11/19/07 Public Hearing
Enclosed Bay/Estuaries-SQO
Deadline: 11/30/07 by 12 p.m.

INDUSTRIAL ENVIRONMENTAL ASSOCIATION

Leaders of Environmental Responsibility

VIA EMAIL (commentletters@waterboards.ca.gov)

RECEIVED
NOV 30 2007
SWRCB EXECUTIVE

November 30, 2007

Jeanine Townsend
Acting Clerk to the Board
Executive Office, State Water Resources Control Board
P.O. Box 100
Sacramento, CA 95812-0100

Dear Ms. Townsend:

The Industrial Environmental Association is an organization representing manufacturing, technology and biotechnology companies with member facilities throughout the State of California.

We have reviewed the State Water Resources Control Board's Draft Staff Report detailing the proposed adoption and implementation of sediment quality objectives (SQOs) for enclosed bays and estuaries of California, including the Appendices, and available supporting documentation for the recommended methods. To the extent possible, we have generated both general and specific comments on the technical merits of the Board's recommendations (please see attachment). Unfortunately, it is not possible to perform a complete, scientific evaluation of the Board proposal, given the limited information provided. Our primary comment on the Draft Report is that it falls well short of the level of documentation, justification, and validation that would be required to evaluate, let alone justify, a new technical policy with such sweeping implications. Nevertheless, we have identified a number of serious technical problems with the proposed process, as we understand it. Our chief concerns include:

- Reliance on several novel evaluation tools and indices of adverse effects that have not been adequately reviewed or validated to form the basis of policy.

- Reliance on overly simplistic, categorical labels to characterize highly complex, quantitative relationships between sediment chemistry and biology.

- Failure to explain or justify thresholds of putative effect.

- Use of simplistic methods to combine disparate lines of evidence without any apparent justification or analysis of accuracy.

- Lack of any reliance on, or even acknowledgment of, the importance of appropriate reference data and conditions in sediment characterization.

- Failure to adequately incorporate evaluation of non-chemical stressors and other factors that may influence various lines of evidence.

- Absence of any apparent prediction tolerances or uncertainty analysis for most lines of evidence.

- Absence of any true sediment chemistry evaluation. Indices of toxicity and benthic community effects are misrepresented as indices of sediment chemistry.

- Inappropriate reliance on theoretical model-driven predictions of toxicity and benthic community effects, when measured toxicity and benthic community endpoints are available.

- Use of inappropriate and/or inadequately justified statistical methods in the development of novel indices of toxicity and benthic community impacts.

In short, the Draft Report makes many poorly supported recommendations, many of which are inconsistent with good scientific practice and current standards in sediment assessment and characterization. We do not believe the Draft Report is adequate to even permit a proper review of the Board's proposed methods. It is certainly inadequate to serve as the basis for a Statewide SQO process.

Meaningful evaluation of the Board technical proposals requires review of the information used to develop and calibrate the proposed assessment methods, including but not limited to details and data used to generate the many quality thresholds upon which the method relies. We respectfully request that the Board make these details and data publicly available, and extend the comment period to allow the review and analysis necessary for meaningful public comment on the proposed SQOs.

Comments on Draft SQO Process

The following comments on the proposed adoption and implementation of sediment quality objectives (SQOs) for enclosed bays and estuaries of California are based on a review of the Draft Staff Report (the report) issued by the State Water Resources Control Board (the Board), dated September 27, 2007 (SWRCB 2007).

General Comments:

1. The proposed methods for deriving and applying SQOs include multiple tests and lines of evidence (LOEs). In addition to standardized, widely used "off-the-shelf" assessment tools (e.g., standard toxicity tests), several novel methods were developed by the Board solely for this purpose (e.g., chemical score index or CSI). Other published methods were adapted or modified specifically for the current use (e.g., California logistic regression model or LRM). It is not possible to fully assess the scientific basis, validity, or relevance of all of the proposed tools, based on the limited documentation provided in the report. In order to scientifically evaluate the proposed methods, it would be necessary to fully review the data and theoretical basis of all methods proposed, and to examine the validation process and interpretation of results that have led the Board to propose their adoption. At best, the report makes passing reference to some selection criteria employed (in choosing logistic regression approaches for example), but in no case does the report contain the technical backup, data, or detail to enable any reviewer to perform an independent evaluation of the proposal. Much more detail is required to adequately assess the proposed methods, as well as adequate time to review it.

