Background Information in Support of June 23 Conference Call With the Scientific Steering Committee

Selection of Sediment Chemistry Indicators

June 22, 2006

At the February 28-March 2 meeting of the Scientific Steering Committee (SSC), the Sediment Quality Objectives (SQO) science team presented the results of statistical analyses that evaluated the performance of various chemical indicators (sediment quality guidelines applied to chemical mixtures). An analysis of the indicators' correspondence with amphipod mortality was presented along with a recommendation to use the CAPmax as an indicator of chemical exposure with respect to sediment toxicity. The SSC expressed support for the indicator comparison and selection process, but requested documentation that the results based on bootstrapped analyses were consistent with results obtained without bootstrapping. In addition, the SSC requested that the science team provide a full suite of comparison results for of all of the candidate indicators, rather than results for the subset of indicators presented at the meeting.

Preliminary results of the evaluation of indicators for describing chemical exposure with respect to benthic community disturbance were also presented at the SSC meeting. The SSC endorsed the concept of using a combination of toxicity-based and benthos based indicators to describe chemical exposure and recommended that the science team complete its development and evaluation of chemical indicators based on benthic community response. The SSC also directed the science team to conduct its evaluation using data from the final benthic indicators selected by the benthic development team.

This document presents the results of the chemical indicator evaluation for both toxicity and benthic community response conducted in response to the SSC recommendations. Results are presented for a variety of established and new or calibrated candidate indicators, each providing a measure of exposure based on the mixture of contaminants present in the sample. A summary of the candidate indicators and statistical methods is provided in Appendix A.

Evaluation of indicators based on toxicity

The complete results of the statistical analyses for 10 candidate toxicity-based chemical indicators are shown in Tables 1-4. Each indicator was evaluated for correlation with amphipod mortality, and for two measures of accuracy in predicting the toxicity response classification: agreement and weighted kappa. Both statewide and regional versions of the indicators were evaluated.

Documentation of the effect of the bootstrapped statistical analysis procedure is also shown in Tables 1-4. The correlation and agreement for each indicator were calculated using both bootstrapped and nonbootstrapped analyses. Both sets of analyses showed similar trends in performance, indicating that the bootstrapped results were representative of the chemical indicators' performance when applied to other types of data distributions.

The results of both the correlation and classification accuracy analyses were used to evaluate the candidate chemical indicators and select the recommended approach. The revised results are in agreement with the results presented at the February SSC meeting, that the statewide version of CAPmax has the best overall performance. The CAPmax had the highest correlation with mortality among the indicators developed for statewide application (Table 1). Results for the two measures of classification accuracy (weighted kappa and agreement) also showed that the statewide version of CAPmax was in the best performing group of indicators for both the north and south regions (Table 2). Several other statewide indicators (TCS, CA ERM, Consensus, NOAA ERM) also had a relatively high level of classification accuracy, but the results were not consistent between the north and south data sets.

The SSC recommended use of a statewide indicator of chemical exposure, unless a region-specific indicator was shown to provide substantially better performance. Analyses of indicator correlation and classification accuracy were conducted using region-specific indicators to investigate this issue. These analyses included the same statewide indicators evaluated previously, but used region-specific thresholds, and also included additional indicators that were calibrated to each region (indicated by a "Nor" or "So" prefix). While some of the regional indicators had relatively high performance, the results were not consistently better than CAPmax among performance measures or regions. For example, the correlation analyses showed that several region-calibrated indicators had relatively high correlations in each region (Table 3), but the statewide CAPmax was the only indicator with a high correlation for both the north and south regions. The regional classification accuracy analyses for the north produced mixed results: the weighted kappas for TCS and NorTCS (Table 4) were greater than the statewide CAPmax value (Table 2), but there was no difference in agreement. Both measures of classification accuracy for the best performing regional indicators in the south were also similar to those obtained for the statewide CAPmax. The use of regional thresholds for the CAPmax (Table 4) did not produce a marked improvement in classification accuracy over a single set of statewide thresholds (Table 2).

