Development of Sediment Quality Objectives for California Bays and Estuaries

Workplan for:
Development of Chemistry Indicators

October 20, 2004

Background

A primary focus of the California State Water Resources Control Board's effort to develop sediment quality objectives is protecting aquatic life from the direct effects of contaminants associated with the sediments in enclosed bays and estuaries. This element of the project will use a multiple line of evidence approach that integrates three lines of evidence (LOE; i.e., chemistry, toxicity, and benthic community health) to provide a robust assessment of sediment condition (Chapman *et al.* 2002). A key aspect of this effort is the development of chemistry indicators that describe the magnitude of contaminant exposure and correspond to ecologically relevant impacts to aquatic life, such as toxicity and benthic community impairment. Activities to develop chemistry indicators predictive of contaminant bioaccumulation are described in a separate workplan.

Various approaches to developing chemical indicators have been used throughout the nation, resulting in a wide variety of sediment quality guidelines (SQGs) based on either empirical or mechanistic relationships with biological effects (Barrick *et al.* 1988, Long *et al.* 1995, Swartz 1999, Di Toro and McGrath 2000, Fairey *et al.* 2001, Field *et al.* 2002). Chemical SQGs are used by environmental managers to summarize complex chemistry data and make assessments regarding the potential for biological impacts (U.S.EPA 2001). As a component of a multiple line of evidence approach, chemical SQGs can help distinguish between biological impacts that are likely due to contamination instead of confounding factors (e.g., grain size).

The use of chemical SQGs is often accompanied by substantial uncertainty and controversy, as no single SQG approach is able to account for all of the factors that influence contaminant effects. Empirical SQGs are based on the analysis of large field data sets and do not indicate causality or directly account for variations in contaminant bioavailability and the effects of contaminant mixtures. Mechanistic SQGs do not account for the effects of unmeasured chemicals, are not available for some chemicals of concern, and may not be protective of long-term or sublethal effects. Studies have shown that chemical SQGs are predictive of the incidence and magnitude of biological effects in instances of high/low contaminant concentrations, but SQGs have also been shown to have high error rates when applied to samples containing intermediate levels of contamination (Long *et al.* 1998, Fairey *et al.* 2001). The predictive ability of SQGs has also been shown to vary among datasets from different regions (Fairey *et al.* 2001, Crane *et al.* 2002), which complicates the selection of the most reliable approach for a given application.

This workplan addresses technical issues involved in developing chemical indicators that are based on various SQG approaches, while other workplans for this project address the additional indicators (benthic macrofauna and toxicity) included in the multiple line of evidence approach.

Objectives

This project has four scientific objectives:

1. Identify geographic, geochemical, or source-related factors that may influence the relationships between sediment contaminant concentrations and biological effects.
2. Develop a chemical indicator that considers variations in contaminant exposure and bioavailability.
3. Develop a chemical indicator that predicts the potential for contaminant-related biological effects to the maximum extent possible.
4. Establish numeric response levels for the chemical indicators that are protective of benthic organisms in bays and estuaries and that correspond to important levels of biological effect.

General Approach

Four tasks will be conducted in order to accomplish these objectives. Task 1 consists of preparing a dataset of high quality matched chemistry, toxicity, and benthic community information that will be used in subsequent tasks. Statistical analyses will be used to determine whether the data should be stratified on the basis of geographic or other factors and an independent validation dataset will be established. A variety of candidate indicators for exposure and effects will be developed or refined in Task 2 and applied to the dataset. Task 3 includes the comparison and evaluation of the candidate indicators in order to select those best suited for use in the chemistry LOE. The focus of Task 4 is to establish numeric response levels associated with different degrees of exposure or impact for the recommended indicators. These response levels will be incorporated into the multiple line of evidence framework for assessing direct effects.

The work in this project will be directed towards developing two types of chemical indicators: mechanistic and empirical. The mechanistic indicator category will include approaches that use sediment and contaminant characteristics combined with equilibrium partitioning theory. The empirical indicator category will include approaches based on the analysis of matching sediment chemistry and bioeffects data to derive relationships that predict the likelihood of observing adverse effects. The inclusion of both mechanistic and empirical indicators into the chemistry LOE has several advantages. First, the use of multiple indicators that are derived using different methodologies will reduce the chance that significant contamination will be missed due to a lack of sensitivity in any one indicator. Second, the use of a mechanistic indicator provides greater independence in this LOE relative to the toxicity and benthic community LOEs, which are based on some of the same information that will be used in developing the empirical indicator. Finally, the inclusion of an empirical indicator provides enhanced applicability in situations where the data are not available to support use of a mechanistic indicator.

