An Evaluation of the Approach for Assessing Risks to the Benthic Invertebrate Community at the Portland Harbor Superfund Site

Preliminary Draft

Prepared for:

U.S. Environmental Protection Agency
Oregon Operations Office
805 SW Broadway, Suite 500
Portland, Oregon 97205

and

Parametrix, Inc.
33972 Texas Street SW
Albany, Oregon 97321

Prepared - *September, 2008* - by:

D.D. MacDonald
MacDonald Environmental Sciences Ltd.
#24 - 4800 Island Highway North
Nanaimo, British Columbia V9T 1W6

P.F. Landrum
Landrum and Associates
6829 Earhart Road
Ann Arbor, Michigan 48105

MACDONALD
ENVIRONMENTAL SCIENCES LTD.

An Evaluation of the Approach for Assessing Risks to the Benthic Invertebrate Community at the Portland Harbor Superfund Site

Preliminary Draft

U.S. Environmental Protection Agency
Oregon Operations Office
805 SW Broadway, Suite 500
Portland, Oregon 97205

and

Parametrix, Inc.
33972 Texas Street SW
Albany, Oregon 97321

Prepared – *September, 2008* – by:

D.D. MacDonald
MacDonald Environmental Sciences Ltd.
#24 - 4800 Island Highway North
Nanaimo, British Columbia V9T 1W6

P.F. Landrum
Landrum and Associates
6829 Earhart Road
Ann Arbor, Michigan 48105

MESL DOCUMENT NO. MESL-PHR-BICRA-0908-V2

Table of Contents

Table of Contents. I

List of Figures. III

List of Acronyms. IV

1.0 Introduction. 1

2.0 Background 3

3.0 Terms of Reference for this Evaluation. 7

4.0 Recommendations and Associated Rationale. 9

4.1 Scope of this Evaluation. 10
4.2 Recommended Framework for Assessing Risks to the Benthic Invertebrate Community. 11
4.3 Recommended Procedures for Designating Sediment Samples as Toxic or Not Toxic. 15
4.4 Recommended Procedures for Developing a Reference Envelope for Interpreting Data from Whole-Sediment Toxicity Tests. 20
4.5 Recommended Procedures for Integrating Data on Multiple Toxicity Test Endpoints. 22
4.6 Recommended Procedures for Evaluating Relationships Between Sediment Chemistry and Sediment Toxicity. 25
4.7 Recommended Procedures for Developing Toxicity Thresholds. 27
4.8 Procedures for Evaluating Concentration-Response Models. 28
4.9 Recommended Procedures for Assessing Risks to Benthic Invertebrates. 31

5.0 Summary and Conclusions. 31

6.0 References. 38

Addendum 1 Further Evaluation of the Approach for Assessing Risks to the Benthic Invertebrate Community at the Portland Harbor Superfund Site. A-1

A1.0 Introduction. A-1

A2.0 Responses to Additional Questions. A-1

A3.0 Application of Regional Sediment Evaluation Team (RSET) Process to the Portland Harbor Site. A-14

A4.0 Development of a Reference Envelope for Portland Harbor A-15

A4.1 Approaches to Selecting Reference Locations. A-16

A4.2 Criteria for Identifying Reference Sediment Samples. . . A-17

A5.0 Development of Clean-up Goals for Portland Harbor. . . A-18

A6.0 References. A-19

Table A1 Reliability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca*, and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass). A-21

Table A2 Predictive ability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca*, and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass). A-22

Table A3 Incidence of toxicity to *Ampelisca abdita* and *Hyalella azteca* exposed to whole-sediment samples with various mean probable effect concentration-quotient (PEC-Q) distributions. A-27

Table A4 Biological conditions that occur within the three categories of risk to the benthic invertebrate community in the Calcasieu Estuary, identified using the risk designations assigned to each sample. A-28

Figure A1 Relationship between the geometric mean of the mean PEC-Q and the average survival of the freshwater amphipod, *Hyalella azteca*, in 28-d toxicity tests (data source: MacDonald *et al.* 2002; dashed lines represent 95% prediction limits).. . . . . A-29

List of Figures

Figure 1 Scatter plot showing the relationship between amphipod (*Hyalella azteca*) survival and biomass (n = 76). F-1

Figure 2 Scatter plot showing the relationship between amphipod (*Hyalella azteca*) survival and midge (*Chironomus dilutus*) survival (n = 76). F-2

Figure 3 Scatter plot showing the relationship between amphipod (*Hyalella azteca*) survival and midge (*Chironomus dilutus*) biomass (n = 76). F-3

List of Acronyms

BERA	-	baseline ecological risk assessment
CERCLA	-	Comprehensive Environmental Response, Compensation, and Liability Act
COPC	-	chemical of potential concern
DW	-	dry weight
ERA	-	ecological risk assessment
ESB-TU	-	equilibrium partitioning-based sediment benchmark-toxic unit
f_{oc}	-	fraction organic carbon
FPM	-	floating percentile model
iAOPC	-	initial area of potential concern
LOE	-	line-of-evidence
LRM	-	logistic regression model
LWG	-	Lower Willamette Group
MSD	-	minimum significant difference
NOAA	-	National Oceanic and Atmospheric Administration
PAH	-	polycyclic aromatic hydrocarbon
PCB	-	polychlorinated biphenyl
PEC-Q	-	probable effect concentration-quotient
PEL	-	probable effect level
PRG	-	preliminary remediation goal
PRP	-	potentially responsible party
QAPP	-	Quality Assurance Project Plan
RSET	-	Regional Sediment Evaluation Team
RI/FS	-	remedial investigation/feasability study
SEM-AVS	-	simultaneously extracted metals minus acid volatile sulfide
SFF	-	Sustainable Fisheries Foundation
SQG	-	sediment quality guideline
SQV	-	sediment quality value
TEL	-	threshold effect level
TIE	-	toxicity identification evaluation
TMDL	-	total maximum daily load
USEPA	-	United States Environmental Protection Agency
WOE	-	weight-of-evidence

1.0 Introduction

The Portland Harbor Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) site is located in Portland, Oregon and includes about 11 miles of the lower Willamette River and surrounding upland areas that discharge to the river. The Willamette River is a major tributary to the Columbia River. As part of the overall remedial investigation/feasability study (RI/FS) that is being conducted at the site, assessments of the nature and extent of contamination, of risks to ecological receptors, and of risks to human health have been ongoing for some time. These assessment activities are being led by the potentially responsible parties (PRPs) through work conducted by the Lower Willamette Group (LWG).

As part of the RI/FS process, the LWG is conducting a baseline ecological risk assessment (BERA) of the Portland Harbor site. According to the baseline problem formulation that has been developed for the site, the BERA is intended to assess risks to aquatic plants, benthic macroinvertebrates, bivalves, decapods, fish, amphibians, aquatic-dependent birds, and aquatic-dependent mammals (USEPA 2008). Importantly, the problem formulation document identifies the assessment endpoints and the measurement endpoints that will be evaluated in the BERA. For benthic macroinvertebrates, the BERA is intended to provide a basis for assessing effects on the survival, growth, and reproduction of benthic invertebrates associated with exposure to contaminated sediments and transition zone water (i.e., pore water) in Portland Harbor. The measurement endpoints that were identified to support evaluation of the status of the assessment endpoint include (USEPA 2008):

- Whole-sediment toxicity;
- Whole-sediment chemistry;
- Surface-water chemistry;
- Pore-water chemistry; and,
- Invertebrate-tissue chemistry.

A number of procedures have been identified for interpreting the data collected in the study area relative to evaluation of this assessment endpoint. For example, the LWG (2004) identified provisional toxicity reference values for use in the ecological risk assessment process. In addition, LWG described procedures for estimating risks to benthic invertebrates using sediment toxicity tests (LWG 2005a) and using predictive models based on sediment toxicity tests (LWG 2006). More recently, United States Environmental Protection Agency (USEPA) identified specific analytical procedures for interpreting these data in the problem formulation document and supporting documentation (USEPA 2008). While there are many similarities among the various data interpretation procedures that have been identified to date, LWG and USEPA have had some difficulty in coming to agreement on the details of these approaches to data analysis.

Both LWG and USEPA recognize that resolving differences regarding the data analysis process for assessing risks to benthic invertebrates could be challenging. For this reason, LWG and USEPA have agreed to solicit an independent evaluation of the various approaches that have been proposed to date to provide a perspective that could help to identify a mutually-acceptable path forward. More specifically, Don MacDonald and Peter Landrum were retained by Parametrix, Inc., on behalf of the LWG and USEPA, to conduct such an evaluation of approaches for assessing risks to the benthic community at the Portland Harbor site. This document presents the background information (Section 2.0) and terms of reference (Section 3.0) that were provided by USEPA. In addition, this document summarizes the recommendations that are offered to LWG and USEPA for assessing risks to benthic invertebrates using the data and information that have been collected at the site (Section 4.0). Responses to each of the seven questions posed by USEPA in the terms of reference are provided in the Summary and Conclusions (Section 5.0) of this document.

2.0 Background

As indicated above, LWG and USEPA agreed to have Don MacDonald and Peter Landrum conduct an independent evaluation of the various approaches for assessing risks to benthic invertebrates at the Portland Harbor site. To facilitate this evaluation, the various documents pertaining to the benthic invertebrate portion of the BERA, prepared by LWG or USEPA, were provided to these reviewers. In addition, the reviewers were provided with access to the data and information that have been collected to date at the site. Furthermore, additional background information was provided by USEPA, as follows:

Portland Harbor Work Plan: Due to the large size of the Portland Harbor site (approximately 11 river miles), USEPA and the Lower Willamette agreed to use sediment and bioassay results to "develop a predictive model of chemical-to-effects to assess risk from bulk sediment." This approach was not described in the programmatic work plan (April 2004) but rather in the technical memorandum - Estimating Risks to Benthic Organisms using Sediment Bioassays (March 18, 2005). This technical memorandum specified the sediment bioassay tests that would be used at the site (10-day *Chironomus* and 28-day *Hyalella*), the endpoints (growth and mortality) the hit/no-hit designation (10% and 25% difference from control for the two mortality endpoints, 25 and 40% difference from control for the *Hyalella* growth endpoint, and 20% and 30% difference from control for the *Chironomus* growth endpoint), and the approaches that would be considered to develop predictive relationships [1) sediment quality values (SQVs) derived using database percentiles, 2) SQVs derived using consensus-based values, 3) a quotient method, 4) the floating percentile method, and 5) logistic regression analysis]. It was agreed that each predictive relationship would be evaluated using measures such as false positive and false negative reliability rates.

Round 2 Data Collection: In 2004, 233 sediment bioassay tests were performed on sediment samples collected from the Portland Harbor site. Sample locations were selected to ensure that bioassay tests were performed across a range of contaminant concentrations and sources. Results were presented in the Round 2A Data Report - Sediment Toxicity Testing (April 8, 2005). Results are presented in this report and are also available in Query Manager, a database developed and maintained by National Oceanic and Atmospheric Administration (NOAA).

Preliminary Evaluation of Benthic Toxicity Results: Once the Round 2 Bioassay results were received, USEPA and the LWG embarked on a series of discussions to determine which predictive model(s) to apply at the site. The LWG presented an analysis that suggested that the Probable Effect Concentration-Quotient (PEC-Q) approach was not a reliable predictor of sediment toxicity at the site and that the predictive models should focus in on the floating percentile and logistic regression models. It was agreed that the models would consider three different hit/no-hit thresholds - 10%, 20% and 30% difference from control. The LWG also raised concerns about the reliability of the *Hyalella* growth endpoint in the floating percentile model.

Benthic Interpretive Report: On March 17, 2006, the LWG submitted the Interpretive Report: Estimating Risks to Benthic Organisms using Predictive Models Based on Sediment Toxicity Tests. This report presented an evaluation of the floating percentile and logistic regression models as well as a comparison to existing SQVs. The stated goal of the predictive model is "to derive SQVs that are sufficiently reliable for predicting benthic toxicity within the study area" and to develop a line-of-evidence "for identifying areas where chemical concentrations in sediment may pose a risk to benthic invertebrates."

On July 6, 2006, USEPA commented on the Benthic Interpretive Approach. The LWG responded to these comments on September 1, 2006. In the LWG response to comments, there were a number of comments that the LWG identified as category 1 - strongly disagree; cannot accept. In particular, the LWG disagreed with USEPA's comment to include the *Hyalella* growth endpoint in the floating percentile model and to consider effects level 1 (10% difference from control) in the development of the predictive models. In addition, the LWG agreed to the use of the alternative logistic regression model using a larger, non-site specific, freshwater database for the *Hyalella* 28-day growth and survival test as a complimentary line-of-evidence (LOE) to the floating percentile model. The LWG also agreed to use the revised logistic regression model based on the *Hyalella* pooled endpoint and the floating percentile model based on *Chironomus* growth, *Chironomus* mortality and *Hyalella* morality endpoints as separate LOEs in assessing risks to the benthic community.

