

# Calculating Background Levels for Ecological Risk Parameters in Toxic Harbor Sediment

## Authors/Organizations

Christopher J. Leadon<sup>1</sup>, Thomas R. McDonnell<sup>2</sup>, Janet Lear<sup>3</sup>, and David Barclift<sup>4</sup>

<sup>1</sup>Naval Facilities Engineering Command Southwest (NAVFACSW), 1220 Pacific Highway, San Diego, CA 92132; <sup>2</sup>Brown and Caldwell, 400 Exchange, Irvine, CA 92602; <sup>3</sup>Brown and Caldwell, 1230 Columbia Street, Suite 400, San Diego, CA 92101; <sup>4</sup>U.S. Geological Survey, Patuxent Wildlife Research Center, Duty Station: U.S. Environmental Protection Agency Region III, 1750 Arch Street (3HS41), Philadelphia, PA, 19109-2029.

## Abstract

Establishing background levels for biological parameters is necessary in assessing the ecological risks from harbor sediment contaminated with toxic chemicals. For chemicals in sediment, the term contaminated is defined as having concentrations above ambient background and significant human health or ecological risk levels. For biological parameters, a site could be considered contaminated if levels of the parameter are either more or less than the background level, depending on the specific parameter. Biological parameters can include chemical concentrations in the tissue of ecological receptors, bioassay test results, bioaccumulation test results, and benthic community analyses. Chemical parameters can include a variety of potentially toxic chemicals in sediment. Indirectly, contaminated harbor sediment can impact shellfish, fish, birds, and marine mammals, and human populations.

This paper summarizes the methods used to define background levels for chemical and biological parameters from a survey of ecological risk investigations of marine harbor sediment at California Navy bases. Background levels for regional biological indices used to quantify ecological risks for benthic communities are also described. Generally, background stations are positioned in relatively clean areas exhibiting the same physical and general chemical characteristics as nearby areas with contaminated harbor sediment. The number of background stations and the number of sample replicates per background station depend on the statistical design of the sediment ecological risk investigation, developed through the data quality objective (DQO) process. Biological data from the background stations can be compared to data from a contaminated site by using minimum or maximum background levels or comparative statistics.

**Key words:** contaminated harbor sediments, ecological risk, bioassessment, and background reference.

## Introduction

The United States Navy has many installations at which sediment investigations may be necessary. To avoid costly cleanup of large sediment volumes, the Department of the Navy (DON) relies on thorough investigations of contaminated sediment through their Environmental Restoration (ER) program, a cleanup program similar to Superfund. Information on how background levels were determined for ecological risk assessment (ERA) parameters in harbor sediment from several California Navy bases was surveyed and summarized in this paper. The California Navy bases include installations on San Francisco Bay, Port Hueneme, Mugu Lagoon, San Pedro Bay, and San Diego Bay. As well as providing an interesting discussion of various methods, the objective of this paper is to provide a useful reference for readers seeking information on how to calculate background for ecological risk assessments of harbor sediments.

The term background is used here to represent the range of physical, chemical, and biological conditions that are generally ubiquitous in the vicinity of the site to be studied. Background provides a reference against which site-specific conditions can be compared. Background levels for chemical and ecological risk parameters must be established during assessments of harbor sediments to distinguish between ecological conditions resulting from the release of a contaminant at a site and ecological conditions resulting from naturally occurring or non-release anthropogenic conditions. This paper will present a limited review of physical and chemical background evaluations pertaining to the biological characterization of background.

The collection of background samples is necessary to supply data for the calculation of site-specific background levels. Background levels for ecological risk parameters are used in screening ERA's as one of the two fundamental screening levels used to screen a sediment site in or out for further investigation as a contaminated site. The other major screening tool is ecological-risk guideline values which are levels of chemical or biological parameters that have been shown to be associated with ecological risks in similar areas.

The DON's Ecological Risk Policy (DON 1999a) describes the process of Ecological Risk Assessments (ERAs) in three general steps: Tier 1, Screening Risk Assessment; Tier 2, Baseline ERA; and Tier 3, Evaluation of Remedial Alternatives. Calculating background levels for biological parameters is usually part of baseline ERAs in Tier 2. Frequently, applicable background levels for chemical parameters have already been established at another nearby toxic harbor sediment site, so that background screening for chemical parameters can be done in Tier 1 of the ERA process. If established information on background levels for chemical parameters is not available from a nearby sediment site, sampling for chemical and biological parameters and calculation of background levels takes place in Tier 2. Sampling investigations for ERA's in the Navy IR Program follow the seven-step data quality objective (DQO) process (U.S. EPA, 1994) to determine the type, quality, and quantity of data needed to support decisions.

Background levels for ecological risk parameters are used in baseline ERAs to define an incident of an adverse biological condition at a site. Whether an area of concern has an adverse condition is important when using a preponderance-of-evidence approach with multiple lines of evidence,

such as the sediment quality triad approach (Chapman 1990). The triad approach includes assessments of chemical concentrations in sediment or biological tissue, sediment toxicity bioassays, and benthic community analyses. Background levels of chemical contaminants in the tissue of biological receptors can also be important in hazard index calculations for shellfish, fish, and marine birds and mammals. The concentrations of contaminants in the tissue of fish and shellfish are often connected to human-health risk assessments of harbor sediment. The food chain in harbor sediment ranges from minute benthic invertebrates, such as 2-5 mm long amphipods in the sediment, through bottom-feeding fish and birds to higher trophic-order fish, birds, marine mammals, and human fishermen.

This document supplements another DON background sediment guidance document, which provides detailed instructions for evaluating chemical characteristics of background sediment (DON 2002a). This document will focus on the biological, or ecological risk, parameters of background levels for harbor sediment. Discussion is also included regarding the selection of background locations.

