Peer Review of *Platte River Recovery Implementation Program Data Synthesis Compilation:*

Weight of Evidence Approach to Assessing Relationships between Flow and Interior Least Tern Forage Fish Abundance, Foraging Behavior, Productivity, and Dietary Requirements

Draft Summary Report

January 2016

Louis Berger

TABLE OF CONTENTS

This page intentionally left blank

1.0 INTRODUCTION

1.1 Background

The Platte River Recovery Implementation Program (Program) is intended to address issues related to endangered species and the loss of critical seasonal habitat in the Platte River in central Nebraska by managing land and water resources using the principles of adaptive management (AM). The application of AM to the Platte River will provide benefits for four protected species (i.e., Whooping Crane, Interior Least Tern, Piping Plover, and Pallid Sturgeon).

The Executive Director's Office (EDO) of the Program prepared a data synthesis compilation document (hereafter referred to as "forage fish synthesis document") of information related to the relationship between forage fish abundance and least tern productivity. The document is intended to inform the use of Program land, water, and fiscal resources to achieve one of the Program's management objectives: increasing production of the tern and plover from the Associated Habitat Reach (AHR) along the central Platte River (the Associated Habitat Reach consists of a 90-mile reach of the Platte River in central Nebraska from Lexington to Chapman). The forage fish synthesis document is a compilation of six sections with unique objectives and analyses for testing Program hypotheses and answering associated PRRIP "Big Question" #8 – Does forage availability limit tern and plover productivity on the central Platte River?

1.2 Purpose and Scope of Peer Review

The purpose of this review is to provide a formal, independent, external scientific peer review of the information presented in the forage fish synthesis document. Reviewers were charged with reviewing the synthesis document from their particular area of expertise and assessing the sufficiency of the document's conclusions regarding outstanding questions of scientific uncertainty. Factors to be addressed include the scientific merit of the sections' technical analyses and conclusions. The peer reviewers were tasked with ensuring any scientific uncertainties are clearly identified and characterized, and the potential implications of the uncertainties for the technical conclusions drawn are clear.

Specifically, the PRRIP requested that reviewers consider and respond to the questions listed below, at a minimum, in their reviews.

1. Does the forage fish synthesis document adequately address the overall objective of the document, which is to present lines of evidence for broader examination of the relationship between forage fish abundance, river flow, and tern productivity on the central Platte River?

2. Do the authors of the forage fish synthesis document draw reasonable and scientifically sound conclusions from the information presented? If not, please identify those that are not and the specifics of each situation.

3. Are the statistical methods used in each section appropriate?

4. Are potential biases, errors, or uncertainties appropriately considered within the methods sections of this document and then discussed in the results and conclusion section?

5. Are the conclusions drawn supported by the synthesis of data reported in Sections 2-5?

6. Did we utilize the data we had as effectively as possible to address the hypotheses "*forage fish abundance below 800 cfs limits least tern productivity*?" and "*there is a direct relationship between flow and forage fish abundance and species diversity*?" and answer the Big Question "*Does forage availability limit tern productivity on the central Platte River*"? If not, what additional analyses do you recommend?

2.0 PEER REVIEW PROCESS

Louis Berger was retained by the PRRIP to facilitate the peer review process. Louis Berger's responsibilities in the peer review process included 11 steps:

1. Develop a clear understanding of the required expertise of each position;
2. Conduct a search for potential candidates;
3. Contact prospective candidates to screen for criteria and conflict of interest;
4. Obtain CVs/resumes, biographical sketch forms, and signed "no-conflict-of-interest" statements from all candidates;
5. Compile a summary report describing recruitment process and candidate qualifications;
6. Communicate with reviewers regarding the selection process;
7. Discuss the scope and charge with the EDO;
8. Participate in an organizational conference call with the reviewers;
9. Distribute materials and commence review;
10. Compile all peer review comments into a spreadsheet and summarize in a summary report; and
11. Submit spreadsheet and summary report to the EDO and facilitate communication between the EDO and reviewers.

2.1 Selection of Reviewers

The Program requested that peer review panel members comprised the following areas of expertise: tern ecology (e.g., tern life history, bioenergetics); small fish ecology (e.g., Cyprinids or semi-arid river small fish); and ecological statistics. In identifying candidates, Louis Berger gave preference to individuals with experience in the Great Plains.

In August 2015, Louis Berger submitted a report to the Program that summarized the qualifications of nine candidate reviewers. In September 2015, the Program's Governance Committee selected three reviewers from that list. The panel comprised the following individuals (see Appendix B for biographical sketches):

Dr. Brian Gray, ecological statistics
Dr. Christopher Hoagstrom, small fish ecology
Dr. Sara Schweitzer, tern ecology

2.2 Document Review and Report Development

Following final approval of the three reviewers, Louis Berger initiated the review by distributing the files to the reviewers, including: the document to be reviewed; the scope of work and schedule for the peer review; the Program's Adaptive Management Plan; and State of the Platte Reports from 2012, 2013, and 2014. Louis Berger staff held a kickoff calls with the three reviewers on November 3 and 6, 2015 to discuss the scope of work, deliverables, and schedule, and answer any questions.

Reviewers conducted their independent desktop reviews between November 3 and December 17, 2015. Louis Berger contacted the reviewers individually to obtain clarification on their comments until December 21, 2015. Reviewers submitted the following deliverables:

1. Responses to the questions listed in Section 1.2;
2. Ratings of the sections in five different categories, as well as an overall recommendation; and
3. Specific comments on the text, by line number (optional).

Upon receipt of the deliverables, Louis Berger compiled the specific comments into a spreadsheet, organized by section and line numbers, which was submitted to the PRRIP as a separate deliverable. Louis Berger summarized reviewer responses to the questions in this summary report, which also includes their ratings and recommendations. Individual reviewer comments are included as Appendix A. As described in the PRRIP Peer Review Guidelines, reviewers can choose whether they would like their review comments to be anonymous or attributed. Because one reviewer preferred anonymity it was applied to the entire review, thus reviewers were each assigned a number (i.e., Reviewer 1, 2, or 3), in no particular order. The reviewers had the opportunity to review the draft summary report to ensure their comments were captured accurately prior to its submittal to the Program.

3.0 RESULTS

3.1 Responses to General Questions

Below are brief summaries of the individual reviewers' responses to the six questions posed by the PRRIP. This section is not intended to be a comprehensive summary or to be redundant with the individual comments in Appendix A, but rather attempts to capture some of the primary comments in each reviewer's response to the individual questions, as well as any themes that emerged or comments that were raised by more than one reviewer independently. For the reviewers' full comments see Appendix A and the comments spreadsheet.

Question 1: Does the forage fish synthesis document adequately address the overall objective of the document, which is to present lines of evidence for broader examination of the relationship between forage fish abundance, river flow, and tern productivity on the central Platte River?

Reviewer 1 responded that the document presents "putative lines of evidence"; however, not all of the lines of evidence address the hypotheses in the AM Plan. For example, Reviewer 1 noted that the clearest example is in Section 2 where three tested hypotheses (i.e., T2, T2a, and TP-4) are interpreted to mean that "mean daily flows on a specific day should be expected to determine the number of prey fish available on that same day" (p. A-3); however, the assessment approach does not address the three hypotheses. The reviewer noted that this is important because Section 2 has major implications for the entire study, as demonstrated by other examples in sections 3, 4, and 5.

Reviewer 2 interpreted "lines of evidence" to mean those that may be defended by peer-reviewed literature and responded "no" to this question (p. A-39). Reviewer 2 provided a number of examples to support this response, including the overall lack of clarity in the hypotheses and throughout the document; concerns about the validity of inferences; and conclusions that are too strong and lack sufficient evidence (see comments marked with "*1" on pages A-29 through A-39).

Contrary to reviewers 1 and 2, Reviewer 3 responded that the document *does* address the overall objective thoroughly and cited examples in each of the sections where objectives and hypotheses were clearly stated (p. A-41 to A-42).

Question 2: Do the authors of the forage fish synthesis document draw reasonable and scientifically sound conclusions from the information presented? If not, please identify those that are not and the specifics of each situation.

Building on comments in response to Question 1, Reviewer 1 responded "no" to this question, citing examples in Section 2 where analysis of forage fish sampling data and interpretation of forage fish abundance are unreasonable and scientifically unsupported (p. A-7 through A-13). This section also ignores important aspects of forage fish ecology, limiting its usefulness in assessing forage fish availability for least terns. Because Section 2 provides a foundation for the synthesis document, these shortcomings have implications for the findings in sections 3, 4, and 5 as well.

Reviewer 2 responded that the reasonableness of the conclusions is "unclear (due to lack of clarity in the document)," and oftentimes the conclusions are not scientifically sound (p. A-39). The reviewer listed numerous examples in response to this question, including: need to defend validity and/or utility of chosen models; lack of information about method of model fitting; and unsupported assumptions and inferences (see comments marked with "*2" on pages A-29 through A-39).

In contrast to reviewers 1 and 2, Reviewer 3 found that overall the conclusions were reasonable and scientifically sound with a few exceptions in sections 1, 2, 3, 4, and 6. For example, in Section 3 Reviewer 3 noted that though the model did not find significant relationships between flow rate or fish density on least tern plunge-rate or fish-capture-probability, the authors stated that flow rate could influence tern foraging behavior (p. A-42 to A-43).

Question 3: Are the statistical methods used in each section appropriate?

Reviewer 1 responded "no" to this question, given that Section 2 only utilizes a logarithmic model, but does not provide justification for its selection. The reviewer recommended that a linear model also be presented and the results of the two models compared. Similarly, Section 3 does not sufficiently describe the rationale for model selection (p. A-13 to A-15). These comments echo those from Reviewer 2 in response to Question 2.

Similar to Question 2, Reviewer 2 indicated that the lack of clarity in the document impedes a comprehensive response to this question, but in one to two sections the response is "no." In addition to examples listed above under Question 2 (e.g., need to defend model selection), the reviewer provided a number of examples to support this response, including: lack of sampling design details; need to explain how prediction errors were addressed; and failure to address sources of variability (see comments marked with "*3" on pages A-29 through A-39).

Reviewer 3 deferred to the statistician for more thorough assessment of the statistical method; however, they noted the "analyses and models seem appropriate and current" (p. A-43). The reviewer also identified an instance in Section 4 where the text regarding the best model is inconsistent.

Question 4: Are potential biases, errors, or uncertainties appropriately considered within the methods sections of this document and then discussed in the results and conclusion section?

In regards to Section 2, Reviewer 1 responded that the biases and uncertainties surrounding forage fish sampling are not even mentioned and the analyses and results do not correspond with the goals of the section. The reviewer listed a number of potential biases, errors, and uncertainties that should be considered and discussed in this section (p. A-15 to A-16). Reviewer 1 also discussed a number of other limitations in sections 3, 4, and 5 in response to this question (p. A-16 to A-17).

Reviewer 2 did not find potential biases, errors, or uncertainties to be appropriately considered in the synthesis document and cited a number of examples, most of which are also referenced in response to Question 3 (p. A-40). Some examples include: need to evaluate whether beta2 estimator may be biased; uncertainties regarding estimation of beta1 and beta2; and need to further evaluate potential errors (e.g., classical or Berkson's errors) (see comments marked with "*4" on pages A-29 through A-39).

Reviewer 3 found the methods to be "strong and well documented" overall, but mentioned a few areas that require further explanation. Specific examples include: explanation of how least tern data were captured; discussion of how variability in the sampling protocol was (or was not) accounted for; and discussion of differences between Caspian terns and least terns (p. A-43 to A-44).

Question 5: Are the conclusions drawn supported by the synthesis of data reported in Sections 2-5?

Reviewer 1 focused their response on the single conclusion drawn in Section 6 and stated that the second part (i.e., forage fish abundance and least tern productivity increase as flows decrease to 5.7 m^3s^{-1}) is not supported by the data, as discussed in response to earlier questions. The reviewer suggested a more accurate conclusion that is supported by the results; however, it does not address the stated hypotheses. In addition, Reviewer 1 found the first part of the conclusion (i.e., there is no evidence to support the hypotheses) to be inaccurate because, as noted in earlier comments, the hypotheses were not actually tested by the study; the reviewer also suggested a revision of this part of the conclusion (p. A-17 to A-19).

Reviewer 2 responded "no, or at least not well" to this question and cited the same examples provided for Question 2 to support this conclusion (p. A-40).

Reviewer 3 found the overall conclusions to be supported by the data synthesis and recommended that the authors also comment on other possible impacts of low flow (e.g., increased predator and human access to sandbars) and positive impacts of higher flows during nonbreeding season (e.g., vegetation removal, sand accretion) (p. A-44).

Question 6: Did we utilize the data we had as effectively as possible to address the hypotheses "**forage fish abundance below 800 cfs limits least tern productivity*****?" and "*****there is a direct relationship between flow and forage fish abundance and species diversity*****?" and answer the Big Question "*****Does forage availability limit tern productivity on the central Platte River*****"? If not, what additional analyses do you recommend?***

Reviewer 1 responded "no" and referenced additional recommended analyses as described in response to other questions. Specifically, the reviewer noted that forage fish data were not used appropriately or accurately and the flow regime was not analyzed in a meaningful way (p. A-19 to A-20).

Reviewer 2 also responded "no" and listed a number of examples in support, most of which were already cited in response to questions 2 and 3. Additional examples include: need to match flows from 1999 – 2000 with multi-year fish counts; need to address how flow can be a realistic surrogate for multiple variables; and need to clarify whether modeling a binomial or Bernoulli process (see comments marked with "*6," "*3," and/or "*2" on pages A-29 through A-39).

Reviewer 3 found data to be used effectively, but noted a few instances were clarification is needed. Specifically, the authors should explain how they accounted for positive bias associated with opportunistic observations of least terns, and clarify how they accounted the influence of turbidity on tern foraging success (p. A-44 to A-45).

3.2 Ratings and Recommendations

Reviewers rated the document using a rating system provided by the Program where 1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor. Below is a table summarizing the ratings of reviewers 2 and 3:

Table 3-1. Reviewer 2 and 3 comprehensive ratings of overall document, by category.

Category	Reviewer 2	Reviewer 3
Scientific soundness	4	2
Degree to which conclusions are supported by the data	4	1
Organization and clarity	5	2
Conciseness	3	1
Important to objectives of the Program	4	1

Reviewer 1 separated their ratings by section of the document as follows:

Table 3-2. Reviewer 1 ratings of each section of document, by category. Right column shows average rating for each category, rounded to the nearest whole number.

Category	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6	Overall Average
Scientific soundness	2	5	4	5	4	5	**4**
Degree to which conclusions are supported by the data	2	5	4	2	2	5	**3**
Organization and clarity	2	4	3	3	3	3	**3**
Conciseness	2	5	3	3	3	3	**3**
Important to objectives of the Program	1	1	4	1	2	1	**2**

Reviewers were then asked to provide their recommendation to either accept the sections, accept them with revisions, or deem them unacceptable. Reviewer 1 divided their recommendations by section; their recommendation was “Accept with revisions” for sections 1, 3, and 5, and they found sections 2, 4, and 6 to be “Unacceptable.” Both reviewers 2 and 3 recommended that the document be accepted with revisions. Reviewer 1 listed a number of specific changes needing to be made, within each section, for them to accept the document (p. A-20 through A-25). Throughout their peer review comments, Reviewer 2 identified those they considered most essential to address with an asterisk (p. A-29 through A-39). Lastly, Reviewer 3 noted that required revisions are summarized in their responses to questions 1, 2, and 4 and described the most critical change to be made on page A-45.