The statewide version of CAPmax is recommended for use as a toxicity-based chemical indicator for both the north and south regions. This recommendation is consistent with the SSC's guidance to use a statewide indicator, unless a substantial improvement in performance can be demonstrated through the use of a regionally calibrated indicator.

Evaluation of indicators based on benthos

In contrast to the toxicity-based chemical indicator results, a recommendation regarding a benthos-based chemical indicator was not presented at the February SSC meeting because the final analyses had not been conducted. The results of the final performance analyses are included in this document, as requested by the SSC (Tables 5 and 6). These analyses used the final classifications of benthic community condition based on the combination of four benthic indices recommended by the benthic indicator development team.

The same process was used to evaluate the performance of the benthos-based indicators as was described for the toxicity-based chemical indicators. There were less data available for analyses of the benthic response indicators, which limited some aspects of the indicator evaluation in the north. The north validation data set consisted of 25 samples, which was not sufficient to support bootstrapped analyses. There were sufficient data for these analyses in the south validation data set. Statistical

analyses were only conducted on a regional basis because the benthic assemblages and community indicators differed between the north and south.

NorBCS and SoBCS are recommended for use as benthic community-based indicators for the north and south regions, respectively. While the relative ranking in the performance measures varied among several indicators, SoBCS placed consistently in the top-performing group of south indicators and NorBCS had one of the highest correlations in the north. The NorBCS and SoBCS are also the only indicators developed using benthic community response data (other indicators had only the thresholds optimized for the benthos) and thus represents the most direct way to implement the SSC's recommendation to use a combination of toxicity-based and benthos-based chemical indicators to classify sediment toxicity. Alternatively, the Consensus approach could be used in the north with similar effectiveness. The Consensus approach has less flexibility with respect to future revisions/additions, as it is based on the mean of several chemical-specific guidelines. Thus it will be more difficult to revise the Consensus indicator in response to data for additional contaminants of concern, such as pyrethroid pesticides. NorBCS and SoBCS are amenable to the inclusion new chemical contaminants and can be easily updated as new data become available.

Table 1. Nonparametric correlation (Spearman) of statewide chemical indicators with amphipod mortality. Values in the highlighted cells are within the 90th percentile of the highest median correlation for the bootstrapped analysis. Analyses were conducted on the combined data for the north and south validation data sets.

		Correlation coefficient (r)	
Region	Indicator	Bootstrapped	Not Bootstrapped
State	CAPmax	0.35	0.40
State	TCS	0.27	0.34
State	NOAA ERM	0.25	0.29
State	Consensus	0.25	0.30
State	NatPmax	0.22	0.29
State	CA ERM	0.20	0.22
State	SQGQ1	0.16	0.23

Table 2. Classification accuracy of statewide chemical indicators for amphipod mortality. Values in the highlighted cells are within the 90th percentile of the highest median correlation for the bootstrapped analysis. Analyses were conducted using statewide thresholds applied to data from each region separately.

Statewide Thresholds		Weighted Kappa	% Agreement	
Region	Indicator	Bootstrapped	Bootstrapped	Not Bootstrapped
North	TCS	0.22	36	42
North	CA ERM	0.21	33	34
North	CAPmax	0.20	38	40
North	SQGQ1	0.13	35	40
North	Consensus	0.12	28	28
North	NOAA ERM	0.12	27	27
North	NatPmax	0.11	35	39
South	CAPmax	0.25	35	35
South	Consensus	0.22	36	36
South	NOAA ERM	0.21	38	38
South	NatPmax	0.18	34	33
South	CA ERM	0.15	34	36
South	TCS	0.15	29	29
South	SQGQ1	0.10	28	27

Table 3. Regional nonparametric correlation (Spearman) of chemical indicators with amphipod mortality. Values in the highlighted cells are within the 90th percentile of the highest median correlation for the bootstrapped analysis. Separate analyses were conducted on the north and south validation data sets.