   In particular, no confidence limits or analysis of uncertainty associated with the predictions of the proposed methods, or inherent in the integration of such diverse endpoints into a single score, is included in the report. If such an analysis has not been performed, it is essential this be done and adequately evaluated prior to adoption or implementation of the SQO process.

   The report is primarily devoted to documentation of the purposes, objectives, and intended benefits to the public of ecological SQO development. However, the underlying scientific justification for the proposed assessment methods is only superficially described. The Appendices to the report include a step-by-step guide to application of the proposed methods, and an example calculation, but little in the way of rationale for selection of specific recommended methods. This is particularly the case with respect to novel sediment evaluation tools that were developed specifically for this purpose (e.g., CSI), or modified from the original application (e.g., LRM). Implementation of the proposed sediment characterization process requires many explicit and implicit assumptions that are not adequately explained or justified in the report.

2. The proposed assessment method is based on integrating multiple LOEs. Three primary LOEs (sediment chemistry, toxicity, and benthic community impacts) are

independently evaluated, then combined to create a final station assessment for each sediment location evaluated. Each primary LOE is itself derived from multiple secondary LOEs. While based on quantitative data inputs (e.g., chemical concentrations, percent survival of test organisms, benthic community indices), all LOEs are evaluated on a semi-quantitative, categorical basis, on an integer scale of 1 to 4. Divisions between categories are arbitrary, without apparent rigorous technical basis, and appear designed to promote rigid decision-making rather than application of sound professional judgment in the interpretation of SQOs. Furthermore, the proposed manner of combining LOEs (generally taking an average of multiple values and rounding up to the next whole number) sacrifices accuracy and resolution for the sake of simplicity. The result is that very small incremental differences in input data can result in major differences in the final LOE scores due to the arbitrary placement of category thresholds (see chemistry score comment 4 for an example).

3. Despite the use of reference conditions to establish a baseline for assessment of impairments of beneficial uses, both in established regulations and in practice (see section 4 of the report), the SQO assessment methods that are described do not incorporate any comparison to reference conditions. By failing to incorporate comparisons to reference conditions, the assessment methods are not grounded in practical reality.

   The goals of the SQO program, as stated in the report, are to establish methods to evaluate conformance to a "protected condition" (section 1.9). This "protected condition" is not defined in section 1.9, but section 5.6 provides several alternative definitions of a "protective condition". However, the definitions in section 5.6 are all based on the results of the assessments of multiple lines of evidence. If the "protective conditions" defined in section 5.6 are taken to be equivalent to the "protected condition" that is defined as a goal in section 1.9, then the consequence is that the goal is defined by the results of the analysis—clearly an inappropriate situation. The "protected condition" defined in section 1.9 must be independent of the methods used to evaluate conformance to that goal. To illustrate why this is so, consider that the assessment methods might be highly inaccurate, resulting in a substantial under-prediction or over-prediction of actual effects; nevertheless, following section 5.6, the lowest assessment categories would be defined as the protected condition, even if those categories actually corresponded to major adverse effects (in the case of under-prediction of effects) or if the highest categories corresponded to a low level of adverse effects (in the case of over-prediction of effects). Because the "protective condition" defined in section 5.6 is not appropriate as the "protected condition" defined in section 1.9, the "protected condition" established as a goal in section 1.9 remains undefined.

   Furthermore, bays and estuaries are not uniform environments: they contain significant spatial variations in physical and biological conditions. As a consequence, uniform sediment quality assessment methods cannot be applied to all sampling locations within a bay or estuary unless those methods incorporate evaluation of a reference condition relevant to each sampling location.

The report should be revised to clearly establish appropriate reference conditions as the protected condition, and the assessment methods must also be revised to incorporate comparisons to reference conditions.

