Sustainable Ecosystems Institute

Independent Science Review of the Habitat Assessment and Monitoring Program Study Plan for the Missouri River

Sustainable Ecosystems Institute
PO Box 80605
Portland OR 97280
Website http://sei.org
Tel 503 246 5008
Email: sei@sei.org

December 2005

Sustainable Ecosystems Institute

Independent Science Review of the Habitat Assessment and Monitoring Program Study Plan for the Missouri River

Science Review Panel

Dr. Nancy Auer
Dr. Henrietta Jager
Dr. Jim Quinn
Dr. David Secor

Project Lead: Lisa Sztukowski
Scientific Facilitator: Dr. Steven Courtney

Sustainable Ecosystems Institute is a non-partisan organization of scientists dedicated to using their technical expertise to solve ecological problems. Headquartered in Portland, Oregon, the Institute works nationally and internationally. SEI specializes in independent scientific review. Visit http:// sei.org and http:// sei.org/peerrev.html for more details. Contact SEI at sei@sei.org

Cover photo courtesy of USGS (http://infolink.cr.usgs.gov/Photos/index.html)

Independent Science Review of the Habitat Assessment and Monitoring Program Study Plan for the Missouri River

TABLE OF CONTENTS

Introduction

Program information

The Habitat Assessment and Monitoring Program (HAMP) constitutes a major, comprehensive effort to document and assess ongoing programs of habitat creation on the Missouri River. This cooperative effort (involving federal, state and other cooperators) supports the overall program for the restoration and management of the river. Monitoring is a key component of science-based adaptive management - HAMP will be a crucial element in the overall monitoring strategy.

Other elements of monitoring include fish population and community status (initial review completed 2004) and their relationship to hydrological measurements. This review focuses solely on the HAMP habitat monitoring element.

Independent Science Review

Independent Science Review (ISR) is an important mechanism to ensure scientific quality control, to determine whether all relevant information and opinion has been included, and to provide assurance to decision-makers that proposed management activities are well-founded in a scientific consensus. Activities by the US Army Corps of Engineers (USACE) on the Missouri River are complex, expensive, and likely to have long-reaching effects. It is therefore highly appropriate that ISR be used by the parties to assess their programs *in advance* of major commitments of resources.

Previously, Sustainable Ecosystems Institute (SEI) carried out a major ISR of the proposed monitoring strategy for Pallid Sturgeon on the Missouri River (SEI 2004). Key findings of that review were:

- The monitoring and assessment program is well-conceived, and constitutes a credible start on a long-term monitoring program.
- Element by element, the components of that program are sound.

- The components could be better integrated, and tied to information needs of decision-makers.
- Power analysis should be considered in design of the program.
- Statistical analyses are likely to be complex, and a full statistical support element is essential.

Detailed assessments and recommendations are given in more detail by SEI (2004).

In response to this ISR, USACE assessed the comments of the review panel, and provided responses to each (C. Fleming pers.comm., M. Wildhaber pers.comm.). Several important changes were made, including the inclusion of statistical support, and the non-random selection of bends in response to concerns regarding power and replication (SEI 2004b). Hence USACE and the other co-operators appear to have used ISR in an appropriate fashion, and to have integrated critiques into concrete program changes. However it is not clear that all recommendations have yet to be fully addressed. In particular, although Mark Wildhaber has been included as a design statistician into the Pallid Sturgeon and other monitoring elements, there is as yet no strategy to obtain further programmatic help (C. Fleming pers.comm.).

In carrying out the current review of the HAMP element, and following directions from USACE, SEI elected to carry out three linked reviews. The first of these (this document) addresses design issues concerning habitat monitoring. The second and third reviews will address hydrological models, and biological data linked to these models. SEI has retained four independent reviewers to address habitat monitoring design. These scientists are experts in data analysis, fish (particularly sturgeon) ecology, habitat assessment, and adaptive management. Three of the four reviewers previously assessed the Pallid Sturgeon program, thus ensuring consistency between reviews, and a degree of familiarity with the materials. In this first review phase, hydrologists were not included. However these are scheduled to participate in the subsequent assessments.

USACE HAMP materials were sent to each reviewer for independent assessment (see attached reviews, below). The team as a whole then discussed their respective reviews, and developed an overall consensus which forms the basis for this report. The USACE materials and reviewer comments were also sent to Dr. Zach Peery, who had acted as a statistical consultant on the previous Pallid Sturgeon review. His comments and recommendations are also included here. In general, the views of the four reviewers, of Dr. Peery, and of SEI staff are all congruent; hence the parties may have some confidence that the views expressed here represent a good consensus among scientists experienced with reviewing similar complex data acquisition and analysis issues. We provide here the panel overview, together with individual panelists' comments and recommendations.

Scope of Review

The scope of the review was to provide both a detailed analysis of particular design issues, and an overview of program effectiveness and direction. Specifically, we were charged with addressing these issues:

- Is the conceptual model sound?
- Is the strategy valid and meaningful in the context of big river restoration?
- Are the objectives feasible?
- Are the sampling design, protocol and objectives relevant and valid?
- Does the effort enable us to detect Pallid responses to habitat manipulations over time?
- Does the program adequately address scale (i.e. basin, reach, bend, macro, micro)?
- Provide a list of recommendations identifying strengths, weaknesses and suggestions to improve design, protocols and objectives.

Responses to major review issues

- Is the conceptual model sound?

Overall, the panel thought that the conceptual model was useful, but should be developed further, particularly with increased linkages to physical, chemical or other biological factors.

- Is the strategy valid and meaningful in the context of big river restoration?

This is a good strategy, at least comparable to other big river and big ecosystem restoration efforts. The Before-After-Control Impact (BACI) design is appropriate, although not the only possible alternative.

- Are the objectives feasible? Are the sampling design, protocol and objectives relevant and valid?

Most of the goals appear feasible, but the panel has concerns with statistical design, and strongly urges additional statistical work and support.

- Does the effort enable us to detect Pallid Sturgeon responses to habitat manipulations over time?

This is hard to address at this time, and stronger linkage to predictions of the conceptual model are recommended.

- Does the program adequately address scale (i.e. basin, reach, bend, macro, micro)?

