May 22, 2006

Background Information in Support of May 25 Conference Call With the Scientific Steering Committee
Selection of Benthic Indices for use as the Benthic Indicator

At their 28th February - 2nd March meeting, the benthic development team presented their recommendation that a combination of three indices be used as the benthic indicator for the California Sediment Quality Objectives (SQO). The Scientific Steering Committee (SSC) concurred with the use of a combination of indices, but asked for additional information to support the selected combination. Specifically, the SSC asked for documentation of the performance of this combination relative to all other possible combinations. This document provides that information.

At the SSC meeting, five index approaches that had been calibrated to California data were presented:

- The Benthic Response Index (BRI), which was originally developed for the southern California mainland shelf by Smith *et al.* (2001) and extended into California bays and estuaries by Smith *et al.* (2003) and Ranasinghe *et al.* (2004). The BRI is the abundance-weighted average pollution tolerance score of organisms occurring in a sample.

- The Relative Benthic Index (RBI), which was originally developed for estuarine applications in California's Bay Protection and Cleanup Program (Hunt *et al.* 2001). The RBI is the weighted sum of (a) several community parameters (total number of species, number of crustacean species, number of crustacean individuals, and number of mollusc species), and abundances of (b) three positive and (c) two negative indicator organisms.

- The Index of Biotic Integrity (IBI), which was developed for freshwater streams and adapted for estuarine applications by Weisberg *et al.* (1997), Van Dolah *et al.* (1999) and Thompson and Lowe (2004). The IBI identifies community measures that have values outside a reference range.

- The River Invertebrate Prediction and Classification System (RIVPACS), which was originally developed for British freshwater streams by Wright *et al.* (1993) and applied in estuaries and bays for the first time in this project. The approach compares the assemblage at a site with an expected species composition determined by a multivariate predictive model that is based on species relationships to habitat gradients (Van Sickle *et al.* 2006).

- The Benthic Quality Index (BQI), which was originally developed for the west coast of Sweden by Rosenberg *et al.* (2004) and applied in the USA for the first time in this project. The BQI is the product of the logarithm ($base_{10}$) of the total number of species and the abundance-weighted average tolerance of organisms occurring in a sample. Species tolerance scores are calculated differently than for the BRI, instead based on relationships of the abundance distributions to Hurlbert's (1971) expected number of species.

The success of these indices was evaluated by comparison with the consensus of nine benthic experts, who classified the condition of 36 samples. These samples were selected by rank ordering the California SQO database according to chemical concentration (using ERMQ) and then randomly selecting sites from within quartile groups, so that a range of benthic conditions was likely to be encountered. Twenty-four of the sites were from euhaline coastal embayments in southern California with the remainder from polyhaline San Francisco Bay. The experts were provided species abundances, together with depth, salinity and sediment grain size information and asked to classify each site into one of four condition categories: reference, low disturbance, moderate disturbance or high disturbance.

Assessments of the 36 samples by the benthic indices and all possible index combinations were compared to the consensus expert condition assessment and evaluated in three ways:

1. Status classification accuracy with respect to a two-category status classification, in which the expert classification was expressed as good (reference or low disturbance) or bad (moderate or major disturbance). This mimics the evaluation approach used in most previously published benthic indicator development efforts.

2. Categorical classification accuracy, which was evaluated with respect to the four categories used in the SQO multiple line of evidence integration (reference, low disturbance, moderate disturbance or high disturbance).

3. Bias in category designation, which is the sum of differences between index (or index combination) and the consensus categorical classification of the experts. Positive values indicate a tendency to score samples as more disturbed than the expert consensus, while negative values indicate a tendency to score samples as less disturbed. Larger values indicate stronger bias.

The results are presented in Table 1. For context, Table 2 provides the same measures where the experts were assessed relative to the answers of their peers.

Individually, none of the indices fared as well as the experts. RIVPACS did the best, with a 91% correct status classification, a 71% correct category classification and a low bias. This was better than two of the nine experts, but not as good as the median expert. The BRI also had a 91% correct status classification, which equaled or exceeded that of four experts, though its categorical classification accuracy was less than that of the lowest expert. None of the other indices had a status classification accuracy that exceeded that of the lowest expert, but all of them had at least a 75% correct status classification, a rate that has been used frequently in other estuarine systems to assess fidelity of indices.

Index combinations (with indices combined using the same algorithm as used for the toxicity indicator) generally performed better than individual indices, and combinations of three or more indices generally performed better than combinations of two. Several index combinations of three or more yielded similar results, but the index combinations that performed best were #26, a four-index combination of the BRI, the RBI, the IBI and RIVPACS; #29, a four-index combination of the BRI, the BQI, the IBI and RIVPACS; and #24, a three-index combination of the BRI, the RBI and RIVPACS. These combinations had the highest status classification accuracy, the highest category classification accuracy and a relatively low bias. These combinations outperformed the average expert for status classification, though they were outperformed by 5 of the 9 experts for the categorical classification.

Indicator Recommendation

We recommend adoption of index combination #26 as the SQO Benthic Indicator. This recommendation differs from that made at the February meeting, in combination #16 was recommended (the same combination as #26, but without RIVPACS). The February recommendation was based on familiarity of the development team with the three indices in combination #16 and because their lesser experience with RIVPACS (applied for the first time here to estuaries) could present a challenge in technology transfer to the Regional Boards. However, we agree with the SSC's observation that RIVPACS performed best among all of the indices individually and also did well in combination. Subsequent to the SSC meeting, the team has improved its familiarity with RIVPACS and now feels comfortable recommending the adoption of combination #26.

