Scientific and Statistical Review of Appendix C, Scientific Investigation Program & Monitoring Plan, Final Administrative Draft of Biological Assessment, 2-Gates Fish Protection Demonstration Project

Ken B. Newman[a], Pete Smith[b], Pete Adams[c], Randy Baxter[d], Mike Chotkowski[e], Steve Culberson[a1]

October 28, 2009

Summary of Top Concerns

1. Quantitative forecasts of the magnitude of possible benefits of 2-Gates to Delta Smelt and other listed species are lacking. This is important because without knowing the size of the expected benefits it is difficult to evaluate the adequacy of the monitoring and evaluation plan. If the expected benefits are small, then it might be difficult to detect them within the term of this experiment as it is presently conceived.

2. Monitoring plans for other listed species including Green Sturgeon, steelhead, winter run Chinook salmon, spring run Chinook salmon, and Longfin Smelt lack development. In the absence of additional detail, it is difficult to determine whether the project will produce the information necessary to determine its effects on these species.

3. The time required to empirically demonstrate benefits of 2-Gates, with a specified level of precision, is not discussed. We are concerned that the time required could exceed the term of the experiment as it is presently conceived.

4. Gate operations undertaken when Delta smelt are in the small larval stage (< 20 mm), potentially a 6-8 week period, do not have an accompanying data collection and monitoring plan. Potential negative impacts to this life history stage would not be detected quickly, or perhaps at all.

Preliminaries

The purpose of this review is to evaluate scientific and statistical aspects of the science investigation program and monitoring plan of the Biological Assessment of the 2-Gates Fish Protection Demonstration Project (Project), as that program and plan is described in Appendix C, downloaded 21 October 2009. There are four parts to this review: (1) Background of relevant biological theories (as related to Delta Smelt (DSm)) and hydrology theories that underly the project; (2) Listing the 5 Hypotheses to be tested; (3) Describing the Data Collection, Generation, and Analysis procedures aimed at testing the hypotheses; (4) Critique of the Hypotheses and Data related procedures. We note that while this review is primarily an evaluation of the proposal, thus critical, we do in some places (e.g., Appendices A, B, and C) include ideas that might lead to alternative hypotheses and statistical and sampling procedures.

[1]a. USFWS; b. Independent Consultant; c. NOAA; d. CDFG; e. USBR. We thank Jeff Stuart, Dan Welsh, Dave Contreras, and Steve Slater for their assistance. For correspondence, the lead author can be contacted at `ken_newman@fws.gov`.

1 Biology and Hydrology Theories

The Project is a planned experiment to examine how the installation and operation of two manipulable barriers or gates installed in the Central Delta would affect the hydrodynamics of the Delta (as defined by flow, turbidity, salinity, and temperature, at least, by time and by location) and how the hydrodynamics would subsequently influence adult DSm and their progeny (as defined by spatial distribution, movement, entrainment at the SWP and CVP facilities, at least). Deliberate manipulation of the Delta's hydrodynamics and DSm, via gate operations, will occur from the months December through June. Gate operations from December to mid-March are aimed at influencing adult DSm, both pre-spawning and spawning (p C25). Later operations are aimed at influencing the progeny, larvae and juveniles. Other ESA listed species (steelhead, Chinook salmon, and green sturgeon) will also be impacted by gate operations and monitoring plans need to be developed for these species also.

Adult Delta Smelt. The relevant conceptual model, "theory", about DSm begins in December with the majority of pre-spawning adults residing in the western Delta, in and near Suisun Bay. These fish have been growing, yolking up, and producing gametes, preparing for reproduction. The fish will at some point migrate upstream toward less saline waters, i.e., into the interior Delta. Upstream migration is often triggered by the first freshets of the winter (around winter solstice) which cool the water and make it more turbulent[2]. The date of migration can vary from late December to early February. They complete the journey in roughly one week. Depending upon the timing of the upstream migration, they can spend up to two months upstream prior to spawning.

What happens during the period prior to spawning, in particular how the adults are spatially distributed, is crucial to population dynamics, i.e., survival and births. One theory is that the probability of an adult occupying a particular location is partially a function of turbidity levels, roughly put, the more turbid the more desirable the location. It may also be, to varying degrees, a function of flows, export levels, temperatures, and salinity. The spatial location of a given adult has direct bearing on the probability of survival. The higher the density of predators, say striped bass or largemouth bass, for example, the lower the probability of survival. The probability of entrainment (and likely death) is also a function of location—more directly a function of geography, flows, and export levels (at least). Locations of particular interest for the Project include Old and Middle Rivers due to their proximity to and physical connectivity to export facilities. General objectives behind 2-Gates are to manipulate gate operations so as to lessen the probability of entrainment of adult DSm.