Adequately addressing many different scales in one program is a difficult task; the general approach is reasonable, but stronger statistical support is recommended.

- Provide a list of recommendations identifying strengths, weaknesses and suggestions to improve design, protocols and objectives.

See below

Review of HAMP program

The review team was unanimous in recognizing that USACE and their cooperators are tackling a major restoration and monitoring program with diligence, appropriate fore-thought, and a considerable level of sophistication. Many comparable efforts elsewhere have proceeded without adequate 'up-front' consideration of design, or of eventual analysis needs. The approach taken here, with extensive design discussions, consultation, and review, is admirable and should serve as a model for other efforts.

The comments and reviews offered here should be seen against this backdrop of support for the strategy being employed by USACE. The review panel has several substantive concerns about the details of the program as set out here; however the panel strongly supports the measured, thoughtful and adaptive way in which the monitoring strategy for the Missouri River is being developed.

Overall, the review panel feels that the HAMP strategy, while conceptually sound and based on appropriate designs, is as yet not fully articulated. There appears to be a wealth of information that has not yet been fully integrated into the protocols (as evidenced by discussions between panel members and HAMP participants). Similarly, the conceptual model appears to be appropriate, but is not yet complete or fully integrated with this planning document – several panel members argued that it should be placed earlier in the document and serve as the basis for all that follows. Given these overall comments, the HAMP strategy document should best be seen as a 'work in progress', requiring further refinement and perhaps further review.

The review panel strongly supports the experimental approach and the BACI design as proposed. This is an entirely appropriate way to tackle the distinct issues addressed by this program (which differs from that of the Pallid Sturgeon monitoring program previously reviewed, SEI 2004). However the panel notes that this is not the only approach to issues at the habitat scale, and that other techniques might be worth considering as supplemental efforts. The panel felt that the HAMP strategy was on '*a*

right track, not necessarily *the* right track' – that is that the BACI approach should be considered alongside alternatives including regressions, paired analyses, more heavily stratified designs, plus the use of investigative efforts such as individual-based models (IBM) and telemetry, and energetics. Given that the USACE have selected only one of these alternatives, it is perhaps appropriate that the HAMP strategy document sets out in more detail why this decision was made – developing such a section in the document might aid the co-operators in assessing the value of incorporating some of these alternative or supplementary techniques.

Throughout the individual and group reviews of the HAMP strategy, the panelists returned again and again to a critical issue – the central role of statistical design concerns, and the relative lack of statistical support for the program. The panel recognizes the many logistical constraints placed on this program, as well as the current budgetary constraints of USACE. The panel also knows that USACE staff are fully aware of the importance of strong statistical support, and that Dr. Wildhaber has been diligently supplying this in the time he is available. However the panel is concerned that USACE has yet to provide additional commitments to this issue. The panel feels strongly that a much more developed design effort is warranted, and that this is best achieved by a team of statistical advisors. Further, such statistical advice should be seen as a primary driver of the monitoring strategy. There is a lingering tendency in the program (as currently described) for instance to use many gear types, and alternative approaches in an exploratory fashion. While the panel supports (to an extent) adaptive monitoring, the panel is concerned that important considerations such as consistency, power, and replicability are being compromised in the mix. These statistical concerns are not options – they are essential. A firm statistical hand, able to argue for consistency across States and time, will be necessary to ensure that the HAMP program optimizes its efforts and benefits. Several options are available to ensure better statistical support (increased staff time, use of outside consultants, oversight panels, etc) – the panel urges USACE to regard this as an item of top priority.

The panel felt that the current HAMP documents do not yet provide a good indication of how shallow water habitat is expected to influence Pallid Sturgeon. This will be important to develop as the program addresses the linkage between hydrologic and biological data. The documents should also address whether shallow water habitat created in the short term persists into the long term. Is one goal to address the effectiveness of different structures? Perhaps this will be discussed in the upcoming hydrological modeling documents – the panel would like to see this discussed in some format.

The panel agreed together that gear needs to be consistently deployed in order to be useful. If particular gear needs to be used to sample one type of habitat, it must be used consistently (even when no catches are expected) to permit any real statistical inference. Every additional gear type added will greatly add to the logistical constraints and at least potentially will reduce effort used in other sampling – hence there is a very strong analytical cost to using too many gear types. This point was made previously by the review team (SEI 2004) but is still valid here.

There was some concern that there would be non-independence of data (including down-stream effects). These issues should be addressed in the design effort.

The panel also identified some opportunities that might be worth exploring, notably exploratory simulations based on past historic flows.

Another design issue that might be profitably discussed is sampling during the period of habitat modification – this represents disturbance rather than successful habitat change. Should these data be included in the analysis, and if so, how, and what information do they provide?

The panel discussions included extensive debate over the relative merits of an adaptive approach to monitoring (Goal 6). Initially, from the written reviews, it appeared that there was some disagreement among the panelists. However in discussion, it rapidly became clear that all the panel members had the same strong reservations about unstructured

changes to a program in ‘mid-stream’. Essentially the issue is again one of statistical power and consistency – changes in design can rob a program of its ability to infer success or failure of a management action. At the same time, the panel also recognizes that this program is still in an exploratory phase, and that new information could greatly increase the effectiveness of monitoring. This creates a tension between ‘future effectiveness’ and ‘current level of inference’. The panel notes that few programs have yet successfully used adaptive monitoring – these few have mostly employed nested designs and subsamples, which provided estimates of variance, leading to decisions on dropping particular elements. The panel felt that a program that countenanced major design modifications along the way could be ‘a recipe for disaster’. In particular, success evaluations relying heavily on a BACI paradigm cannot provide reliable inference if the “before” measurements differ substantially in protocol from the “after”. The panel recommends that the adaptive element of monitoring needs strong statistical oversight, and that the co-operators attempt to decide *up-front* which methods can potentially be altered (eliminated?) along the way without compromising analytical power. Dr. Peery’s comments are congruent with the panels’ – he recommends extensive, early analysis of data as the best way to ensure that adaptive monitoring does not lead to loss of power.

