Integrated Environmental Assessment and Management — Volume 8, Number 4—pp. 589–596

Framework for Interpreting Sediment Quality Triad Data

Steven M Bay and Stephen B Weisberg*

Southern California Coastal Water Research Project, 3535 Harbor Blvd., Suite 110, Costa Mesa, CA 92626, USA

(Submitted 30 January 2009; Returned for Revision 1 April 2009; Accepted 17 July 2010)

Editor's Note

This article represents 1 of 6 papers describing development and evaluation of a sediment quality assessment framework to support implementation of California's new sediment quality objectives for bays and estuaries, which became effective in 2009. Over thirty scientists collaborated on this effort by the California State Water Resources Control Board, which resulted in the establishment of one of the first statewide programs in the US to fully incorporate the sediment quality triad for regulatory applications.

ABSTRACT

Integration of multiple lines of evidence (MLOE) data in a sediment quality triad assessment can be accomplished by means of numerous approaches, with most relying on some form of expert best professional judgment. Best professional judgment (BPJ) can be problematic in application to large data sets or in a regulatory setting where the assessment protocol needs to be transparent and consistently reproducible. We present a quantitative, objective framework for integrating the results of triad-based assessments and test its efficacy by applying it to 25 California sites and comparing the resulting classifications with those of 6 experts who were provided the same data. The framework is based on integrating the answers to 2 questions: 1) is there biological degradation, and 2) is chemical exposure high enough to potentially result in a biological response? The framework produced results that matched the median classifications of the experts better than did 5 of the 6 experts. Moreover, the framework was unbiased, with samples that differed from the median expert response evenly divided between those classified as more or less impacted. The framework was also evaluated by application to a set of sites from known degraded and reference areas, which the framework distinguished well. Although any framework needs to be flexible to supplemental data when they are available, the framework presented provides an objective means for using a triad-based approach in large-scale assessments for which relying on expert input for every sample is impractical. Integr Environ Assess Manag 2012;8:589–596.

Keywords: Sediment quality assessment California Data integration Bays

Special Series

INTRODUCTION

Sediment quality assessments for the effects of chemical contamination on benthic macroinvertebrates frequently use a triad of chemical concentration, sediment toxicity, and benthic community condition data (Long and Chapman 1985). These multiple lines of evidence (MLOE) are used in combination because sediments are a complex matrix and chemical concentration data alone fail to differentiate between the fraction that is tightly bound to sediment and that which is biologically available. Toxicity tests improve on chemical measurements because they integrate the effects of multiple contaminants and provide an empirical assessment of bioavailability. However, toxicity tests are typically conducted under laboratory conditions, using species that may not occur naturally at the test site; this makes it difficult to interpret the ecological significance of the results when used alone. Benthic community condition is a more direct ecological indicator of in situ sediment quality because benthic macroinvertebrates are resources at risk of sediment contamination. Relying solely on benthic community assessment to assess sediment quality is problematic, because the benthos are potentially affected by noncontaminant factors, such as hypoxia or physical disturbance (Dauer et al. 2000).

Several approaches have been developed for integrating MLOE data (Chapman et al. 2002). These integration approaches rely on a similar suite of indicators for each LOE but differ in how the LOEs are combined into a single assessment. Some are based on combinations of binary responses for each LOE; others use a more complex statistical summarization. Some approaches weight the 3 LOEs equally, whereas others weight them differently. Furthermore, effects thresholds must be determined for each LOE for incorporation into an integration framework, and such thresholds may not be available, or they may vary between program LOEs; these thresholds are particularly important when the integration is based on a binary decision for each LOE. As a result of these limitations, most triad applications rely to some degree on best professional judgment (BPJ) of experts to interpret the data based on their experience (Burton et al. 2002; Chapman and Anderson 2005). Despite the many decisions inherent in the integration of LOEs, BPJ has been found to be reasonably repeatable for interpretation of triad data (Bay, Berry, et al. 2007). Thus, BPJ can be an acceptable means of integration for site-specific assessments, but it is not easily applicable to large-scale assessments in which many sites are involved. It is also problematic in a regulatory setting in which the assessment protocol must be objective, transparent, and consistently reproducible over time and space (Forbes and Calow 2004). The State of California is develop-

* To whom correspondence may be addressed: steveb@sccwrp.org
Published online 3 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/ieam.118

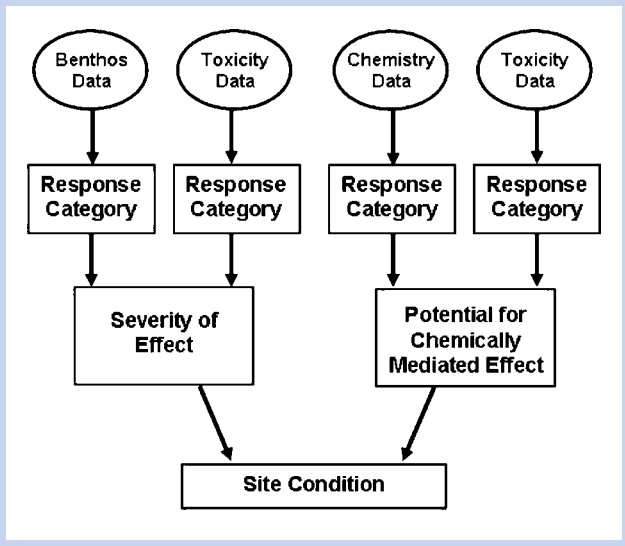

Figure 1. Conceptual model for integration of multiple lines of evidence in the assessment framework.

