SEDIMENT CLASSIFICATION METHODS COMPENDIUM

Prepared by

U.S. Environmental Protection Agency
Sediment Oversight Technical Committee

EPA Work Assignment Managers

Beverly Baker and Michael Kravitz
Office of Science and Technology
Washington, DC 20460

CHAPTER 14

National Status and Trends Program Approach

Edward R. Long
Coastal Monitoring and Bioeffects Assessment Division
National Oceanic and Atmospheric Administration
7600 Sand Pt. Way, NE, Seattle, WA 98115
(206) 526-6338

Donald D. MacDonald
MacDonald Environmental Sciences, Ltd.
2376 Yellow Point Road, R.R. #3, Ladysmith, BC, Canada VOR 2EO

Sediment quality criteria based on multiple methods have been recommended for broad applications in the United States (USEPA/SAB, 1989; Adams *et al.*, in press). The approach used by the National Status and Trends Program (NSTP) of the National Oceanic and Atmospheric Administration (NOAA) to develop informal, effects-based guidelines involves the identification of the ranges in chemical concentrations associated with biological effects based on a weight of evidence from many studies. In this approach, the data for many chemicals are assembled from modeling, laboratory, and field studies to determine the ranges in chemical concentrations that are rarely, sometimes, and usually associated with toxicity. The data from many of the studies of the individual approaches described elsewhere in this document are compiled and examined to develop no-effects, possible-effects, and probable-effects ranges (Figure 14-1).

14.1 SPECIFIC APPLICATIONS

14.1.1 Current Use

The NSTP Approach was used initially to develop informal guidelines for use by the National Status and Trends (NS&T) Program (Long and Morgan, 1990; Long, 1992). NOAA analyzes sediments from numerous locations nationwide as a part of its monitoring program. The guidelines were developed as tools for identifying locations in which there is a potential for toxicity to living resources for which NOAA is the federal steward. Areas in which chemical concentrations often exceeded the guidelines were identified as high priorities for investigations of toxicity with biological tests.

Environment Canada evaluated many candidate approaches to the development of sediment quality guidelines and elected to develop its national guidelines using the NSTP Approach (MacDonald and Smith, 1991; MacDonald *et al.*, 1991). The Florida Department of Environmental Regulation elected to use the NSTP Approach to develop state sediment quality guidelines as a part of its sediment management strategy (MacDonald, 1992). The California Water Resources Control Board will use the NOAA guidelines in its initial evaluations of ambient chemical data. Following that step, data from field studies, laboratory bioassays, and equilibrium partitioning models will be used to develop sediment quality objectives (Lorenzato *et al.*, 1991). Finally, the International Council for Exploration of the Sea Study Group on the Biological Significance of Contaminants in Marine Sediments has elected to adopt the NSTP Approach in the development of guidelines for participating nations (Dr. Herb Windom, Working Group on Marine Sediments, ICES, personal communication).

Guidelines developed with the NSTP Approach were used by NOAA to identify chemicals that occurred in concentrations that were sufficiently high to warrant concern and to identify sampling sites and areas in which there was a potential for toxicity (Long and Morgan, 1990;

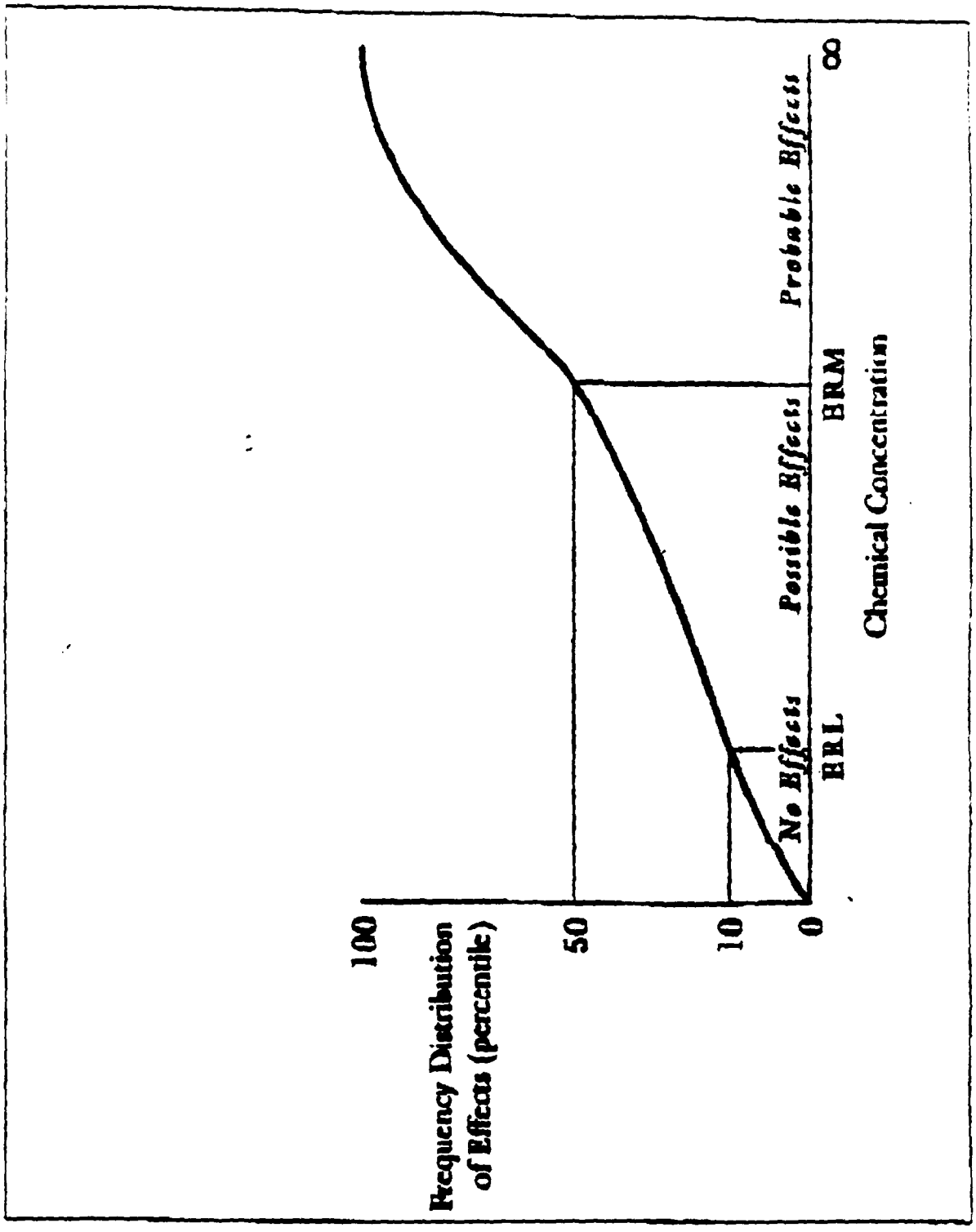

Figure 14-1. Conceptual outline of the relationship between the NS[illegible] proach guidelines and the eco-effects, possible effects, and probable-effects ranges in chemical concentrations.