4. The limited value of sediment chemistry relative to direct measurement of biological effects is recognized in section 5.5 of the report. Variations in bioavailability and in exposure are acknowledged to impede the inference of biological effects from chemistry data alone, and the report concludes that "As a result of the factors described above, sediment quality indicators based on pollutant concentrations in sediment have only limited utility when used by sediment managers unless bolstered by effects data such as toxicity and benthic community disturbance." In contrast to these statements, the method defined in the report for combining LOEs (Attachment B of Appendix A) gives equal weight to chemistry relative to the biological lines of evidence. Equal weighting of the chemical and biological lines of evidence is contrary to the stated intent to give priority to direct measurements of biological effects.

Comments on Sediment Chemistry Scoring:

The sediment chemistry LOE is derived by combining two independently calculated scores, the CSI score and Pmax score (Pmax is the maximum probability of toxicity from a sample constituent, as predicted by the LRM). Each score is an integer, ranging from 1 to 4. The two values are averaged, then rounded up to calculate the final chemistry score, which is interpreted as reflective of chemical exposure (minimal, low, moderate, or high). There are several flaws inherent in this method:

1. As a primary LOE, the chemistry score should reflect information not included in the two other LOEs (benthic community and toxicity). However, the two components of the chemistry score are, in fact, simple functions of predicted benthic community effects and toxicity. CSI is nothing more than a relative measure of the likelihood of benthic community impacts at a station, as predicted by the benthic response index (BRI). The LRM score is simply a relative measure of predicted amphipod toxicity, based on a logistic regression of chemical concentration and toxicity from a database of laboratory bioassay data. Given that BRI and amphipod toxicity are used to determine the benthic community and toxicity LOE scores respectively (together with other benthic community indices and toxicity endpoints), the proposed measure of sediment chemistry includes no new information that is not accounted for by other LOEs. BRI and amphipod toxicity are effectively double-counted by using them to define the sediment chemistry LOE.

   The chemistry LOE should instead be based solely on comparison of chemical measurements to a representative (regional or local) background condition, not on biological effects.

2. Moreover, the use of theoretical BRI and toxicity indices for any purpose makes little sense when empirical measures of both benthic toxicity and BRI are

available. The proposed approach requires the use of predictive models to estimate BRI and toxicity at stations for purposes of evaluating the chemistry LOE, then requires direct calculation of BRI and direct measurement of sediment toxicity to evaluate the benthic community and toxicity LOEs. There is no reason to rely on a highly uncertain prediction when direct measurements are available. Because the CSI and LRM outputs are simply a surrogate for non-site-specific benthic community and toxicity responses, they should always be overridden by site-specific benthic community and toxicity data, and therefore have no value to assess the chemistry LOE or the two biological LOEs .

3. The CSI is a novel approach to characterizing sediment chemistry that was developed by the Board for the express purpose of setting SQOs. It employs a relatively obscure statistical parameter, Cohen's kappa coefficient (Cohen 1960) to causally associate concentration ranges of individual chemicals in sediments with BRI ranges, and to set the weighting factors used to calculate station CSI values from constituent chemical concentrations. This is a non-standard application of the kappa statistic, which was developed and has traditionally been used in the fields of education and medicine to measure the agreement between two "raters" (e.g., two academic testers' findings or two physicians' diagnoses). Although considered useful as a way to test rater independence, the validity of the kappa statistic as a quantitative measure of the level of agreement has been criticized (Brennan and Prediger 1981, Sim and Wright 2005). Reliance on kappa in the current application to quantitatively establish relationships between measured variables, and to scale the relative importance of components within mixtures may well be inappropriate, although it is difficult to fully assess without more detail about its use than the Board has provided. At the least, a thorough technical justification for the proposed statistical approach is needed, beyond the limited information presented in the report, or in the cited CSI method document (Ritter et al. 2007). The method developers should also explain why more standard and commonly used methods to characterize relationships between dependent variables, such as Spearman's correlation, were not used. The use of a metric such as kappa, which has poorly defined statistical characteristics, could be avoided if all data were kept in their original form, on a continuous scale, and not converted to a highly quantized ordinal scale. This unnecessary scale transformation not only discards information that is present in the data, but also leads to the use of semi-quantitative measures such as kappa, which is subject to varying interpretations.

