Development and evaluation of sediment quality guidelines based on benthic macrofauna responses

Kerry J. Ritter, Steven M. Bay, Robert W. Smith[1], Doris E. Vidal-Dorsch and L. Jay Field[2]

Abstract

Toxicity-based sediment quality guidelines (SQGs) are often used to assess the potential of sediment contamination to adversely affect benthic macrofauna, yet the correspondence of these guidelines to benthic community condition is poorly documented. This study compares the performance of five toxicity-based SQG approaches to a new SQG approach based on changes in benthic community condition. Four of the toxicity-based SQG approaches, effects range median (ERM), logistic regression modeling (LRM), sediment quality guideline quotient 1 (SQGQ1), and Consensus, were derived in previous national studies, and one was developed as a regional variation of LRM calibrated to California data (CA LRM). The new SQG approach, Chemical Score Index (CSI), was derived from southern California benthic community condition data. The chemical-specific guidelines for each approach were applied to an independent validation set of matched chemical concentration, amphipod mortality, and benthic macrofauna abundance data for southern California. Respective results for each SQG approach were then combined into a summary statistic describing the overall contamination magnitude (e.g., mean quotient) and assessed in accordance with a set of thresholds in order to classify stations into four categories of biological effect. Results for each SQG approach were significantly correlated with changes in sediment toxicity and benthic community condition. Furthermore, cumulative frequency plots and effect category thresholds for toxicity and benthic community condition were similar, indicating that both types of effect measure had similar sensitivity of response to contamination level. The benthos-based CSI had the highest percent agreement for benthic community condition; differences were relatively minor by comparison to differences observed for other SQG approaches.

Introduction

Resource managers often use SQGs to interpret contamination data with respect to impacts on sediment dwelling organisms. The most widely used SQGs are empirical guidelines based on the statistical analysis of matched sediment chemistry and biological effects data. Although the benthic community is often the focus of concern in sediment quality assessments, few SQGs based on benthic community effects are available; those that are available are typically based on data sets with limited geographic scope (e.g., Puget Sound, Hong Kong) and may not be applicable to other regions (Barrick *et al*. 1988, Leung *et al*. 2005, Kwok *et al*. 2008). Consequently, SQG development is usually based on the analysis of large databases that use laboratory sediment toxicity tests as the primary measure of biological effects (Wenning *et al*. 2005).

Various statistical approaches have been used to develop empirical SQGs. Commonly used approaches include ERM (Long *et al*. 1995), LRM (Field *et al*. 1999, Field *et al*. 2002), SQGQ1 (Fairey *et al*. 2001), apparent effects threshold (AET; Barrick *et al*. 1988), and consensus (MacDonald *et al*. 2000, Swartz 1999). ERMs are chemical values corresponding to the 50th (median) percentile of biological effects concentrations compiled from a variety of study types and sources. LRM indicators model the probability that a sample will be toxic at a given chemical concentration. An individual LRM is

[1] *Deceased*
[2] *National Oceanic and Atmospheric Administration, Seattle, WA*

developed for each chemical in the analysis. The SQGQ1 is calculated using SQGs compiled from multiple sources. Consensus guidelines combine several types of SQGs in order to reduce the uncertainty of using a single SQG approach.

The association of empirical SQGs with toxicity has been documented in many studies (summarized in Wenning *et al*. 2005). Association is often demonstrated by a high incidence of toxicity when multiple guidelines are exceeded (Long *et al*. 1998). A stronger association is obtained when an integrative SQG index based on an empirical approach, such as the mean ERM quotient (mERMQ), is used to describe the magnitude of contamination (Long *et al*. 2000, Long *et al*. 2006).

The relevance of toxicity-based SQGs for assessing the potential for benthic community effects is uncertain due to a paucity of analyses and differences in methodology between studies. Although studies have demonstrated that contamination indices derived from toxicity-based SQGs are well correlated with benthic community disturbance (Hyland *et al*. 2003, Wenning *et al*. 2005), the characteristics of this relationship have been documented for only a few SQG approaches, often using dissimilar methods of analysis. Recent studies also indicate that benthic community degradation may be occurring at much lower contaminant levels than those associated with increased incidence of toxicity identified by laboratory toxicity tests (Hyland *et al*. 2003, Kwok *et al*. 2008), suggesting that toxicity-based SQGs and commonly used thresholds for their interpretation may not be protective of benthic communities.

Two strategies are available to increase confidence in the assessment of sediment contamination effects on benthic community condition. First, the accuracy of existing toxicity-based SQGs might be improved through site-specific calibration. Second, benthos-based SQGs can be developed for regions of interest; such region-specific indices would then provide more relevant and accurate methods for assessing the potential that chemical contamination will impact benthic communities. The goal of this study was to evaluate these two strategies using California sediment contamination data. This study used a data set of matched chemistry, toxicity, and benthic macrofauna measurements to develop a new sediment quality index based on the response of California benthic community assemblages to contamination. These data were also used to calibrate five types of toxicity-based SQG indices with respect to benthic community responses. Each type of index was then evaluated based on its ability to describe the potential for sediment contamination to impact benthic communities.

METHODS

This study assessed the relationship of six empirical SQG approaches by applying them to matched chemical concentration, amphipod mortality, and benthic macrofauna abundance data for southern California (Figure 1). Five of the approaches were based on toxicity responses: four approaches were developed in previous national studies (ERM, LRM, SQGQ1, and Consensus); one approach was a regional variation calibrated to California data (CA LRM). The sixth approach, CSI, was developed for this study and based on the association between chemical concentration and benthic community disturbance. An index of overall contamination (i.e., mean quotient, mean score, or maximum probability of toxicity) was calculated for each approach. Thresholds relating each SQG index to toxicity or benthic response categories were derived using a standardized statistical approach. Each SQG index was evaluated by determining two measures of

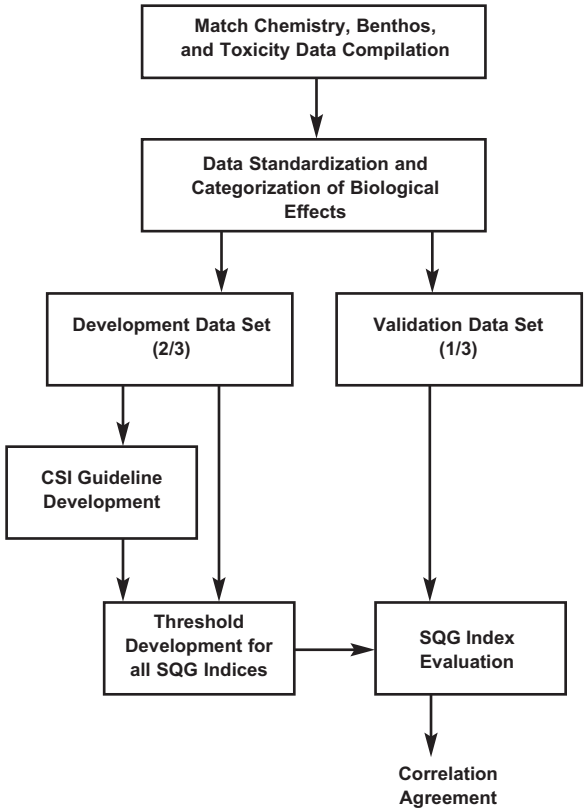

Figure 1. Schematic of data analyses.

association: correlation, and percent agreement between the calculated effect categories and observed biological endpoints.