The development of chemical indicators will focus on 15 priority contaminants or contaminant categories (Table 1). These contaminants are classified as high priority because they have been associated with aquatic life impairment in previous studies in California bays and estuaries. Other contaminants may be included in the chemical indicators resulting from this project, depending upon the availability of data and the presence of an association with biological impacts.

Table 1. Priority chemicals of concern for development of chemistry indicators. The following COCs have been identified based upon occurrence on the 2002 303(d) List for impairing sediments within bays and estuaries.

Metals	Organics
Cadmium	Chlordane
Copper	Chlorpyrifos
Lead	DDT
Mercury	Dieldrin
Nickel	Mirex
Tributyltin	ppDDE
Zinc	PAHs
	PCBs

Work Description

Task 1: Prepare Datasets. This task will create a series of analysis and validation datasets that contain matched chemistry and biological effects data (toxicity or benthic community condition) from the California SQO database. The SQO database includes data that may not be relevant or acceptable for this study, such as results from offshore samples, analyses with high detection limits, or studies that did not measure important parameters. Tasks 2-4 will use these datasets to develop and evaluate various chemistry tools and it is important that data of suitable quality and completeness are analyzed in each case. A series of five subtasks are involved in preparing the dataset.

1.1. Evaluate data quality and completeness. A series of screening steps will be used to select studies and results of high quality and appropriate location. The first screening step will retain only those studies containing analyses of samples of surface sediment (upper 30 cm) located in bays and estuaries. The second screening step will evaluate the chemistry data for completeness and detection limits. Only samples for which at least 10 chemicals were measured, with the analytes including metals and PAHs at a minimum will be retained. In addition, chemical concentration data will be excluded for individual analytes if the method detection limit is greater than the ERL (effects range low). The toxicity and benthic community data will be screened for quality and completeness using procedures described in the workplans specific to those lines of evidence.

1.2. Calculate derived values. The SQO database contains raw chemistry data for most samples, such as concentrations of individual PAH compounds or PCB congeners and replicate measurements. These data will be processed and summarized as necessary to support the development, application, and evaluation of the various SQG approaches. Derived values such

as chemical group sums, SQG quotients, and estimated values for missing or nondetect data will be calculated. Data flags will be included in the dataset to clearly identify any estimated concentration values so that subsequent analyses can either use or exclude these values.

1.3. Normalize data. Variation in sediment characteristics such as grain size, geologic source, and organic carbon can mask important patterns or affect contaminant bioavailability. For example, the naturally occurring background concentration of metals is dependent on the geologic source and grain size of the sediments. Sediment metals concentration will be normalized to iron or grain size using the reference element normalization approach (Schiff and Weisberg 1999). The metals normalization analyses will be conducted separately for southern California, San Francisco Bay, and other northern California locations in order to determine if different relationships are present. This normalization will reduce variability due to geologic factors and thus enhance patterns due to anthropogenic activities. The trace organics data may also be normalized to TOC or other sediment constituents identified by the workgroup.

1.4. Determine geographic strata. The relationships between sediment chemistry and biological effects may vary among geographic regions or habitats due to variations in factors such as chemical mixtures, sediment characteristics, or the type of toxicity test used. Substantial variability in these relationships may influence the predictive ability of some candidate SQGs and it may be necessary to stratify the datasets on the basis of geography prior to the development or evaluation of the chemical indicators. Multivariate techniques such as principal components analysis (PCA) will be used to identify the presence of geographically distinct patterns and determine their significance on chemistry and bioeffects patterns. Decisions regarding whether to stratify the data will be based on the results of these analyses.

1.5. Establish validation dataset. An independent subset of bay and estuary data will be selected and set aside for use in validation of the candidate chemical indicators. Separate validation datasets may be established for various geographic regions, depending upon the results of Task 1.4. Each validation dataset will include approximately one third of the amount of data used for development of the indicators and these data will meet the same selection criteria applied to the analysis dataset. The validation data will be derived from two sources, either selected randomly from the data contained in the SQO database or from new studies not yet included in the database (e.g., southern California Bight'03 regional survey, San Francisco Bay 2003 RMP).

