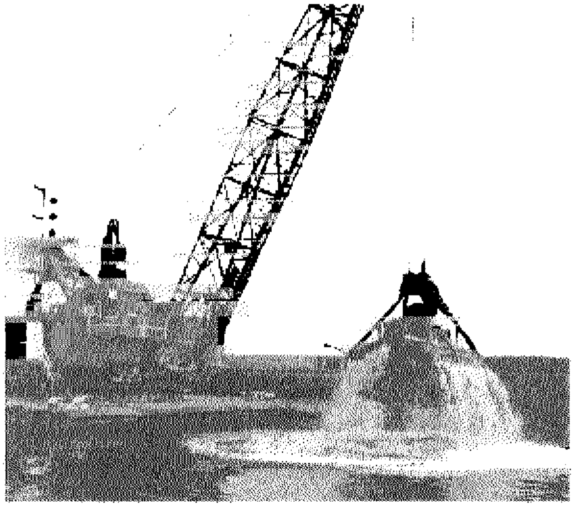
NJMSC-W-00-002 C2

Proceedings: Sediment Toxicity Risk Assessment: Where Are We and Where Should We Be Going?

Michael P. Weinstein and W. Scott Douglas, Editors



Conference on Dredged Material Management: Options and Environmental Considerations

Massachusetts Institute of Technology
Cambridge, Massachusetts

3-6 December 2000

NEW JERSEY MARINE SCIENCES CONSORTIUM

Office of Maritime Resources
NJ Department of Transportation

Sea Grant
National Sea Grant College Program

Preface

By the middle of this century, global trade is expected to triple, with 90% of the weight and 80% of the value of all international goods transported by water (National Ocean Conference 1998[1]). To ship these goods, larger vessels will be required, in turn, requiring expanded ports and deeper navigational channels, some of the latter exceeding 45 ft (15 m) in depth. Nationwide, construction and maintenance of these channels requires the dredging of more than $400*10^6$ yds^3 ($305*10^6$ m^3) annually, with the volume projected to increase in many areas. The Port of New York and New Jersey is no exception. The Port boasts over 250 miles (400 km) of engineered waterways, requiring $2\text{-}4*10^6$ yds^3 ($1.5 - 3.0*10^6$ m^3) of annual maintenance dredging. Planned channel deepening to accommodate traffic projections will require the additional dredging of over $50*10^6$ yds^3 ($38*10^6$ m^3) of sediment over the next 10-15 years. These water highways are essential for sustained economic growth; e.g., the Port of NY and NJ now adds over $30 billion annually to the region's economy and creates hundreds of thousands of direct and indirect jobs.

Unfortunately, sediments that settle into shipping channels also become sinks for pollutants. Contaminant discharges result in the accumulation of heavy metals and persistent organic compounds in the fine sediments of harbors and waterways. Petroleum hydrocarbons and their derivatives, polychlorinated biphenyls (PCBs), dioxins and furans, pesticides, mercury, lead, and chromium, among others are often found at elevated concentrations in the harbor bottom. Since 1972, the US Environmental Protection Agency and the US Army Corps of Engineers have required that dredged material be tested for potential toxicity prior to disposal at open ocean sites. Recent improvements in the assessment of dredged materials proposed for ocean disposal and increased public awareness and sensitivity to the issue of contamination has resulted in a dramatic decrease in open ocean disposal and placement of dredged materials in either confined disposal facilities, or after decontamination, incorporated as feedstock into varied "beneficial uses". However all of these processes remain significantly more expensive than conventional disposal and threaten the continued economic viability of many ports.

What remains unknown, however, are the true ecological risks and other costs/benefits associated with decisions to dispose of dredged materials, whether in the ocean or upland. As the science of ecological risk assessment improves, decision-makers will ultimately have better tools to address the management of dredged materials. It was the purpose of this Workshop -- *Sediment Toxicity Risk Assessment: Where Are We and Where Are We Going?* -- to review the status of ecological risk assessment and discuss what scientists mean when they say, "these muds are toxic", and the corollary, "what is worrisome about dredged materials?". An expert panel was convened on the last day of the Conference, and a series of "challenge" questions posed that were intended to focus the discussion and meet the goals of workshop. Panel members included: *Bruce Brownawell*, SUNY, Stony Brook; *Dominic M. Ditoro*, Manhattan College; *Kay T. Ho* and *Wayne R. Munns, Jr.*, U.S. Environmental Protection Agency; *Peter M. Chapman*, EVS Environment Consultants; and *Keith Solomon*, University Of Guelph. Ms. Elizabeth "Bitsy" Waters moderated the session.

[1] National Ocean Conference - Oceans of Commerce, Oceans of Life; June 11-12, 1998, Naval Postgraduate School, Monterey, California.

To address the issue of sediment toxicity, the New Jersey Marine Sciences Consortium (NJMSC) through its *New Jersey Sea Grant College Program* and the New Jersey Maritime Resources co-hosted this session at the *Conference on Dredged Material Management*, Massachusetts Institute of Technology, Cambridge, MA, on 3-6 Dec 2000. This volume is comprised of five invited papers that will appear separately in *Marine Pollution Bulletin* and an edited transcript of the facilitated Workshop. One of these peer-reviewed papers -- *Issues in Sediment Toxicity and Ecological Risk Assessment* -- is a synthesis article prepared by the workshop panel from the morning discussion. The edited transcript is also reproduced herein. A series of "challenge" questions guided the discussion, and care was taken to constrain the topic to ecological risk assessment and refrain from introducing human health concerns into the dialogue:

<u>Research/Technical</u>

1. How Do *Scientists* Define Sediment Toxicity?
2. How Do We Establish Baselines for Toxicity (Reference/Background)?
3. How Do We Select Appropriate End-Points (e.g., SQC or Bioassays; Tiered Approach or Integrated-SQC)?
4. How Do We Evaluate Ecological Significance of Endpoints or Bioassays?
5. How Can the Magnitude of Uncertainty Be Quantified, Reduced, and/or Managed?

<u>Science-Based Management/Policy</u>

6. What Type of Information Does a Manager Need from the Scientific Community
7. What Type of Information Does the Scientific Community Need From Managers?

<u>Regulatory Decisions</u>

8. Can Sediment Toxicity Measurements be Applied Nation-Wide?

Michael P. Weinstein
New Jersey Marine Sciences Consortium

W. Scott Douglas
New Jersey Maritime Resources

Acknowledgements

The Workshop Organizers gratefully acknowledge the *Sea Grant College Programs* of Connecticut, Delaware, Mississippi-Alabama ("Habitat Theme Team"), MIT and Rhode Island for their efforts in underwriting and hosting the Conference. We also thank the Port Authority of NY&NJ for defraying the cost of videotaping the facilitated Workshop. Kim Kosko, New Jersey Sea Grant, Director of Communications was also instrumental in making the workshop a success.

This publication was supported by the National Sea Grant College Program of the U.S. department of Commerce's National Oceanic and Atmospheric Administration under NOAA Grant # NA76-RG0091. The views expressed herein do not necessarily reflect the views of any of those organizations. **NJSG-02-482**.

Table of Contents

Issues In Sediment Toxicity And Ecological Risk Assessment[2]

PETER M. CHAPMAN[1], KAY T. HO and WAYNE R. MUNNS, JR.[2], KEITH SOLOMON[3], MICHAEL P. WEINSTEIN[4]
[1]*EVS Environment Consultants, 195 Pemberton Avenue, North Vancouver, BC, Canada V7P 2R4, Phone: 604-904-4005, E-Mail: pchapman@attglobal.net*
[2]*U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, 27 Tarzwell Drive, Narragansett, RI USA 02882*
[3]*Centre for Toxicology, University of Guelph, Bovey Building, Gordon Street, Guelph, ON Canada N1G 2W1*
[4]*New Jersey Marine Sciences Consortium, Sandy Hook Field Station, Building # 22, Fort Hancock, NJ USA 07732*

Abstract

This paper is based on a facilitated Workshop and Roundtable Discussion of key issues in sediment toxicology and ecological risk assessment (ERA) as applied to sediments that was held at the Conference on Dredged Material Management: Options and Environmental Considerations. The issues addressed included how toxicity is defined and perceived, how it is measured, and how it should be used within the context of ERA to support management decisions. The following conclusions were reached regarding scientific considerations of these issues. Toxicity is a measure of hazard and not a risk *per se*. Thus, toxicity testing is a means but not the end to understand risks of sediments. Toxicity testing cannot presently be replaced by chemical analyses to define hazard. Toxicity test organisms need to be appropriate to the problem being addressed, and the results put into context relative to both reference and baseline comparisons to understand hazard. Use of toxicity tests in sediment ERAs requires appropriate endpoints and risk hypotheses, considering ecological not just statistical significance, and recognizing that hazard does not equate to risk. Toxicity should be linked to population and community response to support decision-making, assessing possible genotypic adaptations that can influence risk estimates, and addressing uncertainty. Additionally, several key scientific issues were identified to improve future sediment ERAs, including the need to improve basic understanding of ecological mechanisms and processes, recognition of variability in the assessment process, and an improved focus and ability to assess risks to populations and communities.

Keywords: sediments, toxicity, ecological risk assessment, dredging

Introduction

Contaminated sediments in water bodies may be affecting ecosystems, resources and human health; they are certainly having economic effects as the significance of such contamination and the need for sediment remediation are assessed (Ingersoll et al., 1997; Moore et al., 2001). There

---

[2] (Reprinted with permission from *Marine Pollution Bulletin)*

is clear need for continued scientific dialogue around all aspects of the assessment and management of contaminated sediments, and in particular, the use of toxicity tests to estimate risk (Chapman, 1995).

This Viewpoint paper is derived from a discussion held at a facilitated Workshop and Roundtable Discussion (*Sediment Toxicity Risk Assessment: Where are we, and Where Should we be Going?)* at the Conference on Dredged Material Management: Options and Environmental Considerations, held on December 6, 2000 at the Massachusetts Institute of Technology. Panelists at the Workshop were asked to consider the state-of-the-science with respect to aspects of sediment toxicity risk assessment and to identify how such assessments can be improved. Here we offer thoughts on some key issues in the fields of sediment toxicology and ecological risk assessment (ERA), and identify areas requiring additional future attention. Although discussed in the context of dredged material management, the issues and scientific opinions are germane to sediment evaluation in general. Key contributors to the Workshop who are not authors are listed in the Acknowledgments.

Sediment Toxicity

Of primary concern in most sediment evaluations, including those involving dredged materials, is the toxicity of those sediments. Several issues surround our understanding of sediment toxicity, including how it is defined and perceived, measured, and interpreted. These issues influence the management decisions that are based on toxicity test results.

Definition of Toxicity

There are clear differences between how toxicologists define "sediment toxicity", and how this term is viewed by non-scientists (Chapman, in press). The word "toxic" to toxicologists has a specific meaning; i.e., an adverse response was elicited in a specific test. Most toxicologists also recognize that there is not a direct extrapolation between laboratory toxicity test results and field or ecological effects. Lay people most often do not have the training to understand the myriad of mitigating factors that may occur between laboratory toxicity tests and field results, or to understand the difference between hazard and risk. To most lay people the connotation of the word "toxic" is often alarming, has many negative associations, and is often interpreted as "bad".

But "good" and "bad" are human value judgements. For example, the Great Lakes were pristine prior to the extensive urbanization and industrialization of the basin. They were subsequently contaminated with phosphorous, and changed to a different, but apparently stable state. They are now contaminated with exotic species (e.g., zebra mussels, sea lamprey, Eurasian gobies) and have changed to yet another apparently stable but different state. Each of these three states differ, and are populated by different organisms; the determination that only the original, pristine environment was not "bad" is a purely human value judgement.

In some cases, regulatory or political acts have served to confuse the public and have led them to equate hazard with risk. For instance, Canada has promulgated the Canadian Environmental Protection Act (CEPA, 1999), whereby substances are called "CEPA toxic". Politicians have enshrined the word "toxic" in an act of Parliament to equate toxicity with risk. Such actions incorrectly reinforce to the public that the word "toxic" always means "bad", and that observed toxicity equates to actual harm. A more accurate description would have been to introduce the concept of contaminants presenting a *potential* risk.

Thus, characterizing toxicity is not the end, but only a means to achieve the end - which is to determine whether or not there is or will be an environmental problem and what may be causing it. Toxicity tests provide an indication of potential environmental harm. Toxicity is only a symptom; it is indicative of an ecological response, but what it actually means depends on the situation. To use an analogy from human health, genetic mutations also have an alarming negative connotation; however, our bodies undergo mutations constantly. Mutation and repair are part of a natural cycle in all organisms. A mutation itself is not cause for alarm, it is only when the rate of mutation is sufficient to overwhelm our natural repair systems that there may be a problem. Toxicity, like mutation, has only potential risk, yet toxicity test endpoints are incorrectly being used as risk endpoints themselves.

There appears to be general agreement within the scientific community that toxicity cannot be defined solely on the basis of chemistry. Toxicity has to be defined as a biological response to a particular test exposure. It is operationally defined by the specific toxicity test, because the test matrix and performance conditions can affect the availability of toxicants to organisms. Although there is no certainty that any particular sediment will be toxic without toxicity testing, predicting sediment toxicity correctly and consistently without conducting toxicity testing is most likely possible at the present time only at the extremes of exposure to the toxicant; i.e., extremes of contamination or non-contamination. In these cases chemical analysis-based assessment may be sufficient for decision-making, especially if bioavailability is taken into account (Adams et al., 1985; Adams, 1987; Di Toro et al., 1991). When concentrations of the substance(s) are neither very small nor very great, only toxicity testing can determine whether biological effects will occur. Typically, chemical analyses primarily serve to assess contaminants of potential concern that may be responsible for observed toxicity. However, such relationships would be correlative, not causal, particularly since non-chemical factors such as sediment composition can also result in adverse responses (Reynoldson et al., 1997). Further, over-reliance on chemical based assessments would make us vulnerable to errors such as "looking under the lamp post because the light is good there"; i.e., if we only look for chemical stressors for which we can conveniently analyze, we will never find the new or different stressors that may cause an effect. This is of particular importance with the increased occurrence of non-conventional contaminants such as pharmaceuticals and personal care products (PPCPs). The bottom line is clear: toxicity currently is defined as a biological response that is best measured directly.

Appropriate Toxicity Test Organisms

Test species used to evaluate sediment toxicity should provide an appropriate indication of the hazards of chemical stressors in the sediment. Amphipods are generally acknowledged as the test organisms of choice for many sediment toxicity assessments, and amphipod toxicity test results can correlate positively with changes in benthic communities (Swartz et al., 1994; Long et al., 2001). However, different amphipods are used in different regions and sometimes even in different situations, without necessarily determining their appropriateness for those regions or situations. Sometimes testing decisions are based solely on convenience. For instance, the estuarine amphipod *Ampelisca abdita* cannot be cultured but must be field-collected. This can lead to variable test results in terms of meeting quality assurance/quality control (QA/QC) acceptance criteria, since organism condition cannot be guaranteed. In contrast, the amphipod *Leptocheirus plumulosus* can be cultured and the test organisms may more readily meet test QA/QC requirements if organism health is guaranteed by optimal culture conditions. Thus, there is increasing usage of *L. plumulosus* in sediment testing programs. However, *L. plumulosus* is less sensitive to copper and some other contaminants compared to *A. abdita* (Schlekat et al., 1995; Ho et al., 1997; McPherson and Chapman, 2000) and thus is not appropriate as an indicator of hazard nor risk in all cases. The interchangeable use of these and other test organisms ignores the fact that even species within the same genus can have very different sensitivities to different toxicants (Chapman et al., 1982).

The need for appropriate test organism sensitivity cannot be overemphasized. The difficulty in identifying hazard lies in choosing appropriately sensitive organisms when one has no or little prior knowledge of the toxicants involved. If we attempt to optimize organism sensitivity to known chemical stressors, we may altogether miss toxicity due to overlooked or non-conventional contaminants. As we cannot always anticipate the cause of toxicity, a battery of organisms with varying sensitivities is the best solution for comprehensive assessments. Similar arguments apply to the test endpoints measured in these organisms.

Linking Toxicity with Causal Stressors

Using toxicity in the context of predicting adverse effects on populations and ecosystems needs to be "proven". For instance, the fact that organisms are not found in sediments where they are expected to live does not necessarily imply exclusion due to toxicity, although this can and does occur. Other factors such as sediment physical characteristics or competitive interactions could explain such exclusions (DeWitt et al., 1988, Chapman and Wang, 2001). Demonstration of the role that chemical contamination might have played in the absence of particular organisms requires evidence of cause and effect. Cause and effect relationships may be demonstrated through application of Toxicity Identification and Evaluation (TIE) methods (Ho et al., 2001). TIE is an approach that incorporates both toxicity testing and simple chemical manipulations in a logical, iterative fashion to identify the cause(s) of toxicity (Norberg-King et al., 1991; Ankley et al., 1992; Burgess et al., 1996). The approach has been used successfully to identify the cause(s) of toxicity in a number of sediments (Ankley et al., 1990; Schubauer-Berigan and Ankley, 1991; Ho et al., 1997).

Uses of Reference and Baseline Toxicity Tests

Definitions of reference and baseline sediments vary between research program and researchers. To avoid confusion, the definition of these two sediments needs to be clearly stated for each project. Some researchers use these definitions interchangeably, while others define reference sediments as the model or ideal sediment for comparison, and the baseline sediment as an historical sediment from the area that is free from anthropogenic influences. While these two definitions are not the same, they are not mutually exclusive; a reference sediment may be a baseline and a baseline sediment may be used as a reference. Here we use these two terms interchangeably.

In typical sediment assessments, the toxicity of test sediments is compared to that of reference sediments, or to a reference condition. In a technical sense, the reference sediment would be the test sediment absent of all the chemicals that might be a problem (i.e., an uncontaminated but otherwise identical sediment). This would allow an evaluation of whether chemicals in the sediment pose a hazard. However, such an ideal situation seldom exists in the real world. The knowledge base is never sufficient to understand all of the features that constitute the criteria for choosing the reference site. As noted by Dr. Dominic Di Toro during the Panel discussion at this Workshop: "…if [we] knew enough about the problem to be able to make a choice, [we would] not need the reference site!"

Reference site comparisons within sediment assessment are, by definition, only estimates of incremental hazard even though statistical data comparisons are undertaken. When comparisons are made, they should be appropriate to the evaluation being conducted. For instance, a harbour should not be compared with a pristine area unless a societal decision has been made to eliminate that harbour and make the area pristine. Comparisons should be based on realistic baselines, management goals and societal expectations. There is precedence for adopting different baselines (e.g., "urban estuaries") reflecting management goals that depend on the expected qualities and different uses of water bodies. In contrast to strict reference comparisons, baseline comparisons can be more informative to decision-making, provided that adequate data exist.

Comparisons to reference or baseline conditions can also be a basis for restoration. However, choosing a single reference point, such as the least disturbed condition or a relatively undisturbed condition, is usually an unrealistic target. The restoration goal should be bracketed by a *range* of acceptable conditions as defined by society and refined by science (Weinstein et al., 2001). The comparison then becomes a problem of determining whether the toxicity of the test sediments falls within that range.

Toxicity Tests in Ecological Risk Assessment of Sediments

In the interpretation of toxicity responses in ecological risk assessment (ERA), the focus shifts from comparisons to extrapolations, in particular extrapolation of laboratory or other constrained toxicity testing results to the unconstrained field situation, and from the responses of individuals to those of populations and communities. Thus, for ERA, the response, not the comparison, is the major feature.

ERAs are typically tiered (Hill et al., 2000), and are iterated until uncertainties have been reasonably elucidated and reduced and a decision can be made in the context of the specific receiving ecosystem(s). The most useful ERAs are supported by research and monitoring. For instance, toxicity test endpoints need to be extrapolated to the field, and monitoring needs to validate expectations from the ERA such as a prediction of no appreciable harm.
ERAs need to address the answers to two equally important questions relative to sediment risks:

- What stressors are of concern in the sediments and do they represent a hazard to valued ecological receptors?
- What is the likelihood that these stressors will adversely impact these receptors?

To do so, toxicity test endpoints need to be linked to specific risk hypotheses that relate stressors to possible hazards, and the results of those tests interpreted in the context of the population or community we are trying to protect.

Appropriate Toxicity Endpoints and Risk Hypotheses

In addition to species selection, the appropriate endpoints to measure in sediment toxicity tests depend on the question being addressed in the ERA. For example, if the sediment is believed to have endocrine disruption activity, it would be inappropriate to choose a test endpoint that did not address a response mediated by the endocrine system. If the stressor affects reproduction, the obvious focus should be on any reductions in the reproductive rate as opposed to endpoints not linked to reproduction. Toxicity tests also require a reproducible test endpoint that can be accurately, predictably and reliably measured.

However, extrapolation of test results to the environment and measuring the success or failure of management decisions may require different endpoints because what is to be protected and what is measured are often different due to practical considerations. Such "assessment endpoints" are selected at the initial, problem formulation stage of an ERA (US EPA, 1992). These endpoints reflect an ecological component (e.g., species) or function to be protected, and when appropriate, are expressed at the population level. The challenge is to determine the relationship between the population to be protected and the response that is being measured, and determine when that response represents an unacceptable impact. We must distinguish between effects (either measured or predicted) and "significant adverse impacts" (e.g., as per the U.S. National Environmental Policy Act and related law internationally), as to what constitutes an undesirable risk.

A known mechanistic connection is desirable to allow for testable risk hypotheses that link the stressor to an adverse impact, and to relate measured effects back to a population endpoint. This is not always easy as there is not always enough information available. A key component of hypothesis formulation is the identification of stressors of potential concern. The conceptual framework developed in the problem formulation stage is then used to examine exposure routes and determine not only what organisms are most likely affected, but also what mode of action the stressors are likely to have, relative to further testing.

Statistical vs. Ecological Significance/Hazard vs. Risk

The term "significant" has both a scientific and a common meaning. When scientists use the term "statistically significant", it does not necessarily mean "biologically or ecologically important". It simply means that there is a stated level of statistical confidence that two things (such as toxicity measured in a test and a reference sediment) are different. On the other hand, "ecologically significant" in ERA means that a measured or observed response has important implications to ecologically-based assessment endpoints. However, ecological significance is difficult to establish, and the usual default in decision-making is to rely on statistical significance. Given our current state of understanding of ecological significance, this can lead to the use of somewhat arbitrary "bright lines" against which measured responses are compared.

For instance, in dredged material assessments in the United States, a significant difference and a 20% reduction in amphipod survival between test and reference sediment can be used as a basis for decisions regarding suitability for aquatic disposal (US EPA and US ACOE, 1998). Other jurisdictions and agencies use different "bright lines", including the sediment concentration that kills 50% of the test organisms, as a threshold for predicting adverse population level impacts. Although statistical significance may have a role in evaluating the hazard of contaminants in sediments, it does not directly support quantification of ecological risk. Despite this, regulatory decisions must and will continue to be made using the information available. Clearly, "bright lines" should be replaced with more meaningful decision criteria as our understanding of the ecological significance of structural and functional responses to stressors improves.

Recognition of the distinctions between hazard and risk (and statistical vs. ecological significance) is critical to managing sediments to meet policy and societal goals. Hazard is a possibility; risk is a probability. Almost everything has a possibility of occurring, but as a society we want to know how likely that occurrence will be. Further, we want to know how important it will be in terms of adverse impacts on valued components of the environment. Put another way, decisions should be made on the basis of the expected magnitude and extent of adverse impacts on assessment endpoints, and not just on simple observations of toxicity in a test.

In the United States, the testing guidance for dredged material disposal (US EPA and US ACOE, 1991, 1998) specifies that actual risks need only be evaluated in unusual, special circumstances; decisions about acceptability for aquatic disposal can be made earlier in the tiered evaluation process based solely on chemistry and the results of toxicity testing. Thus, this guidance encourages evaluation of hazard rather than risk. Further, there typically is no regulatory distinction made in the decision tree between a situation involving a few hundred and one involving few hundred *million* cubic meters of project material. The criteria, based on statistical significance, are exactly the same: if there is no statistical difference between the test and reference toxicity results, open water disposal is acceptable. If there is a statistical difference, open water disposal is not acceptable. Clearly, the volume of project sediment together with its toxicity are important determinants of the ecological impact the material will have upon disposal. Failing a toxicity test only means a test failure, not that there necessarily will be a problem in the environment.

Moreover, toxicity tests should not be regarded as perfect models for predicting what will occur in the environment. They only measure what occurs under specified test conditions. Conditions in the real environment tend to be much different than conditions during toxicity testing. Thus the results measured in toxicity tests only provide an indication of what is or what could be happening in the environment, and should be interpreted accordingly.

Linking Toxicity to Population and Community Response

Sediments are presently regulated primarily on the basis of the responses of individuals (i.e., in toxicity tests), which is a conservative approach based on the presumption of accumulated effects (the "death of a 1000 cuts"). In contrast to human health risk assessment, and except for threatened or endangered species, in ERAs we are generally not trying to protect individuals but rather populations and communities. Moving towards protection of populations poses new issues, such as how to predict effects on populations. Toxicity at the individual level, as demonstrated in the laboratory, generally needs to be shown to have, or to have the *potential* for, population level impacts (e.g., Coull and Chandler, 1992; Swartz et al., 1994; Hunt et al., 2001). Population modeling approaches using toxicity test data can address population dynamics of *test* species (Kuhn et al., 2000, 2001, in press), but not necessarily those of other valued populations in the receiving ecosystem. At a minimum, predictive linkages are required to extrapolate between what is being measured in toxicity tests and the ecological responses of assessment endpoints in the environment. A process-driven, mechanistic understanding of those linkages facilitates accurate extrapolation of test results to assessment endpoint risk.

Assessing Genotypic Adaptations

Populations of organisms exposed to toxicants at elevated exposures are often able to persist and even thrive. This appears to be due either to homeostatic acclimation or to adaptation by genetic selection for resistant individuals (Dahl and Blanck, 1996; Klerks, 1999; Nacci et al., 2001). Acclimation has an immediate metabolic (energetic) cost. Adaptation may not have an immediate cost, but it may have a long-term consequence because the genetic structures of species are modified and genetic traits are being lost or gained. Acclimation and adaptation apply both to natural (e.g., oil seeps, mineral deposits) and to anthropogenic contamination.

ERAs need to evaluate the roles that acclimation and adaptation play in mitigating risks predicted from toxicity tests, and in doing so, recognize the differences between acclimation and adaptation. They should take into account whether there are any costs associated with genetic selection for resistance, and recognize differences in sensitivity between resistant and non-resistant individuals of the same species. These factors will influence the accuracy and certainty of risk predictions and thereby influence decisions regarding management of contaminated sediments.

Addressing Uncertainty

A key issue in risk assessment is uncertainty, and in particular how uncertainties should be weighted in risk management decisions. Generically, the amount of uncertainty tolerated is inversely proportional to the importance of the problem and the cost of making the wrong decision. More effort should be applied to reduce uncertainty for an important problem with many potential downsides, than when the problem is relatively less important. Thus, the issue is one of determining how much uncertainty can be tolerated and how conservative (i.e., protective) to be relative to the potential for adverse impacts. Clearly the size and severity of the potential impacts needs to be considered. When such impacts might be very large, for instance destruction of rare and endangered species, the level of conservatism should be high. In cases where the potential impacts are small, for instance bioaccumulation by fish of one or more chemicals but low risk to the fish or to consumers of those fish, the level of conservatism can be lower. Society has neither the time nor the resources to address all issues exhaustively. However, when unknowns are revealed, and the problem warrants, uncertainty often can and should be reduced. In this regard and at present, the toxicity of sediments can be established with greater certainty than can the hazards and risks of many other environmental stressors.

Summary: Key Issues for Science to Address

ERAs for sediments tend to be very case- or situation-specific, but there is a need for generic procedures to address broader questions such as: What are the ecological risks associated with a particular management option for dredged material? It would be useful to develop a suite, or a menu, of possible tests or tools to serve as screens or indicators of potential risk, due to various kinds of stressors potentially found in sediments. In developing such procedures, the following issues are key to sediment ERA involving toxicity testing:

- Recognition that baselines for comparisons often include measures of humans in the landscape; i.e., "urbanized" or otherwise altered ecosystems.
- Recognition of the fact that detection of chemicals at increasingly lower levels in sediments equates to neither toxicity (hazard) nor increasing risk of adverse ecological impacts.
- A more basic understanding of mechanisms, both biological and chemical, to support extrapolation of hazard to risk.
- Integration of both temporal and spatial variability into sediment toxicity testing and ERA.
- Validation and calibration of ERA models, ensuring that they address variability.
- Field validation of ERA predictions (e.g., monitoring to assess the validity of decisions)
- Moving beyond protection of individuals, to protection of populations and ecosystems by changing the emphasis from statistical or regulatory significance to ecological significance, from hazard (the possibility of adverse impacts) to risk (the probability of adverse impacts). Fundamental research is needed to support extrapolation of toxicity endpoints to population endpoints.
- A societal paradigm shift that adequately reflects an appreciation of risks and the meaning of scientific terms such as "toxic" and "significant" to the *public*. An alternative is semantic changes, for example referring to "responses" rather than "toxicity".

Implementation of ERA to sediment problems faces perceptual hurdles. The regulated community tends to perceive risk assessments as expensive, lengthy and difficult to interpret. The environmental advocacy community mistrusts risk assessment and worries that such assessments are not sufficiently detailed nor site-specific, and that all toxicity test failures will be mitigated by other ERA factors and therefore an ERA will comprise a "license to dump". It is important to promote an adaptive management approach that allows acceptance of a reasonable level of risk and that rapidly incorporates improvements in scientific understanding into regulation and assessment schemes. These issues need to be addressed if we are to make meaningful progress in dealing with contaminated sediments.

Acknowledgments

We thank the New Jersey Marine Sciences Consortium for supporting the Facilitated Workshop Discussion during the Conference on Dredged Material Management and to the audience for attending and contributing. The following contributed significantly to that discussion, and thus to this article: Elizabeth Waters, Bruce Brownawell, Dominic Di Toro, Susan Kane Driscoll, Tom

O'Connor, Frank Steuer-Lauridsen, Scott Douglas, Kristen Milligan, Helder Costa, Walter Berry, Paul Linnan. The authors of this article are some of the members of the Panel for that discussion as well as the convenor of that workshop (MPW). Although two of us (KTH and WRM) are employees of the US EPA, the ideas described in the paper do not necessarily reflect the position or the policy of the Agency and no official endorsement should be inferred. This is NHEERL contribution number AED-02-011, and New Jersey Sea Grant contribution number NJSG 02-475.

References

Adams, W.J. (1987). Bioavailability of neutral, lipophilic organic chemicals contained on sediments: A review. In K.L. Dickson, A.W. Maki and W.A. Brungs *Fate and Effects of Sediment-bound Chemicals in Aquatic Systems* (pp. 219-245). Pergamon Press. New York.

Adams, W.J., Kimerle, R.A. and Mosher, R.G. (1985). Aquatic safety assessment of chemicals sorbed to sediments. In R.D. Cardwell, R. Purdy and R.C. Bahner *Aquatic Toxicology and Hazard Assessment: Seventh Symposium* (pp. 429-453). STP 854. 854. American Society for Testing Materials. Philadelphia.

Ankley, G.T., Katko, A. and Arthur, J. (1990). Identification of ammonia as an important sediment-associated toxicant in the lower Fox River and Green Bay, Wisconsin. *Environmental Toxicology and Chemistry* **9**: 312-322.

Ankley, G.T., Lodge, K., Call D.J., Balcer, M.D., Brooke, L.T., Cool, P.M., Kreis, R.G. Jr., Carlson, A.R., Johnson, R.D., Niemi, G.J., Joke, R.A., West. C.W., Giesy, J.P., Jones, P.D. and Fuying, Z.C. (1992). Integrated assessment of contaminated sediments in the lower Fox River and Green Bay, Wisconsin. *Ecotoxicology and Environmental Safety* **23**: 46-63.

Burgess, R.M., Ho, K.T., Morrison, G.E., Chapman, G. and Denton, D.L. (1996). Marine toxicity identification evaluation (TIE) procedures manual: Phase I Guidance Document. EPA 600/R-96/054, Office of Research and Development, Washington, DC.

CEPA. (1999). Canadian Environmental Protection Act 1999. Ottawa, Ontario.

Chapman, P.M. (1995). Sediment quality assessment: Status and outlook. *Journal of Aquatic Ecosystem Health* **4**: 183-194.

Chapman, P.M. (in press). How toxic is toxic? *Marine Pollution Bulletin*.

Chapman, P.M. and Wang. F. (2001). Assessing sediment contamination in estuaries. *Environmental Toxicology and Chemistry* **20**: 3-22.

Chapman, P.M., Farrell, M.A. and Brinkhurst, R.O. (1982). Relative tolerances of selected aquatic oligochaetes to individual pollutants and environmental factors. *Aquatic Toxicology* **2**: 47-67.