## **Ecological Risk Assessments and Data Quality Objectives**

Frequently, applicable background levels for chemical parameters have already been established at another IR site or background area at a military base, so that background screening for chemical parameters can be done in Tier 1 of DON's Ecological Risk Policy for a site. The planning and characterization of background levels for biological parameters are part of the data quality objective (DQO) process used to plan baseline ERAs in Tier 2, Steps 4 through 7. Within Tier 2, Step 4 includes Study Design/DQOs, Step 5 is Verification of Field Sampling Design, Step 6 is Site Investigation and Data Analysis, and Step 7 is Risk Characterization.

ERAs follow the seven-step DQO process as do all sampling investigations in the Navy's IR Program. The use of background data for chemical or ecological risk parameters is part of DQO Steps 2 (Identify the Decision), 5 (Develop a Decision Rule), and 7 (Optimize the Sampling Design). Background screening is mentioned in DQO Substep 2.2, Develop a Logic Diagram of Study Questions and Actions (Bilyard et al. 1997). Screening against background is one of the study questions in a logic diagram of a typical ERA. Screening against background is also part of DQO Substep 5.3, Specify the Action Level for the Study. A background data set for a site is listed as a programmatic aid for Substep 5.4, Develop a Decision Rule. A comprehensive review of background data is part of Substep 7.1, Review DQO Outputs and Existing Environmental Data in Step 7, Optimize the Sampling Design.

## **Selecting Background Stations in Harbor Sediment**

Background stations are chosen in areas outside a contaminated site. Locations with existing data may be evaluated or data may be collected from potential background locations. In addition, regulatory agencies or other stakeholders, may propose specific background locations.

To establish locations of background stations in harbor sediment, the qualities and characteristics of possible locations must be evaluated using preliminary data. The background location should

share as many similar attributes as possible with the site sediment. This paper will present a limited review of physical, chemical, and biological considerations for selecting background locations.

Project designers must determine the number of background station locations. The number of background station locations depends on the statistical design, which is developed through the DQO process. This process will determine the type of qualitative or statistical testing that will be used, the error probabilities, and the number of samples to be analyzed.

Background stations should preferably be located in areas with physical, chemical, and hydrological characteristics similar to the contaminated sediment site. It is also desirable to have background stations that are similar to the site with regard to ecological characteristics, such as habitat type and aquatic communities (U.S. EPA 2000). Samples from both the background stations and the contaminated sediment site should be analyzed for the same target analytes, the chemicals of potential ecological concern (COPECs).

**Proximity to known sources.** Background locations should be selected that are not in proximity to known sources of contaminants to avoid influence from these features on the chemical and biological characterization of background conditions. Permitted point sources such as wastewater discharges should be avoided. Stormwater sources may be less visually evident than wastewater discharges, but they may have the potential to influence the nearby sediment.

**Proximity to site.** Background locations should be selected near the study site, yet not so close as to be influenced by the site itself. Background stations should be located in areas that reflect the ambient anthropogenic background conditions of a bay or harbor but not the contaminant levels of the site under investigation.

The number of background stations used in the surveyed ER Program projects in California summarized here for ecological risk parameters, such as bioassay samples, ranged from one to ten, with five being the most common. This represents the probable number of background stations that would be needed at a new toxic harbor sediment site when statistics, regulatory requirements, and habitat characteristics are all taken into consideration. Five replicate background samples were collected at the ER site that had only one background station. This was the only background station among those surveyed in the survey of California Navy harbor sediment sites where replicate samples were collected in the field. For the remaining stations, replicate samples were prepared in the laboratory from single bioassay samples collected in the field.

**Depth and hydrology.** Hydrological characteristics that should be similar include depth of water, salinity (i.e., freshwater, estuarine, or marine), and flow (i.e., lentic or lotic).

**Grain size and organic carbon.** Background stations and the site should exhibit similar physical characteristics, such as sediment grain size and share general chemical characteristics, such as sediment organic carbon content and sulfides.

**Chemical content.** Evaluation of background biological parameters usually accompanies evaluation of background chemical concentrations. While selection of the target chemical compounds, those for which analyses will be run, is typically a site-specific effort, target chemical compounds usually include metals, chlorinated pesticides, polychlorinated biphenyls,

volatile organic compounds, and semi-volatile organic compounds. Other general sediment characteristics are often evaluated, such as grain size, total organic carbon, total sulfides, and acid volatile sulfides and simultaneously extracted metals.

**Toxicity characteristics.** Ideally, background conditions would not exhibit toxicity responses by test organisms in laboratory tests. However, ambient conditions may influence the laboratory toxicity response. Potential background locations for which toxicity data are available can be compared to laboratory control data and a protocol-specific performance threshold. The objective of this approach is to avoid a conclusion of statistically significant toxicity based on laboratory negative control data when the actual difference is low. Based on work similar to Thursby et al. (1997), Phillips et al. (2001) evaluated more than 1,000 samples from the California Bay Protection and Toxic Cleanup Program using t-tests and calculating a minimum significant difference (MSD). Phillips et al. (2001) selected the 90th percentile MSD as a toxicity threshold value. Thresholds are given as percent of negative control and are available for three amphipods (survival), two mollusks (development), a polychaete (survival and growth), and a sea urchin (fertilization and development). By presenting this value as a percent of negative control results (i.e., normalizing by the negative control results), organism response bias that may be present in a particular bioassay test can be minimized.

MSDs are calculated for each t-test comparison using the following equation:

$$MSD = t_{critical} + \left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^{-1/2}$$

where

MSD = minimum significant difference

$t_{critical}$  =  $t$  value from the standard statistical table, for  $\alpha = 0.05$  and the appropriate degrees of freedom

$s_1^2, s_2^2$  = variances for treatments (control and field sample)

$n_1, n_2$  = numbers of replicates for treatments (control and field sample)

**Benthic community, bioaccumulation, and wildlife risk.** Generally, selection of background locations does not include an evaluation of the benthic community, bioaccumulation, or wildlife risk. The assumption is often made that if the previously stated objectives for the selection of background locations based on physical, chemical, and toxicity data are achieved then the benthic community, bioaccumulation, and wildlife risk will be acceptable.