Combination #26 was selected over combination #29, which performed equally, because of concerns with the technical foundation of the BQI. The BQI assumes a monotonic decrease of species richness with pollution exposure, which is inconsistent with the Pearson-Rosenberg model. The assumption is also inconsistent with our observation about species richness along pollution gradients in California's nearshore waters. When the BQI was applied to these well-defined pollution gradients, it performed poorly in comparison with the other candidate indices. Because of concerns with the indices technical foundation and its failure to work well in a system with well-defined gradients that provide a more direct test of the index, we were reluctant to adopt it for SQO application.

Combination #26 was also selected over combination #24 (same indices as in combination #26, except it excludes the IBI), which also performed equally well. This was done because the IBI has already been published and has become somewhat established for these applications in San Francisco Bay. This IBI approach has also been widely accepted in other parts of the country and its inclusion, at no loss of indicator efficiency, will facilitate its continuing evolution in California.

We anticipate that over time, as additional data are collected through the SQO process, the benthic indicator will be refined, strengthened and streamlined. The number of indices that are included for these assessments will likely decline as the indices that are strongest and most well accepted by the community become more apparent. However, we agree with the SSC's suggestion that retaining at least some combination of indices that measure different attributes of the benthos is advisable.

We look forward to the SSC's feedback on our recommendation.

Table 1. Classification Accuracy and Bias for Indices and Index Combinations. Classification accuracy is presented for status ("good" vs "bad") and for four classification categories. Each of 36 samples was assessed into one of four numeric categories by the index or index combination and the result compared with consensus categories from an independent assessment by nine benthic ecologists.

No. of indexes	#	Measure	Status Classification Accuracy (%)	Category Classification Accuracy (%)	Category Bias
One	1	BQI	85.7	68.6	7
	2	BRI	91.4	62.9	-4
	3	IBI	75.9	55.2	-9
	4	RBI	80.0	57.1	13
	5	RIV	91.4	71.4	3
Two	6	BQI, BRI	85.7	65.7	8
	7	BQI, IBI	85.7	65.7	5
	8	BQI, RBI	82.9	54.3	16
	9	BQI, RIV	82.9	68.6	11
	10	BRI, IBI	88.6	68.6	-1
	11	BRI, RBI	88.6	62.9	12
	12	BRI, RIV	88.6	71.4	7
	13	IBI, RBI	80.0	54.3	11
	14	IBI, RIV	91.4	71.4	3
	15	RBI, RIV	82.9	54.3	16
Three	16	BRI IBI RBI	91.4	74.3	1
	17	BQI BRI IBI	91.4	71.4	0
	18	BQI BRI RBI	88.6	77.1	6
	19	BQI BRI RIV	94.3	74.3	3
	20	BQI IBI RBI	88.6	68.6	7
	21	BQI IBI RIV	94.3	77.1	2
	22	BQI RBI RIV	88.6	71.4	6
	23	BRI IBI RIV	91.4	68.6	-3
	24	BRI RBI RIV	94.3	80.0	3
	25	IBI RBI RIV	94.3	74.3	3
Four	26	BRI IBI RBI RIV	94.3	80.0	5
	27	BQI IBI RBI RIV	88.6	71.4	6
	28	BQI BRI RBI RIV	88.6	77.1	8
	29	BQI BRI IBI RIV	94.3	80.0	5
	30	BQI BRI IBI RBI	88.6	77.1	8
Five	31	All	94.3	77.1	4

Table 2. Classification Accuracy and Bias for Expert Consensus Results. Highest, average and lowest values for nine benthic ecologists. Classification accuracy is presented for ("good" vs. "bad") status and for four classification categories. Bias is the sum of category differences between the experts and the consensus (median) of the experts.

Consensus Contribution	Status Classification Accuracy (%)	Category Classification Accuracy (%)	Category Bias
Highest values	97.1	91.4	+7, -4
Average value	92.4	81.9	2.8
Lowest values	85.7	71.4	0

Literature cited

Hunt, J.W., B.S. Anderson, B.M. Phillips, R.S. Tjeerdema, K.M. Taberski, C.J. Wilson, H.M. Puckett, M. Stephenson, R. Fairey and J.M. Oakden. 2001. A large-scale categorization of sites in San Francisco Bay, USA, based on the sediment quality triad, toxicity identification evaluations, and gradient studies. *Environ. Toxicol. Chem.* **20**:1252-1265.

Hurlbert, S.H. 1971. The nonconcept of species diversity: A critique and alternative parameters. *Ecology* **52**:577-586.

Ranasinghe, J.A., B. Thompson, R.W. Smith, S. Lowe and K.C. Schiff. 2004. Evaluation of benthic assessment methodology in southern California bays and San Francisco Bay. Southern California Coastal Water Research Project. Westminster, CA. Technical Report 432.

Rosenberg, R., M. Blomqvist, H.C. Nilsson, H. Cederwall and A. Dimming. 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. *Mar. Pollut. Bull.* **49**:728-739.

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. *Ecological Applications* **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. Extending the Southern California Benthic Response Index to Assess Benthic Condition in Bays. Southern California Coastal Water Research Program. Westminster, CA. Technical Report 410.

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

Van Dolah, R.F., J.L. Hyland, A.F. Holland, J.S. Rosen and T.R. Snoots. 1999. A benthic index of biological integrity for assessing habitat quality in estuaries of the southeastern USA. *Mar. Environ. Res.* **48**:269-283.

Van Sickle, J., D.D. Huff and C.P. Hawkins. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshwater Biol.* **51**:359-372.

Weisberg, S.B., J.A. Ranasinghe, L.C. Schaffner, R.J. Diaz, D.M. Dauer and J.B. Frithsen. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries* **20**:149-158.

Wright, J.F., M.T. Furse and P.D. Armitage. 1993. RIVPACS: a technique for evaluating the biological water quality of rivers in the UK. *European Water Pollution Control* **3**:15-25.