overlooked."]

The analysis of effects on individuals seems strong, the analysis of effects to the populations are reasonable (as far as I can judge with incomplete information), and the effects to habitat should be strengthened or better explained why they appear to be limited in scope. The habitats of the listed species are well known, and the modeling generates the key habitat features, so a more rigorous analysis of habitat effects seems doable from the point of view of available information. It may be possible to use the information that is already present and elevate the habitat analysis, at least to next level of more definitive statements and to relate it better to the effects on species that are mediated through habitat (which I think is most of them). Is a reason that the species analysis are extensive and habitat-specific analyses appear limited is because there is not a clear distinction made between effects on species through habitat and effects on habitat and many got analyzed as effects on species? Or are there other reasons the habitat analyses seemed limited in scope?

The idea of using of the available conceptual models by NMFS is very good. Those conceptual models were developed with an extensive process and can greatly help ensure that effects are "thoroughly" analyzed and "risks are not overlooked." However, I think that more can (and should) be done with the conceptual models than what I saw in the sections I reviewed. They can be used to view multiple effects within a region, trace major effects across regions (and hence through the life cycle), and be especially useful in the S&I. Perhaps the conceptual models can be used directly with the summary tables in each region and then in the S&I as a check-off that major effects and possible cascading effects through the life cycle were considered? The conceptual models provide a ready opportunity to ensure the analyses are thorough and risks are not missed in a robust manner. The multiple conceptual models reflect many people's expertise and overlaying them makes their use even more robust than a single conceptual model.

There were many effects considered and that can hinder transparency when you get to the S&I and also can "numb" the reader so that the major effects get diluted by the presentation. Through my skimming of the results (and mostly with the Delta region), the increased losses related to exports and OMR appears as a major impact of the PA (Section 2.5.5.8.3.1.1.1; Table 2.5.5.-26, page 194 section 2). The S&I should have a format that allows for easy identification of the major effects across regions. Also, and NMFS discusses this, there is also the possibility of impacts on the population-level from many small effects on individuals ("death by a thousand cuts"). There is also

the possibility related to the "thoroughness" issue about non-linear responses and thresholds. Are these types of responses being considered, and if not, they should be noted? The life cycle modeling should help with the nonlinear and threshold issues.

Much of the analysis effects on individuals use the 82-year run of CALSIM2 that captures hydrological variation. Does this time period and the use of CALSIM2 output sufficiently capture rare but important extreme situations such as prolonged droughts? Are the results for the critically-dry year type similar enough to a drought? How about a consecutive 3-4 years of drought? According to the USGS, there have been droughts of 5-7 years, including the recent 2012-2016 drought. The CALSIM2 simulation was 1920 to 2002 so the recent drought is not included? The BiOps states in many places about the large effects of the recent drought. For example (Section 2.2, page 8), the BiOp states *"CDFW has documented critically low Spring-run Chinook adult returns to Mill and Deer Creeks for the fourth consecutive year, due in part, to one of California's most severe and prolonged droughts on record (December 2011 to March 2017)."* To illustrate the issue about time periods, the salvage density method used 1995 to 2009, which is different than the CALSIM2 simulation time period.

Climate change seems unevenly treated or the unevenness appears because of how it's documented. The text from the BiOp on climate change was incomplete. Climate change was sometimes explicitly included (e.g., via the BA modeling to 2030) and sometimes seemingly ignored. For example, was climate change included in the egg mortality calculations, with the DPM (Delta Passage Model) – perhaps many analyses did include climate change since whatever analyses used BA model output used conditions extrapolated to 2030. Climate change was not included in many of the exposure estimations? For example, it seems climate change was not included in Figure 2.5.5.-18 that shows the migration timing for brood years 1994 to 2017 for juvenile winter Chinook; another example to illustrate is Table 2.5.5.-5 that shows the temporal occurrence of winter-run Chinook in the Delta. [Also note the difference in years used versus the CALSIM2 simulation period]. In addition, the references used with climate change (Section 2.4.1.5) appear outdated, especially given the rapid increase in available information over the past 5-10 years.

1.9.3 Have the appropriate analytical tools (i.e., models) been used for the analysis and what, if any, additional currently available tools should have been considered? Were available models appropriately applied and interpreted in the analysis?

[The key phrases here are "models" and "appropriately applied and interpreted"]

In general, the appropriate models were used for the effects analysis at the individual level. This statement has several assumptions and nuances. First, this does not mean that the models and modeling were perfect; that is, new models were not developed for specific questions. For example,

many people recognize that CALSIM2 being a planning (nor truly predictive) model with a monthly time-step is not ideal. These assumptions are carried through all analyses, as the first model in the chain of linkages is CALSIM2 so everything follows from that. However, the questions being asked of the models for the effects analysis on individuals are long-standing and the models used have been themselves evaluated to various degrees. The collection and linkages among the models, assuming the passing of outputs to become inputs and the processing of the outputs to get summary measures of outputs, reflects very good use of the available models. The inclusion of the Anderson model for egg mortality and life cycle models demonstrate that the collection of models, while resembling the same models as used in past analyses, has some flexibility to accommodate new models as they emerge on the scene or in response to recommendations.