Round 2 Report: On February 21, 2007, the LWG submitted the Comprehensive Round 2 Site Characterization Summary and Data Gaps Report. In the Round 2 Report, the evaluation of benthic risks considered the floating percentile model - effect levels 2 and 3 for the *Chironomus* growth, *Chironomus* mortality and *Hyalella* morality endpoints and the logistic regression model at the effect level 2 for the pooled *Hyalella* and *Chironomus* endpoints. Although the Round 2 report utilized the logistic regression model for the identification of Round 2 Chemicals of Potential Concern (COPCs; see Table 9.3-1 of the Round 2 Report), the logistic regression model was not used to develop initial areas of potential concern (iAOPCs) due to the following concerns:

- Irreproducibility of the logistic regression model;
- The predictive ability of the *Hyalella* growth endpoint; and,
- The reduction in predictive accuracy when combining the two models.

In addition, the logistic regression model as applied by Jay Field of NOAA relied on approximately 400 samples collected outside Portland Harbor. The LWG has objected to the inclusion of this data into the logistic regression model - especially if the data can not be made available to the LWG. USEPA has stated that the non-site data must be made available to the LWG if we are to use if for site decision making.

USEPA considered the logistic regression model and the *Hyalella* growth endpoint in our evaluation of benthic risks for the purpose of identifying Round 3B data gaps. However, during the finalization of the field sampling plan for sediment toxicity testing, USEPA and the LWG could not reach agreement on the use of the *Hyalella* growth endpoint in the application of the predictive models and instead agreed to identify sediment sampling locations, in part, based on an evaluation of the empirical *Hyalella* growth toxicity testing. It should be noted that approximately 50 additional samples were collected for toxicity testing in the fall of 2007. These data are available but have not yet been evaluated.

BERA Problem Formulation: On February 15, 2008, USEPA submitted the Problem Formulation for the Baseline Ecological Risk Assessment to the LWG. The purpose of the problem formulation was to guide the development of the baseline ecological risk assessment. Relevant risk hypotheses from the Problem Formulation include:

- Do contaminant concentrations in bulk sediments from Portland Harbor exceed sediment quality benchmarks for the survival, reproduction or growth of benthic macroinvertebrates?

- Is the survival or growth of benthic macroinvertebrates as predicted from bulk sediment chemistry below acceptable thresholds as determined by the use of modeling techniques such as logistic regression modeling or floating percentile modeling?

- Is the survival of benthic invertebrates, as indicated by the survival of the amphipod *Hyalella azteca* and the midge *Chironomus tentans* exposed to whole sediments from Portland Harbor below biological effect thresholds which represent minor, moderate, or severe levels of unacceptable effect?

- Is the growth or biomass of benthic invertebrates (*Hyalella azteca* and *Chironomus tentans*) exposed to bulk sediments from Portland Harbor below biological effect thresholds which represent minor, moderate, or severe levels of unacceptable effect?

The problem formulation required evaluation of the empirical toxicity results at the 10%, 20% and 30% difference from control level and the floating percentile model at the 20% and 30% effect level. In addition, the problem formulation required a substitution of the *Hyalella* growth endpoint with a total biomass endpoint, suggested pooling of endpoints to improve model performance, recommended incorporation of the Round 3 Data into the models, and recommended reconciling the chemicals evaluated in the two models to the extent possible.

Current Status - Post Problem Formulation Discussions: Following submittal of the problem formulation by USEPA, a series of discussions took place in an effort to resolve discrepancies between the Round 2 Report, the Problem Formulation, and previously submitted documents, such as the benthic interpretation report and the 2005 Technical Memorandum - Estimating Risks to the Benthic Community using Sediment Toxicity Tests. A number of approaches were considered including adjusting the effect levels for the *Hyalella* growth endpoint and incorporation of the RSET one-hit/two-hit approach into the floating percentile model.

Ultimately, USEPA and the LWG have not been able to reach agreement on the hit-no-hit threshold for application of the predictive models. USEPA and the LWG have agreed to substitute the total biomass endpoint for the growth endpoint for both *Hyalella* and *Chironomus*. Further, USEPA and the LWG have a tentative agreement to use the 10%, 20% and 30% difference from control for the empirical data but even this agreement is tied to agreements on the use of the predictive models.

3.0 Terms of Reference for this Evaluation

Because the LWG and USEPA have not been able to reach agreement, we have requested your assistance as an impartial reviewer to review the existing data and make recommendations about the evaluation of the empirical toxicity. Specifically we request that you evaluate the existing data and the state of the science to answer the following questions:

- What hit/no-hit criteria should be applied to the empirical sediment toxicity tests?
- What pooling of endpoints, if any, should be applied for use in each of the predictive models? Pooling may include pooling the growth (total biomass) and mortality endpoints for each test organism (2 endpoints) or both test organisms (1 endpoint) and the application of the RSET one-hit/2-hit criteria.
- What hit/no-hit criteria should be applied for the logistic regression and floating percentile models? Note that one, two or three criteria may be applied to each endpoint and each model. However, this will increase the amount of work required to develop the models.
- Should non-site data be considered in the development of the logistic regression model?
- Once the models have been run, what analysis, if any, should be performed to optimize model performance?
- Should the predictive models be used at all given their reliability?
- How should the results of the predictive models be used, in conjunction with other site data, in a weight-of-evidence (WOE) evaluation aimed at assessing risk to the benthic community?

Please provide supporting information for all recommendations.

4.0 Recommendations and Associated Rationale

Ecological risk assessment (ERA) represents an essential element of the overall RI/FS process, which is designed to support risk management decision-making for Superfund sites. More specifically, ERA provides risk managers with key information for managing contaminated sites by estimating and describing risks to ecological receptors associated with exposure to contaminated environmental media. Such information helps risk managers and other interested parties understand the ecological significance of environmental contamination at the site. The ERA process also results in determination of the concentrations of COPCs that represent thresholds for adverse effects on the selected assessment endpoints. This latter information is essential for evaluating the efficacy of the remedial alternatives that are proposed to address concerns regarding risks to ecological receptors utilizing habitats in the vicinity of Superfund sites.

At many Superfund sites, concerns relative to effects on human health and ecological receptors associated with exposure to contaminated media are focused primarily on contaminated sediments. While surface-water resources may also be contaminated, the COPCs in this medium generally originate from sediments or upland activities (e.g., point-source discharges of wastewater and non-point source releases of COPCs). When the COPCs originate from upland sources, other programs (e.g., total maximum daily load; TMDL) represent the most direct means of addressing contamination issues. Otherwise, active sediment management is needed to improve water quality conditions (i.e., when surface water is being degraded by sediment quality conditions). In addition, the tissues of aquatic organisms can be contaminated to such an extent that their consumption poses risks to ecological receptors and/or human health. In these cases, sediment-associated COPCs are frequently the primary source of the tissue contamination. Therefore, aquatic ERAs need to be designed to provide risk managers with the information they need to manage contaminated sediments. From our perspective, the Portland Harbor site does not appear to be an exception to this rule. That is, the BERA for the Portland Harbor site must be designed and implemented in a manner that provides risk managers with the information needed to effectively manage contaminated sediments.

4.1 Scope of this Evaluation

Contaminated sediments can pose unacceptable risks to ecological receptors for two main reasons. First, contaminated sediments can be directly toxic to the organisms that utilize benthic habitats at the site (i.e., microbiota, aquatic plants, benthic invertebrates, benthic fish, sediment-probing birds). Second, sediment-associated COPCs can accumulate in the tissues of aquatic organisms and, in so doing, adversely affect the organisms that feed on these prey species, either directly or indirectly through food web transfer. We understand that procedures for assessing the risks associated with exposure to bioaccumulative COPCs at the Portland Harbor site have been developed and are currently under review. Accordingly, this review is focused on evaluating the approaches that have been proposed by LWG and/or USEPA for assessing risks to benthic invertebrates at the Portland Harbor site (i.e., risks associated with toxicity to benthic invertebrates associated with exposure to contaminated sediments). More specifically, this evaluation is intended to provide the LWG and USEPA with recommendations on the following topics:

- Framework for assessing risks to benthic invertebrates;
- Procedures for designating sediment samples as toxic and not toxic (i.e., hit and no hit);
- Procedures for integrating data on multiple toxicity test endpoints;
- Procedures for evaluating relationships between sediment chemistry and sediment toxicity;
- Procedures for developing toxicity thresholds for sediment;

- Procedures for evaluating concentration-response models (e.g., logistic regression and floating percentile models; and,
- Procedures for assessing risks to benthic invertebrates.

Each of these topics are discussed in the following sections of this document. In addition, the recommendations offered on these topics were used to provide responses to each of the seven questions that were posed in the terms of reference for this evaluation.

4.2 Recommended Framework for Assessing Risks to the Benthic Invertebrate Community

The problem formulation document (USEPA 2008) describes the framework that is preferred by USEPA for assessing risks to benthic invertebrates associated with exposure to contaminated environmental media at the Portland Harbor site. The preferred approach utilizes data on multiple measurement endpoints to assess risks to benthic invertebrates, including:

- Whole-sediment toxicity;
- Whole-sediment chemistry;
- Surface-water chemistry;
- Pore-water chemistry; and,
- Invertebrate-tissue chemistry.

The analysis plan included in the problem formulation document describes how information from each LOE will be used to estimate risks to benthic invertebrates. This framework relies primarily on whole-sediment chemistry and whole-sediment toxicity data. More specifically, sediment samples are classified into one of four effect levels (i.e., 0, 1, 2, and 3) based on the observed control-adjusted response rate. In addition, each sediment sample is classified into one of four effects levels (i.e., 0, 1, 2, and 3) based on the results of the logistic regression model (LRM) and based on the floating percentile model (FPM). Under certain circumstances, the framework calls for adding an additional point to the classification score generated using the LRM or the FPM. The highest score generated by evaluating the toxicity data, the LRM, or the FPM is then used to designate the potential risk to benthic invertebrates or potential for benthic toxicity, as follows:

Classification Score	Potential for Benthic Toxicity
Blank	No Data
0	Unlikely
1	Low
2	Medium
3	High
4	Very High

A WOE framework is also described in the problem formulation document. Application of this framework is dependent on evaluating each LOE and assigning a weight that reflects scientific reliability and relevance. This information will then be used to identify and rank the LOEs for each receptor that provide the most scientifically-reliable indication of the status of each assessment endpoint from exposure to COPCs at the site and, hence, which might be the most useful for making management decisions (USEPA 2008).

The approach for assessing risks to benthic invertebrates described in the problem formulation document is not unreasonable. However, the framework could be refined to simplify the process for conducting the benthic risk assessment. More specifically, we recommend the following framework for classifying sediment samples into multiple categories based on the risks that they pose to benthic invertebrates:

- For sediment samples for which acceptable whole-sediment toxicity data are available (i.e., at minimum, the results of 10-d tests with midge, *Chironomus dilutus*, and 28-d tests with amphipods, *Hyalella azteca*; endpoints: survival and biomass), use only the existing toxicity data to classify samples into risk categories based on the observed effects on the toxicity test organisms used to evaluate the status of the benthic invertebrate community (i.e., the results of the predictive modeling should not be used to evaluate risks to benthic invertebrates for these samples). In this way, risks to benthic invertebrates can be evaluated directly based on the results of toxicity tests to either midge or amphipods. This approach will eliminate the possibility that samples will be predicted to be toxic using one or both of the predictive models (and thereby assigning an elevated risk score), when toxicity test results demonstrate that the sample is not toxic. At any location where LWG or USEPA disagrees with the classification that is assigned using this approach, toxicity identification evaluation (TIE) and/or other procedures may be conducted to provide additional information for identifying the factors that are causing or substantially contributing to the observed toxicity.

- For sediment samples for which acceptable whole-sediment toxicity data are not available (i.e., only whole-sediment chemistry data are available), use the most reliable of the predictive models to predict toxicity to benthic invertebrates associated with exposure to Portland Harbor sediments. If only limited toxicity data are available for the sediment sample, select the higher of the risk classifications from the predictive model results and the toxicity test results. This will provide a conservative basis for assessing risks to benthic invertebrates (i.e., which would tend to over-estimate rather than under-estimate risks). For any location where LWG or USEPA disagrees with the classification that is assigned using this approach, supplementary toxicity testing may be conducted to provide a more reliable basis for assessing risks to benthic invertebrates at the site.

This simplified approach to benthic risk assessment is based on the premise that whole-sediment toxicity tests are likely to provide more reliable information for evaluating effects in benthic invertebrates associated with exposure to Portland Harbor sediments than would predictive modeling. It also recognizes that the two predictive models may have different capabilities for correctly classifying sediment samples from Portland Harbor as toxic or not toxic. Accordingly, the risks to benthic invertebrates are likely to be assessed more accurately if the most reliable predictive model is used to predict sediment toxicity. It is important to acknowledge the possibility that neither of the predictive models can accurately classify sediment samples as toxic and not toxic across the entire site. In this event, it may be necessary to develop supplementary predictive models that can be used to more accurately predict toxicity for the areas that the LRM and/or FPM are shown to be less reliable. Alternatively, supplemental toxicity testing could be conducted in such areas to provide the information needed to accurately assess risks to benthic invertebrates.

At certain locations, risk managers may require additional information (i.e., beyond the risk classification for a sediment sample) to assist them in making sediment management decisions. For example, additional information may be needed when sediment samples have elevated chemistry, but are found to be not toxic to the selected toxicity test organisms and endpoints. In these cases, further data analysis and/or further sampling may be required to explain the lack of toxicity in these samples. In other cases, sediment samples may have low chemistry, but are found to be toxic to the selected toxicity test organisms/endpoints. In these cases, further data analysis and/or further sampling may be required to identify the factor or factors that are causing or substantially contributing to the observed toxicity.