Data

Data consisting of 441 samples of matched toxicity, chemistry, and benthic macrofauna abundance data were compiled across 6 regional monitoring and research surveys in southern California embayments. The data were screened to select high quality and comparable information. All stations were located in enclosed bays or harbors at subtidal depths, and only data from surficial sediment (top 30 cm or less) were selected. The database included stations from Santa Barbara Harbor, in the north, to San Diego Bay, in the south. More information on the studies used to populate this database can be found at http://www.sccwrp.org/view.php?id=519.

Toxicity data was obtained by solid phase 10-day amphipod survival tests using *Rhepoxynius abronius* or *Eohaustorius estuarius* and standardized methods (USEPA 1994). Toxicity data were screened to ensure that mean control survival was at least 85% and overlying water ammonia concentrations were below species-specific criteria (USEPA 1994). Amphipod survival was normalized to the control response for each test batch by expressing the results as a percentage of the control survival. Each toxicity result was assigned to one of four categories of toxicological response; categories were based on classification systems used in other studies (Long *et al*. 2006). The toxicity categories specific to each test species were based on analyses of minimum significant difference and magnitude of response (percent of control survival; Bay *et al*. 2007). The categories for *E. estuarius* were: Nontoxic (≥90% survival), Low Toxicity (≥82%), Moderate Toxicity (≥59%) and High Toxicity (<59%). The categories for *R. abronius* were: Nontoxic (≥90% survival), Low Toxicity (≥83%), Moderate Toxicity (≥70%) and High Toxicity (<70%). The magnitude of toxicity varied widely in the dataset; however, most samples were classified as Nontoxic or Low Toxicity (Figure 2).

Benthic community condition was determined based on species abundances collected from 1-mm sieves. Taxonomic inconsistencies among programs were eliminated by cross-correlating species lists, identifying differences in nomenclature, and resolving discrepancies by consulting the taxonomists from each program. The benthic community condition of each sample was classified into one of four categories of biological response: Reference, Low Disturbance, Moderate Disturbance, and High Disturbance. The condition categories were determined by the median of four benthic indices calibrated to southern California embayments (Ranasinghe *et al*. 2007): Benthic Response Index (BRI; Smith *et al*. 2001, Smith *et al*. 2003), Index of Benthic Biotic Integrity (IBI; Thompson and Lowe 2004), Relative Benthic Index (RBI; Hunt *et al*. 2001), and River Invertebrate Prediction and Classification System (RIVPACS; Wright *et al*. 1993, Van Sickle *et al*. 2006). All categories of benthic condition were represented in the database, but most samples were classified as Reference or Low Disturbance; only 4% had High Disturbance (Figure 2).

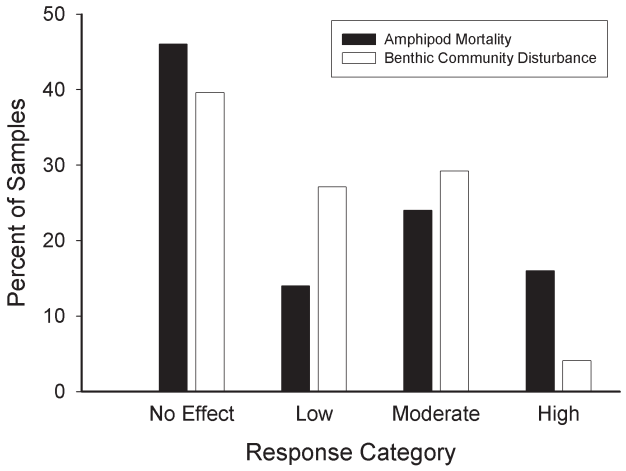

Figure 2. Distribution of sediment toxicity and benthic community disturbance categories in the data set. N = 441.

The chemistry data selected for analysis had to meet several screening criteria that included a review of the data quality assessment by the study authors, use of comparable extraction/digestion methods, and measurement of a minimum suite of contaminants that included multiple metals and polycyclic aromatic hydrocarbons (PAHs). Standardized sums of PAHs, dichloro diphenyl trichloroethane and degradation products (DDTs), polychlorinated biphenyls (PCBs), and chlordanes were calculated using a consistent methodology for all samples. Low molecular weight PAHs were calculated as the sum of acenaphthene, anthracene, biphenyl, naphthalene, 2,6-dimethylnaphthalene, fluorene, 1-methylnaphthalene, 2-methylnaphthalene, 1-methylphenanthrene, and phenanthrene. High molecular weight PAHs was the sum of benzo(a)anthracene, benzo(a)pyrene, benzo(e)pyrene, chrysene, dibenz(a,h)anthracene, fluoranthene, perylene, and pyrene. Total PAHs was

the sum of Low PAH and the High PAH values. Total PCBs was calculated from the sum of congeners 8, 18, 28, 44, 52, 66, 101, 105, 110, 118, 128, 138, 153, 180, 187, and 195. The congener list was a subset of that used by the National Oceanic and Atmospheric Administration (NOAA) Status and Trends program; the sum was multiplied by a correction factor of 1.72 to approximate the value obtained using the larger NOAA list. Total DDTs represented the sum of p,p'-DDT, o,p'-DDT, p,p'-DDE, o,p'-DDE, p,p'-DDD, and o,p'-DDD.

Data were estimated for values reported as below reporting limits based on multiple regression imputation, using SAS PROC MI (SAS Institute Inc, NC,). Imputation produces lesser bias than conventional approaches for estimating nondetect data, such as substituting one-half of the reporting limit (Helsel 2005). Estimation was conducted in a sequential stepwise fashion by contaminant type. Metal data were estimated first, followed in order by pesticides, PAHs, and PCBs. Estimated values were constrained to always be less than the study reporting limit. The imputation method was also used to estimate total organic carbon (TOC) for the purposes of calculating the SQGQ1 and Consensus quotients. Estimated values were not used in calculations for any other analytes missing in the data sets, except when needed to calculate standardized sums of PAHs, PCBs, or pesticides. For example, a value for phenanthrene was estimated for a sample that contained data for other PAHs in order to use the standardized method to calculate the PAH sums, but the estimated phenanthrene value was not used individually to calculate summary SQG values for that sample.

The cumulative frequencies of samples with sediment toxicity or benthic community disturbance (i.e., affected) and those without biological effects (unaffected) were plotted in relation to the mERMQ (see below for calculation method). For this analysis, samples classified as Nontoxic or Low Toxicity were regarded as "unaffected" for sediment toxicity. Samples classified as Reference or Low Disturbance were regarded as unaffected with respect to benthic community condition. Affected samples were defined as those with either a moderate or high category of biological effect.

The resulting data set contained a 1 to 2 orders of magnitude range of concentration for each constituent (Table 1). Approximately two-thirds of the data (293 samples) were used for SQG development and threshold calibration. An independent validation data set, containing one-third of the data (148 samples) randomly selected from throughout the range of benthic community response, was used for performance evaluation of the SQG approaches.

Table 1. Distribution of chemical contaminants in the southern California data set.