Larval and juvenile DSm. Spawning occurs during the late-winter and spring months (late-March or early April through mid-May), with most spawning occurring during April and May when water temperatures are between 12-18° C. For the first three days after hatching the larvae are buoyant, and near the water surface. Hatch timing (slight neap-hatch tendency), negative buoyancy and limited swimming behavior of larvae help maintain them near where they hatched. As the larvae mature they become semi-buoyant and grow to a length of 15-25 mm within 40 days. Mobility increases rapidly as larvae complete air bladder development (ca 14 mm) and fin ray (ca 20 mm) development by about 25 to 40 days post hatch. With development of an air bladder, larvae can vertically migrate to affect their location in the Delta and as fin rays develop, overall motility improves. The larvae gradually move from the spawning areas to rearing areas (downstream towards low salinity zones, e.g., the area

[2]From the Appendix: Movement is triggered by "abrupt changes in flow and turbidity associated with the first large precipitation event of the season' in the basin"; turbidity $\geq$ 12-15 NTU "is correlated with and may be a functional cue for the annual spawning migration".

of X2 where the salinity is 2 ppt); these rearing areas will vary from year to year depending on the hydrology.

Within 70 days, they have grown to 30-40 mm in length and are considered juveniles. Young juveniles can move well vertically within the water column and develop modest ability to move laterally. It is believed that these abilities allow them to ride tidal currents to more favorable environmental locations. As south Delta water temperatures approach 25° C, juveniles move downstream to more saline water of the lower rivers and Suisun Bay. What causes that movement and how long it takes is unknown.

The biological theory behind 2-Gates is that the larvae are largely helpless in terms of spatial location and movement, they will go wherever the water takes them, acting like neutrally buoyant particles—somewhat at odds with above statements regarding their ability to maintain themselves near where they hatched. Thus their mortality will largely be a function of the location where they were born and future water movement patterns and whatever predators, toxic contaminants, or man-made structures might kill them.

Delta hydrodynamics. Hydrodynamics can be defined in terms of water flow, turbidity, salinity, and temperature by time and area. The primary flow sources are the Sacramento River and two of its tributaries (Mokelumne and Cosumnes Rivers) entering the Delta from the north, and the San Joaquin River from the south. In the north Delta near Walnut Grove, a portion of Sacramento River water is diverted into the Delta Cross Channel[3] (DCC) and Georgiana Slough where it is combined with the (normally) lesser flows of the Cosumnes and Mokelumne Rivers before entering the central Delta through the terminus of the Mokelumne River and through Little Potato Slough. When export flows in the south Delta exceed San Joaquin River flow, the Sacramento River water becomes an important source of exported water and arrives at the export facilities by flowing south (upstream) through the interior Delta channels of Old and Middle Rivers. The San Joaquin River water is diverted to the export facilities first from the east through the head of Old River, and then from the north, mostly down Middle River, after flowing through Turner and Columbia Cuts.

When high inflows to the Delta occur in winter, and especially during any first large pulse flow, the Sacramento and San Joaquin Rivers carry large sediment loads into the Delta causing turbidity levels to rise significantly. Historically, the onset of high adult DSm salvage at the fish facilities has occurred when turbid Sacramento River water reaches the south delta through transport by the reverse flows on Old and Middle Rivers. When a connection forms between turbid Sacramento River water in the central Delta with water in the south Delta it is referred to in the two-gate documents as a "turbidity bridge". The 2-Gates operations in winter are meant to delay, or possibly avoid entirely, the formation of this turbidity bridge.

Water exports and position of the DCC gates are currently the dominant human-induced influences on flow dynamics (and salinity and turbidity and temperature) in the Delta. For most conditions of even relatively low exports, reverse flows occur in the Old and Middle Rivers. The installation of 2-Gates, on Old River and in Connection Slough, would add a third means of manipulating hydrodynamics.

The strategy for operating the Old River gates that has so far been evaluated is referred to as "open on ebb operations". These operations are designed to use the strength of the oscillating tidal flows to "pump" water to the north on Old River and prevent the net movement of water from Franks Tract into Old River and towards the pumps. In compensation for any change in net flows on Old

[3]During winter and most of spring the DCC gates are typically closed for protection of outmigrating juvenile salmon so Georgiana Slough becomes the primary diverted flow from Walnut Grove.

River due to the gate operations, the flows on Middle River will necessarily adjust in the opposite amount. The gate operations can thus cause a clockwise recirculation of net flows between the central and south Delta.

The planned manipulations of 2-Gates fall into two categories, those hypothesized to protect adult DSm and those aimed at larval and juvenile DSm. The general idea behind gate manipulations is to keep DSm at all life history stages out of harms way, into areas with a lower probability of entrainment at the water export facilities. The operations in January and February are for adult DSm and are aimed at creating a turbidity barrier of sorts (a zone of relatively low turbidity) in the Old and Middle River flow corridor, with more turbid water to the north and less turbid water to the south. The intent is to keep adults in areas with lower probabilities of entrainment. This ability to control turbidity is constrained by how negative the Old and Middle River flows are and the magnitude of the sediment loads entering the Delta; combined Old and Middle River flows more negative than 2-3000 cfs during periods of especially high sediment inflows to the Delta will make such controls difficult to achieve.