Long *et al.*, 1991; Long and Markel, 1992). It was presumed that the potential for toxicity was relatively high in areas where numerous chemicals exceeded the upper bounds of the guidelines. Likewise, it was assumedthat the potential for toxicity was relatively low in areas where none of the chemical concentrations exceeded the lower bounds of the guidelines. In those regions with the highest potential for toxicity, NOAA has implemented regional surveys of toxicity, using a battery of biological analyses and tests.

Also, NOAA has used the guidelines in assessments and prioritization of hazardous waste sites (Dr. Alyce Fritz, NOAA Hazardous Materials Response and Assessment Division, personal communication). Other agencies and consultants have used the guidelines as a means of placing ambient chemical data into perspective with respect to the potential for toxicity (for example, Squibb *et al.*, 1991 for New York/New Jersey Harbor; Mannheim and Hathaway, 1991 for Boston Harbor; Soule *et al.*, 1991 for Marina Del Rey). The Florida Department of Environmental Regulation has used the guidelines as informal tools for interpreting ambient chemical data and for identifying regional priorities for sediment quality management (MacDonald, 1992).

14.1.2 Potential Use

Potential uses of the guidelines are as follows:

- Identification of potentially toxic chemicals in ambient sediments;
- Ranking and prioritization of areas and sampling sites for further investigation;
- Assessment of potential ecological hazards of contaminated sediments;
- Design of spiked-sediment bioassay experiments;
- Description of the kinds of toxic effects previously associated with specific concentrations of chemicals;
- Quantification of the relative likelihood of toxicity over specific ranges in chemical concentrations; and
- Identification of the need for sediment management initiatives.

14.2 DESCRIPTION

14.2.1 Description of Method

The NSTP Approach involves a simple evaluation of available data to identify three ranges in concentrations for each chemical:

- No-Effects Range: The range in concentrations over which toxic effects are rarely or never observed;
- Possible-Effects Range: The range in concentrations over which toxic effects are occasionally observed; and
- Probable-Effects Range: The range in concentrations over which toxic effects are frequently or always observed.

These ranges are identified by evaluating information from numerous studies in which matching biological and chemical data were developed. The specific steps in the method are:

(1) Compile matching chemical and biological data from laboratory spiked-sediment bioassays, equilibrium-partitioning models, and field studies and determine the chemical concentrations associated with no observed effects and those associated with adverse effects.

(2) Enter the data into a database, including the type of biological test performed, the adverse effect(s) measured, the chemical concentrations associated with observations of either effects or no effects, the type of study method and approach, and the degree of concordance between the

measure of effects and the concentration of the chemical.

(3) For those analytes for which sufficient data exist, prepare data tables sorted according to ascending chemical concentrations.

(4) Arithmetically determine the no-effects range, possible-effects range, and probable-effects range for each chemical.

The steps taken to select and screen candidate data sets are described in Section 14.2.1.2.3. The approach is intended to encourage periodic updates as new data become available.

Two slightly different methods have been used to determine the three chemical ranges. First, two percentiles in the chemical concentrations associated with toxicity were derived by Long and Morgan (1990): the lower 10th percentile and the 50th percentile (median). The lower 10th percentile was identified as the Effects Range-Low (ERL), and the median was identified as the Effects Range-Median (ERM). In their evaluation of the ascending data tables, Long and Morgan (1990) used only the chemical concentrations that had been associated with toxicity (i.e., the "effects" data). The conceptual basis for this approach and the three ranges are illustrated in Figure 14-2.

Later, MacDonald (1992) identified the three ranges with a method that used both the concentrations associated with biological effects (the "effects" data) and those associated with no observed effects (the "no-effects" data). In this method, a threshold affects level (TEL) was calculated first as the square root of the product of the lower 15th-percentile concentration associated with observations of biological effects (the ERL) and the 50th-percentile concentration of the no-observed-effects data (the NER-M). A safety factor of 0.5 was applied to the TEL to define a No-Observable-Effects Level (NOEL). Next, a Probable-Effects Level (PEL) was calculated as the square root of the product of the 50th-percentile concentration of the effects data (the ERM) and the 85th-percentile concentration of the no effects data (the NER-M).

Neither of these methods is preferred or advocated over the other. The significant feature of this approach is the use of a weight of evidence developed in the ascending tables, not in the specific method of using the data tables. In addition to the two methods described here, many others could be applied to the ascending data tables to derive guidelines. The method used by MacDonald (1992) considered both the "effects" and "no-effects" data, whereas that of Long and Morgan (1990) used only the "effects" data. Different percentiles in the ascending data were used in the two methods. Despite these differences in the methods, the agreement between the NOELs and ERLs and between the PELs and the ERMs was very good, usually within a factor of 2.

In both documents, the lower of the two guidelines for each chemical was assumed to represent the concentration below which toxic effects rarely occurred. The range in concentrations between the two values was that in which effects occasionally occurred. Toxic effects usually or frequently occurred at concentrations above the upper guideline value.

As an example, Figure 14-2 compares the frequency distribution of toxic effects and no-effects data associated with concentrations of naphthalene to the ERL and ERM concentrations for naphthalene. Long and Morgan (1990) reported the ERL as 340 ppb dry wt. and the ERM as 2100 ppb dry wt. for naphthalene, based on an ascending data table of 49 data points. These guidelines defined three ranges of chemical concentrations: the no-effects range (0-340 ppb); the possible effects range (340-2100 ppb); and the probable-effects range (>2100 ppb). Only 10.5 percent of the chemical concentrations below the ERL were associated with toxic effects, suggesting that toxicity is unlikely below the ERL concentrations. In contrast, 81 percent of the chemical concentrations between the ERL and ERM values were associated with the toxic effects and 93 percent of the data points were associated with toxicity at concentrations above the ERM value.

