Development of Sediment Quality Objectives for California Bays and Estuaries

Scientific Steering Committee Meeting
February 28 – March 2, 2006
Southern California Coastal Water Research Project

Notes from Closed Session:

General Comments: Overall, the SSC was very impressed with the progress made on this project, and we wish to reaffirm that the Committee thinks that the general MLOE approach being proposed for the Sediment Quality Objectives is both appropriate and complete for performing comprehensive assessments of sediment quality. We have some specific comments to offer relative to each LOE and the general framework.

Chemistry LOE

Concern remains that, because this is intended to be an effects-based set of SQOs, this LOE is not completely independent but based necessarily on toxicity and benthic responses. This LOE needs to have some connection to measures of biological effects so that bioavailability issues and exposure-response relationships can be properly addressed. It may be possible to develop ambient/background concentrations as supplemental information to address the lack of independence in the chemistry LOE and to provide a reality check on concentrations of some trace metals and organic compounds. Some metals (e.g., mercury in San Francisco Bay) and some organic compounds (e.g., PAHs in relatively inert soot or fly ash) may occur in sediments as a result of natural processes. The effects of regional geology on chemistry should not be ignored and effects-based standards that are lower than background geological levels for metals should be avoided.

All subsequent documentation of the chemistry LOE should explain and recognize that potentially toxic chemicals in nature always occur as mixtures, the chemical composition of which varies significantly among samples, coves, bays, estuaries, and regions. As a consequence of this variability, relatively poor correlations between measures of individual chemical exposure and biological response as observed by the Science Team on a statewide or regional basis are to be expected. Therefore, it becomes very important to develop SQOs that account for exposures to mixtures of varying composition, as opposed to individual chemicals. The various sets of candidate SQO approaches that the Science Team is evaluating correctly accounts for the presence of chemical mixtures. However, it is important that the documents produced by the Science Team recognize both the strengths and weaknesses of standards based on multi-chemical indices as recently summarized (Long et al 2006).

- There was no recommendation to change the current Chemistry LOE strategy at this time, with the exception of the lack of complete independence as stated above. The SSC encourages efforts to develop improved methods for this LOE that more directly address bioavailability/exposure and to develop a more independent LOE. However, the SSC recognizes that this effort would require substantially more time and resources and should

not be performed at this time; although, regional geology should be considered when evaluating stations as previously stated.

- The current validation approach used a subset of data with an even distribution, which may not reflect the real world distribution of sediments. The Science Team should provide additional documentation/validation using a dataset that is representative of actual data distributions to verify that validation results are accurate when applied to real world data. Verbal discussion suggested that this has already been conducted and would be supplied.
- Validation of the Benthic SQG should be based on final benthic LOE categories to verify the selection of the best SQG.
- The SQG should consider use of different weights for selection of thresholds. 'Low' and 'High' thresholds may be more important than the 'Moderate' threshold. Threshold selection should consider end user performance objectives, if it doesn't result in clearly ineffective SQG values (e.g., high efficiency for non-toxicity or non-impaired benthos).
- The incorporation of sub-SQGs based on toxicity and benthos is a good idea, and the Science team should continue to use both strategies for this LOE.
- There is a need for more analysis to support the choice of specific SQG combinations to use. The use of regional versions of the SQGs is acceptable in principal, but has to be thoroughly justified to demonstrate that it is needed on the basis of substantially improved performance relative to statewide versions.
- EqP benchmarks are qualitatively different from empirically derived SQGs, in that they address a different question. For example, they should be used for supplemental diagnostic information that will lead to more causal insights for a station.
- Although organic carbon normalization of the data has not been shown to improve exposure-response relationships with these data sets, it is important to continue to collect, whenever possible, organic carbon measurements and other sediment characteristics that affect bioavailability and may be of subsequent diagnostic value (e.g., acid volatile sulfides, black carbon).

Toxicity LOE

The overall logic in the development of the toxicity LOE is robust. A thorough job was done and an impressive amount of new information has been assembled. The incorporation of multiple tests with both acute and sublethal endpoints is encouraged. However, subsequent documentation of this LOE would be strengthened if citations were included of previous comparisons in the performance of various estuarine toxicity tests. Any new reports also should include either documentation of or a rationale for the toxicological and ecological relevance of each candidate test.

- Revise the definition for the 'Low' toxicity threshold. This definition should include results that are significantly different from the control, but have a response that does not exceed the minimum significant difference (MSD). As currently defined in the Toxicity Indicators report (pg 18), nontoxic sediments should be those that are not different from the negative performance controls.
- Provide an expanded description of the narrative intent of each toxicity category. What are they intended to represent?