Gate manipulations aimed at larval/juvenile DSm occur in the latter half of March and then again in June. The aims here are to create a hydrodynamic environment that is labeled "dispersive mixing", where by means of opening the Old River gates mostly on the ebb flows (and mostly leaving the Connection Slough gates closed) juveniles are flushed westward from Franks Tract into the confluence area and Suisun Bay. The modeling studies have shown that the success of the dispersive mixing at keeping the larval/juvenile DSm out of harm's way depends a great deal on where they are residing in the Delta. Because Middle River flows become significantly more negative during gate operations, when DSm are located in areas east of Franks Tract, the gate operations can increase entrainment of those fish.

2 Hypotheses

Five hypotheses are listed. The first two are related to flows and turbidity and how gate operations might affect both. The third and fourth are related to adult DSm migration and salvage. And the fifth is related to how one type of gate operation can generate the particular flow patterns that characterize dispersive mixing.

1. H1: 2-Gates Project operations can control net flows in Old River to achieve a predictable balance of flows in both Old and Middle Rivers.
2. H2: 2-Gates Project operations can balance net flows between Old and Middle Rivers, as indicated in H1, to maintain a low turbidity region in Old and Middle Rivers.
3. H3: Migration of pre-spawning adult DSm from the Suisun Bay into the Delta and freshwater habitats occurs when initial winter storm events increase Sacramento River turbidity in the Delta to above a threshold of 12-15 NTU.
4. H4: Maintaining a low turbidity region in Old and Middle Rivers reduces adult delta smelt salvage.
5. H5: Open-on-ebb operations increase dispersive mixing between the south-central Delta and the the lower San Joaquin River through Franks Tract-False River.

3 Data Collection and Analysis Procedures

The interested reader might want to look at Appendix A which describes an ideal setting for analyzing the effects of 2-Gates. Such a scenario can help put in perspective what is proposed and what one would ideally like to know.

3.1 Testing Hypotheses

The procedures for testing each of the 5 hypotheses are summarized below.

1. H1: Two different procedures will be carried out. One procedure is to compare observed flows (measured at fixed-site water quality monitoring sites) to flows predicted by the RMA model (p C36). The other procedure is labeled a BACI experiment and it compares *simulated* flow output from the RMA model between with (Impact) and without (Control) 2-Gates operations.

2. H2: Essentially the same procedures as for H1: comparisons of model predictions with field observations (at fixed-site water quality monitoring sites) and the BACI experiment, but now using turbidity.

3. H3: The procedure is summarized as "A time series at fixed sites. Single field event monitored over a tidal cycle." (p C-33). Based on pp C38-39, the time series appears to include turbidity measurements and trawl-cam counts of Delta smelt (with such counts made at Dekker Island and Jersey Point). It is not clear whether or how the RMA model is used.

4. H4: Essentially the same as for H1, but salvage at the fish facilities is the outcome variable. The field observations will be based on samples of salvage taken at the facilities.

5. H5: And again, much the same as for H1, but the variables are flow and salinity (in False River). (We note that we did not study details given in Attachment A.)

3.2 Data collection, generation, and analysis

Data collection for hypothesis testing is a combination of using existing surveys and monitoring programs (for abiotic and biotic variables) (pp C51-C61), expanding some programs and adding new programs (pp C61-C66). The real data (flow, turbidity, salinity, salvage, DSm numbers) needed for the hypothesis testing will be collected using a combination existing surveys and monitoring programs, extensions of some of these programs, and some new programs.

For H1, H2, H3, H4, and H5 hypothesis testing, simulated data (flow, turbidity, salinity, particles, etc) are generated using the RMA model. Randomness in the outcome variables is generated in various ways; e.g., Attachment E.

4 Critique of Hypotheses and Data Collection and Analysis Procedures

Before evaluating the above specific hypotheses and analysis procedures, we think the project proposal needs to include some general extremely important features. It seems reasonable to have some *a priori* notion of the benefit of the project, and given the possible benefit, to then know how likely one is to

detect that benefit given the inherent natural variability, model uncertainty, and sampling error. Thus we argue that the 2-Gates project proposal needs to include two additional features:

1. Quantitative statements of the anticipated benefit to the Delta smelt population. The benefits, or lack of harm, to other listed species needs similar quantification.
2. Estimates of the time it would take to determine, with some degree of confidence and under different levels of sampling effort and natural variation, whether or not the anticipated benefit has occurred, or the time needed to estimate the effects with a specified precision. Relatedly, if it turns out that the effects are negative, then the time to detect such effects needs to be as short as possible to avoid further damage to the resource.

Given the ESA status of most of the species, we think the project needs to show a beneficial effect at a certain minimum level and that the precision of the estimated effect needs to be at a certain level as well (both levels to be decided by policy-makers). Appendix B provides more discussion and technical details regarding how one might provide such quantitative assessments.

4.1 Hypotheses

Imprecision of hypotheses. Each of the hypotheses suffers from some degree of ambiguity.