Recommendations

Each panelist made individual assessments and recommendations, provided below. The following are the main group recommendations:

1. Stronger statistical support is critical, and this program should not push forward without ensuring adequate effort (through additional staff time, an outside advisory panel, an expert consultant group, or other means).
2. The current HAMP strategy, while essentially on a good track, needs to be better articulated, and the protocols need to be explained more fully.
3. The conceptual models should be developed further.
4. Caution is urged regarding adaptive monitoring (goal 6), absent strong statistical guidance and oversight.

Individual reviews

Reviewers:

Prof. Nancy Auer, Michigan Technological University
Dr. Henrietta Jager, Oak Ridge National Laboratory
Prof. Jim Quinn, University California at Davis
Prof. David Secor, Chesapeake Biological Laboratory

Statistical commentary

Dr. Zach Peery, Moss Landing Biological Laboratory

Review of the Habitat Assessment and Monitoring Program Study Plan

Jim Quinn

General Comments:

The general approach of the Habitat Assessment and Monitoring Program Study Plan to evaluating the effectiveness of habitat modification on pallid sturgeon habitat use, demographics, and populations appears sound, if, to date, not very well fleshed out. The goals are appropriate, and the hypotheses stated are subject to direct investigation. The core of the proposed monitoring program is built around a Before-After-Control-Impact (BACI) design, a quasi-experimental analysis that is becoming increasingly used in large studies where full experimental replication is infeasible (e.g., port and estuary management). Instead, the studies are built around paired sites (e.g., river bends) monitored before and after a major management manipulation (in this case, notching levees, creating chutes, etc.), where for each manipulated site, the ecological response of an unmanipulated site acts as a control to the same measurements taken before and after the action (notching, etc.) at the manipulated site.

A BACI design is an appropriate approach, but not the only one available. For example, bends could be chosen randomly from the full population of bends (probably in a design stratified by river segment or region as well as presence or absence of an engineered treatment) with the relative impacts of a variety of freely varying drivers of fish populations (flow, depth, temperature, water chemistry, productivity...) all assessed simultaneously by regression or analogous statistical models. In this case, there are effectively two paired controls.

The principal advantage of a paired design is that it focuses the assessment on impacts of the drivers (levees, weirs, chutes, dikes...) most under the control of the USACE and collaborating agencies. For any given level of sampling effort, this probably increases the statistical power with which effects of construction activities can be detected, which presumably is of particular interest to policymakers.

Downsides include that it is unlikely than bends with engineered structures are representative of the river ecosystem as a whole, making extrapolating from the bend scale to larger scale measures more difficult. (I recognize that the document states the intention of working around this issue, e.g., on p. 8, but the degree to which it can be done clearly has not yet been worked out.) A paired design may also make it more difficult to mesh the monitoring with the ongoing monitoring approach for sturgeon (and other big-river fish) populations, as most of the sites for the latter were chosen from a stratified design to be representative of the overall population of bends. (By the same token, the hydrological, geomorphological, and other non-fish attributes to be measured under Goal 1.1 will be representative of constructed sites rather than of the river system as a whole, which may make them less useful for other kinds of investigation and management use.) If we presume that the most readily paired sites are often near one another, there may also be more problems with independence of measurements than if they were placed randomly or systematically along the river, and impacts of proximity/distance may be more difficult to estimate. A general challenge of BACI studies is that alternative possible site-pairs are rarely equally good matches for all of the physio-chemical drivers assumed to be important, so controls for some variable are much less good than for others. Care is needed to tease out the contributions of poorly-controlled extrinsic variability in variables not well treated by the experimental design.

As with other pallid sturgeon monitoring in the Missouri basin, gathering enough data to answer questions about fish behavior, demographics, and population changes with enough statistical power to detect all but the largest changes may prove challenging, and the authors freely admit that both some experimentation with multiple gears and other methods and ongoing power analysis will be needed to adaptively adjust protocols and

effort levels for effectiveness. I assume that this will be done jointly with the other fish monitoring programs in the study region and take advantage of the protocol testing that has already been done there. Last year's SEI review noted that several core methods (e.g., trawls and trammel nets) appeared to have greater abilities to detect population changes than did other gears, at least at current effort levels, and the proposed plan should certainly start with the assumption that the Pallid Sturgeon Population Assessment Protocols core methods will be shared (pp. 16-17).

That review also suggested that the population assessment programs consider more telemetering and tracking of individual adults, allowing the deployment of individual-based model (IBM) approaches to inferring population change. In other words, one tries to assess whether the survivorship, energetics and reproductive output of an individual female is enough to replace herself in the population ($R_0 > 1$) rather than whether the year-to-year population change is positive or negative. IBMs are especially appropriate for the hypotheses addressing habitat requirements and preferences of sturgeon.

I was a little surprised that there was not more discussion of how to assess the degree to which constructed methods will interact with the modifications in dam releases planned under the BO, since I presume that pulses of higher flow, different seasonality, and higher variability will all substantially change the features of enhanced habitats that will be created under this program.

Obviously, the present document is more conceptual a conceptual framework for developing a habitat assessment and monitoring plan than it is a protocol proposal, and I can't comment on the particular feasibility of suitability of particular proposed methods, field or analytical, until they are better defined. As occurred with the population assessment protocols, I suspect it is reasonable to expect 2-3 years of somewhat exploratory protocol testing, but I would urge the Corps to solidify the methods fairly quickly, as Before-After methods can only work if the same things are being measured in the same way on both sides of the construction events. This might suggest erring in the direction of making relatively simple (especially sensored) measurements in a lot of

places around experimental bends at the expense of more complex measurements that can only be done occasionally by expensively equipped field crews. (If physical measurements are being taken by remote sensing, that probably also simplifies both telemetering pallids and interpreting the meanings of their movements in IBMs.

Some specific notes:

Goal 1*:* The bend and meso-habitat scale are fairly well defined. It is unclear to me how the coarse-scale analysis is conceived (e.g., for measuring the importance of distance effects.)

Goals 1 and 2: I assume that the list of agreed-upon physical attributes and metrics to be measured at the bend and meso-habitat is more extensive and explicit than listed here. It would strengthen the document to state what measures and metrics the Corps knows will be used and which are still under discussion.