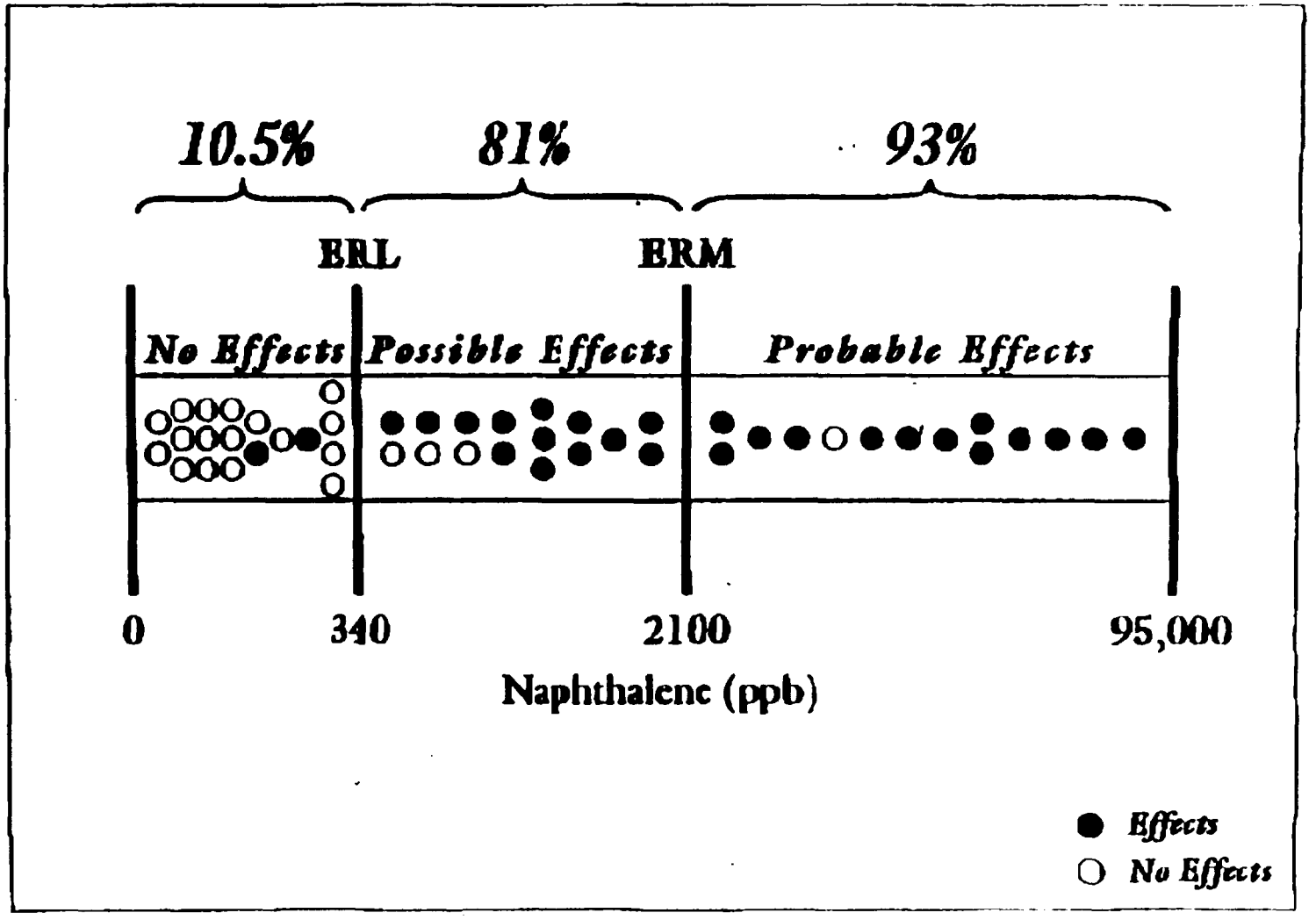

Figure 14-2. Frequency distribution of naphthalene concentration associated with toxic effects below the ERL value, between the ERL and ERM values, and above the ERM value (from Long and Morgan, 1990).

14.2.1.1 Objectives and Assumptions

The objective of the NSTP Approach is to provide informal, effects based guidelines that are based on a weight of evidence and reported as ranges in concentrations. The guidelines are based on chemical concentrations associated with measures of biological effects, thereby providing toxicological and/or biological relevance to the guidelines. They are based on data from multiple studies and research methods, thus providing a weight of evidence. In recognition of the variability in the kinds of data that are available, they are presented as ranges instead of absolute values, thereby providing a flexible interpretive tool with broad applicability. They are presented along with all of the supporting evidence in ascending tables, providing the user an interpretive framework for comparison with ambient data.

In this approach it is assumed that the data from all individual studies are equal in weight and credibility, although they may have involved very different methods and test endpoints. It is assumed that the methods used by the individual investigators were reasonably accurate. Most important, it is assumed that as the concentrations increase, the potential for toxicity also increases, thereby providing a conceptual basis for identifying the ranges in concentrations frequently associated with no toxic effects and those frequently associated with toxic effects. The guidelines can be formulated to account for site-specific factors that control bioavailability (see Section 14.3.1.1).

14.2.1.2 Level of Effort

14.2.1.2.1 Type of Sampling Required

The NSTP Approach relies on the use of a database compiled from a wide variety of sediment quality assessments. The database currently contains over 800 entries generated by the three major approaches to the establishment of effects-based guidelines: equilibrium-partitioning models; laboratory spiked-sediment bioassays; and various assessments of matching, field-collected, sediment chemistry, and biological effects data. The NSTP Approach was specifically designed to use existing data, therefore eliminating or minimizing the need for additional sampling. However, evaluation of the regional applicability of the guidelines could, in some cases, require further site-specific investigations, the magnitude of which could vary considerably.

14.2.1.2.2 Methods

The methods for deriving numerical sediment quality guidelines using the NSTP Approach are summarized in Section 14.2.1. Also, these methods are described by Long and Morgan (1990) and MacDonald (1992).

14.2.1.2.3 Types of Data Required

The NSTP Approach was intended to integrate a diverse assortment of information into a single database to support the derivation of numerical guidelines. Consequently, data from numerous modelling, laboratory, and field studies were collated into one database. Ideally, the data used to establish guidelines should include entries from all three of these types of approaches. Suitable data were available from a wide variety of sources. While collection and analysis of these data sets were labor-intensive, subsequent, incremental updates of the database should be relatively simple and inexpensive.

The data compiled from numerous studies were entered into the Biological Effects Database for Sediments (BEDS) by MacDonald (1992). All of the compiled data were fully evaluated prior to incorporation into the BEDS to ensure internal consistency in the database. The screening procedures used to support the development of the BEDS were designed to ensure that only relevant and high-quality data were used to derive the guidelines. No subjective biases were employed in screening the data; as many sources of data were included as possible. Candidate data from each study were evaluated to determine the acceptability of the experimental design, the test protocols, the analytical methods, and the statistical procedures that were used. Only data in which there were matched measures of sediment chemis-

try and biological effects were included. The database included only those data in which either statistically significant biological results were obtained or in which major differences in the biological results between samples were reported.

The BEDS currently includes over 800 data entries, mainly data from studies performed throughout North America. It was developed jointly by NOAA, Florida Department of Environmental Regulation, Environment Canada, and MacDonald Environmental Services Ltd.

In the evaluation of candidate data from field studies, only those data were used in which at least a 10-fold difference in the concentrations of at least one chemical among the samples was reported. Once this criterion was met, the data from many of the field studies were evaluated to determine the mean chemical concentrations in toxic samples (i.e., significantly different from controls) and those in nontoxic samples or in samples with relatively depauperate benthic communities (i.e., those with low abundance or species richness) versus those with more robust communities. Further, those mean concentrations in biologically effected samples that exceeded by twofold or more the mean concentrations in the background, reference, or nonaffected samples were assigned an asterisk in the ascending tables. The asterisks symbolized that a biological effect was noted and that there was a strong association between the chemical gradient and the biological gradient. Concentrations associated with nontoxic reference conditions were noted as "no effects." Those in which there was no concordance between the measures of effects and chemical concentrations were noted as "no gradient" or "no concordance." The concentrations derived in the modeling and spiked-sediment bioassays were always assigned asterisks. The concentrations with asterisks were used as "effects" data by both Long and Morgan (1990) and MacDonald (1992).