Coull, B.C. and Chandler, G.T. (1992). Pollution and meiofauna: Field, laboratory and mesocosm studies. *Oceanography and Marine Biology Annual Review* **30**: 191-271.

Dahl, B. and Blanck, H. (1996). Pollution-induced community tolerance (PICT) in periphyton communities established under tri-*n*-butyltin (TBT) stress in marine microcosms. *Aquatic Toxicology* **34**: 305-325.

DeWitt, T.H., Ditsworth, G.R. and Swartz, R.C. (1988). Effects of natural sediment features on survival of the phoxocephalid amphipod, *Rhepoxynius abronius*. *Marine Environmental Research* **25**: 99-124.

Di Toro, D.M., Zarba, C.S., Hansen, D.J., Berry, W.J., Swartz, R.C., Cowan, C.E., Pavlou, S.P., Allen, H.E., Thomas, N.A. and Paquin, P.R. (1991). Technical basis for establishing sediment quality criteria for nonionic organic chemicals using equilibrium partitioning. *Environmental Toxicology and Chemistry* **10**: 1541-1583.

Hill, R.A., Chapman, P.M., Mann, G.S. and Lawrence, G.S. (2000). Level of detail in ecological risk assessments. *Marine Pollutution Bulletin* **40**: 471-477.

Ho, K.T., McKinney, R.A., Kuhn, A., Pelletier, M.C. and Burgess, R.M. (1997). Identification of acute toxicants in New Bedford Harbor sediments. *Environmental Toxicology and Chemistry* **16**: 551-558.

Ho, K. T., Burgess, R. M., Pelletier, M. C., Serbst J. R., Ryba S. A., Cantwell, M. G., Kuhn A. and Raczelowski P. (2001). An overview of toxicant identification in sediments and dredged materials. *Marine Pollution Bulletin* (in press).

Hunt, J.W., Anderson, B.S., Phillips, B.M., Tjeerdema, R.S., Taberski, K.M., Wilson, C.J., Puckett, H.M., Stephenson, M., Fairley, R. and Oakden, J. (2001). A large-scale categorization of sites in San Francisco Bay, USA, based on the sediment quality triad, toxicity identification evaluations, and gradient studies. *Environmental Toxicology and Chemistry* **20**: 1252-1265.

Ingersoll, C.G., Dillon, T. and Biddinger, G.R. (1997). *Ecological Risk Assessment of Contaminated Sediments.* SETAC Press. Pensacola, FL.

Klerks, P.L. (1999). Acclimation to pollutants by the grass shrimp *Palaemonetes. pugio*: Individual contaminants vs. mixtures. *Ecotoxicology* **8**: 277-286.

Kuhn, A., Munns, W.R. Jr., Poucher, S., Champlin, D. and Lussier, S. (2000). Prediction of population-level response from mysid toxicity test data using population modeling techniques. *Environmental Toxicology and Chemistry* **19**: 2364-2371.

Kuhn, A., Munns, W.R. Jr., Champlin, D., McKinney, R., Tagliabue, M., Serbst, J. and Gleason, T. (2001). Evaluation of the efficacy of extrapolation population modeling to predict the dynamics of *Americamysis bahia* populations in the laboratory. *Environmental Toxicology and Chemistry* **20**: 213-221.

Kuhn, A., Munns, W.R. Jr., Serbst, J.R., Edwards, P., Cantwell, M.G., Gleason, T., Pelletier, M. and Berry, W. (in press). Evaluating the ecological significance of laboratory response data to predict population-level effects for the estuarine amphipod *Ampelisca abdita* . *Environmental Toxicology and Chemistry* **21**.

Long, E.R., Hong, C.B. and Severn, C.G. (2001). Relationships between acute sediment toxicity in laboratory tests and abundance and diversity of benthic infauna in marine sediments: a review. *Environmental Toxicology and Chemistry* **20**: 46-60.

McPherson, C.A. and Chapman, P.M. (2000). Copper effects on potential sediment test organisms: The importance of appropriate sensitivity. *Marine Pollution Bulletin* **40**: 656-665.

Moore, M.T., Testa, S. III, Cooper, C.M., Smith, S. Jr., Knight, S.S. and Lizotte, R.E. Jr. (2001). Clear as mud: The challenge of sediment criteria and TMDLs. *Water Environment and Technology* **13**: 49-52.

Nacci, D.E., Gleason, T.R. Gutjahr-Gobell, R., Huber, M. and Munns, W.R. Jr. (2001). Effects of chronic stress on wildlife populations: a modeling approach and case study. In M.C. Newman, M.H. Roberts, Jr., and R.C. Hale *Coastal and Estuarine Risk Assessment: Risk on the Edge* (pp. 247-272). CRC Press/Lewis Publishers. New York.

Norberg-King, T., Mount, D.I., Amato, J.R., Jensen, D.A. and Thompson, J.A. (1991). Toxicity identification evaluation: Characterization of chronically toxic effluents, Phase I. EPA-600/6-91/005, Washington, DC.

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E. and Rosenberg, D.M. (1997). The reference condition: a comparison of multiple and multivariate approaches to assess water quality impairment using benthic macro invertebrates. *Journal of the North American Benthological Society* **16**: 833-852.

Schlekat, C.E., Scott, K.J., Swartz, R.C., Albrecht, B., Antrim, L., Doe, K., Douglas, S., Ferretti, J.A., Hansen, D.J., Moore, D.W., Mueller, C. and Tang, A. (1995). Interlaboratory comparison of a 10-day sediment toxicity test method using *Ampelisca abdits*, *Eohaustorius estuarius* and *Leptocheirus plumulosus*. *Environmental Toxicology and Chemistry* **14**: 2163-2174.

Schubauer-Berigan, M.K. and Ankley, G.T. (1991). The contribution of ammonia, metals and nonpolar organic compounds to the toxicity of sediment interstitial water from an Illinois River tributary. *Environmental Toxicology and Chemistry* **10**: 925-939.

Swartz, R.C., Cole, F.A., Lamberson, J.O., Ferraro, S.P., Schults, D.W., DeBen, W.A., Lee, H.I. and Ozretich, R.J. (1994). Sediment toxicity, contamination and amphipod abundance at a DDT and dieldrin contaminated site in San Francisco Bay. *Environmental Toxicology and Chemistry* **13**: 949-962.

(US EPA) US Environmental Protection Agency. (1992). Framework for ecological risk assessment. EPA/630/R-92/001. Washington, DC.

(US EPA and US ACOE) US Environmental Protection Agency and US Army Corps of Engineers. (1991) Evaluation of dredged material proposed for ocean disposal: testing manual. EPA/503R-91/001. Washington, DC.

(US EPA and US ACOE) US Environmental Protection Agency and US Army Corps of Engineers. (1998). *Evaluation of Dredged Material Proposed for Discharge in Waters of the U.S. – Testing Manual.* EPA-823-B-98-004. Washington, DC.

Weinstein, M.P., Teal, J.. M., Balletto J. H. and Strait K. A. 2001. Restoration principles emerging from one of the world's largest tidal marsh restoration projects. *Wetlands and Ecological Management* **9**:387-407.

Facilitated Workshop Discussion

CONFERENCE ON DREDGED MATERIAL MANAGEMENT: OPTIONS AND ENVIRONMENTAL CONSIDERATIONS

SEDIMENT TOXICITY AND RISK ASSESSMENT TOOLS: WHERE ARE WE AND WHERE SHOULD WE BE GOING?

Facilitated Workshop Discussion[3]

6 December 2000
Massachusetts Institute of Technology
Cambridge, Massachusetts

[3] Original Panel Discussion transcript prepared by O'Brien and Levine Court Reporting Services, Boston MA. Edited by the New Jersey Marine Sciences Consortium. Any errors, omissions, or misrepresentations in this edited version are wholly those of the NJMSC.

PANEL MEMBERS

ELIZABETH "BITSY" WATERS, Moderator

BRUCE BROWNAWELL
SUNY, Stony Brook

DOMINIC M. DITORO
Manhattan College

KAY T. HO
U.S. Environmental Protection Agency
Narragansett

WAYNE R. MUNNS, JR.
U.S. Environmental Protection Agency
Narragansett

PETER M. CHAPMAN
EVS Environment Consultants

KEITH SOLOMON
University of Guelph

FACILITATED ROUNDTABLE DISCUSSION

(Introduction by Michael P. Weinstein, Workshop Chair)

DR. WEINSTEIN: Good afternoon, we are about to begin a dialogue among the panel and the audience. We have a court reporter and the idea is to get a complete transcript and edit it down to an Executive Summary-type of document from which the Panel and I will prepare a synthesis article. One ground-rule, please be succinct in your questions and your responses so that we can capture the essence of today's discussion. I ask us all to keep to the science in addressing the question: What is worrisome about these sediments that we are trying to manage? Let us focus on ecological risk assessment just to constrain the discussion and to have control over the document that we are trying to produce. We will stay away from human health -- not to denigrate that important area -- but for just purposes of conservatism. Let me introduce our facilitator, Ms. Elizabeth "Bitsy" Waters who I think you will find will do an excellent job of keeping us on track. And please, I will ask everyone to be responsive to Bitsy, she has great ability to keep the discussion focused.

MS. WATERS: Thank you very much. Let us start with the Panel: How does a scientist define "sediment toxicity." When a scientist talks about sediments being toxic, what do they mean?

DR. CHAPMAN: What makes sediment toxic? Simply, it is a [biological] response that indicates toxicity. But, the issue is not just toxicity. The issue is one of contamination, which includes all of the chemicals out there. And the ultimate issue is one of pollution: Is it polluted? The fact it is toxic may not mean anything in the real world. It may only be a symptom. Whether it is part of the disease or not is another totally different question.

DR. DITORO: There are basically two meanings, and I think some of the confusion about sediment toxicity resides in the fact that there are these two meanings. The first meaning is technical. The word "toxic" to toxicologists has a specific meaning; i.e., a response was elicited in the test that was being applied. And that is a technical use of a term which also has nontechnical connotations. But to the body politic, "toxic" means bad and something to be avoided. I think a lot of the confusion in the lay press is that the word "toxic" is used with the latter meaning, i.e. something to be avoided, as opposed to the former meaning which is simply a response elicited in a test.

DR. BROWNAWELL: It is a question that does not have a simple answer. But what comes to my mind especially, in the coming years, is that we are going to have increasingly more sensitive tools [e.g., gene expression]; very sensitive measures of elicitation of a biochemical response in organisms that one could call "toxic." Any time you have an adaptation to such a response, there are some energetic and potentially competitive costs associated with that. As a citizen, the question is: When does that matter? And in risk assessment, there are some tools that can allow us to address the question. But from my view, the answer depends partly on how big an area that response is affecting. For example, is it a regional response? In which case, maybe we ought to do something about it? Or is it a small footprint in the middle of a large area? It is a matter of scale. This makes it challenging for dredged material management because we want to set criteria for sediments regardless of what we are going to do with them. From a testing perspective, I am not sure that is an appropriate way to look at things. And certainly, using risk assessment tools is the way to probably make some ground on that question [of response area].

DR. HO: I am probably the only environmental toxicologist up here. As an environmental toxicologist when someone says "toxic" to me, I instantly picture a small jar with amphipods floating on the top. And that is coming from the laboratory side. But, despite the conceptual problems that we may have with it, it is fairly well correlated with what we see in the field. There are pluses and minuses, but they are not bad tools to assess toxicity in the field, in terms of what is really happening. As our first cut and as a reasonable tool, I think that our toxicity [criterion], 60 percent survival, that type of thing, is pretty good.

DR. SOLOMON: We should also ask the question: When is something not toxic? Perhaps someone from the audience can address this? A point that struck me earlier comes from the work of Reynoldson and Day [et al.]; the idea that if organisms do not live in the sediment, it does not necessarily mean it is toxic. It may mean that physically it is not an appropriate habitat for that kind of organism. So we have to be very careful about what we reference things to. What is the control sediment? And is the organism adapted to inhabit that type of sediment?

MS. WATERS: If I put Dr. DiToro's two definitions together, one based on technical considerations; the other based on the public perception: anything toxic is bad. And if I add Dr. Chapman's comment: yes, it is toxic, but what problems is it causing? Is there going to be a long-term terminology problem among you scientists, saying, this is toxic and then trying to say, but do not be worried about it? If we move towards a risk assessment approach to handling this issue, is that a problem? And are there some ways that we might grapple with it so that the minute you say "it is toxic", no one has bailed out and stopped listening.

DR. CHAPMAN: Remember that modern medicine is based on toxicity. When we take vaccines and when we take shots, we are basically poisoning ourselves to a limited extent. Toxicity, per se, does not necessarily mean bad. It is toxicity in the context of the symptoms. Toxicity indicates something. What it indicates depends on the situation. Take an analogy from human health. It may be that you have a problem with your heartbeat. Does it actually mean that you have something wrong with your heart? It may be stress. It may be something else. You can not just go to the symptoms. You have to look further. And the problem we are getting into in legislation, in the public's mind, is that toxicity is the end, when it is only a means to achieve an end.

DR. DITORO: Bitsy's comment was: Should we change the word? We have a word that has a public perception which is completely negative, and we use it in a technical way. When we are talking among ourselves, we know what we mean. But when it gets translated and it has a connotation that we do not mean, we have two choices: Either we try to educate 200 million people, or we change our business and call it something else.

DR. CHAPMAN: What would you call it?

DR. DITORO: I am not sure. "Response?" It is a test response.

DR. SOLOMON: In Canada, we have this dreadful thing called the Canadian Environmental Protection Act (CEPA), where substances are called "CEPA toxic". They have actually enshrined that word in an act of Parliament. It actually means they [these contaminants] present a risk. It does not mean they are toxic. They are totally misusing the word. And it is a great concern to me because it is actually saying to the public that the word "toxic" always means "bad." Do not do that here. If you have the opportunity, do not put it in the regulations.

MS. WATERS: Audience?

DR. DRISCOLL: Susan Kane Driscoll from Menzie-Cura & Associates. I would argue that a working definition of "sediment toxicity" is an adverse effect on survival, growth or reproduction that is often assumed to have population level effects. And there certainly are examples where significant effects on survival, growth and reproduction occur, but the point is: when do you move from toxicity testing to risk assessment? You have to provide better evidence that these effects that we are seeing are having population, or have the *potential* [emphasis added]for population level effects. Up until now, we have just taken the precautionary approach and said, "well, it could". So that is good enough and we will stop there. And now we are calling for more evidence to show that those adverse effects do, indeed, cause population effects. And I think we are kind of stuck in the middle. We need to take the next step, if we are going to move into a risk assessment arena.

MS. WATERS: You are assuming "toxic" means not just an effect, but an adverse effect?

DR. DRISCOLL: Right. Dr. DiToro's definition might be a response to a toxicant of any type. But I think a more accurate definition that people use more often includes consideration of survival, growth and reproduction. And certainly, other people would argue for sub-cellular level effects too, some of which can be tied to higher order effects, but some of which cannot. But I think you are on, perhaps more formal ground, if you stick to survival, growth and reproduction. At least that is the way we have often used it in the risk assessment world.

DR. MUNNS: Endpoints aside, I think an issue that we have not laid out yet, is the idea that above some level of a response, it is toxic; below some level of a response, it is not. I do not mean that in a threshold sense. What I mean is we often apply somewhat arbitrary criteria. I think this is a bit of what you are trying to get at, Dr. Driscoll; e.g., if we see greater than, for instance, 20 percent mortality and it is different from the control, we say "that is toxic." So Dr. DiToro's definition of replacing the word "toxic", with "response", I think is more of a risk assessment view; you speak about exposure response relationships instead of something [just] being toxic. You want to understand the level of the response, not just that it is above or below some threshold necessarily.

MS. WATERS: In summary, would you simply call this a "response," rather than a "toxic effect"?

DR. DITORO: It is fairly easy to change the language as long as we all agree that that it is more precise. But I am not a card-carrying toxicologist, so I cannot make these judgements.

DR. SOLOMON: One of the issues, of course, is that all of the people in this room understand the difference between "toxic" and "risk." You all know what we are talking about when we say something is toxic. The problem, when you move outside of this room and you deal with regulators and the public, is that they lose the meaning that we have here. And we can continue to talk this secret language in and amongst ourselves, but I think we have to be more careful about how we deal with these issues when we deal with the public or when we deal with information transfer outside of our scientific group.

DR. CHAPMAN: I would argue that it is not so within the group that we have right here. I have seen too many cases where scientists have messed things up. A good example is the issue of contamination versus pollution. I have lost count of the meetings I have stood up and corrected scientists when they say things are polluted and all they have is chemistry to go on. "Contamination" simply means something out of place or present at too high a level. Fresh water contaminates a sea from a river. That is not pollution. Pollution is contamination causing an effect. You have to have contamination to have pollution. But "contamination" does not mean "pollution." Pollution is what we are after. By the same token, toxicity only becomes important when it really has effects; as Dr. Driscoll pointed out, at the population level. Yes, we want to look at survival, reproduction and growth because that is our best link from individuals to populations, but that still is not a fully secure link. We need to go from hazard to some hybrid of hazard to risk. That is where we need to end up.

DR. HO: I am not sure that I agree with that. I would not say that toxicity operates only at the population level because, first of all, I do not think that we can easily make that link from what we see in a jar to whether or not it affects a population. Maybe it does not have a population level effect. Maybe it has a community effect. Or maybe it has some other type of effect. But until we can make that scientific link with some certainty, I would not totally negate toxicity just because you cannot say it has a population level effect.

DR. CHAPMAN: I am not negating it. But when you address toxicity in risk assessment there is a lot of uncertainty and you are being very cautious. And as you use risk assessment, you are getting more information in trying to define toxicity, and so on. So it depends, first of all, on what level of protection you want to achieve; secondly, on how bad the toxicity might be. We are not necessarily trying to protect individuals in terms of ecology. We are in terms of human health. In ecology, we would want to protect a *group* [emphasis added] of individuals. Whether they call them "populations" or "communities", it is basically the same thing. Or do you want to protect individuals?

DR. MUNNS: I would say you also want to protect individuals under certain circumstances. Some of those are legislatively dictated, like threatened and endangered species. But if we are going to push risk assessment, then it all depends upon how we couch the assessment endpoints.

DR. CHAPMAN: Yes.

DR. MUNNS: But you cannot arbitrarily say it is always populations and above. You have to understand the endpoints we are evaluating those risks against. That is not just an ecology decision. It also has social implications and economic implications.

DR. BROWNAWELL: I have a question that I am not comfortable with my own answer to: How do we consider the risk associated with genotypic adaptations to contaminants? More and more, when we look carefully, we find that organisms exposed to toxicants at elevated levels are doing fine at the population level, but there has been some genetic selection for resistant species; or resistance in the gene pool. And presumably, that comes at some unknown cost. The question is: How do we place that into risk assessment?

DR. MUNNS: I think that is an aspect to consider because someone provided a definition [of toxicity] earlier that included observing a response to a test. The measured response depends on the organisms that are used in the test itself. In the case of predisposed or preselected genetically resistant organisms, sediment, or any other environmental medium, is going to be rated as toxic in one test, whereas in another test with a different set of [resistant] organisms it is not. So I think it does play into the definition. I am not sure how to resolve it.

DR. CHAPMAN: It acknowledges the difference between acclimation and adaptation. Acclimation does have an immediate cost. Adaptation does not have an immediate cost, but it may have a long-term cost because species are being displaced. This is a big issue. In fact, I have a series of papers commissioned with Klerks and Millward for a community based ecological risk assessment to look at this whole question of adaptation and how we deal with it. I am certain it is not just connected to anthropogenic issues. There are lots of areas that are highly mineralized in the world where organisms have adapted incredibly. And the level of contaminants they can tolerate naturally, without man's interference, is unbelievable.

DR. SOLOMON: I would add that natural toxins existed long before anthropogenic toxins. Plants produce lots of natural toxins that they use to defend themselves. And many organisms have, I presume, initially acclimated to them. And then, of course, evolutionary pressures selected for survival those that could tolerate these toxins. So the original acclimation was supported by the selection of genetic traits and the adaptation that allowed them to survive without expending excess energy apart from the original innate change. We have a history of that. And I do not know that we need to differentiate between anthropogenic and natural processes here. Both are happening.

DR. O'CONNOR: Tom O'Connor with NOAA. Toxicity is certainly operationally defined by the test. But inherent in this question is whether or not a scientist can define toxicity just on the basis of chemistry. Can we resolve that?

DR. CHAPMAN: No.

MS. WATERS: Do you want to elaborate a little bit?

DR. CHAPMAN: The answer is no.

DR. O'CONNOR: Does anyone not think the answer is no?

(No verbal response from panel.)

DR. O'CONNOR: Good. One thing is resolved here. Toxicity has to be defined on the basis of a biological response in a test that we arbitrarily define at the beginning of a project.

MS. WATERS: Having said no, toxicity cannot be determined by chemistry alone, would you respond affirmatively to the general statement [that toxicity will be operationally defined by the test] without saying exactly what the biological test will be and what it will be evaluating?

DR. SOLOMON: Yes. That is a reasonable statement with the qualifier. And the reason is that the matrix may affect the availability of the substance to the organism. And I think that is why it is true.

DR. O'CONNOR: Yes.

DR. DITORO: In the most simplified case it essentially becomes one to one: Pure water, cloned organisms that are all genetically identical, a simple chemical that does not have a lot of complexity. But from a definitional point of view, there is no doubt: Toxicity is a biological response.

DR. DRISCOLL: Let's not forget about the margins. For any particular sediment, you cannot say with certainty that you will be able to predict whether or not there is going to be toxicity. But for any particular sediment that is at the margins, i.e. highly contaminated at one end or not highly contaminated at the other end, you can probably say yes [it is toxic or not toxic] with a certain degree of probability. It becomes a question of resources. Do you want to spend the money to answer the question, or are there a lot of cases where you know what the answer is going to be? And I think that the Port Authority [NY & NJ] can tell you that for many samples, they do not even bother to do the sediment toxicity test because they know it is going to be toxic. So, for a lot of projects, they do not do any testing. I think that this is already being implemented. It appears that the models have some utility. It is the gray area where sediment toxicity tests are more useful.

MR. STEUER-LAURIDSEN: I am Frank Steuer-Lauridsen from Lyngby, Denmark, another international participant in this conference. I got the feeling that we all answered no to the fact that chemicals or chemistry was not important here.

(Panel all responds no, that was not their intent.)

MR. STEWELL: Good. Because I want to say I think chemistry is very important in defining toxicity. Toxicity is a bilateral thing. You have a biological response. And you have something causing that biological response. It can be a physical stressor. It can be a chemical stressor. Usually, it is a chemical stressor. And that is part of the definition for toxicity in sediments. Sediments, as such, are not toxic. Some environmental factor in the sediment provides the toxic effect that we measure. Can the panel comment on that?

MS. WATERS: Does anyone want to respond relative to the previous yes/no discussion [on chemistry]?

DR. BROWNAWELL: I would like to respond to Dr. O'Connor's question [can we define toxicity on the basis of chemistry alone?] because I was the one that brought it up earlier when I said I would like to see a future in which we did less toxicity testing for management purposes for dredged materials and more chemistry-based risk assessment. And I do not think we are at that stage yet because we do not even know what the contaminants are that we need to be looking at. But I can certainly imagine a future, a not too distant future, when we could have as much or more reliability and predictability based on chemistry risk assessment than we will have on a simple arthropod toxicity test. And that is my take on the system. But we are not there today.

MS. WATERS: [To all] do we agree on that? And if so, someone please state it. If not, I will simply acknowledge that we largely agree, but with some uncertainty.

DR. BROWNAWELL: I would say that we are clear. At the present, in this day of science, we depend on toxicity testing of some sort to measure a biological response. At some point, perhaps we can move the chemistry forward -- and Dr. O'Connor has a lot to do with that where we can rely more on chemistry and understand more mechanistically. We cannot at this point in time. And I think that is basically what you are saying. And I definitely agree with you. We definitely need the chemistry because we cannot just have toxicity without knowing what is going on. The only caveat I would add is that grain size can be toxic and sediment themselves can sometimes be toxic in and of themselves.

MS. WATERS: The next question is how we establish a baseline for toxicity. In some ways, this is not totally unrelated to the discussion we have just had. But it tends to move us forward. Do we have baselines that we use now? Do we need new ones? What might they be?

DR. MUNNS: Well, certainly in the US, an evaluation process compares the toxicity of the test material to some reference condition. That is a baseline in a sense. And I will go back to something I said before: As we move more and more towards ecological risk assessment, I think we will probably end up being less concerned with comparisons to the reference condition and more concerned with responses that can be extrapolated to the field. It is not quite the same thing as saying that management's concern will go in that direction because there are other factors, e.g., anti-degradation policies and policies to improve the environment. But I think from a purely risk assessment standpoint, I would argue that it is the response to the test material itself, not a comparison to a reference.

MS. WATERS: Other thoughts?

DR. CHAPMAN: I disagree. A harbor is a harbor is a harbor and will always be a harbor. We get too caught up sometimes in trying to clean up the stuff that cannot be cleaned up and putting resources into situations where they are wasted. I believe reference comparisons are important in a societal reference framework because you need to be able to say, "okay, this area is going to be a harbor." It is never going to be totally clean. There is nothing we can do to totally clean it up. What is the most reasonable clean up goal that we can afford, as a society; that we can agree to? And then make reference to comparisons based on that goal instead of what is happening now. And a lot of reference comparisons, I believe, are quite inappropriate because they are the cleanest of the clean in all cases, which should not always be the goal. And I remember an EPA publication that was held up four years because clarification was needed on what was meant by "reference" [for the comparisons].

MS. WATERS: Does that response suggest a flexible baseline concept, i.e., a baseline that is different depending on the kind of situation you are looking at?
DR. HO: That is a question that is still really up in the air, and it is a question that I think society and local communities are deciding. It is more than a managerial question and possibly a local question.
DR. BROWNAWELL: And there is management precedence for having different expected qualities for different uses of water bodies. And that is basically in line with what Dr. Chapman has said, but there are situations mandated by laws, as we learned today, where we have to do comparisons to reference sites. And recently, I have been thinking about the HARS, the Historic Area Remediation Site. It is a real technical problem to think about how you do controls for the HARS site. You have a pile of mud in the middle of a bunch of sand. That is one problem. How, for example, do you do comparisons in terms of indigenous species? And the other problem: it is not a clean, pristine area. So there are a number of technical issues [if they are going to implement the sort of legislation that we talked about today] on how one defines "baseline." And once you define a "baseline", what are the approaches that are available for making some reasonable comparisons? To compare a sediment sample with one species of polychaete to a sediment sample from another site, perhaps with a different species of polychaete, is a waste of financial resources. Perhaps for PCBs [polychlorinated biphenyls] it might be okay, but for just about anything else, it is a waste of time.
DR. HO: Why PCBs?
DR. BROWNAWELL: Because PCBs are pretty well assimilated by all polychaetes, and you do not see big variations in the BCFs [bioconcentration factors] of PCBs; for example, between different species and different environments. But for other contaminants, you have different abilities among species in their adaptive responses, in terms of storage and other protective mechanisms. Different species have different abilities to cope with contaminants and to assimilate contaminants. If you are going to mix them just because they are in the same phylum [related group of organisms], it may sound good, but to me, it is a waste of time.
MS. WATERS: Audience?
MR. DOUGLAS: Scott Douglas, New Jersey Maritime Resources. Is there a difference between a baseline and a goal; i.e., a remedial goal or a management goal? I ask, because I sometimes think we get confused on which one we are talking about. We have references. We have baselines. And we have goals. Thoughts?
DR. DITORO: I think the question is unanswerable.
DR. CHAPMAN: It is a legislative question.
DR. DITORO: Yes. When somebody asks: How do you pick a reference site or reference toxicity or reference whatever?; it presupposes that you know enough about the problem to be able to understand what features constitute the criteria for choosing the reference site. But if you knew enough about the problem to be able to make the choice, you do not need the reference site! As a consequence, it is impossible to answer the question. One can speculate on Long Island Sound, for example. Someone says: everyone knows that the sediment out there [in Long Island Sound] has the following characteristics and therefore is different. And so it presupposes the knowledge of causation which is not there. Thus, the problem is unanswerable. And therefore, I have never had any trouble with these regulatory schemes which compare reference site to impacted sites. It is strictly nonscientific.
MS. WATERS: So you are using "baseline" and "reference" interchangeably? Correct?
DR. DITORO: Essentially.

DR. BROWNAWELL: There are places where people talk about baseline as being pre-Industrial Revolution or Revolutionary era.
DR. DITORO: The original sediment criteria for the Great Lakes were based upon going to pre-anthropomorphically contaminated sediments and making chemical measurements and saying, that is it. That is what we want to get back to. Those were the original Jenson criteria that were established for sediments in the prehistory of sediment criteria.
MS. WATERS: Is that what you are calling the goal Mr. Douglas?
MR. DOUGLAS: I am not making judgements. I am asking the Panel.
DR. DITORO: I think the way the question is phrased and the way it is used in a regulatory context makes no logical sense because there is no operational way of establishing the criteria for a reference site. If you actually knew what constituted a reference site, you could choose one explicitly. Then you know enough about the situation, so you do not need a reference site. You would understand all the mechanisms that affect the population.
DR. MUNNS: I would add that that is the conundrum not only for dredged material, but any kind of situation where we are evaluating contaminated sediments. I can remember numerous arguments in the program about just that issue.
MS. MILLIGAN: Kristen Milligan, I am from Clean Ocean Action. I am a little confused about our definitions of reference and baseline. The operational definition of "reference" is typically a [toxicity] testing definition. You have your test sediments; you have your reference sediments and your control sediments. And most of the time that is how I am exposed to the term "reference sediments." The "baseline," is an entirely different consideration even though reference sediments may be intended to be baseline. I think baseline refers to your intent; i.e., what is the question you are asking? I tend to look at the world as: what is our question for baseline? And I think a reference should also be considered in this realm, as well. If your question is remediation, your baseline or reference might be very different than if your question was; e.g., placement impacts at a disposal site. If you are going to be accepting a certain level of cumulative impact, your baseline may be different than if you have a remedial goal. Likewise, if your question is: are we going to have a toxic response?, the baseline is going to be very different. So I think the definition question is entirely dependent on the context of how you are asking it. I do not know if we can come up with one set definitions for either "reference" or "baseline."
DR. CHAPMAN: To me, baseline and reference are real easy. "Baseline" is what it was. "Reference" is what you would like it to be, want it to be or are willing to accept it being. That is it, period. For example, take a sewage treatment plant before it goes in, baseline is what it is then. In terms of dredged material disposal, baseline might be what it was before you started dumping. The reference condition you compare it to is: what you are willing for it to be, which might not be as good as it was before or might be better depending on the situation. I agree with Dr. DiToro that it is not a scientific decision. It is a societal decision. I take this point as a purist and say we do not know everything, but we sometimes have to make decisions. And I think in many cases, we can make some reasonable decisions and say: okay, if this is what we want this area to be like in the future, what is the range of water quality conditions that are most reasonably going to approximate those conditions?
DR. DITORO: Let me clarify something. The idea of a baseline or reference condition started when comparative testing was brought to bear on the dredged material program. You would collect a baseline or reference sediment, and then you would test it against the dredged material. And then you had a framework for responses that might tell you something. The technical

definition is this: the reference sediment is the sediment absent of all the chemicals that are in it which might be a problem. And in a hypothetical world, that is what you are trying to find because then you can evaluate whether or not the chemicals in the material for disposal are doing anything. What you are looking for is a sediment that is identical in every way except for the fact that it is contaminated. That is the definition. If I could test this against the dredged material or a control; same idea, right?

DR. CHAPMAN: No.

DR. DITORO: Hold it a second. One step at a time. That is the easiest conceptual picture of a baseline or reference sediment. If I am going to do a test, I need some [frame of reference].

DR. CHAPMAN: It is wrong [in my opinion], but please finish your point.

DR. DITORO: Okay. That is what I wanted. I then ask the question: where can I find such a sediment? And at that point, the problem becomes clear because you do not know the characteristics that make up everything except the constituents in the sediments. And therefore, it is a nonoperational definition. You cannot go get them [reference sediments]. That is what I mean. You, as a benthic ecologist, can look at it [the sediments]and say, It looks to me like that is the same as that. But you do not really know that. And therefore, the notion is nonoperational. That is what I mean.

DR. MUNNS: Even though I am not a toxicologist, my understanding of the various uses of the terms "reference" and "controls" differ in that controls are intended to do what I just heard you describe, Dr. DiToro.

DR. DITORO: Yes.

DR. MUNNS: Controls are for all those factors that may be occurring that you can not physically control for.

DR. DITORO: Correct.

DR. MUNNS: Whereas a reference, I am starting to like the definition we heard from the audience earlier of something that you want to manage towards, whatever that is! Good, bad or ugly, it is something you want to manage towards.

DR. DITORO: But I ask you: Where do I find one [reference sediment]?

DR. MUNNS: Well, you have to decide. Society has to decide.

DR. DITORO: No, no. I asked you where (near the place) that I am addressing this problem do I go to find one [reference sediment]?