ing a framework for standardizing such triad-based assessments as part of establishing sediment quality objectives. The present study describes that framework and evaluates the extent to which assessments within the framework are consistent with that of experts employing BPJ on the same data.

METHODS AND MATERIALS

Integration framework

The MLOE integration framework was developed through an iterative process that incorporated input from 3 types of end users: water quality regulators, dischargers, and sediment quality research scientists. Input from regulators, dischargers, and other stakeholders was facilitated through an Advisory Committee that included representatives of the types of agencies that would ultimately be responsible for using the framework as part of monitoring and permit compliance programs. A separate Scientific Steering Committee, made up of experts outside California, was also used to review the conceptual and technical aspects of the framework and ensure that its design represented sound science. Specifics of the framework, including LOE classification criteria, evolved over multiple committee meetings and were refined solely based on the policy, implementation, and scientific issues identified by the agencies and scientists, rather than being calibrated to maximize agreement with prior frameworks or assessment results.

The framework integrates 3 LOEs to assess sediment quality at a site (i.e., specific location or station) and involves a 3-step process (Figure 1). First, the response for each LOE is assigned into 1 of 4 response categories: 1) no difference from background conditions, 2) a small response that is barely distinguishable from background conditions, 3) a response that is clearly distinguishable from background, and 4) a large response indicative of extreme conditions.

Second, the individual LOEs are combined to address 2 key questions from a typical risk assessment paradigm: is there biological degradation at the site, and is chemical exposure at the site high enough to potentially result in an adverse biological response? To answer the first question, the benthos and toxicity LOEs are integrated, to assess the severity of effects (Table 1). The effects assessment is equivalent to the benthic condition in most cases, except where there is extreme disagreement between the benthos and toxicity LOEs. Benthos is given greater weight in this assessment because it is the ultimate endpoint of interest (Chapman 2007). Implicit in this framework is the assumption that the benthic condition indices and study design have sufficient power to detect the presence of an ecologically meaningful effect.

The second question arises because the biological response may be attributable to factors other than chemical contaminants. This framework is intended to assess impacts on sediment quality attributable to anthropogenic chemical contamination, as opposed to impacts from physical or biological processes. The potential that effects are chemically mediated is assessed using the sediment chemistry and toxicity LOEs (Table 2). Chemistry is the more direct measure, but toxicity is given equal weight because of the potential presence of unmeasured chemicals and the unknown bioavailability of the measured chemicals (Ingersoll et al. 2005).

The third data integration step combines the severity of effect and potential for chemically mediated effects to assign a site into 1 of 6 impact categories (Table 3):

- *Unimpacted:* Contamination has been assessed as not responsible for any significant adverse impacts to benthic macroinvertebrates at the site.
- *Likely Unimpacted:* Contamination is not expected to cause adverse impacts to benthic macroinvertebrates, but some disagreement among LOEs reduces certainty that the site is unimpacted.
- *Possibly Impacted:* Contamination at the site may be causing adverse impacts to benthic macroinvertebrates, but the level of impact is either small or uncertain because of disagreement among LOEs.

Table 1. Severity of effect classifications, derived from the benthos and toxicity LOEs

	Toxicity LOE category			
Benthos LOE category	**Nontoxic**	**Low toxicity**	**Moderate toxicity**	**High toxicity**
Reference	Unaffected	Unaffected	Unaffected	Low effect[a]
Low disturbance	Unaffected	Low effect	Low effect	Low effect
Moderate disturbance	Moderate effect	Moderate effect	Moderate effect	Moderate effect
High disturbance	Moderate effect[a]	High effect	High effect	High effect

[a]Extreme disagreement between LOEs is present. Review of additional information about the sample before making an assessment is recommended.

Table 2. Potential that effects are chemically mediated classifications, derived from chemistry and toxicity LOEs

Chemistry LOE category	Toxicity LOE category			
	Nontoxic	Low toxicity	Moderate toxicity	High toxicity
Minimal exposure	Minimal potential	Minimal potential	Low potential	Moderate potential[a]
Low exposure	Minimal potential	Low potential	Moderate potential	Moderate potential
Moderate exposure	Low potential	Moderate potential	Moderate potential	Moderate potential
High exposure	Moderate potential[a]	Moderate potential	High potential	High potential

[a]Extreme disagreement between LOEs is present. Review of additional information about the sample before making an assessment is recommended.