3.3 Other Specific Comments

The reviewers submitted 150 other specific comments, which Louis Berger compiled into a spreadsheet, organized by section and line number, along with reviewer number; this spreadsheet will be used by the PRRIP in preparing responses to the comments.

4.0 REFERENCES

The citations for references recommended by the reviewers are included in their individual comments in Appendix A.

5.0 APPENDICES

Appendix A: Individual Reviewer Comments

Appendix B: Reviewer Biographical Sketches

This page intentionally left blank

APPENDIX A: INDIVIDUAL REVIEWER COMMENTS

Reviewer 1 peer review comments

I have completed a review of the forage fish synthesis document. As requested, my comments are organized in three sections:

1. Responses to the six questions provided in the scope of work for the review;
2. Overall ratings and recommendations on the forage fish synthesis document sections;
3. Page/line-specific comments.

The review was a challenge as many aspects of the methods and results are not clearly explained and in discussion of findings there is generally limited interpretation of findings or use of relevant literature/existing ideas for support. Thus, it is largely left to the reader to infer what was done and what it might mean.

It seems to have been assumed that readers would be essentially familiar with all of the methods used.

Further, although this is presented as a lines of evidence approach, clear connections between sections are not always made and potential disagreements among sections are not resolved (especially from Sections 3 through 6).

I hope that my comments and suggestions are useful for improving the methods used and the overall presentation and interpretation of findings and conclusions. If I can be of further assistance, please let me know.

Response to Question 1

1. Does the forage fish synthesis document adequately address the overall objective of the document, which is to present lines of evidence for broader examination of the relationship between forage fish abundance, river flow, and tern productivity on the central Platte River?

No, although putative lines of evidence are presented in the forage fish synthesis document (FFSD), not all lines address the hypotheses presented in the Adaptive Management Plan.
Section 2

The clearest example is in Section 2, which also provides a fundamental relation used in subsequent sections, so shortcomings in Section 2 have major implications for the entire study. Section 2 purportedly presents a test of priority hypotheses T2 and T2a from the Adaptive Management Plan:

"T2-Tern productivity is related to the number of prey fish (<3 inches) and fish numbers limit tern production below 800 cfs from May-Sept.
T2a-Flow rates influence the number and species diversity in tern prey base (fish)."

In the Adaptive Management Plan, these hypotheses are linked to the Conceptual Ecological Model hypothesis TP-4:

"TP-4: Existing river flows influenced by drought, floods, hydrocycling, etc., do/do not provide a sufficient forage base (invertebrate/fish recruitment, survival, and correct composition) throughout the central Platte River study reach for populations of terns and plovers during the nesting season."

Although the rationale is not explained, justified, or interpreted in Section 2 of the FFSD, these hypotheses are apparently interpreted to mean that mean daily flows on a specific day should be expected to determine the number of prey fish available on that same day. Thus, the analysis presented attempts to establish a relation of fish abundance on a given day to average river discharge on that same day. Not only is this an unorthodox approach for assessing fish populations, it fails to address hypotheses T2, T2a, and TP-4.

The number of fish present in a river does not change daily in response to changes in river discharge. Fishes are adapted to withstand variations in discharge. Thus, the number of fish present in the river on a given day is better predicted by the number of fish present on the previous day than the change in discharge between days. Given the same number of fish present from one day to the next (give or take), the effect of increasing discharge is to dilute the number of fish present. This could make fish less detectible or less capturable via seining, which is what was studied in Section 2. The reason this alternative issue was the focus of Section 2 is not explained or interpreted. Incorrectly, resulting evidence is presented as if it addresses the target hypotheses, although it does not.

The relation found, that fishes were more difficult to capture at higher flows, makes sense as the result of a simple dilution effect caused by increased river discharge (although this is not discussed). However, this does not address the hypothesis that fish populations were lowered due to increased river discharge. It also provides no information on how species diversity was affected by river discharge (Hypothesis T2a). The reason is that forage-fish populations change gradually (except perhaps in an extreme event, such as during complete channel drying). Changes typically occur over two temporal scales: seasonal and annual (Schlosser 1985). On a seasonal scale, forage-fish populations are usually lowest in late winter and early spring due to winter mortality. Populations increase during summer and into autumn through reproduction and recruitment. On an annual scale, forage-fish populations boom in years during which streamflows are favorable for reproduction and recruitment, whereas they crash in years with unfavorable streamflows.

From what I can tell from the Adaptive Management Plan (and from what makes sense within the discipline of stream-fish ecology), hypotheses T2, T2a, and TP-4 are collectively aimed at understanding the annual scale of forage-fish production. They also specifically mention species diversity, which is not addressed in Section 2 and would also be expected to also vary on seasonal and annual scales. Wording in the hypotheses and from supporting figures in the Adaptive Management Plan (also provided in Section 1 of the FFSD) refers to the annual least tern breeding season as the period of interest. Wording in the hypotheses emphasizes fish population size (i.e. numbers of fish) over density, whereas in Section 2 of the FFSD, only daily density is addressed. The hypotheses (especially TP-4) specifically mention drought, floods, etc., which are not addressed in Section 2 of the FFSD. A quote from the USFWS Biological Opinion (2006), provided in Section 1 of the FFSD (line 92), refers to "such a flow event", which indicates an emphasis on the potential for an event within the overall flow regime of the least tern breeding season, not just flows on a single day, to influence subsequent forage-fish abundance.

Supporting graphs provided in the Adaptive Management Plan (copied into Section 1 of the FFSD, page 6) indicate a focus on the total number of individual fish present (not their density) and on fish-species diversity. These are not daily metrics, but are seasonal to annual. The figures also indicate a focus on the entire least tern breeding season (May-Sept.), not just on individual days.

Further evidence that hypotheses T2, T2a, and TP-4 are focused on annual patterns in forage-fish populations comes from the 2006 Biological Opinion. In the Riverine Foraging Habitat portion of the Environmental Baseline section (page 205) it is stated:

"Periodic low summer flows, coupled with high temperatures, are believed to be a critical factor in determining the abundance and diversity of the central Platte River forage fish community."

Section 2 of the FFSD does not address relations of the occurrence of periodic low summer flows. It also does not address effects of temperature.

Emphasis on the importance of flow regimes to fish populations (rather than importance of daily discharge) has a long history in the science of stream-fish ecology. Many studies document effects of the annual flow regime on reproduction and recruitment of forage fishes. For instance, Starrett (1951) showed that streamflow during the reproductive season could dramatically affect forage-fish populations. Schlosser (1985) further detailed seasonal and inter-annual trends in forage-fish abundance and showed that flow regimes strongly influenced recruitment. More recently on the Great Plains, Durham & Wilde (2009) showed that some fishes require flowing water for recruitment and Hoagstrom et al. (2008) showed that populations can decline and recover over consecutive years of unfavorable (i.e. dewatering) or favorable (i.e. stable flow) conditions, respectively. Closer to the Platte River, Moore & Thorp (2008) concluded moderate flows were ideal for fish recruitment because they provided high habitat heterogeneity and abundance. Consistent with hypotheses in the Biological Opinion and Adaptive Management Plan, studies in Great Plains rivers have shown that streamflow intermittence combined with high temperatures can cause heavy mortality of riverine forage fishes (e.g., Durham et al. 2006). Many other studies could be mentioned, but a literature review is beyond the scope of this peer review. In any case, literature on population dynamics of forage fishes in streams and rivers is consistent with hypotheses described in the Adaptive Management Plan. But, the hypothesis actually tested in Section 2 of the FFSD--how does mean daily discharge relate to forage fish density on that day--is unrelated to those of the Adaptive Management Plan and is of unclear relevance to forage fish production, yet this apparent disconnect is neither explained nor discussed.

The importance of the a priori hypotheses T2, T2a, and TP-4 is further addressed in the "Additional Hypotheses Identified" section of the Adaptive Management Plan (page 3):

"-Support fish/aquatic community (prevent loss of richness of aquatic species, especially fish and mollusks; support biological processes, which sustain fish and aquatic organisms dependent on certain flows; support spawning fish and other responses of the aquatic community; influence fish reproductive behavior and the availability and quality of spawning, nursery, and rearing habitat, including backwater habitat of fishes; maintain and prevent loss of the native fish community and will promote survival of fish young-of-the-year; promote critical stages in the life cycles of fishes and other aquatic organisms)"

The emphasis in Section 2 of the FFSD is only on daily density of forage fishes in relation to daily discharge. There is no mention of processes that sustain forage-fish populations. A properly focused study would have considered how annual flow regimes through the Associated Habitat Reach affected fish populations in the study area. This study would need to address over-winter survival of fish that could be both potential prey in the spring and brood stock to produce a new year class of forage fish for the ensuing summer, fall, and winter. Such a study would also need to address annual reproduction and recruitment that would affect young-of-year survival and growth in summer, autumn, and winter.
Factors known to affect these processes include (as mentioned in the hypotheses) floods and droughts.

The specific focus on 800 cfs as a minimum flow in hypotheses T2, T2a, or TP-4 is apparently associated with the possibility that river discharge below this threshold may not: (1) maintain stable water temperatures, (2) support adequate nutrient cycling, or (3) allow adequate fish movement in response to changing environmental conditions (Figure T2 in Adaptive Management Plan and Section 1 of the FFSD). Thus, a study focused on these hypotheses would evaluate the effects of days with discharge < 800 cfs on subsequent fish populations. Further, it would also address (if possible) effects of days with discharge < 800 cfs on water temperatures, nutrient availability, and fish mobility. The data available may not be adequate to address each of these questions, but then that should be noted and evidence available should be considered with caution until all these factors can be considered with relevant data.

Given that the flow versus density relation established in Section 2 has no clear relation to hypotheses T2, T2a, and TP-4, subsequent analyses based on this relation are, by definition, focused only on the relation of daily fish density to daily river discharge, not on the impacts of flow regime on fish populations. Thus, they also do not address hypotheses T2, T2a, or TP-4.

Section 3

This putative line of evidence suggests forage fish abundance does not influence least tern foraging success. However, this is based on the failure of the instantaneous fish density versus instantaneous river discharge relation (from Section 2) to predict least tern foraging success. As discussed above, this relation does not link flow regime through the Associated Habitat Reach with forage-fish population success. It only shows that fishes are more difficult to seine as discharge increases. Thus, analyses in Section 3 only show that dilution of forage fish by increasing discharge does not affect least tern foraging success. Whether or not forage-fish population change affects least tern foraging success was not assessed. Indeed, it was not established that forage fish populations varied during the study period, in which case there would be no reason to expect a relation. If new analyses (e.g., year-to-year comparisons) were conducted and were able to detect variation in forage-fish populations among years, then this information could be compared with associated least tern foraging data to seek a possible relation of one with the other.

Section 4

The lines of evidence approach breaks down in Section 4 because the apparent finding that lower river discharge, < 802 cfs, favors least tern fledging (lines 665-667) counters the apparent finding in Section 3 that river discharge probably has no effect on foraging success (lines 511-520). There could be an explanation for this apparent contradiction, but none is provided. Thus, as presented, one line of evidence does not follow from or lead to the next and there is an issue of internal inconsistency within the FFSD.

Section 5

The forage-fish abundance part of Section 5 is not appropriate for use in a bioenergetics model for least terns. First, the model only establishes a relation between daily discharge and daily capture efficiency with seines. This is neither a predictive model of fish abundance from year to year (as discussed elsewhere in this review), nor is the method used to develop this model without methodological concerns (also discussed elsewhere).

Further, in Section 3 it is suggested that river discharge has no effect on foraging success (lines 522-520). If this line of evidence were taken at face value, then there would be no reason to include a forage-fish- abundance relation in the bioenergetics model. But apparently, the finding in Section 3 is not applied in Section 5, where the forage-fish abundance relation is incorporated. This does not seem to follow a lines of evidence approach. If Section 3 is correct and river discharge does not affect least tern foraging success, then the availability of forage fish to least terns could be treated as a constant in the bioenergetics model.

If analyses for previous sections (especially Section 2) were re-done in ways that actually tested hypotheses T2, T2a, and TP-4, it might be possible to include a forage-fish abundance relation in the bioenergetics model. This would require an understanding of how flow regime, especially low-flow events, potentially cause forage-fish kills that reduce forage fish abundance during the breeding season.

References

Durham, B. W., Wilde, G. R., & Pope, K. L. (2006). Temperature-caused fish kill in a flowing Great Plains river. The Southwestern Naturalist, 51(3), 397-401.
Durham, B. W., & Wilde, G. R. (2009). Effects of streamflow and intermittency on the reproductive success of two broadcast-spawning cyprinid fishes. Copeia, 2009(1), 21-28.

Hoagstrom, C. W., Brooks, J. E., & Davenport, S. R. (2008). Spatiotemporal population trends of Notropis simus pecosensis in relation to habitat conditions and the annual flow regime of the Pecos River, 1992-2005. Copeia, 2008(1), 5-15.

Moore, S. L., & Thorp, J. H. (2008). Coping with hydrogeomorphic variations in a prairie river: resiliency in young-of-the-year fishes. River Research and Applications, 24(3), 267-278.

Schlosser, I. J. (1985). Flow regime, juvenile abundance, and the assemblage structure of stream fishes. Ecology, 1484-1490.

Starrett, W. C. (1951). Some factors affecting the abundance of minnows in the Des Moines River, Iowa. Ecology, 13-27.

Response to Question 2

2. Do the authors of the forage fish synthesis document draw reasonable and scientifically sound conclusions from the information presented? If not, please identify those that are not and the specifics of each situation.

Section 2

No, information presented in Section 2 of the FFSD is based on forage-fish abundance by seine haul, but is interpreted to represent an estimate of overall forage-fish abundance in the Associated Habitat Reach. This is an unorthodox and problematic use of the fish-survey data. Stream and river fish distributions are highly dynamic and can vary greatly from one mesohabitat to another. In addition, the vagaries of sampling with seines can lead to high captures in fortuitous circumstances (e.g. where fishes are easily cornered or surprised) and low captures under the opposite, more challenging circumstances.

For these reasons, fish sampling to estimate abundance uses a standard approach of conducting multiple habitat samples within a study site (Rabeni et al. 2009). An attempt is made to sample multiple representatives of all mesohabitats available at a site to ensure representativeness of the cumulative site sample. Individual seine samples are pooled to determine an overall density estimate either by number per seine haul, if seine effort is standardized, or by number per area seined (i.e., density). This is not the approach used in Section 2 of the FFSD, where seine hauls are treated as individual estimates of forage-fish abundance.

Use of individual seine hauls to represent overall forage-fish density could be considered an instance of pseudoreplication. In any case, the relation presented in Section 2 of the FFSD cannot be considered representative of forage-fish population size. What it evaluates, instead, is how fish concentration (and catchability) within individual seine hauls differs with river discharge. Not surprisingly, the data in Figure 6 (page 18) show that higher concentrations of forage fishes occur in seine samples at lower discharge. This is because as flow recedes, fishes concentrate in remnant, favorable habitats. Simultaneously, lower water velocities and depths make seining easier. Thus, seine hauls among mesohabitats remaining at lower river discharge are more likely to produce high numbers as presented in Figure 6.

This does not mean there are more fish in the river that day than a few days before when river discharge was higher. It just means that fishes are more concentrated at lower river discharge. As discharge increases, there is a dilution effect (discussed above). Also, as a greater amount of suitable habitat becomes available, fishes become more widely dispersed. Finally, as already mentioned, seining becomes more challenging in deeper and swifter water. None of this is mentioned or discussed in Section 2 (although it should be).