Success criteria: Criteria 1 and 2 are really about achieving adequate statistical power to detect changes at levels (5% degradation or improvement in pallid reproduction???) that are needed to guide management (e.g., funding and regulatory) decisions and inform the adaptive management process. One would like to think success is achieving high power to inform at relatively low cost.

Adaptive Management: Still elusive. The process chart is quite sensible, but is too general to infer how the adaptive management process will work in practice. A governance structure, appropriate shared (probably distributed) data management and analytical tools, and some independent advisory process (at least for credibility and the appearance of objectivity) will all be important, but I realize that is not ready for this document.

Program Development (p. 8):

"The plan seeks to incorporate (to the degree possible) existing sampling protocols and efforts for the other active Missouri River Programs (BSNP Mitigation, Missouri National Recreational River, Pallid Sturgeon Population Assessment, Pallid Sturgeon Research Monitoring & Evaluation) to ensure that these programs are complimentary to one another."

This is clearly a key to success, and I know from the previous SEI review that a lot of thought has gone into protocol and data sharing, but the discussion here is too brief to assess how this plan proposes to fit in.

Goals and Objectives (p. 10): This would be a good place to describe the potential synergies between constructed habitat enhancement and the impacts of changes in dam and facility operations. Presumably both will have adaptive management processes that should inform each other, and might come up with combined adjustments that are more than the sum of the parts.

Goals and Hypotheses (p. 10ff):

1.1, 1.2, 2.2 and others: The null hypothesis under a BACI model is really that the after-minus-before measures (rate of change) in the controls do not differ from those in the paired experimental sites. The absolute population densities, pool depths, etc., could be higher or lower, so long as they stayed that way over time.

2.1: Personally, I would not spend too much time on edge effects and fragmentation. That kind of questions gets investigators into documenting polygon properties rather than point properties (GIS rather than databases), which is certainly an order of magnitude more challenging for field crews. It is not obvious from theory, or as far as I know the sturgeon literature, what effect small-scale fragmentation has on such large mobile vertebrates anyway, so it is not clear that mapping habitats at the polygon level will yield

information useful for management. (And I say this as a GIS guy who does that kind of analysis.)

1.1b, 2.1b and others: If the hypothesis of interest is (construction value) > (control value), the converse/null is construction $\leq$ control, not construction = null as stated. (This is a nicety of expression -- the idea is clear enough.)

Goals 3 - 6: Bullets are not stated as hypotheses, as in 1 and 2.

Goal 3: This is an important goal, but I have no idea how it will be achieved.

Data Analysis (p. 14): It will save a lot of time, money, and stomach lining to develop data exchange standards and vocabularies early. Individual programs will no doubt continue to have idiosyncratic datasets, variable names and codes, etc. for legacy reasons or because of the demands for interoperability with different sets of partners, but everybody should think early about how to translate their data into a format/schema that everybody agrees to support for information sharing and synthesis. The USGS's NBII and EPA's NEIEN programs are work-in-progress frameworks for doing some of this, and adapting their standards and tools (or at least understanding their mistakes) might help launch a set of data sharing standards that will be more or less interoperable with a wide range of management and scientific information systems. If it hasn't already been done, it might be worth establishing a separate data management working group and involving people from some set of outside networks (such as NBII, NEIEN, LTER, etc.) with similar goals in designing the strategy.

Gears:

The statement

"Then, based on the habitat designation the team will use a suite of appropriate gears to sample those habitats. It is our intention to throw as many different gears as possible at

these created habitats to ensure that we are sampling the habitats and organisms effectively."

worries me as a potential diffusion of effort in a situation where statistical power using the core methodologies may be marginal for a very rare fish. At very least, it should be closely coordinated with similar experimentation among the population assessment programs.

Review Comments on: Habitat Assessment and Monitoring Program
– 9/15/2005 version Study Plan.

From Nancy Auer

Although a great deal of work has produced this September 15 version of the plan it seems to me to still be in some preliminary development stage. Illustrations and references are incomplete. Comments have been included from one individual but I have no idea who this individual is or how those comments should be considered. I have proceeded and made the following comments.

General Comments:

a) If you really don't have priorities in the way goals are listed then use some other system to order them (not numeric numbers, perhaps simply a, b, or c, some simple symbols). Using numbers instantly prioritizes your goals even if that is not the intent.

b) With habitat modifications described in 1.2a – will there be a measure of the dewatering (will water that moves into new areas seep into soil faster and cause drying faster and fish stranding, especially larvae or juveniles?). Is there a way to measure water exchange behind these modifications? Will there be measures of sediment accumulation near these modified structures?

c) How will you account for possible cumulative affects of many upstream modification actions on modification success further downstream? – i.e., higher sediment loads further downstream?

d) Are all fish to be evaluated and released alive? What are cumulative effects of fish removal if that occurs, again as success is measured at lower river modifications?

e) The figure on page 17 – habitat example – makes no sense to me at all.

f) Has it been considered that these river bend modifications could enhance food organisms or debris loading (which are indirectly important to pallid sturgeon) but pallid sturgeon may not be captured in these areas? In the BO, research showed that larval sturgeon were encountered at island tips along with woody debris. How will you measure "biological attributes" produced at one place (modifications) but actually "needed" at another place for cover, concentrating food, etc. for pallid? Are island tips a focus of any portion of the study? This plan only focuses on the river bend modifications. What about sampling areas that develop downstream as a result of these modifications, again island tips?

g) This plan of study does not seem to address the desire to normalize the hydrograph and temperature conditions for spring spawning pallid sturgeon and for their resulting offspring. Perhaps flow assessment is a separate document? With flow changes taking place in 2006, how will these impact habitat evaluations, etc.?

<u>Specific Comments:</u>

a) Under Goal 1.2b there are 'target species' and 'life stage' terminology used. Why not identify the target species and life stages that are to be targeted? What are the target species? Under Goal 2.2b – there appears "life stage richness"?

b) The comments from Fleming include several hypotheses on "obligate native big river fish." Again what species will be important to focus on? Who are the "obligate native big river fish?" or target fish?

c) Although this is a pallid sturgeon plan, is there no intent to investigate food items (invertebrates) or other aquatic species (mussels) that may be used as preliminary indicators of improving habitat?

d) In the plan there is no intent to evaluate spread or monitoring of exotic species.