## Background Levels for Chemicals in Sediment

Evaluation of background biological parameters usually accompanies evaluation of background chemical concentrations. While selection of the target chemical compounds, those for which analyses will be run, is typically a site-specific effort, target chemical compounds usually include metals, chlorinated pesticides, polychlorinated biphenyls, volatile organic compounds, and semi-volatile organic compounds. Other general sediment characteristics are often evaluated, such as

grain size, total organic carbon, total sulfides and acid volatile sulfides, and simultaneously extracted metals. This document presents only a brief summary of steps for describing sediment background chemical concentrations since this subject has been thoroughly addressed in other documents (DON 2002a).

There are three general approaches to characterize the background chemical concentrations and to compare them to site chemical concentrations: exploratory data analysis, geochemical background analysis (the geochemical method), and comparative statistical analysis (the comparative method) (DON 2002a). The exploratory data analysis method examines distribution patterns for an upper limit background threshold. The use of threshold levels to represent background levels in screening ERAs with the exploratory data analysis method has been shown to result in a high false-positive rate (i.e., declaring a site contaminated when in fact it is not) (DON 2002a). Instead of using the exploratory data analysis method to determine background levels by calculating upper threshold limits, use of either the comparative or geochemical method is recommended (DON 2002a).

The geochemical method evaluates association relationships between concentrations of COPECs and concentrations of naturally occurring background chemicals, such as aluminum and iron, or natural sediment characteristics, such as grain size or organic carbon. This method can be used when it is not possible to identify a background area, because the method does not require background area data for comparison (DON 2002a). The comparative method compares chemical concentration distributions for the site and for background locations using statistical estimates of the data distributions. With the comparative method, the statistical comparative tests are either parametric or nonparametric. Parametric tests assume the background and site data can be represented by specific statistical distributions. Nonparametric tests do not make this assumption (DON 2002a). Some of the comparative statistical tests evaluate the similarity of extreme site and background concentrations, while other tests evaluate aspects like central tendencies (medians or means of the site and background data).

The procedures for calculating background levels for ecological risk parameters in sediment are similar to the comparative method used to screen chemical concentrations in sediment. However, instead of being a comparison of the whole data distribution from the background stations to the data distribution from all the site sampling stations as a group, the data from each individual site station for an ecological risk parameter is compared to the combined data for that parameter from the background stations. The comparison of site chemical concentrations to background concentrations often takes place in the screening phase of an ERA, while the comparison of site ecological risk data to background reference station data is part of a baseline ERA.

## **Bioassays**

Sediment toxicity tests, which include an array of assays of various biological organisms and response endpoints (e.g., survival, development), provide specific information for the selected species under laboratory conditions. Bioassay protocols are available for the major taxonomic groups of sediment organisms, such as polychaete worms, crustaceans, mollusks, and

echinoderms (DON, 2002b). Sediment evaluations from the general region of the site should be reviewed for common types of bioassays used. Using the same bioassay tests from other regional studies provides the benefit of a broad database for comparison and interpretation of results. For example, use of the 10-day amphipod (crustacean) survival test with *Rhepoxynius abronius* would be appropriate for testing marine sediment in many United States coastal areas.

The DON presents useful summaries of the available bioassay methods for various habitats, media, and response endpoints (DON 2002b, 2003). Principal response endpoints include survival, growth, and development. Media alternatives include bulk sediment, water (e.g., porewater), and sediment-water interface. Endpoints, test organisms, and media should be selected in accordance with the project conceptual site model and the DQOs. For purposes of this discussion, the desired bioassay endpoint will be evaluated (i.e., survival, development, and growth) in contrast to the opposite endpoint (i.e., mortality, impaired development, and impaired growth).

Selection of bioassays should be based on applicability to selected assessment endpoints, suitability to site environment (e.g., temperature and salinity), and desired range of sensitivity. The number of background station samples or bioassay analyses should be determined using the DQO process, assuming that a statistical comparison of the background data and site data is desired. Each background station represents one background sample.

Before implementing the selected bioassays, potential confounding factors should be considered. Several confounding factors have been reviewed by the DON (2000). Ammonia and sulfides are toxic to a wide range of marine organisms, and both occur as a result of natural bacterial action on decaying sediment-organic matter in marine sediment. General sediment characteristics such as grain size and total organic carbon can interfere with bioassay test results, typically due to the sensitivities of selected test species. Porewater characteristics, such as salinity and dissolved oxygen, may interfere with test results if they are different from the characteristics of the overlying water. Additionally, salinity and dissolved oxygen may interfere with test results due to their effect on other parameters such as ammonia.

Before analyzing bioassay data for background conditions, the bioassay data should be compared to the laboratory control data. Several types of laboratory control data are collected during bioassays tests to aid in the interpretation of results. Positive controls or reference toxicity tests are implemented to assess the responsiveness of the test organisms. The organisms should respond to a given toxicant according to an established dose-response curve. Negative controls are implemented to assess the health of the test organisms. The negative controls should show that in the absence of a toxicant and under normal laboratory conditions, the test organisms display a strong normal response for the measured endpoint. The timeline control is unique to bioaccumulation bioassay tests in that it records the amount of COPECs present in the tissue of the test organisms prior to exposure to site sediment.

Bioassay data exhibit a fundamental difference from chemical data because of the replicates for each sample. Typically, a single sediment sample is collected from an individual background location. Chemical analysis generates a single value per target chemical. Conversely, bioassay samples are analyzed using replicates prepared in the laboratory. The replicate samples are used

to test for a statistically significant difference between the distribution of replicate test values for the site and the distribution of replicate values for the background location.

ER sites have characterized background bioassay results using the arithmetic mean and various threshold values. As previously stated, the use of some threshold levels to represent background levels has been shown to result in a high false-positive rate (declaring a site contaminated when in fact it is not) (DON 2002a). Comparison of background data with site data can be evaluated using statistical methods. A significant difference between the background results and the site results indicates an adverse response at the site that cannot be attributed to variation in the background sediment conditions. A thorough presentation of statistical methods is available in DON guidance (DON 2002a).