- Redefine the high threshold to use the mean of 99th MSD and 75th toxic distribution. The Science Team should consider omitting the use of the double dose value as it is not as robust and the foundation for the approach is not as good as that for the other two approaches.
- Other test species/methods: The approach should not discourage the use/development of additional lethal or sublethal test methods. New tests should be allowed as the 3rd or 4th tests in the toxicity LOE, provided adequate data are available to develop thresholds in a manner consistent with the program.
- The test advocate has the responsibility to justify the incorporation of additional tests, which must meet the objectives of the program and key characteristics of desired tests for the program. The Science Team should establish required characteristics, such as the toxicological and ecological relevance of each candidate test and the appropriate sample size requirements for an acceptable MSD calculation.
- At this point, the use of pore water tests is not recommended for the toxicity LOE. Other applications of these tests (e.g., TIE) may be useful.
- The SSC supports the addition of *Rhepoxynius abronius* as a primary test option, provided that sensitivity to grain size is addressed and does not compromise the results.
- The SSC is divided on the inclusion of *Ampelisca abdita* as an optional test species. This species has been frequently used and has proven quite useful in many studies, but some labs have had difficulty performing tests with it. It is not always readily available year-round, and it has shown a lack of sensitivity to moderately contaminated sediments, especially in San Francisco Bay and Puget Sound. Additional side-by-side comparisons in the performance of this test are encouraged to evaluate future inclusion of it by the State.
- Additional information should be developed and communicated concerning the use of *Leptocheirus* and *Neanthes* with regard to their geographic distribution and historical use in California.
- Use of the median for test data integration is supported. The strategy of selecting the most protective category (i.e., higher category) in the case of indeterminate median is also supported.

Benthic LOE

The SSC approved of using the median of the combination of the three indices to set the values for the benthic community LOE. The three indices measure slightly different attributes of the benthos, so they are not duplicative. The SSC was concerned with the decision to drop BQI as an index and requested more information/justification for this decision as this index appeared to be among the best performing with the data presented.

- The Science Team should provide additional documentation that shows the accuracy of the recommended combination of indices relative to other combinations; this information is needed to justify selection of the preferred combination for use in this LOE.
- Documentation of the rationale for exclusion of the BQI and RIVPACS indices is needed. The BQI has theoretical concerns, and the RIVPACS has lack of support for development.
- The Science Team should identify needs for future research for benthic LOE development (e.g., benthic community signatures).

Direct Effects Multiple Lines of Evidence Framework

The SSC favors the use of the alternative approach for combining the lines of evidence as it addressed previous concerns of the SSC to ensure that sites were identified as impacted, whether or not the cause was from contaminants.

- The SSC in general found that greater weighting for benthos was appropriate in principle, but the Science Team needs to confirm that this strategy is consistent with State regulations (i.e., toxicity does not have to be weighted equally in order to comply with regulations). However, there was concern that such weighting could lead to eventual dropping of one of the other LOE. This would not be appropriate as there is information that is captured in the chemistry and toxicity LOEs that may be missed by the benthic LOE. The strength in the overall MLOE approach lies in reliance upon a weight of evidence to classify sediments based on information from all three effects-based LOEs.
- There was a desire to insure inclusion of the classification of "inconclusive" for some of the combinations of data where the data do not make sense and are conflicting. Some SSC members suggest alternate classifications for some LOE combinations on a philosophical basis. The boxes where such a condition exists would be at the corners where toxicity and benthos disagree and where chemistry is high but toxicity is low. The other box where chemistry is low and toxicity is high should be rated at least moderate to cover the condition where unmeasured chemicals could be an issue. Presence of these extreme combinations **must** be flagged and identified as in need of further investigation/review of raw data. The SSC would like the Science Team to conduct analysis that compares alternate framework results compared to typical two LOE assessments (Chemistry and Toxicity). At the end of the day, a comparison was provided and feed back on individual boxes was to be provided to S. Weisberg for further analysis.
- It was encouraging to see such good agreement among the panel of six experts, given their differing professional experiences. It was also encouraging to see that several elected to invoke the "inconclusive" classification. This supports the need for inclusion of such a classification.

Indirect Effects (IE) framework:

- Substantial progress has been made in this framework since the last SSC meeting, however, it is not as well developed as the direct effects frameworks and the IE methods will not likely be ready for application by the anticipated deadlines.
- The IE document does not explicitly state many of the key assumptions in the framework. They need to be explicitly stated before the SSC can review and endorse the program. It is important to understand the sources and magnitude of uncertainties involved with this framework. The report should provide a rationale for selecting only nonpolar organic compounds and mercury as chemicals of concern. The term "wildlife" should be defined to establish the limits of the applications of these SQOs.
- The SSC has adequate risk assessment expertise to review the products, but there is concern that there has been insufficient involvement of risk assessors as members of the Science Team. In addition to providing support to the science team in the development of the LOE, the presence of a risk assessor on the team will demonstrate that there has been adequate consideration of the science and assumptions required in such analysis.