I noticed both non-native and invasive fish species were mentioned in the Fleming comments as well. It will be important to monitor them and any interfering influence they may have on recovery efforts – again perhaps list expected ones.

e) For linkage attributes it seems important to clarify and consistently use similar terms for physical, chemical, and biological attributes. They are all termed attributes under Goal 2 but that changes to physical and chemical parameters linked to biological attributes in the small paragraph right underneath Goal 2.

Define Linkage Attributes As:

Physical	**Chemical**	**Biological**
Turbidity	DO	Woody debris
Temperature		Submerged vegetation

There are some indices that use measures of a whole suite of variables to classify stream/river characteristics – have they been considered?

Also what determines biologically beneficial? (see Goal 2.1a).

Review of "Habitat Assessment and Monitoring Program Study Plan"

H.I. Jager

Background

Last year, we were asked how the earlier sampling design would perform with respect to the goal of detecting overall trends. Power analysis suggested that adequate power for pallid trend detection was possible with fewer subsamples and more bends sampled, but only at the largest spatial scale and not within state boundaries. Thus, our recommendation was to shift effort to a coarser spatial scale than bend-level (within bend) sampling. Other key recommendations we made were to focus on fewer gear types and to involve a design statistician.

This new document describes a different experiment that focuses on detecting the effects of micro (structure) scale habitat modifications. This experiment has a different objective from that of our previous review, and therefore requires a new design to meet its objectives. Whereas the previous objective was one of estimating trends for an entire population, best met by a sampling design, this study's objective is hypothesis testing, best met by an experimental design (assuming that there is no interest in regional extrapolation). A BACI (actually, a BACIP - paired) design is proposed to test the null hypothesis of no effect. Creation of shallow-water habitat (SWH) and increased biological productivity are the two main goals.

Critique of Goal Definitions

Study Goals and Objectives. I would expect this section to begin with a discussion of what the large-river habitat looked like historically, followed by an outline of the conceptual model in biologists minds of how creating shallow-water habitat may lead to increased pallid densities. A diagram of this would be really helpful. This is especially important because those not familiar with this paradigm are unlikely to understand why SWH might benefit a species that ostensibly needs deeper habitat. I'm not sure I would

start with a discussion of uncertainties, unless it is made clear how the goals fit into them – i.e., by reducing uncertainty.

One issue is that the plan does not explicitly lay out what the desired change in habitat is, how it will be measured, what the measured biological responses are early enough. In the sections below are comments on each of the goals.

<u>Goal 1.1a</u> Are physical treatments producing the quantity and quality of habitats described in Appendix A.

My understanding of the desired future condition is closer to the mosaic of habitat types that previously existed in the Missouri River; a less-channelized river with more-diverse mosaic of habitats, including floodplain and off-channel SWH. Is the goal to recreate natural river dynamics that maintain SWH over the long term, or is the goal to produce a short term build up of sediment in the main channel that will erode away after the first big flood? Since there is not much sediment recruitment except at confluences, creating shallow habitat in one place might have adverse consequences elsewhere, unless the sinuosity is increased and/or habitats are created away from the main channel. I am not a fluvial geomorphologist (I can hardly pronounce it), but the goal of increasing shallow-water habitat (SWH) during a ~5-y time horizon may not be the best measure of whether this longer-term goal will been met (although I understand that the BiOp requires a specific acreage of SWH to be created). In an earlier study, Thompson (2002) found that many habitat improvement structures were much more effective in the short term than in the long-term.

I would refine performance measures to distinguish between short-term and persistent SWH. I did a very quick review of river restoration literature (see below), and found that guidelines exist that might help define habitat-related goals. However, they are vague and it would be a good idea to involve a fluvial geomorphologist who could translate these restoration goals into measurable quantities. For example, the presence of riparian vegetation is suggested as an index of persistent SWH.

Brief Review of River Restoration Literature. Scientific guidelines have been proposed for ecologically successful river restoration (Palmer et al. 2003). This plan is consistent with some of the restoration guidelines, but not others. The use of a BACI design is consistent with their recommendation for clearly testing the ability of the modifications to meet ecological goals. The standards for successful restoration differ from those in Goal 1.1 in several respects. Rather than measure potentially short-term changes in the amount of shallow-water habitat (presumably the net change), these guidelines focus on restoring natural geomorphic processes. However, it is not clear what the quantitative measure of this would be. Riparian vegetation is also proposed as a measure of successful river restoration--in particular, forest management that allows riparian trees to mature as future sources of in-channel wood (Palmer et al. 2003, Johnson 2002). Grading, breaching and levee widening are also viewed by Palmer et al. (2003) as ecologically effective restoration activities that reconnect the channel with its floodplain.

Recommendation 1. Involve a fluvial geomorphologist with big-river experience that can help to define quantitative measures of persistent SWH. Perhaps someone like Matt Kondolf (UC Berkeley) or David Galat. Alternatively, if the desired future condition is a dynamic one in which the location of SWH will change over time, the whole sampling design must scale up from the local focus on structures to the whole system. The plan could include a paragraph justifying these measures.

Recommendation 2. Conduct a review of previously published studies in the scientific literature to determine whether the proposed habitat improvement structures have been found to move the river habitat closer to the desired future condition. This will be important for choosing the length of time post-modification.

Recommendation 3. Consider extending the post-alteration monitoring to monitor both short- and long-term responses. For example, perhaps skip monitoring the year of the modification, monitor each year for a few years, then perhaps again after ten or fifteen years. This will provide both a short and longer-term assessment of benefits.

<u>Goal 1.1b</u> No differences in fish species composition, richness and relative abundance of targeted life stages.

This goal could be made more specific. It should relate directly to the conceptual model/hypothesis about how SWH influences pallids in a positive way (hypothetical Figure 1 recommended above). For example, the two hypotheses I can think of are 1) More SWH will increase juvenile pallid sturgeon survival and, ultimately, population size; and 2) More SWH will increase the abundance of prey for older pallid sturgeon, leading to an increase in population size.