- *Likely Impacted:* Evidence of contaminant-related impacts to benthic macroinvertebrates is persuasive, in spite of some disagreement among LOEs.
- *Clearly Impacted:* Sediment contamination at the site is responsible for clear and severe adverse impacts on benthic macroinvertebrates.
- Inconclusive: Disagreement among the LOE suggests that either the data are suspect or additional information is needed before a classification can be made.

The decision process for determining the site assessment category is based on a foundation of some evidence of biological effect in order to classify a site as impacted. Additionally, elevated chemical exposure must be in evidence for a site to be classified as chemically impacted.

Application of the framework

Application of the framework involves measuring sediment chemistry, toxicity, and the benthic community condition at each site, using standardized methods. The response of each measurement is compared to response ranges to categorize each of the individual LOEs into 1 of 4 possible response categories (Table 4). A brief description of the methods that comprise each LOE follows. A full description of each standardized method may be found in a technical guidance manual for the framework (Bay et al. 2009).

Chemistry. A combination of 2 sediment chemistry indices is used to determine the magnitude of chemical exposure at each site: the logistic regression models calibrated to California data (CA LRM) and the chemical score index (CSI). The CA LRM uses a set of logistic regression models to predict the probability of sediment toxicity based on the concentrations of 12 sediment contaminants (Bay et al. 2008; Field et al. 2002; USEPA 2005). To determine the probability of toxicity for the target constituents, the concentration data for each are entered in the following logistic regression equation:

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

where p is the probability of observing a toxic effect; B0 is the intercept parameter (contaminant-specific); B1 is the slope parameter (contaminant-specific); and x is the log of the concentration of the contaminant of interest.

The maximum p value (Pmax) among the target analytes is used to classify the magnitude of chemical exposure for the CA LRM index.

The CSI describes the magnitude of chemical exposure with respect to the potential for benthic community disturbance (Ritter et al. 2008). This index is calculated by comparing the concentration of 12 contaminants with a set of contaminant-specific thresholds to produce a benthic disturbance category score for each contaminant. Each score is adjusted by multiplying by a weighting factor based on the strength of association between the contaminant and benthic community response. The final CSI index is the weighted mean of all scores.

$$\mathrm{CSI} = \sum_{i=1}^{12} (w_i \times \mathrm{cat}_i) / \sum_{i=1}^{12} w_i,$$

where cat_i is the predicted benthic disturbance category for chemical i; w_i is the weight factor for chemical i; and Σw_i is the sum of all weights.

Table 3. Multiple lines of evidence (MLOE) sample classifications, derived from intermediate classifications described in Tables 1 and 2

Potential for chemically mediated effects	Severity of effect			
	Unaffected	Low effect	Moderate effect	High effect
Minimal potential	Unimpacted	Likely unimpacted	Likely unimpacted	Inconclusive
Low potential	Unimpacted	Likely unimpacted	Possibly impacted	Possibly impacted
Moderate potential	Likely unimpacted	Possibly impacted or inconclusive[a]	Likely impacted	Likely impacted
High potential	Inconclusive	Likely impacted	Clearly impacted	Clearly impacted

[a]Inconclusive category applies when: chemistry = minimal exposure; benthos = reference; and toxicity = high. Other LOE combinations represented by this cell are classified as Possibly impacted.

Table 4. Ranges of values used to define each LOE indicator category

		Category			
LOE	**Indicator**	**Minimal**	**Low**	**Moderate**	**High**
Chemistry exposure	CA LRM (*P*max)	<0.33	≥0.33 to <0.49	≥0.49 to ≤0.66	>0.66
	CSI (mean)	<1.69	≥1.69 to ≤2.33	>2.34 to ≤2.99	>2.99
		Nontoxic (%)	**Low (% control)**	**Moderate (% control)**	**High (% control)**
Toxicity	*Eohaustorius*[a]	≥90	<90 to ≥82	<82 to ≥59	<59
	Leptocheirus[a]	≥90	<90 to ≥78	<78 to ≥56	<56
	Rhepoxynius[a]	≥90	<90 to ≥83	<83 to ≥70	<70
	Mytilus[b]	≥80	<80 to ≥77	<77 to ≥42	<42
	Neanthes[c]	≥90	<90 to ≥68	<68 to ≥46	<46
		Reference	**Low**	**Moderate**	**High**
Benthos disturbance	BRI	<33	≥33 to < 51	≥51 to <70	≥70
Southern California	IBI	0	1	2	≥3
	RBI	>0.27	≤0.27 to 0.16	≤0.16 to >0.07	≤0.07
	RIVPACS	≥0.9 to <1.1	≥0.74 to <0.89	>0.31 to <0.74	≤0.31
			≥1.1 to <1.27	≥1.27	
	BRI	<22.3	≥22.3 to <33.4	≥33.4 to <82.1	≥82.1
San Francisco Bay	IBI	≤1	2	3	4
	RBI	>0.43	≤0.43 to >0.29	≤0.29 to >0.19	≤0.19
	RIVPACS	≥0.68 to <1.32	≥0.32 to <0.68	>0.15 to <0.32	≤0.15
			≥1.32 to <1.68	≥1.68	

Separate benthic index ranges are used for southern California and San Francisco Bay habitats.
[a]0-d survival test.
[b]2-d embryo survival and development test.
[c]28-d growth test; all classification ranges are expressed as percentage of control.