14.2.1.2.4 Necessary Hardware and Skills

The primary skills required to derive guidelines are associated with the development of the database. Expertise is required to evaluate the suitability of the biological and chemistry data, using the screening criteria. This process requires experience in the evaluation of sediment data and the methods that were used to develop the data.

The database has been developed on a personal computer and is readily transferable to other systems, but requires knowledge of the use of a computer. The database provides a means of storing and accessing all of the information that relates chemical concentrations to adverse biological effects. This information can be manipulated in this environment or exported into other formats.

14.2.1.3 Adequacy of Documentation

The NSTP Approach was documented by Long and Morgan (1990), in which the approach was peer-reviewed both within and outside NOAA. A second printing of the document was issued in 1992, following further review. A synopsis of the approach was described in a scientific journal (Long, 1992). The approach has been described orally in numerous technical and scientific forums. MacDonald and Smith (1991) and MacDonald *et al.* (1991) described the application of the approach in the development of guidelines for Canada. MacDonald (1992) described the use of the approach in a statewide sediment management strategy for Florida.

14.2.2 Applicability of Method to Human Health, Aquatic Life, or Wildlife Protection

The guidelines are intended to provide an estimate of the potential for adverse biological effects of sediment-associated contaminants on benthic organisms, based on a weight of evidence from analyses performed with multiple species and/or biological communities. They accommodate and rely on the data from tests of acute and chronic toxicity and from analyses of benthic community structure. The guidelines are based on data from many different areas and oceanographic regimes, thereby broadening their applicability. Currently, the data entered into the BEDS are from only marine and estuarine areas.

The guidelines provide a means of numerically estimating the percent frequency of biological effects over the three ranges of concentrations. The ascend-

ing tables accompanying the guidelines also provide a supplementary basis for interpreting new ambient chemical data. Also, these tables provide a visual and statistical means of estimating the relative degree of certainty in the guidelines.

The guidelines are not intended to be used for the protection of human life or wildlife. Rather, they are intended to be used in estimating the potential for adverse effects among benthic communities.

14.2.3 Ability of Method to Generate Numerical Criteria for Specific Chemicals

Long and Morgan (1990) reported numerical guidelines for 41 chemicals, including 12 trace metals, 18 polynuclear aromatic hydrocarbons (PAHs), and 11 synthetic organic compounds. MacDonald (1992) developed guidelines for 9 trace metals, total PCBs, 13 PAHs, 3 classes of PAHs, and 2 pesticides.

Conceptually, guidelines derived using this approach could be developed for any toxic chemical, provided sufficient data exist and provided the toxicity of the chemical is dose-responsive. Long and Morgan (1990) assigned a high degree of confidence to guidelines for chemicals for which data existed from many different approaches, different regions, and in which there was a good agreement in the data from different studies. MacDonald (1992) calculated guidelines only for those chemicals for which there was a minimum of 40 data points, after determining the minimum amount of data necessary to calculate reliable and consistent values. These minimum data requirements were established by iteratively calculating guidelines using data sets of increasing size (e.g., 4 to 60 data points) and determining when the estimate of the guidelines stabilized.

14.3 USEFULNESS

14.3.1 Environmental Applicability

14.3.1.1 Suitability for Different Sediment Types

The NSTP Approach can be applied equally to any sediment type that occurs in freshwater, estuarine, and marine environments. Since the database that supports the guidelines contains information from a wide variety of sediment types, the resultant guidelines are considered to be widely applicable. An increasing amount of information suggests that the bioavailability, and, therefore, toxicity, of many contaminants is controlled by such factors as TOC, AVS, and grain size. The BEDS currently accommodates the data for these variables, and, consequently, the guidelines could be normalized to the appropriate factors that control bioavailability. However, insufficient information currently exists to derive guidelines that are expressed in these terms. It is anticipated that future revisions of the guidelines will be expressed in these terms, thereby increasing their applicability.

Partly to increase the suitability of the guidelines to different sediment types, they are expressed as ranges in concentrations, not absolutes. These ranges provide a basis for evaluating chemical concentrations in the different types of sediments represented in the BEDS. In addition, the ascending data tables used to generate the guidelines can be examined to calculate frequency distributions of effects and no effects within each range of concentrations. These frequency distributions can be used as estimates of the probability of toxic effects.

14.3.1.2 Suitability for Different Chemicals or Classes of Chemicals

The approach can be applied to a wide variety of chemicals for which analytical methods are available. Thus far, numerical guidelines have been developed by Long and Morgan (1990) and by MacDonald (1992) for 43 and 28 chemicals or classes of chemicals, respectively. Data are included in the BEDS for over 200 chemicals or classes of chemicals. Guidelines could be developed for all of these substances when sufficient information becomes available.

14.3.1.3 Suitability for Predicting Effects on Different Organisms

Since the database compiled from many different studies is based on tests or analyses performed with many different species, the guidelines are widely applicable to benthic organisms.

In addition, the species studied in each investigation is(are) listed in the database; therefore, species-specific applicability can be evaluated by the users. Furthermore, the ERL values often are based on data from relatively sensitive species or life stages, and, therefore, can be used as guidelines suitable for the protection of sensitive species.

14.3.1.4 Suitability for In-Place Pollutant Control

Numerical sediment guidelines developed using the NSTP Approach can be used in a variety of ways as a tool in pollutant control. Specifically, these assessment tools respond to regulatory requirements by:

- Providing a basis for evaluating existing sediment chemistry data and ranking areas of concern and chemicals of concern in terms of their potential for causing toxicity and
- Identifying the need for further investigations, such as biological testing, to support regulatory decisions.

As is the case with all of the other approaches that rely on data collected in the field, the guidelines derived using the NSTP Approach integrate information obtained from studies of complex mixtures of contaminants and thereby consider their interactive effects. Consideration of the effects of contaminant mixtures is an advantage in the assessment of in-place pollutants in real-world conditions. However, this approach also relies on and gives equal weight to the data from equilibrium-partitioning models and laboratory spiked-sediment bioassays performed with single chemicals (see Section 14.2.1.1).

14.3.1.5 Suitability for Source Control

A reasonable amount of confidence in sediment quality guidelines is needed to justify using them in source control actions. Since the guidelines are developed with a weight of evidence compiled from many different studies, they provide a credible and defensible basis for evaluating contaminants in real-world conditions. The guidelines provide an efficient basis for identifying priority chemicals and priority areas that would benefit from source controls. In addition, the ascending tables provide a basis for estimating the probability of observing adverse effects at sites of interest, reducing the probability of effects through source controls, and evaluating the improvements in sediment quality following the implementation of source control measures.

14.3.1.6 Suitability for Dredged Material Disposal Applications

Neither the numerical guidelines nor the frameworks that have been developed for their application are intended to replace accepted testing protocols for dredged material disposal evaluations. Nonetheless, these guidelines can provide relevant tools for estimating the potential for adverse biological effects of contaminants associated with solid-phase sediments.

14.3.2 General Advantages and Disadvantages

14.3.2.1 Ease of Use

The approach has the advantage of relying on existing data. Therefore, guidelines can be developed relatively quickly and easily.