- H1: How is a "predictable balance of flows" defined? What does this mean? There is a complex interplay between the current state of the system, the actions taken (gate operations), and the resulting state of the system. Predictability refers to the RMA model: how close are its predictions of the future state of the system (given the model inputs, which includes gate operations) to what actually occurs? What actually occurs is a stochastic process, a space (s) and time (t) indexed random variable and the model ideally should produce a distribution of flow values, $f_{s,t}$. For p percent of the time, the actual flows should fall in the middle p percent of the prediction distribution.
- H2: What is meant by a "low turbidity region" in Old and Middle Rivers? Ideally the hypothesis would specify a range of absolute turbidity values by location and time, $Tu_{s,t}$, given a particular set of current state values (current flows, turbidity, etc) and particular sequence of gate operations. Less ideal would be a range of relative turbidity values; e.g., the turbidity in some regions is 20% of the turbidity in other regions. An alternative restatement would be that 2-Gates could balance net flows and maintain a low turbidity zone *more effectively than OCAP operations alone.*
- H3: Migration of pre-spawning adult DSm could be defined more sharply; e.g., the movement of at least p% of the Suisun Bay population to specific locations in the Delta, or specific salinity zones. Does this hypothesis imply that migration will not occur otherwise? It is possible that turbidity would not rise above 12-15 NTUs during the months of December through February during a particularly dry winter—then many or most of the fish might migrate without turbidity cues.
- H4: Again (re: H2) one needs to define low turbidity region (turbidity values and exact locations), and to quantify what is meant by reduction in adult DSm salvage (in percentage terms?). One possible rephrasing is "Through 2-Gates operations maintaining an equal or lower turbidity region over a longer period of time in Old and Middle Rivers reduces adult delta smelt salvage over OCAP operations alone." The degree of reduction in salvage would still need to be stated.

- H5: How exactly is dispersive mixing defined (between south-central Delta and the lower San Joaquin River through Franks Tract-False River) and how is an increase (in mixing) defined?

Missing hypotheses. The primary concerns of natural resource agencies like USFWS, NOAA Fisheries, and CDFG center on the effects of 2-Gates on fish and wildlife. However, only two of the five hypotheses involve biotic variables, in particular just Delta Smelt. Of equal concern is a lack of hypotheses about other listed species and other life history stages. Issues for other species are listed in Section 5, but the issues have not been framed as particular hypotheses to test and/or effects to estimate.

- Larval and Juvenile DSm: There are no hypotheses about effects on Larval and Juvenile DSm. Perhaps H5 can be viewed as indirectly related to juvenile DSm if one views the salmon as non-volitional entities that move much the same as neutrally buoyant particles in the water (and where salinity affects survival perhaps). I The question being begged is "How well do the physical variables of flow and salinity serve as proxies for larval and juvenile DSm?". In this regard, two other hypotheses could be posed:
  - H6: Increased dispersive mixing reduces entrainment (and salvage) of larval/juvenile DSm by p% or greater.
  - H7: The greater the dispersive mixing (to the maximum that 2-Gates can effect), the greater the reduction in larval/juvenile entrainment (regardless of Delta inflow or DSM larval/juvenile distribution).
- Hypotheses (or effects) that are more directly related to DSm population levels and dynamics are lacking as well. In particular hypotheses about changes in survival probabilities by particular life history stages and times and locations, about fecundity levels or reproductive success, about probability of entrainment, and about movement probabilities would all be more directly relevant to each species longterm success and viability.
- Salmonids: Hypotheses about project effects on out-migrating steelhead, winter run Chinook salmon, and spring run Chinook salmon are needed. There is particular concern that San Joaquin origin steelhead, and also non-listed fall run Chinook salmon, may be blocked during out-migration by gate closures. Winter run and spring run Chinook salmon from the Sacramento River may be drawn in the project area (via Georgiana Slough, for example) and be impacted. There is concern that adult salmonids, particularly steelhead, may be affected by the project in their fall return migration. Hypotheses about changes in predation on juvenile salmonids in particular reaches, e.g. in the proximity of the gates, before and after gate installation would be of interest.
- Green Sturgeon: Hypotheses about effects on green sturgeon are needed. Little is know about green sturgeon movement in the delta either as juveniles or adults. In particular it has been observed that juvenile sturgeon, ages 1-3, 200-750 mm in length end up in salvage at the SWP and CVP fish facilities. Adults are also caught by fisherman in the delta.
- Longfin Smelt: Hypotheses about effects on longfin smelt are needed. For example, H_o: longfin smelt larval entrainment at the south Delta export pumps is not changed based on 2-Gates operation Jan-Mar as compared to OCAP, versus H_a: longfin smelt entrainment is reduced [this should be a 1 tailed test; we want to see improvement because we believe that as many or more larvae may be swirling around the south Delta, potentially facing increased predation].

- Other species: While striped bass is an introduced species, given its recreational value, predicted effects on them are of interest.

4.2 Data collection and analysis

Data collection. Regarding the collection of field data (of DSm and hydrology variables):

1. The absence of temporally and spatially proximal sampling of small Delta smelt larvae during the March gate operations is a serious hole in the data and quite worrisome. The fact that operations are taking place over an extended period and that little to no information on the potential harm is being gathered allows for the possibility of serious harm to Delta smelt.