DR. MUNNS: That is a site-specific answer.

DR. DITORO: But how do you collect it [reference sediment]? How do I decide that [place you go to collect] reference sediment?

DR. BROWNAWELL: What are the criteria?

DR. DITORO: What are the criteria? There is not a sign that says, this is a reference setting.

DR. MUNNS: From a scientific standpoint, I would agree.

DR. DITORO: Good.

DR. MUNNS: But I mean you cannot decide.

DR. DITORO: Yes.

DR. MUNNS: I will change hats from sediment toxicity to restoration ecology. So, let me use a restoration ecology analogy.

DR. DITORO: That is better.

DR. WEINSTEIN: Mike Weinstein from the Marine Sciences Consortium. I think I heard the whole discussion couched mainly in a singular framework: that of baseline. I am about to go across the hall, and talk about restoration of coastal degraded salt marshes to some new stable endpoint. And we have reference marshes that we are looking at. But, as I suggested to a speaker the other day, if your frame of reference is the least disturbed condition or a relatively undisturbed condition, then you are choosing an unrealistic target. The reference has to have some kind of bound; i.e., the restoration goal has to be reflected in some *range* [emphasis added] of acceptable conditions; e.g., the urban condition,, the extended urban condition, taking the urban mud and putting it out somewhere in the ocean and everything in between that becomes an acceptable endpoint. In other words, we have to define the limits of the bound, in a societal context, and then get the best science involved. And I think that scientists can participate greatly in that process and then put it in the public purview to get consensus on what is an acceptable bound. And that is the target [goal] you are looking for. You are not going to go back to prehistory. On the other hand you do not want the "black mayonnaise" of the inner harbors and poorly flushed areas. That is probably unacceptable, but you have an acceptable range in between. How do you get there?

MS. WATERS: Our next related question is how do we select appropriate endpoints? We were given several possibilities this morning. People are obviously thinking about this. So to the panel: What kinds of endpoints should we be talking about?

DR. HO: It depends on the question. The endpoint that you might choose is to monitor sediment toxicity. For example if a sediment has endocrine disrupter activity, it would be silly to choose an endpoint does not address that particular effect. Although a test for that effect is not in the normal suite of regulatory driven toxicity tests, there are other tests out there that are appropriate.

DR. SOLOMON: Yes. This morning, Dr. Munns discussed an ecological risk assessment framework that originally came out in 1992. It is at the problem formulation stage that you select your assessment endpoints; these are the things you are trying to protect: e.g., populations, individuals (endangered or anthropomorphic species), etc. Usually, these are expressed at the population level [in an ecological context]. For example, we will not tolerate more than a 10 percent reduction in the population or the holding capacity of that environment. The challenge is to figure out the relationship between the population you are trying to protect and the effect that is being caused. If it is an endocrine-mediated response and it affects reproduction, you obviously would not want it to reduce the reproductive rate too much. And you need to have that mechanistic connection known so you can set up your hypothesis and then use the concentration or the measured effects of those particular substances to relate back to the population endpoint. And that is not always easy. We do not always have enough information to do that. But it is certainly addressed in the problem formulation stage of that whole approach.

DR. CHAPMAN: I agree with Keith. One of the things you do in hypothesis formulation is identify stressors of potential concern. And I use the word "stressors" instead of "contaminants" because they do not necessarily need to be contaminants. Do your conceptual diagram. Look at the exposures and determine what mode of action they [stressors] are likely to have. And that will help guide you in terms of further testing. If, for example, you have mercury as a big issue, you worry about acute effects, but you also worry about food chain effects. So you are guided by what is there, what may be occurring, and design your test and endpoints appropriately.

MS. WATERS: Is this something in risk assessment you are doing anew every time, or are you working within an established menu of endpoints?

DR. MUNNS: I think there is a tension between doing risk assessment the way we have been describing it (very case-specific or situation-specific) and the need to have generic procedures at hand to address the broader question: Are there ecological risks associated with this management option for dredged material? I think that tension is probably going to lead to a suite, or a menu as you have just called it, of possible tests to serve as screens or indicators of potential risk, due to various kinds of stressors. We now look for test species and test conditions that allow us to pick up not only metals toxicity, but organic [compound] toxicity as well. If we can expand this effort to include other chemical stressors [that may not work through narcosis or other more common means, like endocrine disrupters], I would support the idea of a menu. That adds costs, so I do not know where the balance lies.

MS. WATERS: Audience?

MS. COSTA: **HELDER** Costa, from Blasland, Bouck & Lee. We heard several discussions where the chemical contamination was down a few layers in terms of importance, in terms of ecological effects. Is the question on the floor about endpoints more in terms of what protects populations? My sense, from Dr. Chapman's comment on this question, is that we are going to start with chemicals. What are the chemicals of concern for a given site and what are our ecological endpoints, or effect endpoints, based on those chemicals? Or do we start at: Here is our population. Here are the potential points of vulnerability and then work back to the chemicals?

DR. CHAPMAN: I think it is both. Dr. Solomon put one side on it, and I put the other side on it, but you need both sides to make the whole.

DR. BROWNAWELL: I would like to address the question from a pragmatic viewpoint: What are we doing now in terms of endpoints for dredged material disposal? And what are the things that we might be doing better? We currently have two biological endpoints. We have bioaccumulation and in this part of the country, it is *Macoma balthica* and *Nereis virens*. And we have acute toxicity for *Ampelisca abdita*. In terms of bioaccumulation tests, *Macoma* is probably a fine species to use. Dr. Chapman brought up the point before: Sometimes we use the approved species not because they are the best ones, but because they were the convenient species around at the time these tests got going. With *Ampelisca* I would say we are in the same boat, however, the species is not easily cultured. Everyone appears to get his or her animals from the same supplier in Rhode Island. And the shipments from the supplier are variable in terms of animal condition. Sometimes we cannot use a shipment because the animals are sick and dying. Other times, we start a test and three days later, all of the animals are dead. We have wasted a lot of time and money. *Ampelisca* is not a very hardy species. Other research scientists are using *Leptocheirus* more and more in this part of the country. I am not saying *Leptocheirus* is the best [species to use], but an animal like *Leptocheirus* allows you to control its nutritional status, its size, and its health to get more reproducibility in the tests. This certainly makes sense. So that is the acute toxicity side. From the bioaccumulation side, I said *Macoma* seems pretty good. The nice thing about bivalves is that they tend to be very poor metabolizers of a number of hydrocarbons, especially aromatic hydrocarbons. So it is nice to have a bivalve. But one of the things that we are learning for more strongly absorbed contaminants is that there are large differences among animals in assimilation ability. And we have not done a very good job at evaluating interspecies differences yet. And my guess from Heisenburg's Uncertainty Principle, is that we want is a very small head-down deposit feeder that has a lot of surfactant in its gut. It is very small and does not perturb the system, but displays maximal accumulation of strongly absorbed hydrophobic contaminants. I have no idea about metals. But there are better species than *N. virens* to use. My last point is that by simply relying on an acute toxicity endpoint, we have no way of knowing whether a chemical is 90 percent of the concentration [or some other fraction] of a no-effects concentration. We simply do not know how close we are to acute toxicity when something passes an acute toxicity test. And if we want to do ecological risk assessment, we must do something that is a little bit more relevant to the population. Perhaps it is a reproductive test that we need to work on and agree upon. If we want to look at ecological endpoints, we need something other than an acute toxicity test.

MS. WATERS: Audience?

DR. DRISCOLL: The approach that we have now is very conservative. If there is a significant difference in bioaccumulation or toxicity between your site sediments and your reference sediments, then you failed the test. The rule of thumb is that, on average, the difference between the two is 20 percent, then that typically constitutes a "significant" difference. And people will want to say, "well what does that really mean?" Does a 20 percent difference mean anything at the population level? We do not have an answer to that. New York [ACOE] is using the concentration that kills 50 percent of the test organisms as a thresh-hold probably severe enough to have population level effects. So now for new HARS material, they are using the LD50 for certain contaminants to say that a 50 percent level is the thresh hold for concern. I think we need to move beyond just saying "it is significantly different", to looking more towards population level effects, though it is going to be difficult to get there. It will probably be another rule of thumb or another line that someone draws, either at the LD50 or someplace else, I do not know. But there certainly is a lot of reason to move beyond just the "significantly different" cutoff.

MR. DOUGLAS: There are [at least] two types of endpoints. So far, we have discussed endpoints that we generally measure in a toxicity test. When we were talking about the definition of "toxicity", [earlier] we did not talk about the definition of the "toxicity test." I suggest that a toxicity test is a model for predicting what will occur in the environment. And it is not very useful. We have not talked much about are endpoints in the environment. I suggest that when we are doing a toxicity test, we need to have a very reproducible endpoint, something that we can accurately, predictably and reliably measure. However, when we are talking about extrapolating test results to the environment and trying to measure the success or failure with our management decisions, we need to have another kind of endpoint. And these endpoints need to be much more sensitive.

DR. SOLOMON: What you are trying to protect and what you are measuring, are often two different things because there are practical considerations. You may be trying to protect a population, but it is much easier to do laboratory tests or bring in sediment samples and see what is happening [than try to look at the population directly]. But the other issue is that you can choose statistically different responses between reference or controls and use them to test situations. The fact that there may be a difference does not necessarily mean that the outcome is bad. The "badness" is a human judgement. We do a value judgement. The Great Lakes went from pristine conditions, say 500 years ago, to heavily polluted with phosphorous. And now, they are heavily "polluted" with organisms, "foreign organisms": zebra mussels, sea lamprey and Eurasian gobies. Each of those states are stable in a sense. Each of them is different from the other. But whether they are bad or good is a human value judgment. Are the Great Lakes now bad because the water is so clear we can see the bottom in Lake Erie all over the lake just because of the zebra mussel? Nature has no value judgements, only humans do. The difficulty is to sort it all out: if I go and dump a highly flocculent, organic rich sediment over sand, I end up with a different environment, and I will get different organisms inhabiting that environment. But is that necessarily bad? That is the value judgement.

MS. WATERS: Does anyone else want to address Mr. Douglas' question? Do you need two different kinds of endpoints? One is the endpoint from specific tests of some kind. The other is the longer-term endpoint you are looking for in, perhaps, ecological settings or ecological terms. Are those different? Do we need to define both?

DR. SOLOMON: They are different. This is addressed in the guidance for ecological risk assessment. It has come out of the Risk Forum at EPA.

DR. O'CONNOR: I think this is the point: among these first five [challenge] questions we do not seem to get beyond characterizing the material. And if that is true, then we have not even begun to do risk assessment. A risk assessment requires more than doing toxicity tests on the material. It requires more than doing chemistry on the material. If that is all we do, we have not really done much. The questions are okay, but there should be a Question A. Before we get into this, what is at risk? Who are we trying to protect when we are disposing of dredged materials? Until we have defined the problem, the rest of this does not mean very much. We are still going to use those [existing] rules that we have to do it.

MS. WATERS: Let us take a few minutes to talk about what is at risk and why we should be paying attention to the rest of these questions when we are disposing of dredged material? Is that what you would like to talk about for a minute?

DR. O'CONNOR: That would be fine with me.

DR. CHAPMAN: Well, several potential concerns. The first is the physical effect of dredging and disposal. The second is the possibility that chemical contamination can lead to toxicity that may cause pollution. The magnitude that has been alluded to is a third concern. Are we talking about only a small area? If so, does it matter? What is the significance [of this magnitude] relative to other things that we might be doing, could be doing, should be doing? But the issues, as I see it, are habitat and contamination issues.

DR. BROWNAWELL: But what tends to drive things at the end, are human health concerns more often than not.

DR. CHAPMAN: Yes.

DR. BROWNAWELL: Persistent, bioaccumulated substances are certainly the big issue at the HARS site, especially for PCBs. This is what drives the criteria for disposal. More often than not, the issues are human health. Because it is so hard to grapple with, the ecosystem does not seem to have a strong constituency. The matter is also pretty fuzzy in terms of what we are trying to protect; is it ecosystem function or is it an endangered species? It is very difficult for either the citizenry or the legislature or the scientists to get their arms around who/what it is that we are trying to protect.

MS. WATERS: Does anyone wish to comment further on a risk assessment approach for managing dredged materials?

DR. DITORO: Yes. Perhaps the way to about it is to look at the current situation, based essentially on sediment testing. Statistically "significant", by the way, is another one of these words that has scientific meaning and also a [separate] meaning in the body politic. When scientists use the term "statistically significant", it does not necessarily mean "important". It simply means that there is a level of statistical confidence that you have in the result. So I do a test. Has it ever occurred to you why five replicates were chosen between the controls and the test sediments?

AUDIENCE MEMBER: No.

DR. DITORO: It has to do with detection, whether or not you can detect a [significant] difference or not. If I chose a thousand controls and a thousand samples of dredged material to compare, I would never discharge anything because the power [sensitivity] of the tests go up as the square root of the number of tests, and the results would always differ significantly. But there is no underlying reason for having done that [choosing a particular number of replicates]. Five is simply a compromise. As a consequence, if someone looks at the way dredged materials are currently evaluated, they might wonder that there must be a better way of doing it. For example, as Dr. O'Connor has said and others have said earlier, there is no difference between a teaspoonful and a hundred million cubic yards of disposal material. The criteria [for disposal] are exactly the same: test five samples. If there is a statistical difference, do not dispose of these sediments. If there is no statistical difference, you can discharge. It makes no sense. In fact, many people who work in the environmental business regard the regulations for dredged materials disposal as being atypical, they do not look anything like, for example, the regulatory framework for waste load allocation, nutrient management and pesticide control. In these frameworks, the *amount* matters. In dredged material, the amount does not matter. Strange! As a consequence, the risk assessment paradigm will hopefully provide a logical framework for managing sediments and addressing the question: Should I be worried about discharging a teaspoonful or a hundred million cubic yards of contaminated material here, there or somewhere else? That is the name of the game. That is why we are gathered here. And I think we should try and answer the questions in that regard. So are the endpoints that we have for toxicity tests sufficient to apply risk assessment paradigms to managing dredged materials? I would suggest that as we go through these questions, we put a punctuation at the end of them: For the use of risk assessment for evaluating the suitability of discharging dredged material. Maybe that is what we are missing.

DR. O'CONNOR: Thank you.

DR. DRISCOLL: A big questions in moving to a risk assessment approach is: are we really going to start making decisions on a population basis? I think we are currently regulating on the basis of individuals. This is certainly a conservative approach. And some people have argued that that is an appropriate place to be because of the implications of cumulative risks. The "death of a 1000 cuts", as some people like to say. But if we move towards protection of populations there are new questions; e.g., how do you define a population of arthropods.

DR. WEINSTEIN: I think I hear two things in that, Dr. Driscoll. One is the point that was brought up earlier about the test species not being the assessment endpoint. The other is: Do the endpoints that we measure, i.e., the "measurement endpoints" in bioassays and toxicity tests relate in any predictable way to ecosystem responses? With respect to the selection of "the assessment endpoints", the jury is still out. I appreciate the nod to the population modeling that has been underway for several years, but I think that it is a little bit misplaced still because it is still addressing population dynamics of *test* species, not of assessment endpoints in the receiving ecosystem. And if we consider Dr. O'Connor's comment that what we really need to be looking at are risks in the receiving environment, then at a minimum we need to make those statistical linkages between what we are measuring and ecosystem responses. At best, we need some sort of a process-driven, mechanistic understanding of what those linkages are.

MS. WATERS: Most people seem to be saying we are less concerned with the individual than with the population. You will be looking, as you can, at evaluating effects on populations and, the larger environment, rather than an individual fish.

DR. CHAPMAN: Yes and no. I will put forward an argument that the Inland Testing Manual and the Green Book, to some extent, actually inhibit risk assessment. If you look at the wording in Tier 4 (where risk assessment comes in) it says those are for unusual, special circumstances. It is not for every circumstance. My interpretation is that you are encouraged to do more in terms of hazard than in terms of risk and to make decisions based on chemistry and on toxicity testing

without getting out of the box and seeing what is really occurring. I would argue that strongly. And I would be interested to see anybody refute that. Until the regulations change, we will not be going to population [level analysis]. There is no impetus to do that.
DR. BERRY: Walter Berry, USEPA. Following up on Dr. Chapman's comments. I wonder, if to a very real extent, we are not asking the wrong questions. I think it is generally valid that the people in this room want to move towards a risk assessment approach to the management of dredged material. That is what we have seen in several of the presentations.
MS. WATERS: That is what I have been told.
DR. BERRY: If we talked to everyone who was here [at this conference]-- unfortunately, including many people who are not in this room, I suggest that there would not be consensus [that risk assessment was] the way to go for two reasons. One is because the regulated community has a perception that risk assessment is incredibly expensive, that it is site-specific, and that it is a really long process. They feel that the current testing regimen that they go through now is already prohibitively expensive. It takes a long time. And the answers are pulled out of a black box. But if we changed the regulations, which many people feel is impossible to do, and went to a risk assessment approach, that it would take even longer. It would be much more expensive. And it would be even more difficult to interpret. That is on the regulated side. There are also some people on the environmental side who have a mistrust of risk assessment and feel that a risk assessment can be made to say anything that we want it to say, and that it can be misused and become a license to dump. So both of those people need to somehow be convinced and shown that, in fact, risk assessment is a way to go. That we need to find a way to do risk assessment based on disposal sites that is not so case-specific for each dredged material. That it now takes years and a tremendous amount of resources. Somehow we must convince the environmental community, and others, that we can find a fair way to do risk assessment but not as an excuse to dump or as an excuse to ratchet down so low that nothing can be disposed of.
MS. WATERS: Dr. Berry, do you think that this is possible to do? You would be satisfying two groups with very different concerns, if it could be done. One that is concerned that it will cost too much and be too detailed and too site-specific; and the other that it will not be detailed and site-specific enough!
DR. BERRY: Yes, I think so. It seems to me that the sort of framework that Dr. Munns proposed earlier [in his formal paper presentation], is useful for making decisions on where to dispose of something. You take everything into account that you can think of, work through the process, and then identify suitable places for it [dredged material]to go. Taking all sorts of things into consideration is a no-brainer. But what you need to do is to make people understand *how* that will be done, and you have to develop some kind of atmosphere of trust. I am from the government and I am here to help you, so I believe that it can happen.
MS. WATERS: Well, let me turn to the panel and see whether this is in fact the challenge; i.e., to find a way to do this? Are we going to be able to get there?
DR. SOLOMON: I think the last speaker [Dr. Berry] brought up some good ideas. Nonetheless, look at the simple toxicity test. If you fail it, all that means is you failed a toxicity test. It does not mean you necessarily have a problem. If you passed it, the likelihood is that there is not going to be a problem with what you are doing. So those situations where you pass, you are okay; but if you fail, then you need to go to something more realistic where you consider the area involved or the volumes involved, or as Dr. DiToro was saying just now, the quantities involved. That level of risk assessment, hopefully, would not need to be applied in every situation. The appropriate approach would focus the costs where the costs surely should be focused. And those are the

areas where the risk is greatest. Another approach is to go forward with simple testing endpoints, and use the results of those tests to say, yes, we have a problem. We are going to stop whatever we are doing and we are going to do something else. But we need to recognize there may be ecological costs and human health costs with doing that something else. And there are countervailing risks that we need to consider all the time as well and costs associated with over regulation. The benefits of dredging a canal for better navigation, in terms of society, are reasonably large. We need to weigh that against the "disbenefits" of putting that material somewhere else.

MS. WATERS: Anyone else wish to comment?

DR. MUNNS: Your comments are a little unfair, I think, Dr. Berry, in the sense that they have elicited this response [from previous speaker(s)]. One thing I did not mention in the earlier talk was that the current evaluation procedure is explicitly tiered. And if there is insufficient information for any step in that tier, you can kick it out of the process and make a go or no-go kind of decision. When I think of tiering and risk assessment, I think of them very differently. I did not get a chance to explain that in that last flow diagram. But you can tier the assessment itself [in any one of those assessment boxes] in a way that you keep going until the uncertainties associated with the assessment are still too large to alter the cost of making the wrong decision. If that condition exists, you want to go through the loop of risk assessment again. So any time that you are certain enough of the answer to make a decision, you stop; whatever that decision is. I am not sure I explained that well enough, but I mean that is a fundamentally different way of tiering than the current evaluation process.

MS. WATERS: Sort of iterative risk assessment? You do it to a certain stage. If you are comfortable, you make your decision. And if you are not, you go to the next level.

DR. MUNNS: That might be one way to summarize it. Another is that the assessments are done in the context of the receiving ecosystem. You do that [the assessments] for all possible combinations of receiving ecosystems. That starts pushing against the issue of cost. To balance that, we may have to put more structure into the process, i.e., make rules to say that you do not need to look at everything in the universe. You can then apply rules to narrow down the process to much more manageable levels and then apply tiering to minimize costs as well. To answer the question you asked of Dr. Berry, I too think that is doable. I think that kind of approach is doable. That does not mean we can do it today, but I think that conceptually it is doable. And the research approaches to get there are starting to be laid out.

DR. BROWNAWELL: I would like to follow up on Dr. Berry's comment. He talked about two kinds of skeptics to risk assessments: those who are worried that it is going to delay and add cost to the process; those who think it is a license to dump. I would consider myself a third type of skeptic. I would like to get some feedback from the panel and the audience. I am one of those people that trusts things that they do not know very much about. So, I have full confidence that Dr. Munns and Dr. Solomon can model the risks associated with contaminants. But the question I have for Dr. Chapman and others is: Do we have the state of knowledge; i.e., where do you think we are at and how far do we need to go before we have confidence in the risk assessments? We talked about the need for extrapolating our toxicity endpoints or laboratory testing endpoints to population effects. Where are we on that? Is this a tractable problem? And two, is it a tractable problem to validate some of these ideas in terms of monitoring success or expectations from dredging operations? I do not have a solid appreciation of whether we are at a state of the science where we can do a risk assessment that is going to be supported by monitoring and research.

MS. WATERS: Panel?

DR. CHAPMAN: In terms of population levels there has certainly been some useful modeling work that is very promising. And certainly, if we go back to basic ecology, there are some great tools that we can borrow and use to move things forward a great deal. A good example is the use of ordination. Ordination had been used for about 50 years in the plant ecology community before other groups adopted it. There are lots of good examples, but they are in different disciplines and we are not taking advantage of them. I think we can move forward very quickly if we are willing to adopt some of these techniques. I think the major impediment is answering the questions, focusing on what is out there and what needs to be done. I will cite a real example. In 1990, the Green Book came out. I was on the science advisory board that reviewed it. I had all sorts of thoughts; so I went into it and wrote what I thought was appropriate and different. And they looked and said, well, that is great, but you cannot do it this way. And over four years, it went back and forth, and now it is looking very much like the original draft. The big problem we face is ourselves. Remember Pogo, we have met the enemy and he is us! We have all these great ideas and we develop all these scientific tools, but, unless we can change the glacial way in which things are looked at, and make these major leaps forward, nothing will happen except inside the journals. And that is the real problem.

MS. WATERS: The door has been opened by Dr. Chapman saying that there is scientific information out there that is not getting used. Comments?

MR. LINNAN: My name is Paul Linnan, and I work for the Pennsylvania Department of Environmental Protection. I am a public administrator. I am not a scientist. I am going to state things in a down-to-earth way. The reason that I am here is that I am the project manager for a mine site in Pennsylvania that you have heard reference to from time to time at this conference. I deal with people on a community level, out in the Commonwealth of Pennsylvania. What I think folks in your position need to remember – perhaps sometimes we lose sight of -- is when people out in the country hear the word "toxic," [when referring to dredged materials] they think: "This stuff kills people -- or kills things." If we leave it around here long enough, it is either going to kill us or we are going to have children in future generations with birth defects. I mean that is what the word "toxic" means to them. What I feel we need to do -- perhaps not an endpoint, but as a desirable goal -- is to be able to tell people, through science [and I realize these are social decisions as well] that we have examined everything we can about the issue, and here is how it affects humans, and here is how it affects other plants and animals. And as Dr. Solomon suggested, there are going to be tradeoffs. Is it more important to have shipping lanes than it is to have harbor bottoms that are no longer disturbed by our dredging? You may find people that come in on both sides of that issue. I suspect when push comes to shove that people will realize it is more important to have shipping because that keeps us all alive. And when it gets down to it, we are all predators, and we are going to do what it takes to keep ourselves and our families and our society alive. But why I bring this to the forefront now is because I came to hear how I can explain the risks to the community associated with what we are doing. I have learned a lot at this conference. But when we are talking about toxicity or populations, people are immediately going to think about humans and not populations of amphipods in the ocean. How do we explain to the [human] population, in general, that what we are doing is either going to hurt them or it is not? I intend to buy the tape we are making of this. And I intend to show it to as many people as I can.

MS. WATERS: We are trying figure out where we are in the process of being able to do just that with regard to the disposal of dredged material. How close is risk assessment to being a viable tool for managers with a project like that [risks associated with disposal of dredged materials in abandoned mines]? Dr. Chapman?

DR. CHAPMAN: I think the tools are there and can be used. I have certainly been involved in these cases; e.g., one involving a sewage outfall in the City of Victoria. It is something that you may be aware of. It still causes heartache to a lot of people. The City of Victoria discharges untreated sewerage into the Strait of Georgia. The Strait of Georgia is across from the United States. That has caused a lot of issues. We found that there was, indeed, toxicity around the outfall. But it was limited to the size of about a football field. We also found low-level effects on reproduction and other [chronic] effects. We found some contamination which seemed to be due to mercury which we could isolate as being largely coming from dentists and could be obviated. 1,4-Dichlorobenzene was detected in the "hockey pucks" from urinals. And all this information was put before the public at public meetings, and "toxic" was explained and so on. And the public decided that at this point in time, they were not going to spend the money on sewage treatment, but they would rather spend the money improving the hospitals and doing other things to improve human health. I am not saying it is a right or wrong decision, but this was the general public that made it [by referendum]. This was decided by hundreds of thousands of people. They made the decision. A lot of the country does not like it, but they have made their decision. So it can be done.

MR. LECHICH: I am involved with a work group that is looking at revising the bioaccumulation values, or approaches to bioccumulation for the harbor [New York / New Jersey}. The working group is using a risk-based approach. One of the things that we are going to find most difficult to deal with is dealing with uncertainties in the risk-based approach. I will put out a very simple question: How do you deal with uncertainties? But to make it fair, in terms of how you weigh them in establishing a protective [conservative] endpoint versus a "liberal" endpoint.

MS. WATERS: Uncertainty was an element that several of you identified this morning; i.e., how do we handle uncertainty?

DR. DITORO: I can make a generic response: the amount of uncertainty tolerated is inversely proportional to the amount of resources you have to spend on the problem. If it is an important problem with many potential downsides, you can do much more to reduce uncertainty, in relative terms, than if you are time limited and/or if you have very limited resources. The difficulty about uncertainty is not knowing about what you do not know. If you know you do not know and you know things that you can do to help clarify things, then that is good. You can expend resources and do clever things and hopefully reduce the uncertainty. The really hard part of addressing uncertainty is where you do not know you do not know. In which case, you can not do much more than go out into the field and see if you can make measurements which somehow bear relationships to the predictions you are trying to make. The HARS issue is an interesting one. The HARS is the old Mud DumpSite for dredged materials in New York. The question there is: What value should the bioaccumulation tests be set at in order to make sure that the fish are not overly contaminated? The question that occurred to me when I heard about it is this: Has anyone actually measured concentrations in fish living in the HARS to see whether or not the model predictions mean anything? And as far as I can tell, the answer is no. So simple things like going out and measuring the effects that you think are happening can make a lot of sense. The hard part about uncertainty is when you are forecasting into the future or you are forecasting into situations where you do not have any data. That is where it really gets difficult. But if you are trying to predict a situation or evaluate a situation where you already have a prototype, then the sensible thing to do is to make some measurements. So my answer to whoever asked the question about uncertainty is, go out and measure the "bejezus" out of it.

DR. CHAPMAN: The other aspect of the uncertainty question is: how much uncertainty can you tolerate and where do you stand on the side of being overprotective versus under protective? That is a critical issue. I think the only way you can answer that is by looking at the potential downside. Take the example of global climate change. I know little about it, but as a member of the general public, it scares the heck out me. I do not think we should tolerate any uncertainty. We should try to reduce that because the potential downside is an incredible change to the earth and everything on it, including me and mine. So that is a big downside. Translate that to dredging. If you have got a downside where you could potentially wipe out a rare and endangered species, you are not going to tolerate a lot of uncertainty. You want to be very protective. If, however, the downside is much less critical, for instance some but not most fish in a population bioaccumulate one or more contaminants above levels [body burdens] where there is a low level of risk to people eating lots of those fish, how protective do you need to be? This is not a clear-cut situation. If you apply the Precautionary Principle, perhaps you are still highly protective. But you cannot be highly protective in all situations. You have to take into account the relative scale and level of risk. We have neither the time nor the resources to address all issues. We must prioritize. And again, I repeat from my talk, I have seen too many cases where we are spending too much money on things that do not matter at all compared to things that do matter going by the wayside.

MS. WATERS: With the time remaining, let us ask each member of the panel where you think we are with risk assessment tools for addressing issues with dredged materials management. What do you put at the top of the list for where we go next and what we need to focus our attention on?

DR. HO: When people gave their talks, most listed research needs and I do not think we need to reiterate those here, but we do need more basic understanding of mechanisms, both on the biological and the chemical side. The issues of pore-water testing and how we use the information from those tests as well as whole sediment tests in risk assessments for dredged materials and sediments are important. Also, the talks on risk assessment and risk assessment models, identified the general lack of model validation; going back out to the field to verify if the models that people are proposing to use actually work. How does field validation reflect upon the models' utility? What needs to be changed in the model? Those are very basic things that I think we need.

MS. WATERS: Good.

DR. MUNNS: Probably our biggest problem is attaining the management and societal shifts we need, in the way we view the problem. The tools, at least conceptually, are there. That is not to say there is not a lot of research that is still needed. I am a firm believer that for the scale of the problem that we are dealing with, there will be technological solutions in the near future that we can use to address those problems. But I am not as convinced that we can create the shift in approach that many us [automatically] assume while we talk about research needs and paradigms for conducting risk assessments for dredge materials. The likely area most in need of attention is attaining the paradigm shift [in public perception]needed to address this issue.

MS. WATERS: Thank you.

DR. BROWNAWELL: I will summarize some research priorities. One is that it is difficult to see what the next steps in sediment toxicity assessment will be until we know what is causing some of the sediment toxicity that we see. I think that is a tractable, but difficult problem.

Further, it would be useful, if we want to do valid, long-term risk assessments, to look at issues associated with the dynamics of sediment contaminant interactions. What is not available today might be available tomorrow in terms of long-term risk. When you transport something or you store something, that is also a tractable problem. The problem I am sure that is tractable, is the big issue that Drs. Chapman and Ho have talked about. And that is either validating risk assessment predictions or doing more fundamental studies to extrapolate toxicity endpoints to population endpoints. Finally, it is not so much the scientific community, nor the press, nor regulatory agencies that we need to work on. We need to work much harder in terms of conveying risks to the *public*. I have mentioned before that we probably spend a billion dollars on "no detects" intentionally, because we do not know how to convey detects to the public in terms of "toxic chemicals". The fact that the EPA and other regulatory agencies have "toxics" programs is probably not a good choice of words when we talk about these issues to the public. It draws a line, in terms of, "is there a good toxic"? So to me, that is a big societal challenge in the future. As we are able to measure biologically active substances at increasingly lower [trace] levels the key question becomes: what do we do with these more sensitive measurements? The press always talks about the numbers of detects. The management and regulatory communities intentionally avoid detects for good reason. I think we are going to have a "crossing of the paths" that we are going to have to deal with. There is a precedent, though. We deal with cancer risk and radiation all the time and other types of hazardous assessments. It is those things with "toxic chemicals" connotation that the public is not going to deal with very well.

MS. WATERS: Thank you Dr. Brownawell.

DR. SOLOMON: It would be very helpful if we could integrate both temporal and spatial variability into these issues of [dredged materials] disposal or risk assessments related to dredged materials. I think we have some ideas as to how that might be done. I agree with Dr. Ho that we need to validate or calibrate some of these risk assessment models. That would be a reasonably good starting point. It means collecting some additional data. It means trying to deal with those data, to describe variability and then calibrate that against what we see out in real-world situations. Using experimental systems, whether they be *in situ* or microcosm-type would have some advantages, particularly as you deal with things on a temporal scale. You recognize that sediments have high binding constants and slow kinetics of movements and you recognize the need for extended time frames. You are not dealing with rapid, within-season events, as we have seen traditionally with other chemicals like pesticides. We are dealing with much longer time scales.

MS. WATERS: Thank you.