Second, the heavy (almost complete) reliance by NMFS on the modeling done in the BA means that one is assuming the BA methods are most appropriate and done correctly. The test of this is whether if NMFS had done their own modeling from scratch, how closely would it have mimicked the modeling reported in the BA? If the answer is that NMFS' own modeling would have overlapped greatly with the BA modeling, then the use of the BA modeling is quite reasonable and, in fact, is a good way to remove uncertainties that arise between two sets (BA and BiOps) of potentially competing modeling analyses.

Some additional clarification about how the effects are reported would help increase clarity. For example, the DPM results are very important. How well are the flows, velocities, and temperature projected for baseline and under the PA within the domain of applicability of the statistical relationships that comprise the DPM? In the Delta regional analysis, the presentation of DPM results themselves is confusing. The text refers to percentages in several ways. In the text, there are summaries that report *"The absolute differences in modeled median through-Delta survival ranged from +0.3 to -0.2 percent difference the PA and COS."* How can absolute differences be a percent? Do you mean percent so 0.3 is a difference in survival of 0.003 or are fractions, percents, and percent changes being mixed? A value of 0.3 is a fraction, as a percent it is 30% and then there is also the reporting as an absolute difference but is correctly reported as a percent because the two quantities are themselves percents. There is also reporting of relative changes [(x-y)/y x 100] as a percent. Later in the same section with the Perry modeling results, the BiOp says *"… through-Delta survival is approximately 45 percent in November for the COS compared to approximately 30 percent for the PA (middle panel), with a difference in through-Delta survival of about 12-15 percent (bottom panel)."* So now this is in units of daily survival fractions but expressed as fractions in the figure and as a percent in the text? These are therefore absolute differences? Another example of issues related to reported results are that the IOS results reported in the BiOp are different from those presented in the BA, but the BiOp says the results came from BA.

There is a tendency to focus on the sign of the difference between COS and PA (this is using the probability of the difference less than zero), which is a nice way to display the direction of effects but equally counts small and large differences. I suggest many of the small differences are within the uncertainty of the models and may be better as no difference. While the probability approach is fine, the results should also always use the magnitudes of these differences relative to the COS value. The magnitudes of the direction of these effects seems to get progressively lost with each summarization of results step. For example, the summary on page 30 (Section 2.5.5) eventually never mentions that the magnitudes of the differences, while biased in a direction, are relatively small in magnitude (I am assuming NMFS would call them small to illustrate this).

The results of the IOS (Interactive Object-Oriented Simulation) and WRLCM need much more explanation and context. While there is some explanation of what the causes of the few differences between COS and PA were, the preponderance of results was no difference between the COS and PA. These no-difference predictions require explanation, even perhaps more so than when differences were predicted. There is an enormous list of effects by region and then no differences predicted in the life cycle modeling. Of course, this is power of life cycle modeling but there seems to be disconnect between the many effects by region and the small responses predicted by the life cycle models. I believe the life cycle modeling is new and was not part of 2009 BiOp and so is also new to the S&I. In my view, the conceptual models and the life cycle modeling should play major roles, with the extensive tables, in the S&I.

A feature of linked models is that the next model in the chain has its own assumptions and uncertainties but also inherits those of the previous models that feed in information. It is possible when averaging and disaggregating model outputs to make them on the spatial and temporal scales of the next model in the chain that critical signals of PA effects can be lost. Basically, one can lose a change in environmental conditions by having to average outputs (many spatial boxes to one large spatial box) or when converting to finer scales (assumed variability from historical data used to make a monthly value have daily values).

The description of the modeling in the BiOp refers to the BA and, for a few of the modeling analyses, to the methods used in the WaterFix EIS. While this is understandable, it would greatly help to not just see the predictions from the various models but also to see the key inputs to the models to show how the effects of PA get transmitted to each model in the linked chain. For example, several models showed practically no difference between COS and PA (e.g., egg mortality, WRLCM). One can go the BA and see very nicely the effects of the PA versus COS for CALSIM2 outputs and other models used by the BA. I think showing the differences in the outputs in the models as they are used as inputs for the other models (averaged and disaggregated to match each subsequent model's scales) would be helpful. Are small effects predicted by the later models in the link (e.g., egg mortality, life cycle) due to essentially the high similarity of COS and PA values for

inputs or for other reasons (i.e., assumptions and calculations within the models). Knowing this influences the confidence we have in predictions. Was it the averaging or assumed disaggregation that removed the PA effect upon input to the next model (lower confidence in the predicted no response) or was it, for example, a non-linear relationship in the receiving model well-grounded in empirical information (higher confidence in predicted no response). Egg mortality as a function of temperature is a good example to consider.

Really, PA is not a good label for the modeling output, especially for the life cycle models, as it really is more like "PA-generated Temperature and Flow." The almost universal use of model predictions being labelled "PA" can lead to misinterpretation because different models considered various subsets of PA effects and not all effects. The description "holistic" to describe the life cycle modeling is also misleading; it holistic in that the modeling is full life cycle but it is not holistic in that not all of the PA effects were represented.