Chemical	Units	Percentile			N
		10th	50th	90th	
Cadmium	mg/kg	0.1	0.3	1.1	441
Copper	mg/kg	14.3	68.8	211.0	441
Lead	mg/kg	10.4	34.4	101.0	441
Mercury	mg/kg	<0.1	0.2	0.8	441
Zinc	mg/kg	61.7	164.0	315.0	441
DDD, total	µg/kg	0.1	22.1	1.9	408
DDE, total	µg/kg	0.2	6.1	144.0	408
DDT, total	µg/kg	0.1	1.0	9.1	406
alpha-Chlordane	µg/kg	0.3	1.1	4.2	404
gamma-Chlordane	µg/kg	0.4	1.4	6.2	380
HMW PAH, total	µg/kg	77.3	339.0	3723.0	441
LMW PAH, total	µg/kg	19.1	77.1	525.0	441
PCB, total	µg/kg	6.1	21.1	116.0	441

Toxicity-based Sediment Quality Indices

National SQGs

The ERM values used in the analyses were obtained from Long *et al*. (1995). The mERMQ for each sample in the data set was calculated by dividing each chemical concentration by its respective ERM and averaging the individual quotients (Long *et al*. 2000). The subset of ERM values used to calculate the mERMQ (Table 2) was the same as that used in previous mERMQ performance studies (Long *et al*. 2000). The SQGQ1was calculated as described by Fairey *et al*. (2001). The SQG values used in the analysis are listed in Table 2.

The Consensus SQG values for PAHs and PCBs were midpoint effect concentrations obtained from Swartz (1999) and MacDonald *et al*. (2000), respectively. Values for DDTs, dieldrin, arsenic, cadmium, chromium, copper, lead, mercury, nickel, silver, and zinc were obtained from Vidal and Bay (2005). The mean Consensus quotient was calculated by dividing each chemical concentration by its respective SQG (Table 2) and averaging the individual quotients.

The LRM approach was based on the statistical analysis of paired chemistry and amphipod toxicity data from studies throughout the United States (Field *et al*. 1999, Field *et al*. 2002). The logistic model is described by the following equation:

Table 2. Chemical values for individual sediment quality guidelines used for data analyses. Values for the effects range median (ERM) were taken from Long *et al.* (1995). Mean sediment quality guideline quotient (SQGQ1) values taken from Fairey *et al.* (2001). Consensus midpoint effect concentration values taken from Swartz (1999); MacDonald *et al.* (2000); and Vidal and Bay (2005). Concentrations are on a dry weight basis except where noted.

Chemical	Units	ERM	SQGQ1	Consensus
Arsenic	mg/kg	70.0		55.0
Cadmium	mg/kg	9.6	4.2	5.9
Chromium	mg/kg	370.0		224.9
Copper	mg/kg	270.0	270.0	225.0
Lead	mg/kg	218.0	112.2	222.3
Mercury	mg/kg	0.7		0.6
Nickel	mg/kg	51.6		67.6
Silver	mg/kg	3.7	1.8	3.4
Zinc	mg/kg	410.0	410.0	357.1
2-Methylnaphthalene	µg/kg	670.0		
Acenaphthene	µg/kg	500.0		
Acenaphthylene	µg/kg	640.0		
Anthracene	µg/kg	1100.0		
Benzo(a)anthracene	µg/kg	1600.0		
Benzo(a)pyrene	µg/kg	1600.0		
Chrysene	µg/kg	2800.0		
Dibenz(a,h)anthracene	µg/kg	260.0		
Dieldrin	µg/kg	8.0	8.0	7.0
Fluoranthene	µg/kg	5100.0		
Fluorene	µg/kg	540.0		
Naphthalene	µg/kg	2100.0		
p,p'-DDE	µg/kg			
Phenanthrene	µg/kg	1500.0		
Pyrene	µg/kg	2600.0		
Total Chlordane	µg/kg		6.0	
DDTs, total	µg/kg	46.1		25.4
PAH, total	µg/kg		1,800.0*	1,800.0*
PCB, total	µg/kg	180.0	400.0	0.5

* µg/g organic carbon basis

selected for use (Table 3). The maximum probability of toxicity obtained from the individual models (P_{max}) for each sample was used as the index of overall contamination.

Regional SQGs

A version of the LRM approach calibrated using California data (CA LRM) was also evaluated, as recent analyses showed that the CA LRM had greater association with toxicity in southern California sediments relative to other SQG approaches (Bay *et al*. 2008). Regional calibration consisted of developing logistic regression models for individual chemicals based on California data sets using the methods described in USEPA (2005). These models used <80% control-adjusted amphipod survival as the definition of a toxic sample. The specific models included in the CA LRM approaches were selected from a library of candidate models that included national models (Field *et al*. 2002), as well as models derived using the California data sets. The selected models were chosen based on the goodness of fit with the observed probability of toxicity (Table 3). Models with high false positive rates were not included.

Benthos-based Sediment Quality Index

CSI was derived using matched chemistry and Southern California benthic macrofauna data. The CSI describes the overall level of chemical exposure in terms of the potential for benthic community disturbance. Two types of data are combined to calculate the CSI: a set of individual chemical exposure scores that describe the association between the concentration and benthic community disturbance, and a set of weighting factors that reflect the relative strength of each chemical-benthos association. The index is the weighted average of scores for 13 chemical contaminants:

$$p = e^{B0+B1\,(x)} / (1 + e^{B0+B1\,(x)})$$

where: p = probability of observing a toxic effect; B0 = intercept parameter; B1 = slope parameter; and x = concentration or log concentration of the chemical.

The chemical-specific models used in this study were based on an analysis of the accuracy for predicting toxicity for 37 candidate models. Models for 18 chemicals having low rates of false positives were

$$CSI = (\Sigma w * score) / \Sigma w$$

where: score = chemical score based on the concentration of a contaminant, and w = weighting factor for that contaminant.

Chemical scores are determined by comparing each chemical concentration to a set of three chemical-specific guidelines (G1, G2, and G3) that classify chemical concentration into one of four exposure categories that correspond to four levels of commu-

Table 3. Logistic regression parameters for the regional and national approaches compared in this study. National values are taken from Field *et al*. (2002), CA LRM values are from Bay *et al*. (2008). B0 = intercept; B1 = slope; T50 = calculated concentration corresponding to a toxicity probability of 0.5. Concentrations are on a dry weight basis.