DR. DITORO: I think the present regulatory system is clearly broken. The level of protection that is afforded by the regulatory procedures are unknown. One of the characteristics that suggest it are that the magnitude of the dredging project, i.e., the amount of material that is going to be disposed has no effect on the decision. Therefore, very large projects, that in a risk assessment framework, you would look at much more carefully, are afforded the same look as a tiny project suggesting that it may be overprotective in one case and under protective in another. Another problem is that the present regulatory structure [for managing dredged material] does not bear any relationship to other regulatory programs in the nation. TMDLs, the way we deal with effluents, the way we handle pesticides and all the rest are reasonably well-understood regulatory frameworks where we are allowed to discharge materials into sanctuaries and streams and reservoirs under regulated situations. The situation with open-water disposal of dredged materials is politically highly charged. As a consequence, there is less flexibility in the political process. So, the major problem is not on the side of risk assessment frameworks that make sense, rather the major problem is to convince the regulatory authorities to change their behavior. And if we can get past that, I would say that we have the technical tools to do a better job than we are currently doing are there.

MS. WATERS: Thank you.

DR. CHAPMAN: I would say the big challenge from a scientific point of view, and also from a societal point of view, is to change the emphasis. Right now, we emphasize the statistical or regulatory significance. That is the basis for the Green Book. Is it statistically different from what the regulations allow? What happened to ecological significance? For goodness sake, if we cannot bring that back, we have no place to go. And we are going to continue spending money on trivial issues and driving everyone mad, and lawyers are going to get rich[er]. I think that is the major issue we need to focus on. I was sitting beside a lady at lunch, who is not in this session, and she was telling me how she cannot exceed one ERL for mercury. And that this is stalling a whole project. One ERL! That is crazy. As a public, we are confusing hazard with risk. And I think that is something that has been brought up, at least from the terminology perspective. Risk has a probability; hazard does not. Hazard has a possibility. Everything has a possibility of happening. I think too often we are looking at possibilities instead of probabilities. That does not make much sense. My final comment. I think there is a real disconnect between scientists who stay scientists and scientists who become regulators. Because scientists who stay scientists continue their training to take risks, to see if we are going to be wrong. We learn by being wrong. Regulators cannot afford to be wrong. But I think they should [be able to be wrong]. Not on things that matter [determined by consensus], but certainly in those cases where the significance of making a wrong decision is not going to be the end of the universe. There should be a chance to try something different, try something that might be better, take a risk that might make things better. So when it comes to big issues, we can deal with those more appropriately. As long as scientists are risk takers and regulators are not, you are going to have this enormous disconnect. And we are never going to have the science catch up to the regulations. It is always going to lag far behind. Those are my two issues.

MS. WATERS: Dr. Weinstein?

DR. WEINSTEIN: I would like to thank our superb panel, and the audience for a constructive dialogue. I know you are all worn out. I hope that we will be able to extract out the good stuff from this discussion for the published proceedings. I am also hoping that there is going to be a next step. Where do we go from here with regards to what we did these three days? So thank you very, very much. Ms. Waters, thanks for a terrific job. And the folks who did the taping and the monitoring and the typing, thank you all very, very much.

(End of panel discussion.)

Technical Papers

(Reprinted with permission from *Marine Pollution Bulletin)*

Integrating Toxicology and Ecology: Putting the "Eco" Into Ecotoxicology

PETER M. CHAPMAN
EVS Environment Consultants, 195 Pemberton Avenue, North Vancouver, BC, Canada V7P 2R4, pchapman@attglobal.net

Abstract

Environmental toxicology has been and continues to be an important discipline (e.g., single-species testing for screening purposes). However, ecological toxicology (ecotoxicology – more realism in tests, test species and exposures) is required for predicting real world effects and for site-specific assessments. Ecotoxicology and ecology have shown similar developmental patterns over time; closer cooperation between ecologists and toxicologists would benefit both disciplines. Ecology can be incorporated into toxicology either extrinsically (separately, e.g., providing information on preselected test species) or intrinsically (e.g., as part of test species selection) - the latter is preferable. General guidelines for acute and chronic testing and criteria for species selection differ for ecotoxicology and environmental toxicology, and are outlined. An overall framework is proposed based on ecological risk assessment, for combining ecology and toxicology (environmental and ecological) for decision-making. Increased emphasis on ecotoxicology represents a shift from reductionist to holistic approaches.

Keywords: ecology, ecotoxicology, sediments, risk assessment.

Introduction

Toxicological studies of the environment can be mostly characterized as environmental toxicology. Such studies are conducted independently from ecological considerations, and perhaps subsequently compared to ecological studies in a burden-of-evidence approach (e.g., Ingersoll et al., 1997). Consideration of ecology is generally extrinsic rather than intrinsic. In other words, tests are, in many cases, conducted with organisms that can readily be obtained, cultured, and tested. The ecological significance of the test organisms is a secondary consideration. Thus, for example, freshwater rainbow trout toxicity tests are used in Canada even for effluents discharging into marine waters.

Arguably a paradigm shift is occurring, with ecological toxicology (ecotoxicology) assuming increasing importance. The purpose of this paper is to detail the importance of ecotoxicology, for decision-making related to ecosystem protection, and encourage this paradigm shift. This paper begins by discussing the status and progress of ecology relative to aquatic toxicology in general, then proceeds to discuss the differences between ecotoxicology and environmental toxicology, key ecotoxicological issues, a specific example (estuarine sediments), and ecological risk assessment (ERA). It finishes by providing specific recommendations for the better integration of ecology and aquatic toxicology.

Ecology – Progress and Comparison to Aquatic Toxicology

Ecology focuses on interactions between organisms, distributions and abundance of organisms, the functioning of biological populations and communities, and the processes that affect all these parameters (Andrewartha and Birch, 1954). Ecologists study interactions between organisms and their environment at all levels, from the individual organism through to the ecosystem. This includes factors governing the geographic distributions of species and that influence abundance and other characteristics of individual populations. The primary purpose of ecological investigations is "to *understand* and *explain* natural phenomena, ecological processes and, therefore, the resultant patterns of distribution, abundance, diversity and interactions of species" (Underwood et al., 2000).

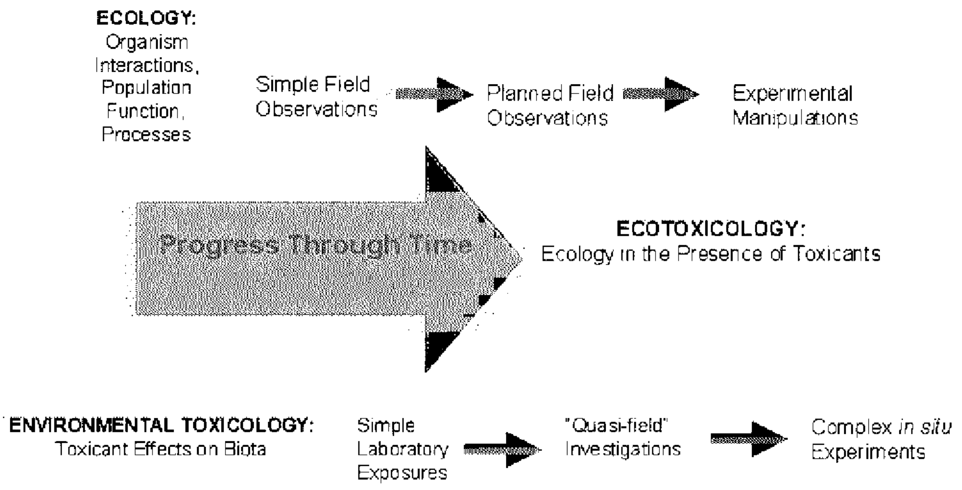


Fig. 1. The development of ecology, environmental toxicology, and ecotoxicology. Initial approaches have not been supplanted but rather have been complemented by subsequent approaches.

Ecology began with simple observations (natural history and description), that were then complemented by planned investigations, and later by experimental manipulations; this mirrors the progress of toxicology (i.e., from only simple laboratory exposures to complementary complex *in situ* experiments – Figure 1). A major focus of ecology is the general principles that structure natural communities (Menge, 2000). Manipulative experiments in ecology data back at least 70 years (e.g., Hatton, 1932). Experimental aquatic ecology, particularly in the marine environment, has blossomed in the past few decades to become a solid scientific discipline (Castilla, 2000; Underwood, 2000). Classic experiments have been conducted to test hypotheses concerning, for example: competition, predation, succession, perturbation, resilience, and species richness. In addition to manipulative experiments, the range of study approaches used by ecologists includes descriptive observations, laboratory experiments, and mathematical models.

Major ecological paradigms have been developed including the controversial (Power and Mills,1995) keystone species concept (Mills et al., 1993); community resilience/ecologically alternative states (Sutherland, 1974); and the influence of biotic factors (carnivores and herbivores) on patterns of biomass (Polis, 1999). The keystone species concept, if correct, may be particularly germane for aquatic ecotoxicology as discussed later. A key-stone species is simply a species whose impact on its community or ecosystem is disproportionately large relative to its abundance (Power et al., 1996), thus its loss from or addition to a system would change community composition, structure or function sufficiently to arouse concern (Power and Mills, 1995). Keystone species differ from species that are dominant in terms of biomass or abundance, which latter are critical for the maintenance of the structure and dynamics of communities. However, keystone species may not exist in all environments. For instance, in wetland plant communities the relative importance of a species to community structure and function is strongly correlated with the species' overall abundance, and there is a great deal of functional redundancy within guilds of wetland plant species.

There are also areas where both ecology and aquatic toxicology have similarities in terms of deficiencies. For instance, there has been little experimental work on ecology related to detrital webs (Castilla, 2000), which mirrors a similar lack of progress with regards to bacterial bioassays in aquatic toxicology. Similarly, many ecological experiments are designed to minimize (or ignore) rather than measure natural variability (Chapman, 2000a), which again is a similar situation to aquatic toxicology (Baird et al., 1996).

Another similarity between ecology and aquatic toxicology relates to the use of "quasi-field" and field experiments. Both have attempted to make laboratory conditions more realistic. For instance, aquatic toxicologists use microcosms and mesocosms (Solomon, 2001); ecologists similarly transfer animals into the laboratory with patches of natural habitat (e.g., Della Santina and Naylor, 1994), take the laboratory into the field (e.g., Colombini et al., 1994) or conduct transplantation experiments (Underwood, 2000). Arguments regarding top-down as opposed to bottom-up assessments are as rife in ecology as in aquatic toxicology (Baird et al., 1996; Menge, 2000; Underwood, 2000), but do not necessarily involve a good understanding of each others' disciplines.

There are also areas where ecology and aquatic toxicology should be more similar but are not. One of these is the issue of functional redundancy. In other words, when and under what circumstances do communities contain functionally-analogous species, such that the disappearance of one species from the community entails no measurable loss of functionality (Tilman, 1997)? As noted by Duarte (2000), functional redundancy is a property of the particular species present rather than the number of species. Thus, *Trichodesmium* species, the main pelagic nitrogen fixer in the ocean, plays a key role that cannot be assumed by other species within the same community. In contrast, functional redundancy does occur between closely-related species such as seagrass communities and is best determined related to specific functions rather than, for example, gross morphology (Padilla and Allen, 2000). Any relationship between species richness and functional redundancy is likely indirect (Duarte, 2000). Risk assessments are beginning to consider the issue of functional redundancy, however this is not considered in environmental toxicology as part of single-species tests conducted under laboratory conditions (Solomon, 2001). Calow (1996) has raised this issue of structural redundancy with the recommendation that "protecting structure should, in general, protect function – this is a kind of ecological precautionary principle." But it has not yet been recognized as a key issue in aquatic toxicology. Similarly, population dynamics is one of the most important branches of ecology and one of the most relevant to ecotoxicology.

Environmental Toxicology and Ecotoxicology

The domain of toxicology in general (including both environmental toxicology and ecotoxicology as defined below), includes understanding the types of effects caused by chemicals, the biochemical and physiological processes responsible for those effects, the relative sensitivities of different types of organisms to chemical exposures, and the relative toxicities of different chemicals and chemical classes. While controlled laboratory experiments using single "indicator" species have served well in the past and continue to provide a mainstay for toxicology (e.g., screening large numbers of substances and environmental media to identify those that may be hazardous), more complex studies and better choice of test species are essential complements for present and future studies if we are to predict toxicity to wild organisms under actual exposure conditions.

Ecotoxicology comprises the integration of ecology and toxicology (Chapman, 1995; Baird et al., 1996; Figure 1). Its objective is to understand and predict effects of chemicals on natural communities under realistic exposure conditions. Theoretical insights and methods drawn from ecology are needed to achieve this objective. Very real differences exist compared to environmental toxicology as summarized in Table 1.

Table 1. Environmental Toxicology Compared to Ecological Toxicology (Ecotoxicology).

Environmental Toxicology	**Ecological Toxicology**
Laboratory issues primary (e.g., collection, culturing, holding, testing)	Ecological issues primary (e.g., importance in: food chain, community structuring)
Individual species tests	Combined species tests
Cost of testing a major issue	Cost of an incorrect decision the major issue
Simple tests	Complex tests
Chemicals of primary concern	Chemicals only one issue, and not necessarily the most important issue
Toxicologists only	Toxicologists and ecologists, and other disciplines as necessary (e.g., microbiologists).

For the latter, laboratory issues are primary (e.g., collection, culturing, holding, testing) rather than ecological issues (e.g., importance in the food chain / community structuring and function). Environmental toxicologists generally test individual species rather than combined species. Testing mixtures of species can result in reduced toxicity compared to testing individual species (e.g., Table 2).

Table 2. Decrease in acute toxicity for mixed species of aquatic oligochaete worms compared to individual species (from Chapman et al., 1982).

Species	NaPCP	Cd	Hg
L. hoffmeisteri and *T. tubifex* compared to *L. hoffmeisteri* alone	76%	241%	22%
L. hoffeisteri and *T. tubifex* compared to *T. tubifex* alone	53%	81%	64%

Interestingly and similarly, ecologists have found "an enhanced functional performance of mixed communities over that expected from a simple additive contribution of the community members" (Duarte, 2000). Environmental toxicologists worry about the cost of testing whereas the ultimate concern should be the cost of an incorrect decision. Their tests are simple whereas the environment is complex. Testing focuses on chemicals even though chemicals are only one issue, and not necessarily the most important one. As noted by (Chapman, 1995), habitat loss, introduced (exotic) species, and nutrient enrichment are all more significant global environmental insults than are toxic chemicals. So too is global climate change, including ozone depletion. And finally, toxicologists conduct environmental toxicology testing, generally without involving other scientific disciplines such as ecology.

Basically, those conducting environmental toxicology have become too concerned that their tests work and provide data. The focus is not on problem solving related to the ecology, but rather problem solving related to the tests themselves. This focus is understandable given regulatory, contractual, and other imperatives that require relatively simple, reproducible, and successful tests. Penalties may accrue to contract laboratories that do not meet set performance criteria for stipulated toxicity tests. But, the costs of an incorrect decision are arguably much larger than the costs of testing, though perhaps not as immediate.

The importance of ecotoxicology is readily demonstrated by a few examples. The first example involves unpublished studies at a marine site whose sediments were very highly contaminated with chlorophenols. Sediment toxicity testing with infaunal amphipods (*Rhepoxynius abronius*) recorded almost total mortalities in these sediments. Thus, by the measures of sediment chemistry and toxicity the sediments were both highly contaminated and highly toxic. However, surprisingly, benthic infaunal studies indicated that apparently healthy if not overly abundant communities existed in these sediments, including amphipods. Further investigations revealed the reasons for this discrepancy. Basically, the standard sediment grab samples taken for analyses had combined and mixed different sediment depths. In the actual environment heavily contaminated sediments were overlain by 1-2 cm of clean sediment that had been successfully recolonized by a diverse benthic community. Ecological input (i.e., ecotoxicology not solely environmental toxicology) was critical for fully evaluating management options.

A second example involves the Southern California Bight and infaunal benthic communities impacted by PCB and DDT around a sewage outfall discharge. Laboratory studies with spiked and field collected sediments indicated impaired reproduction at environmentally realistic concentrations (Murdoch et al., 1997), however such effects were not distinguishable in standard benthic infaunal surveys, whose results were dominated by organic carbon enrichment effects. It has been proposed that such populations may not be self-sustaining but rather dependent on outside immigration for their persistence (Chapman, 1995). This hypothesis is similar to that proposed by Menzie (1984) of diminished recruitment related to substrate modifications and inhibition of larval settlement for planktonic larvae. Ecotoxicology is required to test these hypotheses and to determine if they are correct and, if so, whether or not it matters.

Another example of the need for ecological toxicology is the issue of endocrine disruptive substances (EDCs). An EDC can be defined as an exogenous agent that interferes with the synthesis, storage/release, transport, metabolism, binding, action, or elimination of natural hormones responsible for the maintenance of homeostasis and the regulation of developmental processes (Cooper and Kavlock, 1997). The impact of EDCs on ecological systems is a new area of research; the current "model" of the endocrine system is mammalian and is still not completely defined in terms of its mechanisms and effects. Research on the effects of EDCs on non-mammalian species generally relies on the use of one or two model organisms, exposed to short-term, higher concentration, aqueous exposures, or exposure of isolated cell cultures to establish that endocrine disruption is occurring. Such studies comprise environmental toxicology not ecological toxicology. While additional single-species tests are required to attempt to determine mechanisms responsible for toxicity, this should not be the only type of testing undertaken. For example, very little research involves multiple species for long-term durations at environmentally relevant concentrations; data from standard test species such as *Daphnia* or rainbow trout are used to represent numerous phyla of aquatic organisms that have very different metabolisms and hormonal systems (e.g., Kashian and Dodson, 2000). There have been very few measures of endocrine disruption involving sediment exposures, bioavailability, or dietary uptake by organisms associated with sediment (Depledge and Billinghurst, 1999), though it has been suggested that uptake from sediment organisms can result in endocrine disruption in bottom feeding fish (Hecht et al., 2000). And, in fact, the classification of specific chemicals as estrogenic is still debatable (Soto et al., 1995).

Ecological toxicology is also required for substances that have different modes of acute and chronic toxicity. For instance, selenium is acutely toxic via water column exposures, however chronic toxicity occurs via dietary exposures and, in the aquatic environment, is primarily restricted to fish and waterfowl (Chapman, 1999). Thus standard environmental toxicology tests, which typically involve aqueous exposures, are not sufficient to determine the risks posed by selenium contamination.

However, similarly, ecology is not alone capable of determining what is occurring in the environment related to contamination. In particular, ecology alone cannot determine: relationships between organisms and contaminants (or other stressors); how contaminants and other stressors change community structure in terms of both direct (e.g., toxicity) and indirect (e.g., food chain) effects. To adequately assess and protect environmental quality, it is critically important to determine how stressor(s) affect different organisms and populations. Such information does not come from either environmental toxicology or ecology alone, but rather from their combination into ecological toxicology.

Key Ecotoxicological Issues

There are two key issues specific to ecotoxicology: acute and chronic responses; and, criteria for species selection.

Acute and Chronic Responses

Toxicological testing with acute and chronic responses often involves several individual species and endpoints. The results are used in some form of weight-of-evidence assessment, but without clear guidance as to how to use/interpret differential responses and intensities of response. The general assumption (which is not true in all cases as discussed above), is that the primary route of exposure is aqueous. Thus standard toxicity testing is routinely based on concentrations (e.g., LC/EC_{50} determinations) in the external medium (e.g., water, sediment).

However, increasingly it is becoming apparent that the dose, that is, the material associated with biological tissues, is a much better predictor of effects (Chapman and Wang, 2000).

Primary emphasis should be on three key testing parameters. First, test taxa should be most similar to resident taxa ***and*** of ecological relevance and importance. Second, exposure routes need to be direct and relevant. Third, taxa to be tested need to have proven to be appropriately sensitive to contaminants/stressors of concern. Testing the "most sensitive" species will not necessarily protect the majority of species (Chapman, 1998, 2000b); in fact, generally sensitive species do not seem to exist (Calow, 1996). Further, arguably the primacy of responses from "worst" to "less bad" is: mortality ? reproductive (fecundity) or growth effects ? other sublethal endpoints (e.g., behaviour, biomarkers). If you are dead you cannot reproduce or grow; if you cannot reproduce or grow your populations are self-sustaining.

Biomarkers (e.g., induction of metallothionein, mixed function oxidases, stress proteins) only provide an indication of exposure, and have not yet been linked directly to impacts at the organism level, let alone at the level of populations and communities (Cormier and Daniel, 1994; Decaprio, 1997). Bioindicators involve assessments of whole organisms, involve field data from multiple levels of biological organization, and have been linked directly to impacts (McCarty and Munkittrick, 1996; Munkittrick and McCarty, 1995; Power and McCarty, 1997; Vigerstad and McCarty, 2000). Pending their further development, biomarkers belong in environmental toxicology, while bioindicators belong in ecotoxicology.

The highest credibility in ecotoxicological testing will be derived from tests which measure mortality and reproductive or growth effects (e.g., both acute and energetically-based chronic effects), and use ecologically significant taxa similar to or related to resident taxa, which are likely to be exposed and which are appropriately sensitive. Calow (1996) suggests that "the direct effects of toxicants on survival and reproduction are more important than indirect action due to adjustments in predator and/or prey competitor-competitor interactions."

Criteria for Species Selection

Improvements are also needed in the manner in which we select species for testing (Table 3). Typically, we choose an organism that is economically or ecologically important. While the latter criterion is necessary, it does not go far enough if we are doing more than screening. For better predictions and for site-specific assessments, we need to test the equivalent of "key-stone" species, ideally for the area being assessed. Test species should be identified by community-based studies. Presently we test organisms that are widely available as this involves less effort; we should not hesitate to test organisms that are only reasonably available, even if this involves more effort, where these organisms are more appropriate. We focus on organisms that are easily cultured in the laboratory and genetically stable; again, this should not be our only focus. If testing a particular organism will greatly improve our assessment, we should not hesitate simply because somewhat more effort is required to cultivate that organism, so long as the extra effort required is not excessive compared to the information that will be obtained. As noted by Calow (1996), "the state of a few particular species in communities is likely to be more important than effects on a large number of species for community structure and function." Two characteristics that are not commonly considered for toxicity test organisms, but which should be, are their ability to be tested with other species and that the endpoints of such testing be ecologically and toxicologically relevant. Instead of worrying about consistent, measurable responses to toxicants and a linear dose-response, we should instead focus simply on being able to conduct testing in the laboratory or field, as appropriate. Linear dose-responses may be the exception rather than the rule (Chapman, 1998, 2000b); we should not shrink from this reality.

Table 3. Standard compared to recommended criteria for test species selection.

Standard Criteria	**Recommended Criteria**
An important ecological group (based on taxonomy, trophic level or niche)	Dominant or key-stone species (ideally for area being assessed); identified by community-based studies
Widely available (less effort)	Reasonably available (more effort)
Easily cultivated in laboratory and genetically stable	Can be reasonably cultivated from laboratory or collected from the field
Physiology, genetics, taxonomy, behavior, etc. well known	Not specified
Can be tested with other species/taxa	Not specified
Endpoints ecologically and toxicologically relevant	Consistent, measurable response to toxicants
Can be tested in laboratory or field	Resistant to disease and physical damage, can be handled in laboratory

Further, predictive and site-specific testing should more often involve mixtures of species rather than solely individual species, with appropriate selection of both individual and combined species. Such testing is important for several reasons. First, interactions affect toxicity responses. For instance, Chapman et al. (1982) found that testing mixed species of aquatic oligochaete worms, using species that typically coexist, resulted in lower toxicity than when individual species were tested (Table 2). Second, real environment interdependencies are not fully understood. These complex tests do not replace single-species tests, but are useful for detailed assessments of specific substances (e.g., pesticides) and field investigations of contaminated environmental media. However, they have their own problems. For instance, while field mesocosm studies will be closer to reality than single species tests, they have other disadvantages (e.g., lack of statistical power for detecting effects, strong edge influence).

And, testing should not be restricted to the laboratory (cf. Figure 1). The laboratory does not and cannot duplicate the field (laboratory testing can be under- or over-protective: Chapman, 2000b). Further, individual surrogate species responses are not related to all trophic levels, keystone species, populations, or ecosystem functioning responses. As noted above, mixed species tests involving, for instance, microcosms or mesocosms, are more realistic (though more difficult to interpret). Finally, in the field multiple responses can combine to produce an end result that would never be predictable from simple laboratory tests. For instance, food limitations and toxicant inputs can combine to magnify impacts.

Example: Estuarine Sediments

An example of the need for ecotoxicology rather than environmental toxicology is provided by estuarine sediments (Chapman and Wang, 2001). Estuaries are ecologically critical breeding, rearing and feeding areas. They are also physico-chemically unique with variable salinity gradients as well as strong gradients in pH, DO, Eh and particulates. Salinity gradients fluctuate temporally and spatially, particularly in salt wedge estuaries, for both sediment interstitial (i.e., pore) waters and for overlying waters. Salinity effectively controls contaminant partitioning. For instance, at high salinity hydrophobic organic chemicals are removed from the water column to the sediments in a "salting out" process, sorbing to formed particulate organic material (POM). For inorganic chemicals, two counteractive processes apply: desorption so that inorganics are flushed out of estuaries, or coagulation, flocculation and precipitation such that the sediments are a major repository. Because it controls contaminant partitioning, salinity also controls contaminant bioavailability. For instance, partitioning to particles favors sediment feeders. And salinity also controls faunal distributions directly, related to salinity tolerances and preferences.

Bioavailability predictions such as equilibrium partitioning (EqP) and acid volatile sulphides/selectively extracted metals (AVS/SEM) are not applicable to estuaries for two reasons. First, sediments are dynamic and there is no quasi-equilibrium state. Second, particle ingestion is an important exposure route. Further, sediment quality values (SQVs) have been derived for either fresh or marine waters; there are no SQVs that have been derived specifically for estuaries. Comparisons of existing SQVs to estuaries are questionable at best. In fresh or marine environments, interstitial (pore) water measurements provide useful information, however this is not the case in estuaries. Contaminants can occur in the overlying waters or on particles, and not all dissolved contaminants are bioavailable.

Estuarine benthos exhibits the "paradox of brackish water" (Remane, 1934). Basically, for most ecological factors the largest number of species occurs at intermediate values, however this is not the case for salinity. Rather, the largest number of species occurs in fresh and marine waters, with fewer species at intermediate salinities. Truly estuarine benthic infauna (existing between salinity ranges of about 5 to 8 g/L [ppt] and about 15-20 g/L) are generally r-selected. That is, they are small, with rapid/high reproduction/development and a low competitive ability. Because estuarine ecosystems are so dynamic, the benthos tends to be naturally "disturbed" as salinity fluctuations in particular render these communities highly variable. Biological surveys are difficult as there are no true reference sites. Instead, gradient approaches must be used to deal with salinity differences. And, in salt wedge estuaries, it is not unusual for tens of square kilometers of bottom sediments to show seasonal interstitial salinity differences related to seasonal differences in the extent and duration of the salt wedge. Whereas overlying water salinities can fluctuate daily or even hourly, interstitial waters in muddy sediments are much more conservative. This phenomenon results in seasonal up- and down-stream movements of fresh, estuarine and marine benthos over distances that may exceed ten kilometers related to seasonal changes in river flow and salt intrusion (Chapman and Brinkhurst, 1981). The end result is that habitat use over a year is greater than can be determined using a single sampling period "snapshot in time".

Estuarine sediment toxicity tests also do not generally consider the salinity that the organisms are actually exposed to or how this affects contaminant bioavailability. For reasons outlined above and detailed in Chapman and Wang (2001), estuarine sediment toxicity testing should be conducted at *in situ* interstitial salinities using estuarine organisms. However, most testing is done at manipulated salinities using freshwater or marine organisms. Basically, test salinities are adjusted to suit the test organisms that are available rather than using test organisms appropriate to the salinity conditions.

To date a total of 19 taxa have been used in estuarine sediment toxicity tests: 12 crustaceans, of which 8 are amphipods; 1 fish; 1 polychaete; 4 molluscs; 1 bacterium (Microtox®). This is a phylogenetically limited list. In fact, two North American amphipod species are favored for estuarine sediment toxicity testing: *Eohaustorius estuarius* (free burrowing, west coast species – relatively insensitive to copper [McPherson and Chapman, 2000]) and *Leptocheirus plumulosus* (open tube dweller, east coast species). As noted previously, few of these above 19 taxa are truly estuarine. None can reproduce across the full estuarine salinity gradient. Some obvious candidates for such testing, for instance estuarine aquatic oligochaetes, have been ignored (Chapman, 2001).

Clearly the single species toxicity tests conducted in estuaries to date are not ecotoxicological, but rather are representative of simplistic environmental toxicology approaches. Community level toxicity tests that may be considered ecotoxicological have been conducted, but there have been relatively few of these. To date such tests basically comprise three different types: field-collected sediments frozen then thawed, with exposures occurring in the laboratory either with pelagic larvae, or with the addition of meiofauna-rich sediment; field-collected sediments kept unfrozen and either spiked in the laboratory with contaminants or tested intact in microcosms; and, artificial sediments spiked and placed into the field.

Combining Ecology and Toxicology: Ecological Risk Assessment (ERA

An ERA is basically a process that evaluates the potential for adverse ecological effects that ***may occur*** as a result of exposure to contaminants or other stressors. It basically provides a framework or systematic means for gathering, organizing and evaluating scientific information to support management decisions. It recognizes, considers and reports uncertainties in estimating adverse effects of stressors.

An ERA, at least in North America, basically consists of four sequential components. First is the Problem Formulation/Hazard Identification phase, where goals and procedures are defined and available information is summarized. It is not inappropriate to use environmental toxicology in this preliminary phase of an ERA; it is inappropriate to use it in any further phases. Predictions, such as those based on the Equilibrium Partitioning (EqP) approach also belong here. The Exposure Assessment identifies exposure concentrations (emissions, rates, pathways), bioavailability, sensitive species/populations. This phase requires ecological information and knowledge. In particular, it requires the relative scaling of temporal and spatial processes affecting chemical contaminants (e.g., distribution and persistence), habitat (e.g., heterogeneity, fragmentation, movement of chemicals and organisms), and organisms (e.g., rate of population change) (Jepson and Sherratt, 1996). The Effects Assessment identifies the nature/character of the hazard. This is where ecological toxicology is required for a correct assessment with minimal uncertainty. Here too fit other "tools", particularly those related to cause-and-effect, such as toxicity identification and evaluation (TIE - Ho, 2001), reverse TIE and critical body residues. Good chemistry is required throughout and, in particular, the later stages of an ERA depend heavily on understanding of sediment chemistry and methods that allow for better control of contaminant exposure in laboratory toxicity and bioaccumulation tests. Natural variability must also be factored into the Effects Assessment phase: natural systems generally do not adhere to the equilibrium conditions projected by theory or assumed in study designs (Wiens, 1996). The final stage of an ERA, Risk Characterization, brings all the information from the other stages together to estimate risk based on exposure compared to effects and summarize major uncertainties (Munns et al., 2001).

What is critically important to recognize is that the type of risk assessment that is appropriate, and the type of data that are needed, depend on the objective of the assessment. Environmental toxicology will often be sufficient for risk assessments performed for generic assessments such as evaluations of most new and existing substances; it will rarely be sufficient for site-specific assessments or for realistic predictions of environmental effects.

Final Comments

Current relatively simple (environmental toxicology) tests remain useful for screening purposes but not for realistic predictions or for site-specific assessments. For the latter cases, ecology needs to be combined with toxicology both extrinsically (in a weight of evidence approach) and intrinsically (ecotoxicology). For example, in terms of determining the Predicted No Effect Concentration (PNEC) necessary for risk characterization (Predicted Effect Characterization or PEC divided by the PNEC), the latter combination offers the greatest reduction in uncertainty and increase in realism (Figure 2).

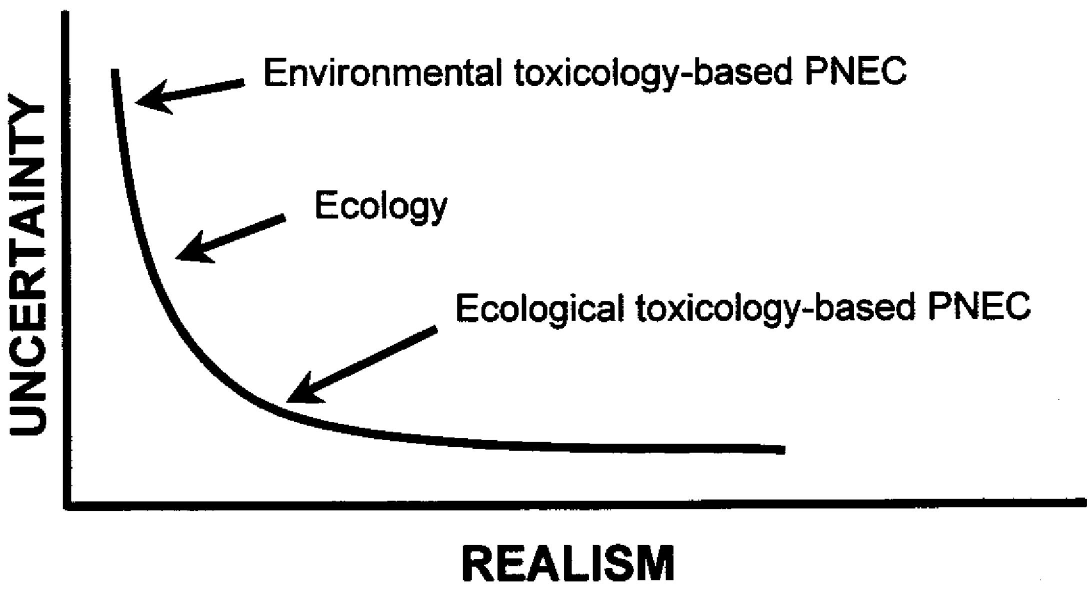


Fig. 2. Uncertainty and realism related to environmental toxicology, ecology, ecotoxicology.