1.10 To the extent that reviewer expertise allows informed review of Central Valley water temperature guidance (see Additional References below as needed):

[I did not respond to these questions due to time limitations.]

1.10.1 Does the EPA (2003) water temperature guidance protect Chinook salmon on CVP rivers and creeks, and what implications do newer studies have for considering effects on salmon?

1.10.2 How appropriate is the application by Anderson (2018) of age-dependent thermal mortality and spatially-dependent background mortality to understanding early life history of winter-run Chinook salmon and temperature management planning? Are the effects, including uncertainties, of this new approach captured in the analysis?

1.11 To the extent that reviewer expertise allows informed review of analyses of effects of Delta conditions:

[I did not respond to these questions due to time limitations.]

1.11.1 How well are the near-field, mid-field, and far-field effects described for different potential volitional and entrainment migration paths in the Delta (e.g., north Delta, Sacramento River, central Delta, San

Joaquin River, south Delta, salvage, etc.) for different species and different basins?

1.11.2 *How well does the period of record in the Delta Salvage Model (1995-2009) reflect the conditions of the proposed action given the change in Old and Middle River management (from 2009 when NMFS' 2009 Opinion was issued and implemented)? What period of record does the panel recommend to generate a seasonal pattern of loss for use in comparing between the operational scenarios (i.e., PA and COS)?*

2 Additional Thoughts, Concerns, and Suggestions for Improvements to the Analyses

[There are no additional comments beyond those stated as part of my responses to charge questions.]

3 References

3.1 Materials Provided Prior to the Review

Anderson, J.J., R.T. Kneib, S.A. Luthy and P.E. Smith, 2010. *Report of the 2010 Independent Review Panel (IRP) on the Reasonable and Prudent Alternative (RPA) Actions Affecting the Operations Criteria and Plan (OCAP) for State/Federal Water Operations*. Prepared for the Delta Stewardship Council, Delta Science Program. December 9, 2010.

CALFED Science Review Panel, 2009. *Independent Review of a Draft Version of the 2009 NMFS OCAP Biological Opinion*. January 23, 2009.

Gore, J.A., B.P. Kennedy, R.T. Kneib, N.E. Monsen, J. Van Sickle, and D.D. Tullos, 2018. *Independent Review Panel (IRP) Report for the 2017 Long-Term Operations Biological Opinions (LOBO) Biennial Science Review*. A report to the Delta Science Program. January 2018.

Kneib, R.T., J.J. Anderson, J.A. Gore, M.S. Lorang, J.M. Nestler and J. Van Sickle, 2013. *Report of the 2013 Delta Science Program Independent Review Panel (IRP) on the Long-Term Operations Biological Opinions (LOBO) Annual Review*. Final report submitted to the Delta Stewardship Council, Delta Science Program. December 7, 2013.

Kneib, R.T., J.J. Anderson, J.A., Gore, M.S. Lorang and J. Van Sickle, 2011. *Report of the 2011 Independent Review Panel (IRP) on the Implementation of Reasonable and Prudent Alternative (RPA) Actions Affecting the Operations Criteria and Plan (OCAP) for State/Federal Water Operations.* Final report submitted to the Delta Stewardship Council, Delta Science Program. December 9, 2011.

Kneib, R.T., J.J. Anderson, J.A., Gore, M.S. Lorang and J. Van Sickle, 2012. *Report of the 2012 Delta Science Program Independent Review Panel (IRP) on the Long-Term Operations Opinions (LOO) Annual Review*. Final report submitted to the Delta Stewardship Council, Delta Science Program. December 1, 2012.

Kneib, R.T., J.J. Anderson, J.A., Gore, N.E. Monsen, G. Schladow, and J. Van Sickle, 2014. *Independent Review Panel (IRP) Report for the 2015 Long-Term Operations Biological Opinions (LOBO) Annual Science Review*. A report to the Delta Science Program. December 2015.

Kneib, R.T., J.J. Anderson, J.A., Gore, N.E. Monsen, J.M. Nestler, and J. Van Sickle, 2014. *Independent Review Panel (IRP) Report for the 2014 Long-Term Operations Biological Opinions (LOBO) Annual Science Review*. A report to the Delta Science Program. December 2014.

National Marine Fisheries Service, 2009. *Endangered Species Act Section 7 Consultation Biological Opinion and Conference Opinion on the Long-Term Operations of the Central Valley Project and State Water Project.* File Number 2008/09022. June 4, 2009.

National Marine Fisheries Service, 2017. *Endangered Species Act Section 7(a)(2) Biological Opinion, Magnuson-Stevens Fishery Conservation and Management Act Essential Fish Habitat Response, and Fish and Wildlife Coordination Act Recommendations for the California WaterFix Project in Central Valley, California.* NMFS Consultation Number WCR-2016-5506. June 16, 2017.

U.S. Bureau of Reclamation, 2019. *Reinitiation of Consultation on the Coordinated Long-Term Operation of the Central Valley Project and State Water Project Central Valley Project, California, Mid-Pacific Region, Final Biological Assessment*. January 2019.

3.2 Supplemental Materials Review -

None

3.3 Data Assessed

None