To address hypotheses T2, T2a, and TP-4, seine samples should be combined within each study site. That is, sampling units should be site-scale samples (not individual seine hauls, which are mesohabitat- scale samples). If desired, the relation of fish density to mean daily discharge the day of sampling could be used to correct fish density estimates so that they are less biased by sampling conditions (using the relation presented in Section 2).

Regardless, site-scale abundance estimates (one per year) should then be evaluated with regard to the flow regime of the preceding year. Necessarily, this analysis should include the number of days < 800 cfs, as this was a specific hypotheses to be tested. Ideally, total number of days < 800 cfs, number of separate events < 800 cfs, and consecutive days < 800 cfs per event would all be tested individually as possible predictors of forage-fish abundance. However, it would also be good to consider other factors such as flood events and lower discharge thresholds (e.g., 500 cfs), because these may also be important predictors (e.g., see references in the previous section). This would be useful to determine whether 800 cfs is truly the most valuable predictor, or if some other metric is actually more meaningful. There is extensive literature on interactions of fish populations with flow-regime characteristics. Helpful reviews are available in volumes on freshwater fish ecology (e.g., Matthews 1998; Ross 2013).

It is also important to determine whether individual site samples are generally representative of the Associated Habitat Reach. This could be done by conducting among-site-and-year comparisons for all sampling sites. A cluster analysis or multivariate technique (e.g. NMDS) using a similarity index (e.g. Sorensen Index) could be used for this.

In Section 2 it was assumed that all study sites were representative of the entire Associated Habitat Reach, but this may not be the case. It is well known that fish assemblages differ in different portions of the Platte River depending on habitat conditions, which in large part reflect local impacts of modifications by humans (e.g., Lynch & Roh 1996). Further, maps of the Associated Habitat Reach (Figures 1 & 4) indicate there are multiple features likely to influence streamflow and habitat conditions (i.e., hydropower return, diversion dam), so similarity in fish abundance patterns among study sites must be established first before information among sites can be combined. If it is found that reaches are distinct, then distinct reaches should be defined and analyses should be conducted separately by reach.

The relative importance of distinct forage-fish reaches should also be considered in light of patterns of least tern foraging among reaches. Documentation of different spatial trends in fish populations could be critical for understanding vagaries of least tern foraging behaviors (as discussed in Section 4). For instance, if least terns concentrate foraging in an area with a distinct fish assemblage, then factors in that specific area associated with forage-fish abundance or availability may provide important insight into needs of least terns. These possibilities are overlooked in analyses that fail to consider among-site differences (as in Section 2).

This issue is actually broached later in the FFSD (Section 4, lines 577-593), where unique conditions at the Kearney Canal Diversion are described. If this study site has a unique fish assemblage, it should not be grouped with other sampling locations. Later in Section 4 (lines 615-654), analytical results are interpreted to suggest that this area is a particularly important foraging area, unlike any other in the Associated Habitat Reach. This strongly supports the argument that forage-fish assemblages should not be presumed to be identical throughout the Associated Habitat Reach. This background and discussion should be moved into Section 1 of the FFSD and should then be considered in analyses conducted within all subsequent sections, not just in Section 4.

Another important detail has to do with coverage of mesohabitats during fish sampling. In Section 1 of the FFSD, it is described that fish sampling was conducted in "open channel areas". These are described as within "...the flowing portion of the active channel area..." (line 198). What specific open channel habitats were sampled and what specific closed channel areas were not sampled is not clearly explained. This makes it unclear what the fish sampling data actually represent.

The list of forage fishes (Section 1, page 12) includes species commonly associated with flowing water and active channel areas. These species include red shiner, sand shiner, and bigmouth shiner (although red shiner will also associate with slackwater habitat). But, the list also includes species that commonly use slackwater habitats adjacent to the active river channel. These species include brassy minnow, western mosquitofish, and plains killifish. Which of these species is most important for least tern?

Which was most prevalent in the forage-fish samples? Were habitats used by all species sampled equally? Which habitats were actually used by foraging least terns?

Although in Section 3, considerable verbiage is expended to explain why fish abundance could not be determined instantaneously at a given location (which is certainly true, lines 371-381), what is overlooked is that the mesohabitat within which least terns forage should be readily observed and, if desired, characteristics there could easily be measured and the range of habitats typically fished by least terns could be quantified (e.g., the range of depth-velocity combinations typically fished could be determined and then these areas could be sampled to understand forage-fish availability at a given time; in addition, with knowledge of local fluvial geomorphology, the relation of suitable foraging habitat to river discharge could be modeled). With this information, sampling could then focus on understanding the abundance of forage fish in mesohabitats that are actually fished. This knowledge could produce important insight into the link between forage-fish abundance, river discharge, and least tern foraging.

However, the production of forage fish does not necessarily rely only on mesohabitats actually fished by least terns. Forage-fish production requires spawning habitats, nursery habitats, and juvenile-adult habitats (i.e., habitat complementation; Schlosser 1995). Even though least terns might not forage there, fish populations found in these habitats represent the potential food base because without recruitment, there will be no juveniles or adults to occupy the open-channel habitat where least terns reportedly forage. For instance, shallow slackwater habitats are critical nurseries for forage fishes (e.g., Moore & Thorp 2008). Further, if least terns rely on all the species listed in Section 2 of the FFSD, it suggest that all mesohabitats available (except perhaps deep pools or very swift channels) are relevant for producing forage-fish populations, whether or not least actually forage in all of those habitats.

Thus, although it would be good for forage-fish sampling to include information on fish populations from mesohabitats actually fished by least terns, in order to determine overall forage fish abundance, there should either be forage-fish sampling within mesohabitats fished by least terns throughout the breeding season or single yearly samples should include all mesohabitats to estimate not just the abundance of forage fishes in fished habitats at that time, but the overall population of forage fishes present that year. Even so, a single yearly sample will be unable to represent fish abundance very accurately because whatever time of year sampling occurs will not necessarily represent other seasons. For instance, a small population in the spring could still produce a large population in the autumn if conditions are ideal for reproduction and recruitment. In this case, a spring or early summer sample might accurately represent the spring, but not the autumn. An autumn sample would have the opposite bias.

From information provided in Section 2, it is not clear what exactly the fish sampling data represent. This should be explained and assessed in more detail before the data are used to model fish abundance. It would be relevant to know what mesohabitats were actually sampled, the abundance of each species collected, and the size distribution of each species collected. For instance, if the species collected only were open-water species and if the sizes collected only were relatively large juveniles and adults, then the samples would not represent the entire food base available to least terns (because least terns consume more than just open-water species) and they also would not necessarily represent the previous or subsequent food base available over the entire breeding season (because the abundance of smaller recruits that would produce forage later in the year was not assessed and the abundance of forage fish in previous months is unknown). In this case (for example), fish surveys would truly represent only "instantaneous" samples and would not represent the food base over an entire least tern breeding season, just during one part of it. Assuming least terns need adequate forage prior to and throughout the breeding season (not just during part of it), it is unclear whether the forage-fish sampling data can be used in a meaningful way to represent forage fish available throughout the breeding season. This potential shortcoming should be addressed in the FFSD and should be accounted for in the analyses, if possible. If this cannot be done, then the available data may be inadequate for the task at hand.

In summary, to understand forage-fish availability, it must first be recognized that forage fishes are short lived and, as a result, population fluctuations occur within and among years based on vagaries of reproduction and recruitment, which are subject to ambient conditions, especially the seasonal and annual flow regime. Flow-regime characteristics can elevate or reduce instream productivity, which may indirectly affect forage-fish production. Flow-regime characteristics may also directly affect forage fish production by providing ideal or unfavorable conditions for reproduction and recruitment, which may vary among species. Thus, the identity of species may also be relevant both with regard to their relative importance as least tern forage and their differing responses to flow-regime characteristics. Section 2 as written not only has a series of methodological shortcoming already discussed, but it completely overlooks all of these important aspects of forage-fish ecology and thus provides little if any insight into forage-fish availability for least terns.

Section 3

The conclusion that forage-fish abundance does not influence least tern foraging success (lines 465-470) is unsupportable because it is based on the model from Section 2, which only models the dilution effect of river discharge on forage fish densities on a given day. Models M1 and M3 in

Section 3 do not incorporate actual differences in forage-fish populations and, in fact, analyses in Section 2 did not assess whether forage-fish populations varied or not during the study period (as discussed in comments on Section 2). They only demonstrated that higher river discharge at the time of sampling was associated with reduced fish density in seine collections. Thus, the implication of findings in Section 3 is only that least terns can forage with equal effectiveness at a range of river discharges, despite the fact that higher streamflows appear to dilute the density of forage fishes. This suggests that least terns are well-adapted fish predators, which is not surprising since they have dealt with variable streamflows throughout their evolutionary history.

To determine whether forage fish abundance was associated with changes in least tern foraging success, it would first be necessary to establish that forage-fish abundance was variable in the study period. To do this, forage fish abundance in site-scale samples (seine hauls combined) should be compared among years, perhaps corrected for differences in sampling efficiency associated with discharge at the time of sampling (using the relation found in Section 2). If differences were found, then variation in least tern foraging efficiency could be compared among years vis à vis forage-fish abundance. As discussed above, this analysis should be conducted separately for each sampling site, unless justification can be provided for combining different sites into a single sample (i.e., what length of river does each forage-fish sample site represent?).

Section 4

Again, the river discharge-fish abundance relation established in Section 2 is only a measure of fish population dilution via increasing discharge. Thus, the conclusion (lines 665-667) that river discharge < 802 cfs was associated with higher fledging probability simply suggests that more concentrated fish populations are more favorable for least tern foraging (a logical conclusion). This finding, however, contradicts that in Section 3 where it was concluded that increased river discharge did not affect least tern foraging success. Thus, it would be prudent to include a discussion of the possibility that the relation found in Section 4 could be spurious. Or else, there must be some flaw in the relation found in Section 3.

Available evidence suggests findings in Section 3 are likely to be incorrect. According to the Biological Opinion (USFWS 2006, page 105):

"Nesting coincides with the timing of lowest flows in major river systems (historically, in the summer), providing easy access to forage species for least tern (Dugger 1997)."

This supports the findings in Section 4, but contradicts the findings in Section 3. This also suggests the FFSD would benefit from additional use of available literature on least tern breeding and foraging (i.e., Dugger 1997 is not cited in the FFSD).

In any case, it should not be implied that this study assessed how a forage-fish shortage might affect fledging success. There is no point in the FFSD where it was established that forage fish populations were ever depressed, simply because this possibility was never assessed. Thus, this study did not address effects of forage-fish population declines on least tern foraging or production.

Section 5

The bioenergetics model seems extremely simplistic and tenuous. The perception/conclusion that forage-fish abundance is constant among years and throughout the breeding season is simply unrealistic (as discussed elsewhere). Thus, there is no assessment, within the FFSD, as to whether forage fish populations suffering from a fish kill could lead to inadequate foraging for least terns. This issue is not even mentioned. Concerns expressed in the Biological Opinion (2006), that low flows and high temperatures could affect least tern production by limiting available forage, is simply not addressed.

References

Lynch, J. D., & Roh, B. R. (1996). An ichthyological survey of the forks of the Platte River in western Nebraska. Transactions of the Nebraska Academy of Sciences, 23: 65-84.
Matthews, W. J. (1998). Patterns in Freshwater Fish Ecology. Chapman & Hall, New York.

Moore, S. L., & Thorp, J. H. (2008). Coping with hydrogeomorphic variations in a prairie river: resiliency in young-of-the-year fishes. River Research and Applications, 24(3), 267-278.

Rabeni, C. F., Lyons, J., Mercado-Silva, N., & Peterson, J. T. (2009). Chapter 4: Warmwater fish in wadeable streams. Pages 43-58 in Bonar, S. A., Hubert, W. A., & Willis, D. W. (editors). Standard Methods for Sampling North American Freshwater Fishes. American Fisheries Society, Bethesda, Maryland.

Ross, S. T. (2013). Ecology of North American Freshwater Fishes. University of California Press, Berkeley.

Schlosser, I. J. (1995). Critical landscape attributes that influence fish population dynamics in headwater streams. Hydrobiologia 303, 71-81.

Response to Question 3

3. Are the statistical methods used in each section appropriate?

Section 2

No, the regression analysis in Section 2 provides only one possible relation without justification. If there is a reason to believe that only a negative logarithmic relation of fish density to river discharge is present, then the prior studies that have established this relation should be summarized and cited. If not, then alternative possibilities should be considered. For instance, a null model (no change in density versus river discharge) should be considered simply as a way to establish the validity of alternative models (they must outperform the null model to be useful). In addition, a linear model seems most reasonable because an increase in discharge should dilute the ratio of fish biomass to water volume in a linear fashion. Thus, a linear model should be presented and its performance should be compared to the logarithmic model. Other alternatives may also be reasonable and should be considered. In any case, the negative logarithmic model should be justified and if, upon further analysis, it proves to be the best fitting model, then the rationale for why this model makes sense should be provided (i.e., what mechanism creates this relation?). There is no discussion of this at present.

Statistics of model fit should also be provided. Without them, it is difficult for the reader to fully understand the strength of the models presented. These statistics should be provided and discussed in detail (e.g., Burnham & Anderson 2002).

However, as already discussed, these analyses don't actually address the hypotheses T2, T2a, or TP-4, so these technical issues may be moot once conceptual issues are addressed. But, they are mentioned here in the interest of thoroughness.

Section 3

As in Section 2, the basis of model development and selection is not provided. There are no references associated with the models presented, so it appears these models were developed in this study for the first time. If so, then there should be an increased effort to justify and explain the models used.
However, it also seems that it would be highly beneficial to assess multiple models rather than assuming a priori that a single model (presented without justification) is useful.

Section 3 is superior to Section 2 in the sense that there is more discussion of analytical methods and ultimately multiple models are discussed. However, the meaning/justification for each model is not presented in detail. Further, as Bayesian methods are not familiar to everyone and the reference provided is "in press" and may not be readily available, it would be a major improvement if the metrics used for model comparison were explained in a conceptual manner so that readers could better understand the results. As it is, "scores" are provided as results (lines 415-418), but the manner in which scores are compared, interpreted, and evaluated is not explained. Thus, the conclusions reached based on these scores seem vague. Scores should be explained to the degree that readers can understand how they indicate model accuracy as well as how they are compared amongst each other.

Additional metrics of model fit may be needed for this discussion. A table comparing various features of fit exhibited by the multiple models would be highly beneficial.
Section 4

In Section 4 (lines 564-566), it is claimed, again, that fish abundance on a given day is determined by discharge on that day. This claim demonstrates a lack of understanding of stream-fish ecology (as already discussed). Stream fish populations are not determined on a daily basis. They are determined by factors associated with survival, reproduction, and recruitment that occur over scales of weeks to months to years. Hypotheses T2, T2a, and TP-4 do not imply that fish populations are determined on a daily basis by daily discharge. This is simply a misconception/disconnect applied mistakenly throughout the entire FFSD. Thus, as in previous sections, methods in Section 4 do not address hypotheses T2, T2a, and TP-4 (despite claims that they do).

In addition, neither data nor rationale are provided to justify emphasis on only the 7-day period prior to the date of fate determination. What is the biological significance of 7 days? This should be explained, preferably with reference to prior studies that have established its significance. As presented, it appears to be a haphazard inference with no biological basis.

One advantage of Section 4 is that a multi-model inference approach is apparently employed. However, the overall study design and set of competing alternative models is not very clearly described (again, a table would help). Burnham & Anderson (2002) provide specific instructions for presenting the results of analyses used in Section 4 (including example tables) and it would be highly beneficial if these instructions were followed throughout the FFSD.