4.3 Recommended Procedures for Designating Sediment Samples as Toxic or Not Toxic

At the Portland Harbor site, a number of whole-sediment toxicity tests have been conducted to evaluate the effects on benthic invertebrates associated with exposure to contaminated sediments. More specifically, 10-d whole-sediment toxicity tests with the midge, *Chironomus dilutus*, and 28-d whole-sediment toxicity tests with the amphipod, *Hyalella azteca*, have been conducted on over 300 sediment samples from the study area (Endpoints: survival and growth for both tests). In addition, information on the survival and growth of oligochaetes (*Lumbriculus variegatus*) and Asiatic clams (*Corbicula fluminea*) exposed to Portland Harbor sediments during 28-d bioaccumulation tests provides additional information for assessing sediment toxicity. Interpretation of the results of these toxicity tests requires a procedure for designating the samples as toxic (hit) or not toxic (no hit) to benthic invertebrates.

A number of approaches can be used to interpret the results of whole-sediment toxicity tests with benthic invertebrates. These approaches can be classified into four general categories, including control comparison approach, minimum significant difference (MSD) approach, reference envelope approach, and the multiple category approach. Each of these approaches are briefly described below:

- ***Control Comparison Approach*** - Application of the control comparison approach involves statistical comparison of the responses of test organisms exposed to site sediments to the responses of test organisms exposed to control sediments. Treatments that have responses that are significantly different from those observed in the control treatment(s) are designated as toxic.

- ***Minimum Significant Difference Approach*** - Application of the MSD approach is dependent on the completion of power analyses with data from multiple studies for a specific toxicity test. These results are used to identify the MSD (or minimum detectable difference) from the control treatment. Treatments with response levels greater than the MSD are designated as toxic (Thursby *et al.* 1997; Phillips *et al.* 2001).

- ***Reference Envelope Approach*** - Application of the reference envelope approach involves collection and testing of sediment samples from a number of reference sites within or nearby the study area. In this context, a reference sediment sample is considered to be whole-sediment obtained near an area of concern used to assess sediment conditions exclusive of the materials of interest (i.e., COPCs; ASTM 2007). The results of the toxicity testing conducted on these samples can be used to develop a reference envelope (i.e., normal range of responses of test organisms exposed to reference sediments, as defined by ASTM 2007). Sediment samples with response levels that fall outside the normal range of responses (e.g., survival below the 5th percentile for the reference samples) are designated as toxic.

- ***Multiple Category Approach*** - Application of the multiple category approach involves classifying sediment samples into various groups (e.g., not toxic, low toxicity, moderate toxicity, or high toxicity), based on the magnitude of the observed response. The results of statistical comparisons to the negative control results are also used to classify sediment samples into the various categories.

According to the information presented in the problem formulation document, a multiple category approach has been selected for interpreting the results of whole-sediment toxicity tests conducted using sediments obtained from Portland Harbor. More specifically, sediment samples will be classified into effects level 0, 1, 2, or 3 if control-adjusted response rates are >90%, 80 - 90%, 70 - 80%, and <70% respectively. In order for effects to be considered significant, the response must be statistically-significantly different from the negative control response at the $p < 0.05$ level.

Recently (2007), the Sustainable Fisheries Foundation (SFF) convened a workshop in Victoria on behalf of the B.C. Ministry of the Environment to explore the question of how to interpret the results of sediment toxicity tests (SFF 2007). At this workshop, participants agreed that site-wide ecological risk assessments represent the most important applications of whole-sediment toxicity data. More specifically, it was agreed that the results of the toxicity testing program that is implemented at a site should support the development of site-specific toxicity thresholds (i.e., to support development of preliminary remediation goals and/or clean-up goals). In this context, workshop participants agreed that designation of samples as toxic or not toxic is not necessarily required early in the site assessment process. Rather, the magnitude of effect data can be used directly in the development of concentration-response relationships for COPCs at the site. The magnitude of effect data can also be used to classify sediment samples into risk categories, without having to designate individual sediment samples groups as toxic or not toxic. This approach to the interpretation of whole-sediment toxicity data was considered to be desirable because no information is lost during the interpretation process. Hence, workshop participants generally agreed with the approach that has been described for use in Portland Harbor (USEPA 2008).

Workshop participants also recognized that interpretation of toxicity test results may necessitate designation of individual sediment samples as toxic or not toxic (e.g., hot spot identification, evaluation of the spatial extent of toxicity). In these cases, workshop participants agreed that a step-wise approach should be used to interpret the results of individual toxicity tests. We have reviewed the approach suggested by workshop participants and refined it to recommend a toxicity designation process for the Portland Harbor site that consists of the following steps:

- Conduct whole-sediment toxicity tests in accordance with standardized protocols, as described in the project Quality Assurance Project Plan (QAPP);
- Evaluate the validity of each whole-sediment toxicity test. The project data quality objectives, which are documented in the QAPP, should define the performance criteria for measurement data that will be used to evaluate toxicity test acceptability. At minimum, such performance criteria should define the acceptable range of negative control and positive control (i.e., reference toxicant) results. Evaluation of potential test interferences should also be conducted during this step in the process (e.g., comparison of ammonia and hydrogen sulfide levels to lowest observed effect concentrations for the test species, conducting Spearman Rank correlation analysis);
- Compare the results obtained for each sediment sample to the negative control results for the corresponding batch of samples. Sediment samples for which the measured response is significantly greater than that for the negative control (i.e., a one-tailed statistical test would be used) should be tentatively identified as toxic;
- Compare the toxicity test results obtained for each sediment sample to the reference envelope developed for the corresponding toxicity test endpoint. Sediment samples that were tentatively identified as toxic based on the previous step of the process (i.e., based on comparison to the results for the negative control treatment) would be designated as toxic if the measured response is greater than the lower limit of responses for reference sediment samples (e.g., if the reference envelope for amphipod survival in a 28-d whole-sediment toxicity test is 77 to 98%, then sediment samples for which amphipod survival is less than 77% would be designated as toxic). In general, control-adjusted response rates for reference sediment samples should be used to develop the reference envelope because the negative control results for multiple batches of samples are likely to be different; and,

- Sediment samples that are designated as toxic using both the reference envelope and control comparison approaches should be identified as those that pose the highest risks to the benthic invertebrate community. Sediment samples for which the response of the test organism falls within the reference envelope should not be designated as toxic and should be considered to pose the lowest risks to the benthic invertebrate community.

Participants at the SFF workshop also indicated that the MSD approach can be used to designate sediment samples as toxic or not toxic. While the MSD approach could also be applied at the Portland Harbor site, MSDs have not yet been developed for the four toxicity tests that have been used to evaluate the toxicity of sediments at the site. While such MSDs are currently under development, they are unlikely to be available within the time frame required to support the Portland Harbor BERA (C.G. Ingersoll, United States Geological Survey. Personal communication).

All of the participants at the SFF workshop recognized that the results of individual whole-sediment and pore-water toxicity tests may be used within a WOE framework for evaluating risks to the benthic invertebrate community associated with exposure to contaminated sediments. Workshop participants agreed that such WOE evaluations require information on the magnitude of toxicity in addition to, or instead of, toxicity designation information. Hence, it was generally agreed that the information on the magnitude of the response be retained to support further analyses of the toxicity data (i.e., WOE evaluations). Such WOE evaluations can be used to classify sediment samples into categories based on the magnitude of risk that they pose to benthic invertebrates. However, such categories are not relevant for determining if individual samples are toxic or not toxic.

4.4 Recommended Procedures for Developing a Reference Envelope for Interpreting Data from Whole-Sediment Toxicity Tests

Based on the information that was provided to support this evaluation, a multiple category approach has been proposed by USEPA (2008) for the Portland Harbor site. We believe that the reference envelope approach will complement the multiple category approach by providing a robust and defensible basis for designating sediment samples from the study area toxic or not toxic. Therefore, it is recommended that LWG and USEPA include the reference envelope approach in the process that will be used to interpret the results of whole-sediment toxicity tests conducted with sediment samples from Portland Harbor (as described in Section 4.3).

In general, application of the reference envelope approach necessitates identification of candidate reference sites as part of the overall sampling program design. Accordingly, LWG (2005b) indicated that whole-sediment toxicity testing would be conducted on a total of six upstream ambient stations "to place the results for the study area in a regional context". While these data represent an important element of the overall sediment sampling program, they may not be sufficient to define reference conditions for the Portland Harbor site. Our experience at other sites suggests that about 15 sediment samples are needed to adequately characterize variability in the responses of toxicity test organisms associated with exposure to reference sediments. It is understood that three rounds of toxicity testing have already been completed and that both LWG and USEPA have an interest in completing the BERA in a timely manner. Therefore, the following procedure is recommended for developing reference envelopes for the toxicity test endpoints that have been used to characterize sediment quality conditions at the Portland Harbor site:

- Identify sediment samples from the study area that are representative of reference conditions. Candidate reference sediment samples can be identified on an *a posteriori* basis by applying a series of criteria for sediment chemistry and sediment toxicity. More specifically, the following criteria for whole-sediment chemistry are recommended for identifying candidate reference sample (USEPA 2003; 2005; MacDonald *et al.* 2007):

- All measured metals, polycyclic aromatic hydrocarbons (PAHs), and polychlorinated biphenyls (PCBs) occur at concentrations below conservative sediment quality guidelines (SQGs);
- Mean $PEC\text{-}Q_{DW} < 0.1$;
- $\sum ESB\text{-}TU_{PAHs} < 0.1$; and,
- $(\sum SEM\text{-}AVS)/f_{oc} < 130$ µmol/g.

Candidate reference samples that meet the criteria for whole-sediment chemistry should be further evaluated to confirm that they were not toxic to sediment-dwelling organisms. More specifically:

- Control-adjusted response rate should not exceed the MSD for each toxicity test endpoint; or,
- In the absence of MSD values, control-adjusted response rate should not exceed the Tier II levels applied in the National Sediment Inventory (USEPA 2004);

These biological criteria should be applied to ensure that samples for which the biological response may have been adversely affected due to the presence of unmeasured COPCs (or COPCs for which SQGs are not available) are not used in the reference envelope calculation. Sediment samples that meet both the chemical and biological criteria should be selected as reference samples for the study area.

- Determine the normal range of toxicological responses for each toxicity test conducted and endpoint measured. The reference envelope is commonly calculated in a manner such that it encompasses 95% of the variability in the response data. While several procedures can be used to calculate the reference envelope, we recommend calculating the lower limit of the reference envelope as the 5th percentile of the control-adjusted response data for each toxicity test and endpoint. It is recommended that the response data be log-transformed prior to calculating the 5th percentile response level. The normal range of reference responses spans the range from the 5th percentile value to the maximum value in the data set.

- Designate sediment samples with control-adjusted effect values lower than the lower limit of the normal range of control-adjusted responses in reference samples (i.e., lower than the 5th percentile) as toxic for the endpoint under consideration (see Appendix E2 of the MacDonald *et al.* 2002 for a more detailed description of these procedures).

As indicated in Section 4.3, the criteria for statistical difference from the control would also need to be met to designate a sediment sample as toxic using the reference envelope approach. It is important to note that application of this approach results in the designation of toxicity on an endpoint-by-endpoint basis. Therefore, a single sample can be designated as toxic for certain endpoints and not toxic for other endpoints. This reflects differences in species sensitivity and response to different mechanisms of toxic action, as represented by the mixture of contaminants in the sediments.

4.5 Recommended Procedures for Integrating Data on Multiple Toxicity Test Endpoints

The concept of pooling multiple endpoints for a toxicity test and/or multiple endpoints from multiple toxicity tests has been proposed for interpreting the whole-sediment toxicity data for the Portland Harbor site, particularly for use in predictive modeling of sediment toxicity. It is our recommendation that multiple endpoints should not be pooled, either to support interpretation of the whole-sediment toxicity data or to support the development of predictive models. Rather, we believe that each endpoint provides unique information that can be used to support assessment of risks to benthic invertebrates, the development of predictive models, and the derivation of site-specific toxicity thresholds [including preliminary remediation goals (PRGs) and/or clean-up goals].

From a toxicological perspective, organisms can be differentially sensitive to contaminants because of differences in exposure conditions, differences in biotransformation rates, and differences in receptor sensitivities to the active toxicant. This suggests that each endpoint provides information on the response of the toxicity test organism to the mixture of COPCs in the sediments at the site. Such responses may be different from those of other species or toxicity test endpoints, thereby representing a unique response to the exposure. Examples of this can be found in the literature where a species shows responses to different contaminants at different concentration levels, even without considering the differences in exposure conditions (Hwang *et al.* 2004). Figures 1 to 3 provide plots of the relationships between amphipod survival and amphipod biomass, midge survival, and midge biomass at another site in the U.S. These results indicate that the response of the toxicity test organisms are not well correlated with one another. That is, these toxicity test endpoints frequently provide unique information on the toxicity of sediment samples. By refining these plots in a way that conveys information on the COPC mixture in each sample (e.g., which class of COPC has the largest hazard quotient) or geographic location (e.g., area of interest), patterns can emerge that can help interpret the toxicity test results. Such information could be lost if the test results are pooled for different endpoints or different toxicity tests.