The most common comparison technique in the ER Program projects summarized here is use of a t-test with pooled background data and pooled site data. However, this technique does not provide a finding for each individual station, which is often the basis of the evaluation. Individual stations have been evaluated using four techniques.

The first technique is the t-test ( $p < 0.05$ ) comparing the distribution of the laboratory replicates for a site station with the distribution of the pooled replicates from all of the background stations. In this case, the laboratory replicates from individual background stations are grouped together as if they were replicates of a single background station.

A second technique is using the 95th percent lower predictive limit (LPL) as a threshold value for bioassay results from background stations. Most bioassay data are presented as survival. The background data can be seen as a bell-shaped curve with a mean at the center of the curve. Samples from contaminated sites would result in lower survival than samples from background stations. The lowest survival data would be off to the left of the bell-shaped curve. Therefore, an LPL of the background population is needed. The threshold value representing the minimum acceptable background bioassay survival would be the 95 percent LPL, on the left side of the bell-shaped curve.

LPL values were calculated from the project data set as:

$$LPL = \bar{x} - t\sigma\sqrt{\left(1 + \frac{1}{n}\right)}$$

where

- $LPL$  = lower predictive limit
- $\bar{x}$  = mean of background samples
- $t$  = one-sided  $t$  statistic, alpha of 0.05
- $\sigma$  = standard deviation of background samples
- $n$  = number of background samples

A third technique is using the reference envelope tolerance limit as a threshold value for bioassay results from background stations. The reference envelope lower tolerance limit (LTL) values were calculated from a regional data set from San Francisco Bay (Hunt et al. 1998) as:

$$LTL = \bar{x} - (g_{\alpha,p,n} * \sigma)$$

where

- LTL = lower tolerance limit
- $\bar{x}$  = mean of background samples
- g = g statistic (estimated using a bootstrap technique)
- $\sigma$  = standard deviation of background samples
- $\alpha$  = alpha value 0.05
- p = percentile value 10
- n = number of background samples

The fourth technique used to evaluate a single station is a relative percent difference (RPD) of 20 percent between the site station bioassay result and the background threshold value (minimum or mean). The RPD method was used in an effort to identify a biologically significant difference between the site data and the background data, which may have been identified as statistically different but in fact is quite small. However, there is no validation that a 20 percent RPD is biologically significant.

Some IR Program projects have used an additional comparison to evaluate toxicity based on the protocol-specific precision performance of the bioassay test. As described earlier, the MSD approach (Thursby et al. 1997, Phillips et al. 2001) eliminates findings of toxicity when the results are statistically different from the laboratory controls but not different based on the protocol-specific performance threshold.

## Bioaccumulation

Bioaccumulation studies measure the concentration of chemical residue in an organism's tissue. Bioaccumulation is the increase of a chemical's concentration within an organism due to exposure to the chemical in the organism's environment. In an aquatic environment, mechanisms of bioaccumulation include direct absorption from the sediment or water medium and absorption from ingested food items. If elimination mechanisms such as excretion are more efficient than the uptake due to exposure mechanisms, bioaccumulation does not occur.

Bioaccumulation may be assessed by using models or by direct measurement. Due to the high uncertainties associated with bioaccumulation models, the direct or empirical measurement of bioaccumulation is often selected for baseline ERA. The direct measurement approach may be accomplished by using laboratory-exposed organisms, resident site-exposed organisms, or caged transplanted site-exposed organisms.

Project design objectives determine whether the measured tissue is from laboratory organisms, resident organisms of the site, or transplanted organisms. The project design objectives also determine if the measured tissue will be from specific organs, such as liver and muscle, or from a

whole-body homogenate. United States Environmental Protection Agency (U.S. EPA) provides a thorough discussion of advantages and disadvantages of the various approaches for assessing bioaccumulation (U.S. EPA 2000).

Laboratory-exposed organisms are selected to model certain species or groups of species that occur at a site. The DON presents useful summaries of the laboratory organisms available for bioaccumulation studies (DON 2002b, 2003). Common organisms include a clam (*Macoma nasuta*) and a worm (*Nereis virens*). The laboratory studies measure the chemical uptake by standard laboratory organisms exposed to site sediment under laboratory conditions. These results may then be used as surrogate data for site-specific organisms under site-specific conditions.

Alternatively, organisms under site-specific conditions can be measured directly by sampling the resident site-exposed biota. Fish sampling is often conducted to assess the chemical residue in the tissue of resident biota. Even though the biota are collected from the site, it is often difficult to determine what percentage of the exposure has been site-specific. Fish, for example, are generally quite mobile, and the tissue residue may be due to complex exposure conditions from site and off-site sources. The results of the field collection efforts are often inconsistent between sample locations, and target species may be available at one site but not at another.

Transplanted site-exposed organism studies control some of these uncertainties by confining organisms to cages to assure that exposure is site-specific and that a sufficient quantity of the target species is collected. Special consideration for the design and placement of the field exposure equipment (e.g., cages) should be taken to avoid any interference with site activities, such as vessel traffic, or site organisms that may prey on the test organisms.

A bioaccumulation background condition can be described for each of the available approaches. Selection of background locations is important so that the exposure conditions are as similar as possible to exposure conditions in resident site-exposed organisms or caged, transplanted site-exposed organisms.

Of the sediment IR sites reviewed, most used a 28-day laboratory-exposed clam to assess bioaccumulation. When pooled site data were compared to pooled background data, significantly different bioaccumulation was determined using a t-test or a Wilcoxon rank sum test. The t-test method compares the two groups of data with certain assumptions, such as the data in each group approximating a normal distribution. The Wilcoxon rank sum test method compares the two groups of data with few assumptions and uses the rank value of the data. When site sample locations ( $n = 1$ ) were individually compared to pooled background data, elevated bioaccumulation values were determined with a threshold value based on the background data such as the maximum, 95 percent upper confidence limit (UCL), and 90th percentile. However, as noted earlier, comparison of site chemical concentrations to background threshold values (such as the maximum, 95 percent UCL, or 90th percentile of the background data) may have unacceptable error.