2. Data would need to be gathered for the additional hypotheses, H6 and H7, that we have proposed.

3. Adequacy of the spatial coverage of the sampling (sampling locations) and sampling frequency in terms of achieving desired levels of precision needs to be examined (see Appendix B.2 for a way to do this). (A particular concern discussed by the authors was whether sampling for juvenile DSm in the Mokelumne River would be adequate.) We also recognize that the efficiency of the salvage facilities is several orders of magnitude greater than any net sampling approaches—methods for effectively combining data from multiple sources deserve additional thought.

4. Discussion is needed as to what would be done when "take" limits on a species is reached, and further sampling is prohibited.

5. Relatedly, the trawl cam is an unproven experimental sampling device. Whether it has undesireable impacts on species needs to be considered.

6. Costs of additional sampling effort needs to be summarized (in a table).

7. Gear efficiency, or in general, the probability of capturing fish given that they are present deserves study (this is true of many existing studies, as well).

Regarding sampling for other fish species (again associated costs need to be provided):

1. For green sturgeon: Permanent installation of DIDSON could help understand the impact of gate operations on both adults and juveniles. Increased acoustic tagging would also be useful.

2. Details on acoustic tagging for salmonids (both from the Sacramento and the San Joaquin systems) are needed. Exactly what parameters would be estimated, desired precision of estimates, approximate number of tags and receivers would be useful information.

Soundness of statistical testing procedures. The BACI analysis of RMA simulated data that forms part of the testing for 4 of the 5 hypotheses (H1, H2, H4 and H5) does not seem in fact to be a BACI analysis. Simulation during the Before period for the Control and Impact settings appears to be done under identical conditions and the output should not differ on average; and as much seems to be admitted in places (e.g., p C40, line 3: "... should then show similar mean values for the Control and Impact runs in the Before period"). Hence there really is no "Before" to the analysis. Regarding the comparison of runs in the After period, differences in output (be it for flows, turbidity, or salinity) are expected and the magnitude of the differences can be quantified by comparing probability distributions. There needs to be no statistical testing in this case—one can simply report the degree of difference

or overlap in the distributions. Whether this has any scientific merit, however, depends on the other half of the "testing" procedure, i.e., the validation of model predictions using field data. We note also that the analysis of simulated data alone could presumably be done now, and having the results in the Project proposal would be quite useful.

Comparisons of model predictions to field-based observations are extremely sensible. Explicit statements of the underlying null hypotheses being tested are lacking, however. One could hypothesize that reality corresponds to model predictions, i.e., that is H_o, and the alternative is that reality is different (H_a), and various goodness-of-fit testing procedures could be used. However, we are certain that H_o is not true *a priori*, and it may be better to specify what degree and type of differences between model predictions and reality are acceptable. For example, one might state that if the correlation between predicted flow and observed flow (for a set of locations and a set of times) is at least 0.80, then that is acceptable (noting that correlation would just describe the degree of association not the size of absolute errors). Or one might say that if the average absolute errors are no larger than some particular size, then the model is acceptable.

For testing hypothesis H5, as mentioned previously, whether comparisons of model output of turbidity and flows with actual observations has any relevance to what is happening to larval and juvenile DSm is an unanswered question. Testing the suggested hypotheses H6 and H7 would provide some information about relevance.

In Appendix C we contrast the general testing approach considered in the proposal (validating the accuracy of a predictive model by making comparisons with field observations) with alternative "classical" experiment designs. We do so to make clear that we recognize the inherent difficulties of testing the effects of 2-Gates, and to make clear that some degree of model-driven assessment seems necessary.

4.3 Adaptive Management

Due to time limitations we are giving short shrift to the notion of adaptive management (A.M.) for the 2-Gates Project, but a few comments can be made. Our evaluation is made using the DOI Technical Guide on AM (Williams, Szaro, and Shapiro, 2007) as a reference document. By definition AM "is a systematic approach to improving resource management by learning from management outcomes" (Williams et al. 2007). The two phases of AM are the Set-up Phase followed by the Iterative Phase.

The Set-up Phase begins by bringing together stakeholders, which in this case would include water users and contractors, natural resource agencies, and environmental NGOs. The stakeholders then work together to identify "clear, measurable and agree-upon management objectives"; in this case protecting and rebuilding the populations of listed species might be an objective. Third, "identify a set of potential management actions for decision making"; this would include not installing 2-Gates, installing 2-Gates and operating it one particular way, installing 2-Gates and operating it another way, and so on. Fourth, "identify models that characterize different ideas (hypotheses) about how the system works"; project proponents have postulated models and hypotheses, but they need to be both sharpened and expanded upon, and competing models and hypotheses from other stakeholders need to be formally written. Fifth, "design and implement a monitoring plan to track resource status and other key resource attributes"; this, we think, is far from complete.