Chemical	Units	LRM			CA LRM		
		B0	B1	T50	B0	B1	T50
Cadmium	mg/kg	-0.34	2.51	1	0.29	3.18	1
Copper	mg/kg				-5.59	2.59	145
Lead	mg/kg	-5.45	2.77	94	-4.72	2.84	46
Mercury	mg/kg				-0.06	2.68	1
Zinc	mg/kg	-7.98	3.34	245	-5.13	2.42	132
1-Methylnaphthalene	µg/kg	-4.14	2.1	94			
1-Methylphenanthrene	µg/kg	-3.59	1.75	112			
2,6-Dimethylnaphthalene	µg/kg	-4.05	1.9	133			
2-Methylnaphthalene	µg/kg	-3.76	1.78	128			
Acenaphthene	µg/kg	-3.62	1.75	116			
Acenaphthylene	µg/kg	-2.96	1.38	140			
Benzo(b)fluoranthene	µg/kg	-4.54	1.49	1107			
Biphenyl	µg/kg	-4.11	2.21	73			
alpha-Chlordane	µg/kg				-3.41	4.46	6
Dieldrin	µg/kg	-1.17	2.56	3	-1.83	2.59	5
Fluoranthene	µg/kg	-4.46	1.48	1034			
Fluorene	µg/kg	-3.71	1.81	114			
HMW PAH, total	µg/kg				-8.19	2	12506
LMW PAH, total	µg/kg				-6.81	1.88	4127
Naphthalene	µg/kg	-3.78	1.62	217			
trans-Nonachlor	µg/kg				-4.26	5.31	6
p,p'-DDD	µg/kg	-1.9	1.49	19			
p,p'-DDT	µg/kg				-3.55	3.26	12
Phenanthrene	µg/kg	-4.46	1.68	455			
PCB, total	µg/kg	-3.46	1.35	368	-4.41	1.48	945

nity disturbance defined by the BRI: Reference (≤G1), Low Disturbance (>G1 and ≤G2), Moderate Disturbance (>G2 and ≤G3), and High Disturbance (>G3). The chemical exposure categories are given a numerical score of 1, 2, 3, or 4 corresponding to a category of Reference, Low, Moderate or High, respectively. The chemical score for each contaminant is then multiplied by a weighting factor that corresponds to the strength of association between the chemical and biological effect. These products are then summed across the 13 constituents in the sample and divided by the sum of weighting factors, producing the mean weighted CSI. An example of the CSI calculation is shown in Table 4.

The CSI chemical-specific guidelines were selected by applying an optimization procedure based on maximizing the agreement between chemical scores and the four effect levels of the benthic response index (Ranasinghe *et al*. 2007). Agreement was measured by the weighted kappa statistic (Cohen 1960, 1968) with weights based on the linear weighting scheme of Cicchetti and Allison (1971). These weights are commonly used with ordinal data (i.e., ordered categorical data) and are the default in many statistical programs including SAS. With this weighting scheme, the highest (full) weight is given to perfect agreement (e.g., chemical score = 2 and BRI = Low Disturbance. A relatively high weight (i.e., more partial credit) is given when the disagreement is relatively small (e.g., chemical score = 2 and BRI = Moderate Disturbance), while a lower weight is given when disagreement is relatively large (e.g. chemical score = 2 and BRI = High Disturbance).

The guideline optimization procedure was conducted on multiple subsamples of the development data set. The subsampling procedure relied on bootstrapping, where approximately even distributions of

Table 4. Example of the calculation of the CSI from a chemistry data set; the chemical scores were obtained using the guidelines listed in Table 5.

Chemical	Units	Concentration	Chemical Score	Weight	Weight x Score
Cadmium	mg/kg	0.3	3	38.33	114.99
Copper	mg/kg	100.0	3	100.00	300.00
Lead	mg/kg	25.0	1	88.37	88.37
Mercury	mg/kg	0.3	2	30.31	60.62
Zinc	mg/kg	500.0	3	97.85	293.55
HMW PAH, total	ug/kg	450.0	2	16.12	32.24
LMW PAH, total	ug/kg	125.0	2	5.16	10.32
alpha-Chlordane	ug/kg	1.0	2	55.00	110.00
gamma-Chlordane	ug/kg	1.5	3	58.48	175.44
DDD, total	ug/kg	2.5	2	46.19	92.38
DDE, total	ug/kg	100.0	3	31.41	94.23
DDT, total	ug/kg	6.5	3	16.27	48.81
PCB, total	ug/kg	20.0	2	55.18	110.36
Sum				638.67	1531.31

CSI = $\sum$(w * score)/$\sum$w = 1531.31/638.67 = 2.4

unaffected (i.e., Reference or Low BRI category) and affected (Moderate or High BRI category) samples were randomly subsampled for analysis. This step was necessary in order to reduce the tendency of skewed data distributions to provide misleading kappa results (Kraemer and Bloch 1988, Fienstein and Cicchetti, 1990). Kappa is least susceptible to providing misleading results when distributions are even (Canran *et al*. 2005, Lantz and Nebenzahl 1996). PROC SURVEYSELECT was used to randomly select (with replacement) 60 affected and 60 unaffected samples from subsampled data. The subsampling process, (i.e., bootstrapping) was repeated 50 times in order to ensure that nearly all samples were represented in at least one subset.

The optimal set of three guideline values (G1, G2, G3) for each chemical in each bootstrap sample was selected by computing weighted kappa statistics for a large set of possible candidates. These candidates were selected by choosing all permutations of three guidelines, taken at 5% increments of the concentration range. In addition, distances between individual guidelines within each set were constrained to be no less than 10% of the range for each chemical. These conditions ensured that optimization converged and resulting guidelines within a set were not too close to one another. For each data subset, the set of three guidelines that yielded the highest kappa statistic was selected. The median of the guidelines across all subsets was chosen to be the final guideline value for the Low (G1), Moderate (G2), and High (G3) categories.

The CSI weighting factor for each chemical was determined by applying the final set of guidelines to all 50 bootstrap samples and computing the weighted kappa for each sample. The median kappa was calculated, representing the overall relative strength of association between the variations in chemical concentration and benthic community condition. The kappa values were then normalized to the largest median kappa across 13 chemicals for use as weighting factors in calculation of the CSI.

Threshold development for all SQG indices

Determining the agreement of sample classification based on SQGs with those based on toxicity or benthos requires the application of category thresholds for each SQG index. Such thresholds are generally unavailable for these SQG approaches or vary in the method of development. The thresholds used in this study were developed for each SQG approach using a consistent methodology so that differences in performance would reflect inherent differences among approaches, rather than variations in how thresholds were assigned.

Two sets of three thresholds were established for each SQG approach. One set of thresholds defined four ranges of index values that corresponded to toxicity categories based on amphipod survival: Nontoxic,

Low Toxicity, Moderate Toxicity, and High Toxicity. The other set of thresholds was based on benthic community condition categories (median of four indices): Reference, Low Disturbance, Moderate Disturbance, and High Disturbance. Each set of thresholds was selected on the basis of optimizing the percent agreement between the four index ranges and the corresponding biological effect categories and used the same methodology as described for CSI guideline selection.

Selection of the benthic community-based SQG index thresholds used the SQG development data set described previously. The toxicity-based thresholds were based on a larger development data set (n = 887), which was used for a related study to calibrate and evaluate various toxicity-based SQG indices (Bay *et al*. 2008). This data set contained many of the same samples contained in the benthic community data set and also additional samples from the same water bodies for which benthic community data were not available. This larger data set was used to provide the most robust set of threshold optimization analyses possible and to enhance comparability between the two studies.

SQG index evaluation

SQG index performance was evaluated by quantifying the strength of association between chemistry and biological effect (i.e., toxicity or benthic community condition) in terms of both correlation and categorical classification accuracy. Correlation was measured as the nonparametric Spearman's correlation coefficient between the SQG index value (i.e., mean quotient, mean score, or P_{max}) and amphipod mortality or community condition category.

Categorical classification accuracy was quantified as the percent agreement between the four SQG index categories (determined by applying the thresholds derived using the development data set) and the biological effect categories (toxicity or benthos). Percent agreement was calculated as:

$$A = (Nc/Nt)*100$$

where: A = percent agreement;
Nc = number of samples correctly classified;
and Nt = total number of samples.