## Benthic Community Structure

In its most fundamental presentation, benthic community analysis enumerates and identifies every individual organism within the sampling area. However, because the results of such an analysis can present more than 100 species and more than 1,000 individuals, interpretation of results can be complex. These data can be combined according to various theories to present estimates of diversity. Other indices, often based on regional data, are available that are constructed around estimates of the likelihood of representative species to occur along a gradient of sediment quality. Multivariate techniques, such as cluster analysis, are also available.

The benthic invertebrate community can have a very irregular distribution as certain populations respond to varying conditions in small areas. Additionally, benthic invertebrate populations respond to stressors other than the COPECs, such as temperature, salinity, and physical disturbances. Therefore, interpreting the benthic community data can be a significant challenge. Consequently, this line of evidence, which is the most direct measure of the biological characterization of a site, is often omitted from project designs. This is why background stations are so important for the interpretation of benthic invertebrate community structure data from toxic harbor sediment sites. At sites where suitable background stations are available, the non-COPEC responses are matched at the background and site stations and can then be removed from consideration as COPEC effects. When included in ERAs of sediment, benthic community parameters add a direct non-laboratory estimate of benthic community conditions that can be useful for comparisons to other ecological risk parameters at sediment sampling stations.

The typical characterization of benthic community includes the following indices:

Abundance	=	total number of individual specimens identified in a sample
Species richness	=	total number of distinct taxonomic identifications per sample (nominally, the number of species)
Dominance (Swartz et al. 1985)	=	number of species accounting for 75 percent of the total abundance
Shannon-Wiener diversity (Shannon and Weaver 1962)	=	$-1 * \sum p_i * \ln(p_i)$ where $p_i = \frac{\text{abundance of species } i}{\text{total abundance}}$
Evenness (Pielou 1977)	=	$\frac{-1 * \text{Shannon - Wiener}}{\ln(\text{species richness})}$
Margalef's species richness	=	$\frac{\text{species richness} - 1}{\ln(\text{total abundance})}$

(Margalef 1951)

Comparison of distributions for benthic community indices using statistical tests such as the t-test is not often conducted for sediment assessment projects. However, this comparison is not uncommon for monitoring programs. For those benthic community indices with background levels by definition above levels at a contaminated site, threshold values for the indices have been estimated using the minimum background threshold value. However, as noted earlier, comparison of site data to background threshold values, such as the minimum value of the background data, may have unacceptable error. The LPL, presented in the discussion of the bioassay data, may present a more appropriate estimate of the background threshold than a minimum value. For those benthic community indices with background levels by definition less than levels at a contaminated site, threshold values for the indices have been estimated using the maximum background threshold value. An upper predictive limit (UPL) may present a more appropriate estimate of the background threshold than a maximum.

Confidence in interpreting benthic community data can be gained by using multiple assessment tools based on substantially different approaches. Since most diversity indices are based on the distribution of the abundance across the species richness, they share the same strengths and weaknesses. Two independent indices have been developed in southern California: the relative benthic index based on regional data from San Diego Bay (Anderson et al. 1997, Fairey et al. 1996) and the benthic response index based on regional data from southern California (Smith et al. 1998). Both of these indices present threshold values that identify a degraded benthic community.

The relative benthic index (RBI), developed as part of the California Bay Protection and Toxic Cleanup Program, was initially presented by Fairey et al. (1996) and subsequently revised by Anderson et al. (1997) and Fairey et al. (1998). The index averages three subindices that are scaled relative to the range of values addressed. Indices such as the RBI may be implemented for a time until a new index is proposed. The RBI was used for several sediment investigations in California (Jacobi et al. 1998, Anderson et al. 1997, 1998, Phillips et al. 1998, Hunt et al. 1998, Fairey et al. 1996), but appears to be giving way to a new index, the benthic response index.

$$\text{Relative Benthic Index} = \frac{\sum \left( \begin{array}{l} \text{Species Richness Index} \\ + \text{Crustacea Richness Index} \\ + \text{Indicator Species Index} \end{array} \right)}{3}$$

The benthic response index (BRI), developed as part of the Southern California Bight Regional Monitoring Program, was initially presented by Smith et al. (2001) and subsequently revised by Smith et al. (2003) for application to southern California bays. The BRI is the average pollution tolerance for species occurring at the sample location weighted for species-specific abundance. Pollution tolerance values for each species were developed using multivariate analysis and are generally based on effects-range median data and amphipod toxicity data. Background levels for

a regional parameter such as the BRI are based on the accumulated data from harbor sediments in southern California that had been used initially to construct the index (Smith et al. 2003).

$$\text{Benthic Response Index} = \frac{\sum_{i=1}^n (p_i \sqrt[3]{a_{si}^f})}{\sum_{i=1}^n \sqrt[3]{a_{si}^f}}$$

where

n = total number of individual specimens identified in a sample

p<sub>i</sub> = pollution tolerance of species i

a<sub>si</sub> = abundance of species i

f = exponent coefficient

Cluster analysis is a multivariate analysis intended to detect groupings in data. This procedure can be applied to site and background reference locations according to similarity of benthic community structure. The Bray-Curtis similarity index is commonly used with cluster analysis to provide groupings of stations based on benthic community data. The results of cluster analysis are often expressed graphically in a dendrogram that facilitates the interpretation of the results.

Only two of the sediment investigations at California Navy bases evaluated the *in situ* benthic community. Both investigations evaluated the benthic community data from the contaminated site using cluster analysis with the Bray-Curtis similarity index and threshold values (minimum or maximum values) for several community indices. Background levels for benthic community indices were determined by using the appropriate minimum or maximum values for the indices from reference stations. One of the studies included calculation of the benthic response index.