Section 5

As already mentioned, the model produced in Section 2 does not allow for any seasonal or annual variation in forage-fish populations. Thus, forage fish abundance is assumed throughout to be constant except as affected by daily discharge, which dilutes fish populations and possibly makes fishes less available to least terns. Hence, the bioenergetics model is only relevant for the fish abundance level represented by the average of all forage-fish samplings across years (these were the data used in Section 2). If these collections truly represent forage-fish populations and, by inference, fish populations in the Associated Habitat Reach are invariable, then the model could be meaningful, but there was no assessment of annual variation within the forage-fish survey data (and those data were used inappropriately, by individual seine haul), so there is no way to know what actual trends existed among years. Thus, the basis for determining forage-fish abundance for bioenergetics modeling needs to be re-assessed, beginning with a re-assessment of methods used in Section 2 (as already discussed).

References

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.

Response to Question 4

4. Are potential biases, errors, or uncertainties appropriately considered within the methods sections of this document and then discussed in the results and conclusion sections?

Section 2

No, as already discussed at length, the biases and uncertainties associated with forage-fish sampling are not even mentioned, much less considered or discussed. More critically, the actual goals of Section 2 are not clearly presented and the analyses and results presented do not articulate with the stated goal.

Potential biases, errors, uncertainties:

1. How can a single seine haul be legitimately used to represent the entire forage fish population in the Associated Habitat Reach? Given that this is an unacceptable methodology for fish- population sampling, how can its use be justified in this case?
2. If question 1 can be adequately addressed (likely it cannot be), does the negative logarithmic model outperform null and linear models?
    a. What other reasonable models could be considered?
    b. What specific statistics of model-fit support the conclusion that the negative logarithmic model both provides a meaningful fit and outperforms alternative models?
    c. If it does outperform other models, why does this make sense?
3. Can a single annual sample of forage fishes truly represent overall forage-fish abundance throughout the least tern breeding season given that forage-fish populations are dynamic and the breeding season is long?
4. What fishes (species and sizes) did forage-fish sampling actually represent?
    a. Are these species and sizes of importance to least terns?
    b. Do these represent only a part of the forage fish population at a site (due to habitat bias in sampling) or do they represent all forage fishes?
5. Can forage-fish samples from different study sites be combined (i.e., are they quantitatively similar in species and size-class composition)? Or, are there distinct reaches within the Associated Habitat Reach, each with distinct fish assemblages, possibly associated with human- constructed features (e.g., diversion dams, irrigation returns, hydropower returns)?
    a. If there are distinct reaches, how do they correspond to patterns of least tern foraging?
6. If questions 3-5 can be addressed, how do the number of days (cumulative) < 800 cfs during the previous year correspond to forage-fish abundance at the time of sampling in each distinct reach of the Associated Habitat Reach? What about the frequency and duration of events < 800 cfs?
    a. Are there other metrics of river discharge (e.g., flood events, lower thresholds of discharge) that outperform the 800 cfs threshold metric?

Section 3

There is some discussion of using data from two different study designs for least tern foraging activity (systematic versus opportunistic). Despite this, results from the two separate designs are combined. Then, although the null model provides the best fit, there is substantial discussion of a model with a poorer (but similar) fit to the null. While the overall conclusion of this section seems reasonable, that river discharge does not affect tern foraging success (over the range of flows studied), the amount of attention given to the model with the poorer fit (M2) seems to imply that there is evidence for an effect of flow (although there probably is not). This issue is complicated by the fact that the M2 model seems to be influenced by the opportunistic samples (Figure 10). If so, then this relation is not the result of a foraging success - river discharge relation, but instead is an artifact of the combined use of data from two different sampling designs, in which case it has no meaning in the study context.

To assess this, the two data sets should be assessed separately before being combined. It is mentioned in Section 1 (line 239) that these data were combined to bolster sample size, so apparently the modeling performed in Section 3 could not be done separately on individual data sets. Even if this is true, it should be possible to simply plot the two different data sets and make some visual or simple statistical comparisons (i.e., compare regression slopes or shapes) to determine if they can be appropriately combined. If this is not possible, then it is not clear if there is any validity to the results presented in Section 3. Perhaps a more simple statistical approach should be used that is more appropriate for the limited data available and recommendations should be made regarding what data are needed to support more sophisticated analyses.

If the analysis were left to stand, then it should be made evident that the data are marginal at best for addressing the section hypotheses and findings should be considered preliminary and only useful as a guide for future studies, not as a basis for management decisions. Further, relatively low model fits produced for models including forage-fish abundance (lines 417-418, models M1 & M3) are based on the relation established in Section 2. This relation, as already discussed, simply reflects the concentration of fishes at a given discharge, assuming fish populations are constant (this model doesn't allow for intra- or interannual variation in population size). Again, it models a simple dilution-effect relation and is not an indicator of fish population size. Thus, it is possible this relation is too simplistic to model the relation of forage-fish abundance to least tern foraging success. That is, if fish abundance did differ by season or year (which is possible, but unknown), then these factors were not accounted for in the model. This limitation alone could explain why no relation of least tern foraging success with fish abundance was found. Because the relation presented in Section 2 did not actually assess variation in fish-population size, its use in Section 3 does not actually assess the influence of forage-fish abundance on least tern foraging success.

Section 4

As already mentioned, the fact that Section 4 suggests lower river discharge purportedly increases fledging success seem to contradict the fact that Section 3 suggests river discharge does not affect foraging success. Thus, the actual mechanism of how river discharge affects fledging success is unclear.

If Section 3 is correct, then the effect of river discharge on fledging success must not be related to least tern foraging. Perhaps lower river discharge was associated with some other factor that benefitted least terns. These possibilities should be discussed and the validity of the assumption that river discharge can be used as a surrogate for forage-fish abundance (lines 564-565) should be re-considered.

Section 5

Tremendous uncertainly in the bioenergetics model is admitted, but the potential flaws in the forage- fish abundance estimates are not considered. In short, as discussed elsewhere in this review, the forage-fish assessment provided in the FFSD does not address the target hypotheses and does not use appropriate approaches for assessing forage-fish abundance. The emphasis only on how daily river discharge affects daily captures of forage fishes does not provide an understanding of whether or not limiting flow events (e.g., low discharge coupled with high temperatures) can reduce forage-fish abundance to a level low enough to limit annual least tern production.

Response to Question 5

5. Are the conclusions drawn supported by the synthesis of data reported in Sections 2 – 5?

This question seems to reference Section 6. There seems to be a single conclusion drawn in this section (lines 849-852):

"Results of our approach indicate there continues to be no evidence to support the relationships postulated in Hypotheses T2 and T2a. Instead, our results indicate forage fish abundance and least tern productivity increase as flows decrease to 5.7 m3s-1 (200 ft3s-1)."

The conclusion in the second sentence is not supported in this study. The results show that forage fish are more abundant in individual seine hauls as river discharge decreases. As discussed above at length, this does not indicate that fish population numbers are increased (as postulated in Hypotheses T2 and T2a). Rather, results from this study simply show that decreasing discharge is associated with increased capture efficiencies of forage fishes. This is a reasonable result, but has no bearing on how flow regime affects forage-fish production. A more accurate and supportable version of the second sentence would be:
Our results indicate that forage fishes become more concentrated in their within-site distribution as river discharge decreases. This may make them more available to least terns and increase fledging success in periods with lower river discharge, but our analyses indicate increasing discharge did not decrease least tern foraging success, so it is possible there is some other explanation for increased fledging success that is unrelated to forage-fish density.

This finding appears to support previous studies cited in the Biological Opinion (USFWS 2006, page 105):

"Nesting coincides with the timing of lowest flows in major river systems (historically, in the summer), providing easy access to forage species for least tern (Dugger 1997)."
This may be a useful or interesting result with regard to least tern management, but it does not address Hypotheses T2 and T2a.

The conclusion in the first sentence quoted above misrepresents the significance of the present study. It implies that this study directly assessed Hypotheses T2 and T2a and failed to provide support. However, as discussed at length above (in answers to previous questions), this is not the case. A more accurate and supportable version of the first sentence would be:

The relationships postulated in Hypotheses T2 and T2a remain untested. In order to test them, there needs to be determination of site-scale spatial variation in forage fish assemblages within the Associated Habitat Reach to identify unique sub-reaches, if present. Also, documentation of mesohabitats used during least-tern foraging would allow forage-fish sampling to better assess forage availability within habitats specifically used by least terns. If sampling to estimate forage-fish abundance for least terns was focused on habitats fished by least terns, the sampling frequency should be seasonal (e.g., monthly), to ensure that forage availability is assessed throughout the breeding season. Alternatively, if less frequent (i.e., quarterly) forage-fish sampling is desired, it should be focused on all mesohabitats that potentially contribute to the least tern food base, including nursery habitats (i.e., slackwaters) important for production of forage fishes, whether or not they are actually fished by least terns.

Sampling that documented fish-population cycles throughout the least tern breeding season in the Associated Habitat Reach overall and specifically within habitat typically fished by least terns would allow development of a sampling methodology that accurately represents forage- fish abundance and availability to least terns in each sub-reach for the entire period that affects breeding success. Then, assessment could be made of the relation of flow-regime characteristics to the size of each forage-fish population (species by species) in each sub reach, throughout the least tern breeding season, including an assessment of how the frequency and duration of events < 800 cfs affect abundance of each species. Once these factors are determined and a multi-year record of forage fish population data is available, patterns of forage-fish abundance could be compared with spatial patterns of least tern foraging, food choice (species and size) by least terns, temporal patterns of least tern foraging success, and annual patterns of least tern production to determine how flow regime influences least tern production. At present, there is not adequate information available to assess these hypotheses.

It may be that using the present data some of these issues can be addressed and suggestions to remedy this are provided above. In that case, perhaps conclusions could be drawn to specifically address Hypotheses T2 and T2a. At present, this study did not assess those hypotheses and thus the conclusions should not imply that it did.

Response to Question 6

6. Did we utilize the data we had as effectively as possible to address the hypotheses "forage fish abundance below 800 cfs limits least tern productivity?" and "there is a direct relationship between flow and forage fish abundance and species diversity?" and answer the Big Question "Does forage availability limit tern productivity on the central Platte River"? If not, what additional analyses do you recommend?

No. Additional analyses are described and recommended in detail in preceding sections. In particular, the forage-fish data were not used appropriately or effectively. Fish abundance should have been reported as a cumulative density by annual sample at each site (not by individual seine haul). Sites should not have been combined unless shown to be quantitatively similar.

Similarly, the flow-regime was not analyzed in meaningful ways and, as a result, the effect of the hypothetical 800 cfs threshold was not assessed. Once the suitability of forage-fish data for representing forage-fish abundance was established (if possible, with uncertainties acknowledged), the potential for days < 800 cfs, number of events < 800 cfs, and duration of repeated events < 800 cfs to influence forage-fish abundance in a given year should have been assessed.

Instead, the instantaneous effect of river discharge on fish density within seine hauls was assessed using improper methods. This does not assess whether flow regime affects forage-fish production.

Several other conceptual and methodological issues are assessed above in the section by section reviews.

Peer Review Rating & Recommendation

SECTION 1 RATING (1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor)

Category Rating Score
Scientific soundness 2
Degree to which the study approach addresses the project's objectives 2
Organization and clarity 2
Conciseness 2
Important to objectives of the Program 1

SECTION 1 RECOMMENDATION (Check One)

Accept
Accept with revisions X
Unacceptable
SECTION 1 REQUIRED CHANGES

1. More discussion is needed on hydrologic and fluvial geomorphic differences within the study area.
   1.1. Discussion in Section 4 on the Kearney Diversion Canal location should be moved to Section 1 and thereafter considered in all analyses of all subsequent sections.
   1.2. Potential for forage-fish assemblages and least tern foraging to differ among sub-sections of the Associated Habitat Reach (and existing evidence) should be discussed/presented in Section 1 and analyzed throughout the FFSD.
      1.2.1. Forage-fish patterns of distribution and abundance could be presented here preliminary to more detailed assessment in Section 2.
      1.2.2. There is no value in assuming forage-fish populations and least tern foraging is evenly distributed throughout the Associated Habitat Reach unless this has been established by prior study (which should be explained in Section 1).

1.2.2.1. Assuming uniformity without justification creates potential to overlook important spatial patterns in forage-fish availability and least tern foraging preferences that could be critical for sustaining or restoring least tern production.

SECTION 2 RATING (1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor)

Category Rating Score
Scientific soundness 5
Degree to which the study approach addresses the project's objectives 5
Organization and clarity 4
Conciseness 5
Important to objectives of the Program 1

SECTION 2 RECOMMENDATION (Check One)

Accept
Accept with revisions
Unacceptable X

SECTION 2 REQUIRED CHANGES

1. Correctly analyze forage-fish monitoring data at the site scale (seine samples combined by site/date) to make population comparisons:
    1.1. among sites to determine if the forage-fish assemblage is contiguous or segregated into sub- assemblages;
    1.2. among years to determine if populations fluctuated or not; if populations did not fluctuate, then the only additional analysis (i.e., section) that would be needed in the FFSD would be the bioenergetics section (5). That is, in the absence of population fluctuation, the only relevant question remaining is whether the existing population is adequate to support least tern production;
    1.3. analyses focused on individual species would be more informative as individual species are not expected to respond identically to flow-regime changes. If this is not done, justification for combining species (presumably linked to least-tern foraging ecology) should be provided;
    1.4. species diversity should also be estimated by site for annual samples (Hypothesis T2a).
2. If there is evidence for population fluctuation, then analyses could be conducted to determine if flow-regime differences among years correspond to forage-fish abundance differences. This should be an orthodox type of analysis that recognizes that a flow regime over time (i.e., throughout the year or season) affects forage-fish populations by influencing survival, reproduction, and recruitment over weeks and months and from year to year (not day to day). That is, the focus on daily fish abundance in relation to daily discharge (the same day) is not an indicator of how flow regime influences forage-fish abundance.
    2.1. Analyses need not be restricted to, but must at least address how instances of river discharge < 800 cfs relate to forage-fish abundance (Hypothesis T2).

2.2. Temperature should also be considered if data are available (USFWS 2006).
2.3. Effects on species diversity should also be considered (Hypothesis T2a).

SECTION 3 RATING (1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor)

Category Rating Score
Scientific soundness 4
Degree to which the study approach addresses the project's objectives 4
Organization and clarity 3
Conciseness 3
Important to objectives of the Program 4

SECTION 3 RECOMMENDATION (Check One)

Accept
Accept with revisions X
Unacceptable

SECTION 3 REQUIRED CHANGES

1. Estimates of forage-fish abundance should be based on annual, site-scale (seine hauls combined at each site) estimates. The model presented in Section 2 does not assess how flow regime affects fish abundance, just how daily river discharge affects daily fish density. A new model that addresses annual forage-fish abundance versus annual flow regime should be developed and used in Section 3.
   1.1. Weaknesses of forage-fish sampling only once per year and only in open-channel habitats for estimating abundance of forage fishes at only 4 to 5 sites over the entire breeding season should be discussed and acknowledged.
   1.2. Sites should be analyzed separately unless it can be shown that the forage-fish assemblage is uniform between adjacent sites (which could then be combined).
2. Systematic and opportunistic samples of least tern foraging should not be combined unless it can be shown they represent a similar trend. Otherwise, they should be analyzed separately because combining them when they are incompatible could create spurious results that reflect only the effect of combining incompatible data sets (which appears to be the case in the data presented).
3. More rationale should be provided to justify use of regression models presented (with references if appropriate).
4. Statistics of model fit, with interpretation, should be provided.
   4.1. It should be clear which model(s) are best supported and most parsimonious.
   4.2. All supportable models should be used in subsequent analyses.
   4.3. Rationale for interpreting supported models should be provided (with references if appropriate).