		Correlation coefficient (r)	
Region	Indicator	Bootstrapped	Not Bootstrapped
North	CAPmax	0.39	0.45
North	NorTCS	0.38	0.47
North	TCS	0.37	0.49
North	NOAA ERM	0.31	0.37
North	NorCAPmax	0.27	0.34
North	SQGQ1	0.25	0.30
North	Consensus	0.23	0.28
North	CA ERM	0.22	0.28
North	NorCA ERM	0.22	0.29
North	NatPmax	0.15	0.21
South	CAPmax	0.42	0.42
South	SoCAPmax	0.37	0.39
South	NatPmax	0.33	0.34
South	Consensus	0.31	0.32
South	SoTCS	0.30	0.32
South	TCS	0.28	0.32
South	NOAA ERM	0.28	0.29
South	SQGQ1	0.26	0.29
South	SoCA ERM	0.18	0.20
South	CA ERM	0.18	0.19

Table 4. Regional classification accuracy of chemical indicators for amphipod mortality. Values in the highlighted cells are within the 90th percentile of the highest median correlation for the bootstrapped analysis. Separate analyses were conducted on the north and south validation data sets. Region-specific thresholds were applied to each of the indicators shown.

Regional Thresholds			% Agreement	
Region	Indicator	Bootstrapped	Bootstrapped	Not Bootstrapped
North	TCS	0.31	37	40
North	NorTCS	0.26	32	35
North	CA ERM	0.21	33	35
North	SQGQ1	0.21	33	35
North	NorCA ERM	0.21	33	35
North	NatPmax	0.20	33	37
North	NorCAPmax	0.20	35	37
North	NOAA ERM	0.17	30	33
North	CAPmax	0.16	27	27
North	Consensus	0.15	29	31
South	CAPmax	0.28	40	39
South	Consensus	0.25	39	39
South	NatPmax	0.22	36	36
South	NOAA ERM	0.22	38	37
South	SoCAPmax	0.22	36	37
South	So TCS	0.19	34	33
South	TCS	0.18	35	35
South	SQGQ1	0.18	33	31
South	SoCA ERM	0.16	35	36
South	CA ERM	0.13	33	34

Table 5. Nonparametric correlation (Spearman) of chemical indicators with benthic community condition. Values in the highlighted cells are within the 90th percentile of the highest median correlation for the bootstrapped analysis. Separate analyses were conducted on the north and south validation data sets.

Region	Indicator	Correlation coefficient (r)
		Not Bootstrapped
North	Consensus	0.40
North	NorBCS	0.40
North	SQGQ1	0.30
North	NorTCS	0.24
North	CAPmax	0.23
North	NorCAPmax	0.21
North	NOAA ERM	0.12
North	NatPmax	-0.14
		Bootstrapped
South	NatPmax	0.53
South	CAPmax	0.52
South	SoCAPmax	0.52
South	SoBCS	0.49
South	NOAA ERM	0.47
South	SQGQ1	0.47
South	Consensus	0.46
South	SoTCS	0.45

Table 6. Classification accuracy of chemical indicators for benthic community condition. Values in the highlighted cells are within the 90th percentile of the highest median correlation for the bootstrapped analysis. Separate analyses were conducted on the north and south validation data sets. Region-specific thresholds were applied to each of the indicators shown.

Region	SQG	Weighted Kappa	% Agreement
		Not Bootstrapped	
North	NorCAPmax		52
North	SQGQ1		52
North	Consensus		44
North	NorBCS		40
North	NOAA ERM		40
North	NorTCS		40
North	CAPmax		36
North	NatPmax		36
		Bootstrapped	
South	SoBCS	0.44	53
South	CAPmax	0.41	49
South	NatPmax	0.40	49
South	SoCAPmax	0.33	45
South	Consensus	0.33	44
South	SoTCS	0.29	41
South	NOAA ERM	0.27	40
South	SQGQ1	0.23	36