## Hazard Quotient Evaluation for Wildlife Risk Calculations

Potential ecological risk to aquatic wildlife such as birds and mammals is often assessed using the chemical-specific and receptor-specific hazard quotient (HQ) approach: calculating the ratio of the exposure dose to the screening-level dose. The HQ calculations are described in detail in U.S. EPA (1997) and DON (2001) guidance.

The exposure dose for aquatic wildlife is based on ingestion of food items that may have bioaccumulated chemicals from sediment at the site. The chemical concentrations in the food items are estimated by literature-based bioaccumulation factors, site-specific bioaccumulation factors, or chemical analysis of site-specific food items. Screening-level doses are frequently identified as toxicity reference values (TRVs). HQ values not exceeding 1 are generally interpreted as indicative of acceptable ecological risk. HQ values exceeding 1 in a screening-level ERA are generally interpreted as sufficient cause for further investigation. HQ values exceeding 1 in a baseline risk assessment are generally interpreted as indicative of adverse ecological effects, relative to the specific assumptions made during model development.

The evaluation of background using HQ's is based on the results of the bioaccumulation studies described above. The bioaccumulation studies provide data that may be used to estimate the chemical concentrations present in prey species that larger birds and mammals will consume. Design of the bioaccumulation studies should take into consideration the models that will be used for the wildlife assessment. Relevant design considerations include how well the species used for bioaccumulation measurements represents the wildlife prey species and how well the tissue analyzed represents the portion of the wildlife prey species that is consumed.

Background HQ values, based on background bioaccumulation data, can be compared to HQ values calculated from site data because most of the exposure and toxicity factors, such as ingestion rate and TRVs, would be the same for both HQs. A few exposure factors used to calculate the background HQ and the site HQ (such as site use factor) may be different due to differences in conditions at the background location and at the site.

The site assessment often eliminates or screens out inorganic chemical compounds such as metals based on a statistical comparison of the background and site chemical concentrations. Inorganic compounds with concentrations that are not significantly greater than the background concentrations may be eliminated from further evaluation. On the other hand, all organic compounds reported by the bioaccumulation studies should be evaluated with the HQ approach. Characterization of the site-related risk should include a comparison of the site HQ values to the background HQ values. In a screening-level ERA, it is common to have many chemical compounds with background HQ values that are greater than 1 due to the very conservative nature of the TRVs.

In some circumstances, HQ values for individual chemical compounds are summed for a hazard index value. The intent of summing the individual HQ values is to estimate the cumulative toxic effect of multiple stressor chemicals. The value of the hazard index as an indicator of potential cumulative toxic effects depends on identification of modes of toxicity for each stressor chemical. Only those chemicals with similar modes of action will have the potential for additive toxic effects. See U.S. EPA (1997) and DON (2001) guidance for further discussion of the use of the hazard index value.

## **Discussion**

The methods discussed in this paper for determining background for ecological risk parameters in sediment represent practical methods used at California Navy bases in studies of harbor sediment potentially contaminated with toxic chemicals. Although the selection of background sites is irrespective of high laboratory costs, the high cost of the laboratory methods for ecological risk parameters limits the number of background samples collected for the purpose of decreasing uncertainty. The comparative method is used because it allows comparison of sample data from the possibly contaminated part of a harbor with sample data from background stations located in areas thought to represent typical, noncontaminated, ambient conditions.

The statistical parameters discussed in this paper to represent background ecological risk for sediment include threshold values (e.g., lower predictive limits or lower tolerance limits). These threshold values were not considered as susceptible to false positives as other threshold values

such as maximums and percentiles. It has been shown that representing background levels with threshold values such as maximums and percentiles leads to a greater likelihood of false-positives (i.e., declaring a site contaminated when in fact it is not) (DON 1999b, 2002a).

To avoid possible false positives, a better approach might be to combine the background data set from the background stations with the site data set for exploratory data analysis, as described in Section 2.2 of the DON guidance (DON 2003). In general, the exploratory data analysis in Section 2.2 includes the following steps: 1) determine the probability distribution of the investigated data, 2) compute descriptive summary statistics of measured values, 3) compute representative exposure concentrations for risk screening, 4) identify potential outliers, and 5) determine background ranges (DON 2003). The univariate, post plot, and probability plot analyses presented in Section 2.2.4 of the DON guidance could then be used to estimate more accurately the lower, or an upper, limit of the background range for a particular ecological risk parameter. This estimate of the lower, or upper, background limit for a particular ecological risk parameter would be compared with single data points from the site area during a baseline ERA.

In ERAs of sediment, site data are first screened with a variety of ecological risk screening guideline values such as probable effect levels (PEL's) or effects range-median's (ERM's). If the maximum concentration of the chemical data in the site sediment is greater than the maximum or average concentrations of the ecological risk screening guideline values, the site fails the screening-level ERA and a baseline ERA is performed. In the baseline ERA, a one-to-one comparison is made between a lower or an upper limit representing the background level for a particular ecological risk parameter and a single value of the ecological risk parameter at a site sampling location or from an area of potential concern (AOPC) at the site. Those sampling stations or AOPCs with values of ecological risk parameters indicating more ecological risk than background levels are considered for possible remediation. Sampling stations or AOPCs with values indicating less ecological risk than background limits are recommended for no further action.

## **Conclusions and Recommendations**

The general design of the biological measurements in a sediment investigation should closely parallel the physical and chemical sediment measurements. The similarity of sampling design and sample location will allow for the investigation of association relationships among the data. Selection of background stations should follow the recommendations of the DON (2002a). When planning the background locations, particular attention should be given to sediment grain size and organic carbon content. The number of background stations should be developed according to the project DQO process. Based on a survey of sediment investigations for California Navy bases, five background sediment locations may be considered a minimum number of background stations when statistics, regulatory requirements, habitat characteristics, and cost are all considered.