Index-specific thresholds are applied to each index value to classify the result into 1 of 4 chemical exposure categories: minimal, low, moderate, and high. These thresholds were developed specifically for use in this framework by calibrating the indices to toxicity and benthic community condition data from California. A large number of candidate thresholds were selected using a statistical optimization procedure based on maximizing overall agreement between the chemistry index categories and biological effect categories in subsets of the calibration data set. The optimal set of thresholds was selected by computing the percentage agreement for a large set of possible candidates and selecting the set of 3 thresholds that resulted in the highest overall agreement.

The resulting chemical exposure categories are assigned a score of 1–4 (e.g., minimal exposure = 1) and the mean score from the 2 chemistry indices are used to determine the overall chemistry LOE category. Mean scores are rounded up to the next whole number in the case of intermediate results (e.g., mean score of 2.5 = moderate exposure).

Toxicity. The results of multiple sediment toxicity tests are used to determine the magnitude of sediment toxicity at each site. The tests must include a 10-d amphipod survival test and a sublethal test (e.g., 4-d mussel embryo development or 28-d juvenile polychaete growth). Effects thresholds based on magnitude of response are applied to classify the results from each test into 1 of 4 toxicity categories (Bay, Greenstein, Young 2007): nontoxic (1), low (2), moderate (3), and high (4). The mean toxicity category score from all tests is used to determine the overall toxicity LOE category.

Separate effects thresholds were developed for each toxicity test and were based on a consistent statistical approach. The threshold separating the nontoxic and low categories was equivalent to the minimum acceptable control value specified in each test method. Classification of the results as low or moderate was based on a threshold equal to the 90th percentile minimum significant difference (MSD) for the test, calculated using test response data from California (Bay, Greenstein, Young 2007). The threshold defining high toxicity was the average of 2 effect values: the 99th percentile

MSD and the 75th percentile of responses in samples that were statistically significant from the control.

Benthos. A combination of 4 benthic community condition indices is used to determine the magnitude of disturbance to the benthos at each site (Ranasinghe et al. 2007). The benthic indices include approaches based on community metrics and abundance of individual species:

Benthic response index (BRI): This index was originally developed for the southern California mainland shelf and extended into California's bays and estuaries (Smith et al. 2001, 2003). The BRI is the abundance-weighted average pollution tolerance score of organisms that occur in a sample. The first steps in the BRI calculation are to compute the 4th root of the abundance of each taxon in the sample and then multiply the value by the pollution tolerance score (P) for the taxon. The BRI score is calculated by dividing the sum of the products (abundance $\times P$) by the sum of the abundances for all taxa having pollution tolerance scores.

Index of benthic biotic integrity (IBI): This index was developed for freshwater streams and adapted for California's bays and estuaries (Thompson and Lowe 2004). The IBI compares the values of 4 different metrics with the ranges expected under reference conditions. The specific metrics vary by habitat and include number of taxa, number or percentage of sensitive taxa, number of mollusk or amphipod taxa, abundance of tolerant indicator species, and total abundance. Each metric that is outside the reference range increases the IBI score by 1.

Relative benthic index (RBI): This index was originally developed for California's Bay Protection and Toxic Cleanup Program (Hunt et al. 2001). The RBI is the weighted sum of 1) 4 community metrics related to biodiversity (total number of taxa, number of crustacean taxa, abundance of crustacean individuals, and number of mollusk taxa), 2) abundance of 3 positive indicator taxa (vary by habitat), and 3) presence of 2 negative indicator species (vary by habitat). The community metric values are normalized to habitat-specific maxima and combined with normalized values for the indicator taxa and/or species to produce a RBI score scaled to the expected range of scores in the habitat.

River Invertebrate Prediction and Classification System (RIVPACS): This index was originally developed for British freshwater streams (Wright et al. 1993; Van Sickle et al. 2006) and adapted for California's bays and estuaries. The RIVPACS index calculates the number of reference taxa present in the test sample and compares it with the number expected to be present in a reference sample from the same habitat. First, the probability of the test sample belonging to each of several habitat-specific reference groups is calculated using a linear discriminant function. Next, the identity and expected number of reference species, based on the probabilities of group membership and the distribution of reference species in each group, is determined. In the final step, the number of reference species observed in the sample is counted, and the observed/expected (O/E) RIVPACS score is calculated.