The percent agreement analysis assessed agreement between two sets of four categories, resulting in an expected value of 25% when the agreement was solely due to chance. Values above 25% indicated the magnitude of improvement in classification accuracy due to the SQG index.

The analyses were conducted across 50 bootstrap samplings of 40 unaffected and 40 affected samples. The 90th percentile confidence limits of the highest correlation or percent agreement value were determined from the bootstrap samplings and used to identify the best performing SQG indices. All analyses were conducted using the same independent validation data set, which consisted of data that was not used for the development of benthic community-based guidelines or SQG indices.

RESULTS

CSI Development

There was substantial variability in the relationship between individual chemical concentrations and benthic community condition. Spearman correlation coefficients were highest for some metals (e.g., copper and zinc) and lowest for DDE and DDT (Figure 3). Markedly different correlations were obtained for some contaminants when compared to toxicity. For example, increased cadmium concentration showed a relatively high correlation with increased amphipod mortality but a low correlation with increased benthic community disturbance.

The plots of the cumulative frequencies of stations with and without biological effects against the mERMQ showed a high degree of similarity for toxi-

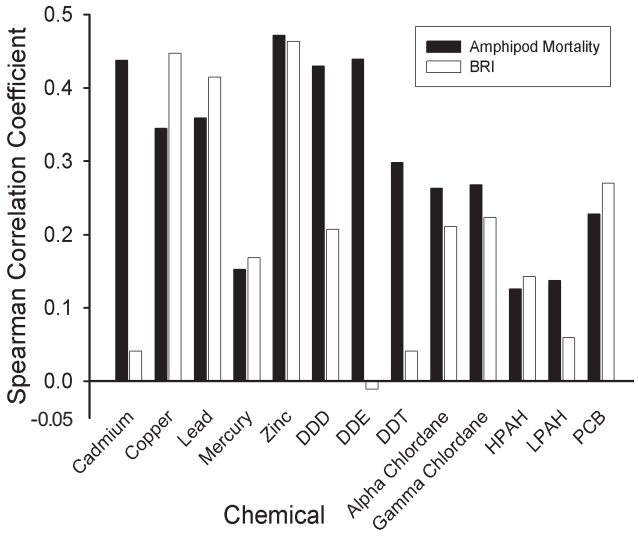

Figure 3. Correlation of chemical concentration with toxicity or benthic community disturbance.

city and benthic condition (Figure 4). Within each type of biological effect measure, values of the mERMQ were always higher for affected samples at equivalent cumulative frequency values. The difference in quotients (i.e., mERMQaffected – mERMQunaffected) for a given percentage was also similar between each type of effects measure, indicating a similar ability of the mERMQ to discriminate between affected and unaffected samples. For example, at a cumulative percentage 50%, the difference in quotients for toxicity and benthic condition was 0.062 and 0.072, respectively. Quotients corresponding to low, moderate, and high probabilities of effects (i.e., 10, 50, 90% cumulative frequency) were also similar for toxicity and benthic condition. For example, quotients corresponding to the 10th percentile of effects were 0.083 and 0.074 for toxicity and benthos respectively. The main differences between the frequency plots occurred at a mERMQ of ≤0.07, were there was a somewhat higher proportion of affected benthos than affected toxicity samples and also relatively little discrimination between affected and unaffected benthic samples.

Plots of representative results from the CSI guideline development analyses are shown in Figure 5 for copper, which had one of the strongest correlations, and high molecular weight PAHs (weak correlation). Analyses for each contaminant usually showed a trend of greater frequency and severity of community disturbance at higher concentrations. A poor relationship was present at low chemical concentrations, where benthic community condition was highly variable, ranging from Reference to High Disturbance (Figure 5). High BRI scores at low concentrations may indicate the presence of samples impacted by stress due to noncontaminant factors, such physical disturbance.

A set of guideline values was developed for 13 contaminants using the statistical optimization procedure (Table 5). The guidelines for copper, lead, and zinc produced the highest kappa values, indicating that classifications based on these metals had the greatest concordance with benthic community disturbance (BRI score) than guidelines for other chemicals. The lowest kappa values and therefore the least concordance with benthic community condition were obtained for the high and low molecular weight PAHs and DDT. Variation in kappa values was generally similar to the pattern observed for Spearman correlation coefficients. The normalized weighting factor for each contaminant, which was based on kappa, varied from 100 (copper) to 5 (low molecular weight PAHs).

SQG Index Associations with Toxicity and Benthos

All of the SQG indices were significantly correlated with changes in sediment toxicity and benthic

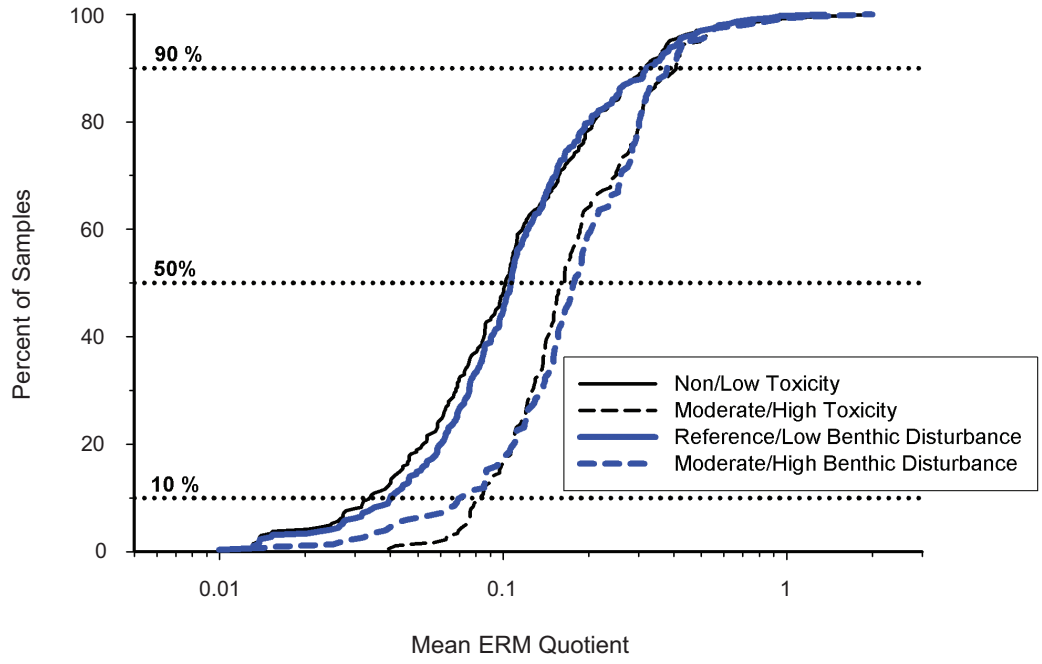

Figure 4. Cumulative percentage of samples showing toxicity or benthic community disturbance versus mean ERM quotient.