Task 2: Develop and Refine SQGs. This task will develop and apply a variety of SQG approaches to the chemistry data set, with the intent of developing chemistry indicators that describe biologically meaningful levels of contaminant exposure and biological effect. This task will utilize both mechanistic and empirical approaches. The emphasis will be on developing indicators that incorporate chemical mixture effects, such as mean SQG quotients or sums of toxic units, since the primary goal of the project is to evaluate overall sediment quality as opposed to identifying the cause of impairment. Four strategies of indicator development and application will be employed, as described in the following subtasks.

2.1. Calculate SQG values using existing approaches. Existing SQG approaches that have been published in the scientific literature or applied by state and local agencies will be applied to the

datasets following the procedures recommended by the original authors. The approaches to be used are listed in Table 1.

Table 1. Existing SQG approaches to be evaluated.

Guideline Approach	Metric	Source
EqP organics	Sum Acute/Chronic TU	U.S.EPA 2001 and 2003
EqP metals	Potential for toxicity	U.S. EPA *in preparation*
ERM	Mean Quotient	Long *et al*. 1995
Consensus MEC	Mean Quotient	MacDonald *et al.* 2000, Swartz 1999, Vidal and Bay. *in press*
SQGQ1	Mean Quotient	Fairey *et al*. 2001
Logistic Regression	Pmax	Field et al. 2002

SQG values based on equilibrium portioning (EqP) for organic pesticides will be derived using acute and chronic toxicity values specified in the California Toxic Rule and procedures described by the U.S EPA (U.S. EPA 2003). EqP SQGs for other organic contaminants (e.g., PAHs) will be derived using toxicity values and procedures described by the EPA (U.S. EPA 2001).

The metrics specified in Table 1 will be calculated for each sample in the SQO analysis dataset (where sufficient data is available) and matched with the corresponding measures of toxicity or benthic community effect. The results will be compared in Task 3 to evaluate the various candidate approaches.

2.2. Refine existing approaches. The empirical approaches listed in Table 1 will be refined in an effort to improve their predictive ability. For example, Fairey *et al*. (2001) showed that the predictive ability of the SQG mean quotient was improved by the use of a subset of SQG values selected from existing approaches. The refinements will include exclusion of specific chemical SQGs having a low predictive value and the use alternative mixture models (e.g. calculation of a mean SQG quotient for selected chemical classes (Ingersoll *et al.* 2001)). This will be an iterative process that will use regression analysis and statistical optimization procedures to identify the most effective combinations of the SQG values.

2.3. Calibrate existing approaches. Empirical SQGs are usually developed using a large dataset compiled from studies conducted in diverse locations and habitats. The chemical:biological relationships present in these datasets are likely to reflect the average conditions in the dataset and therefore may not reflect important relationships present in California. For example, historical industrial discharges of DDT have created widespread sediment contamination in portions of southern California at concentrations above those present in most other coastal areas, yet the sediments are frequently nontoxic to benthic organisms. Most empirically-based SQGs for DDT appear to have low predictive value, possibly as a result of being derived from datasets having much lower relative DDT concentrations.

The predictive ability of some SQG approaches may be improved by adjusting (i.e., calibrating) the values for specific chemicals or the thresholds used to classify samples as impaired to reflect regional patterns. The effectiveness of these refinements will be evaluated for the logistic

regression model approach (LRM). The individual chemical-specific regression models will be adjusted to match contamination and amphipod toxicity patterns present in the California dataset. The predictive ability of the adjusted models will be compared to that of the existing models (based on a national dataset) and the results used to determine whether model calibration provides a substantial improvement in performance of the LRM approach. If model calibration is effective, then the adjusted models will be combined using the Pmax approach to develop a California-specific model that addresses contaminant mixtures. Comparisons of model performance will also be made among subsets of the data in order to determine if model performance is affected by geographic region or study type.

2.4. Develop new approaches. Most empirical SQG approaches use relatively simple procedures to identify effect concentrations and deal with the complex mixtures of contaminants present in the environment (e.g., percentiles and means). Factors influencing the bioavailability of contaminants (e.g., organic carbon, grain size, nonbioavailable metals) are also infrequently considered in empirical SQGs. A portion of the high uncertainty associated with some applications of empirical SQGs may be due to the use of relatively simple models. Multivariate statistical techniques and normalization methods are available that may be able to account for mixture effects or other sources of variability and thus reduce the uncertainty associated with the application of chemical SQGs. The utility of developing a new SQG approach based on the use of multivariate statistics will be explored.