4. The two proposed secondary LOEs for sediment chemistry assessment, CSI and Pmax, appear to be logically inconsistent internally, and with each other. For example, the CSI predicts low benthic disturbance conditions (category 2) will be observed when total DDT concentrations in sediment exceed 0.50 μg/kg, a detection limit level. However, the LRM prediction for amphipod mortality at this concentration is only 1 percent, a level that would seem unlikely to have significant population-level implications. Contrast this with the case of copper. Low benthic effects for copper are predicted at a bulk concentration of 53 mg/kg, a level well within the background range in many locations. At 53 mg/kg copper, the LRM predicts amphipod mortality of 24 percent. The implications of "low

disturbance" are unclear, based on the performance of these indicators for different chemicals. Similar discrepancies exist throughout the four disturbance categories.

5. The category 2 CSI thresholds for pesticides are set so low that they approach normal detection limits. This calls into question the applicability of the logistic regressions used at very low concentrations. Have effects on benthic communities been reliably observed at these concentrations, or are they simply predicted by extrapolation of relationships determined at higher concentrations? Without full access to the data from which the regressions were derived, it is impossible to answer this question.

6. As noted above in the general comments, the proposed categorical scoring scheme forces an arbitrary, low-resolution outcome on all sediment evaluations. This approach artificially simplifies complex, multivariate relationships between chemical and biological effects, but it is unlikely to result in an accurate representation of sediment quality in bays and estuaries, and may create misleading results. When concentrations of even a single chemical are close to a threshold between categories, station scores may be driven by artifacts of the method rather than the data. Consider the following simple example:

    - Two sediment stations, A and B, have identical chemistry data, with the sole exception of zinc, which differs very slightly (4 percent). A third station, C, has double the zinc found in A, and also has an order of magnitude higher HPAH, LPAH, and PCBs:

Chemical	Concentration		
	Station A	Station B	Station C
Cadmium (mg/kg)	0.20	0.20	0.20
Copper (mg/kg)	40	40	40
Lead (mg/kg)	25	25	25
Mercury (mg/kg)	0.10	0.10	0.10
Zinc (mg/kg)	125	130	250
HPAH (µg/kg)	1,500	1,500	15,000
LPAH (µg/kg)	900	900	9,000
Alpha Chlordane (µg/kg)	1.1	1.1	1.1
Gamma Chlordane (µg/kg)	1.4	1.4	1.4
DDDs, total (µg/kg)	3.8	3.8	3.8
DDEs, total (µg/kg)	3.6	3.6	3.6
DDTs, total (µg/kg)	4.6	4.6	4.6
PCBs, total (µg/kg)	250	250	2,500

    - Based on the above data, Pmax is driven by zinc at all three stations: Station A = 0.49 (category 2), Station B = 0.50 (category 3), Station C = 0.66 (category 3)

    - The CSI at each station is calculated below:

Chemical	Station A			Station B			Station C		
	CSI Category	Wt.	CSI	CSI Category	Wt.	CSI	CSI Category	Wt.	CSI
Cadmium	2	38	76	2	38	76	2	38	76
Copper	1	100	100	1	100	100	1	100	100
Lead	1	88	88	1	88	88	1	88	88
Mercury	2	30	60	2	30	60	2	30	60
Zinc	2	98	196	2	98	196	3	98	294
HPAH	3	16	48	3	16	48	4	16	64
LPAH	3	5	15	3	5	15	4	5	20
Alpha Chlordane	2	55	110	2	55	110	2	55	110
Gamma Chlordane	2	58	116	2	58	116	2	58	116
DDDs, total	3	46	138	3	46	138	3	46	138
DDEs, total	2	31	62	2	31	62	2	31	62
DDTs, total	3	16	48	3	16	48	3	16	48
PCBs, total	3	55	165	3	55	165	4	55	220
SUM			**1,222**			**1,222**			**1,396**

- The weighted average CSI values (rounded up) are calculated by dividing the CSI sum by the sum of the weighting factors (always 636):
  Station A = 1.92 (2), Station B = 1.92 (2), Station C = 2.20 (2)

- Finally, the total station chemistry scores are:
  Station A = 2 (low exposure), Stations B and C = 3 (moderate exposure)