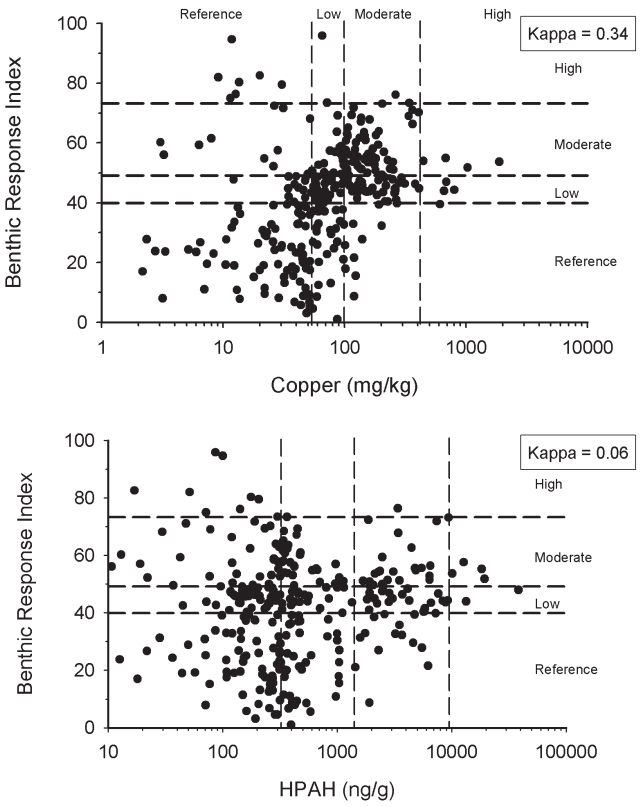

Figure 5. BRI data distributions and CSI guideline values (vertical dashed lines) for copper and high molecular weight PAHs.

community condition (Table 6). The correlations were similar, both among indices and between biological effects measures. The toxicity correlations ranged from 0.42 (SQGQ1) to 0.55 (CA LRM), while the correlations with benthic community condition ranged from 0.46 (Consensus) to 0.53 (National LRM). The correlation between benthic condition and the new CSI approach was intermediate (0.49).

Each of the indices was also significantly correlated with each other. Pairwise correlations ranged from 0.73 (Consensus vs. SQGQ1) to 0.95 (Consensus vs. mERMQ). Correlations of the CSI with the other indices ranged from 0.80 to 0.88.

While overall correlations were similar, the toxicity and benthos data showed a different distribution with respect to the SQG index values. Results for the mERMQ (Figure 6) are typical of those for the other indices. At low index values (e.g., mERMQ <0.1), a low incidence of toxic samples is evident but a greater incidence of samples with an affected benthic community is evident. An opposite pattern is present among samples with the highest mERMQ values: the incidence of affected benthic samples is greater than the incidence of toxic samples.

The SQG index category thresholds developed based on benthic condition were similar to the toxicity-based thresholds (Table 7). The toxicity-based thresholds were almost always lower than those based on benthos, but the differences were small (usually less than 10%). The two High SQGQ1 thresholds showed the greatest variation: 0.80 for High Toxicity versus 1.26 for High Disturbance.

Each of the SQG approaches had similar percent agreement for each type of biological effect (Table 6). All of the agreement values were substantially higher than 25%, the value expected solely on the basis of

Table 5. Guideline values and weighting factors used in the calculation of the mean CSI score. Weights are based on the kappa obtained during threshold optimization for each chemical, normalized to the largest kappa value.

Chemical	Units	Guideline			Kappa	Weight
		Low (G1)	Moderate (G2)	High (G3)		
Cadmium	mg/kg	0.09	0.22	1.66	0.13	38
Copper	mg/kg	52.80	96.50	406.00	0.34	100
Lead	mg/kg	26.40	60.80	154.00	0.30	88
Mercury	mg/kg	0.09	0.45	2.18	0.10	30
Zinc	mg/kg	112.00	200.00	629.00	0.33	98
DDD, total	µg/kg	0.38	2.69	117.00	0.16	46
DDE, total	µg/kg	0.11	4.15	153.00	0.11	31
DDT, total	µg/kg	0.42	1.52	89.30	0.06	16
alpha-Chlordane	µg/kg	0.50	1.23	11.10	0.19	55
gamma-Chlordane	µg/kg	0.54	1.45	14.50	0.20	58
HMW PAH, total	µg/kg	312.00	1325.00	9320.00	0.05	16
LMW PAH, total	µg/kg	85.40	312.00	2471.00	0.02	5
PCB, total	µg/kg	11.90	24.70	288.00	0.19	55

Table 6. Correlation and classification accuracy of SQG index values with respect to toxicity or benthic community condition. Cells marked with an asterisk are within the 90th percentile of the largest value across 50 bootstrap samplings.

SQG Approach	Toxicity		Benthos	
	Correlation (r)	Agreement (%)	Correlation (r)	Agreement (%)
CSI[1]	0.54*	50*	0.49*	53*
CA LRM	0.55*	44	0.52*	49*
National LRM	0.45	46*	0.53*	49*
National ERM	0.45	49*	0.47*	40
Consensus	0.53*	47*	0.46	44
SQGQ1	0.42	40	0.47*	36

[1] All thresholds for CSI were based on benthic community condition.

chance. The CSI had the greatest percent agreement for both toxicity (50%) and benthic community condition (53%). The SQGQ1 had the lowest percent agreement for each effects measure. Results for the other SQG indices varied within a relatively narrow range. There was no consistent trend in percent agreement between the effects measures. Percent agreement was higher for benthos than for toxicity for CSI, CA LRM, and National LRM, but the trend was reversed for the other SQG approaches.

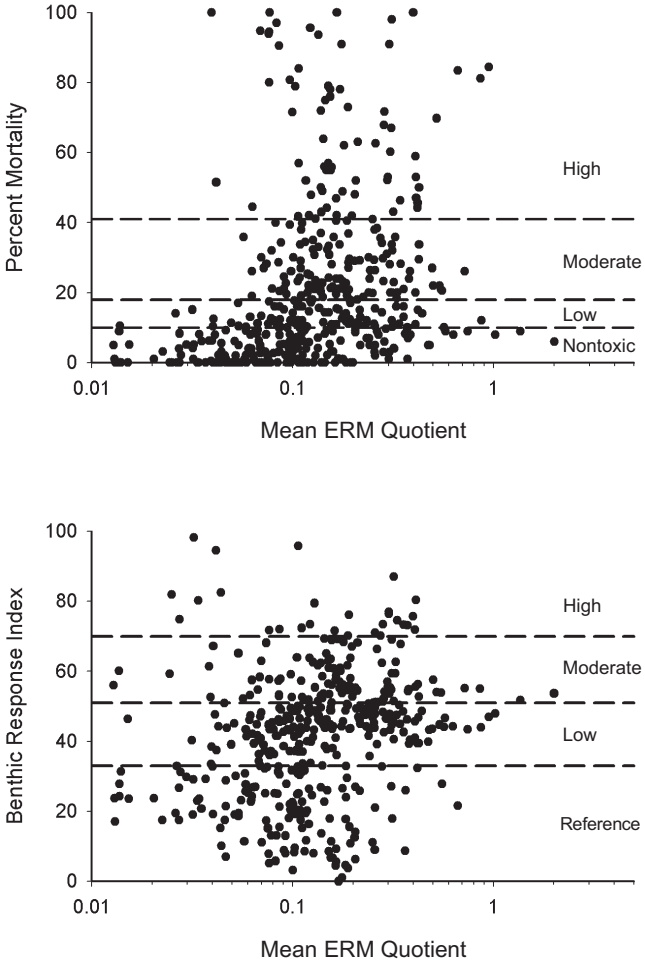

Figure 6. Biological effect data distributions relative to mean ERM quotient.