SECTION 4 RATING (1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor)

Category Rating Score
Scientific soundness 5
Degree to which the study approach addresses the project's objectives 2
Organization and clarity 3
Conciseness 3
Important to objectives of the Program 1

SECTION 4 RECOMMENDATION (Check One)

Accept
Accept with revisions
Unacceptable X

SECTION 4 REQUIRED CHANGES

1. The validity of the 7-day period up to brood-fate determination as the only flow period evaluated should be provided, or analyses should include river discharge over the entire breeding period, or a suite of different periods should be analyzed and compared.
2. Use of only logistic regression models should be justified, or additional models should be analyzed and compared with the logistic models.
3. Mean daily river discharge on a single day is not a surrogate for fish abundance.
   3.1. Section 2 is in need of revision. If valid estimates of annual abundance can be produced for each site, then annual abundance should be used as a measure of forage-fish abundance. 3.1.1.It is irrational to presume that a flow over a 7-day period determines forage fish
4. abundance for least terns over their entire breeding season. Actual site-scale forage fish estimates should be used if available. Otherwise, some relation between a flow period and forage-fish abundance must be demonstrated. Without this, the presumed flow- forage fish relation has no basis and should be abandoned.
   4.1.1.1. Consider a 7-day period including a large rainstorm. If discharge is elevated within this period, is it this event alone that has determined the fish population present? Or, is it possible previous events were also influential and in fact may have had more influence? Is it possible the effect of increased discharge during the 7-day period might actually manifest at a later time due to influence or reproduction and recruitment?
5. To maintain a lines-of-evidence approach, contradictions between findings in Section 4 versus Section 3 must be addressed. If Section 3 indicates river discharge either does not affect least tern foraging (M4) or higher discharge enhances foraging (M2), why would least tern production increase at lower river discharge as concluded in Section 4?
   5.1. Does this mean some other variable correlated with river discharge (not forage-fish abundance) actually increases least tern production?

SECTION 5 RATING (1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor)

Category Rating Score
Scientific soundness 4
Degree to which the study approach addresses the project's objectives 2
Organization and clarity 3
Conciseness 3
Important to objectives of the Program 2

SECTION 5 RECOMMENDATION (Check One)

Accept
Accept with revisions X
Unacceptable

SECTION 5 REQUIRED CHANGES

1. Estimates of forage-fish abundance should be made at the site scale (all seine hauls combined by site and sampling event).
   1.1. Potential influence of sites should be analyzed separately unless justification can be provided that adjacent sites are similar and can be considered a single, contiguous forage-fish assemblage.
2. Disagreement between Sections 3 and 4 (after accurate forage-fish abundance data are incorporated into analyses) should be resolved.
   2.1. If Section 3 is more supportable, then forage-fish abundance could be a constant in the bioenergetics equation because forage-fish abundance apparently has no effect on least tern foraging.
   2.2. If Section 4 is more supportable, then annual estimates of forage-fish abundance by study site should be analyzed versus least tern bioenergetic demands to see if annual abundance was ever low enough to theoretically limit least tern production.
      2.2.1. s would ideally be done separately by study site and forage-fish species, unless it is shown that certain adjacent study sites share the same fish assemblage and/or that certain species follow the same population trends in response to flow regime.
      2.2.2. If annual abundance of some forage fish species at some study site is associate with least tern foraging success, then these findings could be considered in light of Section 4 findings to see if years with predicted energetic limitations actually corresponded with years of lower least tern productivity (i.e., lines-of-evidence approach).

SECTION 6 RATING (1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor)

Category Rating Score
Scientific soundness 5
Degree to which the study approach addresses the project's objectives 5
Organization and clarity 3
Conciseness 3

Important to objectives of the Program 1

SECTION 6 RECOMMENDATION (Check One)

Accept
Accept with revisions
Unacceptable X

SECTION 6 REQUIRED CHANGES

1. Because this section presumes the validity of prior sections, conclusions presented are invalid.
   1.1. Section 2 did not address effects of flow regime on forage-fish abundance, so conclusions that suggest Hypotheses T2 and T2a were not supported are incorrect. Rather, these hypotheses were not assessed.
      1.1.1. uggestions for more accurate conclusions, based on the data provided in the FFSD are provided in answer to peer-review Question 5 (above).
   1.2. Section 4 also relied on seemingly ad hoc methodologies that were not well justified or explained.
   1.3. Sections 3 through 5 all relied on results from Section 2 and need to be revised following revisions in that section.
2. If the analyses in Sections 2 to 5 were revised, then Section 6 would also, necessarily, be revised.

Page/line-specific comments

- line 191 - how representative are these sites of the Associated Habitat Reach?
- line 193 - how does including a site sampled in fewer years affect estimates of fish abundance versus river discharge?
- line 200 - if not all sites were sampled the same number of times and presumably not in the same years, how does this variation affect estimates of fish abundance?
    - Effects of year and day of the year and location should be included in the analysis of abundance versus river discharge to see if there are spatial or temporal influences. It should not be assumed there is no influence (such influences are common and have been documented within the Platte River).
- line 203 - a total was (not were)
- lines 234-237 - how can combining data sets from these separate study designs be justified? As presented, this seems certain to have confounded the analysis.
- line 239 - this choice is not justified, just because sample size was small does not justify combining different data sets, it must be shown that they are compatible prior to their being combined
- line 304 - why was a nonlinear relationship expected?
- line 314 - explain this caution more clearly
- lines 319-320 - explain this uncertainty more clearly

- lines 368-381 - A mesohabitat approach would be useful (although too late to implement for this study). Observing what mesohabitats least terns fish would allow fish sampling to target these areas (would be best if observations were seasonal). Then, such mesohabitats could be sampled for forage fishes to determine their abundance at relevant times and also to determine mesohabitat characteristics (i.e., depth and velocity).
- lines 371-372 - doesn't it seem that least terns search for areas where fish are present prior to plunging? In field observations I have made, it seems clear that least terns fly up and down and back and forth looking for vulnerable/suitable prey. They are able to cover large areas of river channel in almost no time. These discussions seem to suggest least terns have no ability to adjust for variable prey distributions or densities. Isn't there anything known about this? Couldn't this be explained and analyzed in a more realistic manner?
- line 421 - in more detail (not detailed)
- lines 449-450 - this seems to say that the opportunistic samples show a different pattern than the systematic ones (i.e., the two data sets should not be combined)
- lines 456-457 - ditto
- line 489 - there is a little evidence? What does this mean? Sounds speculative rather than evidence based.
- line 500 - awkward "...in two session..."
- line 518 - awkward "...the fact best model..."
- line 537 - what evidence justifies the choice of the 7-day window? The breeding season is reported to be 38 to 50 days. Why are all the other prior days irrelevant? It seems least terns (especially females) might also need forage even before the breeding season begins to build up energy reserves, develop eggs, establish a nest, etc.
- lines 554-555 - here it is admitted that the period over which forage-fish abundance is important is unknown; why don't the analyses include a wider range of possibilities?
- lines 564-565 - flow cannot be used as a surrogate for forage-fish abundance (at least not based on any evidence provided in this document); Section 3 (cited here) does not assess the relation of flow to fish abundance; Section 2 (where this is supposedly assessed) assesses the effect of discharge on captures in individual seine hauls, which is not a measure of overall abundance; also, the relevance of the dilution effect of increased river discharge on least tern foraging is unclear and Section 3 suggests higher flows actually increase foraging success. Is this finding incorporated into the least tern productivity model? If not why not integrate these lines of evidence?
- line 577 - the significance of the Kearney Canal Diversion site should be described in Section 1 and should be considered in all analyses of every section; similarly, any other unique study locations should be considered individually, not collectively
- lines 579-580 - what time of year? what types of fish? what range of depths?
- lines 640-641 - this suggests forage-fish abundance at the Kearney Canal Diversion alone should be assessed in earlier sections; if this site is a primary or critical foraging site, it alone may have a predictive relation with least tern breeding success; when and how many least terns were observed at other fish sampling sites? It seems more data are needed to determine where least terns forage and which least terns from which breeding locations forage in what river reaches.
- line 676 - "top down" approach (not approached)

- line 684 - awkward "...model to estimated average..."
- line 737 - awkward "...multiplied their estimated by the..."
- line 748 - are reported to weigh (not weight)
- line 772 - calculations are apparently based on the assumption that fish populations do not vary over time; it seems to be assumed that only daily discharge affects fish abundance; yet, no basis is provided for this assumption, this is inconsistent with expectations from Hypotheses T2 and T2a, and is extremely unlikely; in any case, Hypotheses T2 and T2a are not assessed in the FFSD, analyses in Section 2 do not assess variation in forage-fish abundance except in relation to discharge at the time of sampling, thus the finding of no effect of forage-fish abundance on least tern production is due to a lack of looking, not a lack of a relation; whether the relation actually exists remains unstudied
- lines 781-782 - given this study design, shouldn't forage-fish abundance at each sampling site be analyzed only with regard to least tern success at the nearest nesting habitat? Is it possible that individual nesting locations have differing success in a given year? Could differing success be associated with differing foraging conditions in nearby sites? By combining all data into a single set without prior analysis, it is possible important site-specific differences have been missed and are confounding the combined analyses.
- lines 849-852 - as discussed above, these conclusions do not accurately represent the findings of the analyses presented

- line 854 - No forage fish data have (not has)
- line 854 - data indicate (not indicates)

This page intentionally left blank

Reviewer 2 peer review comments

Review of PRRIP doc
25 November 2015

Note: Comments that I deemed most essential to address are marked with an asterisk. Numbers following an asterisk denote an association with one or more questions in the "Charge to the panel."

This document addresses whether flow might affect forage fish (for terns) and, hence, tern abundance. Addressing these questions, given the data, will typically be challenging.

*12: A major concern with the report is that it suffers from lack of clarity. Hence, I recommend that the authors seek the assistance of a technical editor. General suggestions include the addition of a list of abbreviations to the beginning of the document, and the movement of lists of numbers to tables (despite its length, the document contains few tables). Other clarity-related issues are addressed below.

*12346: Analytical efforts described in the report would almost certainly benefit from consultations with a statistician on the faculty of a university statistics department. I say "almost certainly" because the potential strengths and weaknesses of the study methods and resulting inferences are partly obscured by the report's lack of clarity. A statistician selected to help guide this report's analyses should have a PhD in statistics or related discipline (eg biostatistics), and have interests in both generalized linear models and making inferences from observational data.

*12346: The authors focus on an integration of processes over the study area and study period but often without attending to processes at finer spatial and temporal scales. In the process, the validity of inferences over the study area and study period may have been compromised. This concern may be at least partly addressed by incorporating sampling design components in models.

Section 1.

1) 63: "test" is used throughout the document. This word often connotes a binary yes/no decision regarding a hypothesis. However, such an approach seems contrary to the ambiguous and complex nature of the presented evidence (as the authors address in the preface; lines 25-33). Hence, I suggest the authors replace "test" with "evaluate" or similar throughout.
2) *1: Define "Big Question" at first mention. The preface mentions the ambiguous term "Big Question" but does not define it. The term is repeated but in plural on line 74. On lines 93, 94, we are provided with a priority hypothesis and sub-hypothesis—but without clear context. The "Big Question" is stated on lines 101-2. However, this statement is ambiguous because it is given in the context of "these hypotheses" (which are not clearly tied to _the_ big question) and because we readers now know that more than one big question exists. If

this 'big question' defines the purpose of this document, it should be defined at first mention. (note that this question is defined in the peer review scope of work.)

3) Data subsection.
   a) the transition from program uncertainties to data in this section is unclear. A reader will be forgiven for not realizing that s/he has made an effectively seamless transition into data/methods section.
   b) The ambiguity about data composition and study designs is perpetuated when, in subsequent sections, datasets are mentioned without reference to their description in Section 1. This disconnect will best be rectified by moving data and sampling design descriptions to the sections in which they are used. (Where data are used more than once, readers may be referred to the section in which those data are first described—with the caveat that aspects particularly relevant to an analysis should be mentioned in the relevant section.)
   c) Further blurring the line between the data (sub)section of Section 1 and the study sections is that Section 1 contains references to analyses and results (p14).
   d) If the authors prefer to continue to treat all data and sampling designs together and separately, then they should consider doing so in a section wholly devoted to data/sampling designs. If this latter approach is pursued, links between the new section and the study sections need to be clear.

4) 95-97: Explicitly define hypothesis and sub-hypothesis as "the hypotheses"—as they are defined on line 250.

5) *12: 297: T2a as written (and repeated on line 250) is very general ("influence" and "related"). However, T2a as depicted in Fig 2 assumes T2a is defined by ascending and descending components. This disagreement needs to be resolved. Section 2 assumes a "non-linear relationship" between FFA and flow in text (line 261) and a monotonic (loglinear) relationship for the modeling phase; both assumptions are congruent with T2a as written but only the former is congruent with that depicted in fig 2. The final paragraph of the report (lines 850+) appears congruent with Fig 2, in that it assumes T2a may be rejected on grounds that FFA decreased with flow as flows approached small values. Given its importance in this report, the T2a hypothesis needs to be stated with more clarity.

6) Fig 2b: "due"□"do"

7) *346: 194-204: provide sampling details (numbers of samples by date, site and year, along with whether sampling was precluded by low or high water) in tabular form; this information may later be used when evaluating whether trends in flow might have been confounded with trends in date of sampling, sample size, low or high water, or combinations thereof.

8) *346: 220: provide number of riverine sites.

9) *2346: 225: clarify whether "only river channel" includes the two sandbar colony sites and, if not, the number of river channel sessions. Comment on whether % successful plunges and % plunges with unknown capture success appeared related to sampling site/location, date (within year) or length of session.

10) *346: 234-240: provide sample sizes for the two designs. Address whether means and sampling variances might be allowed to vary by design.

11) *2346: What proportion of family units are thought to include one chick or two chicks (257)? If either proportion is nontrivial, defend assumption of uniformly three chicks in

family unit. Might some family units contain more chicks at the beginning than at the end of brood rearing?

12) Figure 3 is helpful

Section 2.