Response ranges specific to regional assemblages are applied to the results in order to classify each benthic index result into 1 of 4 disturbance categories: reference, low, moderate, and high. The median of categories for each individual index is used to determine the overall benthos LOE category. The thresholds used for classification were developed using a 2-step process. First, a provisional set of thresholds was identified by the index developer. Second, the thresholds were refined (if necessary) to improve classification accuracy by comparison with a calibration data set of samples with known benthic condition.

The final site assessment category is determined by comparing the response category results for each LOE, based on the relationships shown in Tables 1–3.

Evaluation of the framework

The efficacy of the framework was assessed in 2 ways. The first was to apply it to samples from 25 sites and compare the classifications with those of 6 experts provided the same data. The experts were selected to represent a diverse range of sectors (industry, academia, government), with each having at least 15 years of experience in conducting sediment quality assessments and advising national, state, and local agencies with regard to sediment management and remediation decisions (Bay, Berry, et al. 2007). The experts were asked to classify the samples into 1 of the 6 impact categories described above.

The 25 samples were selected by rank-ordering a California database according to overall chemical concentrations based on the mean effects range median quotient (ERMq) (Long et al. 2006) and then randomly selecting from quartile groups, so that a range of exposure conditions were represented. Twenty-one of the samples were from euryhaline coastal bays in southern California; 4 samples were from polyhaline areas of the San Francisco Bay.

The data provided to the experts included depth, percentage sediment fines, percentage total organic carbon, chemical concentrations, toxicity, and benthic infaunal condition. The experts were not provided site location or any of the calculated values used in the framework. The chemical concentration data were for 11 metals, 21 polycyclic aromatic hydrocarbons (PAHs), chlorinated pesticides (DDTs and chlordanes), and total PCBs (sum of 16 congeners). Mean quotients for 2 empirical sediment quality guidelines (ERM and SQGQ1) and total toxic units for PAHs (based on EPA sediment quality benchmarks) were also provided for context, but no recommendations were given regarding whether or how to use the SQGs. Toxicity data were from a 10-d *Eohaustorius estuarius* mortality test conducted according to standard methods (USEPA 1994). Because not all of the MLOE experts were familiar with California benthos, benthic infauna data were provided as a 4-category consensus BPJ condition assessment developed by 9 benthic experts (Weisberg et al. 2008); the underlying benthic species abundance data were made available on request.

Agreement between the experts and the framework was quantified in 2 ways. First, the overall rate of disagreement was determined. This value was calculated by counting the number of impact categories for which the assessment derived using the framework differed from the median assessment of the experts for each sample, and then taking the sum of the counts across all samples. Second, the bias of the framework

Table 5. Individual sample results for expert and MLOE framework assessments

Sample	Expert median	MLOE framework
1	Unimpacted	Unimpacted
2	Possibly impacted	Possibly impacted
3	Likely unimpacted[a]	Possibly impacted[a]
4	Likely unimpacted[a]	Unimpacted[a]
5	Likely impacted	Likely impacted
6	Unimpacted	Unimpacted
7	Likely unimpacted	Likely unimpacted
8	Likely impacted	Likely impacted
9	Possibly impacted	Possibly impacted
10	Likely impacted	Likely impacted
11	Clearly impacted	Clearly impacted
12	Possibly impacted[a]	Likely unimpacted[a]
13	Possibly impacted	Possibly impacted
14	Likely impacted[a]	Clearly impacted[a]
15	Likely impacted	Likely impacted
16	Possibly impacted[a]	Unimpacted[a]
17	Possibly impacted	Possibly impacted
18	Unimpacted	Unimpacted
19	Clearly impacted	Clearly impacted
20	Clearly impacted	Clearly impacted
21	Clearly impacted	Clearly impacted
22	Clearly impacted	Clearly impacted
23	Unimpacted	Unimpacted
24	Unimpacted	Unimpacted
25	Unimpacted	Unimpacted

[a]Sample for which the assessments differ. Expert median is based on results for experts as reported in Bay et al. (2007).

was quantified as the net of positive and negative differences from the median expert. This value was calculated by incorporating a sign into the count of category differences from the median for each sample, and then summing the counts across all samples. For additional perspective, the framework's agreement with the median of the experts' results was compared with the agreement of each of the individual experts with the median of the other experts.

The second evaluation approach involved determining the extent to which the framework differentiated samples from geographic areas previously designated as toxic hotspots by the State of California with those from reference areas. The hotspot regions were identified by the State's Bay Protection and Toxic Cleanup Program (BPTCP) as the worst in the state, based on a BPJ assessment of sediment chemistry, toxicity, and benthic community condition (Anderson et al. 2001; Fairey et al. 1998; Phillips et al. 1998). The reference sites were identified as areas that were distant from known sources of contamination (e.g., outer portion of embayments) and for which previous surveys had consistently shown low toxicity (defined as >80% amphipod survival) and low chemistry (defined as a mean ERM quotient <0.5). The data used for this evaluation were independent of the data used to identify either the hotspot or reference areas.