- Clarification: The SQO program is not intended to develop a new risk assessment (RA) program for the State of California rather, it is primarily to incorporate existing State RA policies into the SQO policy. The purpose is to develop the connection between sediment chemistry and existing risk thresholds using the risk thresholds to set the limits for indirect SQO values.
- The Science Team should factor the uncertainties associated with key parameters into their analysis. The approach to address these undertainties within the analysis should be clearly documented within the resulting guidance and products.
- The SSC suggests that the sequence of evaluation be altered so that Step 2 should be a chemistry threshold comparison. Bioaccumulation evaluations should be conducted only if existing data are not available to assess bioavailability from sediments and the sediment concentrations exceed the threshold. For well understood contaminants for which a great deal of data exists (e.g., PCBs, DDTs), a water body, embayment or statewide demonstration of bioavailability is sufficient. In many cases, it will be possible to make such a demonstration through reference to published literature. A correlation test is unnecessary and may be inappropriate. A straight forward demonstration that bioavailability is present for some sites in the water body is sufficient. For contaminants for which little is known, or little data exist, a demonstration of the contaminant's bioavailability relationships will be needed.
- Empirically-based BAFs can be biased high if tissue contamination is driven by sources in addition to sediments. The SSC is concerned about relying just on empirical BAFs and a mechanistic model should be used to verify the adequacy of the empirical BAFs. If extreme differences occur between the model and the empirical values careful consideration of the cause should be evaluated and the choice of BAF selected carefully. Failure to consider alternative sources or complications to the BAF relationship may drive sediment thresholds to indicate more of a problem than exists.
- Additional clarification is needed regarding how the tissue data will be used in the assessment: are fish and shellfish used separately, are individual species or composites used? The details of the approach need to be spelled out.
- Care should be taken in describing the BAF calculation so that it is clearly evident what proportion of bioaccumulation is due to sediment.
- The SSC observed that selecting input values for models and risk calculations to ensure a protective conclusion about risk is acceptable practice. However, the Science Team should consider the role of compounding uncertainties in deriving thresholds. Selecting the 95 percentile from distributions for multiple inputs to ensure protection will result in a level of protection more stringent than would be required to protect the 95 fractile of a population. This compounding can result in extremely unrealistic thresholds for sediment. The influence of these choices should be evaluated through performing sensitivity analysis and by evaluating how realistic or unrealistic the conditions are that result in exceeding a derived sediment threshold.
- The process should allow use of additional bioaccumulation test species, not just *Macoma balthica* although the addition of other species needs to be supported with data that demonstrates that the species is appropriate. The Science Team needs to specify the characteristics of the data required to allow additional species.

- Use of field-derived bioaccumulation data, if adequately designed, could be an appropriate substitute for laboratory studies, but care is needed to ensure issues such as gut clearance and route of exposure are considered.
- A discontinuity exists between optimum species selection for tissue evaluation, BAF calculation, and the importance of the species as a prey item. For example, concentrations in salmon and striped bass are important for human health exposure but would not be the optimal species used for determining the BAF. Clear connections between the BAF and the food web, human or wildlife, needs to be described as part of the process.
- Development of a list of standard species for use in the program is suggested as it would simplify analysis and ensure development of an internally consistent database. This standard set could take the form of generic food webs for geographic regions of the state.
- The SSC recommends that the procedure explicitly state that the combined effects of multiple contaminants are not considered in this framework. The implications of this choice should be considered in the program.

Plans for next meeting

- June is a bad month for many and a date of July 10-14 was suggested.
- The purpose of the meeting will be to review the technical and guidance documents produced by the science team, and additional issues raised by committees. Documents for review should be submitted to the SSC a minimum of one month in advance of any review meeting. The Science Team reported that this much lead time will not be possible for some of the documents. Without such lead time the level of review the SSC could perform would likely be limited.
- The SSC requested a formal charge for the next meeting. Discussions led to the following charge for the next meeting and document review: 1) Is the framework as a whole good (MLOE and assessment plan/indicators); 2) Are indicator selections appropriate?; 3) Are thresholds reasonable?; 4) Is the overall program scientifically sound and consistent with current practices?

References:

Long, E. R., C. G. Ingersoll, D. D. MacDonald. 2006. Calculation and uses of mean sediment quality guideline quotients: A critical review. *Environmental Science and Technology*. 40 (6): 1726-1736.