Information from multiple toxicity tests and multiple toxicity test endpoints can, however, be considered together to help prioritize areas of interest within a site that may be considered for source control or other sediment management actions. In such evaluations, each toxicity test endpoint can provide a unique LOE for assessing sediment quality conditions. Sediment samples that are found to be toxic for more than one toxicity test endpoint may be assigned a higher priority than those that are found to be toxic relative to a single toxicity test endpoint. However, it is also important to consider the endpoint measured and the magnitude of the response in such a prioritization process. It is also important to remember that certain COPCs and/or COPC mixtures can be especially toxic to certain test organisms (Schuler *et al.* 2006). Therefore, finding a single significant toxic response using the criteria of significant difference from control and the reference envelope approach would suggest that there are conditions of concern in the sediment (i.e., exposure to such sediments poses potential risks to benthic invertebrates). Risk managers must utilize this information when considering alternatives for addressing such risks (e.g., collecting additional information to further evaluate the nature and extent of contamination, to further evaluate sediment toxicity, to identify the factors that are causing or substantially contributing to the observed effects, monitored natural attenuation, active remediation).

From a modeling perspective, focusing on a single endpoint for each model provides a more consistent data set than an approach that attempts to combine endpoints. Such pooling of endpoints could easily result in conflicting results, where one endpoint provides no hit data and another endpoint provides a hit. This makes the modeling less reliable and more variable than would be the case if each endpoint is considered separately in the development and evaluation of the various models. This problem was clearly evident in the data presented in the LWG (2006) report.

For the purpose of modeling, survival and biomass are the two toxicity test endpoints that should be considered for the amphipod and midge tests. The use of biomass as a substitution for the growth endpoint corrects for the problem that occurs with the growth endpoint when changes in nutrient availability due to reduction in numbers of organisms in a replicate influence the growth of surviving organisms in that replicate (i.e., these types of data are evident in the Round 2 data report). Thus, by making a series of models for the different endpoints, each model can be compared to the existing data to determine which performs the best in terms of correctly predicting the presence and absence of toxicity for each sample (on an endpoint-by-endpoint basis).

As indicated above, each endpoint should be evaluated separately for each model. In addition, both modeling approaches should use the same criteria (i.e., modified reference envelope approach) for what constitutes a hit or no-hit for the toxicity test endpoint under consideration. In this way, the models will be generated using comparable data sets and the outputs of the models can be directly compared. Subsequently, the more reliable models can be identified and selected for use in the BERA. The use of different terms of reference for the two modeling approaches can lead to predictions that have different meanings. There is no toxicological reason to believe that the criteria for selecting endpoints or designating samples as toxic or not toxic should be different for the two models. Thus, for consistency in comparing the utility of the models and for understanding the predictions, we recommend that the same criteria, as outlined above, be employed for both modeling efforts.

4.6 Recommended Procedures for Evaluating Relationships Between Sediment Chemistry and Sediment Toxicity

There are a number of approaches that could be used to evaluate the relationships between whole-sediment chemistry and whole-sediment toxicity at the Portland Harbor site. Based on the information presented in LWG (2006) and USEPA (2008), the logistic regression model and the floating percentile model are the two approaches that are currently being considered and tested for the Portland Harbor site. These models are being developed to provide accurate predictions of sediment toxicity for sediment samples for which only whole-sediment chemistry data are available to evaluate sediment quality conditions. That is, the model must result in the identification of toxicity thresholds for COPCs and/or COPC mixtures that provide a reliable basis for classifying such sediment samples as toxic or not toxic. Accordingly, these models must be able to incorporate all the identified COPCs and toxicity test endpoints within the modeling framework.

The two models that have been identified for use at the Portland Harbor site both have the potential to provide risk assessors with the tools needed to support the BERA (i.e., toxicity thresholds that accurately classify sediment samples from the Portland Harbor site as toxic and not toxic). Therefore, it is recommended that predictive modeling be included in the overall framework that is used to evaluate risks to the benthic invertebrate community at the Portland Harbor site.

The use of matching whole-sediment chemistry and whole-sediment toxicity data from the Portland Harbor site in the development of such predictive models represents a reasonable approach for deriving toxicity thresholds for COPCs and COPC mixtures at the site. However, there is no reason to believe that data from other freshwater sites cannot be used to generate relationships between sediment chemistry and sediment toxicity. While certain data from other sites could be fundamentally different from those for the site (i.e., due to differences in the underlying geology or due to differences in the binding phases that alter contaminant bioavailability), the toxicity thresholds that are derived using the predictive models will be evaluated to determine their performance in terms of predicting toxicity at the Portland Harbor site. The toxicity thresholds that perform the best (i.e., that provide the most accurate basis for classifying sediment samples as toxic and not toxic) should be selected to support the BERA. Therefore, the use of non-site data in model development does not represent a substantive issue relative to application of the various models. On the contrary, by using additional data in model development, the potential for variation in response due to differences in habitat or other factors can be incorporated into the model. Therefore, use of non-site data could improve the models that are developed for the site.

In addition to the two modeling approaches that have been explicitly identified to date, there are other modeling approaches that could be used to describe the matching sediment chemistry and sediment toxicity data from the site (see MacDonald *et al.* 2003; 2005a; 2005b; 2008 for examples). In addition, it may be necessary to develop Area of Interest-specific models to describe such relationships in areas within the site that have unique COPCs, COPC mixtures, or COPC concentration gradients. The need for additional models should be evaluated following the evaluation of the site-wide models that are developed using the LRM and FPM approaches.

4.7 Recommended Procedures for Developing Toxicity Thresholds

There are a wide variety of approaches that can be used to develop toxicity thresholds for COPCs and/or COPC mixtures in sediments. The LRM and FPM approaches that have been selected for use at the Portland Harbor site both have established procedures for deriving toxicity thresholds based on the modeling results. These procedures are reasonable and can be used to establish candidate toxicity thresholds for use in the BERA.

At this stage of the process, it is important to explicitly identify the narrative intent of any toxicity thresholds that are developed using the predictive models. For example, MacDonald *et al.* (2003) developed two types of toxicity thresholds for selected COPCs and COPC mixtures. More specifically, these investigators derived low risk and high risk toxicity thresholds for selected COPCs and COPC mixtures. The low risk toxicity thresholds were intended to identify the concentrations of COPCs or COPC mixtures below which adverse effects on benthic invertebrates were unlikely to be observed (i.e., fewer than 20% of the sediment samples would be toxic to benthic invertebrates). These low risk toxicity thresholds were established at COPC/COPC mixture concentrations that corresponded to a 10% increase in the magnitude of toxicity to selected toxicity test organisms, relative to the average response rates for toxicity test organisms exposed to reference sediment samples. In contrast, the high risk toxicity thresholds were intended to identify the concentrations of COPCs or COPC mixtures above which adverse effects on benthic invertebrates were likely to be observed frequently (i.e., more than 50% of the sediment samples would be toxic to benthic invertebrates). These high risk toxicity thresholds were established at COPC/COPC mixture concentrations that corresponded to a 20% increase in the magnitude of toxicity to selected toxicity test organisms, relative to the average response rates for toxicity test organisms exposed to reference sediment samples. By explicitly establishing the narrative intent of the toxicity thresholds, it is possible to develop criteria for evaluating the performance of the resultant toxicity thresholds that directly reflect the intended uses of the toxicity thresholds. Therefore, it is recommended that the narrative intent of the toxicity thresholds for the Portland Harbor site be explicitly described. In general, the remedial action objectives that are established for the site will provide a relevant basis for determining the narrative intent of the toxicity thresholds.

4.8 Procedures for Evaluating Concentration-Response Models

LWG (2006) identified seven reliability parameters for evaluating existing SQVs and the model predictions, including false positives, false negatives, sensitivity, efficiency, predicted hit reliability, predicted no-hit reliability, and overall reliability. However, it is not clear that the narrative intent of these SQVs was considered during the evaluation process. For example, the threshold effect levels (TELs) and similar values are intended to identify the concentrations of COPCs or COPC mixtures below which adverse effects on benthic invertebrates would be infrequently observed (i.e., in fewer than 10% of the samples). In contrast, the probable effect levels (PELs) and similar values are intended to identify the concentrations of COPC or COPC mixtures above which adverse effects on benthic invertebrates would be frequently observed (i.e., greater than 50% of the sediment samples would be toxic). It is not clear from the analysis presented in LWG (2006) how the narrative intent of the SQVs was considered in the evaluation process. Without considering information on the narrative intent of the SQVs, it is not possible to determine how applicable certain SQVs could be for predicting the presence or absence of sediment toxicity at the Portland Harbor site. Therefore, a suite of candidate SQVs should be identified that are consistent with the narrative intent of toxicity thresholds for the Portland Harbor site and these candidate SQVs should be evaluated using the same criteria and data that are used to evaluate the site-specific toxicity thresholds derived using the LRM and the FPM.

As indicated in LWG (2006), evaluation of the toxicity thresholds that are developed using the LRM and FPM represents the most important part of the predictive modeling process. However, it is essential to establish the narrative intent of the toxicity thresholds that are developed using the predictive models to ensure that the evaluation process is fair and relevant. That is, information on the narrative intent of the toxicity thresholds should be used to establish the criteria that will be used in the evaluation process.

Once the evaluation criteria have been established, the models can be developed and their performance can be evaluated relative to the criteria. Two general types of evaluations are recommended, including reliability of the toxicity thresholds and predictive ability of the toxicity thresholds. In this context, reliability is defined as the ability of the toxicity thresholds to correctly classify the sediment samples that are used to develop the model as toxic and not toxic. In contrast, predictive ability is defined as the ability of the toxicity thresholds to correctly classify sediment samples as toxic and not toxic for an independent data set (i.e., data that were not used in the model development process).

For Portland Harbor, matching sediment chemistry and sediment toxicity data are available for more than 300 sediment samples. Most of these data have been used to develop the existing FPMs and LRMs. However, there is a whole new set of data that has been collected (50 samples) which might be excluded from formulation of the model and used as a validation data set. Alternatively, the entire data set could be split into two sub-sets, one of which could be used to re-develop the models (i.e., using data for about 200 sediment samples) and the second could be used to evaluate the predictive ability of the models (i.e., using the data for about 100 sediment samples). If the second approach is used, it may be useful to stratify the data into quartiles based on sediment chemistry (e.g., mean PEC-Qs) and randomly select 25 sediment samples from each quartile for use in the predictive-ability evaluation. The remainder of the data could be used to develop the models and evaluate their predictive ability.

The criteria that were established by LWG (2006) could be refined prior to evaluating the reliability and predictive ability of the models. More specifically, it may be useful to refine the evaluation criteria to align them better with the remedial action objectives for the site. In this case, a low risk toxicity threshold would be considered to be reliable and predictive if, for example, the incidence of sediment toxicity is low (e.g., < 10%) for sediment samples with COPC or COPC mixture concentrations below the toxicity threshold. In contrast, a high risk toxicity threshold would be considered to be reliable and predictive if, for example, the incidence of sediment toxicity is high (e.g., > 50%) for sediment samples with COPC or COPC mixture concentrations above the toxicity threshold. An intermediate incidence of toxicity might be expected at concentrations of COPCs or COPC mixtures between the low risk and high risk toxicity thresholds. The point is, it is not unreasonable to expect that multiple toxicity thresholds may be required to provide risk assessors and risk managers with the tools that they need to evaluate and manage contaminated sediments at the Portland Harbor site. The results of the reliability and predictive-ability evaluations will provide risk assessors and risk managers with the information that they need to select the tools required to support the RI/FS.

The obvious should also be pointed out. That is, none of the models are without limitations. Neither model can be considered to provide any direct information about cause and effect. Although the Pmax logistic regression model does provide some insight. Both models are making correlations between a gross chemistry value and the observed toxicity response without regard to issues such as bioavailability or the mixture of chemicals at the various stations. This is particularly true for the floating percentile model that does not attempt to address mixture response in any manner but uses the correlations for each chemical to produce a separate acceptable value for a specific chemical. The logistic regression model can use either a sum probability or the more usual probability max approach to incorporate response addition as the likely interaction of compounds in the sediment (Field *et al.* 2002). It would be helpful to present the results of LRM for both Pmax and Pavg because the two versions of the COPC mixture model can provide different information about the sediment samples. The logistic regression approach has been peer reviewed and published to provide additional reliance in its acceptability. However, the model that is selected for use in Portland Harbor should be the one that provides the best predictions of toxicity after fully developing the models and comparing the results to a validation data set.

4.9 Recommended Procedures for Assessing Risks to Benthic Invertebrates

A WOE approach is recommended for assessing risks to benthic invertebrates at the Portland Harbor site (as described in Section 4.1). Models are not perfect and all LOEs should be employed to make the best decision possible about the status of a station. It is particularly important to consider the spatial data if the model predicts a different result than is observed at nearby stations. Then, depending on the importance of the decision to be made, additional sampling and analysis (including additional toxicity testing) may be required.

5.0 Summary and Conclusions

Over the past few years, the LWG and USEPA have prepared a variety of technical reports and engaged in a number of technical discussions in an effort to come to agreement on the procedures that should be used to evaluate risks to benthic invertebrates at the Portland Harbor Superfund Site. While substantial progress has been made in certain areas (e.g., sediment sampling and characterization), there are several issues that have not yet been resolved. This is important because both LWG and USEPA are interested in completing the BERA component of the RI/FS and uncertainty regarding these outstanding issues is likely to impede progress towards this goal.