RESULTS

The framework evaluation categorized the 25 samples the same as the median assessment of the experts for all but 5 of the samples (Table 5). There was only 1 sample for which the framework and median expert assessment differed by more than a single impact category, resulting in a total of 6 category disagreements for all samples. This compares favorably with agreement among the experts (Table 6). Only 1 of the 6 experts had a lesser number of disagreements with the median expert result than did the framework. Many of the remaining experts disagreed with the median at twice the rate of the framework.

The framework also had little bias, with 3 of the samples rated as less impacted compared with the expert median and 2 as more impacted. The overall net bias, which incorporated both the number and sign of the category differences, was −2. Only 2 of the experts had a lesser bias; 3 of the experts had 5 times greater bias.

The framework also did a good job of differentiating the BPTCP hotspots from reference areas. Almost 90% of the samples from predicted reference areas were classified as Unimpacted or Likely Unimpacted and none of these samples was classified as Clearly Impacted (Figure 2). By

Table 6. Summary of categorical assessments for experts and MLOE framework

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Framework
Nr of samples	25	22	25	19	25	22	25
Disagreement[a]	7	16	13	10	15	5	6
Bias[b]	1	−12	11	4	–	−1	−2

[a]Disagreement values for experts represent the total number of category differences between the expert's assessment and the median of all other experts' assessments. Framework disagreement is the number of category differences between the framework and median of all experts.
[b]Bias values reflect the net of positive or negative assessment differences, with positive numbers indicating a bias toward rating the sample as more impacted and numbers closest to zero indicating the least bias overall.

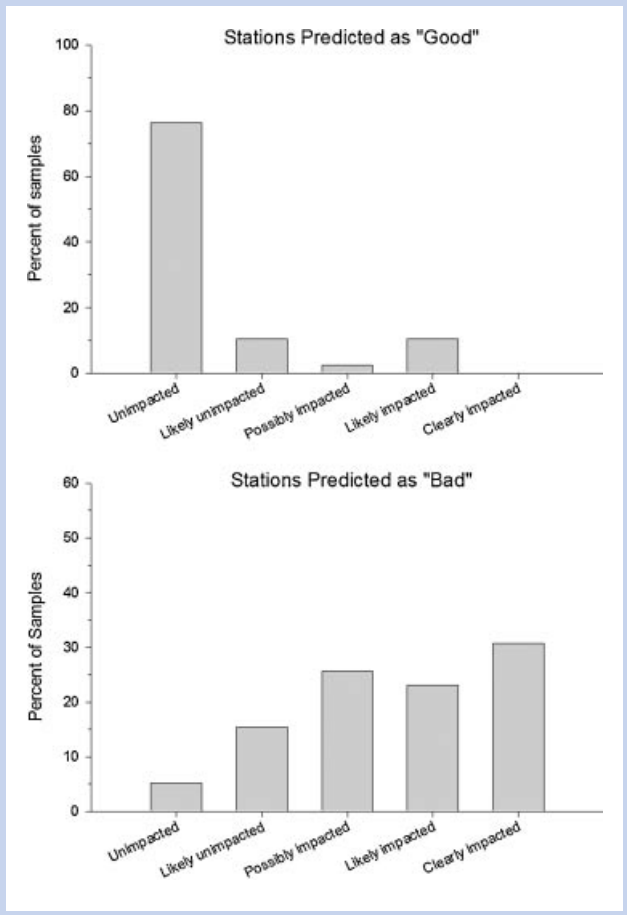

Figure 2. Distribution of MLOE assessment categories for California sites located in locations predicted from previous studies to be either unimpacted or impacted. $n = 38$ for unimpacted samples; $n = 39$ for impacted samples.

contrast, about 80% of the samples from predicted hotspot areas were classified into one of the impacted categories, with more than 50% of the samples classified as Likely or Clearly Impacted.

DISCUSSION

A formulaic approach to data integration has potential shortcomings because additional information about a site would sometimes be factored in by experts but not included in a structured objective assessment. However, the formulaic approach also offers some advantages. It is transparent, so that all parties will reach the same conclusion using the same data. Moreover, it is not susceptible to the individual biases associated with use of BPJ. Such biases, or the need for employing a large team of experts to average out individual biases, would be problematic for large-scale assessments.

The selected framework employs unequal weighting among LOEs, which differs from that of the earliest triad integration frameworks (Chapman 1990). We initially considered a framework based on equal weighting, but believed that the 2-phase assessment approach and its inherent weightings of the different LOEs had a conceptual basis more in line with risk assessment. Subsequent to this study, we attempted an equally weighted framework and found it did not perform as well in reproducing results from the experts. Discussions with the experts after the study indicated that the 2-phase approach mimicked their thought process, as few placed equal weighting in their assessments. Most placed greatest emphasis on benthos because it is the ultimate endpoint of interest and weighted chemistry least because of uncertainties associated with potential exposure from unmeasured chemicals.