The original efforts by Long and Morgan (1990) and MacDonald (1992) to assemble the databases used to develop the guidelines were labor-intensive. Numerous reports and data sets were located, and a huge amount of data was entered into spreadsheets. However, these data now exist in a centralized, computerized database, the BEDS. Subsequent derivations of guidelines based on iterative expansions of the BEDS database should be relatively quick, easy, and inexpensive.

The guidelines are easily used and interpreted. Chemical data can be readily compared with the guidelines and with the ascending tables. The frequency of occurrence of toxicity over the no-effects, possible-effects, and probable-effects ranges can be calculated and compared with the chemical data.

Sediments in which numerous chemicals occur at concentrations that fall within the probable-effects ranges have a higher probability of being toxic than those in which most of the chemical concentrations are within the no-effects range. This type of simple interpretation makes the guidelines very easy to use.

14.3.2.2 Relative Cost

The original effort of Long and Morgan (1990) involved roughly one year of labor. The confirmation and expansion of the database by MacDonald (1992) involved more than another year of labor. The costs of subsequent iterations of the guidelines based on further expansions of the database would vary with the amount of data entered and the number of chemicals. The calculations of the guideline values themselves are very simple and quick. Also, the guidelines can be used very quickly and easily.

If the necessary data are not available for entry into a database, then the costs to generate them could be relatively high. If initiated *de novo*, modeling, bioassay, and field studies necessary to generate sufficient data could vary considerably in costs and time, depending on the amount of data needed.

14.2.3 Tendency to Be Conservative

The predictive capabilities of the guidelines have not been independently quantified. The protectiveness of the guidelines could be increased by considering data only from chronic sublethal endpoints or by applying a numerical safety factor, such as was applied in the Florida guidelines (MacDonald, 1992). Also, the guidelines would become more conservative if data were included only from areas in which sediments were highly bioavailable.

14.3.2.4 Level of Acceptance

The NSTP Approach has been published by NOAA, following an in-house and outside peer review. It has been published in a peer-reviewed scientific journal. The approach has been used by Environment Canada and Florida Department of Environmental Regulation in the development of their respective guidelines. It has been adopted by a committee of the International Council for Exploration of the Sea for use by member nations. The State of California has adopted a similar approach to the development of sediment quality objectives (Lorenzato *et al.*, 1991).

The numerical guidelines developed by use of the approach have been used by NOAA to compare and rank the potential for toxicity at monitoring sites nationwide, within San Francisco Bay, and within Tampa Bay. Approximately 1500 copies of the report by Long and Morgan (1990) have been distributed. Users of the report have compared ambient concentrations with the guidelines in assessments of hazardous waste sites, analyses of prospective dredge material, evaluations of survey and monitoring data, and estimates of ecological risk (for example, Manaheim and Hathaway, 1991; Soule *et al.*, 1991; Squibb *et al.*, 1991). NOAA routinely uses the guidelines in its estimates of ecological risk at National Priority List hazardous waste sites. The guidelines have been used as a basis for interpretation of chemical data in a

14.3.2.5 Ability to Be Implemented by Laboratories with Typical Equipment and Handling Facilities

The spreadsheets and database needed to generate the guidelines can be prepared with a personal computer and need not be very complicated. Entry of data into the database and the generation of the ascending tables are very simple. The calculations of the guidelines can be performed manually, on a desk-top calculator or a personal computer. The database can be supplemented with new data as they become available. Implementation of the approach can become more laborious and complicated if the necessary data must be generated *de novo*.

14.3.2.6 Level of Effort Required to Generate Results

As outlined in Section 14.3.2.2, the level of effort required in the development of the original set of guidelines was relatively high. Subsequent iterations of the guidelines for other purpos

other chemicals, or for the same chemicals following additions to the database would be relatively easy. Entry of new data points from spiked-sediment bioassays, equilibrium-partitioning models, or apparent effects thresholds into the database would require only a few minutes. Manipulation-of-raw matching data from biological and chemical analyses performed in a field study would require from a few hours to several days, depending on the size of the data set, followed by entry of the data points into the database.

14.3.2.7 *Degree to Which Results Lend Themselves to Interpretation*

The guidelines and the ascending data tables on which they are based can be used in a number of ways. First, the data from analyses of ambient samples can be compared visually with the two numerical guidelines to determine whether the ambient concentrations exceed either of the guidelines. Second, the ambient concentrations can be compared with the data in the ascending tables to determine the kinds of toxic effects that have been observed in previous studies at the concentrations of concern. Finally, the frequencies of toxicity in the no-effects, possible-effects, and probable-effects ranges can be used to predict the probability of toxicity associated with any contaminant concentration.

The guidelines developed thus far with this approach do not account for the effects of factors that control bioavailability of the toxicants. This is not a weakness of the approach; rather, it is a weakness of the available data. Nevertheless, this weakness may hinder interpretation of ambient data with the guidelines. The BEDS database includes a provision for entering data from analyses of acid volatile sulfides and total organic carbon (and other potential normalizers) and, therefore, would lend itself to recalculation of guidelines normalized to these factors once the necessary data become available.

An important strength of this approach is that it provides the user some flexibility in the use and interpretation of the guidelines. All of the data are provided in ascending order for the user to see and evaluate. The degree of certainty in the data can be assessed and judged by the user. Ranges in concentrations are provided, instead of rigid, single absolute values.

One of the most attractive features of this approach is the estimation of the probability of biological effects, based on the frequency distributions of effects for each chemical. For example, the data in the BEDS database indicate that only 5.8 percent of the chemical concentrations within the no-effects range for cadmium (0 to 1 mg/kg) determined by MacDonald (1992) were associated with adverse biological effects (Figure 14-3). These data suggest that there is a low probability of observing adverse effects within this range. Within the probable effects range for cadmium (>7.5 mg/kg), roughly 68 percent of the database entries were associated with adverse effects. These data suggest that there is a relatively high probability of observing adverse effects within this range. Positive concordance between frequency of effects and chemical concentrations should inspire confidence in the guideline values.

Evaluation of the guidelines for mercury reveals that a lower level of confidence should be placed on the guidelines for this element. The data in the BEDS database indicate that within the no-effects range (0 to 0.1 mg/kg), roughly 7 percent of the entries were associated with adverse effects (Figure 14-4). However, frequency distributions of effects are similar within the possible-effects range (0.1 to 1.4 mg/kg) and the probable-effects range (>1.4 mg/kg), namely 30.1 percent and 33.3 percent, respectively. Therefore, it is more difficult to adequately determine the unacceptable levels of mercury in sediments than with, say, cadmium.