1) Overview
   a) This section could be substantially improved by attention to links with data and sampling design, by fuller explanation of methods, by better defense of the model (and probably elaboration of the number of models) and, absent additional evidence, a decrease in implied confidence in study conclusions.
2) Introduction
   a) *2: Summarize FFA-flow associations found by other investigators. Address why an instantaneous rather than lagged FFA-flow association is expected.
3) Methods
   a) *346: 264-7: describe data and sampling designs here, or reference where this info might be found. Section 2 opens with reference to data and sampling designs that are not tied to the relevant description in Section 1. for clearer reading and to better address concerns with the data, sampling designs and analytical methods, the descriptions of the data and associated designs in section 1 should be moved to section 2 or be cited from section 1.
   b) *1346: Attention to the sampling design addressed in section 1 reveals that fish counts were obtained from four sites from 7 years, and from an additional site for 6 years. Yet these design aspects are omitted from the proposed model. It seems reasonable to imagine that means and possibly dispersion parameters varied by site and year. hence, absent strong evidence to the contrary, authors they should allow these sources of heterogeneity. Failure to do so, given the presence of this heterogeneity, will incorrectly address precision and possibly lead to bias. If numbers of samples per site per year exceeded 1, then the authors should also evaluate the potential importance of site-year effects.
   c) Sites and years will need to be treated as fixed if models are fit under a frequentist paradigm and as either fixed or random if under a Bayesian paradigm. Site-year effects may be treated as random under either paradigm. Note that these potential sources of heterogeneity can’t realistically be dismissed based on p values (which, with rare exception, evaluate alternative rather than null hypotheses).
   d) investigators should evaluate whether variation of site-year effects changed with year, site and/or flow.
   e) If investigators permit random site-year effects, then they will need to decide whether they are interested in marginal or subject-specific effects of flow (and present results accordingly).
   f) 274: evaluate whether phi might be other than constant (eg change with site, year and/or flow).
   g) provide sampling details (numbers of samples by date, site and year), along with whether sampling was precluded by low or high water, in a table, and either here or in section 1; evaluate whether trends in flow might have been confounded with

trends in date of sampling, sample size, low or high water, or combinations thereof. See similar note for lines 194-204.

h) 267: “results”□”data”

i) *234: 270+: the authors begin with a FFA-flow assumption, and define it as loglinear. Allowance for evidence to the contrary appears to be confined to depictions of expectations (which, given the assumed model, cannot permit non-monotonicity; Fig 5), of raw data (Fig 6) and posterior predictive checks. And yet the authors appear interested in whether FFA and flow are related and, if so, whether the association is non-monotonic (sections 1, 6). Given this and the strength of this section’s stated conclusions (304+), why not fit a no-flow model and a model that permits a non-monotonic FFA-flow association? Fitting the latter model (eg with a quadratic flow term) would permit greater support (or otherwise) for the ascending limb statement in section 6 (851- 2). Relative support for the multiple models might be evaluated using an information criterion (as done in section 4). The report elsewhere describes an intercept-only model (ie a model analogous to a no-flow model; line 593).

j) *3: 277: beta1 as stated is either ill-defined or undefined. Line 91 equates no flow with dry channel. can fish persist when flow = 0? If not, beta1 = log(expected catch at a flow of 0 cms) = log(0) is undefined. Further, Figs 5 and 6 imply and line 499 state that the minimum flow is 5 cms. Hence, beta1 at flow = 0 represents an extrapolation. Consider centering or log- transforming flow.

k) *36: 278: the data section (lines 242-5) indicates flows were obtained for 1999-2000. If so, how are flows matched with the many years over which fish counts were obtained? Flows should correspond to the date of sampling (whether lagged or otherwise). Clarify.

l) *2346: 278: state whether differences between flows at fish sampling locations and nearest gaging stations are expected to be more than trivial and, if so, whether they are thought to represent classical or Berkson’s errors—or a combination thereof. If classical or a combination, then evaluate whether they should be treated as additive or multiplicative and also whether they might be heteroscedastic—and then attempt to address them.

m) 280: note that phi at 112.5 m2 and at 1 m2 scales may be very different.

n) 282+, Figs 5, 6: for clarity, authors should describe how expectations and predictions differ

o) *234: the authors must describe their method of fitting their forage fish model. From lines 282- 4, readers may infer that models were fitted under a Bayesian paradigm. If so, the readers need information about priors, algorithm, burn-in period, computer package, etc.

4) Results:

a) 286: “we found expected [FFA] … decreased….” This may work for a statistician. However, the use of the term “expected” in this context may convey to managers and the public that your model revealed truth—even while the finding is model and data dependent. Suggest rephrasing.

b) *234: 286: provide beta1 and beta2 estimates (with CIs)

c) *34: 288: provide phi with CI. Report whether phi appeared to vary with flow, year, flowxyear (and, if so, permit such variation).

d) 289: insert “at low flows” after “densities”

e) *23: 291. Fig 6 alone seems insufficient to support this statement. What, for example, if phi were permitted to decrease with flow? Greater support for “is adequate” may be supplied by adding the two models proposed above

f) Fig 5: define “CIs” (credible intervals).

5) Discussion

a) “broader inference” might better read “unbiased inference to the sampled population” or similar

b) 304: “we expected and found a nonlinear relationship.” See comments above re nonlinear and non-monotonicity.315: as implied by “a fish” and “individual” (315-16), insert “fish-to-fish and” before “sample”

c) *2346: 314-20: this seems close to a fatal concession. Do the authors have counts from multiple seines from each of multiple site-date or site-discharge conditions? If yes, then report (with results) whether qualitative inference on catchability might be obtained from these pseudoreplicates. If no, then say so.

6) Summary

a) *234: 322: the study didn’t “establish” a relationship (which, if present, would have preceded the modeling exercise). Replace with more neutral term (eg “provide evidence of”). Concede that the authors cannot preclude that the results might be explained by confounding with catchability.

b) 323-325: “opposite to that hypothesized”. See earlier comments on T2a.

c) *2: 325-6: this inference is not directly supported in section. Omit—or provide evidence in favor of statement in results and discussion.

Section 3

1) Introduction

a) *2: Provide what others have found or written on this topic.

b) *1: Define Districts, Big Question and “this premise.”

c) *2: 334: define “hypotheses”

2) Methods

a) *2346: The validity or at least utility of the fish abundance model needs to be better defended before relying on it to supply predictions for a follow-up study.

b) *2346: In the event that the fish-flow model becomes better defended, the authors will still need to address that fish CPUEs at a given plunge are likely to be estimated very imprecisely (Fig. 6). Measurement errors will need to be addresses.

c) *2346: 339: define a session—does it, for example, include observations from more than one location? Why doesn’t the index also address location and year?

d) *234: 341: address whether external or internal evidence suggests phi should be permitted to vary with flow, year or location.

e) *346: 344: “at the location” or at the gaging station nearest the location?

f) 349: “model two”□”model equation 4.2 (or 3.2)”

g) *2346: Mention whether plunges might result in catches exceeding one fish and, if so, how this was addressed (given the binomial assumption)

h) *2346: Address the potential for extra-binomial variation and whether any such variation depends on year, location and/or flow.
i) *2: 358-9: support your assumptions with references (either here or in introduction)
j) *2346: 358+: flow is a surrogate for multiple latent variables (depth, velocity and water clarity) and is the sole predictor of FFA. Address how one predictor can realistically and simultaneously perform these several roles.
k) *34: 382+: provide priors, algorithm, burn-in period, statistical modeling package, etc.; explain how prediction errors were addressed [including whether this is accomplished via an embedded FFA-flow model—as may be implied by 1)]; if an embedded FFA-model is not used, address how the latent FFA-flow relationship defined in eqn 4.4/3.4 was addressed (eg via structural equation models); describe how uncertainties in estimates of pi and lambda were incorporated in estimates of gamma (and, explicitly, how credible intervals on gamma were estimated).
l) 383, and in figures: “hierarchal”→”hierarchical”
m) *3: clarify use of “model” as singular (“our model”, lines 388, 393, 399) with statement that two models were fitted (line 336). The statement that implies more than two models were fitted (“best two models”, line 393) presumably relates to model selection. However, this also needs to be clarified. the use of “our model” on line 399 may relate to a model of gamma; if so, make this explicit.
n) Equation 4.6/3.6 and lines 415+ imply cross-validation was performed for only one model. Why not for others?
o) *3: 389-393: functional form does not appear to be addressed in the results
p) 390: reference to equation 3.2 begins a hint (continued at 401, 413) that equation numbering in section 2 and the first part of section 3 should’ve read 2.? and 3.?, respectively.
q) 393: multiple models of each outcome as yet undescribed
r) 401: mu→lambda

3) Results
   a) *234: Provide parameter estimates with CIs; include NB dispersion parameter.
   b) *234: 416, 421: “neither influenced” seems too strong. Perhaps the evidence supplied suggest, instead, merely greater relative support for the model with neither covariate (which may not imply null effects for both)
   c) 419: higher?
   d) *234: 423: this paragraph implies, absent caution on the part of the reader, the utility of model M2. But this assumption is in doubt—as the previous paragraph makes clear. Address this concern at the beginning of this paragraph.
   e) 454: I don’t think I’ve seen Fig S1

4) Discussion
   a) 486+: the authors appear to confuse model assumptions and model results with reality in this paragraph—but not later. A few tweaks could fix this concern. Specifically, does the performance of model 4 suggest that neither flow nor fish density affect the outcomes of interest? or, rather, that predictions are not clearly improved by the addition of either (given data and limited model set)? The phrase

“relatively small” (491) and the hunch of “small influence” in 505-7 and 515-520 both suggest a safer approach.

b) *12: Address whether the results presented in this section support the Big Question and/or the study hypotheses.

5) Summary

a) 517: credible

6) Overview: this section reads moderately well. The authors need to improve their explanations, and more tightly integrate the intro, methods, results and discussion sections. While it appears that the authors need to address multiple modeling/statistical concerns, this appearance may partially reflect lack of clarity.

Section 4.

1) Introduction.

a) The intro in this section is relatively well developed.

b) *1: Define “big question” and “this premise”.

c) Move the opening sentence (on data collection) to the methods section.

2) Methods

a) readers are unlikely to remember tern data collection and sampling design details from 22-23 pages earlier. Hence, the portions of the data collection and sampling design details from Section 1 that pertain to tern productivity need to be moved or reiterated here.

b) 548: for consistency, use pi (as in section 3) for the binomial parameter (eta is often used as the linear predictor of a GLM/GLMM).

c) *2346: 548: see comments for section 3 on addressing the sampling design. For this model, we might expect pi to vary by pair, pairxyear (presumably brood), habitat, year and maybe habitat×year. if so, then these sources of variability should be assumed and so incorporated into the model (with the exception of pair effects—if, as I suspect, that information is not available). Extra-binomial variation at the brood scale may be handled parametrically or nonparametrically and, as mentioned earlier, inference may be marginal or subject-specific. In this setting, failure to address sources of variability may lead to biased parameter estimates.

d) 550: how the number of broods (486) was determined needs to be explained here. On encountering this number, I returned (unsuccessfully) to section 1 to find its source. Turns out that the source of this number is provided subsequently, in the results section (lines 601+).

e) *2346: the authors need to defend the exclusion of the 12% of broods with unknown fate. In particular, to defend that the status ‘unknown fate’ represents a missing at random process and so may be treated as ignorable.

f) address that broods were monitored only in open channel habitat (as revealed on line 623)

g) *2346: The numbers of broods in all three categories (fledged, failed and unknown) should be provided by categories that relate to sampling and which have hitherto been addressed, namely year and habitat type. Line 644 suggests a site component to the brood monitoring design. If so, broods should be reported by site, year and habitat type.

h) *346: 550: the authors need to clearly state whether they are modeling a binomial or Bernoulli process. based on page 10, the authors can't estimate the probability of chick fledging but only the probability of nest success; line 549 implies a binary process (that "a" (singular) chick fledged from a given brood); line 661 states that the analyses in this section pertain to broods rather than to chicks in broods.

   1. If the authors were modeling a binary process, then they need to clearly say so. Further, they need to say why they weren't able to model the probability of chick fledge—which outcome would presumably have more power to detect effects than binary brood fledge status.
   2. If brood fledge status was modeled, then the generating model in 5.1 should be presented as that of a Bernoulli (ie a binomial with index 1).

i) 550-568: move all but decision to use flow to discussion

j) 564: evidence of "high correlation between flow and expected fish abundance" remains implicit but not explicit; address this in Section 2. Strengthening the study of the association between expected fish abundance and flow in section 1 would help the study described in Section 4.

k) *234: 562-564: the expectations of interest appear to represent those of an area with maximum horizontal dimension <=3 km. by contrast, the expectations from Section 2 correspond to flows integrated over the >100 km-long study area. if so, these two uses of expectations aren't commensurate. This mismatch might be corrected by using "small-scale" and "large-scale" or similar for predicted and expectation, respectively.

l) *346: 572: why a single value? Lines 541-3 imply that a minimum flow will be calculated for each brood (from the 7-day period prior to the fate of each brood being determined). If so, the authors will have multiple minimum flows. Clarify.

m) 573: "influenced"□"was associated with" or similar (as currently described, analyses in this report don't support causal inference)

n) 578: clarify whether "the authors" refers to Sherfy et al or to the authors of the FFA report.

o) *234: 584: defend use of a linear rather than nonlinear term for distance to Kearney Canal. For example, the probability of terns flying to Kearney Canal might decrease monotonically with distance—up to a threshold (beyond which terns might not fly)

p) *234: 594: the data were collected in multiple years. Hence, the authors should explain why they didn't treat year as a potential block.

q) *3: 595-6: explain why used frequentist methods for this analysis when used Bayesian methods for analyses described in previous two sections

r) 615-621: put deviance, AIC, AIC model weights and parameter estimates (with confidence intervals) for all models in a table.

s) *234: 636-654: these comments eliminate the AIC best model from consideration.

t) 636-654: this paragraph points to a potential criticism of the analyses in sections 2 through 4: The authors' focus on an integration of processes over the study area has been at the expense of attention to fine-scale heterogeneity. In the process, the validity of inferences over the study area may have been compromised. This concern may be at least partly addressed by addressing sampling design attributes in models.

u) *123: The discussion needs to attempt conclusions re hypotheses/models. The AIC best model has been excluded on scientific grounds. However, that model may have naively modeled distance from the Kearney Canal and, if so, a modification of that model may need to be considered. If not, the authors are left with the ambiguous situation of three models with AIC values that fall within only 3 AIC units. Perhaps an elaboration of the final sentence (669-71) represents an appropriate conclusion.

3) Summary

a) Much of this subsection represents discussion rather than summary

b) *: 657-659: this is new information and so should be moved.

c) *23: 667: the validity of the noted model is in doubt. Hence, the use of a causal statement appears inappropriate.

d) 669-701: this may be new material (similar statements in 644+ but for only a single site). if so, move.

4) Overview. section 4 reads as a draft; it requires attention to clarity (eg to ensuring material reflects subsection headings, and to ensuring context for statements), to modeling details, and to conclusions.

Section 5.

1) Intro.
   a) Define “priority hypothesis T2”
2) Methods
   a) *3: 684: describe this model
   b) *34: 689: SD and SE both mentioned; clarify
   c) *34: 717, 8: “we showed”. The FFA-flow model was merely evaluated, and then effectively only against itself. hence, the study described in Section 2 does not support the strength of this statement (altho elaboration of that study such as is recommended above might permit a weaker statement, such as, “We provided evidence in support of a loglinear FFA-flow association” or similar).
   d) *234: 722: _if_ the model reflected by model equation 6.3 is defensible, then these parameter estimates still require evaluation. We have not, to my knowledge, been provided with credible intervals for beta1^ and beta2^. Whether the beta2 estimator might be biased also needs to be evaluated.
   e) *2346: 722-765: The authors need to convince their readers that the approach selected, after adjusting for uncertainty, provides useful inferences (the comments re uncertainty propagation at lines 793+, given the importance of the study, are unconvincing). At a glance, this uncertainty would include that arising from imprecise estimation of beta1 and beta2 from the FFA-flow model, number chicks in brood—if that number ever differs from

3, error from assuming all fish were red shiners, average total weight for male and female red shiners of age class 0, wet wt:dry wt proportion, red shiner calorie content, daily energy expenditures for terns and tern chicks, tern weights, whether the form of the Nagy relationship is correct, uncertainty in the Nagy estimates, Roby metabolic energy requirements for tern chicks, adult-tern allometric assumption—perhaps more). Table 2 and Fig 12 should then incorporate the new estimates, with estimated uncertainty (eg approximations to 95% confidence intervals).

f) *4: 738: Clarify use of "lowest weight" when this value is provided as a mean with nonzero SE on line 737

3) *234: Results, discussion and summary. Revise after addressing uncertainty in assumptions.