The miniscule difference between A and B (which is well within the range of analytical variability) have the same net effect on the final chemistry scores as the drastic differences between A and C. Although this example was contrived to make a point, many other combinations that result in similar nonsensical outcomes are possible. Such artifacts are inherent in any scoring scheme that is based on quantum thresholds in concentration, such as that proposed by the Board. The sensitivity of the process to small changes near category boundaries is exacerbated by the requirement to round all scores up to the next highest category. When two adjacent numbers are averaged and rounded up (e.g., Category 2 Pmax and Category 3 CSI), **the net effect of the prescribed process is use of the highest value, not an average at all**. If such a scheme is relied upon to characterize sediment chemistry, situations such as the example above can be expected to occur. This will cause confusion rather than clarity, and will not result in sound decision-making or transparency of the regulatory process.

Comments on Sediment Toxicity Scoring:

The toxicity LOE is derived from at least two approved laboratory bioassays conducted with sediment samples. At least one amphipod survival endpoint and one sublethal endpoint (*Neanthes* growth or *Mytilus* development) must be measured in separate bioassays. The results of the bioassays are scored on an integer scale from 1 to 4, based on a comparison to control results. The scores from all available tests are then averaged, and rounded up to the next whole number. There are several flaws inherent in this method:

1. As noted in the general comments, the proposed method fails to rely upon, or even include comparison to reference response levels. Negative controls are used primarily as a quality assurance tool in environmental toxicology, to establish dose response baselines and evaluate signal to noise ratio for the measured endpoints. Full evaluation of the environmental significance of toxicological data, especially in the case of a complex, multi-chemical exposure scenario like sediment toxicity, requires a comparison to endpoints measured under appropriate reference conditions.

2. The method implies that all tests are equally reliable and predictive of sediment toxicity. This is not necessarily the case. Sublethal endpoints are inherently more variable than mortality, particularly endpoints that require subjective determination (i.e., "normal" development in bivalve larvae). Weighting all tests equally effectively results in a toxicity LOE with the variance of the least reproducible study.

3. As with chemistry characterization (see comment #4 on chemistry scoring above), all toxicity scoring schemes that generate discontinuous, qualitative classifications from continuous, quantitative observations or data result in loss of measurement accuracy and resolution, and are likely to introduce artificial distinctions among samples that are not reflective of meaningful differences in toxicity.

Comments on Benthic Community Scoring:

The benthic community LOE is assessed by calculating four indices from abundance data at each station: BRI, relative benthic index (RBI), index of biological integrity (IBI), and the river invertebrate prediction and classification system (RivPACS). Each of the four indices is calculated, assigned a score (integer from 1 to 4), then the median value of the four is taken as the final station score for benthic community.

1. As with toxicity data, full interpretation of benthic community data requires comparison to appropriate reference stations. Only in this way can index performance and sensitivity be evaluated. Failure to adequately assess the reference benthic condition renders any conclusions reached using the proposed approach questionable.

2. No adequate justification is offered by the Board for their recommendation to use these four, and only these four, indices of benthic community disturbance. Two of the four indices (IBI and RivPACS) were developed for very different habitats (freshwater riverine systems) than those for which they are proposed here. Beyond a single station example, little detail is provided on how to apply and interpret these freshwater methods, and no validation of their reliability and relevance in estuarine and marine environments is included in the report.

3. Even some basic procedural steps involved in calculating the various benthic community indices are incompletely documented and explained in the report. For example, calculation of IBI requires identification of pollution tolerant and pollution sensitive taxa. Inter-species sensitivity difference is a continuum, not a binary condition. Categorization of taxa at any site as pollution tolerant or sensitive requires guidance (i.e., a comprehensive definition of each category, justification for the categorization guidance, and validation of the decisions made).

4. As with chemistry and toxicity characterization (see comments above), all benthic community scoring schemes that generate discontinuous, qualitative classifications from continuous, quantitative observations or data result in loss of measurement accuracy and resolution, and are likely to introduce artificial distinctions among samples that are not reflective of meaningful differences in benthic community structure or chemical impacts. Furthermore, the breakpoints that are used to convert the continuous index values into discrete categories are not adequately explained or justified. These breakpoints appear to be arbitrary, and appear not to have been validated against a range of actual sites.