The Iterative Phase has three steps: (*Decision Making*) "Select management actions based on management objectives, resource conditions, and understanding", (*Monitoring*) "Use monitoring to track system responses to management actions", and (*Assessment*) "Improve understanding of resource dynamics by comparing predicted and observed changes to resource status". The 2-Gates proposal

does contain elements of all three steps. Re: Decision Making—actions include gate operations aiming to protect adult Delta smelt and then operations aiming to protect larval and juvenile Delta smelt. How gate operations might change with different flow levels, with different export levels, with different *a priori* estimates of abundances of different species, by area, however, would benefit from more thought. Re: Monitoring—many ideas have been put forth, using Spring Kodiak trawl data for adults, using 20 mm survey and salvage data, to name a few. We remain quite concerned about holes in the data collection and monitoring process for Delta smelt during the small larval stage and some ambiguity in the monitoring process for other listed species. Re: Assessment—the plan to compare field data, both abiotic and biotic, with RMA model predictions, and then to adjust, improve, and update parameters in the RMA model is stated repeatedly. Details on how to quantitatively assess the actual benefits for listed species populations, including Delta smelt, however, are lacking.

Appendices

A Idealized Scenarios

An idealized scenario for analyzing the effects of 2-Gates is described. Such a scenario can serve as a basis for evaluation of the proposed data collection and analysis procedures. There are many dynamic spatial-temporal processes, biotic and abiotic, defined on the Delta. For the purposes of this report the geographical region of interest is just the waterways of the Delta, not the land. The biotic processes considered here are Delta Smelt, Salmon (distinguished perhaps by run and origin), Green Sturgeon, Steelhead, and Longfin Smelt. The abiotic processes are water volume, flow (amount and direction), turbidity, salinity, and temperature.

In the idealized scenario the value of any variable at any location at any given time is known, i.e., there is perfect knowledge of the state of the system. In addition, one would have stochastic models, systems of equations, which could predict the future state of the system conditional on the current state and on management actions and other inputs (e.g., a rainstorm) and quantify the degree of uncertainty about these states.

With regard to the water variables, at any given location s and time t one would know the volume ($v_{s,t}$), flow ($f_{s,t}$), turbidity ($Tu_{s,t}$), salinity ($S_{s,t}$), and temperature ($Te_{s,t}$). Further, one would also know the probability distributions of the values of these variables at given later point in time, conditional on the current values and other relevant inputs. The probability distributions would be defined in terms of functional relationships with other variables; e.g., precipitation, air temperatures, export levels, position of Delta Cross-Channel gates, etc. In other words one could calculate expected future values along with prediction intervals.

With regard to the fish, one would know the number of fish, $n_{k,s,t}$, of each species k (perhaps distinguished by life history stage or age) at any given spatial location s at any given time t. Furthermore, one would know, per species, location, and time, the probability of survival ($\phi_{k,s,t}$), the probability of movement ($m_{s,t\to s',t'}$) to any given location s' by time t' ($t' > t$), and the probability of producing progeny by time $t+1$. These probabilities would be defined in terms of functional relationships with variables, abiotic and/or biotic; e.g., the probability of an adult Delta Smelt moving from a particular location on Old River on January 12 is a function of flow, turbidity, and salinity at that location and in surrounding locations. Furthermore there would be explicit linkages from one life history stage to the next, and from one cohort to the next.

The biological and hydrological theories that the Project is based on can be viewed as particular explanatory and predictive models about water variables and fish (particularly Delta smelt). For example, closing of the gates during January and February is predicted to cause turbidity levels in locations "relatively" near, and south of, the gates, in Old and Middle Rivers, to be lower than would be observed with gates open (and absent). This is the gist of H2 (and H1), that gate operations can maintain a low turbidity zone in Old and Middle Rivers. Ideally one would have a predictive model that would take the current state of the system ($v_{s,t}$, $f_{s,t}$, $Tu_{s,t}$, etc) and relevant future inputs (gate operations, export levels, dam releases, precipitation) and predict the state of the system over a sequence of days, in particular predict $Tu_{s,t'}$ for locations in the proximity of the gates.

We also note that besides the above "process" model, one can formulate various observational models that would reflect how data are generated from sampling the output from the process model. This would be an example of a hierarchical model and it would be useful for assessing how different sampling schemes and intensities might affect the quality of estimates (e.g., Appendix B). One ad-

vantage of the hierarchical framework is that estimation of unknown parameters of the process model based upon observed data can be integrated with the process model.

To adequately and fairly contrast an idealized scenario, particularly an idealized hierarchical model, with the Project proposal requires a more careful study of the RMA model, including examination of the spatial-temporal resolution of the modeling of abiotic and biotic variables. One point that can be made is that the RMA model does not contain a full life history model in that there is no between year linkage in the abundances, by time and area. Each run of the model is essentially a fresh start and the "knock-on" consequences of year-specific survival and birth rates are not accounted for in the simulation output. So, for example, the consequences of a particularly disastrous mortality event in year t is not being predicted or examined for year t+1. Another point is that linkage between historical survey data and the parameter values used in the RMA model is not readily apparent; i.e., the degree to which historical data were used to specify model parameters is unknown (but in fairness to the proponents such details may lie elsewhere in the proposal).