Section 6.

1) *1: 814-41: this information provides essential context for the entire report, and so should be reproduced in Section 1
2) *1234: 850: a claim of "no evidence" is a strong claim, and one that I don't think the supplied evidence and inferences can support.
   a) A similar statement applies to the claim of "no evidence" at line 837. The authors would do well to follow the more ambiguous language provided in their quotes from the EIS and BO.
3) 851: for clarity, repeat T2a and reference Fig 2
4) *1234: 856-858: these conclusions are too strong.
   a) For T2:
   b) These conclusions seem to contradict summary statements in lines 515-7.
   c) Section 3 provides weak evidence in favor of a unimodal association between catch rate and flow.
   d) The results from section 4 were inconclusive.
   e) The results from section 5 need to be supplemented with measures of imprecision prior to interpretation.
5) The study supports T2a as supplied in text (line 97) but provides some evidence against T2a as depicted in Fig 2 (this latter evidence most obviously being the data depicted in Fig 6). The FFA- flow analyses could be elaborated to become more conclusive even while hypothesis T2a needs to be clarified.

From the scope of work:

1. Does the forage fish synthesis document adequately address the overall objective of the document, which is to present lines of evidence for broader examination of the relationship between forage fish abundance, river flow, and tern productivity on the central Platte River?
   a. Answer: If, by "lines of evidence," you mean "lines of evidence that may reasonably be defended in the peer-reviewed literature," then I'd say, "No." See review comments preceded by *1.

2. Do the authors of the forage fish synthesis document draw reasonable and scientifically sound conclusions from the information presented? If not, please identify those that are not and the specifics of each situation.

   a. The answer to the "reasonable" question is, in my opinion, typically unclear (due to lack of clarity in the document) while the answer to the "scientifically sound conclusions" question is, "Often not." See review comments preceded by * and 2.

3. Are the statistical methods used in each section appropriate?
   a. The answer to this question is not entirely clear (due to lack of clarity in the document). For at least one and probably two sections, I will say 'no' while, for other sections, my answer is 'maybe.' See review comments preceded by * and 3.

4. Are potential biases, errors, or uncertainties appropriately considered within the methods sections of this document and then discussed in the results and conclusion sections?
   a. Answer: given the phrase "appropriately considered," I'd say, "No." See review comments preceded by * and 4.
   b. Given that statistical methods address uncertainty and that uncertainty affects conclusions, '4's often co-occur with '3's and, to lesser extent, with '2's

5. Are the conclusions drawn supported by the synthesis of data reported in Sections 2 – 5?
   a. Answer: No, or at least not well. See review comments in sections 2 through 6 and which are also preceded by * and 2.

6. Did we utilize the data we had as effectively as possible to address the hypotheses "forage fish abundance below 800 cfs limits least tern productivity?" and "there is a direct relationship between flow and forage fish abundance and species diversity?" and answer the Big Question "Does forage availability limit tern productivity on the central Platte River"? If not, what additional analyses do you recommend?
   a. Answer to data question: No. See review comments preceded by * and 6. Use of data is a statistical issue, and also affects conclusions. Hence, '6's tend to co-occur with '3's and '2's.
   b. Answer to additional analyses question: See review comments preceded by * and 3.

RATING:
Please score each aspect of this manuscript using the following rating system: 1=excellent, 2=very good, 3=good, 4=fair, 5=poor.

Scientific soundness 4

Degree to which the monitoring approach addresses the project's objectives 3

Organization and clarity 5

Conciseness 3

Importance to objectives of the Program 4

RECOMMENDATION (check one)
Accept

Accept after revision X

Unacceptable

Reviewer 3 peer review comments

Platte River Recovery Implementation Program Forage Fish Synthesis Document Peer Review

Charge to Panel

1. Does document adequately address overall objective of the document – to present lines of evidence for broader examination of the relationship between forage fish abundance, river flow, and tern productivity on central Platte River.

Yes, the document addresses the overall objective thoroughly in its sections. In the Preface, it is stated clearly that the information in this report focuses "… on informing the use of Program water and fiscal resources to achieve one of the Program's management objectives: increasing production of interior least tern …." Later (lines 42-46), it's clarified that analyses of forage fish, foraging ecology, tern productivity, and flow data were used to test three inferred relationships in hypotheses, as well as an examination of a bioenergetics model. Within Section 1, the authors clearly identify "The Big Question" – does forage availability limit tern productivity on the Platte River (lines 101-102).

After providing information on how data were collected, the authors clearly state how the data were synthesized to test three inferred relationships in the hypotheses (lines 250-254). Further, they explain how the data were used with information in the literature, to develop a bioenergetics model to predict numbers of Least Tern family units that could be supported in the AHR at various flows.

Section 2 and its first paragraph introduce the reader to priority hypothesis focusing on the relationship between Least Tern productivity and availability of forage fish. The first focus of this section is an assessment of the relationship between forage fish abundance and flow rates.

Section 3 and its first paragraph, introduce the reader to the objective of this portion of the evaluation and what the authors hope to determine. They state their hypotheses (expectations) – flow and forage availability influence foraging behavior and success of Least Terns.

Section 4 provided a clear objective in its first paragraph – determine impacts of flow on Least Tern productivity within the AHR, as well as the authors' expectations or hypotheses.

Section 5 developed a bioenergetics approach to estimating forage fish needs of Least Terns. They objective was clearly stated in the last sentence of the introductory paragraph; however, no hypotheses or expectations were given.

Overall, the authors did a very good job of clearly stating the Big Question and each objective they addressed to provide answers to the question. Further, for all but one objective (bioenergetics), the authors' hypotheses were provided and later, discussed in the latter parts of each section. The final section of the documents does a good job of making conclusions about *associations among flow rates, fish abundance, and Least Tern productivity. The data and their analyses indicate that river flows ≤800 ft3s-1 do not affect Least Tern productivity negatively.*

2. Do authors draw reasonable and scientifically sound conclusions from the information presented?

Yes, for the most part, the authors use their data and the analyses to provide summaries and conclusions that are reasonable and scientifically sound. Where I had questions or recommendations, I provided comments and related them to the line(s) in the text.

For example, in Section 1 (lines 234-240), the authors should clarify better how they accounted for the positive bias associated with observations taken opportunistically on detected flocks of Least Terns relative to data from systematic, unbiased observations from pre-selected locations.

Also in Section 1 (lines 242-245), often, increased flow rates increase turbidity, which reduces terns' ability to detect and obtain forage fish. How did the authors account for turbidity and its influence on terns' foraging success? Did they account for such variables as they came to their conclusions about terns' ability to detect and catch fish?

In Section 2 (lines 322-326), it appears that assessment of data detected an impact of flow rate on fish catchability; however, an estimate of this impact was not determined definitively. Given this uncertainty, it cannot be stated that catchability changed with flow rate; it can only be stated that the analysis of the data was inconclusive. This situation should trigger a more intensive study of this question.

In Section 3 (lines 465-470), in their discussion, the authors conclude that prey availability (or fish density) did not affect tern plunge rate, probability of fish capture, or fish capture rate – perhaps the estimate of fish density was too inexact (estimate relative to tern location while feeding [fish samples were not taken at exact same location or time as foraging terns]); or Least Tern foraging behavior is independent of fish densities (as recorded; range of densities); or fish capture rate by terns was independent of flow rate, as recorded in this portion of the study. This conclusion and assessment should be revisited.

Further (lines 484-485), the authors' analyses determined the best, most parsimonious model and found that neither flow rate or fish density significantly affected tern plunge-rate or fish-capture-probability. Then, on line 497, it's stated, it seems reasonable that flow could influence Least Tern foraging behavior. Because the best model found no significant impact of flow rate on tern plunge-rate or fish-capture-probability, there can be no further statement about flow rate affecting tern foraging. If it's thought the model is incorrect, perhaps better or more data are needed for further analyses. The results thus far, however, do not detect a significant influence of flow rate on tern foraging.

Later (lines 505-509), the authors continue to try to say that flow rate has an impact on tern foraging behavior, in contrast to analyses of data and model results.

Finally (lines 511-520), in the summary, it's restated that analyses and modeling found no effect of flow rate on tern foraging behavior; however, as written, the authors are still trying to find an effect.

Although the results may be disappointing to the authors, and perhaps intuitively or biologically, it seems that flow rate should affect tern foraging behavior and success, if the statistical models do not find an effect, given the data they had to analyze, then it cannot be stated that there was an effect. It can be stated that the data were scant or different, additional data (variables) need to be collected, or the model needs to be corrected; but, it cannot be stated that there was an effect.

In Section 4 (lines 647-648), the authors should address other factors that affect tern productivity such as flooding of nests and loss of chicks and nearly-fledged young to flooding; peoples' activities in and near colonies (see Ducey 1985); and predation. Those factors are affected by flow rates, channel depths, and elevation of sand bars and other nesting sites. Addressing this information would strengthen and improve the conclusions. In the summary and conclusions (line 665), the authors should also consider that at low flows, predators have greater access to sand bar and other tern nesting habitat, and may affect productivity. At low flows, losses to flooding are low or absent; thus, this factor is likely not a threat to productivity.

Relative to Section 6, the study, analyses, and conclusions are good. They support an assessment that Least Tern productivity in the study area, is not adversely affected by river flow rates <800 ft3/sec. The authors should comment on other possible impacts of low flow, including greater access to sand bars by predators and people. Further, greater river flows during the nonbreeding seasons often benefit terns and plovers by removing vegetation on sandy areas, and creating new accretions of sand bars and islands for better nesting habitat.

3. Are statistical methods used in each section appropriate?

A statistician will have a better assessment of statistical methods used. However, the analyses and models used seem appropriate and current.

Specifically, see lines 607-618 and Fig. 11 – Please clarify – is equation number wrong in the text (Line 615)? The figure caption says Eq. 5.5 was best model; text says Eq. 4.6 was best. I could not find Eq. 4.6 in the text.

4. Are potential biases, errors, or uncertainties appropriately considered with the methods sections of the document, then discussed in results and conclusions sections?

Overall, the methods are strong and well documented. There were specific portions of the methods that may need further explanation, however.

In Section 1 (lines 176-178), the authors should provide more explanation regarding how field biologists were able to record accurate data on each Least Tern (LETE) pair and their brood. As colonial birds, semi-precocial chicks leave the nest ~2 days after hatching, and often, chicks from more than one pair form creches. Thereafter, unless each adult and each chick is marked with clear identifying bands or other marks, it cannot be determined if each LETE pair's chicks have *survived to fledge. Often, biologists use a less accurate estimate; they determine number of pairs, then record number of fledged chicks in the colony. Productivity is estimated via the relationship between the number of fledged young and number of adult pairs in the colony. This section is written as if each pair and its chicks were followed specifically and accurately. If this was so, then further description is needed to explain how the accurate data were obtained.*

Further (lines 191-196), the sampling protocol used to obtain estimates of fish abundance and diversity seems to include a lot of variability (e.g., sites from which sampling occurred; time of each sampling event; number of days of each sampling event). Was this variability accounted for in the treatment of the data or use of co-variables?

In Section 3 (lines 360-362), the authors should clarify what is meant by low flow being associated with shallow depths and greater flows being associated with greater depths. This discussion seems to assume that shallow water results in low abundance of fish. Perhaps a figure, a cross section of the river in the project area, would help explain the situation that is being described. Perhaps there is low slope rather than shallow edges and pools, with a deep channel. Overall, this discussion could use more clarification so readers not intimately familiar with the study area can picture what is being described.

In Section 5 (lines 746-756), the authors complete a very good study of bioenergetics. They should consider and discuss differences between the Caspian Terns used for estimates of metabolic rates, and Least Terns. Differences to consider include body size, body surface area, salt water vs. fresh water environments, and number of eggs laid / chicks produced by each (typically one for CATE, and 1-3 for LETE). Each of these factors may affect the use of CATE as examples for LETE, and the authors should address this concern.

5. Are conclusions drawn supported by the synthesis of the data?

Yes, the overall conclusions from this work, and reflecting on others' work (e.g., Sherfy et al. 2012), are supported by the synthesis of the data. The study, analyses, and conclusions are good. They support an assessment that Least Tern productivity in the study area, is not adversely affected by river flow rates <800 ft3/sec. The authors should comment on other possible impacts of low flow, including greater access to sand bars by predators and people. Further, greater river flows during the nonbreeding seasons often benefit terns and plovers by removing vegetation on sandy areas, and creating new accretions of sand bars and islands for better nesting habitat.

6. Were data used as effectively as possible to address the hypotheses?

Data were used effectively to address hypotheses. There were a few areas within the document where further clarification would remove questions about the data.

In Section 1 (lines 234-240), the authors should clarify better how they accounted for the positive bias associated with observations taken opportunistically on detected flocks of Least Terns relative to data from systematic, unbiased observations from pre-selected locations.

Also in Section 1 (lines 242-245), often, increased flow rates increase turbidity, which reduces terns' ability to detect and obtain forage fish. How did the authors account for turbidity and its influence on terns' foraging success? Did they account for such variables as they came to their conclusions about terns' ability to detect and catch fish?

Peer Review Rating & Recommendation

Category	Rating
Scientific soundness	2
Degree to which the monitoring approach addresses the project's objectives	1
Organization and clarity	2
Conciseness	1
Important to objectives of the program	1

Recommendation	Check One
Accept	
Accept with revisions1	XX
Unacceptable	

1Revisions or responses to comments and questions in General Comments section

The overall rating of the report is Accept with Revisions, and if the authors address the concerns and questions highlighted in numbers 1, 2, and 4 above, the document will be Acceptable. The most concerning portion the authors should address is the assessment of flow rate and fish density impacts on tern foraging behavior and success (lines 465-470, 484-485, 505-509, 511-520). If this section is re-examined and the uncertainties addressed, the document will be strengthened significantly, as there was a lot of uncertainty in these conclusions.

Report Review Guidelines

The major strength of the document is its thorough assessment of questions regarding Least Tern productivity and the riverine environment (flow rates; forage fish availability and abundance). Strengths include the many approaches used to address the problem statement or question. Because of these multiple approaches, the conclusions are more reliable and convincing. A weakness lies in its not openly addressing factors that were not included and those that should be examined further. These weaknesses and opportunities for further comment are noted specifically in the General Comments section. The conclusions drawn are sound, but may be enhanced further by considering specific comments provided in the General Comments section.

General Comments

1. Scientific Soundness (make sure any scientific and technical uncertainties are clearly identified and characterized, and implications of any uncertainties for the technical conclusions drawn are clear)
Notes on scientific and technical uncertainties –

Lines 176-178. The authors should provide more explanation regarding how field biologists were able to record accurate data on each Least Tern (LETE) pair and their brood. As colonial birds, semi-precocial chicks leave the nest ~2 days after hatching, and often, chicks from more than one pair form creches. Thereafter, unless each adult and each chick is marked with clear identifying bands or other marks, it cannot be determined if each LETE pair's chicks have survived to fledge. Often, biologists use a less accurate estimate; they determine number of pairs, then record number of fledged chicks in the colony. Productivity is estimated via the relationship between the number of fledged young and number of adult pairs in the colony. This section is written as if each pair and its chicks were followed specifically and accurately. If this was so, then further description is needed to explain how the accurate data were obtained.