Exploratory analyses using subsets of data from the SQO database will be conducted in an effort to develop a SQG approach having improved performance, relative to those developed in Subtasks 2.1-2.3. Two strategies for SQG development will be investigated. The first strategy will use multivariate statistics to identify key contaminants and effects thresholds from the complex mixture of chemicals typical of field data. Principal Components Analysis (PCA) will be used to identify patterns and groupings of abiotic variables (e.g., contaminants, TOC, and grain size) in the dataset. Multiple regression will then be used to identify those factors (groupings of abiotic variables) that are most closely associated with toxicity or benthic impacts. Plots or regression techniques will then be used to identify contaminant thresholds associated with various degrees of adverse biological effects.

The second strategy will use various normalization techniques to reduce the uncertainty in SQG values associated with variations in chemical bioavailability. Metals data will be normalized to a reference constituent such as iron or grain size that corresponds to natural variation in background concentrations. Various normalization strategies for nonpolar organics that include parameters such as organic carbon and the octanol-water partition coefficient will be evaluated for their effectiveness in improving the correspondence between concentration and biological effects.

The development of new SQG approaches will be a lower priority than Subtasks 2.1-2.3 and will not interfere with the completion of the other tasks. The utility of these new approaches will be evaluated through preliminary analyses using subsets of data. If either of these strategies yields encouraging results, then additional effort will be expended to develop a new approach using the full SQG development dataset.

Task 3: Evaluate SQG Approaches. The activities described in Task 2 will result in multiple candidate approaches for interpreting sediment chemistry data. The performance of the candidate approaches will be documented and compared in Task 3 in order to identify those approaches that have the best utility for use as chemistry indicators. Evaluation of the candidate approaches will consist of the following steps:

1. **Compare overall discriminatory power of each approach.** The overall ability of each approach to correctly classify impacted (i.e., toxic or impaired benthos) and unimpacted samples will be evaluated through the use of receiver operating characteristic (ROC) curves (Shine et al. 2003). The ROC method expresses the overall discriminatory power of each approach on a common scale (curve area), which facilitates the comparison of different methodologies. In addition, the ROC method does not require the identification of a specific application threshold (e.g., mean quotient = 1.0). The ROC curve area will be calculated for each candidate approach using a common dataset. Approaches will be ranked on the basis of their area, and those with the best discriminatory power will be selected for further evaluation.
2. **Identify applications.** There are multiple possible likely applications of the chemistry indicators, such as identifying sites with low contaminant exposure/probability of impact versus identifying sites with a high contaminant exposure/probability of impact. Previous studies have shown that the performance of various SQG approaches can vary with respect to the desired application (Vidal and Bay *in press*). Thus, the candidate indicators cannot be effectively evaluated until the desired applications are known. The specific types of SQO applications to be included in this program will be determined based on input from the SWRCB, SSC and Advisory Committee. It is expected that two primary applications will be desired: the identification of likely unimpaired sites and the identification of likely impaired sites.
3. **Calculate performance metrics.** Four basic measures of performance will be calculate for each candidate SQG approach that is selected following Steps 1 and 2 (Shine *et al.* 2003):

   Sensitivity, the rate at which a SQG correctly classifies an impacted sample (toxic or impaired benthos) as impacted;
   S**pecificity**, the rate at which a SQG correctly classifies an unimpacted sample as unimpacted;
   Positive predictive value, the likelihood that a sample exceeding a threshold is impacted; and
   Negative predictive value, the likelihood that a sample below a threshold is unimpacted

   The performance metrics will be calculated for a range of potential application thresholds (e.g., specific SQG quotient values) for each indicator in order to describe the range of performance that can be obtained.
4. **Specify performance objectives.** Each of the performance metrics described in Step 3 relate to a different aspect of SQG performance and some will be of greater significance for specific applications. For example, the specificity and negative predictive value of a chemistry indicator are likely to be most important for an application intended to protect

aquatic life. Some of these metrics will be inversely related to one another (i.e., good performance in one aspect will be accompanied by poor performance in another aspect). The desired range of performance for each metric and their priority must be determined in order to identify the most suitable SQG approaches for each type of application identified in Step 2. The performance objectives for each type of SQO application will be determined based on input from the SWRCB, SSC and Advisory Committee.