B Quantitative assessment of 2-Gates Benefits

We recommend two additions to the proposal.

1. Quantitative statements of the anticipated benefit to the Delta smelt population. The benefits, or lack of harm, to other listed species need similar quantification.

2. Estimates of the time necessary to determine, with some degree of confidence and under different levels of sampling effort and natural variation, whether or not the anticipated benefit has occurred, or the time needed to estimate the effects with a specified precision. If it turns out that the effects are negative, then the time to detect such effects needs to be as short as possible to avoid further damage to the resource.

B.1 Quantifying the Anticipated Benefit

Quantifying the anticipated benefit, prior to conducting an experiment, could be done several different ways. Here we just discuss one approach and that is to quantify spatially- and temporally-local effects via simulation. For example, to quantify the benefit of gate operations aimed at adult delta smelt, suppose that all delta smelt have migrated upstream (denoting the total by N_{t_0}), and spawning is about to begin. Partition the area occupied by the smelt into K regions. The N_{t_0} total fish have distributed themselves in a particular way, perhaps based upon initial conditions related to turbidity (with fraction p_i in region i, $i = 1, 2, \ldots, K$). Historical data could be used to generate some data-based estimates of the fractions p_i. Given these initial conditions and a particular set of gate operations, the RMA model could be used to generate the predicted number[4] surviving to a specific time period t_1, say, is N_{t_1}. Thus the expected survival probability between time period t_0 and time t_1 for a particular set of gate operations is:

$$S_G = \frac{N_{t_1}}{N_{t_0}}$$

[4]We recognize that the RMA model's particle tracking simulation will not likely accurately characterize changes in "natural mortality" associated with the gates themselves and new circulation patterns in south Delta—this approach just gives some approximate measures of effect.

Letting S_{NG} be the expected survival without 2-Gates, differences in the two expected survival probabilities could serve as measures of the benefits of 2-Gates, say B:

$$B = S_G - S_{NG}$$

Comments.

- Given the influence of initial conditions, a range of initial conditions could be systematically explored and the resulting range of benefits predicted. We note that this same procedure could be carried out for a third case, where the gates are installed but are locked open.

- Related to the previous comment, if the initial spatial allocation of fish does not match well with reality, the predicted benefit could be badly biased. For example, suppose that there are just 2 regions and 2 time periods. For each region there are 2 parameters, the fraction of the population in that region (p_1 and 1-p_1), and the probability of entrainment for that region (say e_1 and e_2). Starting with N_{t_0} fish, the predicted number surviving, N_{t_1}, (not entrained, ignoring other sources of mortality) is

  $$N_{t_1} = N_{t_0} * p_1 * (1 - e_1) + N * (1 - p_1) * (1 - e_2).$$

  The degree of error in these predictions is a key concern. Suppose N_{t_0}=1000 and that the true parameter values are $p_1 = 0.7$, $e_1 = 0.6$, and $e_2 = 0.1$, then the number surviving is

  $$N_{t_1} = 1000 * 0.7 * 0.4 + 1000 * 0.3 * 0.9 = 550$$

  for an overall survival of 55%.

  Suppose that the RMA model has the right e_1 and e_2 values but wrong spatial allocation parameters, namely it is exactly the opposite of the true situation, $p_1 = 0.3$. Now

  $$N_{t_1} = 1000 * 0.3 * 0.4 + 1000 * 0.7 * 0.9 = 750$$

  for an overall survival of 75%.

- We note that what is posed (incorrectly we think) as BACI tests for four of the five hypotheses is similar to what is proposed here. Run the RMA model both with and without 2-Gates *with the same initial conditions* and examine the subsequent predictions of abundances by time and area, and in the aggregate. Differences between the two sets of simulation runs would provide measures of predicted benefits, but as above the degree of accuracy of initial regional allocation and accuracy of probability of entrainment per region are crucial to projections of benefit.

- Lastly, we recommend producing visual displays of predicted abundances by area over time, as these could serve as a useful tool for understanding the dynamics of initial spatial allocation, survival, and movement.

B.2 Estimating and Detecting the Benefit

The primary question is "How long will it take to estimate the 2-Gates benefit with a desired degree of precision, given specific sampling procedures and levels of sampling, and a particular magnitude of benefit or size of "signal?".