Recognizing that several key issues need to be resolved in the near-term to keep the project on schedule, LWG and USEPA agreed to have Don MacDonald and Peter Landrum conduct an independent evaluation of the various approaches for assessing risks to benthic invertebrates at the Portland Harbor site. To facilitate this evaluation, the various documents pertaining to the benthic invertebrate portion of the BERA, prepared by LWG or USEPA, were provided to these reviewers. In addition, the reviewers were provided with access to the data and information that have been collected to date at the site. Furthermore, the reviewers were provided with background information considered to be particularly relevant to understanding the unresolved issues.

This document summarizes the recommendations that are offered by Don MacDonald and Peter Landrum for assessing risks to benthic invertebrates at the Portland Harbor Site. More specifically, Section 4.1 to 4.8 of this document outline the recommended procedures for assessing risks to the benthic invertebrate community at the site. These recommendations are summarized in the following responses to the seven questions that were posed to help structure this review:

1. *What hit/no-hit criteria should be applied to the empirical sediment toxicity tests?*

   Response: The whole-sediment toxicity data should be designated as toxic (hit) or not toxic (no hit) using the modified reference envelop approach (as described in Section 4.3). In this approach, the toxicity of sediment samples is evaluated on an endpoint-by-endpoint basis. A sediment sample is designated as toxic for a specific endpoint if the response of the toxicity test organism exposed to sediment from the site is significantly greater than the response of toxicity test organisms exposed to negative control sediment and if the response falls outside the normal range of responses for reference sediment samples (i.e., outside the reference envelope). It is clear from the information in LWG (2006) that the Level 1 hit/no hit criteria includes samples in the hit category that are not statistically different from reference conditions. This decision likely added variability to the modeling exercise.

2. *What pooling of endpoints, if any, should be applied for use in each of the predictive models? Pooling may include pooling the growth (total biomass) and mortality endpoints for each test organism (2 endpoints) or both test organisms (1 endpoint) and the application of the RSET one-hit/2-hit criteria.*

   Response: Endpoints should not be pooled, either for the purpose of interpreting toxicity test results or for the purpose of developing predictive models and the associated toxicity thresholds. Each endpoint provides potentially unique information about the station and a hit from one endpoint should be sufficient to question the character of the station. Therefore, survival and biomass of midge and survival and biomass of amphipods are the four endpoints that should be evaluated in the predictive modeling process.

3. *What hit/no-hit criteria should be applied for the logistic regression and floating percentile models? Note that one, two or three criteria may be applied to each endpoint and each model. However, this will increase the amount of work required to develop the models.*

   Response: The toxicity designations that are used to support interpretation of the results of the empirical whole-sediment toxicity tests should be used in evaluating both of the predictive models (i.e., LRM and FPM) because there is no toxicological justification for selecting different criteria for different modeling structures.

4. *Should non-site data be considered in the development of the logistic regression model?*

   Response: There is no reason why non-site data cannot be used to develop either the LRM or the FPM. The most important step in the process is to evaluate the performance of the models utilizing the site-specific data. Only those models that have the best performance and least uncertainty should be used in the BERA. The data set for Portland Harbor is relatively small for model development purposes, so it makes sense to use appropriate non-site data if this leads to improved model prediction (performance).

5. *Once the models have been run, what analysis, if any, should be performed to optimize model performance?*

   Response: The performance of the models should be evaluated by determining the reliability and predictive ability of the toxicity thresholds that are derived using the models. While the reliability of the models was evaluated in the LWG (2006) document using seven criteria, these criteria should be refined to better reflect the narrative intent of the toxicity thresholds that are being evaluated and the remedial action objectives that are established for the site. Other candidate sediment quality values should also be evaluated using these site data to determine which ones may be the most reliable for evaluating risks to sediment-dwelling organisms at the Portland Harbor site. The results of such evaluations will provide a basis for determining which model provides the most accurate basis for predicting toxicity at sampling locations for which sediment-chemistry data represent the principal LOE for assessing risks to benthic invertebrates.

It is important to evaluate models equally and consistently using the data from the site. Therefore, model performance should be evaluated on an endpoint-by-endpoint basis. Subsequently, these results can be integrated to determine overall model performance at the site. The uncertainty of the model predictions should be provided as part of the information to allow for improved interpretation of the model prediction.

The reliability of the toxicity thresholds should be evaluated using the data that were used to develop the models. The predictive ability of the toxicity thresholds should be evaluated using an independent data set. In this respect, there should be a portion of the data set that is set aside for model validation that is not used for model development. Testing on an independent data set is generally accepted as the appropriate approach to evaluating model performance. The independent data set should be representative of the data as a whole for both contaminant concentrations and organism response. We recognize that the data set for Portland Harbor is relatively small for the purpose of model development, however; it should be possible to set aside 20 to 30% of the data for a validation set. The size of the Portland Harbor data set is one of the reasons that inclusion of non-site data for the development of the model should be considered.

6. *Should the predictive models be used at all given their reliability?*

Response: Insufficient model development and evaluation has been completed to fully assess the reliability of the predictive models that are proposed for use at the site. Therefore, it is recommended that a systematic model development process be undertaken to create high-quality models. Subsequently, the model results should be evaluated to determine how well the resultant toxicity thresholds predict the presence and absence of sediment toxicity at the Portland Harbor site. If the results of these evaluations show that one or both of the models cannot be applied to reliably predict the presence and absence of sediment toxicity throughout the site, additional toxicity testing should be conducted in the areas where the models are thought to be unreliable. Alternatively, area-specific models might be developed that provide a more reliable basis for predicting sediment toxicity in specific areas.

7. *How should the results of the predictive models be used, in conjunction with other site data, in a weight-of-evidence evaluation aimed at assessing risk to the benthic community?*

   Response: Risks to benthic invertebrates associated with exposure to sediments at the Portland Harbor site should be evaluated differently, depending on the types of data that are available for a sampling location. If the minimum whole-sediment toxicity data (i.e., survival and biomass of midge in 10-d exposures and survival and biomass of amphipods in 28-d exposures) are available for a sampling location, then these data should be used preferentially to assess risks to benthic invertebrates (as stated in LWG 2006). If the requisite whole-sediment toxicity data are not available for a sampling location, then the most reliable predictive model should be used, in conjunction with any toxicity data that are available, to assess risks to benthic invertebrates. In addition, the prediction should be compared to nearby stations of similar characteristics (chemistry, geology, etc.) that include toxicity information to help inform whether to trust the prediction results. Even comparison to stations that are some distance away, but have similar physical/chemical characteristics and have toxicity information, could lead to improved interpretation of the validity of the prediction. Furthermore, the potential for a station to follow a concentration/toxicity gradient can add information about the validity of the prediction. Examination of the data for the samples where chemistry and toxicity are not well correlated can provide additional insights on the bioavailability of COPCs. In any case where the prediction seems questionable, additional chemical and/or toxicity testing is recommended to resolve the issue.

In response to a preliminary review by USEPA personnel, an addendum was prepared to further clarify some of the responses included in this document. This addendum is attached to this document.

6.0 References

ASTM (American Society for Testing and Materials). 2007. Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. E1706-00. Section Eleven: Water and Environment Technology. West Conshohocken, Pennsylvania.

Field, L.J., D.D. MacDonald, S.B. Norton, C.G. Ingersoll, C. Severn, D.E. Smorong, and R.A. Lindskoog. 2002. Predicting amphipod toxicity from sediment chemistry using logistic regression models. Environmental Toxicology and Chemistry 21:1993-2005.

Hwang, H., S.W. Fisher, K. Kim, and P.F. Landrum. 2004. Comparison of the toxicity using body residues of DDE and select PCB congeners to the midge, *Chironomus riparius*, in partial-life cycle tests. Archives of Environmental Contamination and Toxicology 46:32-42.

LWG (Lower Willamette Group). 2004. Portland Harbor RI/FS. Technical memorandum: Provisional toxicity reference value selection for the Portland Harbor preliminary ecological risk assessment. Portland, Oregon.

LWG (Lower Willamette Group). 2005a. Portland Harbor Superfund Site Ecological Risk Assessment: Estimating Risks to Benthic Organisms Using Sediment Toxicity Tests. Prepared by Windward Environmental, TerraStat Consulting Group, and Avocet Consulting . Seattle, Washington.

LWG (Lower Willamette Group). 2005b. Portland Harbor RI/FS. Round 2A data report. Sediment toxicity testing. Prepared by Windward Environmental. Seattle, Washington.

LWG (Lower Willamette Group). 2006. Portland Harbor Superfund Site Ecological Risk Assessment: Interpretative Report: Estimating Risks to Benthic Organisms Using Predictive Models Based on Sediment Toxicity Tests. Prepared by Windward Environmental, TerraStat Consulting Group, and Avocet Consulting. Seattle, Washington.

MacDonald, D.D., C.G. Ingersoll, D.R.J. Moore, M. Bonnell, R.L. Breton, R.A. Lindskoog, D.B. MacDonald, Y.K. Muirhead, A.V. Pawlitz, D.E. Sims, D.E. Smorong, R.S. Teed, R.P. Thompson, and N. Wang. 2002. Calcasieu Estuary remedial investigation/feasability study (RI/FS): Baseline ecological risk assessment (BERA). Technical report plus appendices. Contract No. 68-W5-0022. Prepared for CDM Federal Programs Corporation and United States Environmental Protection Agency. Dallas, Texas.

MacDonald, D.D., R.L. Breton, K. Edelmann, M.S. Goldberg, C.G. Ingersoll, R.A. Lindskoog, D.B. MacDonald, D.R.J. Moore, A.V. Pawlitz, D.E. Smorong, and R.P. Thompson. 2003. Development and evaluation of preliminary remediation goals for selected contaminants of concern at the Calcasieu Estuary cooperative site, Lake Charles, Louisiana. Prepared for United States Environmental Protection Agency, Region 6. Dallas, Texas.

MacDonald, D.D., C.G. Ingersoll, D.E. Smorong, L. Fisher, C. Huntington, and G. Braun. 2005a. Development and evaluation of risk-based preliminary remediation goals for selected sediment-associated contaminants of concern in the West Branch of the Grand Calumet River. Prepared for: United States Fish and Wildlife Service. Bloomington, Indiana.

MacDonald, D.D., C.G. Ingersoll, A.D. Porter, S.B Black, C. Miller, Y.K. Muirhead. 2005b. Development and evaluation of preliminary remediation goals for aquatic receptors in the Indiana Harbor Area of Concern. Technical Report. Prepared for: United States Fish and Wildlife Service. Bloomington, Indiana and Indiana Department of Environmental Management. Indianapolis, Indiana.

MacDonald, D.D., D.E. Smorong, D.G. Pehrman, C.G. Ingersoll, J.J. Jackson, Y.K. Muirhead, S. Irving, and C. McCarthy. 2007. Conceptual field sampling design - 2007 sediment sampling program of the Tri-State Mining District. Prepared for U.S. Environmental Protection Agency. Region VI, Dallas, Texas. Region VII, Kansas City, Kansas.

MacDonald, D.D., D.E. Smorong, C.G. Ingersoll, J.M. Besser, W.G. Brumbaugh, N. Kemble, T.W. May, S. Irving, and M. O'Hare. 2008. Evaluation of the matching sediment chemistry and sediment toxicity in the Tri-State Mining District (TSMD), Missouri, Oklahoma, and Kansas. Preliminary Draft. Prepared for United States Environmental Protection Agency Region 6, and Region 7 and United States Fish and Wildlife Service, Columbia, Missouri. Prepared by MacDonald Environmental Sciences Ltd. Nanaimo, British Columbia. United States Geological Survey. Columbia, Missouri and CH2M Hill. Dallas, Texas.

Phillips, B.M., J.W. Hunt, B.S. Anderson, H.M. Puckett, R. Fairey, C.J. Wilson, and R. Tjeerdema. 2001. Statistical significance of sediment toxicity test results: Threshold values derived by the detectable significance approach. Environmental Toxicology and Chemistry 20:371-373..

Schuler, L.J., P.F. Landrum, and M.J. Lydy. 2006. Comparative toxicity of fluoranthene and pantachlorobenzene to three freshwater invertebrates. Environmental Toxicology and Chemistry 25:985-994.

SFF (Sustainable Fisheries Foundation). 2007. Workshop to support development of guidance on the assessment of contaminated sediments in British Columbia. Prepared for B.C. Ministry of Environment. Victoria, British Columbia.

Thursby, G.B., J. Heltshe, and K.J. Scott. 1997. Revised approach to toxicity test acceptability criteria using a statistical performance assessment. Environmental Toxicology and Chemistry 16(6):1322-1329.

USEPA (United States Environmental Protection Agency). 2003. Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms: PAH mixtures. EPA-600-R-02-013. Office of Research and Development. Washington, District of Columbia.

USEPA (United States Environmental Protection Agency). 2004. The incidence and severity of sediment contamination in surface waters of the United States. National sediment quality survey: Second edition (updated). EPA 823-R-02-013. Office of Research and Development. Washington, District of Columbia.