In contrast, the relationship between fish species composition and pallid population status and the relationship between fish species richness and pallid population status are not clear. One might argue that pallids can probably eat pretty much any fish biomass of a certain size, perhaps with a bias for benthic species. Therefore, less effort could be spent on species identification, more effort on weighing and measuring length, and unidentified fish should definitely be recorded because they contribute to prey biomass. If other information is needed for other purposes, fine, but they don't need to be listed in the plan for this experiment.

<u>Goal 1.2a</u> Which habitat modifications are most effective in creating and sustaining habitat (SWH?)

What I like about this goal is that it distinguishes short-term creation of SWH from persistence of SWH. I am unclear what the experimental design is for the BACI test that compares structures (habitat modifications). Will this be tested using structure-level data? Is there also pairing between structures in the control and treatment bends, or will bends be treated as replicates and pooled for this analysis?

Will the sampling effort at the structure level be the same before and after the modification? I am highly concerned about statements like this one on page 17:

> "It is our intention to throw as many different gears as possible at these created habitats to ensure that we are sampling the habitats and organisms effectively"

It is unclear how the selection of control bends controls for the types and numbers of existing structures. It is also unclear whether structures could be considered as treatments with multiple different modifications within a treatment bend – independence is likely to be a significant issue. I suspect that a statistician would suggest that you sample from a list of structures in both bends, and I'm pretty sure it would be a good idea to have a subset of these having no modifications in either bend, but again, check with an expert.

Goal 1.2b Which habitat modifications are most effective in?

The comments above (Goal 1.2a and 1.1b) also apply here.

Goal 2 Linkage attributes?

Apparently this set of goals deals with a larger suite of response variables or covariates (not clear which), but a smaller sample size. The list includes both static variables (e.g., woody debris) and dynamic variables (e.g, temperature). It seems more logical to me to obtain this baseline information once before the modifications for all sites, and then perhaps at 5-y intervals or some manageable frequency. If this is to be part of the design, equations and ANOVA tables showing the role of these variables in the design are needed. What are the specific hypotheses to be tested?

Goal 3 Assess the effect of habitat covariates on biological response.

This goal attempts to address spatial differences among bends that might influence responses, and the potential interactions among modified structures. Most of the information required to meet these goals is baseline data that can be collected ahead of time and used by a statistician to tailor the design, for example, by avoiding atypical areas or spacing treatment and control bends. I would suggest compiling this information in the first year during a preliminary study. In revising the description of this goal, it would be good to formulate specific hypotheses about how you expect these things to influence biotic response to habitat modification.

Goals 4 and 5 Assess the effects of constructed habitat on "other ecosystem components" and on the recovery of pallid sturgeon at the basin level.

I suspect that a real statistician would agree with the proposal to conduct a separate, complementary effort as part of the pallid trend analysis program, to characterize SWH and structure status for bends drawn as part of the probability sample. A power analysis can tell whether or not there would be adequate power to detect trends between the collection of bends with and without modifications. This integration is a significant problem from a statistical design standpoint, and it's not one I have the expertise to address.

Goal 6 Adaptively manage sampling strategies to reduce uncertainties and increase efficiency

No – bad goal. Changing sampling strategies in mid-stream is likely to invalidate the conclusions of the study and its ability to test hypotheses set out by the sampling design. However, see Downs and Kondolf (2002).

Critique of Experimental Design

This document includes many elements required for a complete experimental design, but not all. The table below lists some essential components of a experimental design:

1. Define the experimental units and provide a plan for developing a list frame of these units and a plan for sampling them;
2. For each objective, identify treatments, distinguishing random and fixed effects; Provide an ANOVA table listing, in a hierarchical manner, sources of variation associate with each treatment, and giving the sample sizes;
3. Identify response variables for each goal or experiment;
4. Provide model equations for the quantities to be used in hypothesis testing (indicate treatments, interactions, if any, and error terms). Note that the same equations are not used for testing hypotheses associated with Goal 1.1 and 1.2 or 1.1 and 2.1. Yet, the design has to be capable of separating these sources of variation.
5. Once candidate designs are identified, conduct power analyses to help choose a design, the number of years before and after modification, and sample sizes.

The basic experimental design proposed for Goal 1.1 appears to be a paired BACI (a.k.a., BACIP). Paired comparisons are good because they increase power, and if the pairs are situated in the same segment, then it may not be necessary to treat river segments as treatments – giving you even more power. However, it is important to clarify whether the six river segments will be spatial treatments in the design. A power analysis will indicate whether detecting improvements at this spatial resolution is possible. With 48 pairs (trios) of bends, a power analysis is needed to determine whether you would have enough power to detect differences at each level of the hierarchy (bends vs. structures) and within vs. among river segments.

It is not clear what the experimental design for addressing Goal 1.2 is. Possibly a "beyond BACI" design with hierarchical sampling (Underwood 1994)? Is the idea to use the bend-level data to address Goal 1.1 and the structure-level data to address Goal 1.2? For each goal, equations and ANOVA tables should show what the treatments are. Discussion with Mark Wildhaber revealed a temporal aspect to the study that is not reflected in the plan, as written. This may have important implications – for example, it

may be necessary to include a 'year of modification' treatment effect in the experimental design.

Recommendation 4. There are a number of alternative designs to consider that could address program goals. The proposed approach is one of hypothesis testing. Ask the statistician to compare the power of a design that uses multiple control bends (possibly those drawn at random and sampled as part of the Pallid Assessment) vs. the paired design. The power analysis will also reveal how many years before and after the modifications are needed to detect an effect. According to Johnson et al. (2005), sampling at multiple pairs of sites (treated as replicates) or how long sampling records before and after the change is recommended to account for natural variation.

Sampling frame

I have not been able to find anywhere in the plan how many bends will be included in the experiment, although there is some mention of the number to be included in physical modeling. However, Mark Wildhaber called and explained a great deal about the experimental design that is not in the plan. For example, he says there are basically three types of bends:
A) control bends that have never been modified (except for some revetment) and will not be in future because the surrounding land is not available for purchase,
B) too-late bends that have already been modified, and
C) treatment bends that the Corps can buy surrounding land and therefore plans to modify.

The full set of bends includes 48 bends, but not all in one year. Possibly 12 or 16 trios will be sampled each year.