14.3.2.8 *Degree of Environmental Applicability*

The guidelines are highly applicable to the interpretation of environmental data. They are generated with data from environmentally realistic field studies, as well as theoretical modeling studies and controlled laboratory experiments. They are generated with data from many different regions in which the mixtures and concentrations

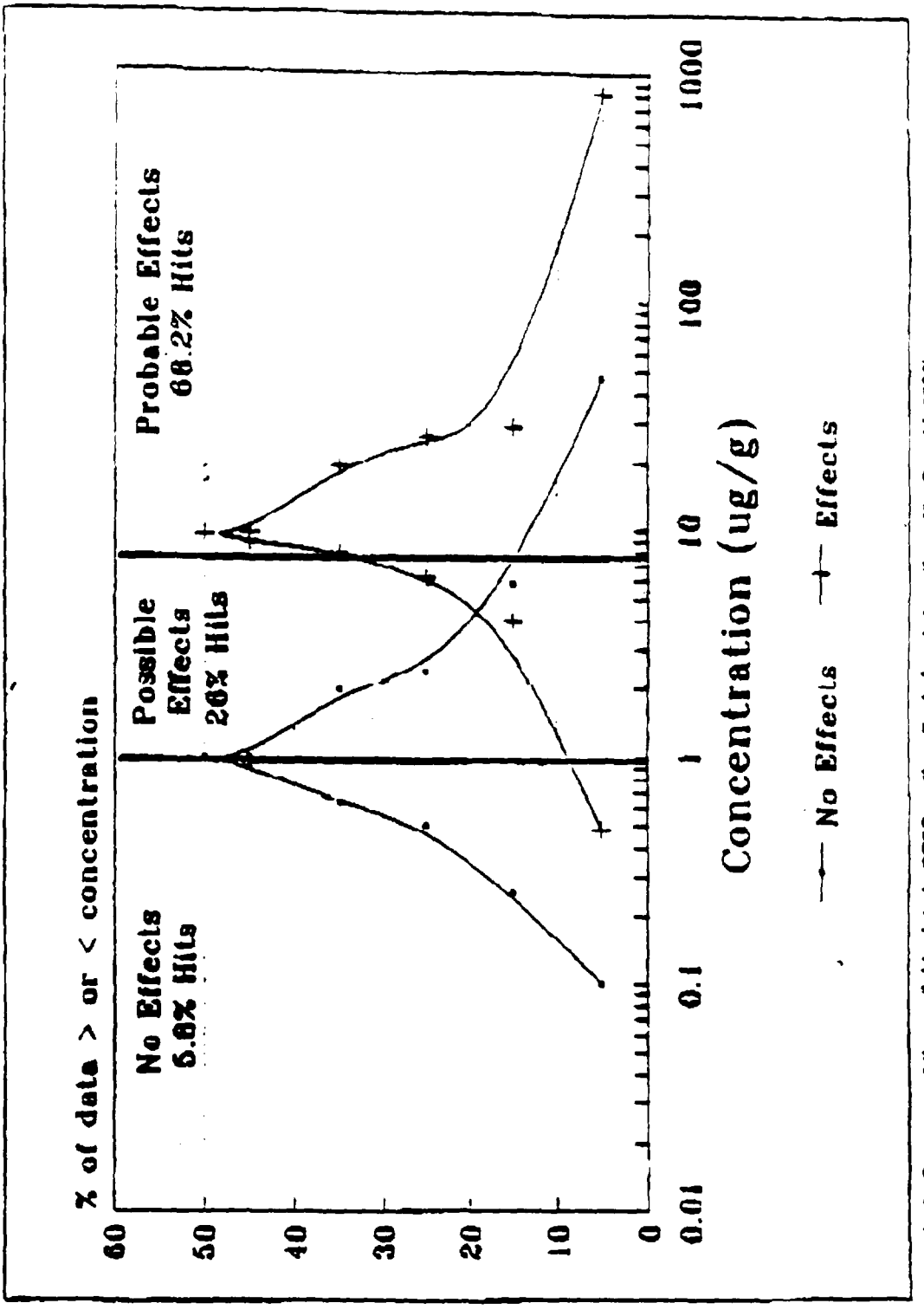

Figure 14-3. Summary of the available data in BEDS on the effect of cadmium (from MacDonald, 1992). Percent frequency of effects data and percent frequency of no effects data are plotted against cadmium concentrations.

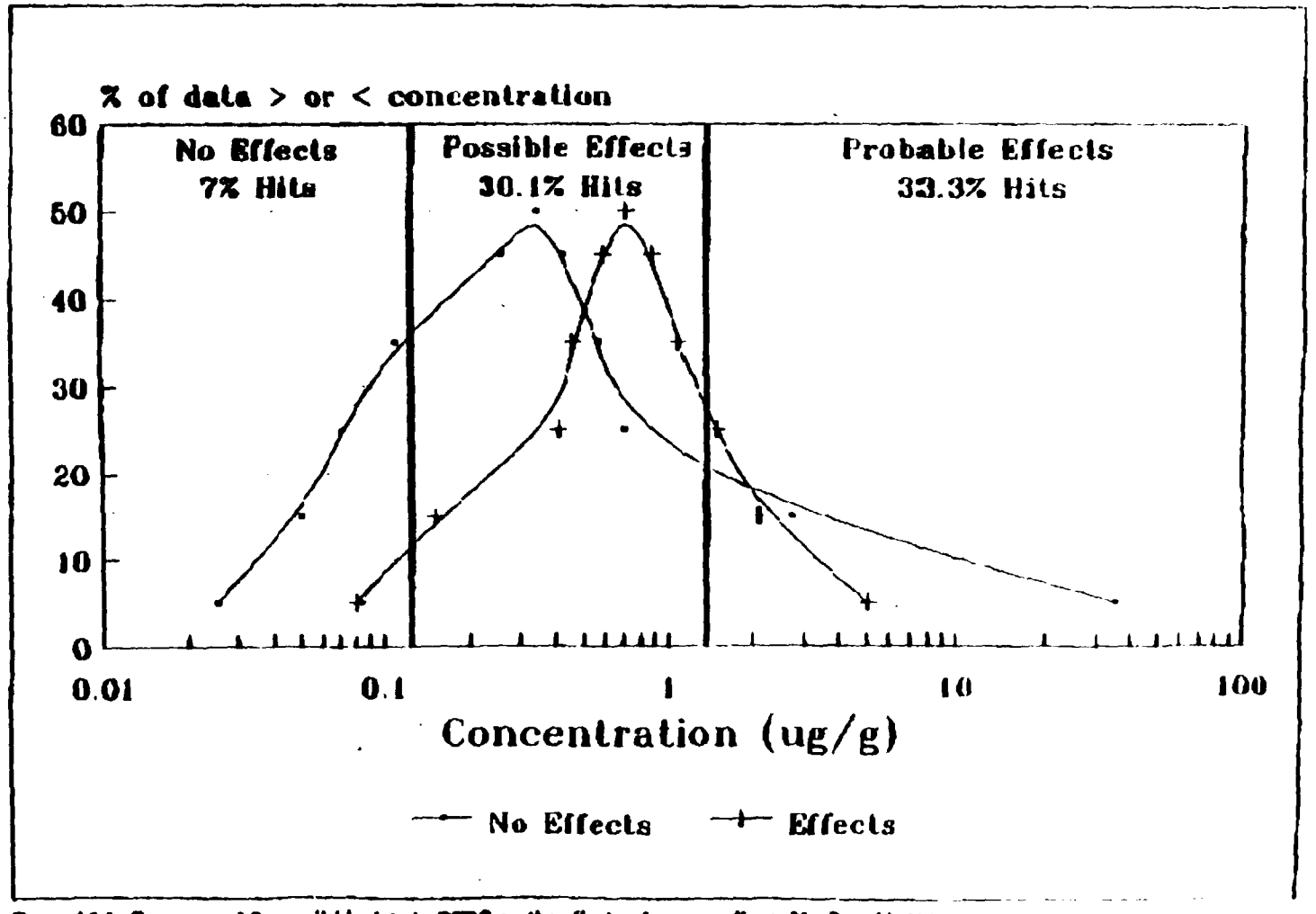

Figure 14-4. Summary of the available data in BEDS on the effects of mercury (from MacDonald, 1992).

of chemicals differ and in which sedimentological properties differ. They are generated with tests using different species with different sensitivities to toxicants. They are universally applicable in North America since they were generated with data from many regions in the United States and Canada. Confidence in the utility of the guidelines is inspired by the weight of evidence from these multiple studies.