Lines 191-196. The sampling protocol used to obtain estimates of fish abundance and diversity seems to include a lot of variability (e.g., sites from which sampling occurred; time of each sampling event; number of days of each sampling event). Was this variability accounted for in the treatment of the data or use of co-variables?

Be sure that consistent and correct terminology is used throughout the document. For example, fish may be abundant but not available. It appears the terms abundant and available are used interchangeably when they should not be, due to quite different meanings.

Lines 234-240. Provide better clarification of authors' accounting for the positive bias associated with observations taken opportunistically on detected flocks of LETE, relative to data from systematic, unbiased observations from pre-selected locations.

Lines 242-245. Often increased flow rates increase turbidity, which reduces terns' ability to detect and obtain forage fish. How did the authors account for turbidity and its influence on terns' foraging success?

Lines 258-267. The authors should recognize that availability and abundance of food to LETE is only one factor affecting productivity. Other factors include predation risk, frequency and type of human activities in or near colonies, over-wash of nests and chicks due to storm events (including hail and high winds), high temperatures, and contaminants. The authors should mention if any of these other factors were considered and if they were insignificant or if they were accounted for in some manner.

Lines 286-291. How did the authors account for the impact of high flow rates on sampling capability and effort? Were fish samples obtained at each flow rate comparable?

These questions arise within these lines, but from line 307-309 and 314-320, the authors address this concern. It would be helpful if the authors recognize the likely impact of flow rate on sampling bias immediately rather than many lines later in the paragraph. The authors should also clarify that flow rates did/didn't affect LETEs' ability to detect and obtain forage fish.

Lines 322-326. It appears that assessment of data detected an impact of flow rate on fish catchability; however, an estimate of this impact was not determined definitively. Given this *uncertainty, it cannot be stated that catchability changed with flow rate; it can only be stated that the analysis of the data was inconclusive. This situation should trigger a more intensive study of this question.*

Lines 360-362. In this section, the authors should clarify what is meant by low flow being associated with shallow depths and greater flows being associated with greater depths. This discussion seems to assume that shallow water results in low abundance of fish. Perhaps a figure, a cross section of the river in the project area, would help explain the situation that is being described. Perhaps there is low slope rather than shallow edges and pools, with a deep channel. Overall, this discussion could use more clarification so readers not intimately familiar with the study area can picture what is being described.

In this same section, the authors should state what factors influence flow rate most often. Is it influenced predominately by operation of the dam? Or is rainfall the greatest influential factor on flow rate?

Lines 363-364. A cross-section diagram / figure showing sampling protocol would be helpful, especially in determining if sampling occurred at differing depths because the hypothesis incorporates depth as a potential factor influencing fish abundance).

It appears that models assume that greater flow rates result in turbidity (line 364); were any Secchi disk or other measurements taken to quantify turbidity / clarity? It's likely that this factor affected terns' ability to detect fish; thus, affected fish availability.

Also at line 364, were flow rates measured at differing depths within each 200-m reach of the channel?

Lines 465-470. In the discussion, concluding that prey availability (or fish density) did not affect tern plunge rate, probability of fish capture, or fish capture rate – perhaps the estimate of fish density was too inexact (estimate relative to tern location while feeding [fish samples were not taken at exact same location or time as foraging terns]); or Least Tern foraging behavior is independent of fish densities (as recorded; range of densities); or fish capture rate by terns was independent of flow rate, as recorded in this portion of the study.

Lines 484-485. Authors' analyses determined the best, most parsimonious model and found that neither flow rate or fish density significantly affected tern plunge-rate or fish-capture-probability. Then, on line 497, it's stated, it seems reasonable that flow could influence Least Tern foraging behavior. Because the best model found no significant impact of flow rate on tern plunge-rate or fish-capture-probability, there can be no further statement about flow rate affecting tern foraging. If it's thought the model is incorrect, perhaps better or more data are needed for further analyses. The results thus far, however, do not detect a significant influence of flow rate on tern foraging.

Lines 505-509. These lines continue to try to say that flow rate has an impact on tern foraging behavior, in contrast to analyses of data and model results.

Lines 511-520. In this summary, it's restated that analyses and modeling found no effect of flow rate on tern foraging behavior; however, as written, the authors are still trying to find an effect.

Although the results may be disappointing to the authors, and perhaps intuitively or biologically, it seems that flow rate should affect tern foraging behavior and success, if the statistical models do not find an effect, given the data they had to analyze, then it cannot be stated that there was an effect. It can be stated that the data were scant or different, additional data (variables) need to be collected, or the model needs to be corrected; but, it cannot be stated that there was an effect.

Lines 557-568. Use caution in using flow rate in study area to predict productivity when it is unknown where terns from the colony are going to feed (assuming that all terns from a colony are feeding only in the study area?).

Line 566. The model assumption that probability of fledging is dependent on flow (as an indicator of fish abundance) should take into account, fish abundance ≠ fish availability. This model should also take into account, flow rate is not associated with tern fish-capture-probability or plunge-rate (Lines 484-485); hence, provisioning of fish to chicks and fledging would not be affected by flow rate.

Lines 607-618 & Fig. 11. Please clarify – is equation number wrong in the text (Line 615)? The figure caption says Eq. 5.5 was best model; text says Eq. 4.6 was best. I could not find Eq. 4.6 in the text.

Lines 636-642. Authors provide good explanation of prior information.

Lines 647-648. The authors should address other factors that affect tern productivity such as flooding of nests and loss of chicks and nearly-fledged young to flooding; peoples' activities in and near colonies (see Ducey 1985); and predation. Those factors are affected by flow rates, channel depths, and elevation of sand bars and other nesting sites.

Line 652. It is inaccurate to speculate about causes of mortality or nest loss if exact cause is not known. Loss could be due to starvation, but also to flooding, predation, or human activities. Unless data were collected that rule out other causes of mortality, it would be unwise to state mortality was due to starvation only.

Lines 665- Summary and Conclusions. Good discussion of effects of low flow and tern productivity. Also consider that at low flows, predators have greater access to sand bar and other tern nesting habitat, and may affect productivity. At low flows, losses to flooding are low or absent; thus, this factor is likely not a threat to productivity.

Lines 746-756. Good study of bioenergetics. Authors should consider and discuss differences between the Caspian Terns used for estimates of metabolic rates, and Least Terns. Differences to consider include body size, body surface area, salt water vs. fresh water environments, and number of eggs laid / chicks produced by each (typically one for CATE, and 1-3 for LETE). Each of these factors may affect the use of CATE as examples for LETE, and the authors should address this concern.

Within this section, there are many grammar, punctuation, and writing errors that should be corrected.

Section 6. Overall, the study, analyses, and conclusions are good. They support an assessment that Least Tern productivity in the study area, is not adversely affected by river flow rates <800 ft3/sec. The authors should comment on other possible impacts of low flow, including greater access to sand bars by predators and people. Further, greater river flows during the nonbreeding seasons often benefit terns and plovers by removing vegetation on sandy areas, and creating new accretions of sand bars and islands for better nesting habitat.

2. Organization and Clarity

The organization of the document is very good. There are some sections where the clarity could be improved upon, and those areas are mentioned above.

3. Conciseness

The writing is good and concise, except where noted. Section 5 had errors that should be corrected.

4. Degree to which conclusions are supported by data

Most conclusions are appropriately supported by data and analyses. However, I noted above specific statements that should be examined again. (see also number 5, above)

5. Cohesiveness of conclusions

Most conclusions are good and cohesive. Any specific comments on conclusions are given above.

Comment on any of the following matters that significantly affected your opinion of the manuscript:

1. Presentation (tightly reasoned argument throughout? Does ms wander from central purpose?)

The document brings the reader back to the main point and big question; thus, the document focuses on the main argument throughout and does not lose its purpose.

2. Methods (are they appropriate? Current? Described clearly and with sufficient detail so someone else could repeat the work?)

The methods are appropriate and current. Specific comments relative to methods are given above.

3. Data Presentation (when results are stated in the text of the manuscript, can you easily verify them by examining tables and figures? Are any of the results counter-intuitive? Are all tables and figures labelled clearly? Well planned? Too complex? Necessary?)

Overall, data presentation is good, and it is clear. There were a few areas that needed clarification (discussed above in my line by line review). It may be that my comments are off base; however, the fact that I was confused indicates that the lines or section need to be re-examined.

4. Statistical Design and Analyses (are they appropriate and correct? Can the reader readily discern which measurements or observations are independent of which other measurements or observations? Are replicates correctly identified? Are significance statements justified?)

The statistical design and analyses seem appropriate and correct. Specific comments are above in my line by line assessment.

5. Conclusions (are conclusions drawn from insufficient evidence? Are the interpretations of the data logical, reasonable, and based on the application of relevant and generally accepted scientific principles? Have alternative hypotheses been overlooked?)

There are some areas of the document that could use further discussion and inclusion of additional ideas. In the line by line review, I pointed out these possibilities.

6. Errors (point out any errors in technique, fact, calculation, interpretation, or style)

In the above line by line assessment, I pointed out any errors I discovered. Overall, I found few errors or questionable information.

7. Citations (cross check between text and list; adequately used?)

Citations were done well.

This page intentionally left blank

APPENDIX B: REVIEWER BIOGRAPHICAL SKETCHES

Brian Gray, USGS Upper Mississippi Ecological Sciences Center

Proposed Peer Review Panel Member for Platte River Recovery Implementation Program

Name	Brian Gray
Title	Statistician
Affiliation	US Geological Survey
Address	2630 Fanta Reed Rd, La Crosse, WI 54603
Phone #	608-781-6234
E-mail	brgray@usgs.gov

Education	BSc (Botany), 1980; MS (Biology), 1993; PhD (Biostatistics), 2001

Unique Qualifications

- Research and consulting statistician for the USACEs' Long Term Resource Monitoring Program (LTRMP) on the Upper Mississippi and Illinois Rivers
- Currently overseeing efforts to estimate temporal trends in ~200 fish, vegetation and water quality variables for the LTRMP
- A statistician who is familiar with avian, and river and stream studies: contributing statistician to 12 peer-reviewed avian publications, and author or coauthor of a further 20 (of 47) peer-reviewed publications that address river or stream issues
- Specialties: analysis of clustered and of observational (nonexperimental) data

Short Biography of Proposed Peer Review Panelist

Brian worked in part as a de facto statistician in a sediment toxicology laboratory in the mid-1990s. That work led to enrollment in a PhD in biostatistics program--during which time he acted as consulting statistician for a marine field station and a hospital. He currently works as a research and consulting statistician at the USGS' Upper Midwest Environmental Science Center, a position he has held since 2001. The USGS position includes a half-time appointment as the statistician for a USACE river monitoring program, a program that oversees the collection and analysis of ecological and environmental data from the Upper Mississippi and Illinois Rivers. Brian's background in the biological and environmental sciences allows him to better understand scientific aspects of studies on which he is asked to contribute statistical expertise.

Brian's work for the USGS has largely focused on methods of making inferences from non-experimental data, such as are typically collected by ecological monitoring programs. Such data may be unbalanced, derive from complex sampling designs, and be collected in ways that risk attributing causation to non-causal variables. He has especially focused on methods of analyzing data that are correlated within groups, such as typically occurs when multiple measurements are obtained from each of multiple years, lakes or other groups. His experience includes the evaluation of hypotheses, the reasonableness of research designs (given objectives), data collection procedures, sampling methods, and methods to analyze data and report statistics. Brian has worked and published with data from invertebrate, plant, fish, bird, water and sediment studies; the context for most of his work has been that of rivers, streams and lakes.

Christopher Hoagstrom, Weber State University

Proposed Peer Review Panel Member for Platte River Recovery Implementation Program

Name	Christopher Hoagstrom
Title	Associate Professor
Affiliation	Weber State University, Department of Zoology
Address	1415 Edvalson Street, Dept. 2505
Phone #	801-626-7486
E-mail	christopherhoagstrom@weber.edu
Education	PhD

Unique Qualifications

Extensive experience dealing with conservation ecology of minnows in rivers on the Great Plains. Detailed understanding of hydrology and reservoir operations in relation to river ecology on the Great Plains.

Short Biography of Proposed Peer Review Panelist

I received a BS in Biology from Ohio Northern University in 1992 and an MS in 1994 from Sul Ross State University. After that I was employed by the U.S. Fish & Wildlife Service from 1993 through 2002 within the New Mexico Fishery Resources Office. While there I worked my way up from a biological technician to a fishery biologist. During this entire time I was intimately involved in and often helped develop and lead detailed field studies. I was also responsible for reporting and presenting research findings. My main responsibilities during that time were field research on the Pecos River and Rio Grande aimed at documenting the status of declining species of minnows and providing recommendations for their conservation.

In 2003 I returned to school at South Dakota State University to pursue a PhD. My dissertation research had to do with relations between fish assemblages and environmental conditions within the Missouri River drainage. Part of this research included detailed studies of fish distributions in the Cheyenne River drainage.

In 2006 I accepted a position as an assistant professor in the Department of Zoology at Weber State University. I have remained there and have been promoted to associate professor and am presently department chair. While here I have continued active research on Great Plains rivers including field work in New Mexico (Pecos River) and continuing work on long-term data bases.

Sara Schweitzer, North Carolina Wildlife Resources Commission

Proposed Peer Review Panel Member for Platte River Recovery Implementation Program

Name	Sara Schweitzer
Title	Waterbird Investigations & Management Project Leader, Wildlife Diversity Biologist
Affiliation	North Carolina Wildlife Resources Commission
Address	106 Ferret Run Ln., New Bern, NC 28562
Phone #	252-639-8435
E-mail	sara.schweitzer@ncwildlife.org

Education	Ph.D., Oklahoma State University, Wildlife & Fisheries Ecology; M.S., Texas Tech University, Wildlife Science; B.S., University of North Carolina, Biology/Chemistry

Unique Qualifications

One part of Schweitzer's dissertation pertained to foraging ecology of Least Terns in the Central Flyway (Salt Plains NWR), and her work has included additional studies of Least Terns in Georgia and North Carolina. Specifically, her work includes surveys of abundance and distribution, monitoring for nesting success, and research relative to site selection, threat factors to nesting colonies and means of alleviating them, as well as impacts of toxicants in the environment to eggs, chicks, and adults.

Short Biography of Proposed Peer Review Panelist

Sara Schweitzer received her Ph.D. at Oklahoma State University through the USGS Cooperative Fish and Wildlife Research Unit, Department of Zoology. Her dissertation was titled, "Abundance and conservation of endangered Least Terns nesting on salt flat habitat." After completing her doctoral program at OSU, she joined the faculty in the Warnell School of Forestry and Natural Resources, University of Georgia. In the Warnell School, she taught graduate and undergraduate classes in wildlife ecology and wetland habitats, she established a research lab with her research associates and graduate students, and she became Director of the Bulgaria Study Abroad program upon returning to UGA from a Senior Fulbright Scholarship program in Sofia, Bulgaria. At UGA, Schweitzer served on numerous committees at the university and school level, contributed greatly to professional societies, and mentored 100s of students, both graduates and undergraduates. In 2010, Schweitzer accepted the opportunity to lead the Waterbirds Investigations and Management Project for NCWRC's Wildlife Diversity Program. In this position, she coordinates waterbird surveys, monitoring, and research by members of the N.C. Waterbird Management Committee (federal and state agencies, and NGOs), and conducts independent research on coastal waterbird species and their habitats. She continues to mentor students as an adjunct Professor and Graduate Faculty member for the Warnell School, serve professional societies, contribute to the scientific literature, conduct reviews for journals and grant organizations, and lead working groups and committees for the Association of Fish and Wildlife Agencies.