I have no idea how the too-late bends will be used in the data analysis, so it might be a waste of time collecting data on these bends. It would certainly be worthwhile running this by a statistician. I am just ignoring this component of the design for the time being.

Hierarchical sampling

It is unclear how the data from bend-level sampling vs. structure-level sampling will be used in the statistical analysis and hypothesis testing. Underwood (1994) considers beyond-BACI designs with multiple spatial scales for situations in which the spatial "footprint" of the treatment effect is not known ahead of time. Again, someone with a PhD in statistics should ensure that the structure-level effects are properly included in the design equations.

One statistical issue is that there is an implicit assumption that changes in one place will not have an effect farther downstream – i.e., that bends are independent. Clearly this is not consistent with principles of fluvial geomorphology. If shallow areas are created upstream, is there a chance that existing shallow habitat farther downstream will be deepened? This is not my area of expertise, but it's an issue that should be addressed by the design, and is reflected in Goal 5. I would think it best to choose treatment bends downstream of control bends (and above the "too-late" bends, if you must include them), but again, check with a statistician – maybe doing this would introduce a bias).
The report doesn't specifically identify a response variable, so maybe pallid densities aren't the response of interest. The figure on page 20 says "Increased use of created habitat by pallids and other native fishes" and "Increased biological production..." Goal 2B refers to "obligate native big river fish abundance" and "non-native or invasive fish abundance".

Recommendation 5. The expected response is one of increased forage fishes for pallids (and possibly habitat for juveniles) – it is highly unlikely that adult pallids would be jumping into shallow habitat. Therefore:
A) I support the decision to focus on summer sampling only.
B) Stick with one or two gears and drop the gear comparison (make it a separate targeted study). In general, it is a bad idea to try to make one design fit all the various objectives that a program has.

Recommendation 6. Define the desired change in habitat and specific response variables, and give justification for how they relate to pallid recovery.

Scaling to the Basin level pallid response

Because bends are not selected in a probabilistic manner, this study as it is currently designed will not be tied formally in any statistical manner to the population survey. The proposed experimental design can address the hypothesis that the habitat modifications do or do not have the desired effect. It is not designed to estimate changes in habitat or biological production for the whole river.

Goal 5 suggests that the design is intended to do both. It is beyond my expertise to suggest how to integrate the designs for the two programs. You would be well advised to consult a specialized design statistician ahead of time – i.e., before selecting bends. Since you have a list of three types of bends, it would potentially be possible to sample from them in a way that would be consistent with the other survey.

According to my statistician-friends, it is a mistake to try to use the same design for multiple objectives. In this case, to test the effectiveness of micro-scale structures and to address the original systemwide question of trends. These should be two separate studies, sharing protocols when it makes sense. If an integrated design that links the two is needed, I would strongly recommend the involvement of specialized statistical expertise for an effort of this size and complexity.

Recommendation 7. Involve a design statistician to help develop a design for both spatial scales and both questions. Unfortunately, there are few ecological statisticians with this expertise. I would call Dr. Scott Overton (OSU emeritus) for recommendations at 541/929-3715 (overton@casco.net).

Literature Cited

Downs, P.W. and M.G. Kondolff. 2002. Post-project appraisals in adaptive management of river channel restoration. Environmental Management 29(4): 477-496.

Johnson, M.J., and 4 authors. 2005. A novel approach to assess the impact of landuse activity on chemical and biological parameters in river catchments. Freshwater Biology 50: 1273-1289.

Palmer, M.A. and 20 authors. 2005. Standards for ecologically successful river restoration. Journal of Applied Ecology 42: 208-217.

Thompson, D.M. 2002. Long-term effect of instream habitat improvement structures on channel morphology along the Blackledge and Salmon Rivers, Connecticut, USA. Environmental Management 29(1): 250-265.

Underwood, A.J. 1994. On beyond BACI: Sampling designs that might reliably detect environmental disturbances. Ecological Applications 4(1) 3-15.

Review by David Secor

Review – Habitat Assessment and Monitoring Program – Study Plan

Overall: Habitat assessment here is based upon evaluating the influence of habitat creation actions (engineered channel and bank modifications) on pallid sturgeon recovery, which is justified through current set of concepts on restoring a channelized hydrograph to one of more diverse flow regime and increased benthic, shoreline and floodplain habitat. I also found the Before/After/Impact/Control (BAIC) study design concept appropriate here in that it would provide alternative means of evaluation in either time (Before/After) or space (Impact/Control). A focus on linkage parameters, albeit somewhat ambiguously defined, also seemed warranted given that these linkage parameters (water quality, litter, other microhabitat features) are dynamic and likely responsive to habitat creation actions. Thus, unlike pallid and other fish community attributes, which would be expected to lag temporally or show nonlinear responses, changes in linkage parameters would be more efficiently measured and detected as response variables.

Unfortunately, the Plan is rather opaque in my opinion due to 1) lack of a clear overall goal statement that is linked to the conceptual model; 2) redundancy, jargon, and organization within goals statement; 3) in the face of uncertainty on effects of habitat creation, a plan that would defer consideration of monitoring, a sampling design, and statistical analysis to some future date; 4) inadequate review of habitat creation actions and their hypothesized effects on pallid sturgeon habitats (App. 1). The report is also incomplete – missing Table 2 and appendices other than App. 1.

Organizationally, I would advise moving the concept model to the front of the document. Further, I am a big believer that a picture is worth a thousand words and would urge reference to conceptual drawing approach, a very useful primer and tools are available at http://ian.umces.edu/ (look under scientific communication and image library). I would also simplifying figures and structures in App. 1 to highlight expected riverine/ecological

responses to the structures. I look at these as central and it is worthwhile considering moving these out of an appendix into section that follows conceptual diagram and goals statement. (Structure of document might be introduction, background, concept, goals, habitat creation actions, tools for measuring effects of habitat creation, approach.) As indicated above, I think linkage attributes should become a more central part of the approach and better explicated throughout the document.

Specific Comments:

Executive Summary – I think here there should be mention of the underlying concept behind habitat creation (restoration of diversity in hydrograph, diversity in alluvial habitats, etc.). I have left goals alone since I comment on them later below. The statement that the process for adaptive framework is undefined is problematic. I think more could be said here that the adaptive framework will rely in iterative review from internal and external scientific committees employing statistical tools such as power analysis to insure that monitoring practices are well matched to evaluation of expected outcomes.