14.3.2.9 Degree of Accuracy and Precision

By iteratively adding and removing different data sets from the ascending tables, MacDonald (1992) determined that a minimum of 40 data sets were needed to develop consistent and reliable guidelines. Clearly, some variability in the guidelines is to be expected as data are added or deleted, but, once the minimum amount of data is compiled, this variability appears to be minimal.

MacDonald (1992) generally doubled or tripled the amount of data in the ascending tables compiled by Long and Morgan (1990) mainly with new data from field studies and laboratory spiked-sediment bioassays. Also, MacDonald (1992) considered only estuarine and marine data, thereby deleting the freshwater data included in Long and Morgan (1990). The effects on the guideline concentrations of eliminating some data and adding a substantial amount of new data are illustrated in Tables 14-1 and 14-2. The ERL and ERM values, based on the Long and Morgan (1990) data tables and the larger MacDonald (1992) tables, are compared by using the methods of Long and Morgan (1990) applied to both data sets.

For 13 aromatic hydrocarbons, the average of the ratios between the two sets of guidelines was 1.5 (1.9 for the ERLs and 1.2 for the ERMs). For eight trace metals, the average of the ratios between the two sets of guidelines was 1.7. The trace metals ERL values changed more than the ERM values (average ratios of 1.9 for the ERLs and 1.5 for the ERMs).

Overall, 7 of the 23 ERL values did not change and the ratios between the two sets of ERL values ranged from 1.0 to 9.4. Also, 7 of the 23 ERM values did not change. Of the 46 values, 14 remained unchanged, 17 increased, and 15 decreased. The overall mean factors of change were less than twofold for both trace metals and PAHs. These observations suggest that the guidelines are not terribly sensitive to the addition of new data once a minimum amount has been compiled. Also, they suggest that the guidelines originally developed by Long and Morgan (1990) generally are substantiated by additional data compiled by MacDonald (1992).

The accuracy of the guidelines in predicting toxicity has not yet been quantified. However, in the Hudson-Raritan estuary, the concentrations of many chemicals quantified in previous studies (Squibb *et al.*, 1992) frequently exceeded the ERM guidelines in the Arthur Kill and rarely exceeded them in the lower Hudson River. In a recent survey funded by NOAA, sediments from the Arthur Kill were extremely toxic to amphipods and other species, whereas the sediments from the lower Hudson River were not toxic.

14.4 STATUS

14.4.1 Extent of Use

The NSTP Approach is being used by NOAA's National Status and Trends Program, by Environment Canada, and by the Florida Department of Environmental Regulation. A variation on the approach is being pursued by the California Water Resources Control Board. Other states and regional districts have inquired about the possible use of the approach.

14.4.2 Extent to Which Approach Has Been Field-Validated

Validations of the guidelines have not yet been quantified. As described in Section 14.3.2.9, the original set of guidelines generally were substantiated by the addition of considerable amounts of new data, largely from field studies performed in many regions. The concordance between predictions of toxicity with the guidelines and actual observations of toxicity has been very

Table 14-1. Ratios Between the Guideline Values for Polynuclear Aromatic Hydrocarbons Determined with Data from Long and Morgan (1990) and Those Determined with Data from MacDonald (1992). *Total number of data points available are listed (with those used to determine guidelines in parentheses).*

Chemical Analyte	MacDonald (1992)	Long and Morgan (1990)	Ratio Between Two Sets of Values	Values Increased (+) Decreased (-)
Polynuclear aromatic hydrocarbons (ppb d.w.)				
Acenaphthene	n=69(30)	n=35(15)	2.0(2.0)	
ERL	16	150	9.4	-
ERM	500	650	1.3	-
Anthracene	n=86(46)	n=39(26)	2.3(1.8)	
ERL	85.3	85	1.0	-
ERM	1100	960	1.1	+
Fluorene	n=86(46)	n=44(28)	2.2(1.7)	
ERL	19	35	1.8	-
ERM	540	640	1.2	-
2-methylnaphalene	n=48(28)	n=31(15)	1.6(1.9)	
ERL	70	65	1.1	+
ERM	670	670	1.0	-
naphthalene	n=87(44)	n=50(28)	1.9(1.6)	
ERL	160	340	2.1	-
ERM	2100	2100	1.0	-
phenanthrene	n=101(51)	n=49(34)	2.1(1.5)	
ERL	240	220	1.1	-
ERM	1500	1380	1.1	+
benzo(a)anthracene	n=81(43)	n=34(30)	1.9(1.4)	
ERL	261	230	1.1	+
ERM	1600	1600	1.0	-
benzo(a)pyrene	n=89(44)	n=43(27)	2.1(1.6)	
ERL	430	400	1.1	+
ERM	1600	2500	1.6	-
chrysene	n=89(46)	n=41(27)	2.2(1.7)	
ERL	384	400	1.0	-
ERM	2800	2800	1.0	-
dibenzo(a,h)anthracene	n=78(31)	n=39(18)	2.0(1.7)	
ERL	63.4	60	1.1	+
ERM	260	260	1.0	-
fluoranthene	n=117(71)	n=61(38)	2.3(1.8)	
ERL	600	600	1.0	-
ERM	5100	3600	1.4	+
pyrene	n=88(50)	n=45(28)	2.2(1.8)	
ERL	665	350	1.9	+
ERM	2600	2200	1.2	+
total PAH	n=78(34)	n=65(34)	1.2(1.0)	
ERL	4022	4000	1.0	-
ERM	44,792	35,000	1.3	+
Mean change in PAH ERLs		1.90		
Mean change in PAH ERMs		1.17		
Overall mean change in PAH values		1.53		

Table 14-2. Ratios Between the Guideline Values for Total PCBs and Trace Metals Determined with Data from Long and Morgan (1990) and Those Determined with Data from MacDonald (1992). *Total number of data points available are listed (with those used to determine guidelines in parentheses).*