USEPA (United States Environmental Protection Agency). 2005. Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms: Metal mixtures (cadmium, copper, lead, nickel, silver, and zinc). EPA-600-R-02-11. Office of Research and Development. Washington, District of Columbia.

USEPA (United States Environmental Protection Agency). 2008. Problem formulation for the baseline ecological risk assessment of the Portland Harbor site. USEPA Region X. Portland, Oregon.

Figures

Figure 1. Scatter plot showing the relationship between amphipod (*Hyalella azteca*; HA) survival and biomass (n = 76).

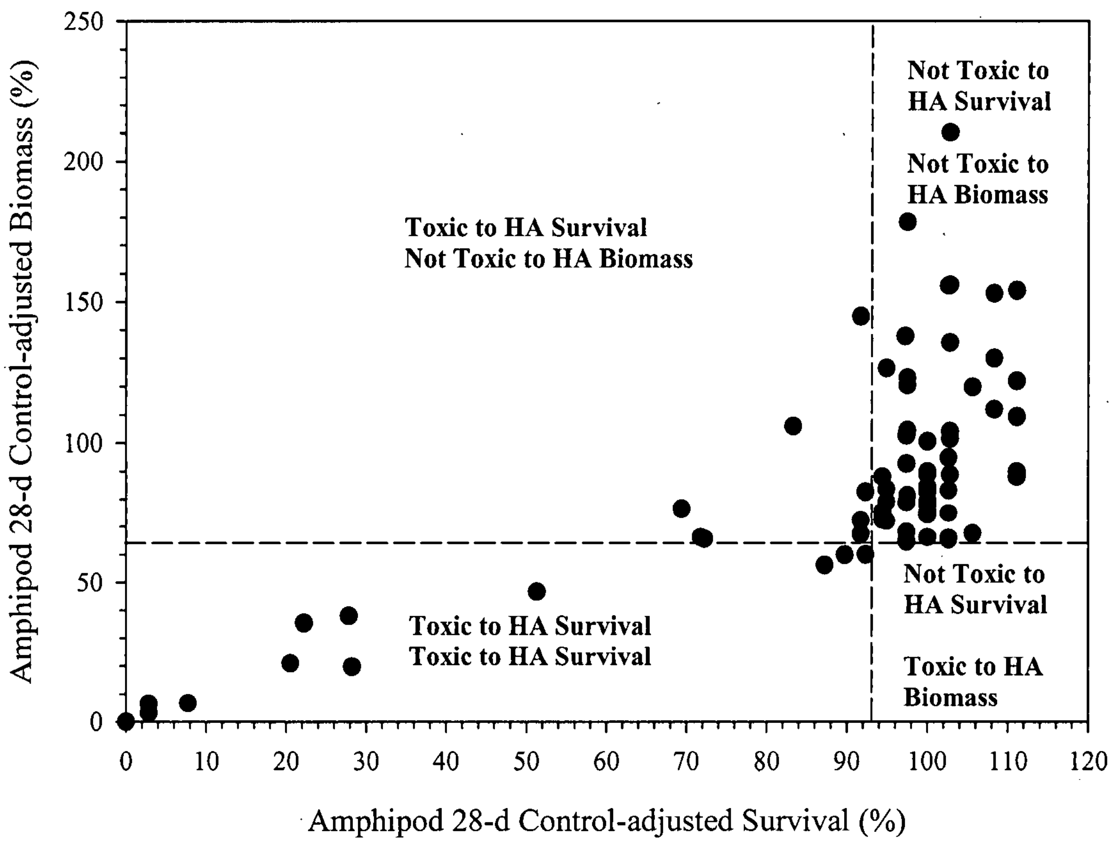


Figure 2. Scatter plot showing the relationship between amphipod (*Hyalella azteca*; HA) survival and midge (*Chironmus dilutus*; CD) survival (n = 76).

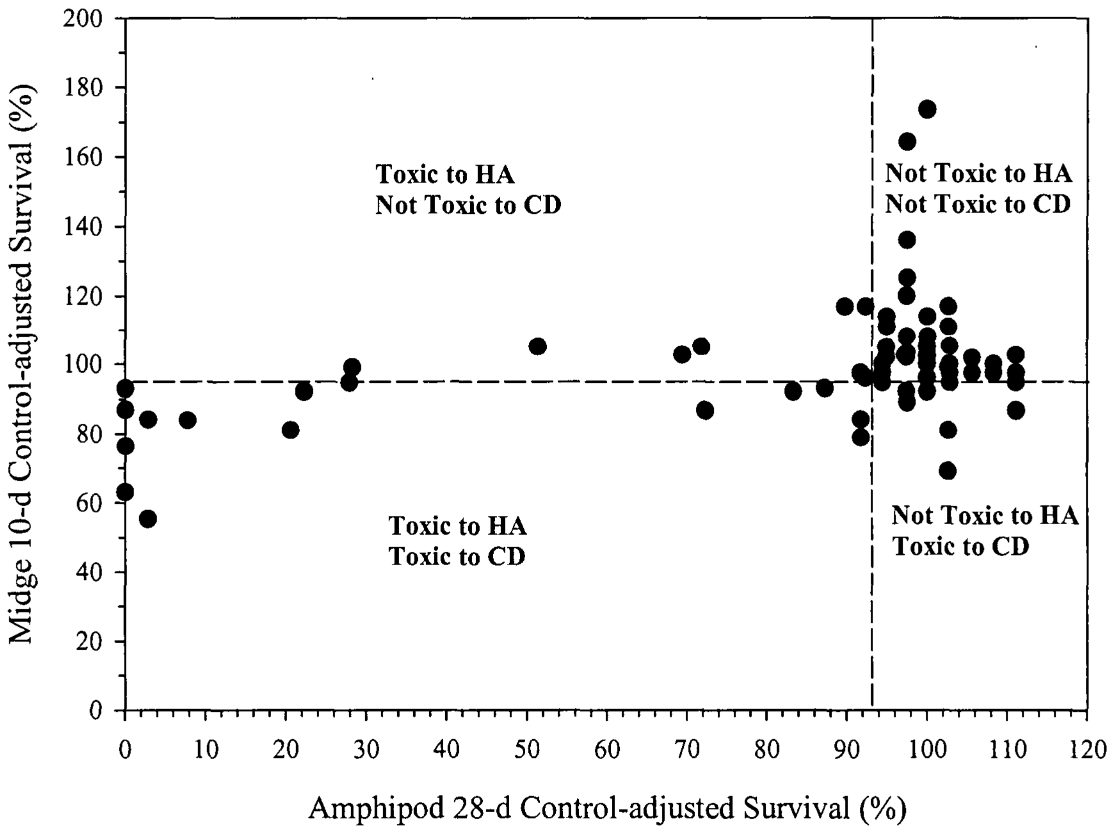
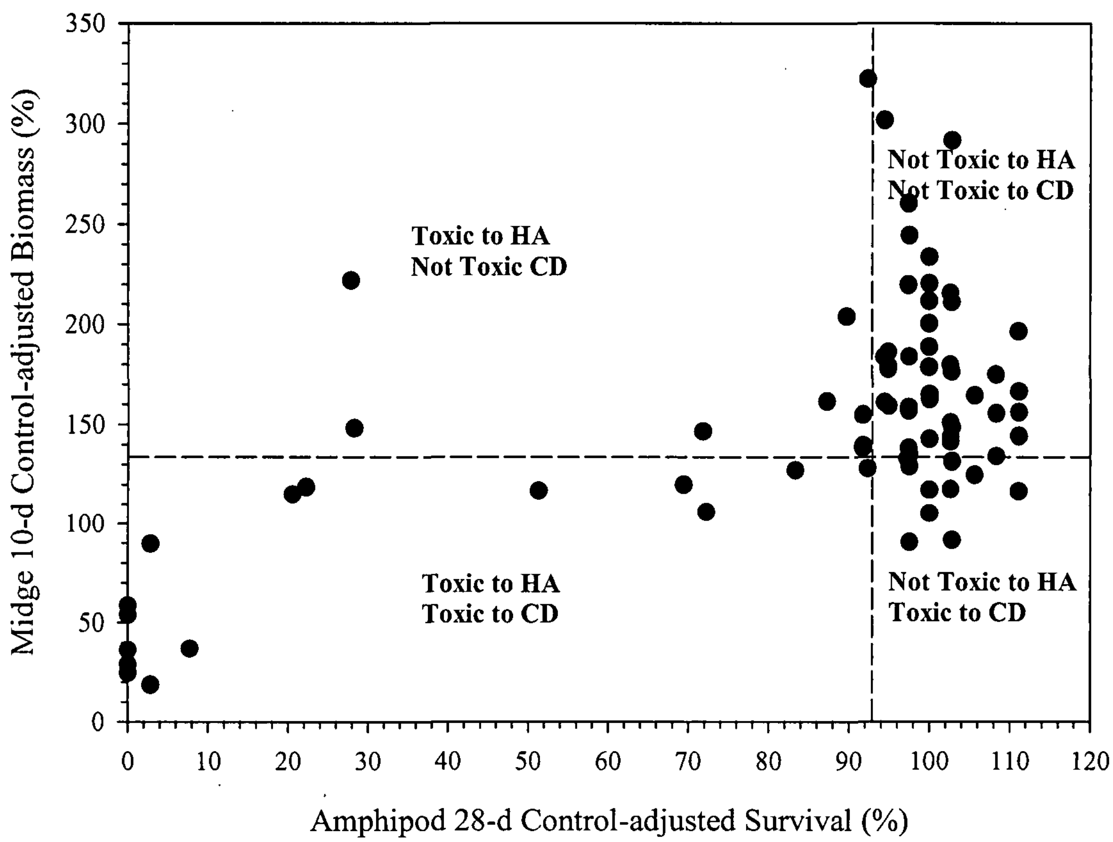


Figure 3. **Scatter plot showing the relationship between amphipod (*Hyalella azteca*; HA) survival and midge (*Chironomus dilutus*; CD) biomass (n = 76).**



Addendum 1

Addendum 1 Further Evaluation of the Approach for Assessing Risks to the Benthic Invertebrate Community at the Portland Harbor Superfund Site

A1.0 Introduction

In response to a request by the U.S. Environmental Protection Agency (USEPA) and the Lower Willamette Group (LWG), Don MacDonald and Peter Landrum conducted an independent evaluation of the approach for assessing risks to the benthic invertebrate community at the Portland Harbor Superfund site (MacDonald and Landrum 2008). Following submission, the document was reviewed by several members of the USEPA Technical Team. This review resulted in the identification several additional questions that needed to be answered to enhance the clarity of the original document. This addendum to the original report is intended to address the additional questions that were posed by the USEPA Technical Team, as well as several issues that were not sufficiently discussed in the original document.

A2.0 Responses to Additional Questions

Four additional questions were posed by the USEPA Technical Team in an effort to achieve greater clarity in the recommendations offered by MacDonald and Landrum (2008). These questions are presented below, along with our responses.

Question 1: In Section 4.6 (Recommended Procedures for Developing ToxicityThresholds), you discuss the "narrative intent" of toxicity thresholds as an important element of developing the specific quantitative threshold values to be used in Portland Harbor. Even though you mention some examples and provide a citation, it was not entirely clear to us what quantitative thresholds should be used to support the "low-risk" and "high-risk" toxicity thresholds, whether two risk thresholds is sufficient, and what specific steps, if any, would need to be taken to use the narrative intent to develop quantitative thresholds. Are different quantitative thresholds needed for each of the four empirical toxicity test results? Also, are these determinations made *a priori* or *a posteriori* to analysis of the Portland Harbor toxicity data, and to what extent are site data used? In general, additional detail regarding the scientific basis and specific steps needed would be helpful.

Response: Section 4.6 of MacDonald and Landrum (2008) describes our recommendations relative to the development of toxicity thresholds for the Portland Harbor site. However, these follow-up questions make it clear that our original text was not sufficiently detailed to enable the reader to fully understand the recommended procedures. For this reason, we would like to offer the following clarifications to make our recommendations more accessible. More specifically, we believe that toxicity thresholds for the Portland Harbor site should be developed using a step-wise process. The steps in this process include:

- Develop remedial action objectives (RAOs);
- Define the purpose of the toxicity thresholds;
- Establish the narrative intent of the toxicity thresholds;
- Establish criteria for evaluating the toxicity thresholds;
- Establish procedures for designating sediment samples as toxic or not toxic;
- Apply the procedures for toxicity designation and assign toxicity designations for each endpoint;
- Develop concentration-response models using the matching sediment chemistry and toxicity data;
- Derive toxicity thresholds;
- Evaluate the reliability and/or predictive ability of the toxicity thresholds.

Each of these steps in the process are briefly clarified in the following sections of this response.

Develop Remedial Action Objectives - RAOs are narrative statements that describe the intent of any remedial actions that are undertaken to protect human health and the environment at a contaminated site. For example, the RAOs for whole sediment at the Portland Harbor site might be to minimize or prevent exposure to whole sediments that are sufficiently contaminated to pose moderate or high risks to the microbial or benthic invertebrate communities. Such RAOs describe the desired future of the condition of sediments at the site relative to the risks that they pose to human health and/or ecological receptors. Therefore, the RAOs provide important guidance to risk assessors on the establishment of the narrative intent of the toxicity threshold that will be used in the Baseline Ecological Risk Assessment (BERA) and/or Feasibility Study (FS).