Discussion

Each of the SQG indices investigated in this study showed utility with respect to describing the potential for impacts on the benthic community, regardless of whether the SQGs were developed and calibrated to toxicity or benthos. This effectiveness was demonstrated by the presence of significant correlations and moderate percent agreements (>25% expected from chance) among four levels of biological impact. These results are consistent with studies conducted in other locations in the United States (Long *et al*. 2006). Mean SQG quotients based on ERMs have been shown to be predictive of marine benthic community disturbance in Florida (MacDonald *et al*. 2004), along the Atlantic and Gulf Coasts (Hyland *et al*. 2003), and in Washington (Long *et al*. 2005). This study documents for the first time that the new CSI index, as well as SQG indices derived from logistic regression models and consensus guidelines, relate comparably to benthic community impacts.

Each of the SQG indices illustrated similar levels of association with toxicity and benthic community disturbance, regardless of the SQG approach (e.g., ERM, LRM) or type of biological data underlying the guidelines (i.e., toxicity or benthos). Such similarity may have been related to two factors: 1) the robust nature and general applicability of empirical SQGs, and 2) the use of consistent methods for

Table 7. SQG index thresholds based on toxicity or benthic community condition. na = not applicable; CSI thresholds were not developed for toxicity.

SQG Approach	Index	Toxicity			Benthos		
		Low	Moderate	High	Low	Moderate	High
CSI	Weighted Mean Score	na	na	na	1.69	2.34	3.00
CA LRM	Maximum Probability	0.42	0.58	0.72	0.43	0.60	0.78
National LRM	Maximum Probability	0.23	0.44	0.61	0.22	0.43	0.65
National ERM	Mean Quotient	0.06	0.12	0.38	0.06	0.13	0.40
Consensus	Mean Quotient	0.14	0.26	0.60	0.15	0.30	0.68
SQGQ1	Mean Quotient	0.16	0.34	0.80	0.13	0.37	1.26

index evaluation. Since the empirical SQGs were developed using large and diverse data sets, these SQGs tend to describe general trends in the data set that are common to multiple regions rather than site-specific variations in contaminant mixtures or bioavailability. As a result, contamination indices based on these SQGs (e.g., mean ERM quotient) would be expected to have a similar (albeit relatively low) degree of association with various measures of biological effect. The use of consistent methods for threshold selection and SQG index evaluation in this study minimized the potential for differences in analysis method to influence the evaluation results. Had thresholds obtained from the literature been used for evaluating the national SQG indices, it is likely that the agreement with biological effects would have been lower and not representative of the maximum performance of each approach. For example, a mERMQ of >1.5 has been identified as indicative of a high risk of biological effects (Long *et al*. 2000) yet the threshold providing the highest concordance between mERMQ and the highest category of southern California sediment toxicity was much lower (0.4).

The cumulative frequency and classification thresholds for toxicity and benthic community condition were nearly identical, indicating that these laboratory and field biological effects measures had similar sensitivity in their response to sediment contamination. This finding differs from other studies, where benthic community responses were reported to be more sensitive. For example, Hyland *et al*. (2003) determined a very high level of risk to the benthos occurred at a mERMQ of >0.361, whereas Long *et al*. (1998) determined that a mERMQ of >1.5 corresponded to a high risk of toxicity. Kwok *et al*. (2008) determined probable effect concentrations for Hong Kong benthos that were lower than most current SQGs.

The disparity in relative sensitivity reported in other studies may be related to several factors. One factor is a difference in the source of the toxicity response thresholds used for comparison. Both Hyland *et al*. and Kwok *et al*. compared benthic community effect thresholds that they derived from regional data sets to toxicity effect thresholds derived by others from different data sets and regions. The toxicity and benthic community thresholds compared in this study were developed using similar data sets and identical statistical approach.

Differences in relative sensitivity of toxicity or benthic community SQG thresholds among studies may also be related to variations in test species or benthic community indices. The toxicity data used by Long *et al*. (1998) to identify SQG thresholds is dominated by survival tests using the marine amphipod *Ampelisca abdita*, whereas the southern California analyses were based on the survival of two other species: *E. estuarius* and *R. abronius*. Tests of split samples indicate that *A. abdita* is less responsive to sediment contamination than *E. estuarius* or *R. abronius* (Bay *et al*. 2007), suggesting that the similarity of toxicity and benthic based thresholds in this study may be due to the use of a more sensitive test species. Different benthic indices were used in each study and their relative sensitivity to contamination may potentially influence threshold selection.

The CSI had the highest percent agreement of the SQG indices evaluated, although the improvement was relatively small compared to the LRM

approach. The high percent agreement for CSI may have been related to several unique attributes: the approach was developed using southern California data and therefore incorporated regional differences in contaminant mixtures or bioavailability, the approach was calibrated directly to a measure of benthic community response instead of toxicity, and finally, the contribution of individual contaminants to the final index value was weighted in proportion to the strength of association with biological effects.

The results of this study suggest three strategies for increasing the utility of SQG indices in the evaluation of sediment quality with respect to benthic community effects. First, the performance of established approaches such as the ERM can be enhanced through the use of thresholds calibrated to local benthic community response patterns. A second strategy is to use a SQG index developed specifically for benthic macrofauna responses, instead of a toxicity-based index. A third strategy consists of using a combination of SQG approaches, where multiple SQG indices based on either toxicity or benthos are combined to describe the overall potential for biological effects. A similar multiple-index approach is frequently used in sediment quality assessments, where various toxicity-based chemical indices are integrated in order to interpret sediment chemistry data (Chapman *et al*. 2002, McDonald *et al*. 2007).

Integrating the results of various types of SQG indices is likely to be the most reliable strategy for assessing the potential for sediment contamination to impact benthic communities. Use of this strategy strengthens the conceptual relationship between sediment contamination and benthic community response, while preserving the ability of the analysis to assess the potential for toxicity, an independent and widely used measure of contaminant effects. The need for a combined analysis supported by the results of this study, which found different chemical-specific and station-specific patterns of response for toxicity and benthic community condition. The greater effectiveness of a multiple index approach has recently been documented for the analysis of benthic community condition, where the integration of results from multiple indices provided a more accurate determination of benthic community condition compared to the use of a single index (Ranasinghe *et al*. 2007). Such a multiple-index approach may provide a more confident method to assess the significance of chemical contamination, considering the limitations of current SQG approaches to predict the bioavailability and biological effects of contaminants under diverse conditions.

Literature Cited

Barrick, R., S. Becker, L. Brown, H. Beller and R. Pastorok. 1988. Sediment quality values refinement: 1988 update and evaluation of Puget Sound AET, Volume 1. PTI Environmental Services. Bellevue, WA.

Bay, S., D. Greenstein and D. Young. 2007. Evaluation of methods for measuring sediment toxicity in California bays and estuaries. Technical Report 503. Southern California Coastal Water Research Project. Costa Mesa, CA.

Bay, S.M., K.J. Ritter, D.E. Vidal-Dorsch and L.J. Field. 2008. Comparison of national and regional sediment quality guidelines for classifying sediment toxicity in California. pp.79-90 *in*: S.B. Weisberg and K. Miller (eds.), Southern California Coastal Water Research Project Annual Report 2008.