Chemical Analyte	MacDonald (1992)	Long and Morgan (1990)	Ratio Between Two Sets of Values	Values Increased (+) Decreased (-)
Polychlorinated Biphenyl (ppb d.w.)				
total PCB	n=128(50)	n=77(33)	1.6(1.5)	
ERL	22.7	50	2.2	-
ERM	180	400	2.2	-
Trace Metals (ppm d.w.)				
arsenic	n=143(27)	n=48(16)	3.0(1.7)	
ERL	8.2	33.0	4.0	-
ERM	70.0	85.0	1.2	-
cadmium	n=261(84)	n=108(36)	2.5(2.3)	
ERL	1.2	5.0	4.2	-
ERM	9.6	9.6	1.0	-
copper	n=221(76)	n=81(51)	2.4(1.5)	
ERL	34.0	70.0	2.0	-
ERM	270	390	1.4	-
chromium	n=197(37)	n=76(21)	2.6(1.8)	
ERL	81	80	1.0	-
ERM	370	145	2.6	+
lead	n=210(73)	n=83(47)	2.5(2.6)	
ERL	46.7	35.0	1.3	+
ERM	220	110	2.0	+
mercury	n=169(42)	n=76(30)	2.2(1.4)	
ERL	0.15	0.15	1.0	+
ERM	0.71	1.3	1.8	-
nickel	n=169(19)	n=66(18)	3.0(1.1)	
ERL	20.9	30	1.4	-
ERM	51.6	50	1.0	-
silver	n=88(25)	n=47(13)	2.0(1.9)	
ERL	1.0	1.0	1.0	+
ERM	3.7	2.2	1.7	+
zinc	n=214(74)	n=78(46)	2.7(1.6)	
ERL	150	120	1.25	+
ERM	410	270	1.5	+
Mean change in PAH ERLs			1.9	
Mean change in PAH ERMs			1.9	
Overall mean change in metals values		1.74		

good thus far, but the degree of concordance has not been quantified. Additional opportunities to field-validate the guidelines will be available in future studies in Tampa Bay, the Hudson-Raritan estuary, and southern California.

14.4.3 Reasons for Limited Use

The NSTP Approach initially was used by NOAA to develop informal guidelines for internal agency use. Therefore, knowledge of and access to the guidelines was limited. As interest in the guidelines increased, they were released in a government document with a limited distribution. Therefore, the main reason for the limited use of the approach has been the limited awareness of its existence. Furthermore, the equilibrium-partitioning approach to national criteria and the most successful regional approach to criteria (apparent effects thresholds in Washington) have received considerable attention. Moreover, the guidelines thus far have not considered the potential for bioavailability or bioaccumulation because of a lack of data.

14.4.4 Outlook for Future Use and Amount of Development Yet Needed

There is significant potential for the expanded use of the NSTP Approach. Canada, Florida, and California currently are using the approach to develop their respective guidelines. Since the approach relies on existing data, other region-specific guidelines could be developed easily, using the data available from specific regions. The approach can be used to validate criteria developed with other single-method approaches. The database can be accessed for specific regions or for fresh, estuarine, or marine waters.

Several types of data are needed to further develop the approach. First, additional data are needed from studies in which TOC, grain size, and acid volatile sulfides were measured. Second, additional data are needed from spiked-sediment bioassays to establish cause-effect relationships. Third, additional data are needed from field studies in which very strong chemical gradients were observed. These studies should include measures of the toxicity and chemical contamination of bulk sediments and pore water. The would benefit from toxicity identification evaluations to identify the causative agents responsible for the observed biological effects (Ankley 1989). A number of large field surveys are under way and being planned by NOAA and will lead to additional data to be included in the database. Once these additional data are available, they could be entered into the database and used to develop updated or new guidelines.

14.5 REFERENCES

Adams, W.J., R. A. Kimerle, and J. W. Barnett Jr. In press. Sediment quality and aquatic life assessment. Envir. Sci. and Technol.

Ankley, G. 1989. Sediment toxicity assessment through evaluation of the toxicity of interstitial water. Environmental Research Laboratory-Duluth. U.S. Environmental Protection Agency, Duluth, MN. 27 pp.

Long, E. R. 1992. Ranges in chemical concentrations in sediments associated with adverse biological effects. Mar. Pollu. Bull. 24 (1): 38-45.

Long, E.R., D. MacDonald, and C. Cairncross. 1992. Status and trends in toxicants and the potential for their biological effects in Tampa Bay, Florida. NOAA Tech. Memo. NOS OMA 58. National Oceanic and Atmospheric Administration, Seattle, WA. 77 pp.

Long, E.R., and R. Markel. 1992. An evaluation of the extent and magnitude of biological effects associated with chemical contaminants in San Francisco Bay, California. NOAA Tech. Memo. NOS OMA 64. National Oceanic and Atmospheric Administration, Seattle, WA. 86 pp.

Long, E.R., and L.G. Morgan. 1990. The potential for biological effects of sediment-sorbed contaminants tested in the National Status and Trends Program. NOAA Tech. Memo. NOS OMA 62. National Oceanic and Atmospheric Administration, Seattle, WA. 175 pp.

Lorenzato, S.G., A. J. Gunther, and J. M. O'Connor. 1991. Summary of a workshop

concerning sediment quality assessment and development of sediment quality objectives. California State Water Resources Control Board, Sacramento, CA. 32 pp.

MacDonald, D.D. 1992. Development of an integrated approach to the assessment of sediment quality in Florida. Prepared for Florida Department of Environmental Regulation. MacDonald Environmental Services, Ltd. Ladysmith, British Columbia. 114 pp.

MacDonald, D.D., and S.L. Smith. 1991. A discussion paper on the derivation and use of Canadian sediment quality guidelines for the protection of freshwater and ma: ie aquatic life. Prepared for Canadian Council of Ministers of the Environment. Environment Canada. Ottawa.

MacDonald, D.D., S.L. Smith, M.P. Wong, and P. Mudroch. 1991. The development of Canadian marine environmental quality guidelines. Report prepared for the Interdepartment Working Group on Marine Environmental Quality Guidelines and the Canadian Council of Ministers of the Environment. Environment Canada. Ottawa, Canada. 50 pp.

Mannheim, F.T., and J.C. Hathaway. 1991. Polluted sediments in Boston Harbor-Massachusetts Bay: Progress report on the Boston Harbor data management file. U.S. Dept. of the Interior, Geological Survey Open File Report 91-331. USGS, Woods Hole, MA. 18 pp.

Soule, D.F., M. Oguri, and B.H. Jones. 1991. Marine Studies of San Pedro Bay, California, Part 20F. The marine environment of Marina Del Rey. October 1989 to September 1990. Submitted to Department of Beaches and Harbors, County of Los Angeles. University of Southern California, Los Angeles, CA. 206 pp.

Squibb, K. S., J. M. O'Connor, and T.J. Kneip. 1991. New York/New Jersey Harbor Estuary Program. Module 3.1: Toxics characterization report. Prepared for U.S. Environmental Protection Agency, Region 2. NYU Medical Center, Tuxedo, NY. 65 pp.

USEPA/SAB. 1989. Evaluation of the apparent effects threshold (AET) approach to assessing sediment quality. U.S. Environmental Protection Agency Science Advisory Board. Report of the Sediment Criteria Subcommittee. U.S. EPA SAB-EETFC-89-027. 16 pp.