However, ecotoxicology needs to be more than "largely toxicology with ecology added as a 'seasoning' as opposed to a 'main ingredient'" (Kareiva et al., 1996). Ecological understanding must be integrated into toxicology for a better, more coherent whole. For example, ecotoxicologists must be concerned both with small-scale variability and with large-scale variability; presently the focus is more on the former than the latter, and generally "either/or".

Key ecotoxicological issues involve the ecological relevance of lower-level effects. For instance, if individuals are killed or impaired, what does this mean to populations, and specifically what level of individual effect can significantly affect populations? Kareiva et al. (1996) note that ecotoxicology is required to answer two critical questions: "(1) how does an organism's rate of population growth or decline change as a function of chemical concentration; and (2) how rapidly can an organism's population recover from brief exposure to toxic compounds that subsequently degrade?"

To address these issues and questions, and to adequately protect populations against contaminants and other stressors, the following must be known and require both ecotoxicologists and population ecologists:

- The individual-level consequences of suborganism effects, including any trade-offs between life-history traits.
- The population-level consequences of individual effects, including any trade-offs between organisms.
- Processes (abiotic, biotic) regulating population size and health.
- Minimum viable population size, and genetic constraints.

Ecotoxicology must involve both observation (focused on ecology) as well as experimentation. Observations provide a basis for determinations, explanations or hypotheses; they also provide new information for hypothesis testing. As noted by Chapman (2000a): "in order to understand how animals respond to habitat, it is necessary to envisage the world from the perspective of the animal in question." This approach has yet to be incorporated broadly into either ecology or toxicology.

Acknowledgments

I thank two anonymous referees for their comments, which substantially improved the final paper.

References

Andrewartha H. G. & Birch, L. C. (1954). *The Distribution and Abundance of Animals*. University of Chicago Press, Chicago, Illinois.

Baird, D. J., Maltby, L., Greig-Smith, P. W. & Douben, P. E. T., eds. (1996). *ECOtoxicology: Ecological Dimensions*. Chapman and Hall, London, U.K.

Calow, P. (1996). Ecology in ecotoxicology: Some possible 'rules of thumb'. In *ECOtoxicology: Ecological Dimensions* (D. J. Baird, L. Maltby, P. W. Greig-Smith & P. E. T. Douben, eds), pp. 5-12. Chapman and Hall, London UK.

Castilla, J. C. (2000). Roles of experimental marine ecology in coastal management and conservation. *J. Exp. Mar. Biol. Ecol.* **250**, 3-21.

Chapman, P. M. (1995). Ecotoxicology and pollution – key issues. *Mar. Pollut. Bull.* **31**, 167-177.

Chapman, P. M. (1998). New and emerging issues in ecotoxicology – the shape of testing to come? *Austral. J. Ecotox.* **4**,1-7.

Chapman, P. M. (1999). Selenium – A potential time bomb or just another contaminant? *Human Ecol. Risk Assess.* **5**,1122-1137.

Chapman, M. G. (2000a). Poor design of behavioural experiments gets poor results: Examples from intertidal habitats. *J. Exper. Mar. Biol. Ecol.* **250**, 77-95.

Chapman, P. M. (2000b). Whole effluent toxicity (WET) testing – usefulness, level of protection, and risk assessment. *Environ. Toxicol. Chem.* **19**, 3-13.

Chapman, P. M. (2001). Utility and relevance of aquatic oligochaetes in ecological risk assessment. *Hydrobiologia* (in press).

Chapman, P. M. & Brinkhurst, R. O. (1981). Seasonal changes in interstitial salinities and seasonal movements of subtidal benthic invertebrates in the Fraser River estuary, BC. *Estuar. Coastal Mar. Sci.* **12**, 49-66.

Chapman, P. M. & Wang, F. (2000). Issues in ecological risk assessment of metals and metalloids. *Human Ecol. Risk Assess.* **6**: 965-988.

Chapman, P. M. & Wang, F. (2001). Assessing sediment contamination in estuaries. *Environ. Toxicol. Chem.* **20**: 3-22.

Chapman, P. M., Farrell, M. A. & Brinkhurst, R. O. (1982). Effects of species interactions on the survival and respiration of *Limnodrilus hoffmeisteri* and *Tubifex tubifex* (Oligochaeta, Tubificidae) exposed to various pollutants and environmental factors. *Water Res.* **16**, 1405-1408.

Colombini, I., Chelazzi, L. & Scapini, F. (1994). Solar and landscape cues as orientation mechanisms in the beach-dwelling beetle *Eurynebria complata* (Coleoptera, Carabidae). *Mar. Biol.* **118**, 425-432.

Cooper, R. & Kavlock, R. (1997). Endocrine disruptors and reproductive development: A weight-of-evidence overview. *J. Endocrin.* **152**, 159-166.

Cormier, S. M. & Daniel, F. B. (1994). Biomarkers: Taking the science forward. *Environ. Toxicol. Chem.* **13**, 1011-1012.

Decaprio, A. P. (1997). Biomarkers: Coming of age for environmental health and risk assessment. *Environ. Sci. Technol.* **31**, 1837-1848.

Della Santina, P. & Naylor, E. (1994). Endogenous rhythms in the homing behaviour of the limpet *Patella vulgata* Linnaeus. *J. Molluscan. Stud.* **60**, 87-91.

Depledge, M. & Billinghurst, Z. (1999). Ecological significance of endocrine disruption in marine invertebrates. *Mar. Pollut. Bull.* **39**, 32-38.

Duarte, C. M. (2000). Marine biodiversity and ecosystem services: An elusive link. *J. Exper. Mar. Biol. Ecol.* **250**, 117-131.

Hatton, H. (1932). Quelques observations sur le peuplement en *Fucus vesiculosus* des surfaces rochenses denudees. *Bull. Lab. Mar. Mus. Hist. Nat.* **9**, 1-6.

Hecht, S., Gunnarrson, J., Boese, B., Lamberson, J., Schaffner, C. & Giger, W. (2000). Uptake and accumulation of 4-nonylphenol in estuarine amphipods in response to refractory and labile organic matter. Presented at the 21st Annual Meeting of the Society of Environmental Toxicology and Chemistry, November 12-16, 2000, Nashville, Tennessee.

Ho, K. (2001). Identification and evaluation of stressors in toxic sediments and dredged materials. *Mar. Pollut. Bull.* [THIS ISSUE]

Ingersoll, C. G., Dillon, T. & Biddinger, G.R. (1997). *Ecological Risk Assessment of Contaminated Sediments.* SETAC Press, Pensacola, Florida.

Jepson, P. C. & Sherratt, T. N. (1996). The dimensions of space and time in the assessment of ecotoxicological risks. In *ECOtoxicology: Ecological Dimensions* (D. J. Baird, L. Maltby, P. W. Greig-Smith & P. E. T. Douben, eds), pp. 43-54. Chapman and Hall, London, UK.

Kareiva, P., Stark, J. & Wennergren, U. (1996). Using demographic theory, community ecology and spatial models to illuminate. In ecotoxicology. In *ECOtoxicology: Ecological Dimensions* (J. Baird, L. Maltby, P. W. Greig-Smith & P. E. T. Douben, eds), pp. 13-23. Chapman and Hall, London, UK.

Kashian, D. R. & Dodson, S. I. (2000). The effects of xenobiotics on sex determination and reproduction in *Daphnia*. Presented at the 21st Annual Meeting of the Society of Environmental Toxicology and Chemistry, November 12-16, 2000, Nashville, Tennessee.

McCarty, L. S. & Munkittrick, K. R. (1996). Environmental biomarkers in aquatic toxicology: fiction, fantasy, or functional? *Human Environ. Risk Assess.* **2**, 268-274.

McPherson, C. A. & Chapman, P. M. (2000). Copper effects on potential sediment test organisms: The importance of appropriate sensitivity. *Mar. Pollut. Bull.* **40**, 656-665.

Menge, B. A. (2000). Top-down and bottom-up community regulation in marine rocky intertidal habitats. *J. Exper. Mar. Biol. Ecol.* **250**, 257-289.

Menzie, C. A. (1984). Diminishment of recruitment: A hypothesis concerning impacts on benthic communities. *Mar. Pollut. Bull.* **15**, 127-128.

Mills, L. S., Soule, M. E. & Doak, D. F. (1993). The key-stone species concept in ecology and conservation. *Bioscience* **43**, 219-224.

Munkittrick, K. R. & McCarty, L. S. (1995). An integrated approach to ecosystem health management: top-down, bottom up or middle out. *J. Aquat. Ecosys. Health* **4**, 77-90.

Munns, W. R. Jr., Berry W. J. & DeWitt, T. (2001). Toxicity testing, risk assessment, and options for dredged material management. *Mar. Pollut. Bull.* [THIS ISSUE]

Murdoch, M. H., Chapman, P. M., Johns, D. M. & Paine, M. D. (1997). Chronic effects of organochlorine exposure in sediment to the marine polychaete *Neanthes arenaceodentata*. *Environ. Toxicol. Chem.* **16**, 1494-1503.

Padilla, D. K. & Allen, B. J. (2000). Paradigm lost: Reconsidering functional form and group hypotheses in marine ecology. *J. Exper. Mar. Biol. Ecol.* **250**, 207-221.

Polis, G. A. (1999). Why are parts of the world green? Multiple factors control productivity and the distribution of biomass. *Oikos* **86**, 3-15.

Power, M. E. & Mills, L. S. (1995). The Keystone cops meet in Hilo. *TREE* **10**: 182-183.

Power, M. & McCarty, L. S. (1997). Fallacies in ecological risk assessment practices. *Environ. Sci. Technol.* **31**, 370A-375A.

Power, M. E., Tilman, D., Estes, J., Menge, B. A., Bond, W. J., Mills, L. S., Daily, G., Castilla, J. C., Lubchenco, J. & Paine, R. T. (1996). Challenges in the quest for keystones. *Bioscience* **46**, 609-620.

Remane, A. (1934). Die Brackwasserfauna. *Verh. Dtsch. Zool. Ges.* **36**, 34-74.

Solomon, K. R. (2001). New concepts in ecological risk assessment: Where do we go from here? *Mar. Pollut. Bull.* [THIS ISSUE]

Soto, A. M., Sonneschein, C., Chung, K. L., Fernandez, M. F., Olea, N. & Serrano, F. O. (1995). The E-screen assay as a tool to identify estrogens – An update on estrogenic environmental pollutants. *Environ. Health Perspec.* **103**, 113-122.

Sutherland, J. P. (1974). Multiple stable points in natural communities. *Am. Nat.* **108**, 859-873.

Tilman, D. (1997). Biodiversity and ecosystem functioning. In *Nature's Services – Societal Dependence on Natural Ecosystems* (G. C. Daly, ed), pp. 93-112. Island Press, Washington, D.C.

Underwood, A. J. (2000). Experimental ecology of rocky intertidal habitats: What are we learning? *J. Exp. Mar. Biol. Ecol.* **250**, 51-76.

Underwood, A. J., Chapman, M. G. & Connell, S. D. (2000). Observations in ecology: You can't make progress on processes without understanding the patterns. *J. Exp. Mar. Biol. Ecol.* **250**, 97-115.

Vigerstad, T. & McCarty, L. S. (2000). The ecosystem paradigm and environmental risk management. *Human Environ. Risk Assess.* **6**, 369-381.

Wiens, J. A. (1996). Coping with variability in environmental impact assessment. In *ECOtoxicology: Ecological Dimensions* (D. J. Baird, L. Maltby, P. W. Greig-Smith & P. E. T. Douben, eds), pp. 55-70. Chapman and Hall, London, UK.

An Overview of Toxicant Identification in Sediments and Dredged Materials

KAY T. HO[1], ROBERT M.. BURGESS[1], MARGUERITE C. PELLETIER[1], JONATHAN R. SERBST[1], STEVE A. RYBA[1], MARK G.CANTWELL[1] ANNE KUHN[1] and PAMELA RACZELOWSKI[2]
[1]*U.S. Environmental Protection Agency, Atlantic Ecology Division, Narragansett, RI 02882, USA*
[2]*University of Rhode Island, Kingston, RI 02892, USA*

Abstract

The identification of toxicants affecting aquatic benthic systems is critical to sound assessment and management of our nation's waterways. Identification of toxicants can be useful in designing effective sediment remediation plans and reasonable options for sediment disposal. Knowledge of which contaminants affect benthic systems allows managers to link pollution to specific dischargers and prevent further release of toxicant(s). In addition, identification of major causes of toxicity in sediments may guide programs such as those developing environmental sediment guidelines and registering pesticides, while knowledge of the causes of toxicity which drive ecological changes such as shifts in benthic community structure would be useful in performing ecological risk assessments. To this end, the US Environmental Protection Agency has developed tools (Toxicity Identification and Evaluation (TIE) methods) that allow investigators to characterize and identify chemicals causing acute toxicity in sediments and dredged materials. Development of these methods for both interstitial waters and whole sediments is nearly complete and a draft guidance document is expected by the end of 2001. To date, most sediment TIEs have been performed on interstitial waters. Preliminary evidence from the use of interstitial water TIEs reveals certain patterns in causes of sediment toxicity. First, among all sediments tested, there is no one predominant cause of toxicity; metals, organics and ammonia play approximately equal roles in causing toxicity. Second, within a single sediment there are multiple causes of toxicity detected; not just one chemical class is active. Third, the role of ammonia is very prominent in these interstitial waters. Finally, if sediments are divided into marine or freshwater, TIEs performed on interstitial waters from freshwater sediments indicate a variety of toxicants in fairly equal proportions, while TIEs performed on interstitial waters from marine sediments have identified only ammonia and organics as toxicants, with metals playing a minor role. Preliminary evidence from whole sediment TIEs indicate that organic compounds play a major role in the toxicity of marine sediments, with almost no evidence for either metal or ammonia toxicity. However, interpretation of these results may be skewed because only a small number of interstitial water (n= 13) and whole sediment (n=5) TIEs have been completed. These trends may change as more data are collected.

Keywords: sediment; toxicity; identification of toxicants, TIE; Toxicity Identification and Evaluation.

Introduction

Toxic sediments pose a risk to aquatic life, human health and wildlife throughout the world. There is an overwhelming amount of evidence that demonstrates chemicals in sediments are responsible for toxicological (Ankley et al., 1989; Chapman, 1988; Giesy and Hoke, 1989; Giesy and Hoke, 1990; Swartz et al., 1989; Williams et al., 1986) and adverse ecological effects (Anderson et al., 1987; Bailey et al., 1995b; Hartwell et al., 1997; Hatakeyama and Yokoyama, 1997; Swartz et al., 1994; Swartz et al., 1982). Frequently, the chemicals causing these effects are present in the sediment as mixtures of organic, metal and other types of contaminants. The ability to identify which class or specific chemical is responsible for toxicity in such complex mixtures is the objective of Toxicity Identification and Evaluation (TIE) methods. These methods combine toxicity testing and simple chemical manipulations in an iterative process that allows the investigator to continually narrow the focus of the investigation on the suspected toxicant until a satisfactory identification has been performed. TIE methods are divided into three Phases: I-- Characterization, II-- Identification and III --Confirmation (Burgess et al., 1996; Durhan et al., 1992; Mount and Anderson-Carnahan, 1988; Mount and Anderson-Carnahan, 1989; Mount et al., 1993; Norberg-King et al., 1991a; Norberg-King et al., 1991b). TIE methods have been widely and successfully used for identifying toxicants in effluents (Amato et al., 1992; Ankley and Burkhard, 1992; Bailey et al., 1995a; Burgess et al., 1995; Burkhard and Ankley, 1989; Burkhard and Jenson, 1993; Jin et al., 1999a; Jin et al., 1999b; Jop and Askew, 1994; Schubauer-Berigan et al., 1993; Wells et al., 1994), fresh (Norberg-King et al., 1991; Riveles and Gersberg, 1999; Steidl-Pulley et al., 1998) and marine (Burgess, 2000; Hunt et al., 1999; Rumbold and Snedaker, 1999) waters, ballast waters (Ertan-Unal et al., 1998), and wastewater treatment plants (Adamsson et al., 1998). These effluent methods have also been adapted for use with interstitial waters (Ankley et al., 1992a; Ho et al., 1997b).

Identification of toxicants in sediments is useful in a variety of contexts. The identification of specific classes of toxicants would be helpful in designing reasonable options for disposal of dredged sediments (Ankley et al., 1992b) and establishing effective sediment remediation schemes. Once a toxicant is identified, steps can also be taken to link a toxicant to a discharger and prevent further discharge. In addition, identification of major causes of toxicity in sediments may guide programs such as development of environmental sediment guidelines and, retrospectively, aid regulators in determining the type of pesticide or manufactured chemical that may cause toxicity in the field. Finally, knowledge of the causes of toxicity which drive ecological changes such as alteration of community structure would be useful in performing ecological risk assessments.

While the examples in this paper are from in-place sediments and not dredged materials, the methods used, and the lessons learned from the research covered in this paper may be applied to dredged materials. The only difference we have noted between in-place sediments and dredged materials is that dredged materials may often have much higher levels of ammonia since they are not subject to the same rates of flushing as surficial sediments.

Interstitial Water TIEs

To date, methods to identify toxicants in sediments have largely been focused on sediment interstitial waters. This is due to the fairly straightforward application of effluent methods to the interstitial water phase. Advantages to performing TIEs on interstitial waters include (1) already existing aqueous methods, (2) the ability to test organisms that are not compatible with a solid matrix (i.e., sediment) and (3) the fact that interstitial waters are a major route of exposure of many toxicants and the primary route for most water-soluble toxicants (Adams et al., 1985). However, there are also factors that make results from interstitial water exposures suspect and these will be discussed later in the paper.

Interstitial water TIEs performed on freshwater sediments have identified ammonia (Ankley et al., 1990; Gupta and Karuppiah, 1996a; Karuppiah and Gupta, 1996; Sprang and Janssen, 1997; Sprang et al., 1996; Wenholz and Crunkilton, 1995), organic chemicals (Gupta and Karuppiah, 1996; Karuppiah and Gupta, 1996; Schubauer-Berigan and Ankley, 1991), and metals (Boucher and Watzin, 1999; Gupta and Karuppiah, 1996a; Gupta and Karuppiah, 1996b; Karuppiah and Gupta, 1996; Schubauer-Berigan et al., 1993; Schubauer-Berigan and Ankley, 1991; Wenholz and Crunkilton, 1995) as bioavailable toxicants. In marine interstitial waters, ammonia and organic chemicals (Ho et al., 1997b; Kuhn et al., 1995) have been characterized as causes of toxicity.

From these relatively limited interstitial water results (n=13) we can propose some hypotheses about causes of toxicity in sediments. First, the causes of toxicity are fairly numerous, that is, there is no one predominant cause of toxicity, such as PAHs, and that metals, organics and ammonia all play a role in about equal amounts in causing toxicity (Figure 1).

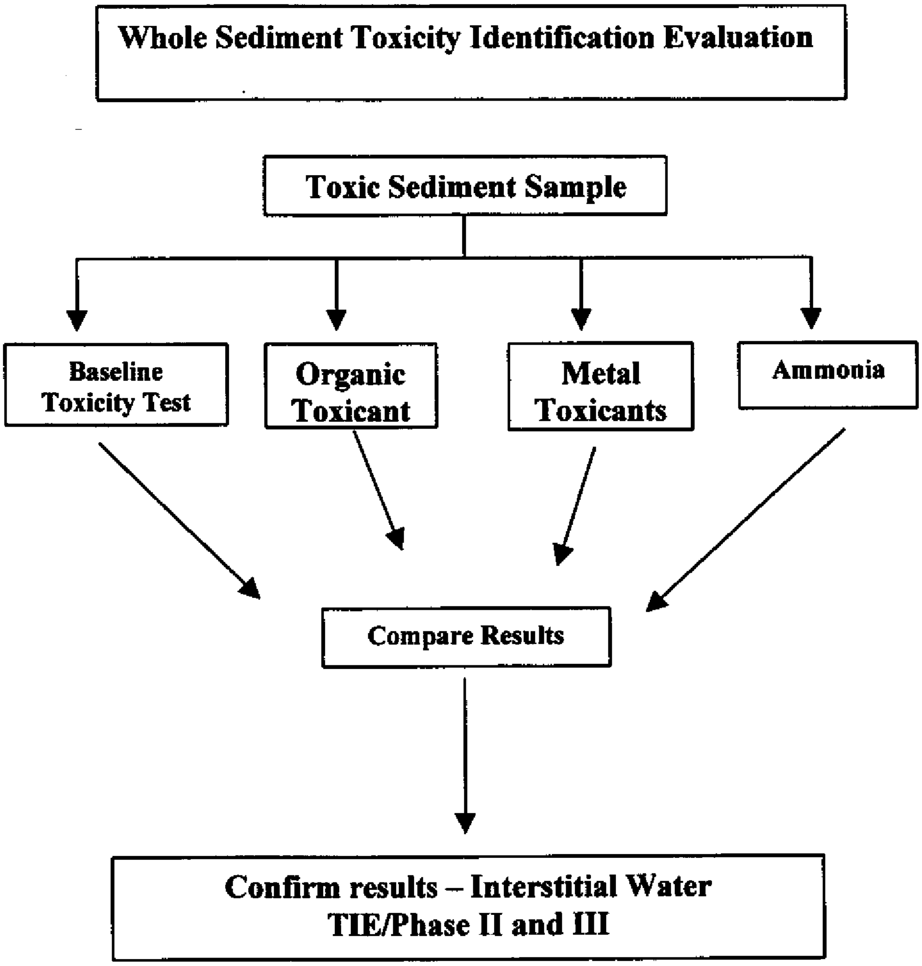


Fig. 1. Flow chart of the whole sediment TIE process.

Second, within a single sediment there are usually multiple causes of toxicity detected; not just one chemical class is active. Third, the role of ammonia is very prominent in these interstitial waters. For years researchers have been aware of the presence of ammonia in sediments, but

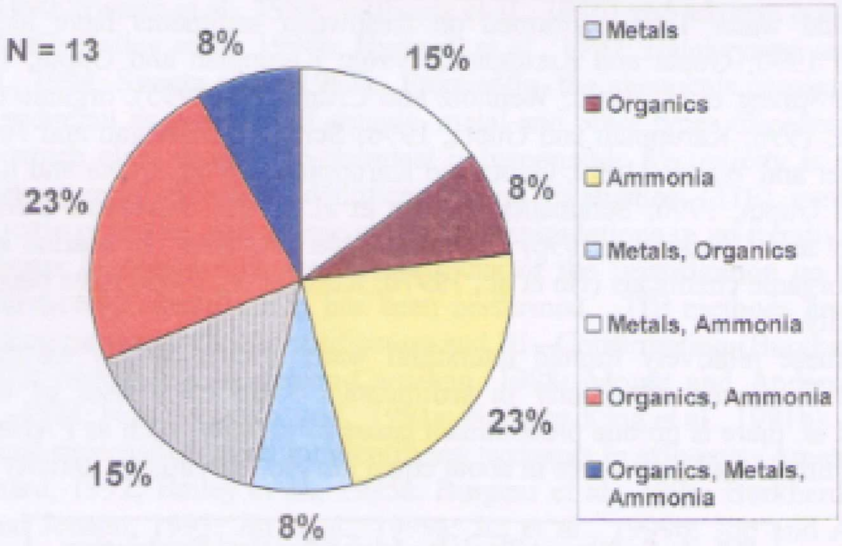


Fig. 2. Causes of acute toxicity in interstitial waters based on TIE results – freshwater and marine sediments.

until the results of these TIEs, the importance of ammonia in sediment toxicity had not been elucidated. However the importance of ammonia in sediment toxicity may be overemphasized due to the interstitial water exposure route which generally over-exposes organisms to water soluble toxicants. Finally, if sediments are divided into freshwater (Figure 2) or marine (Figure 3), TIE results from freshwater sediments indicate that metals, organics and ammonia cause toxicity in fairly equal proportions, while TIE results from marine sediments indicate metals do not play a large role in causing toxicity.

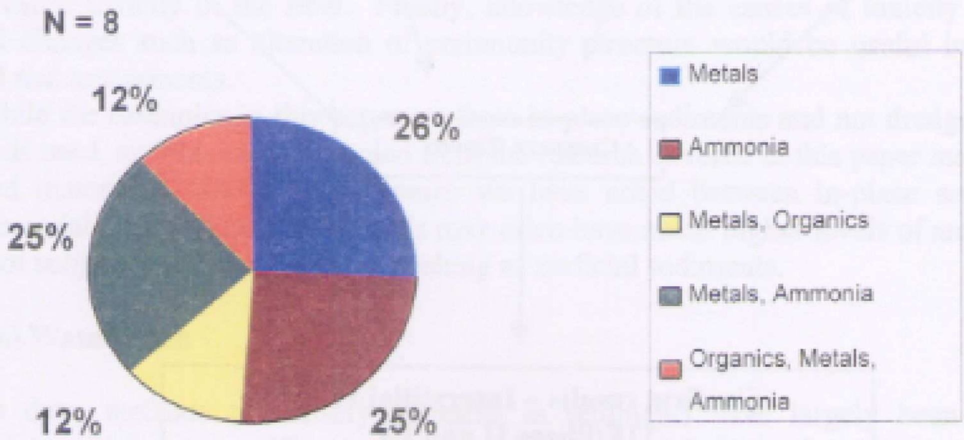


Fig. 3. Causes of acute toxicity in interstitial waters based on TIE results – freshwater sediments.

This may be the result of higher concentrations of sulfides found in marine sediments (Hansen et al., 1996) which bind to many toxic metals reducing their toxicity. This interpretation is supported by acid volatile sulfide (AVS) research which shows that metals may not play as large a role as previously assumed in causing *acute* toxicity in marine sediments (Hansen et al., 1996). It is necessary to stress that these findings address only the *acute* toxicity of metals as TIEs currently do not address chronic or bioaccumulative effects. These trends may change as larger numbers of interstitial water TIEs are performed and as interstitial and whole sediment TIE results are compared to each other and verified in the field.

Issues in Interstitial Water TIEs

As previously mentioned, many reasons exist to perform interstitial water TIEs, however, there are issues that complicate interpretation of results from TIEs that are based upon an interstitial water exposure. These issues include changes in metal toxicity due to oxidation during routine aeration of interstitial water toxicity tests, changes in the interstitial water pH due to CO_2 volatilization, underexposure to high log K_{ow} compounds that may sorb to test containers, overexposure of organisms not normally exposed to 100% interstitial water, and elimination of other routes of exposure, such as sediment ingestion. On the other hand, interstitial water methods may reasonably approximate field exposures for water soluble compounds such as metals and ammonia, particularly for organisms that have interstitial water as their major route of

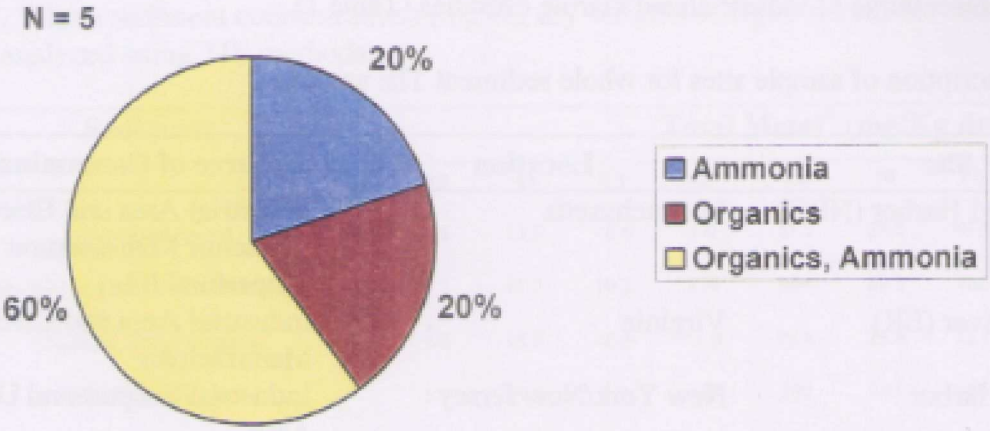


Fig. 4. Causes of acute toxicity in interstitial waters based on TIE results – marine sediments.

exposure.

In-situ interstitial water TIEs (Burton et al., 1996) may alleviate some of the concerns over sorption of high log K_{ow} compounds and oxidation of metals by allowing the interstitial waters to remain in contact with the solid sediment matrix, however, all interstitial water testing eliminates potentially important routes of exposure such as sediment ingestion and direct contact with particles. In addition, in-situ testing combined with TIEs is exploratory and limited to low-energy, shallow water systems. Whole sediment testing and TIEs may be a more "accurate" and realistic mode of exposure to organisms. Whole sediment TIEs include a more realistic exposure for burrowing organisms in a whole sediment matrix. Further, the presence of whole sediment stabilizes the interstitial water reduction and oxidation potential and pH. For all these reasons we have been developing whole sediment TIE methods.

Whole Sediment TIEs

Unlike interstitial water TIEs, whole sediment methods, particularly Phase II (Identification) and Phase III (Confirmation) methods, are still under development. These methods assume there are three major sources of toxicity in sediments: ammonia, metals and non-polar organics. The methods use the green macroalgae *Ulva lactuca* to remove ammonia, cation exchange resins to sequester metals, and powdered coconut charcoal to sorb organics (Figure 1). Details of whole sediment TIE methodologies are found elsewhere (Burgess et al., 2000; Ho et al., 1997a; Ho et al., 2000; Ho et al., 1999). In general, sediments are amended with cation resin or coconut charcoal, or *U. lactuca* is added to the overlying water. The toxicity of the treated sediments are then compared to the toxicity of untreated sediments. The methods used in this research have been validated in several marine sediments; similar methods for freshwater sediments exist (Ho et al., 1997a).

Our laboratory has performed whole sediment TIEs on five marine sediments: New Bedford Harbor (NBH), Elizabeth River (ER), New York Harbor (NYH), Baltimore Harbor (BH) and Bayou Verdine (BV). Three of the five sediments were tested within two weeks of collection; the other two (NBH and NYH) were stored for over 2 years. The sediments are generally representative of industrialized marine estuaries (Table 1).

Table 1. Description of sample sites for whole sediment TIE analyzed.

Site	Location	Source of Contamination
New Bedford Harbor (NBH)	Massachusetts	Industrial Area and Electrical Capacitor Manufacturer (Superfund Site)
Elizabeth River (ER)	Virginia	Industrial Area and Creosote Manufacturer
New York Harbor	New York/New Jersey	Industrial Seaport and Urban Area
Baltimore Harbor (BH)	Maryland	Industrial Seaport and Urban Area
Bayou Verdine	Louisiana	Industrial Area

In four of the five sediments, powdered coconut charcoal removed all of the toxicity; this implies the toxicity is from bioavailable organic compounds (Figure 5). All of these sites contain elevated concentrations of toxic metals, polychlorinated biphenyls (PCBs) and polycyclic aromatic hydrocarbons (PAHs) (Tables 2 - 4).

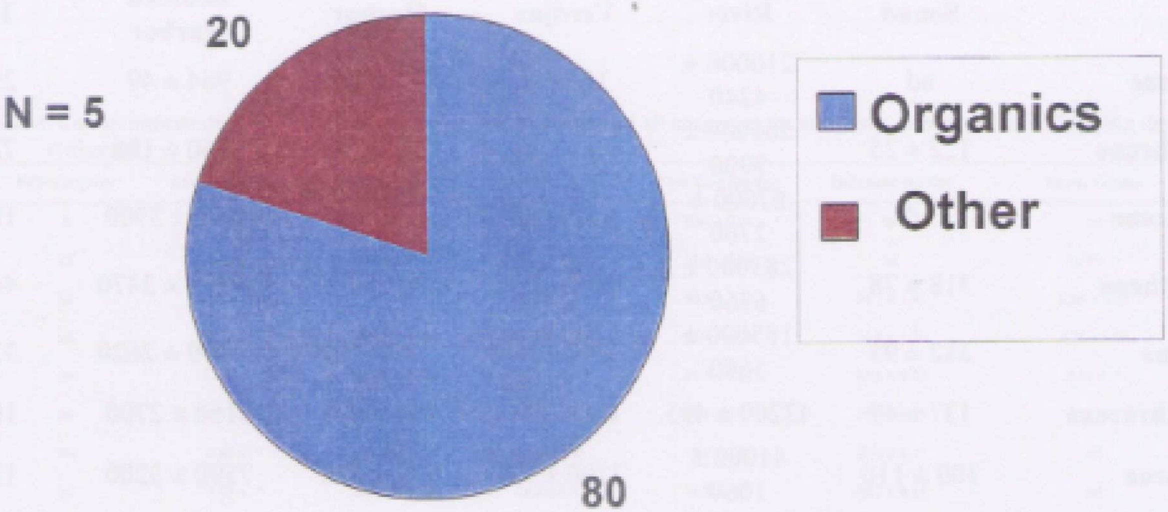


Fig. 5. Causes of acute toxicity in whole sediments based on TIE results – marine sediments.