Analysis of background sediment samples for physical characteristics and chemical constituents should be guided by the project DQO process. Evaluation of the background sediment physical and chemical data should follow guidance of the DON (2002a).

Bioassays should be selected through the project DQO process and include those test species that are applicable to the site conceptual model. Background conditions can be described by the data distribution or by a calculated threshold value. The background data distribution can be statistically compared to the data distribution for site locations using a test such as a t-test. Alternatively, background threshold values may be calculated as the LPL of the background data. Statistical significance identified by either the comparisons of data distributions or a threshold value should be confirmed for relevance by reviewing overall protocol performance with a test such as the MSD threshold.

Bioaccumulation data should be collected according to project DQOs to estimate the potential bioaccumulation for exposure pathways identified in the conceptual site model. The bioaccumulation data can be evaluated with exploratory data analysis and the comparative method.

Benthic community assessment is more complex than the other sediment investigation measurements because none of the available benthic community measurements includes the state of the benthic community. Two fundamental types of data are generated by benthic community measurements: a list of species (or best possible taxonomic level) and an enumeration or count of the individual organisms belonging to the listed species. These data can be summarized into various indices considered reflective of the condition of the benthic community. The results of these indices are typically similar in the resulting interpretation of conditions in benthic communities at a harbor sediment station due to the limited two-variable mathematical structure of these indices.

Therefore, calculating additional indices does not necessarily increase confidence in the data interpretation. The benthic community indices range from simple to complex integrations of the two fundamental benthic community measurements and have from traditional to contemporary origins. However, none of the indices is considered a definitive measurement of the benthic community.

Additionally, a number of factors can affect the state of the benthic community at a given point in time. Therefore, setting narrow, stand-alone measurement endpoints, such as the number of crustacean individuals or the number of molluscan species is impractical. However, setting several measurement endpoints for benthic community performance can overcome the uncertainty attributed to a single endpoint. Selecting measurement endpoints that are based on different data analysis strategies can overcome the uncertainty that may occur with a single strategy. A concurrence of findings from a multivariate technique such as cluster analysis, a regional benthic community index such as the BRI in southern California, and conventional community indices such as abundance and diversity can identify background reference conditions.

Potential risk to wildlife such as mammals and birds is often evaluated according to the HQ approach. HQ values are based on estimated exposures developed from measurement of various site media such as soil, water, or biota (tissue). An evaluation of the background conditions and comparison to site conditions should be completed before estimating the exposure dose and calculating HQ values. Inorganic compounds can be eliminated if site data are statistically less

than background data. HQ values should be prepared for each remaining COPEC and compared qualitatively to HQ values from background data.

## **Acknowledgements**

The authors would like to thank the many people in the Navy Risk Assessment Working Group, at Brown and Caldwell and Bechtel Environmental Inc., and with the Soil and Sediment Contamination: an International Journal who reviewed this issue paper. The views expressed herein are the authors' own and not the official views of the United States Navy or that of Naval Facilities Engineering Command.

## References

- Anderson, B., J. Hunt, S. Tudor, J. Newman, R. Tjeerdema, R. Fairey, J. Oakden, C. Bretz, C. Wilson, F. La Caro, G. Kapahi, M. Stephenson, M. Puckett, J. Anderson, E. Long, T. Fleming, and K. Summers. 1997. Chemistry, Toxicity and Benthic Community Conditions in Sediments of Selected Southern California Bays and Estuaries, Final Report. California State Water Resources Control Board. Sacramento, California.
- Anderson, B., J. Hunt, B. Phillips, J. Newman, R. Tjeerdema, C. Wilson, G. Kapahi, R.A. Sapudar, M. Stephenson, M. Puckett, R. Fairey, J. Oakden, M. Lyons, and S. Birosik. 1998. Sediment chemistry, toxicity, and benthic community conditions in selected water bodies of the Los Angeles Region. Final Report, California State Water Resources Control Board. August.
- Bilyard, G.R., H. Beckert, J.J. Bascietto, C.W. Abrams, S.A. Dyer, and L.A. Haselow. 1997. Using the Data Quality Process During the Design and Conduct of Ecological Risk Assessments. U.S. Department of Energy, Assistant Secretary for Environment, Safety, and Health, Office of Environment, Washington, DC.
- Chapman, P.M. 1990. The sediment quality triad approach to determining pollution-induced degradation. *Sci. Total Env.* 97(8):815–825.
- DON (Department of the Navy). 1999a. Memorandum: Navy Policy for Conducting Ecological Risk Assessments, Chief of Naval Operations, 05 April.
- . 1999b. Handbook for Statistical Analysis of Environmental Background Data, Southwest Division, Naval Facilities Engineering Command (NAVFAC) and Engineering Field Activity West (EFA West), San Diego, California, July.
- . 2000. Confounding Factors in Sediment Toxicology, Navy Guidance for Conducting Ecological Risk Assessments, Issue Papers. Naval Facilities Engineering Command (NAVFAC), (October 24, 2000); <http://web.ead.anl.gov/ecorisk/issue/>.
- . 2001. Navy Guidance for Conducting Ecological Risk Assessments, Revised January 30, 2001; <http://web.ead.anl.gov/ecorisk/process/pdf/index.cfm>.
- . 2002a. Guidance for Environmental Background Analysis, Volume II: Sediment, Naval Facilities Engineering Command (NAVFAC). Prepared by Battelle, Earth Tech, Inc., and NewFields, Inc.
- . 2002b. Implementation Guide for Assessing and Managing Contaminated Sediment at Navy Facilities. Naval Facilities Engineering Command (NAVFAC), Prepared by SPAWAR Systems Center (SSC) San Diego.
- . 2003. Toxicity Tests for Various Habitats and Media, Navy Guidance for Conducting Ecological Risk Assessments, Methods and Tools. Naval Facilities Engineering Command (NAVFAC); <http://web.ead.anl.gov/ecorisk/index.cfm>.
- Fairey, R., C. Bretz, S. Lamerdin, J. Hunt, B. Anderson, S. Tudor, C. Wilson, F. La Caro, M. Stephenson, M. Puckett, and E. Long. 1996. Chemistry, Toxicity, and Benthic Community

Conditions in Sediments of the San Diego Bay Region, Final Report. California State Water Resources Control Board, Sacramento, California.