To begin one needs to consider in some detail (a) how the benefit would be estimated in a given year (or season, say), (b) the precision of that estimate, and (c) the degree of between-year variation (environmental variation) there is in the benefit. Regarding estimating the benefit for a given year, the sampling procedure and the probability of capture by the trawl would first be specified. For example, stratified random samples could be taken at the beginning of the time period (t_0) and end of the time period (t_1) and the benefit estimated from the sample recoveries. For simplicity, suppose 2 samples are taken per region, the same volume of water (v) is sampled per trawl, and that the probability of capture within a sampled volume is 100%. Letting $y_{p,i,j}$ denote the number of fish caught in time period p (either t_1 or t_2) in region i and sample j (j=1 or 2). Then an estimate of abundance at each time period is a stratified random sample ratio estimate:

$$\widehat{N}_p = \sum_{i=1}^{K} V_i r_i$$

where

$$r_i = \frac{y_{p,i,1} + y_{p,i,2}}{2v}$$

Survival could be estimated:

$$\widehat{S} = \frac{\widehat{N}_{t_1}}{\widehat{N}_{t_0}}$$

Repeat this process for two scenarios, one with gates and one without, and the estimate of the benefit is simply:

$$\widehat{B} = \widehat{S}_G - \widehat{S}_{NG}$$

Taking a minimum of two samples per region allows variances to be calculated, thus the precision of the benefit can be calculated. If the resulting estimate of precision, as measured by standard error of $\widehat{B}$ or by a 95% confidence interval width, say, is not considered adequate, one could keep increasing sample sizes, for example, until the desired precision was achieved. Different simulations of the spatial distribution of delta smelt could be carried out to try to get a reasonable picture of the sampling uncertainty due to clustering or low densities.

The whole process would need to be repeated for multiple years to quantify the degree of natural variation and the resulting effect on the ability to estimate the benefits.

We note also that the historical sampling data, from Fall Midwater Trawl, Spring Kodiak Trawl, 20 mm survey, and from salvage sampling could be used to advantage in both checking the approximate quality of simulated abundances (cf B.1) and in determining necessary sample sizes to achieve specified levels of precision. The paucity of recoveries of Delta smelt, particularly in the central Delta, in recent years, many zeros with occasional large recoveries, will increase the variance of estimates. With such low population levels the potential "take" impact might drastically constrain the desired sampling intensity, and subsequent precision.

Another practical complication is that as fish begin to spawn, they change behavior and no longer distribute primarily near the surface. Detecting differences in estimates due to mortality versus those due to males and some females hanging out near good spawning substrate would be difficult.

C Alternative experiment designs and analyses.

The purpose of this section is to show that some degree of model-based assessment is required to estimate the effects of 2-Gates. We begin with an idealized classical experiment design setting where one has n experimental units (say "years") and randomly assigns the "treatment" (2-Gates installed and operating) to n_1 of the n units, and the remaining $n_2 = n - n_1$ units get the "control" (no 2-Gates). The response variables, y's are measured, and the difference in averages, $\overline{y}_1 - \overline{y}_2$, is an estimate of the treatment effect, and something like a t-test can be used to test hypotheses about the underlying true longterm difference in means. The points we wish to emphasize here are (1) *there is no explanatory or mechanistic aspect inherent to this approach*, one can simply report there is an estimated effect of size of $\overline{y}_1 - \overline{y}_2 \pm 2$ standard errors, say, and (2) *due to random assignment any factors that influence the outcome, other than the treatment, have been controlled for*, any idiosyncratic units which might have peculiar responses are just as likely to receive the treatment as the control. These features are very attractive—estimates of the treatment effect are thus unbiased, and "significant" differences can be detected without necessarily needing to explain exactly why they are exist. A scientist will want to understand why, but knowing that there is strong evidence that differences do exist is a reassuring base to work from.

With 2-Gates, the treatment (2-Gates) cannot practically be applied at random and the design is necessarily Before-After, with the control (no 2-Gates) coming first, and 2-Gates following. Ideally environmental conditions in the Before and After periods would, on average, be identical, and effects in any given year would have no affect on other years, i.e., the experimental units are independent. Neither condition will likely hold—the environment of the Delta appears to be in systematic decline (as evidenced by the Pelagic Organism Decline) and with a localized population what happens in one year will affect the population the next year. One response variable, a y, could be the Fall Midwater Trawl index for Delta smelt, which is an approximate population level measure. One could compare $\overline{y}_{Before}$ to $\overline{y}_{After}$, even trying something as questionable as a t-test to attempt to estimate the 2-Gates effect at a population level. Systematic differences in environmental conditions, dependence between observations, and unequal sampling error would make such an estimate of treatment effect, and such a t-test, suspect, however.

Two approaches to dealing with these problems are (i) to incorporate covariates in the analysis, attempting to control for differences in the environmental conditions, and (ii) to use a BACI design. With a BACI design, one would have a second set of experimental units that is measured during the Before and After periods which will always receive the control, while the first set of units is control Before, and treatment After. The second set is "controlling" for systematic environmental differences assuming that those differences are the same for both sets. The problem with Delta smelt is that it is a population endemic to the Delta and one does not have some second population existing in some parallel Delta that could be used as the Control. Whatever is done somewhere at sometime to part of the population will have effects on all the population, at least at some later time.

Given these difficulties, some sort of model-based approach to estimating the effects of 2-Gates appears necessary. Even if one had perfect information about the status of the population (as in the Idealized Scenario, Appendix A), numbers by time and location and life history stage, problems of changes in background environmental conditions require some sort of model-based adjustment.