Canran, L., P.M. Berry, T.P. Dawson and R.G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385-393.

Chapman, P.M., B.G. McDonald and G.S. Lawrence. 2002. Weight-of-evidence issues and frameworks for sediment quality (and other) assessments. *Human and Ecological Risk Assessment* 8:1489-1515.

Cicchetti, D.V. and T. Allison. 1971. A new procedure for assessing reliability of scoring EEG sleep recordings. *American Journal of EEG Technology* 11:101-109.

Cohen J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37-46.

Cohen, J. 1968. Weighted Kappa nominal scale agreement with provision for scale disagreement or partial credit. *Psychological Bulletin* 70:213-220.

Fairey, R., E.R. Long, C.A. Roberts, B.S. Anderson, B.M. Phillips, J.W.. Hunt, H.R. Puckett and C.J. Wilson. 2001. An evaluation of methods for calculating mean sediment quality guideline quotients as indicators of contamination and acute toxicity to amphipods by chemical mixtures. *Environmental Toxicology and Chemistry* 20:2276-2286.

Field, L.J., D.D. MacDonald, S.B. Norton, C.G. Ingersoll, C.G. Severn, D. Smorong and R. Lindskoog. 2002. Predicting amphipod toxicity from sediments using Logistic Regression Models. *Environmental Toxicology and Chemistry* 9:1993-2005.

Field L.J., D. MacDonald, S.B. Norton, C.G. Severn and C.G. Ingersoll. 1999. Evaluating sediment chemistry and toxicity data using logistic regression modeling. *Environmental Toxicology and Chemistry* 18:1311-1322.

Fienstein, A.R. and D.V. Cicchetti. 1990. High Agreement but low Kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology* 43:6:543-549.

Helsel, D. 2005. More than obvious: better methods for interpreting nondetect data. *Environmental Science & Technology* 39:419A-423A.

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. 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. *Environmental Toxicology and Chemistry* 20:1252-1265.

Hyland, J.L., W.L. Balthis, V.D. Engle, E.R. Long, J.F. Paul, J.K. Summers and R.F. Van Dolah. 2003. Incidence of stress in benthic communities along the U.S. Atlantic and Gulf of Mexico coasts within different ranges of sediment contamination from chemical mixtures. *Environmental Monitoring and Assessment* 81:149-161.

Kraemer, H.C. and D.A. Bloch. 1988. Kappa coefficients in epidemiology: an appraisal of a reappraisal. *Journal of Clinical Epidemeiology* 41:959-968.

Kwok, K.W.H, A. Bjorgesæter, K.Y. Leung, G.C.S. Lui, J.S. Gray, P.K.S. Shin and P.K.S. Lam. 2008. Deriving site-specific sediment quality guidelines for Hong Kong marine environments using field-based species sensitivity distributions. *Environmental Toxicology and Chemistry* 27:226-234.

Lantz, C.A. and E. Nebenzahl. 1996. Behavior and interpretation of the kappa statistic: Resolution of the two paradoxes. *Journal of Clinical Epidemiology* 49:431-434.

Leung, K.M.Y., A. Bjorgesater, J.S. Gray, W.K. Li, G.C.S. Lui, U. Wang and P.K.S. Lam. 2005. Deriving sediment quality guidelines from field-based species sensitivity distributions. *Environmental Science & Technology* 39:5148-5156.

Long, E.R., M. Dutch, S. Aasen, K. Welch and M.J. Hameedi. 2005. Spatial extent of degraded sediment quality in Puget Sound (Washington state, USA) based upon measures of the sediment quality triad. *Environmental Monitoring and Assessment* 111:173-222.

Long, E.R., L.J. Field and D.D. MacDonald. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environmental Toxicology and Chemistry* 17:714-727.

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

Long, E.R., D.D. MacDonald, C.G. Severn and C.B. Hong. 2000. Classifying the probabilities of acute toxicity in marine sediments with empirically-derived sediment quality guidelines. *Environmental Toxicology and Chemistry* 19:2598-2601.

Long, E.R., D.D. MacDonald, S.L. Smith and F.D. Calder. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environmental Management* 19:81-97.

MacDonald, D.D., L.M. Di Pinto, L.J. Field, C.G. Ingersoll, E.R. Long and R.C. Swartz. 2000. Development and evaluation of consensus-based sediment effect concentrations for polychlorinated biphenyls (PCB). *Environmental Toxicology and Chemistry* 19:1403-1413.

McDonald, B.G., A.M.H. de Bruyn, B.G. Wernick, J. Patterson, N. Pellerin and P. Chapman. 2007. Design and application of a transparent and scalable weight-of-evidence framework: An example from Wabamun Lake, Alberta, Canada. *Integrated Environmental Assessment and Management* 3:476-483.

MacDonald, D.D., C.G. Ingersoll, D.E. Smorong, H. Greening, R. Pribble, T. Janicki, S. Janicki, S. Grabe, B. Sloane, D. Eckenrod, E.R. Long and R.S. Carr. 2004. Development of an ecosystem-based framework for assessing and managing sediment quality

conditions in Tampa Bay, Florida. *Archives of Environmental Contamination and Toxicology* 46:147-161.

Ranasinghe, J.A., S.B. Weisberg, R.W. Smith, D.E. Montagne, B. Thompson, J.M. Oakden, D.D. Huff, D.B. Cadien, R.G. Velarde and K.J. Ritter. 2007. Evaluation of five indicators of benthic community condition in two California bay and estuary habitats. Technical Report 524. Southern California Coastal Water Research Project. Costa Mesa, CA.

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. Technical Report 410. Southern California Coastal Water Research Program. Westminster, CA.

Swartz, R.C. 1999. Consensus sediment quality guidelines for PAH mixtures. *Environmental Toxicology and Chemistry* 18:780-787.

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

US Environmental Protection Agency (USEPA). 1994. Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods. EPA 600-R94-025. USEPA Office of Research and Development. Washington, DC.

USEPA. 2005. Predicting toxicity to amphipods from sediment chemistry (Final Report). EPA/600/R-04/030. USEPA National Center for Environmental Assessment Office of Research and Development. Washington, DC.

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

Vidal, D.E. and S.M. Bay. 2005. Comparative sediment guideline performance for predicting sediment toxicity in southern California, USA. *Environmental Toxicology and Chemistry* 24:3173-3182.

Wenning, R.J., G.E. Batley, C.G. Ingersoll and D.W. Moore (eds.). 2005. Use of sediment quality guidelines (SQGs) and related tools for the assessment of contaminated sediments. Society of Environmental Toxicology and Chemistry. Pensacola, FL.

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.

Acknowledgements

The authors thank Chris Beegan from the California Water Resources Control Board, and Mike Connor and Bruce Thompson of the San Francisco Estuary Institute for their suggestions on the design of this study. Peggy Myre of Exa Data and Mapping and Ananda Ranasinghe compiled and standardized the data sets. Darrin Greenstein, Jeff Brown, and Diana Young assisted with data analysis. The authors also thank Peter Landrum, Ed Long, Todd Bridges, Tom Gries, Rob Burgess and Bob Van Dolah for their thoughtful review of the ideas contained within the document. Work on this project was funded by the California State Water Resources Control Board under agreement 01-274-250-0.