The framework ranks each LOE on a 4-category response scale, in contrast to the binary framework prevalent in the initial triad integration approaches (Long and Chapman 1985). A multicategory response scale improves on the binary approach because it lessens the all-or-none nature of thresholds that are established with great uncertainty (Batley et al. 2002). The use of 5 response categories for such applications is prevalent in Europe, but ultimately the choice of number of categories becomes a tradeoff between placing great importance on a small number of thresholds and having more thresholds than philosophical bases on which to establish them. For the present study, we chose 4 response categories because we were able to identify a unifying concept for threshold selection across LOEs. The first threshold separates the reference and low-effect categories and is one at which differences from reference initially become apparent. The second threshold is where the differences become substantial enough to be detected with statistical confidence. The last threshold, separating the moderate-effect and high-effect categories, is one where the difference from background is severe and becomes increasingly subjective to select. We stopped at 4 categories because establishing additional thresholds beyond that seemed increasingly artificial.

Our application of the framework involved using multiple indices to summarize the complex chemistry and benthos LOEs. The framework is not dependent on use of multiple indicators within an LOE, but multiple indicators proved helpful in reducing variability associated with individual indices. We did not use multiple indicators for the toxicity LOE because the available California data sets that contained all 3 LOEs included only a single toxicity test. However, we believe that integrating multiple toxicity tests is also advisable to reduce uncertainty in the evaluation of this LOE (Burton et al. 1996).

The assessment framework yields 5 impact categories along a continuous scale, differing from Chapman's (1990) original integration framework, which provided nonlinear independent interpretations for each of 8 LOE combinations. The impact category approach used here resulted from consultation with managers from the regulatory, regulated, and public advocacy sectors who requested that information be reduced to a linear scale that ranks samples, at least categorically, from best to worst. Linearization is scientifically challenging because it confounds several factors: confidence that there is an effect, magnitude of the effect, and likelihood that the effect is chemically mediated. The 2-phase assessment approach described provides the management community with the linear scale they need for large-scale assessments while retaining a relationship to the individual LOEs needed to interpret data from individual samples.

The framework suggested is not the only one possible. Numerous other MLOE integration approaches have been suggested, including those based on multivariate analysis, statistical summarization, logic models, and scoring systems (Burton et al. 2002; Chapman et al. 2002). It is also clear that

when other kinds of data for a site are available, such as toxicity identification evaluations or bioaccumulation testing, they should be incorporated into the assessment process (Chapman and Hollert 2006). However, the framework described was shown to reflect closely the central tendency of assessments conducted by experts provided with the same data, and demonstrates that objective formulaic approaches are viable for large-scale assessments or in a regulatory context in which transparency in the decision process is critical.

Acknowledgment—Work on this project was funded by the California State Water Resources Control Board under agreement 01-274-250-0. The authors thank Chris Beegan from the California Water Resources Control Board and Mike Connor, Bruce Thompson, and Ben Greenfield of the San Francisco Estuary Institute for their thoughts during many discussions about possible frameworks. Jeff Brown assisted with data compilation and statistical 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.

REFERENCES

Anderson BS, Hunt JW, Phillips BM, Fairey R, Roberts CA, Oakden JM, Puckett HM, Stephenson M, Tjeerdema RS, Long ER, Wilson CJ, Lyons JM. 2001. Sediment quality in Los Angeles Harbor, USA: A triad assessment. *Environ Toxicol Chem* 20:359–370.

Batley GE, Burton GA, Chapman PM. 2002. Uncertainties in sediment quality weight of evidence (WOE) assessments. *Hum Ecol Risk Assess* 8:1517–1548.

Bay S, Berry W, Chapman P, Fairey R, Gries T, Long ER, McDonald D, Weisberg SB. 2007. Evaluating consistency of best professional judgment in the application of a multiple lines of evidence sediment quality triad. *Integr Environ Assess Manag* 3:491–497.

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

Bay SM, Ritter KJ, Vidal-Dorsch DE, Field LJ. 2008. Comparison of national and regional sediment quality guidelines for predicting sediment toxicity in California. In: Weisberg SB, Miller K, editors. Annual report. Costa Mesa (CA): Southern California Coastal Water Research Project. p 79–90.

Bay SM, Greenstein DJ, Ranasinghe JA, Diehl DW, Fetscher AE. 2009. Sediment Quality Assessment Draft Technical Support Manual. Technical Report 582, Southern California Coastal Water Research Project, Costa Mesa, CA.

Burton GA Jr, Ingersoll CG, Burnett LC, Henry M, Hinman ML, Klaine SJ, Landrum PF, Ross P, Tuchman M. 1996. A comparison of sediment toxicity test methods at three Great Lake areas of concern. *J Great Lake Res* 22:495–511.

Burton GA Jr, Chapman PM, Smith EP. 2002. Weight of evidence approaches for assessing ecosystem impairment. *Hum Ecol Risk Assess* 8:1657–1673.

Chapman PM. 1990. The sediment quality triad approach to determining pollution induced degradation. *Sci Tot Environ* 97:815–825.

Chapman PM. 2007. Do not disregard the benthos in sediment quality assessment. *Marine Pollut Bull* 54:633–635.