5. Although poorly explained in the report, we have reviewed the method documents for the BRI (Smith et al. 2001, 2003). The BRI approach is an attempt to compress the wealth of information available on benthic macroinvertebrate communities into a single number. However, evaluation of the approach indicates that several of its features are either highly subjective or negatively affected by uncertainty. Although the BRI values represent a continuum of benthic community conditions, variation along this continuum is assumed to be attributable solely to pollution effects, and the manner in which the benthic response thresholds break that continuum into discrete categories is highly artificial and subjective. The use of these artificial thresholds can overestimate the significance of any benthic community alterations, particularly with respect to major community characteristics such as taxa richness, total abundance, and species diversity.

   The BRI approach is also limited by the fact that it ignores study-specific reference conditions and the fact that species replacements can occur in benthic communities without resulting in measurable losses of community function. Many of the species with pollution tolerances may be similar to those of the reference species, and changes in the index may be the result of habitat changes unrelated to toxic chemicals (e.g., sediment grain size, sediment organic content and water currents).

   The varying physical conditions in bays—water depths, sediment grain size, temperature, salinity, and water currents—create a very complex environment, with associated complex variations in the benthic macroinvertebrate community. Some species prefer muddy bottoms, and some prefer sandy bottoms; some prefer deep water, and others prefer shallow water. Species also vary in their tolerance for differing temperature and salinity. Some of these habitat conditions vary systematically throughout an enclosed bay. For example, grain size, temperature, and salinity can all be expected to change systematically with distance from the open ocean. These systematically varying habitat characteristics also can be related to systematic changes in physical–chemical relationships, such as the tendency of some chemicals to naturally occur at higher concentrations on fine particles. Organisms that are naturally found in fine sediments are therefore more likely to have high pollution tolerance scores, and their presence will therefore result in elevated BRI values even in the absence of pollution. Therefore, interpretation of the BRI is only meaningful if an individual station value of the index is compared with an appropriate reference station.

Such uncertainties in the method may profoundly affect the scores and the presumed relationship to pollution effects. Because of these factors, the BRI index (or any other index) should not be incorporated into the SQOs without appropriate scientific justification for all arbitrary scores, category thresholds, and an independent validation of the index using benthic data sets from different bays.

Sincerely,

Patti Krebs

Patti Krebs
Executive Director

References

Brennan, R.L., and D.J. Prediger. 1981. Coefficient kappa: Some uses, misuses, and alternatives. Educat. Psychol. Meas. 41:687-699.

Cohen, J. 1960. A coefficient of agreement for nominal scales, Educat. Psychol. Meas. 20: 37–46.

Ritter K.J., S.M. Bay, R.W. Smith, D.E. Vidal-Dorsch, and L.J. Field. 2007. Development and evaluation of sediment quality guidelines based on benthic macrofauna responses. Southern California Coastal Water Research Project. Costa Mesa, CA.

Sim, J., and C.C. Wright. 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. Phys. Therap. 85:257-258.

Smith, R.W., M. Bergen, S.B. Weisberg, D.B. Cadien, A. Dalkey, D.E. Montagne, J.K. Stull, and R.G. Velarde. 2001. Benthic response index for assessing infaunal communities on the southern California mainland shelf. Ecol. Applic. 11:1073-1087.

Smith, R.W., J.A. Ranasinghe, S.B. Weisberg, D.E. Montagne, D.B. Cadien, T.K. Mikel, R.G. Velarde, and A. Dalkey. 2003. Appendices C and D of benthic response index for assessing infaunal communities on the southern California mainland shelf. Southern California Coastal Water Research Project Authority.

State Water Resources Control Board (SWRCB). 2007. Draft Staff Report. Water Quality Control Plan for Enclosed Bays and Estuaries. Part 1. Sediment Quality. September 27, 2007.

Thompson, B., and S. Lowe. 2004. Assessment of macrobenthos response to sediment contamination in San Francisco Estuary, California, USA. Environ. Toxicol. Chem., 23:2178–2187.