5. **Compare and identify the most suitable approaches.** The candidate approaches will be ranked relative to the performance objectives specified in Step 4. These rankings will be summarized in a series of tables or plots that indicate the relative performance for each type of application. The candidate approaches will also be ranked based on other factors affecting their utility for use in the chemistry LOE. These factors will be determined based on input from the SWRCB, SSC and Advisory Committee, and may include: relevance to project goals (indicators of both exposure and effect), applicability (suitable for a range of applications), feasibility (can be applied to typical chemistry data types), generality (feasible for use in a diversity of regions or habitats). Recommendations regarding the most suitable approaches for use as chemistry indicators will be based on a consideration of all of the above factors.

Task 4: Describe Response Levels. The relationship between each of the recommended indicators (Task 3) and biological responses (toxicity and benthic community condition) will be summarized using graphical methods. The error rates associated with specific indicator thresholds (e.g., false negatives and false positives) will also be calculated. This information will be evaluated in the context of the established policy applications and performance objectives to identify response levels or thresholds that will be incorporated into the multiple line of evidence framework.

Schedule

Task	Activity or Deliverable	Completion Date
1: Prepare Datasets	1.1. Evaluate data quality and completeness	November 2004
	1.2. Calculate derived values	November 2004
	1.3. Normalize data	December 2004
	1.4. Determine regional clusters	December 2004
2: Develop and Refine SQGs	2.1. Calculate SQG values using existing approaches	January 2005
	2.2. Refine existing approaches	March 2005
	2.3. Calibrate existing approaches	April 2005
	2.4. Develop new approaches	April 2005
3: Evaluate SQG Approaches	Evaluate SQG approaches	May 2005
4: Describe Response Levels	Describe response levels	June 2005

Literature Cited

Barrick, R., S. Becker, R. Pastorok, L. Brown and H. Beller. 1988. Sediment quality values refinement: 1988 update evaluation of Puget Sound AET. U.S. Environmental Protection Agency. Seattle, WA.

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.

Crane, J.L., D.D. MacDonald, C.G. Ingersoll, D.E. Smorong, R.A. Lindskoog, C.G. Severn and T.A. Berger. 2002. Evaluation of numerical sediment quality targets for the St. Louis River Area of Concern. *Archives of Environmental Contamination and Toxicology* 43: 1-10.

Di Toro, D.M. and J.A. McGrath. 2000. Technical basis for narcotic chemicals and polycyclic aromatic hydrocarbon criteria. II. Mixtures and sediments. *Environmental Toxicology and Chemistry* 19: 1971-1982.

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 sediment chemistry using logistic regression models. *Environmental Toxicology and Chemistry* 21: 1993-2005.

Ingersoll, C.G., D.D. MacDonald, N. Wang, J.L. Crane, L.J. Field, N.E. Haverland, N.E. Kemble, R. Lindskoog, C.G. Severn and D.E. Smorong. 2001. Predictions of sediment toxicity using consensus-based freshwater sediment quality guidelines. *Archives of Environmental Contamination and Toxicology* 41: 8-21.

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., 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.

Schiff, K.C. and S.B. Weisberg. 1999. Iron as a reference element for determining trace metal enrichment in Southern California coastal shelf sediments. *Marine Environmental Research* 48: 161-176.

Shine, J.P., C.J. Trapp and B.A. Coull. 2003. Use of receiver operating characteristic curves to evaluate sediment quality guidelines for metals. *Environmental Toxicology and Chemistry* 22: 1642-1648.

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

U.S.EPA. 2001. National sediment quality survey: second edition. EPA 823-R-01-01. U.S. Environmental Protection Agency, Office of Science and Technology Standards and Health Protection Division. Washington, D.C.

U.S. EPA. 2003. Procedures for the Derivation of Equilibrium Partitioning Sediment Benchmarks (ESBs) for the Protection of Benthic Organisms: Endrin. EPA-600-R-02-009. Office of Research and Development, Washington, DC.

Vidal, D.E. and S.M. Bay. In press. Comparison of sediment quality guideline performance for predicting sediment toxicity in Southern California. Southern California Coastal Water Research Project 2003-2004 Annual Report. Southern California Coastal Water Research Project, Westminster, CA.