Chapman PM, Anderson J. 2005. A decisionmaking framework for sediment contamination. *Integr Environ Assess Manag* 1:163–173.

Chapman PM, Hollert H. 2006. Should the sediment quality triad become a tetrad, a pentad or possibly even a hexad? *J Soil Sediments* 8:4–8.

Chapman PM, McDonald BG, Lawrence GS. 2002. Weight-of-evidence issues and frameworks for sediment quality (and other) assessments. *Hum Ecol Risk Assess* 8:1489–1515.

Dauer DM, Ranasinghe JA, Weisberg SB. 2000. Benthic community condition in relation to water quality, sediment quality and watershed stressors in Chesapeake Bay. *Estuaries* 23:80–96.

Fairey R, Roberts C, Jacobi M, Lamerdin S, Clark R, Downing J, Long E, Hunt J, Anderson B, Newman J, Tjeerdema R, Stephenson M, Wilson C. 1998. Assessment of sediment toxicity and chemical concentrations in the San Diego Bay region, California, USA. *Environ Toxicol Chem* 17:1570–1581.

Field LJ, MacDonald DD, Norton SB, Ingersoll CG, Severn CG, Smorong D, Lindskoog R. 2002. Predicting amphipod toxicity from sediments using Logistic Regression Models. *Environ Toxicol Chem* 9:1993–2005.

Forbes VE, Calow P. 2004. Systematic approach to weight of evidence in sediment quality assessment: Challenges and opportunities. *Aquatic Ecosyst Health Manag* 7:339–350.

Hunt JW, Anderson BS, Phillips BM, Tjeerdema RS, Taberski KM, Wilson CJ, Puckett HM, Stephenson M, Fairey R, Oakden JM. 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.

Ingersoll CG, Bay SM, Crane JL, Field LJ, Gries TH, Hyland JL, Long ER, MacDonald DD, O'Connor TP. 2005. Ability of SQGs to estimate effects of sediment-associated contaminants in laboratory toxicity tests or in benthic community assessments. In: Wenning RJ, Batley GE, Ingersoll CG, Moore DW, editors. Use of sediment quality guidelines (SQGs) and related tools for the assessment of contaminated sediments. Pensacola (FL): SETAC. p 497–556.

Long ER, Chapman PM. 1985. A sediment quality triad-measures of sediment contamination, toxicity and infaunal community composition in Puget Sound. *Marine Pollut Bull* 16:405–415.

Long ER, Ingersoll CG, MacDonald DD. 2006. Calculation and uses of mean sediment quality guideline quotients: A critical review. *Environ Sci Technol* 40:1726–1736.

Phillips BM, Anderson BS, Hunt JW, Newman J, Tjeerdema RS, Wilson CJ, Long ER, Stephenson M, Puckett HM, Fairey R, Oakden JM, Dawson S, Smythe H. 1998. Sediment chemistry, toxicity and benthic community conditions in selected water bodies of the Santa Ana region. Sacramento (CA): California State Water Resources Control Board.

Ranasinghe JA, Weisberg SB, Smith RW, Montagne DE, Thompson B, Oakden JM, Huff DD, Cadien DB, Velarde RG. 2007. Evaluation of five indicators of benthic community condition in two California bay and estuary habitats Technical Report 524. Costa Mesa (CA): Southern California Coastal Water Research Project.

Ritter KJ, Bay SM, Smith RW, Vidal-Dorsch DE, Field LJ. 2008. Development and evaluation of sediment quality guidelines based on benthic macrofauna responses. In: Weisberg SB, Miller K, editors. Annual report. Costa Mesa (CA): Southern California Coastal Water Research Project. p 91–105.

Smith RW, Bergen M, Weisberg SB, Cadien DB, Dalkey A, Montagne DE, Stull JK, Velarde RG. 2001. Benthic response index for assessing infaunal communities on the Southern California Mainland Shelf. *Ecol Appl* 11:1073–1087.

Smith RW, Ranasinghe JA, Weisberg SB, Montagne DE, Cadien DB, Mikel TK, Velarde RG, Dalkey A. 2003. Extending the Southern California benthic response index to assess benthic condition in bays. Technical Report 410. Southern California Coastal Water Research Project. Westminster, CA.

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

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

[USEPA] United States Environmental Protection Agency. 2005. Predicting Toxicity to Amphipods from Sediment Chemistry (Final Report). EPA/600/R-04/030. Washington (DC): USEPA, ORD National Center for Environmental Assessment.

Van Sickle J, Huff DD, Hawkins CP. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshw Biol* 51:359–372.

Weisberg SB, Thompson BE, Ranasinghe JA, Montagne DE, Cadien DB. 2008. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. *Ecol Indic* 8:389–394.

Wright JF, Furse MT, Armitage PD. 1993. RIVPACS: A technique for evaluating the biological water quality of rivers in the UK. *Eur Water Pollut Control* 3: 15–25.