Introduction – This is a nice section, but I would like to see conceptual diagram revised (see above) and included here. Statement in second paragraph on converting 1500 miles of riverine habitats to 1500 miles of lakes seems confusing with second clause not clearly linked to this first idea. Top of p. 6 – " ..action specific end states" ugh! I urge simplifying language here. Second paragr. P. 6, second to last statement – Statement of uncertainty should be recast and probably form subject of separate paragraph that introduces idea of focusing on linkage parameters. Even though there is uncertainty here, the linkage parameters are supported by the current conceptual models and as indicated above there are other good reasons in terms of monitoring and analysis to focus on these. I would suggest that these linkage parameters could be alternatively termed as the more traditional – indicator variables, upon which other ecosystem restoration programs focus.

Study Goals – second bullet on p. 9. I think this is overstated. Perhaps we are unable to identify ALL habitat requirements, but we know in most trivial case that sturgeon need water! I would rephrase – we remain uncertain on the specific effects of habitat variables (indicator/linkage variables?) and vital rates and behaviors of specific life history stages of sturgeon…

Goals and Objectives

I am a bit confused on what to review here as I have two sets of these – one sent with comments marked upon them. I liked this second one better, but address comments to the first set, which should be more consistent with other drafted sections of the plan. I think the goals move too quickly to the specific – time periods and segments, etc. and then back off specification when it is required – principal linkage variables and fish response variables, integration of the BAIC approach, implementation of power analysis. My suggestion would be to start with an overarching goal statement that reiterates focus on assessing both biotic and physical responses to habitat creation actions, which are expected to benefit pallid sturgeon and related community. Then move to specific goals on 1. macro-scale physical changes – did the habitat creation action create the intended macrohabitat feature? Are spatial arrangements of habitat creation projects (currently goal 3) important? 2. linkage parameters – did the habitat creation action cause changes in turbidity, bottom conditions, flow, etc.? 3. fish community metrics – a. did we see a response of fish community to habitat creation in comparative sense (BAIC approach)? 4. Are fish communities responding to linkage parameters according to our expectations (not really included thus far in list of goals, but could be assessed using general linear model, general additive models, etc.)? For these revised goals 2 and 3, subgoals would focus on assessments at differing temporal and spatial scales as now occurs in this section. Goals 4 – 6 in the original document seem well placed and appropriate; although Goals 4 and 5 seem redundant. For goal 4/5, I might suggest hatchery fish could be used experimentally to evaluate habitat creation projects. For goal 5/6, the team may wish to explicate use of power analysis, particularly for looking at fish community and linkage parameter metrics under BAIC approach.

General Approach

P. 14, second paragraph. I think a list of linkage parameters and other physical data should be included. Last statement of paragr. is problematic. If process for analysis is uncertain, then two or three specific examples or expected outcomes should be highlighted to elucidate the type of analysis that might occur. The impression here is that we will amass a lot of information without consideration of survey design (that is, statistical analysis) and give it over to someone else to sort out. The BAIC approach should provide a framework to outline a few likely types of analysis that will be undertaken in the coming years.

P. 15, I really like the use of controls in time and space. But what if control sites are not correctly chosen? Here, replication would seem important to hedge against confounding underlying variation (climate, system-wide hydrograph) and poor control site collection. Has the team given any consideration on the relative merits of replication in time (longitudinal analyses – repeated measures treatment of data) versus space (cross-sectional analysis – ANOVA type treatment of data)? Unfortunately no treatment sites are listed in Table 1 to evaluate the intended degree of replication.

P. 16, second paragr. The three year's of sampling before treatment seems good, given practical considerations, but would be good to compare this with interannual variance (c.v.) measure from underlying hydrograph. How much of this underlying variance is expected to be captured in 3 yr period?

P. 17, end of first paragr. Throwing out as much gear as possible seems like too much adaptation. As indicated above, I think if the first focus is on linkage (indicator) variables, then these can be consistently measured across sites. I recognize that these may not be directly related to desired pallid sturgeon responses, but it is important to measure some relevant feature at a scale likely to be responsive to habitat creation

actions. Also, there is precedence for using indicator variables as surrogates for fish community metrics in other ecosystem restoration programs. With regards to throwing gear out – what if a novel catfish trap catches sturgeon at a treatment site, will this treatment be deemed a success if that gear has never been used historically or not deployed concurrently at reference sites? I really urge as much standardization in gear deployments as possible. I would match gear types with mesohabitats as has been done in the pallid sturgeon assessment plan.

P. 18, paragraph 2 seems too focused on logistical constraints of putting out as much gear as possible to capture sturgeon. I think this is important for pallid assessment activities, but any balance of effort towards habitat should probably be weighted towards standard measures of linkage variables and standard gear deployments (based upon mesohabitat) for sturgeons and related fish community.

P. 18, last paragr. I think it would be good to urge that initial habitat creations be structured to permit improved evaluation through power analysis, if possible. For instance, perhaps a single type of habitat creation could be analyzed across several segments.

Appendix A. I think this is an important appendix and needs to be modified to fully define acronyms, reduce jargon, and provide simplified figures on structures and expected habitat modifications.

Comments on statistical design issues by Dr. Zach Peery

My two major comments can be summarized as follows.

1) When randomly selecting bends within segment for treatments/controls I would avoid selecting bends corresponding to the 25th and 75th percentiles of bend size. Restricting sampling to bends of these sizes will limit inference of treatment effects to these bend sizes. If treatment effects depend on bend size, only a limited understanding of treatment effects will result. I suggest randomly selecting bends from the full spectrum of bend sizes or randomly selecting bends from within bend size strata (e.g. small, medium and large).

2) I suggest that following the first year of field work, a thorough evaluation of the data is conducted. Specifically, are biological and physical variables distributed in a manner that will allow for robust modeling? For example, if Pallid Sturgeon are rarely detected, which will probably be the case, it may not be possible to conduct detailed habitat evaluations and estimate treatment effects. Consulting with a professional statistician would greatly help this effort.