Fairey, R., J. Downing, C. Roberts, E. Landrau, J. Hunt, B. Anderson, C. Wilson, G. Kapahi, F. La Caro, P. Michael, M. Stephenson, and M. Puckett. 1998. Chemistry, Toxicity, and Benthic Community Conditions in Sediments of the San Diego Bay Region, Final Addendum Report. California State Water Resources Control Board, Sacramento, California.

Hunt, J.W., B.S. Anderson, B.M. Phillips, J. Newman, R. Tjeerdema, M. Stephenson, M. Puckett, R. Fairey, R.W. Smith, and K. Taberski. 1998. Evaluation and Use of Sediment Reference Sites and Toxicity Tests in San Francisco Bay, Final Report. California State Water Resources Control Board, Sacramento, California, San Francisco Bay Regional Water Quality Control Board, California Department of Fish and Game, Institute of Marine Sciences University of California Santa Cruz.

Hunt, J., B. Anderson, B. Phillips, J. Newman, R. Tjeerdema, K. Taberski, C. Wilson, M. Stephenson, M. Puckett, R. Fairey, and J. Oakden. 1998. Sediment quality and biological effects in San Francisco Bay. Bay protection and toxic cleanup program. Final Technical Report, California State Water Resources Control Board. August.

Jacobi, M., R. Fairey, C. Roberts, E. Landrau, J. Hunt, B. Anderson, B. Phillips, C. Wilson, G. Kapahi, F. LaCaro, B. Gwynne, M. Stephenson, and M. Puckett. 1998. Chemical and biological measures of sediment quality and tissue bioaccumulation in the North Coast Region. Final Report, California State Water Resources Control Board. October.

Long, E.R., Field, L. J., MacDonald, D.D. 1998. Predicting Toxicity In Marine Sediments With Numerical Sediment Quality Guidelines. *Environ. Toxicol. Chem.* 17(4):714–727.

Margalef, R. 1951. Diversidad de especies en las comunidades naturales, *Publ. Inst. Biol. Apl. Barcelona* 9:5–27.

Phillips, B., B. Anderson, J. Hunt, J. Newman, R. Tjeerdema, C. Wilson, E.R. Long, M. Stephenson, M. Puckett, R. Fairey, J. Oakden, S. Dawson, and H. Smythe. 1998. Sediment chemistry, toxicity and benthic community conditions in selected water bodies of the Santa Ana Region. Final Report, California State Water Resources Control Board. August.

Phillips, B.M., J.W. Hunt, B.S. Anderson, H.M. Puckett, R. Fairey, C.J. Wilson, and R. Tjeerdema. 2001. Statistical significance of sediment toxicity test results: threshold values derived by the detectable significance approach. *Environ. Toxicol. Chem.* 20 (2):371–373.

Pielou, E.C. 1977. *Mathematical Ecology*, John Wiley and Sons, New York.

Shannon, C.H. and W. Weaver. 1962. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Illinois.

Smith, R. W., M. Bergen, S.B. Weisberg, D.B. Cadien, A. Dalkey, D.E. Montagne, J.K. Stull, and R.G. Velarde. 1998. Benthic response index for assessing infaunal communities on the mainland shelf of Southern California. In: *Southern California Bight 1994 Pilot Project: IV Benthic Infauna*, M. Bergen, S. B. Weisberg, D. Cadien, A. Dalkey, D. Montagne, R. W. Smith, J. K. Stull, and R. G. Velarde, eds., Southern California Coastal Water Research Project, Westminster, California, pp. E1-E86.

- Smith, R. W., Bergen, M., Weisberg, S. B., Cadien, D. B., Dalkey, A., Montagne, D. E., Stull, J. K., and Velarde, R.G. 2001. Benthic response index for assessing infaunal communities on the Southern California mainland shelf. *Ecol. Appl.* 11:1073–1087.
- Smith, R.W., J.A. Ranasinghe, S.B. Weisberg, D.E. Montagne, D. Cadien, T. Mikel, R. Velarde, and A. Dalkey. 2003. Extending the Southern California benthic response index to assess benthic conditions in bays. In: *Southern California Bight 1998 Regional Monitoring Program: VII Benthic Macrofauna*, J. A. Ranasinghe, D. E. Montagne, R. W. Smith, T. K. Mikel, S. B. Weisberg, D. Cadien, R. Velarde, and A. Dalkey, eds., Southern California Coastal Water Research Project, Westminster, California, pp. C1-C42.
- Swartz, R. C., D.W. Schultz, G.R. Ditsworth, W.A. DeBen, and F.A. Cole. 1985. Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall. In: *Validation and Predictability of Laboratory Methods for Assessing the Fate and Effects of Contaminants in Aquatic Ecosystems*, T. T. Boyle, ed., American Society for Testing and Materials, Philadelphia, Pennsylvania.
- Thursby, G.B., J. Heltshe, and K.J. Scott. 1997. Revised Approach to Toxicity Test Acceptability Criteria Using a Statistical Performance Assessment. *Environmental Toxicology and Chemistry* 16:1322–1329.
- U.S. EPA (U. S. Environmental Protection Agency). 1994. Guidance for the Data Quality Objectives Process. Office of Research and Development. EPA/600/R-96/055. September.
- . 1997. Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments, Interim Final. EPA/540/R-97/006, OSWER 9285.7-25. June.
- . 2000. Bioaccumulation Testing and Interpretation for the Purpose of Sediment Quality Assessment, Status and Needs. Office of Water, Office of Solid Waste, Washington, DC, EPA-823-R-00-001, February.