Table 2. Whole sediment concentrations (mg/Kg dry weight) of eight metals for sediments analyzed using TIE methods.

Sediment	Total Metal[a] (mg/Kg dry)							
	Ag	As	Cd	Cr	Cu	Ni	Pb	Zn
Long Island Sound	<dl	13.2	<0.6	5.0	47.7	29.8	40.0	148
New Bedford Harbor (Massachusetts)	6.2	11.2	10.2	474	904	89.5	468	1230
Elizabeth River (Virginia)	<dl	18.0	<0.6	72.9	93.8	23.6	72.1	267
New York Harbor	5.9	29.8	2.4	419	200	141	186	324
Baltimore Harbor (Maryland)	na	na	na	na	na	na	na	na
Bayou Verdine (Louisiana)	na	na	na	na	na	na	na	na

[a]Variability between analytical replicates ~5%.
na = Data not available
<dl= Below detection limit

Table 3. Whole sediment concentrations of polycyclic aromatic hydrocarbons (PAHs).

PAH (μg/Kg dry)	Site					
	Long Island Sound	Elizabeth River	Bayou Verdine	Baltimore Harbor	New Bedford Harbor	New York Harbor
Fluorene	nd	210000 ± 4240	129 ± 1	295 ± 175	964 ± 49	2530 ± *
Phenanthrene	122 ± 25	560000 ± 9900	735 ± 13	790 ± 383	9960 ± 198	7220 ± *
Anthracene	nd	67000 ± 2760	141 ± 18	269 ± 117	4290 ± 3900	1040 ± *
Fluoranthene	318 ± 78	283000 ± 6360	697 ± 81	2180 ± 955	29800 ± 3470	4430 ± *
Pyrene	382 ± 95	185000 ± 5660	1880 ± 269	1770 ± 728	24500 ± 2620	3250 ± *
Benz[a]anthracene	137 ± 47	42200 ± 495	639 ± 30	754 ± 433	6160 ± 2700	1040 ± *
Chrysene	300 ± 110	41000 ± 1060	2190 ± 92	978 ± 513	7590 ± 3200	1320 ± *
ΣBenzofluoranthene	152 ± 37	17300 ± 636	1290 ± 120	1160 ± 611	5930 ± 1480	680 ± *
Benzo[e]Pyrene	282 ± 78	13700 ± 424	1900 ± 163	2300 ± 1630	5930 ± 2020	1100 ± *
Benzo[a]Pyrene	192 ± 86	18400 ± 849	1100 ± 92	926 ± 373	6970 ± 2510	1010 ± *
Perylene	nd	4900 ± 537	205 ± 20	598 ± 114	1540 ± 559	468 ± *
Indeno[123-cd]pyrene	nd	4880 ± 269	414 ± 37	915 ± 390	3520 ± 1840	461 ± *
Dibenz[ah]anthracene	nd	1675 ± 106	311 ± 52	355 ± 286	1090 ± 631	160 ± *
Benzo[ghi]perylene	206 ± *	4660 ± 290	1420 ± 170	1830 ± 1290	3820 ± 1930	594 ± *

nd not detected
* based on one replicate.

Table 4. Concentrations of 23 individual PCB congeners and the sum of all congeners on sediments (mean ± sd, n = 2) (µg/Kg dry sediment).

Table 4. Concentrations of 23 individual PCB congeners and the sum of all congeners on sediments (mean ± sd, n = 2) (µg/Kg dry sediment).

PCB Congener	Long Island Sound	New Bedford Harbor	Elizabeth River	New York Harbor	Baltimore Harbor	Bayou Verdine
8	nd	23000 ± 7500	80.1 ± 8.56	9.82 ± 0.26	nd	nd
18	6.74 ± *	26900 ± 8840	58.7 ± 1.20	13.1 ± 0.78	nd	7.87 ± *
28	5.33 ± 1.18	54400 ± 14500	16.4 ± 9.07	40.5 ± 1.77	14.7 ± 1.41	6.80 ± 0.42
52	3.00 ± *	30800 ± 9330	35.8 ± 2.55	15.9 ± 1.77	14.8 ± 1.20	4.21 ± 1.76
44	2.52 ± *	16200 ± 3460	20.1 ± 0.00	10.4 ± 0.21	13.5 ± 0.78	5.41 ± *
66	3.92 ± 1.59	26200 ± 6010	13.9 ± 3.39	32.1 ± 1.77	19.9 ± 0.49	7.16 ± *
101	4.67 ± *	19800 ± 5020	19.0 ± 4.81	18.6 ± 2.33	24.0 ± 0.21	nd
99	4.51 ± *	20500 ± 5160	44.4 ± 8.41	14.8 ± 0.78	13.7 ± 0.71	nd
110	5.61 ± 0.21	32700 ± 9620	15.0 ± 2.05	32.8 ± 3.46	25.5 ± 0.35	4.54 ± 0.15
151	nd	1870 ± 460	12.4 ± 2.19	8.63 ± 1.17	nd	nd
118	3.53 ± 1.16	19100 ± 4450	13.2 ± 2.69	22.0 ± 1.63	18.5 ± 0.64	3.56 ± *
153	7.19 ± 0.01	12600 ± 3110	20.1 ± 0.21	34.4 ± 1.41	27.1 ± 5.09	4.33 ± 1.71
105	nd	3030 ± 799	11.9 ± 0.14	13.2 ± 0.57	10.6 ± 2.21	nd
138	13.2 ± 11.2	8730 ± 2500	23.9 ± 2.69	37.0 ± 0.49	20.9 ± 0.28	12.4 ± 5.98
187	2.94 ± 0.92	1710 ± 424	7.62 ± 2.45	15.6 ± 0.85	8.41 ± 0.54	2.24 ± 0.57
183	4.08 ± *	789 ± 165	15.0 ± 0.49	8.62 ± 1.96	nd	3.20 ± *
128	3.34 ± *	1530 ± 438	15.4 ± 1.98	9.02 ± 1.68	4.31 ± *	nd
180	4.20 ± 1.99	2200 ± 410	11.3 ± 7.51	24.8 ± 0.71	14.4 ± 0.42	2.68 ± 0.21
170	2.04 ± *	1360 ± 226	nd	10.2 ± 1.04	5.58 ± 0.33	nd
195	2.22 ± *	212 ± 38.2	nd	5.19 ± 0.63	nd	nd
194	3.58 ± 0.81	283 ± 48.8	3.05 ± *	4.41 ± 0.78	13.3 ± *	1.96 ± 0.85
206	nd	324 ± 79.9	13.3 ± *	15.0 ± 0.78	nd	5.60 ± *
209	3.41 ± 0.22	63.3 ± 15.4	23.0 ± 4.81	12.6 ± 1.06	12.2 ± 4.00	1.76 ± *
Sum	86.0	304000	473	408	261	73.7

nd Not Detected;
* based on one replicate.

The concentrations are high enough to cause toxicity (US Environmental Protection Agency, 1991, Ho et al., 1997b, however sediment normalization factors such as acid volatile sulfide (AVS) and organic carbon still need to be applied to assess the bioavailable concentrations. The fifth sediment, Bayou Verdine is from an unusual area which contains a groundwater seep. The seep contains concentrations of ions different enough from seawater to cause ion toxicity to the test organisms. Although metal concentrations were high in many of the sediments, TIE results imply that only organics were bioavailable. The *U. lactuca* manipulation did not change the toxicity of any of these sediments which indicates ammonia was not toxic in these sediments. The lack of ammonia toxicity in any of the whole sediment TIE results was surprising. This early trend in whole sediment TIEs may imply that the large ammonia signal observed in many of the interstitial water TIEs may be an artifact of overexposure to ammonia which is a water soluble toxicant. However, sediments tested so far with whole sediment TIE methods had non-toxic levels of un-ionized ammonia (Miller et al., 1990) in their interstitial waters except for Baltimore Harbor (Table 5). Field verification of results from both interstitial and whole sediment TIEs will allow us to make more informed decisions on the cause(s) of acute toxicity.

Table 5. Interstitial water concentrations of total and un-ionized ammonia. These are based upon a single replicate.

Site	NH_x (mg/l)	NH_3 (mg/l)
Long Island Sound	1.93	0.06
New Bedford Harbor (Massachusetts)	2.57	0.04
Elizabeth River (Virginia)	1.38	0.05
New York Harbor	0.84	0.02
Baltimore Harbor (Maryland)	9.19	0.23
Bayou Verdine (Louisiana)	13.2	0.03

Current Limitations and Abilities in Identification of Toxicants in Sediments

TIEs depend upon both biological testing and chemical manipulations in order to achieve results. The combined use of these two approaches, in part, is what makes these methods so effective. However, TIEs are also affected by the limitations associated with both of these approaches. As toxicity assessment is the basis for TIEs, unless toxicity is assessed correctly, i.e., is relevant or correlated with field toxicity, final TIE results may not be relevant. Our acute toxicity tests are generally not sensitive to bioaccumulative toxicants or toxicants that have chronic effects. In addition, limitations of a TIE are often linked to our inability to chemically discriminate between like organic compounds. That is, it is relatively easy to distinguish metals from organics; however, it is much more difficult to distinguish between organic compounds that are chemically similar. For example, it would be very difficult to distinguish if the cause of toxicity in an oil spill was from a fuel oil # 6 spill or background diesel oil from commercial traffic, as they have the same mode of action and are chemically similar. TIEs can relatively easily confirm cause(s) of toxicity when common pollutants with known chemical formulas such as PCBs, PAHs, pesticides and herbicides are responsible; unusual compounds such as microbial degradation products may be very difficult to identify and confirm. Currently we can fairly easily characterize and identify chemical classes of toxicants in interstitial waters. We can also identify specific toxicants depending upon the chemical and the amount of effort we are willing to invest in the identification. For example, at our EPA laboratory, identification of a known pesticide or "typical" pollutant would be relatively easy, however identification of a pharmaceutical breakdown product would be very difficult. For chemistry laboratories that specialize in pharmaceutical products, the opposite may be true.

Research Needs and Conclusions

Toxicity Identification and Evaluation methods are useful for identifying causes of acute toxicity in sediments and dredged materials. These methods can be relatively easily incorporated into the assessment of dredged materials (Ankley et al., 1992b). Issues that need to be addressed before complete acceptance of interstitial water methods occurs include resolving questions regarding oxidation of metals, changes in interstitial water pH, sorption of high log K_{ow} organics and elimination of other routes of exposure. Whole sediment methods for Phase II, (Identification) and Phase III (Confirmation) need to be developed and verified in the laboratory. Both whole sediment and interstitial water TIEs need to be field validated. As with any toxicity test, extrapolation from laboratory to field toxicity and vice-versa may be problematic but research exists that indicates that laboratory toxicity testing is indicative of field conditions (Swartz et al., 1994; Swartz et al., 1986b; Swartz et al., 1986a; Swartz et al., 1982). Field validation includes confirmation that the toxicant identified in the laboratory is the same toxicant causing adverse ecological changes in the field. A potential experimental design could include selection of an impaired benthic community and identification of the toxicant(s) in sediments from the affected area using whole sediment and interstitial TIEs. Once a toxicant is identified, re-creation of the impaired benthic community by addition of the suspected toxicant in a mesocosm would be strong evidence that the toxicant identified by the TIE method is the same toxicant causing damage to the benthic community.

Using interstitial water TIE methods over the last decade we have learned that there is no one predominant cause of toxicity in sediments; ammonia, metals and organics all play a role in fairly even proportions, except in marine sediments where acute metal toxicity appears to be a minor factor. Interstitial water TIEs have also taught us that within a single sediment usually more than one toxicant is bioavailable and that ammonia may play a larger role in sediment toxicity than was previously expected. Information we are just starting to gather using whole sediment TIE methods implies that organics play the major role in the toxicity of marine sediments.

Disclaimer

This is EPA/NHEERL-AED 01-015. This paper has been technically reviewed by AED; however, it has not been subject to agency-wide peer review and therefore does not necessarily represent the views of the US Environmental Protection Agency. No official endorsement of any aforementioned products should be inferred.

References

Adams, W. J., R. A. Kimerle and R. G. Mosher. (1985). Aquatic safety assessment of chemicals sorbed to sediments. R. D. Cardwell, R. Purdy and R. C. Bahner eds. Aquatic Toxicology and Hazard Assessment: Seventh Symposium. STP 854. 854. *American Society for Testing Materials*. Philadelphia. 429-453.

Adamsson, M., G. Dave, L. Forsberg and B. Guterstam. (1998). Toxicity identification evaluation of ammonia, nitrite and heavy metals at the Stensund Wastewater Aquaculture plant, Sweden. *Water Sci. Techn.* **38**, 151-157.

Amato, J. R., D. I. Mount, E. J. Durhan, M. T. Lukasewycz, G. T. Ankley and E. D. Robert. (1992). An example of the identification of diazinon as a primary toxicant in an effluent. *Environ. Toxicol. Chem.* **11**, 209-216.

Anderson, J., W. Birge, J. Gentile, J. Lake and J. J. Rodgers. (1987). Biological Effects. Bioaccumulation and ecotoxicology of sediment associated chemicals In *Fate and Effects of Sediment-Bound Chemicals in Aquatic Systems.* (K. L. Dickson, A. W. Maki and W. A. Brungs eds), pp. 267-296. Pergamon Press, NY.

Ankley, G. T. and L. P. Burkhard. (1992). Identification of surfactants as toxicants in a primary effluent. *Environ. Toxicol. Chem.* **11**, 1235-1248.

Ankley, G. T., R. A. Hoke, J. Giesy and P. V. Winger. (1989). Evaluation of the toxicity of marine sediments and dredge spoils with the Microtox Bioassay. *Chemosphere* **18**, 2069-2075.

Ankley, G. T., A. Katko and J. Arthur. (1990). Identification of ammonia as an important sediment-associated toxicant in the lower Fox River and Green Bay, Wisconsin. *Environ. Toxicol. Chem.*. **9**, 312-322.

Ankley, G. T., M. K. Schubauer-Berigan, J. R. Dierkes and M. T. Lukasewycz. (1992). Sediment toxicity identification evaluation: Phase I (characterization), Phase II (identification) and Phase III (confirmation) modifications of effluent procedures. US Environmental Protection Agency/Environmental Research Laboratory-Duluth, MN. Draft-technical report. EPA 08-91.

Ankley, G. T., M. K. Schubauer-Berigan and R. A. Hoke. (1992). Use of toxicity identification evaluation techniques to identify dredged material disposal options: a proposed approach. *Environ. Manage* **16**, 1-6.

Bailey, H. C., J. L. Miller, M. J. Miller and B. S. Dhaliwal. (1995). Application of toxicity identification procedures to the echinoderm fertilization assay to identify toxicity in a municipal effluent. *Environ. Toxicol. Chem.* **14**, 2181-2186.

Bailey, R. C., K. E. Day, R. H. Norris and T. B. Reynoldson. (1995). Macroinvertebrate community structure and sediment bioassay results from nearshore areas of North American Great Lakes. *J. Great Lakes Res.* **21**, 42-52.

Boucher, A. M. and M. C. Watzin. (1999). Toxicity identification evaluation of metal-contaminated sediments using an artificial pore water containing dissolved organic carbon. *Environ. Toxicol. Chem.* **18**, 509-518.

Burgess, R. M. (2000). Characterizing and identifying toxicants in marine waters: a review of marine Toxicity Identification Evaluations (TIEs). *Int. J. Environ. Pollut.* **13**, 2-33.

Burgess, R. M., M. G. Cantwell, M. C. Pelletier, K. T. Ho, J. R. Serbst, H. F. Cook and A. Kuhn. (2000). Development of a Toxicity Identification Evaluation (TIE) procedure for characterizing metal toxicity in marine sediments. *Environ. Toxicol. and Chem.* **19**, 982-991.

Burgess, R. M., K. T. Ho, M. D. Tagliabue, A. Kuhn, R. Comeleo, P. Comeleo, G. Modica and G. E. Morrison. (1995). Toxicity characterization of an industrial and a municipal effluent discharging to the marine environment. *Mar. Poll. Bull.* **30**, 524-535.

Burkhard, L. P. and G. T. Ankley. (1989). Identifying toxicants: NETAC's toxicity-based approach. *Environ. Sci. Technol.* **23**, 1438-1443.

Burkhard, L. P. and J. J. Jenson. (1993). Identification of ammonia, chlorine, and diazinon as toxicants in a municipal effluent. *Arch. Environ. Contam. Toxicol.* **25**, 506-515.

Burton, G. A. J., C. Hickey, T. DeWitt, D. Morrison, D. Roper and M. Nipper. (1996.). *In situ* toxicity testing: Teasing out the environmental stressors. *SETAC News* from the Society of Environmental Toxicology and Chemistry. Pensacola, FL . 20-22.

Chapman, P. M. (1988). Marine sediment toxicity tests. J. L. Lichtenberg, J. A. Winter, C. I. Weber and L. Fradkin eds. Chemical and biological characterization of municipal sludges, sediments, dredge spoils and drilling muds. 976. *Amer. Soc. for Testing and Materials* STP. Philadelphia. 391-402.

Ertan-Unal, M., A. B. Gelderloos and J. S. Hughes. (1998). A toxicity reduction evaluation for an oily waste treatment plant exhibiting episodic effluent toxicity. *Sci. Total Environ.* **218**, 141-152.

Giesy, J. P. and R. A. Hoke. (1989). Freshwater sediment toxicity bioassessment: Rationale for species selection and test design. *J. Great Lakes Res.* **15**, 539-569.

Giesy, J. P. and R. A. Hoke. (1990a). Freshwater sediment quality criteria: Toxicity bioassessment. In *Sediments: Chemistry and Toxicity of In-Place Pollutants.* (R. Baudo, J. Giesy and H. Muntau eds.), pp. 265-348. Lewis Publishers, Inc., Boca Raton, FL.

Gupta, G. and M. Karuppiah. (1996a). Toxicity study of a Chesapeake Bay tributary - Wicomico River. *Chemosphere* **32**, 1193-1215.

Gupta, G. and M. Karuppiah. (1996b). Toxicity identification of Pocomoke river porewater. *Chemosphere* **33**, 939-960.

Hansen, D. J., W. J. Berry, J. D. Mahony, W. S. Boothman, D. M. Di Toro, D. L. Robson, G. T. Ankley, D. Ma, Q. Yan and C. E. Pesch. (1996). Predicting the toxicity of metal-contaminated field sediments using interstitial concentration of metals and acid-volatile sulfide normalizations. *Environ. Toxicol. Chem.* **15**, 2080-2094.

Hartwell, S. I., C. E. Dawson, E. Q. Durell, R. W. Alden, P. C. Adolphson, D. Wright, G. M. Coelho, J. A. Magee, S. Ailstock and M. Norman. (1997). Correlation of measures of ambient toxicity and fish community diversity in Chesapeake Bay, USA, tributaries----urbanizing watersheds. *Environ.Toxicol. Chem.* **16**, 2556-2567.

Hatakeyama, S. and N. Yokoyama. (1997). Correlation between overall pesticide effects monitored by shrimp mortality test and change in macrobenthic fauna in a river. *Ecotox. Environ. Safety* **36**, 148-161.

Ho, K. T., R. M. Burgess, M. Pelletier, M. Cantwell, D. R. Mount, E. Leonard, T. Norberg-King, C. West and M. Lukasewycz. (1997). Development of Whole Sediment Toxicity Identification Methods. *Conference Proceedings:* 18th Annual Society of Environmental Toxicology and Chemistry , November, 1997 San Francisco, CA.

Ho, K. T., R. M. Burgess, M. Pelletier, J. Serbst, A. Kuhn, M. Cantwell and S. Ryba. (2000). Sediment toxicity identification and evaluations: New test methods. What's been done? Where are we going? *Conference Proceedings:* 6th Annual NAC/SETAC Meeting , April 2000, Newport RI.

Ho, K. T., A. Kuhn, M. C. Pelletier, R. M. Burgess and A. Helmstetter. (1999). Use of *Ulva lactuca* to distinguish pH dependent toxicants in marine waters and sediments. *Environ. Toxicol. Chem.* **18**, 207-212.

Ho, K. T., R. A. McKinney, A. Kuhn, M. C. Pelletier and R. M. Burgess. (1997). Identification of acute toxicants in New Bedford Harbor sediments. *Environ. Toxicol. Chem.* **16**, 551-558.

Hunt, J. W., B. S. Anderson, B. M. Phillips, R. S. Tjeerderma, H. M. Puckett and V. deVlaming. (1999). Patterns of aquatic toxicity in an agriculturally dominated coastal watershed in California. *Agri. Ecosys. Environ.* **75**, 75-91.

Jin, H., X. Yang, D. Yin and H. Yu. (1999). A case study on identifying the toxicant in effluent discharged from a chemical plant. *Marine Poll. Bull.* **39**, 122-125.

Jin, H., X. Yang, H. Yu and D. Yin. (1999). Identification of ammonia and volatile phenols as primary toxicants in a coal gasification effluent. *Bull. Environ. Contamin. Toxicol.* **63**, 399-406.

Jop, K. M. and A. M. Askew. (1994). Toxicity identification evaluation using a short-term chronic test with *Ceriodaphnia dubia. Bull. Environ. Contam. Toxicol.* **53**, 91-97.

Karuppiah, M. and G. Gupta. (1996). Impact of point and nonpoint source pollution on pore waters of two Chesapeake Bay tributaries. *Ecotoxicol. Environ. Safety* **35**, 81-85.

Kuhn, A., K. T. Ho, M. Pelletier and R. M. Burgess. (1995). Phase I Toxicity Identification Evaluation (TIE) for New York/New Jersey Harbor Estuary Sediments. Atlantic Ecology Division, US Environmental Protection Agency. Draft.

Miller, D. C., S. Poucher, J. A. Cardin and D. Hansen. (1990). The acute and chronic toxicity of ammonia to marine fish and a mysid. *Arch. Environ. Contam. Toxicol.* **19**, 40-48.
Norberg-King, T., E. Durhan, G. T. Ankley and E. Robert. (1991). Application of Toxicity Identification and Evaluation procedures to the ambient waters of the Coulsa Basin Drain, California. *Environ. Toxicol. Chem.* **10**, 881-900.

Riveles, K. and R. M. Gersberg. (1999). Toxicity identification evaluation of wet and dry weather runoff from the Tijuana River. *Bull. Environ. Contam. Toxicol.* **63**, 625-632.

Rumbold, D. and S. Snedaker. (1999). Sea-surface microlayer toxicity off the Florida Keys. *Mar. Environ. Res.* **47**, 457-472.

Schubauer-Berigan, M. K., J. R. Amato, G. T. Ankley, S. E. Baker, L. P. Burkhard, J. R. Dierkes, J. J. Jenson, M. T. Lukasewycz and T. J. Norberg-King. (1993). The behavior and identification of toxic metals in complex mixtures: examples from effluent and sediment pore water toxicity identification evaluations. *Arch. Environ. Contam. Toxicol.* **24**, 298-306.

Schubauer-Berigan, M. K. and G. T. Ankley. (1991). The contribution of ammonia, metals and nonpolar organic compounds to the toxicity of sediment interstitial water from an Illinois River tributary. *Environ. Toxicol. Chem.* **10**, 925-939.

Sprang, P. A. V. and C. R. Janssen. (1997). Identification and confirmation of ammonia toxicity in contaminated sediments using a modified toxicity identification evaluation approach. *Environ. Toxicol. Chem.* **16**, 2501-2507.

Sprang, P. A. V., C. R. Janssen, M. Sabayasachi, F. Benijts and G. Persoone. (1996). Assessment of ammonia toxicity in contaminated sediments of the Upper Scheldt (Belgium): the development and application of toxicity identification evaluation procedures. *Chemosphere* **33**, 1967-1974.

Steidl-Pulley, T., D. W. R. Nimmo and J. D. Tessari. (1998). Characterization of toxic conditions above Wilson's Creek National Battlefield Park, Missouri. *J. Am. Water Res. Assoc.* **34**, 1087-1098.

Swartz, R. C., F. A. Cole, J. O. Lamberson, S. P. Ferraro, D. W. Schults, W. A. DeBen, H. I. Lee and R. J. Ozretich. (1994). Sediment toxicity, contamination and amphipod abundance at a DDT and dieldrin contaminated site in San Francisco Bay. *Environ. Toxicol. Chem* **13**, 949-962.

Swartz, R. C., F. A. Cole, D. W. Schults and W. A. DeBen. (1986b). Ecological changes in the Southern California Bight near a large sewage outfall: benthic conditions in 1980 and 1983. *Mar. Ecol. Prog. Ser.* **31**, 1-13.

Swartz, R. C., F. A. Cole, D. W. Schultz and W. A. DeBen. (1986a). Ecological changes on the Palos Verdes Shelf near a large sewage outfall: 1980 - 1983. *Mar. Ecol. Prog. Ser.* **31**, 1-13.

Swartz, R. C., W. A. DeBen, K. A. Sercu and J. O. Lamberson. (1982). Sediment toxicity and the distribution of amphipods in Commencement Bay, Washington, USA. *Mar. Poll. Bull.* **13**, 359-364.

Swartz, R. C., P. F. Kemp, D. W. Schultzs, G. R. Ditsworth and R. J. Ozretich. (1989). Acute toxicity of sediments from Eagle Harbor, Washington, to the infaunal amphipod *Rhepoxynius Abronius*. *Environ. Tox. Chem.* **8**, 215-222.

US Environmental Protection Agency. (1991). Water Quality Criteria Summary Chart. U.S. Environmental Protection Agency- Office of Science and Technology, Washington DC. Washington, D.C. Summary.

Wells, M. J. M., A. J. J. Rossano and E. C. Roberts. (1994). Textile wastewater effluent toxicity identification evaluation. *Arch. Environ. Contam. Toxicol.* **27**, 555-560.
Wenholz, M. and R. Crunkilton. (1995). Use of toxicity identification evaluation procedures in the assessment of sediment pore water toxicity from an urban stormwater retention pond in Madison, Wisconsin. *Bull. Environ. Contam. Toxicol.* **54**, 676-682.

Williams, L. G., P. M. Chapman and T. C. Ginn. (1986). A comparative evaluation of marine sediment toxicity using bacterial luminescence, oyster embryo and amphipod sediment bioassays. *Marine Environ. Res.* **19**, 225-249.

Toxicity Testing, Risk Assessment, and Options for Dredged Material Management

WAYNE R. MUNNS, JR.†, WALTER J. BERRY† AND THEODORE H. DEWITT‡
†*U.S. Environmental Protection Agency, NHEERL Atlantic Ecology Division, 27 Tarzwell Drive, Narragansett, RI 02882 USA*
‡*U.S. Environmental Protection Agency, NHEERL Western Ecology Division, 2111 S.E. Marine Science Drive, Newport, OR 97365 USA*

Abstract

Programs for evaluating proposed discharges of dredged material into waters of the United States specify a tiered testing and evaluation protocol that includes performance of acute and chronic bioassays to assess toxicity of the dredged sediments. Although these evaluations reflect the toxicological risks associated with disposal activities to some degree, analysis activities are limited to the sediments of each dredging project separately. Cumulative risks to water column and benthic organisms at and near the designated disposal site are therefore difficult to assess. An alternate approach is to focus attention on the disposal site, with the goal of understanding more directly the risks of multiple disposal events to receiving ecosystems. Here we review current US toxicity testing and evaluation protocols, and describe an application of ecological risk assessment that allows consideration of the temporal and spatial components of risk to receiving aquatic ecosystems. When expanded to include other disposal options, this approach can provide the basis for holistic management of dredged material disposal.

Keywords: dredged material disposal; toxicity testing; risk assessment; risk management.

Introduction

Dredging and disposal of marine and freshwater sediments often is required to maintain navigational channels or to remediate contaminated waterways. Dredging projects may also be conducted to provide material for beach nourishment projects, to create or expand wetlands, and for other beneficial uses. In the United States, The Clean Water Act (CWA) governs discharges of dredged material into inland waters and surrounding environs, including coastal waters and all waters landward of the baseline of the territorial sea. The Marine Protection, Research, and Sanctuaries Act (MPRSA) governs the transportation of dredged material seaward of the baseline (in ocean waters) for disposal. Section 404 of the CWA requires that proposed discharges of sediments into aquatic systems must: a) present the least environmentally damaging, practicable management alternative; b) comply with established legal standards; 3) not result in significant degradation of the aquatic environment; and 4) utilize all practicable means to minimize adverse environmental impacts. Lead responsibility for developing guidelines and environmental criteria for evaluating proposed discharges is shared by the US Environmental Protection Agency (US EPA) and the US Army Corps of Engineers (US ACE).

Testing and analysis protocols used in the US to evaluate whether guideline criteria are met for dredged material disposal in inland and open ocean waters are described in the "Inland Testing Manual" (ITM) (US EPA and US ACE, 1998) and the "Green Book" (US EPA and US ACE, 1991), respectively. These protocols are designed to support informed management decisions about the placement of dredged sediments through chemical, physical, and biological evaluations of the dredged material. They specify a tiered testing and evaluation approach that includes performance of bioassays to assess toxicity of the dredged sediments to species inhabiting the disposal site. Both water column and bedded sediment toxicity tests are employed, and sediment bioaccumulation tests are indicated when bioaccumulative chemicals are present in the dredged material at sufficiently high levels. Early tier toxicity tests focus on acute responses, whereas later tier testing (when required) can reflect longer test exposures and evaluate sublethal endpoints. In all cases, the toxicity of dredged material proposed for disposal is assessed against toxicity measured in a suitable reference sediment. As part of this paper, we will describe toxicity tests currently used in dredged material evaluations, and will suggest ways to improve their value to the evaluation process.

Although current US evaluation protocols incorporate both exposure (sediment chemistry and bioaccumulation) and effects (toxicity) components, and therefore reflect to some degree the toxicological risks associated with disposal activities, the focus of analysis activities is limited to the sediments of each dredging project separately. Thus cumulative risks to water column and benthic organisms at and near the designated disposal site are difficult to assess. An alternate approach is to focus attention on the disposal site, with the goal of understanding more directly the risks of multiple disposal events to receiving ecosystems. Here we review the US federal practices for evaluating dredged materials for aquatic disposal, and then describe an alternative management framework based on ecological risk assessment which has two principal advantages over previous approaches. First, it allows specification of receptors and assessment endpoints at the disposal site, recognizes variation in exposure conditions (including the discharge of material from different disposal projects), and considers the temporal and spatial components of risks in the context of the receiving ecosystem. Second, the relative risks and benefits of several management options (i.e., aquatic disposal, upland disposal, treatment, no action, etc.) can be evaluated simultaneously. When applied to sediment management issues other than dredging, this approach permits holistic assessment and management of in-place sediments generally.

Tiered Evaluation Protocol for Evaluating Dredged Materials for Aquatic Disposal in the U.S.

The US is a party to the London (Ocean Dumping) Convention of 1972, which governs disposal activities in the world's oceans. The London Convention is implemented in the US through the MPRSA. The 1996 Protocol to the Convention describes a framework for deciding whether dredged sediments are suitable for aquatic disposal which forms the basis for the US evaluation protocols. By design (US EPA and US ACE, 1992), the ITM and Green Book differ little in their basic approaches to evaluating project sediments, and both utilize a tiered protocol intended to minimize unnecessary testing and evaluation (and thereby costs). Evaluations typically begin with an assessment of existing information, and proceed through activities arrayed by increasing assessment intensity and cost until there is sufficient information about the project sediment to make a factual determination of its acceptability for ocean disposal. Because of the similarly in the ITM and Green Book evaluation designs, the following description focuses on the ITM (US EPA and US ACE, 1998) because it is the more recent (Figure 1).