Define the Purpose of the Toxicity Thresholds - For the Portland Harbor site, numerical toxicity thresholds are required to satisfy two important needs. First, toxicity thresholds are needed to support the BERA. In this application, the toxicity thresholds are needed to classify chemistry-only sediment samples into categories based on the risks that they pose to benthic invertebrates. Second, toxicity thresholds are needed to support the FS. In this application, the toxicity thresholds are needed to establish preliminary remediation goals (PRGs; i.e., risk-based tools for evaluating remedial options at the site) that can be used to evaluate the costs and benefits associated with various remedial options. At other sites, we have endeavored to establish toxicity thresholds that could be consistently applied within the BERA and the FS. In this way, there is a direct linkage between the toxicity thresholds that are used to evaluate risks to benthic invertebrates and the toxicity thresholds that are used to establish clean-up goals (e.g., PRGs; i.e., RAOs inform the narrative intent of the toxicity thresholds, which informs selection of toxicity thresholds based on reliability and predictive ability analyses, which inform the selection of PRGs).

Establish the Narrative Intent of the Toxicity Thresholds - Virtually all approaches to the development of sediment quality guidelines (SQGs) are linked to a narrative that describes the purpose or intent of the resultant SQGs. This narrative intent has been described in various publications and summarized for selected national SQGs in Wenning *et al.* (2002).

Importantly, the narrative intent of the SQGs provides risk assessors with essential guidance on the appropriate uses of the SQGs and relevant information for establishing criteria for evaluating how well the SQGs work at specific sites. For example, a threshold effect level (TEL) is intended to identify the concentration of a chemical of potential concern (COPC) below which adverse effects on benthic invertebrates are likely to be observed only infrequently. Therefore, a TEL should be used to identify conditions where the concentrations of a specific COPC are unlikely to cause or substantially contribute to sediment toxicity. In addition, TELs should be considered to be reliable if there is a low incidence of toxicity (IOT; i.e., <10%) for sediment samples that have COPC concentrations below the TELs for all measured substances. A TEL should not, necessarily, be evaluated to determine how well it predicts toxicity because TELs were not designed for this purpose.

Numerical toxicity thresholds (i.e., site-specific sediment quality values; SQVs) have been identified as important tools for assessing risks to benthic invertebrates at the Portland Harbor site. As such, it would be beneficial to clearly articulate the narrative intent of the toxicity thresholds that will be used in the BERA process and/or to establish target clean-up goals (i.e., PRGs). The narrative intent of the SQVs should be consistent with the RAOs that are established for the site. More specifically, numerical SQVs are required to identify sediment samples at the Portland Harbor site that pose low risks to benthic invertebrates (i.e., below which there would be a low IOT; e.g., <20% of the samples would be predicted to be toxic). Remedial measures are unlikely to be required to address risks to the benthic invertebrate community at locations with COPC concentrations below the low-risk SQVs. In addition, numerical SQVs are required to identify sediment samples that pose high risks to benthic invertebrates (i.e., above which there would be a high IOT; e.g., > 50% of the samples would be predicted to be toxic). Remedial measures may be required to address risks to the benthic invertebrate community at locations with COPC concentrations above the high-risk SQV. Such low-risk and high-risk SQVs would also result in the identification of COPC concentrations that would be predicted to be associated with a moderate IOT; e.g., 20 to 50% of the samples would be predicted to be toxic). Additional data interpretation and/or toxicity testing may be required at locations with

COPC concentrations that fall between the low-risk and high-risk SQVs. This approach would be consistent with the one used in the Calcasieu Estuary to support the derivation of toxicity thresholds for use in the BERA and the FS (MacDonald *et al.* 2002; 2003).

The Calcasieu Estuary example illustrates one option for establishing the narrative intent of the SQGs. It may be that there is a need to establish additional categories for assessing risks to benthic invertebrates in Portland Harbor. For example, the State of California established a set of criteria for placing sediment samples into each of four categories, based on potential for toxicity to benthic invertebrates (i.e., non-toxic, low toxicity, moderate toxicity, and high toxicity). The toxicity thresholds that were established for various COPCs and COPC mixtures reflected the narrative intent of the categories (see http://www.sccwrp.org/sqo/pubs/503_toxicity_indicator_methods.pdf). These thresholds were explicitly developed to facilitate classification of sediment samples into these categories using data on sediment toxicity and/or sediment chemistry (See http://www.sccwrp.org/sqo/pubs/543_ChemToxSQGComparison_Draft_10_24_07.pdf). While the two examples described here illustrate two options for describing the narrative intent of SQGs, the numbers of categories for which the narrative is established depends on the needs of the manager.

It is recommended that SQVs be established for all four of the endpoints (i.e., amphipod survival, midge survival, amphipod biomass, and midge biomass) examined at the Portland Harbor site because the organisms may be differentially sensitive by endpoint and/or by species to different mixtures of chemicals in the sediment. However, the narrative intent of the SQVs developed using the models for each endpoint should be similar, at least at the outset. Following model and SQV development, the reliability and predictive ability evaluations will provide the information needed to determine the relative sensitivity of each endpoint and the level of protection that SQGs derived for various endpoints will afford toxicity test organisms overall.

Establish Criteria for Evaluating the Toxicity Thresholds - Once the narrative intent of the SQVs has been established, it is possible to establish criteria for evaluating the site-specific toxicity thresholds. For the above example, the low-risk SQVs would be considered to be reliable if there is a low IOT (i.e., <20%) for sediment samples that have COPC concentrations below the low-risk SQVs for all measured substances. In contrast, the high-risk SQVs would be considered to be reliable if there is a high IOT (i.e., >50%) for sediment samples that have COPC concentrations above the high-risk SQVs for all measured substances. In addition, a low-risk/high-risk pair of SQVs for a COPC or COPC mixture would be considered to be reliable if there is an moderate IOT when COPC concentrations fall between the two SQVs (i.e., 20 to 50% IOT). This example illustrates the need to establish a direct linkage between the narrative intent of the SQVs and the criteria that are used to evaluate the SQVs.

Establish Procedures for Designating Sediment Samples as Toxic or Not Toxic - Both of the modeling approaches that have been selected for use at the Portland Harbor site rely on hit/no hit designations of the sediment samples used in the development of the predictive models. Section 4.3 of MacDonald and Landrum (2008) describes our recommended procedures for determining if individual sediment samples are toxic or not toxic to benthic invertebrates (i.e., reference envelope approach). This approach can be applied to designate sediment samples as toxic or not toxic for each of the toxicity test endpoints selected for assessing whole-sediment toxicity at the Portland Harbor site. Recommended approaches for selecting reference stations are described in Section A4.1 of this document. In addition, the recommended criteria for identifying reference sediment samples are presented in Section A4.2 of this document. The criteria for evaluating candidate reference samples presented in Section A4.2 supercedes the criteria listed in Section 4.3 of MacDonald and Landrum (2008).

Assign Toxicity Designations to Sediment Samples - As indicated above, MacDonald and Landrum (2008) recommended procedures for designating sediment samples from Portland Harbor as toxic or not toxic. Implementation of these and/or alternate procedures will facilitate designation of each sediment sample from the study area as toxic or not toxic on an endpoint-by-endpoint basis. That is, each sediment sample will have at least four toxicity designations (i.e., based on amphipod survival, amphipod biomass, midge survival, and midge biomass). These toxicity designations should directly support the development of predictive models for each of the four toxicity test endpoints and each of the COPCs/COPC mixtures that are relevant to the site.

Develop Predictive Models (i.e., Concentration-Response Models) - As indicated by MacDonald and Landrum (2008), there are a variety of approaches that could be used to evaluate relationships between whole-sediment chemistry and whole-sediment toxicity at the Portland Harbor site (See Section 4.5). The logistic regression model (LRM) and floating percentile model (FPM) are likely to provide useful tools for evaluating relationships between the concentrations of COPCs/COPC mixtures in Portland Harbor sediments and the responses of benthic invertebrates (i.e.,amphipod survival, amphipod biomass, midge survival, and midge biomass). In addition, the site-specific sediment chemistry and sediment toxicity data could be used to develop concentration-response models based on magnitude of toxicity (MOT; e.g., control-adjusted survival of amphipods). Furthermore, area of interest-specific models could be developed to better explain the relationships between sediment chemistry and sediment toxicity if the site-wide models are not sufficiently reliable to accurately predict the presence or absence of sediment toxicity.

Based on our review of the existing models and their performance, it appears that grain size (i.e., percent fines) is the metric that is best correlated with the responses of benthic invertebrates exposed to Portland Harbor sediments. While these results could reflect the physical effects of grain size, the toxicity test organisms that were selected to evaluate Portland Harbor sediments are not highly sensitive to grain size (USEPA 2000; ASTM 2007). Therefore, it is more likely that percent fines represents a general surrogate for contamination in Portland Harbor sediments. That percent fines is better correlated with sediment toxicity than any of the measured COPCs or COPC mixtures likely indicates that a variety of measured and/or unmeasured substances are causing or substantially contributing to the observed toxicity in these sediments. This information strengthens the position that multiple chemical concentration gradients occur within Portland Harbor sediments. If this is the case, then it is unlikely that site-wide predictive models for individual COPCs or simple COPC mixtures (e.g., tPAHs, tDDTs, tPCBs) will provide highly reliable bases for classifying sediment samples as toxic or not toxic to benthic invertebrates. If this is the case, area of interest-specific predictive models may be required to improve the reliability and predictive ability of the models. Alternatively, other data collection and/or interpretation approaches may be required to support remedial decisions at the site.

Derive Toxicity Thresholds - As indicated above, two modeling approaches have been selected to support evaluation of risks to benthic invertebrates at the Portland Harbor site. Both the logistic regression model (LRM) and floating percentile model (FPM) approaches can be used to derive numerical toxicity thresholds (i.e., SQVs) for individual COPCs and/or COPC mixtures. Both approaches provide information on the probability of observing toxicity to benthic invertebrates based on the measured concentrations of COPCs/COPC mixtures in sediments (i.e., these models are IOT based rather than MOT based).

At other sites that we have worked on (e.g., Calcasieu Estuary, Tri-State Mining District), two types of toxicity thresholds were established to support the BERA and FS processes, including low-risk toxicity thresholds and high-risk toxicity thresholds [as described in Section 4.6 of MacDonald and Landrum (2008)]. Both of these toxicity threshold types were developed to correspond to pre-selected magnitudes of toxicity (MOT; i.e., 10% and 20% increase in the MOT relative to reference conditions, respectively). The MOTs were selected jointly by the risk assessors, the risk managers, and the Natural Resources Trustees, and were considered to be consistent with the RAOs for the sites. The low-risk and high-risk toxicity thresholds were derived from the concentration-response relationships developed for each COPC/COPC mixture-toxicity test endpoint pair of interest at the site (Figure A1; see MacDonald *et al.* 2003; 2005a; 2005b for more information).

It is our understanding that the two modeling approaches selected to support evaluation of risks to benthic invertebrates provide information on the probability of observing toxicity to benthic invertebrates (i.e., IOT rather than MOT). Our experience suggests that toxicity thresholds based on IOT and on MOT can be generally consistent, with a 10% increase in the MOT roughly corresponding to a 20% increase in the IOT. Toxicity thresholds based on a 20% increase in the MOT generally correspond to those based on a 50% increase in the IOT. Therefore, it would not be unreasonable to establish the narrative intent of SQGs for the Portland Harbor site as follows:

- Low-risk toxicity thresholds represent the concentrations of COPC or COPC mixtures below which there is less than 20% IOT to benthic invertebrates;
- High-risk toxicity thresholds represent the concentrations of COPC or COPC mixtures above which there is greater than 50% IOT to benthic invertebrates; and,
- A moderate IOT (i.e., 20 to 50%) should be observed at concentrations of COPCs or COPC mixtures between the low-risk and high-risk toxicity thresholds. A moderate risk would be assigned to sediment samples with concentrations of COPCs or COPC-mixtures that fall within this category.

Such narrative objectives for the toxicity thresholds would provide clear guidance to the modelers relative to the development of toxicity thresholds from the models. In addition, establishment of such narrative objectives for the toxicity thresholds would provide important information for establishing evaluation criteria for determining the reliability and predictive ability of the toxicity thresholds that are developed from the models.

Evaluate the Reliability and/or Predictive Ability of the Toxicity Thresholds - The reliability of the various toxicity thresholds should be evaluated to determine if they can be used to accurately classify sediment samples from the site as toxic or not toxic (i.e., using the matching sediment chemistry and toxicity data that were used to derive the toxicity thresholds). In contrast, the evaluation of predictive ability is conducted using an independent data set (i.e., using matching sediment chemistry and toxicity data that were not used to derive the toxicity thresholds).

At a metals-contaminated site, toxicity thresholds were developed using the results of 28-d toxicity tests with the amphipod, *Hyalella azteca*, and the mussel, *Lampsilis siliquoidea* (i.e., T_{10} and T_{20} values, based on MOT). The results of the evaluation of the reliability of these toxicity thresholds are presented in Table A1. These results show the IOT below each toxicity threshold, the IOT above each toxicity threshold, and the overall correct classification rate for each toxicity threshold. Similarly, the results of the predictive ability evaluation are presented in Table A2.