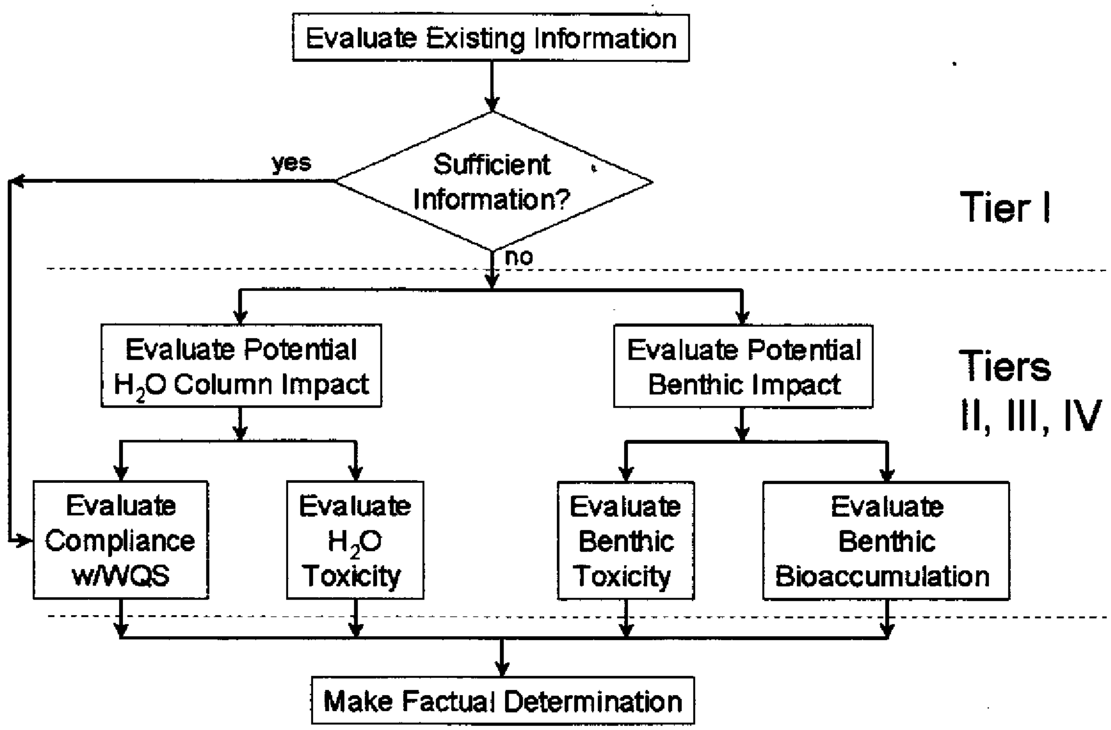


Fig. 1. Overview of the tiered approach to evaluating the potential for environmental impact of aquatic disposal of dredged material (modified from USEPA and USACE, 1998).

The initial activity of assessing existing information (Tier I) allows for dredging projects to be exempted from additional evaluation and testing if that information clearly indicates a lack of potential for adverse environmental impact, or to be excluded from ocean disposal if a potential for adverse impact is obvious. This tier involves assessment of any and all information available concerning the potential for contamination of project sediment and potential adverse effects on aquatic biota. Included is knowledge of possible nearby sources and transport pathways of chemical and other stressors, data from previous physical, chemical, and biological tests, and monitoring information from past disposal activities involving the material. If information available for Tier I evaluation is insufficient to make a factual determination, or the potential for adverse impact is unclear, the evaluation typically proceeds to subsequent tiers.

In the ITM, Tier II assessments focus on the chemistry of the project sediment, addressing contaminant exposure to water column and sediment-dwelling aquatic organisms through use of simple screening tools (Figure 2). To evaluate potential impact to water column species, elutriate tests and simple models are employed to estimate concentrations of dissolved contaminants. These values are compared with State water quality standards in the fashion of hazard quotients; concentrations exceeding water quality standards indicate the potential for adverse impact (toxicity) to water column species. Calculations of bioaccumulation potential relative to that of an appropriate reference sediment, based on chemical partitioning theory, provide information about potential impacts to benthic species. Continuing uncertainty about the magnitude of exposure and its relationship to adverse effects levels indicates the need for additional evaluation.

Actual toxicity testing of project sediments occurs in Tiers III and IV of the ITM protocol (Figures. 1 and 2). Generic evaluations of water column and sediment toxicity are conducted in Tier III using standardized test protocols and species whose responses are intended to represent those of aquatic biota at the disposal site. Quasi-steady state bioaccumulation tests also are performed to support understanding of contaminant bioavailability and to inform considerations of possible trophic transfer. Tier IV assessments are employed only when there continues to be uncertainty about the potential for adverse impact or the lack thereof. These are considered case-specific assessments that are designed to reduce uncertainties remaining after earlier tier assessments. Guidance for toxicity testing in the final two tiers of the evaluation are described below.

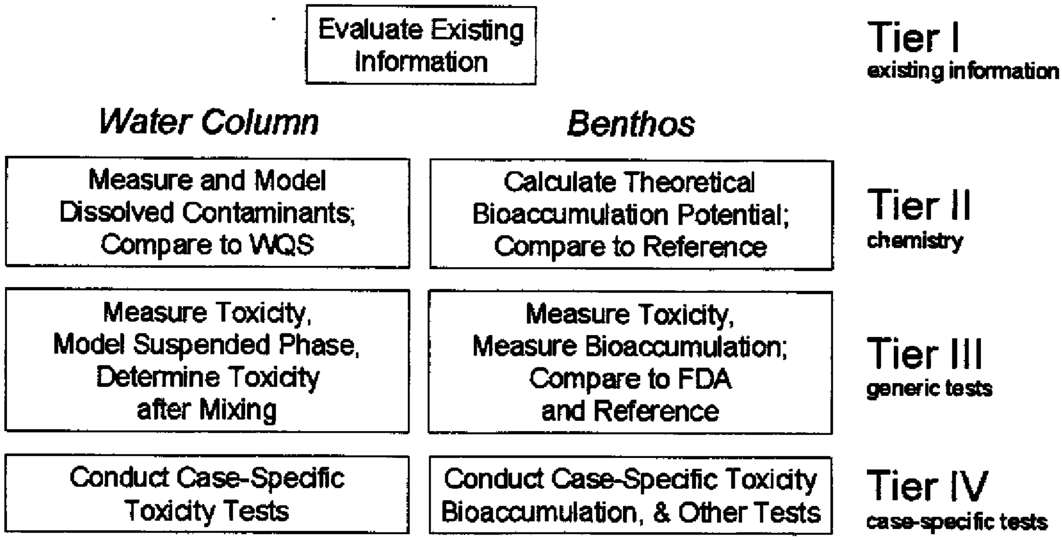


Fig. 2. Tiered testing activities recommended in the Inland Test Manual to support evaluations of the potential for environmental impact in the water column and sediment (modified from USEPA and USACE, 1998).

Toxicity Testing

Current approach and guidance

Both the ITM and Green Book provide guidance and recommendations for performing toxicity tests. This guidance includes considerations for identifying and selecting test species, lists of suitable species and tests that have been developed to support these and other toxicity evaluations, and descriptions of testing designs and conditions. The evaluation manuals also describe appropriate data analysis approaches, and offer guidance for interpreting and presenting test results. Considerations for selecting appropriate reference sites for statistical comparisons of toxicity also are described.

The evaluation protocol involves water column and sediment testing. Both the ITM and Green Book offer candidate test species and testing procedures appropriate for the environmental medium (sediment or water) to be tested and the disposal environment (fresh or saline). In aggregate, a fairly large number of test species is available, representing five phyla and including 21 species of crustacean, 13 species of fish, and seven species of bivalves (Table 1). Tests for additional species also have been developed and are reported in the open literature. In practice, however, the number of species used routinely is much smaller, with arthropods, annelids, and molluscs being the most commonly used. Of these, arthropods generally are the most sensitive in acute tests. Often, only specific life stages of a species are used in these toxicity tests. For example, larval stages of mussels and oysters are used to test water toxicity. The bulk of tests recommended by the ITM and Green Book are acute tests, typically measuring mortality in test organisms exposed over short time spans. In cases where test methods are appropriately developed, a particular species might also be used in chronic tests, and endpoints other than mortality (*e.g.*, individual growth, reproduction) can be measured.

Table 1. Summary of species appropriate for testing potential impacts of dredged material disposal (combined from US EPA and US ACE, 1991 and 1998).

Taxonomic Group		Test Species	Medium Tested		Salinity Tolerance[a]
			water	sediment	
crustaceans	copepod	*Acartia* sp	✓		M
	mysid shrimp	*Americamysis* sp.[b]	✓	✓	M
		Neomysis americana	✓	✓	M
		Holmesimysis costata	✓	✓	M
	grass shrimp	*Palaemonetes* sp.	✓	✓	M,E
	sand shrimp	*Crangon* sp.		✓	M,E
	shrimp	*Farfante penoeus*	✓	✓	M
		Pandalus sp.	✓	✓	M
		Sicyonia ingentis		✓	M
	crab	*Callinectes sapidus*	✓	✓	M,E
		Cancer sp.	✓	✓	M,E
	amphipod	*Ampelisca* sp.		✓	M
		Rhepoxynius sp.		✓	M
		Eohaustorius sp.		✓	M,E
		Grandiderella japonica		✓	M
		Corophium insidiosum		✓	M,E
		Leptocheirus plumulosus		✓	M,E
		Hyalella azteca		✓	F,E
	cladoceran	*Daphnia magna*	✓		F
		Ceriodaphnia dubia	✓		F
insects	midge	*Chironomus tentans*		✓	F
		C. riparius		✓	F
	mayfly	*Hexagenia limbata*		✓	F
molluscs	mussel	*Mytilus edulis*[c]	✓		M,E
		Anodonta imbecillis		✓	F
	oyster	*Ostrea* sp.	✓		M,E
		Crassostrea sp.	✓		M,E
	clam	*Yoldia limatula*		✓	M

Taxonomic Group		Test Species	Medium Tested		Salinity Tolerance[a]
		Protothaca staminea		✓	M
		Tapes japonica		✓	M
annelids	burrowing polychaete	*Nereis* sp.		✓	M
		Neanthes arenaceodentata		✓	M
		Nephthys sp.		✓	M
		Glycera sp.		✓	M
		Arenicola sp.		✓	M
		Abarenicola sp.		✓	M
	oligochaete	*Pristina leidyi*		✓	F
		Tubifex tubifex		✓	F
		Lumbriculus variegates		✓	F
echinoderms	sea urchin	*Strongylocentrotus purpuratus*		✓	M
		Lytechinus pictus	✓		M
	sand dollar	*Dendraster* sp	✓		M
fish	silversides	*Menidia* sp.	✓		M,E
	shiner perch	*Cymatogaster aggregata*	✓		M,E
	sheepshead minnow	*Cyprinodon variegates*		✓	M,E
	fathead minnow	*Pimephales promelas*	✓		F
	pinfish	*Lagodon rhomboids*	✓		M
	spot	*Leiostomus xanthurus*	✓		M
	sand dab	*Citharicthys stigmaeus*	✓		M
	grunion	*Leuresthes tenuis*	✓		M
	dolphinfish	*Coryphaena hippurus*	✓		M
	arrow gobi	*Clevelandia ios*		✓	M
	bluegill	*Lepomis macrochirus*	✓		F
	channel catfish	*Ictalurus punctatus*	✓		F
	rainbow trout	*Oncorhynchus mykiss*	✓		F,E

[a] following ITM (US EPA and US ACE, 1991) classification when given, F = freshwater, salinity ≤ 1‰; E = estuarine, salinity 1-25‰; M = marine, salinity ≥ 25‰.
[b] formerly *Mysidopsis*.

To help ensure the tests provide information useful for making a factual determination of the sediment's suitability for aquatic disposal, the evaluation manuals offer guidance for selecting among potential test species. A number of factors are offered for consideration, including the sensitivity of the species (and life stage tested) to contaminants in the project sediments, its degree of phylogenetic and ecological relatedness to receptors at the disposal site, its preferences and tolerance to the particle size makeup of the test sediment, and so on. Consideration of these factors helps to establish each test species as a surrogate for organisms living at the disposal site. Both the inland and ocean waters protocols recommend that three species representing different phyla (when possible) be tested for water column effects, and that three different "life history strategies", or perhaps more appropriately, three ecological life styles (filter feeding, deposit feeding, burrowing) be represented by species used to assess sediment effects (US EPA and US ACE, 1991, 1998).

Improving testing methods and data extrapolation

Given the wide range of test species available to support the evaluation process, it is likely that species can be selected that are fairly representative of those inhabiting the designated disposal site, at least with respect to ecological life style and taxonomy. However, the conditions of the tests themselves, and the toxicity endpoints measured, require substantial extrapolation to be representative of actual disposal events and potential adverse effects. This is not a shortcoming limited to dredged material testing protocols, but probably is true of most toxicity testing approaches used to assess real-world ecological risk. Keeping in mind the objective of predicting possible ecological effects associated with aquatic disposal in support of the evaluation process, we offer recommendations for improving two interrelated aspects of the testing approach: testing methods and extrapolation of test results.

The majority of tests used in dredged material evaluations are of a short-term nature (i.e., acute), at least in Tier III assessments. They therefore provide information most relevant to one-time, transient exposures. Depending upon the specific mechanisms(s) of toxic effect of contaminants in the dredged material, test exposure durations may be inadequate to elicit more subtle, but still important, effects associated with chronic bedded sediment exposures or reoccurring water column exposures (see Munns and Paul, 1987). This suggests the need to add to the suite of toxicity tests those that evaluate effects evoked by chronic exposures, either through increasing tests durations and expanding the range of endpoints measured, or through developing short-term tests that evaluate endpoints that are correlated with longer-term effects. In the latter case, short-term estimators of chronic response can enhance the value of toxicity testing in Tier III testing while keeping testing costs to a minimum. Currently, chronic testing is viewed as a Tier IV activity that is needed only in special cases (US EPA and US ACE, 1991, 1998); subtle effects may therefore be missed in to course of "routine" evaluations.

Enhanced realism of the toxicity evaluation with respect to actual field exposures could also be achieved by developing understanding of the concentration-time relationships necessary to extrapolate data obtained through short-term, cost-effective testing. Such research would not be part of the evaluation process *per se*, but rather could be applied in the interpretation of routine testing results. Promising results that support this approach have been communicated by Mancini (1983) and others. Additionally, *in situ* toxicity testing can provide a realistic assessment of the toxicity of in-place sediments (be they at the source or disposal sites) (DeWitt et al. 1999; Burton et al. 1996), although *in situ* methods also have limitations (DeWitt et al. 1996).

The endpoints measured in tests used in the evaluation process reflect effects on the demographic characteristics of individuals. In the simplest case, the rate of mortality measured in, say, a 10-day amphipod test may correlate directly with the probability that an individual amphipod would die if exposed to the sediment. However, the broader ecological consequences of this effect, in terms of populations of amphipods and benthic communities at the disposal site, are not always apparent. One potentially valuable approach to address this shortcoming is to develop and employ tools and models that permit extrapolation of the responses measured at the level of the individual to those of populations and communities. Examples of this approach as applied specifically to dredged material assessments are described by Gentile *et al.* (1987) and

Scott and Redmond (1989), and for species used in dredged material toxicity tests by Kuhn *et al.* (2000, 2001). Similarly, incorporation of test endpoints that can be related to other important ecological processes and functions would enhance the ecological relevance of testing in the evaluation process (see Chapman, this issue).

Other recommendations for ways to enhance the value of toxicity testing in the dredged material evaluation process have been offered by Peddicord *et al.* (1997), and by Dillon (1993), Ingersoll (1995), and others for sediment evaluations in general. They include increasing the battery of available test species and methods, improving the interpretive and diagnostic power of tests (Swartz *et al.*, 1994, 1995; Ankley and Schubauer-Berigan, 1995; Ho *et al.*, this issue), and reducing methodological uncertainties in ecological risk assessment applications (Ingersoll *et al.*, 1997). Conceptually, these improvements can easily be incorporated into the current evaluation processes described in the ITM and Green Book.

Risk Assessment and Management of Dredged Material

The primary intent of the tiered evaluation approach is to inform the decision maker to limit potential adverse effects of aquatic disposal. It is clear that the current approach considers aspects of exposure to chemical, physical, and biological stressors present in the project sediment, and that it provides information about the potential effects of those stressors on receptors at the aquatic disposal site. In this regard, the current evaluation approach has some of the features of a risk assessment (Peddicord *et al.*, 1997; also see Solomon, this issue). However, we believe that explicit application of the risk assessment process to disposal evaluations could enhance their value in supporting dredged material management decisions.

Peddicord *et al.* (1997) describe some initial thinking about how to structure risk assessments of aquatic disposal of dredged material. They advocate an important shift from the project-by-project evaluation of the sediment itself as currently mandated by the London Convention to one that focuses on the disposal site. By shifting the context of the assessment from the dredged sediment to the receiving ecosystem, the ability to predict adverse environmental effects arguably is enhanced. In following the ecological risk assessment framework developed by the US EPA (1992, 1998), the features of their approach include identification of assessment endpoints relevant to the disposal site, identification of stressors associated with the dredged sediment and the disposal operation (*e.g.*, burial), and development of conceptual models that link stressors to assessment endpoints. Reflecting the fact that aquatic disposal sites usually receive sediments from more than one project, they describe a conceptual approach for characterizing exposure that weighs the contribution of individual project sediments to total bedded sediment exposure by the spatial extent of each project's "footprint" on the developing disposal mound. Peddicord *et al.* (1997) also promote explicit consideration of effects on assessment endpoints (ecological receptors at the disposal site) through extrapolation of measured responses obtained in toxicity tests and otherwise, and include both direct and possible indirect processes in characterizing ecological effects. By rigorously applying the concepts and approaches of ecological risk assessment, a factual determination can be made of the potential for adverse effects at the disposal site. Much of this thinking is reflected in the guidance for aquatic disposal evaluations released recently by the US ACE (Cura *et al.*, 1999).

Although they did not address it explicitly, the approach described by Peddicord *et al.* (1997) also begins to draw attention to potential risks beyond the immediate environs of the disposal site, in that unplanned exposures (unconfined dispersal and errant releases) are acknowledged in their conceptual risk models. Reinforcing this broader perspective, we propose that formal risk assessment of all facets of dredging operations, from dredging to placement and including transport, would improve the efficacy of the entire process of managing dredging materials. Risks associated with disposal alternatives, including aquatic (open or confined) and upland disposal, and beneficial uses such as creation of wetlands and beach nourishment, and various treatment alternatives all would be considered in the context of those receiving ecosystems and the assessments endpoints appropriate to each. The risk of leaving the sediment in place also could be considered relative to risks of other alternatives. In this manner, decisions about management of dredged materials could be made in a holistic fashion.

It is not difficult to extend this approach conceptually to all problems associated with management of in-place sediments. A crude decision tree, similar in form to those used currently in consideration of aquatic disposal of dredged material (*cf* Figure. 1), is depicted in Figure 3 to support development of this idea. Although it highlights navigational dredging, this figure reflects different pathways of assessment depending upon the impetus for the dredging activity. For instance, assessments of projects utilizing clean sand for beach nourishment likely would not include all of the evaluations and decision points that would be needed in remedial investigations of hazardous waste sites. Common to all pathways, however, is the assessment of risks associated with all phases of the dredging and disposal operation.

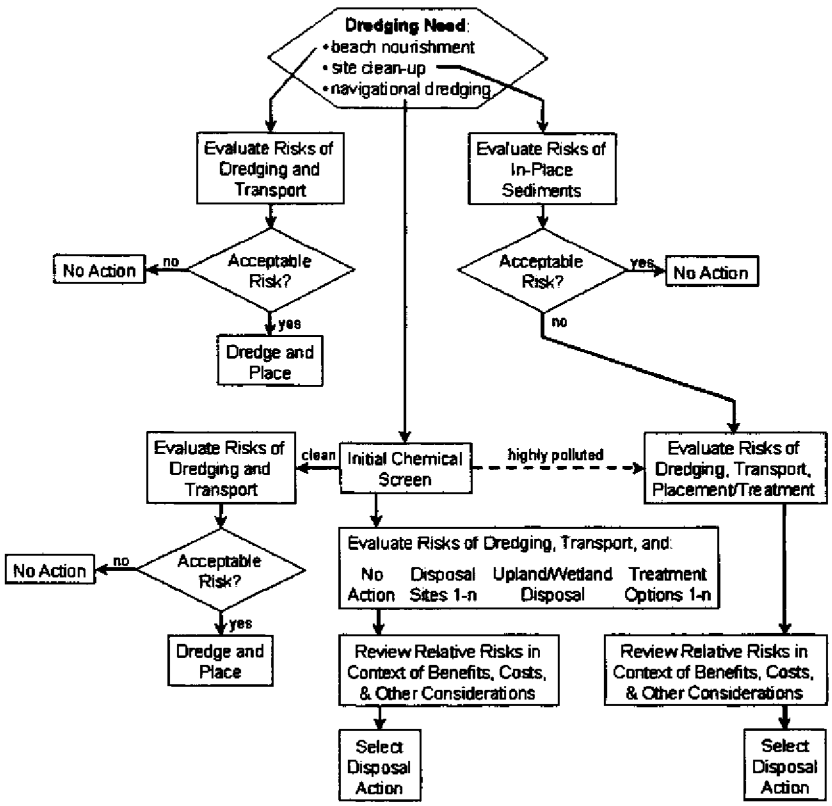


Fig. 3 Risk-based decision tree supporting holistic management of in-place sediments.

The navigational dredging pathway of the proposed decision tree starts with an initial chemical screen. Sediments classified as "clean", perhaps using approaches similar to those in Tiers I and II of the ITM (US EPA and US ACE, 1998), would be acceptable for aquatic disposal, and would need only to pass assessments of dredging and transport risks to go forward. Conversely, highly polluted sediments might be shunted to the "site clean-up" pathway of Figure 3 that requires extensive assessments of risk in a hazardous waste (*e.g.*, Superfund) context. Only those sediments about which substantial uncertainty remained after the initial screen would require further evaluation to understand potential risks to receiving ecosystems. In this regard, the process mirrors to some degree the initial tier of the ITM and Green Book evaluations.

A key feature of the navigational dredging pathway after the initial screen is the explicit focus on relative risks among all available options for managing the sediments. Included among these options are the various placement alternatives of aquatic disposal, disposal at upland sites, and recreation or extension of wetlands. Additionally, multiple sites may be available as aquatic disposal options. Two other options include treatments that would minimize risks associated with other disposal options, and a "no action" alternative of leaving the sediment in place. Ultimately, selection of the preferred disposal action (now broadly defined to include all alternatives) would be based on a comparison of the relative risks of each of the alternatives conducted in the context of the ecological, social, economic, and public health benefits and costs of each. Jurisdictional and regulatory considerations would determine who makes the selection, and it is important to note that the current regulatory approach does not accommodate multifaceted decision processes easily.

Additional work will be required to formalize the assessment procedures for various steps of the decision tree. Ideally, these procedures would be sufficiently rigorous to provide the information necessary for decision making, yet would be efficient and cost effective so as to minimize the burdens associated with comprehensive assessments. Many currently available methods for evaluating the exposure, toxicity, and ecological risk of contaminated sediments can be used in these assessment steps; it clearly is not necessary to start from ground-zero to set up a risk-based framework for managing dredged materials. However, accurate prediction of impacts to key assessment endpoints, including those at the population, community, and ecosystem levels of organization, will require continuation of research efforts currently underway at the US EPA and elsewhere. Despite these research and development needs, we believe that selection of disposal actions based on a full consideration of relative risks and in the context of the benefits, costs, and other considerations of each option should enhance the effectiveness of sediment management to the benefit of society.

Acknowledgments

Some of the ideas communicated herein grew out of earlier discussions with T. Dillon, T. Chase, R. Peddicord, J. McGrath, K. van de Gucthe, and W. van der Schalie during a sediment risk assessment workshop sponsored by the Society of Environmental Toxicology and Chemistry. Valuable discussions also were held with J. Gentile, J. Paul, and K.J. Scott. We thank B.S. Brown, K.T. Ho, N.I. Rubinstein, T.R. Gleason, S.C. Schimmel, and two anonymous reviewers for providing helpful reviews of the manuscript. The ideas and approaches described in the paper do not necessarily reflect the position or the policy of the US EPA, and no official endorsement should be inferred.

Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This is NHEERL contribution number AED-01-019.

References

Ankley, G.T. and Schubauer-Berigan, M.K. (1995) Background and overview of current sediment toxicity identification evaluation procedures. *J. of Aquat. Ecosys Health* **4**, 133-149.

Burton, G.A., Jr., Hickey, C.W., DeWitt, T.H., Roper, D.S., Morrisey, D.J. and Nipper, M.G. (1996) *In situ* toxicity testing: teasing out the environmental stressors. *SETAC News, Learned Discourses* **16**, 20-22.

Chapman, P. (this issue) Integrating toxicology and benthic ecology: Putting "eco" back into ecotoxicology. *Mar. Pollut. Bull.*

Cura, J.J., Heiger-Bernays, W., Bridges, T.S. and Moore, D.W. (1999) Ecological and human health risk assessment guidance for aquatic environments. Technical Report DOER-4, US Army Engineer Research and Development Center, Vicksburg, MS.

DeWitt, T.H., Morrisey, D.J., Roper, D.S. and Nipper, M.G. (1996) Fact or artefact: the need for appropriate controls in ecotoxicological field experiments. *SETAC News, Learned Discourses* **16**, 22-23.

DeWitt, T.H., Hickey, C.W., Morrisey, D.J., Nipper, M.G., Roper, D.S., Williamson, R.B, Van Dam, L. and Williams E.K. (1999) Do amphipods have the same concentration-response to contaminated sediment *in situ* as *in vitro*? *Environ. Toxicol. Chem.* **18**, 1026-1037.

Dillon, T.M. (1993) Developing chronic sublethal sediment bioassays: a challenge to the scientific community. In *Environmental Toxicology and Risk Assessment*, Vol. 2, STP 1216. (J.W. Gorsuch, F.J. Dwyer, C.G. Ingersoll and T.W. La Point, eds.), pp. 623-639. American Society for Testing and Materials, Philadelphia.

Gentile, J.H., Scott, K.J., Lussier, S.M. and Redmond, M.S. (1987) The assessment of Black Rock Harbor dredged material impacts on laboratory population responses. Technical Report D-87-3, US Army Engineer Waterways Experiment Station, Vicksburg, MS.

Ho, K.T., Burgess, R.M., Pelletier, M.C., Serbst, J.R., Ryba, S.A., Cantwell, M.G., Kuhn, A. and Raczelowski, P. (this issue) Identification and evaluation of stressors in toxic sediments and dredged materials. *Marine Pollution Bulletin.*

Ingersoll, C.G. (1995) Sediment toxicity tests. In *Fundamentals of Aquatic Toxicology*, 2nd edition. (G.M. Rand, ed.), pp. 231-256. Taylor and Francis, Washington, DC.

Ingersoll, C.G., Ankley, G.T., Baudo, R., Burton, G.A., Lick, W., Luoma, S.N., MacDonald, D.D., Reynoldson, T.B., Solomon, K.R., Swartz, R.C. and Warren-Hicks, W.J. (1997) Workshop summary report on uncertainty evaluation of measurement endpoints used in sediment ecological risk assessment. In *Ecological Risk Assessments of Contaminated Sediments*. (C.G. Ingersoll, T. Dillon and G.A. Biddinger, eds.), pp. 297-352. SETAC Special Publication Series, Society of Environmental Toxicology and Chemistry, Pensacola, FL.

Kuhn-Hines, A., Munns, Jr., W.R., Poucher, S., Champlin, D. and Lussier, S. (2000) Prediction of population-level response from mysid toxicity test data using population modeling techniques. *Environ. Toxicol. Chem.* **19**, 2364-2371.

Kuhn, A., Munns, Jr., W.R., Champlin, D., McKinney, R., Tagliabue, M., Serbst, J. and Gleason. T. (2001) Evaluation of the efficacy of extrapolation population modeling to predict the dynamics of *Americamysis bahia* populations in the laboratory. *Environ. Toxicol. Chem.* **20**, 213-221.

Mancini, J.L. (1983) A method for calculating effects, on aquatic organisms, of time varying concentrations. *Water Res.* **17**, 1355-1362.

Munns, W.R., Jr. and Paul, J.F. (1987) Dredged material plume dispersal in Central Long Island Sound. In *Aquatic Toxicology and Environmental Fate*, Vol. 11, ASTM STP 1007. (G.W. Suter II and M.A. Lewis, eds.), pp. 82-93. American Society for Testing and Materials, Philadelphia, PA.

Peddicord, R, T. Chase, T. Dillon, J. MacGrath, W.R. Munns, K. van de Gucthe, and W. van der Schalie. 1997. Navigational dredging. In *Ecological Risk Assessments of Contaminated Sediments.* (C.G. Ingersoll, T. Dillon and G.A. Biddinger, eds.), pp. 41-71. SETAC Special Publication Series, Society of Environmental Toxicology and Chemistry, Pensacola, FL.

Scott, K.J. and Redmond, M.S. (1989) The effects of a contaminated dredged material on laboratory populations of the tubiculous amphipod *Ampelisca abdita.* In *Aquatic Toxicology and Hazard Assessment*, Vol. 12, ASTM STP 1027. (U.M. Cowgill and L.R. Williams, eds.), pp. 289-303. American Society for Testing and Materials, Philadelphia, PA.

Solomon, K. (this issue). New concepts in ecological risk assessment: Where do we go from here? *Mar. Poll. Bull.*

Swartz, R.C., Cole, F.A., Lamberson, J.O., Ferraro, S.P., Schults, D.W., DeBen, W.A., Lee II, H., and Ozretich, J.R. (1994) Sediment toxicity, contamination, and amphipod abundance at a DDT and dieldrin-contaminated site in San Francisco Bay. *Environ. Toxicol. Chem.* **13**, 949-962.

Swartz, R.C., Schults, D.W., Ozretich, J.R., Lamberson, J.O., Cole, F.A., DeWitt, T.H., Redmond, M.S., and Ferraro, S.P. (1995) ∑PAH: a model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments. *Environ. Toxicol. Chem.* **14**, 1977-1987.

(US EPA) US Environmental Protection Agency (1992) Framework for ecological risk assessment. EPA/630/R-92/001, Washington, DC.

(US EPA) US Environmental Protection Agency (1998) Guidelines for ecological risk assessment. EPA/630/R-95/002F, Washington, DC.

(US EPA and US ACE) US Environmental Protection Agency and US Army Corps of Engineers. (1991) Evaluation of dredged material proposed for ocean disposal: testing manual. EPA/503R-91/001, Washington, DC.

(US EPA and US ACE) US Environmental Protection Agency and US Army Corps of Engineers. 1992. Evaluating environmental effects of dredged material management alternatives - A technical framework. EPA/842-B-92/008, Washington, DC.

(US EPA and US ACE) US Environmental Protection Agency and US Army Corps of Engineers. (1998) Evaluation of Dredged Material Proposed for Discharge in Waters of the U.S. - Testing Manual. EPA/823-B-94/002, Washington, D.C.

New Concepts in Ecological Risk Assessment: Where Do We Go from Here?

KEITH R. SOLOMON[4] and PAUL SIBLEY
Centre for Toxicology and Department of Environmental Biology, University of Guelph, Guelph, ON, N1G 2W1, Canada.

Abstract

Through the use of safety factors, the use of single-species test data has been adequate for use in protective hazard assessments and criteria setting but, because hazard quotients do not consider the presence of multiple species each with a particular sensitivity or the interactions that can occur between these species in a functioning community, they are ill-suited to environmental risk assessment. Significant functional redundancy occurs in most ecosystems but this is poorly considered in single-species tests conducted under laboratory conditions. A significant advance in effects assessment was the use of the microcosm as a unit within which to test interacting populations of organisms. The microcosm has allowed the measurement of the environmental effect measures such as the $NOAEC_{community}$ under laboratory or field conditions and the application of this and other similarly derived measures to ecological risk assessment. More recently, distributions of single-species laboratory test data have been used for criteria setting and, combined with distributions of exposure concentrations, for risk assessment. Distributions of species sensitivity values have been used in an *a priori* way for setting environmental quality criteria such as the Final Acute Value (FAV) derived for water quality criteria. Similar distributional approaches have been combined with modeled or measured concentrations to produce estimates of the joint probability of a single species being affected or that a proportion of organisms in a community will be impacted in *a posteriori* risk assessments. These techniques has not been widely applied for risk assessment of dredged materials, however, with appropriate consideration of bioavailability and spatial and nature of the data these techniques can be applied to soils and sediments.

Keywords: ecotoxicology, environmental risk assessment, new approaches

Introduction

The application of risk assessment to protect human health has grown over the last 60 years, but it is only during the last 25 years that ecological risk assessment (ERA) has become more widely used. The aim of ERA is the estimation of risk of adverse effects to communities of species in locations that are potentially exposed to pollutants and other substances. ERA can also be used for the prioritization of pollutants or sites for regulatory purposes, as well as in the development of environmental quality guidelines (Solomon and Takacs, 2001). The most common approach in ERA has been the use of single-species test data as surrogates for other species in the environments being assessed. Through the use of safety factors, this approach was adequate for use in protective hazard assessments and criteria setting but, because single-species tests do not consider the presence of multiple species each with a particular sensitivity or the

[4]Corresponding author: Keith Solomon, T 519-837-3320, F 519-837-3861, Ksolomon@tox.uoguelph.ca

interactions that can occur between these species in a functioning community, they are ill-suited to environmental risk assessment. Significant functional redundancy occurs in most ecosystems but this is poorly considered in single-species tests conducted under laboratory conditions.

A significant advance in effects assessment was the use of the microcosm as a unit within which to test interacting populations of organisms. The microcosm has allowed the measurement of the environmental effect measures such as the $NOAEC_{community}$ under laboratory or field conditions(Campbell et al., 1999; Hill, Heimbach, Leeuwangh and Matthiessen, 1994; Van den Brink, Van Wijngaarden, Lucassen, Brock and Leeuwangh, 1996) and the application of this and other similarly derived measures to ecological risk assessment. Similar usefulness has been derived from in-situ community-level risk assessments (Hill et al., 1994).

More recently, effects data from many species have been used in new approaches to ERA (Cardwell, Parkhurst, Warren-Hicks and Volosin, 1993; Parkhurst et al., 1996; Stephan, Mount, Hansen, Gentile, Chapman and Brungs, 1985; Van Straalen, 1990; Van Straalen and Denneman, 1989). Distributions of single-species laboratory test data have been used for criteria setting and, combined with distributions of exposure concentrations, for ERA. Thus, lower centiles of distributions of species sensitivity values have been used in an *a priori* way for setting environmental quality criteria such as the Final Acute Value (FAV), FCV, FSV (Stephan et al., 1985), and HC_5 (Van Straalen and Van Rijn, 1998). Similar distributional approaches have been combined with modeled or measured concentrations to produce estimates of the joint probability of a single species being affected or that a proportion of organisms in a community will be impacted in *a posteriori* risk assessments. These approaches have recently been incorporated in new recommendations for ecological risk assessment for pesticides as suggested through the ECOFRAM process, a joint EPA, Industry, and Academic initiative to develop new methods of risk assessment for pesticides (ECOFRAM, 1999).

While some of these developments have addressed risk assessments of toxic substances in sediments, the use of the techniques has not been widely applied for risk assessment of dredged materials and their disposal on land or in aquatic environments. This paper chronicles these developments in ecotoxicology in the larger framework of the developing science of ecological risk assessment and draw attention to components of the process that could be applied to risk assessment for sediments, dredged material and other similar matrices.

Ecological Risk Assessment in the Context of Ecology

In risk assessment, assessment endpoints and measures of effect can be defined at all levels or organization in ecosystems, from the individual to the community and community (Suter, Barnthouse, Bartell, Mill, Mackay and Patterson, 1993). Most assessment measures in ecological risk assessment are defined at the population, rather than at the level of the individual organism. While risk assessment for human health protection normally focuses on individual humans, organisms in the ecosystem are generally regarded as transitory and, because they are usually part of a food web, are individually unessential for maintaining ecosystem function (Suter et al., 1993). In the case of rare or endangered species such as whales or pandas, individuals have value and are afforded similar protection to that offered to humans. Functional redundancy is essential to the continuance of ecosystems exposed to natural stressors, results from evolutionary pressures from changing and unpredictable environmental conditions and is seen when multiple species are able to perform the same critical functions (Baskin, 1994; Walker, 1992, 1995). In ERA, functional redundancy is the basis for being able to accept some effects in some organisms because these are unlikely to adversely affect the functions of the ecosystem as a whole (Stephan et al., 1985). Therefore, in ERA, some effects on organisms and populations can be accepted, provided that these effects are restricted on the spatial and temporal scale, in other words, only affect a small proportion of locations and are short term in nature.

Single Species Tests, Experimental Microcosms, and *in Situ* Community Responses

Single species tests, such as acute and chronic laboratory tests in organisms cannot take into account effects that involve interactions between populations in communities or those that affect ecosystem function. Field studies in exposed environments and experimentally perturbed systems have been used to assess higher levels of response to exposure to stressors and this has been applied to sediments by a number of workers including the group working at the Canada Centre for Inland Waters in Burlington, ON (Reynoldson, Bailey, Day and Norris, 1995; Reynoldson, Noris, Resh, Day and Rosenberg, 1997; Reynoldson and Rodriguez, 1998; Reynoldson and Zarull, 1993). These *in situ* assessments are useful for characterizing differences between locations, but potential confounders such as the physical properties of the sediments need to be considered. These community metrics also do not allow prediction of responses in advance but rather the detection of responses after contamination has occurred. From the point of view of regulations and the permitting process, *in situ* assessments are less useful.

Multispecies experimental systems (microcosms and mesocosms) where one or more variables can be controlled by the researcher offer more utility in predictive assessments. There is confusion in the use of the terms microcosms and mesocosms but the former are generally considered to be larger than a 10-L beaker but smaller than a 0.1 ha field pond. Mesocosms are generally thought of as being equal to or larger than a 0.1 ha pond and are usually outdoor systems. There are numerous examples of the use of experimental ecosystems to measure effects in communities (Hill et al., 1994). Studies in microcosms provide effect measures that are closer to the assessment measures, that incorporate the summation of responses of many species in the community, that allow for the observation of the recovery of populations and communities, and that allow the observation of responses caused by indirect effects of stressors on insensitive organisms (Solomon et al., 1996). Microcosm studies allow three types of ecologically relevant observations to be made at the population level. These include no effect, the ecosystem-level no-observed-effect-concentrations (NOECs); effects with recovery in the period of observation, no-adverse-effect-concentrations (NOAECs); (Giesy, Solomon, Coates, Dixon, Giddings and Kenaga, 1999, Solomon et al., 1996), and effects with no recovery observed in the period of observation. Depending on the responses observed and the organisms impacted, the (NOAECs) has been used to characterize the Ecologically Acceptable Concentration (EAC, Campbell et al., 1999), or $NOEC_{community}$ (Giddings, Hall and Solomon, 2000; Giddings, Solomon and Maund, 2001; Van den Brink, Hattink, Bransen, Van Donk and Brock, 2000; Van den Brink et al., 1996). These responses include a measure of the resiliency of the system and redundancy of function and are thus more useful in assessments of ecological risks. It should be noted that, although microcosms offer significant advantages over single species tests, they cannot be used to answer all questions about a situation or scenario. Properly used, they can be used to test specific hypotheses derived from laboratory or other data but, because they do not contain all possible species nor experience all possible abiotic variables, they are not surrogates for the environment in general.

The Ecological Risk Assessment Process

ERAs are usually conducted in series of steps or tiers (ECOFRAM, 1999; SETAC, 1994; Suter et al., 1993; USEPA, 1992, 1998). It is normal to divide complex tasks into smaller components that can be more easily managed and, in ERAs this can reduce complexity and narrow the focus to the key issues. In the tiered approach, the initial use of conservative criteria allows substances or situations that truly do not present a risk to be eliminated from the ERA process, thus allowing a shift of resources to situations with potentially greater risk. As one progresses through the tiers, the estimates of exposure and effects become more realistic as uncertainty is reduced through the acquisition of more or better quality data. Thus tiers are normally designed such that the earlier tiers are more conservative, while the later tiers are more realistic. Because the earlier tiers are designed to be protective, failing to meet the criteria for these tiers is merely an indication that an assessment based on more realistic data is needed before a regulatory or risk management decision can be reached.

The simplest forms of ERA processes are classification systems, such as US Environmental Protection Agency's (USEPA) Office of Toxic Substances (OTS) chemical scoring system and the USEPA's hazard ranking system (USEPA-HRS) (Suter et al., 1993). The basic use of scoring systems is to assign a rank or priority to a substance, either from a quantal criterion for a property (above or below a threshold) or the use of multiple criteria which are assigned numerical scores. Correctly used, scoring systems have been employed to rank substances in order of priority for further assessment and this is usually carried out in the initial stages of risk assessment. Further assessment is normally required because the scoring systems commonly make use of worst-case data (the most extreme value); they cannot always handle missing values, weighting, or scaling in clear or appropriate ways; and they take no consideration of exposures other than through estimates of total production, use, or release of the substance into the environment (Suter et al., 1993).

Currently, the most widely used method in ERA is the hazard quotient (HQ) method. In this system, the environmental concentration of a stressor is compared to an effect concentration (Calabrese and Baldwin, 1993; Urban and Cook, 1986). These are simple ratios of single exposure and effects values and may be used to express hazard or relative safety. HQs are normally calculated from the effect concentration of the most sensitive organism or group of organisms and comparing this to the greatest exposure concentration. To allow for unquantified uncertainty in the effect and exposure estimates, the HQ may be made more conservative by the use of an uncertainty factor (CWQG, 1999) or by comparison to predefined criteria (Levels of Concern, LOCs) Urban and Cook, 1986) which may vary, depending on the effect or whether endangered species are likely to be exposed (Urban and Cook, 1986). Because they frequently make use of worst-case or extreme data, HQs are designed to be protective of almost all possible situations that may occur and this may lead to the implementation of expensive management measures for stressors that pose little or no actual threat to humans or the environment (Lee and Jones-Lee, 1995; Moore and Elliott, 1996). As the HQ is based on a ratio of point estimates, it is not proportional to the risk, its use assumes that the conditions of the HQ exist on every occasion and in every location.

Probabilistic approaches to ecological risk assessment (PERA) have been recommended for later tiers in the ERA process (ECOFRAM, 1999; SETAC, 1994). These methods use distributions of species sensitivity combined with distributions of exposure concentrations to better describe the likelihood of exceedences of effect thresholds and thus the risk of adverse effects. The major advantage of PERAs is that they uses all relevant single species toxicity data and, when combined with exposure distributions, allows quantitative estimations of risks. However, the method does have some disadvantages; more effects and exposure data are usually needed, it does not address all sources of uncertainty and has not been widely calibrated against field observations. Although relatively new, the methods of PERA have been described (Campbell, Bartell and Shaw, 2000; Cardwell et al., 1993; Giesy et al., 1999; Jongbloed, Traas and Luttik, 1996; Klaine et al., 1996; Parkhurst et al., 1996; SETAC, 1994; Solomon, 1996; Solomon, Giddings and Maund, 2001; Traas, Luttik and Jongbloed, 1996).

Where PERAs have included studies in the field or in microcosms, it has been consistently noted that effects in the field are rarely observed at concentrations equivalent to lower centiles of toxicity distributions (Giddings et al., 2000; Giddings et al., 2001; Giesy et al., 1999; Hall, Giddings, Solomon and Balcomb, 1999; Solomon et al., 1996; Versteeg, Belanger and Carr, 1999). In fact, ecologically significant effects are sometimes only observed at concentrations exceeding 25th centiles of laboratory-based acute toxicity values (Giddings et al., 2001; Hall and Giddings, 2000).

New Directions for Risk Assessment Related to Dredged Materials

Several important differences between risk assessments in the water column and in sediments will affect the degree to which new approaches to risk assessment can be applied. While these differences may be viewed as barriers at this time, they also present interesting challenges to the scientific community.

Risk assessments that rely on measured concentrations can be significantly affected by interactions between the matrix and the substances of concern. Bioavailability of metals and organic substances is affected by other components in the sediments (Di Toro, Kavvadas, Mathew, Paquin and Winfiled, 2001). These factors usually result in a lowering of effective concentration of the substances of concern, thus reducing risk. Bioavailability can be estimated from models and exposure concentrations can be adjusted to more closely reflect the actual value. These new values can then be used to assess hazard or risk, depending on the amount of data available. One uncertainty in using this approach is the stability of the bioavailability fraction over time. Natural and anthropogenic changes in the sediment environment can change bioavailability and this must be considered in the assessment process.

Bioaccumulative substances present another challenge to risk assessment, particularly where there is the potential for exposure directly as well as through the food chain. The problem with many chemical stressors is that the exposures used in laboratory testing situations are different from those occurring in the ecosystem. Because of the kinetics of uptake, laboratory tests may be too short-term to allow a strongly bioaccumulative substance to attain maximum concentration in the receptor organism. Bioassay results could underestimate the toxicity of the substance in relation to its ultimate potential effect on organisms in the environment. If a certain route of exposure is identified as important, specific tests utilizing this route can be conducted, such as has been recommended for sediment-mediated toxicity of pesticides (ECOFRAM, 1999). PERAs can be conducted for pollutants that bioaccumulate by accounting for secondary exposure through food webs. A methodology for conducting this has been proposed by the Organization for Economic Cooperation and Development (OECD) (Balk, Okkerman and

Dogger, 1995). In principle, effect concentrations in the organisms are compared to exposure concentrations in organisms from the exposed environment. The units are the same and an HQ or a PERA approach can be applied. However good data on the relationship between body burden and effects and a good database of concentrations measured in organisms collected from the environment are required. Use of bioaccumulation and food-web models introduces additional complexity, thus making the risk assessment and, more important, its communication to non-technical decision makers more difficult (Solomon and Takacs, 2001).

As has been pointed out in a number of companion papers, potentially toxic substances seldom occur in isolation, particularly in sediments. Thus organisms are frequently exposed to mixtures of substances, often with different mechanisms of action (Giesy et al., 1999; Lee and Jones-Lee, 1999; Solomon et al., 1996). Where sediments have a constant composition, whole matrix testing can be used as a physical model for assessment of a complex mixture, however, because of temporal and environmental variables deposition of sediments in the environment may be essentially unpredictable, resulting in combinations of concentrations in both the temporal and spatial dimension. Where substances are known to act additively, it is possible to use the toxic equivalents (TEs) or toxic units (TUs) to sum concentrations and assess risks from the mixtures. This approach has been applied to several classes of compounds including dioxins (Ahlborg et al., 1994; Parrot, Hodson, Servos, Huestis and Dixon, 1995), chlorinated phenols (Kovacs, Martel, Voss, Wrist and Willes, 1993), and polyaromatic hydrocarbons (Schwarz et al., 1995).

The TE approach was used to assess the combined risk from atrazine and its metabolites (Solomon, 1999) using probabilistic approaches and similar techniques could be used to assess risks from substances with a common mode of action. This may have utility in assessing risks from complex mixtures of hydrocarbons and similar substances that act through generalized narcotic mechanisms (Lipnick, 1993). Traditionally, these equivalents have been based on responses measured in the same organism, for example, the laboratory rat. This is appropriate if the risks are to be assessed in the same organism or extrapolated to another (humans) with appropriate uncertainty factors. However, potencies measured in one animal may not be the same in another, and wide inter-specific extrapolations, such as from rats to fish, may not be possible (Parrott et al., 1995). This situation becomes more complex when dealing with ecological risk assessments depending on whether TEs are based on responses measured in a single species or on point estimates of potency derived from species sensitivity distributions (Solomon and Takacs, 2001). If the substances interact through response addition, similar approaches would be possible; although, differences in exposure times may introduce additional complexity to the assessment.

The great complexity of mixtures in sediments and soils makes analysis and identification of individual substances difficult and the associated toxicity data required to assess responses in several groups of organisms may be impractically expensive to collect. If common mechanisms of action such as narcosis dominate, it may be possible to use bulk parameters as surrogates for exposure such as the amount of organic matter extractable in non-polar solvents. Coupled with concepts of critical body residue (McCarty, Mackay, Smith, Ozburn and Dixon, 1992), it may be possible to estimate the summed potency of mixtures. Again, this introduces significant complexity into the risk assessment and it may be more efficient to rely on simple physical models such as bioassays.

One of the advantages of the concept of PERA is that it allows variability in exposures and sensitivity of organisms to be addressed in the risk assessment process. One method for displaying this is the exceedence profile (EP) (Solomon and Takacs, 2001) or joint probability curve (JPC) as proposed by ECOFRAM (1999). The EP (Figure 1) illustrates the relationship between the proportion of species affected and the likelihood that their response concentration will be exceeded as a simple vector on a diagram. While this obscures the identity of the species being affected, it does allow for a visual presentation of a complex relationship. The area under the curve can be calculated and provides a risk indicator (the mean risk or the estimated total risk) for purposes of ranking where data from multiple sites or temporal sampling sets are being assessed.

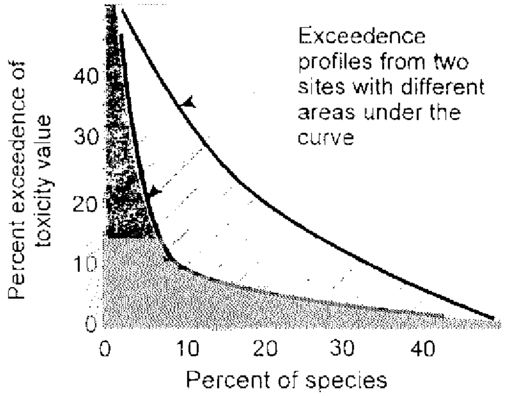


Fig. 1. Illustration of the exceedence profile and the area under the curve as risk indicator.

In assessing risks in sediments and in dredged materials, spatial and temporal variability become much more important than when conducting ERAs in water. While water is a relatively well mixed matrix, sediments are less mobile and can gave relatively large gradients of concentrations in the horizontal and the vertical plane (Figure 2).

Whereas concentrations of many substances in water tend to change relatively rapidly,

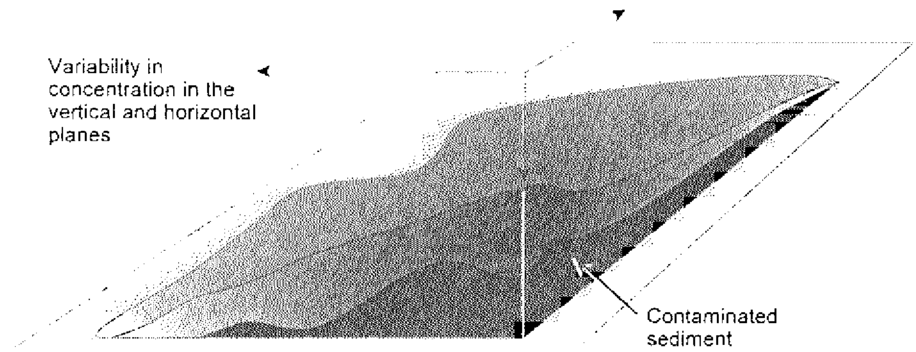


Fig. 2. Illustration of variation in concentration in three dimensions.

concentrations of most substances in sediments change more slowly. This is due to high binding and adsorption to sediment particles and lack of exposure of the substance to degradation either by chemical or biological pathways. Thus, in characterizing variability of exposure concentrations in sediments, the temporal scale is probably less important than the spatial scale. As a consequence, the distributions of exposure concentrations may be best expressed in terms of horizontal or areal dimensions. When comparing these exposures to species sensitivity distributions, it may be more appropriate to consider organisms in classes related to their mobility. Thus, if the area of contamination is small in comparison to the range of the organism, probability of exposure is reduced. If the organism is sessile, the probability of exposure in the contaminated area is great but essentially zero outside this area. In considering habitat and mobility ranges in conducting PERAs for sediments, more use could be made of landscape-level approaches such as have been described in the ECOFRAM process (ECOFRAM, 1999) and used in the risk assessment of pesticides (Travis and Hendley, 2001).

Although PERA approaches are useful for characterizing variability, they are less useful for assessing uncertainty. Uncertainty may come from lack of knowledge, random, or systematic noise and errors in measurements. For complex mixtures in sediments, lack of knowledge of all the potentially toxic constituents is the greatest source of uncertainty. Random or systematic errors in the quantification of toxic compounds known to be present are less likely to be important, however, errors in sampling may be a major source of uncertainty, especially when the site is at depth and not readily accessible. In characterizing toxicity, similar uncertainties may exist, especially as related to lack of knowledge of the sensitivity of organisms not tested. A potentially important uncertainty in the risk assessment process is that of exposure and bioavailability. Substances such as metals may be less bioavailable than indicated from analytical methods that do not address speciation (Di Toro et al., 2001).

Although PERA provides tools to more thoroughly conduct risk assessments and to handle large data sets, other lines of evidence are also important in reaching the final conclusions (Hall and Giddings, 2000). PERA is an improvement over the HQ approach, but it will likely continue to develop as the entire science of risk assessment advances. One of the major hurdles that PERA faces is its acceptance by the public and regulators (Roberts, 1999; Solomon, 1996; Solomon and Takacs, 2001). For the most part, the public and regulators want to know whether an activity is safe or not and prefer being told what will happen, not what might happen (Morgan, 1998). The public demands absolute safety but has less understanding of science, the scientific method, and the fact that science can never give an absolute answer.

This paper has provided several examples of new approaches to risk assessment that have yet to be tested and calibrated for sediments and sediment-bound substances. Sediments present several obstacles to the implementation of new risk assessment methods. Rather than being viewed as a hindrance, these should be viewed as challenges and a stimulation to develop new and better methods that can be easily and broadly applied.

References

Ahlborg, U.G., Becking, G.C., Birnbaum, L.S., Brouwer, A., Derks, H.J.G., Feeley, M., Golor, G., Hanberg, A., Larsen, J.C., Liem, A.K.D., Safe, S.H., Schlatter, C., Waern, F., Younes, M. and Yrjänheikki, E. (1994) Toxic equivalence factors for dioxin-like PCBs. Report on a WHO-ECEH and IPCS consultation, December 1993. *Chemosphere* **28**, 1049-1067.

Balk, F., Okkerman, P.C. and Dogger, J.W. (1995) *Guidance Document for Aquatic Effects Assessment*. Organization for Economic Co-operation and Development (OECD), Paris.
Baskin, Y. (1994) Ecosystem function of biodiversity. *Bioscience* **44**, 657-660.

Calabrese, E.J. and Baldwin, L.A. (1993) *Performing Ecological Risk Assessments*. Lewis Publishers, Boca Raton, FL.

Campbell, K.R., Bartell, S.M. and Shaw, J.L. (2000) Characterizing aquatic ecotoxicological risks from pesticides using a diquat dibromide case study. II. Approaches using quotients and distributions. *Environ. Toxicol. Chem.* **19**, (760-774).

Campbell, P.J., Arnold, D., Brock, T., Grandy, N., Heger, W., Heimbach, F., Maund, S.J. and Streloke, M. (1999) Guidance Document: Higher-tier Aquatic Risk Assessment for Pesticides, p. 179. SETAC-Europe, Brussels, Belgium.

Cardwell, R.D., Parkhurst, B.R., Warren-Hicks, W. and Volosin, J.S. (1993) Aquatic ecological risk. *Water Environ. Tech.* **5**, 47-51.

CWQG. (1999) *Canadian Water Quality Guidelines (and updates)*. Task Force on Water Quality Guidelines of the Canadian Council of Resource and Environment Ministers, Ottawa, ON.

Di Toro, D.M., Kavvadas, C.D., Mathew, R., Paquin, P.R. and Winfiled, R.P. (2001) *The Persistence and availability of metals in aquatic environments*, p. 67. International Council on Metals and the Environment, Ottawa, ON, Canada.

ECOFRAM. (1999) ECOFRAM Aquatic and Terrestrial Final Draft Reports, 1999. USEPA.
Giddings, J.M., Hall, L.W.J. and Solomon, K.R. (2000) An ecological risk assessment of diazinon from agricultural use in the Sacramento-San Joaquin River basins, California. *Risk Anal.* **20**, 545-572.

Giddings, J.M., Solomon, K.R. and Maund, S.J. (2001) Probabilistic risk assessment of cotton pyrethroids: II. Aquatic mesocosm and field studies. *Environ. Toxicol. Chem.* **20**, 660-668.

Giesy, J.P., Solomon, K.R., Coates, J.R., Dixon, K.R., Giddings, J.M. and Kenaga, E.E. (1999) Chlorpyrifos: Ecological risk assessment in North American aquatic environments. *Environ. Contam. Toxicol.* **160**, 1-129.

Hall, L.W.J. and Giddings, J.M. (2000) The need for multiple lines of evidence for predicting site-specific ecological effects. *Human Ecol. Risk Assess.* **6**, 679-710.

Hall, L.W.J., Giddings, J.M., Solomon, K.R. and Balcomb, R. (1999) An ecological risk assessment for the use of Irgarol 1051 as an algaecide for antifoulant paints. *Crit. Rev. Toxicol.* **29**, 367-437.

Hill, I.R., Heimbach, F., Leeuwangh, P. and Matthiessen, P. (1994) *Freshwater Field Tests for Hazard Assessment of Chemicals*. CRC Press, Boca Raton, FL, USA.

Jongbloed, R.H., Traas, T.P. and Luttik, R. (1996) A probabilistic model for deriving soil quality criteria based on secondary poisoning of top predators. II. Calculations for dichlorodiphenyltrichloroethane (DDT) and cadmium. *Ecotoxicol. Environ. Safety* **34**, 279-306.

Klaine, S.J., Cobb, G.P., Dickerson, R.L., Dixon, K.R., Kendall, R.J., Smith, E.E. and Solomon, K.R. (1996) An ecological risk assessment for the use of the biocide, dibromonitrilopropionamide (DBNPA) in industrial cooling systems. *Environ. Toxicol. Chem.* **15**, 21-30.

Kovacs, T.G., Martel, P.H., Voss, R.H., Wrist, P.E. and Willes, R.F. (1993) Aquatic toxicity equivalency factors for chlorinated phenolic compounds present in pulp mill effluent. *Environ. Toxicol. Chem.* **12**, 684-691.

Lee, G.F. and Jones-Lee, A. (1995) Appropriate use of numeric chemical concentration-based water quality criteria. *Human Ecol. Risk Assess.* **1**, 5-11.

Lee, G.F. and Jones-Lee, A. (1999) The single chemical probabilistic risk assessment approach is inadequate—for OP pesticide aquatic life toxicity. *SETAC News* ***19*(6)**, 20-21.

Lipnick, R.L. (1993) Baseline toxicity QSAR models: a means to assess mechanism of toxicity for aquatic organisms and mammals. In *Environmental Toxicology and Risk Assessment STP 1216, 2* (J. Gorsuch, F. Dwyer, C. Ingersoll and L. TW, eds), pp. 610-619. American Society for Testing and Materials, Philadelphia, PA.

McCarty, L.S., Mackay, D., Smith, A.D., Ozburn, G.W. and Dixon, G.D. (1992) Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: neutral narcotic organic. *Environ. Toxicol. Chem.* **11**, 917-930.

Moore, D.R.J. and Elliott, B.J. (1996) Should uncertainty be quantified in human and ecological risk assessments used for decision-making? *Human Ecol. Risk Assess.* **2**, 11-24.

Morgan, M.G. (1998) Uncertainty analysis in risk assessment. *Human Ecol. Risk Assess.* **4**, 25-39.

Parkhurst, B.R., Warren-Hicks, W., Cardwell, R.D., Volison, J., Etchison, T., Butcher, J.B. and Covington, S.M. (1996) *Aquatic ecological risk assessment: A multi-tiered approach to risk assessment.* Water Environment Research Foundation, Alexandria, VA, USA.

Parrott, J.L., Hodson, P.V., Servos, M.R., Huestis, S.L. and Dixon, D.G. (1995) Relative potencies of polychlorinated dibenzo-p-dioxins and dibenzofurans for inducing mixed function oxygenase activity in rainbow trout. *Environ. Toxicol. Chem.* **14**, 1041-1050.

Reynoldson, T.B., Bailey, R.C., Day, K.E. and Norris, R.H. (1995) Biological guidelines for freshwater sediment based on **BE**nthic Assessment of Sedimen**T** (the **BEAST**) using a multivariate approach for predicting biological state. *Austr. Ecol.* **20**, 198-219.

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E. and Rosenberg, D.M. (1997) The reference condition: a comparison of multiple and multivariate approaches to assess water-quality impairment using benthic macro invertebrates. *North Amer. Benth. Soc.* **16**, 833-852.

Reynoldson, T.B. and Rodriguez, P. (1998) Manual of bioassessment of aquatic sediment quality. In *Field methods and interpretation for sediment bioassessment* (A. Murdoch, J. Azcue and P. Murdoch, eds), CRC Press, Boca Raton, FL, USA.

Reynoldson, T.B. and Zarull, M.A. (1993) Ecological Integrity and the Management of Ecosystems. In *An approach to the development of biological sediment guidelines* (G. Francis, J. Kay and S. Woodley, eds). St. Lucie Press, Boca Raton, FL, USA.

Roberts, S.M. (1999) Practical issues in the use of probabilistic risk assessment. *Human Ecol. Risk Assess.* **5**, 729-736.

Schwarz, R.C., Schults, D.W., Ozretich, R.W., Lamberson, J.O., Cole, F.A., DeWitt, T.H., Redmond, M.S. and Ferraro, S.P. (1995) Sigma PAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments. *Environ. Toxicol. Chem.* **14**, 1977-1978.

SETAC. (1994) *Society for Environmental Toxicology and Chemistry. Pesticide Risk and Mitigation. Final Report of the Aquatic Risk Assessment and Mitigation Dialog Group*. 220 p. SETAC Foundation for Environmental Education,, Pensacola, FL, USA.

Solomon, K.R. (1996) Overview of recent developments in ecotoxicological risk assessment. *Risk Anal.* **16**, 627-633.

Solomon, K.R. (1999) Pesticide Chemistry and Bioscience: The Food-Environment Challenge. In *Integrating environmental fate and effects information: the keys to ecotoxicological risk assessment for pesticides* (G.T. Brooks and T.R. Roberts, eds), p. 313-326. Royal Society of Chemistry, London, UK.

Solomon, K.R., Baker, D.B., Richards, P., Dixon, K.R., Klaine, S.J., La Point, T.W., Kendall, R.J., Giddings, J.M., Giesy, J.P., Hall, L.W.J., Weisskopf, C. and Williams, M. (1996) Ecological risk assessment of atrazine in North American surface waters. *Environ. Toxicol. Chem.* **15**, 31-76.

Solomon, K.R., Giddings, J.M. and Maund, S.J. (2001) Probabilistic risk assessment of cotton pyrethroids: I. Distributional analyses of laboratory aquatic toxicity. *Environ. Toxicol. Chem.* ***20***, 652-659.

Solomon, K.R. and Takacs, P. (2001) Species Sensitivity Distributions in Risk Assessment. In *Probabilistic risk assessment using species sensitivity distributions* (L. Postuma, T. Traas and G.W. Suter, eds), Chapter 15, In Press. CRC press, Boca Raton, FL, USA.

Stephan, C.E., Mount, D.I., Hansen, D.J., Gentile, J.H., Chapman, G.A. and Brungs, W.A. (1985) *Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses.*, p. 1-97. US EPA ORD ERL, Duluth MN.

Suter, G.W., Barnthouse, L.W., Bartell, S.M., Mill, T., Mackay, D. and Patterson, S. (1993) *Ecol. Risk Assess.* 538 p. Lewis Publishers,, Boca Raton, FL.

Traas, T.P., Luttik, R. and Jongbloed, R.H. (1996) A probabilistic model for deriving soil quality criteria based on secondary poisoning of top predators. I. Model description and uncertainty analysis. *Ecotoxicol. Environ. Safety* **34**, 264-278.

Travis, K.Z. and Hendley, P. (2001) Probabilistic aquatic risk assessment of pyrethroids: IV. Landscape-level exposure characterization. *Environ. Toxicol. Chem.* **20**, 679-686.

Urban, D.J. and Cook, N.J. (1986) *Standard Evaluation Procedure for Ecological Risk Assessment.* Hazard Evaluation Division, Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, DC.

USEPA. (1992) *United States Environmental Protection Agency. Framework for ecological risk assessment.* USEPA, Washington, DC, USA.

USEPA. (1998) *United States Environmental Protection Agency. Guidelines for Ecological Risk Assessment.* Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC, USA.

Van den Brink, P.J., Hattink, J., Bransen, F., Van Donk, E. and Brock, T.C.M. (2000) Impact of the fungicide carbendazim in freshwater microcosms. II. zooplankton, primary producers and final conclusions. *Aquatic Toxicol.* **48**, 251-264.

Van den Brink, P.J., Van Wijngaarden, R.P.A., Lucassen, W.G.H., Brock, T.C.M. and Leeuwangh, P. (1996) Effects of the insecticide Dursban® 4E (active ingredient chlorpyrifos) in outdoor experimental ditches: II. Invertebrate community responses and recovery. *Environ. Toxicol. and Chem.* ***15(7)***, 1143-1153.

Van Straalen, N.M. (1990) New methodologies for estimating the ecological risk of chemicals. In *Sixth International Congress of the International Association of Engineering Geology* (D. Price, ed), pp.165-173. A.A. Balkema, Rotterdam, Amsterdam, Netherlands.

Van Straalen, N.M. and Denneman, C.A.J. (1989) Ecotoxicological evaluation of soil quality criteria. *Ecotoxicol. Environ. Safety* **18**, 241-251.

Van Straalen, N.M. and Van Rijn, J.P. (1998) Ecotoxicological risk assessment of soil fauna recovery from pesticide application. *Environ. Contam. Toxicol.* **154**, 85-141.

Versteeg, D.J., Belanger, S.E. and Carr, G.J. (1999) Understanding single-species and model ecosystem sensitivity: Data-based comparison. *Environ. Toxicol. Chem.* **18**, 1329-1346.

Walker, B. (1992) Biodiversity and ecological redundancy. *Conser. Biol.* **6**, 18-23.

Walker B. (1995) Conserving biological diversity through ecosystem resilience. *Conser. Biol.* **9**, 747-752.