The Calcasieu Estuary study also provides a useful example for illustrating the importance of conducting the reliability and predictive ability evaluations. In this case, mean probable effect concentration-quotients (PEC-Qs) of 0.24 and 0.45 for amphipod survival (*Hyalella azteca*) were selected as the low-risk and high-risk toxicity thresholds, respectively. The results of the reliability evaluation showed that the incidence to toxicity was generally low below the selected low-risk toxicity threshold (i.e., 18.7% of the samples were toxic to *Hyalella azteca* in 28-d exposures; Table A3). Above the selected high-risk threshold, 69% of the samples were toxic. Because there was a high IOT between the two toxicity thresholds (i.e., 67%), it was concluded that a single toxicity threshold could be used to classify sediment samples into two categories, toxic or not toxic to amphipods in 28-d toxicity tests (i.e., the low-risk toxicity threshold of 0.24 for mean PEC-Q was selected as the toxicity threshold).

This example also provides important information on the predictive ability of the toxicity thresholds (i.e., in terms of predicting toxicity to other toxicity test organisms and endpoints and predicting responses of the benthic invertebrate community). These results show that the selected toxicity thresholds provided an accurate basis for classifying sediment samples from the site as toxic and not toxic based on the survival of another amphipod species (*Ampelisca abdita*) and on the fertilization of sea urchins (*Arbacia punctulata*; Table A4). In addition, many of the benthic invertebrate community structure endpoints showed graded responses for the groups of sediment samples identified by the toxicity thresholds (Table A4). Therefore, the results of the predictive ability evaluation confirmed that the toxicity thresholds could be used to accurately classify sediment samples into low, moderate, and high-risk categories. Interestingly, these results showed that the growth (length) of *Hyalella azteca* did not provide additional information relative to the risks that sediment-associated COPCs/COPC mixtures posed to benthic invertebrates.

Selection and Application of the Toxicity Thresholds - As indicated in the previous section, the results of the reliability and predictive ability evaluations provide essential information for selecting toxicity thresholds for use in the BERA and/or FS. For both the metals-contaminated site and the Calcasieu Estuary, these results can be used directly to identify the toxicity thresholds that meet the narrative intent established earlier in the process. This direct linkage between narrative intent and the performance of the toxicity thresholds makes the selection process relatively straight forward.

For the Portland Harbor BERA, the results of the reliability and predictive ability evaluations will provide the information needed to decide which toxicity thresholds should be used in the BERA and FS processes and to decide how such toxicity thresholds should be used to assess risks to the benthic invertebrate community and/or establish clean-up goals (i.e., PRGs) for the site. As indicated in the Calcasieu Estuary example, it is possible that a single toxicity threshold can be used to conduct risk assessments in the BERA and to establish PRGs to support the FS. The results of these evaluations for the Portland Harbor site could also suggest that it is reasonable to utilize toxicity thresholds for multiple endpoints to provide multiple lines-of-evidence for evaluating risks to benthic invertebrates (i.e., in the sample-by-sample evaluation of sediment quality conditions). For example, the State of California combined multiple lines of evidence to evaluate sediment quality conditions at each sampling station (for more information, see http://www.sccwrp.org/sqo/pubs/545_MLOE_FrameworkValidationDraft_10_15_07.pdf). The same type of approach could be used for the various endpoints, organisms and thresholds to provide a framework for deciding the magnitude of concern about a station. This does not mean that a station with only one threshold exceeded is ignored but rather it would be assigned a lower magnitude of concern than one with multiple thresholds exceeded. In contrast, it may be reasonable to select toxicity thresholds for only one endpoint during the development of PRGs (e.g., the most sensitive toxicity test endpoint, which would be expected to be protective of all other toxicity test endpoints).

Summary - In summary, we recommend that the RAOs and narrative intent of the SQVs be established prior to developing predictive models for the site. This is important for ensuring that the models can be properly optimized to respond to the narrative intent articulated. Establishment of the narrative intent of the SQVs *a priori* will support the development of evaluation criteria that are consistent with management needs at the site (as articulated in the RAOs). In addition, we recommend using data from the Portland Harbor site and/or from other locations in the development of the two models. We further recommend that a portion of the data from the site be set aside for use in evaluating the predictive ability of the models. By doing so, both the reliability and predictive ability of the SQVs can be evaluated. The results of these evaluations should be used to identify the toxicity threshold or toxicity thresholds that ought to be used to classify sediment samples from the site in terms of the risks that they pose to the benthic invertebrate community. These results should also be used to identify the need for area of interest-specific toxicity thresholds and/or other data interpretation approaches to evaluate risks to the benthic invertebrate community associated with exposure to contaminated sediments and to support remedial decisions at the site.

Question 2: In your answer to question #4 (should non-site data be considered in the development of the LRM?), you support use of non-site data. However, would you also support use of non-site data in the development of the floating point model? Most of the discussion regarding use of non-site data between EPA and LWG have focused on the LRM, but in the interests of full clarity, we wanted to know whether you suggested non-site data are also of value to the floating point model.

Response: It would also be acceptable to use non-site data for developing the floating point model. The objective of the modeling process is to develop one or more tools that can be used to accurately classify sediment samples as toxic or not toxic, based on whole-sediment chemistry data alone. Such tools can include generic SQGs or site-specific sediment toxicity thresholds for individual COPCs and/or COPC mixtures. From our perspective, the approach that is used to generate the models and the source of the underlying data that are applied in the modeling process is not particularly relevant. What matters is whether or not the resultant model can be used to accurately classify sediment samples from Portland Harbor as toxic or not toxic (i.e., based on the results of the reliability and predictive ability evaluations). We have described the procedures for evaluating the models in Section 4.7 of the document.

There is one issue that we have some concern about with respect to the use of site data in the development of the models of toxic response versus chemical contamination. The sediment samples that have been collected at the Portland Harbor site include material present within the 0 to 30 cm sediment depth. Hence, the samples include material located beyond (i.e., deeper than) the biologically-active zone [i.e., 9.8 ± 4.5 cm for marine organisms (Boudreau 1998), 0-2 cm to 0-15 cm for nearshore infauna, and 0-2 cm to 0-12 cm for freshwater invertebrates (http://www.sediments.org/sedstab/germano.pdf). The biologically-active depth is tied to the rate of deposition of the sediments (White and Miller 2008).

Inclusion of deeper material in the site sediment samples increases the likelihood that factors such as ammonia and/or hydrogen sulfide have contributed to the observed responses of toxicity test organisms. Thus, the selection of 0-30 cm sediment horizon in the sampling programs could lead to some misleading information on the current surficial conditions and, because of the complications noted above, could result in variability in the development of the relationship between sediment chemistry and toxicity. This issue is also relevant to the selection of sediment samples for inclusion in the reference envelope calculations (see Section A4.0 below).

Question 3: Are there any problems with the *Hyalella azteca* biomass endpoint tests that would preclude their use as an empirical line of evidence in the baseline ecological risk assessment for Portland Harbor?

Response: No. The biomass endpoint is a useful endpoint for evaluating effects on benthic invertebrates associated with exposure to contaminated sediments. While we have only recently started to use the biomass endpoint, our experience at other sites indicates that this endpoint can be among the most useful endpoints relative to quantifying the relationships between COPC concentrations in sediments and the responses of toxicity test organisms. By integrating the survival and weight endpoints, the biomass endpoint can provide useful information for evaluating the effects on amphipods associated with exposure to contaminated sediments at the Portland Harbor site. This endpoint is particularly useful for evaluating sediment samples that have marginal hits for one or both of the underlying endpoints (survival and weight).

Question 4: Are there any reasons the *Hyalella azteca* biomass endpoint empirical results should not be used in the floating percentile models under development for Portland Harbor?

Response: No. We have used the biomass endpoint to develop concentration-response relationships for a variety of COPCs and COPC mixtures. As indicated in Section 4.7 of MacDonald and Landrum (2008), the key is to evaluate the reliability and predictive ability of the resultant models and the associated toxicity thresholds. The results of such evaluations will provide the information needed to determine if the models developed using this endpoint are appropriate for use in the BERA and/or the establishment of PRGs for the site.

A3.0 Application of Regional Sediment Evaluation Team (RSET) Process to the Portland Harbor Site

The RSET process was initiated in 2002 to update the Lower Columbia dredged material evaluation framework (DMEF). More specifically, RSET was established to revise and develop sediment evaluation procedures for the region. This process was intended to result in the development of a northwest regional sediment evaluation framework that could be used by federal and state agencies in Region 10. As part of this effort, RSET is in the process of evaluating the protectiveness of the current suite of bioassays, reviewing and refining biological interpretive criteria, and reviewing and refining sediment screening levels.

Based on our cursory review, the RSET process has the potential to provide useful advice and guidance relative to the evaluation of dredged materials and other sediments. Therefore, it is reasonable to review the results of the RSET process and assess their applicability to the Portland Harbor site. However, it is important to remember that the narrative intent of the sediment screening levels that emerge from the RSET process may not be consistent with the remedial action objectives (RAOs) that are established for the Portland Harbor site. Similarly, guidance provided by RSET relative to the interpretation of toxicity test results may not be consistent with the RAOs. Therefore, the tools that are ultimately used to evaluate risks to the benthic invertebrate community should be selected to meet site assessment and management needs at the Portland Harbor site. In our view, there is no need for site assessment activities to be entirely consistent with RSET guidance or RSET decisions regarding data utilization or interpretation.

A4.0 Development of a Reference Envelope for Portland Harbor

Section 4.3 of MacDonald and Landrum (2008) describes the recommended procedures for developing a reference envelope for interpreting whole-sediment toxicity data from the Portland Harbor site. This section of the original document did not provide sufficient detail to enable risk assessors to establish a reference envelope for the site. The following information is provided to assist readers in better understanding our recommendations for developing a reference envelope for Portland Harbor.

A4.1 Approaches to Selecting Reference Locations

In general, candidate reference locations should be established on an *a priori* basis, based on an understanding of the water body under investigation and the existing data on sediment quality conditions. According to ASTM (2007), a reference sediment sample is defined as whole sediment obtained from an area of concern used to assess sediment conditions exclusive of the materials of interest. Therefore, candidate reference locations should be selected based on their proximity to the study area, using, at minimum, information on whole-sediment chemistry.

At the Portland harbor site, several options are available for identifying candidate reference locations. First, the sediment samples that were collected at the six locations in upstream areas can be considered for use as reference sediment samples. In addition, it may be possible to identify reference sediment samples from the samples that have been collected to date from the Portland Harbor site. Finally, additional candidate reference locations could be identified in upstream areas, within the site boundaries, in downstream areas, in tributaries, or in the Columbia River. In all cases, the whole-sediment chemistry and whole-sediment toxicity data collected at candidate reference locations would need to be reviewed to determine if the sample qualifies as a reference sample [see Section 4.3 of MacDonald and Landrum (2008) and below for criteria for evaluating candidate reference sediment samples]. Only those samples that meet the evaluation criteria should be included in the data set used to develop the reference envelope.

A tiered process is recommended for identifying candidate reference locations for the Portland Harbor BERA. As a first step, the desired number of reference sediment samples for developing the reference envelope should be selected. Based on our experience, about 15 sediment samples are required to adequately characterize variability in the responses of toxicity test organisms associated with exposure to reference sediments. Then, the six sediment samples that were collected upstream of the site should be evaluated to determine if they qualify as reference sediment samples. Subsequently, sediment samples from within the study area that meet the evaluation criteria [presented in Section 4.3 of MacDonald and Landrum (2008) and refined below] should be identified and their locations plotted. Clusters of samples with low chemistry should be selected preferentially as reference samples (i.e., rather than isolated samples) because such clustering increases confidence that the sediments in that geographic area do not contain elevated levels of COPCs. If insufficient numbers of reference samples are not identified using the first two approaches, then it may be necessary to collect additional sediment samples to obtain sufficient data to develop the reference envelope. Because additional sampling would require additional time and resources, this option would be pursued only if the requisite data are not already available from within the existing data set.

A4.2 Criteria for Identifying Reference Sediment Samples

The recommended criteria for identifying reference sediment samples are presented in Section 4.3 of MacDonald and Landrum (2008). These criteria specified the chemical and biological characteristics of sediment samples that would qualify for inclusion in a reference envelope. We have further reviewed these criteria and would like to offer the following refinements (Note: Refinements are shown in bold italics):

Whole-Sediment Chemistry

- All measured metals, PAHs, ***DDTs,*** and PCBs occur at concentrations below conservative SQGs;
- Mean PEC-Q_{DW} < 0.1;
- $\sum$ESB-TU_{PAHs} < 0.1; and
- ($\sum$SEM-AVS)/f_{oc} < 130.

Whole-Sediment Toxicity

- Control-adjusted response rate should not exceed the minimum significant difference (MSD) for each toxicity test endpoint; or,
- In the absence of MSD values, control-adjusted response rate should not exceed the Tier II levels applied in the NSI (USEPA 2004);

Pore-Water Chemistry

- ***Total ammonia ($NH_4^+ + NH_3$), unionized ammonia (NH_3), and hydrogen sulfide (H_2S) concentrations in pore water should not exceed lowest observed effect levels (LOELs) based on the results of water-only toxicity tests conducted with each of the toxicity test organisms.***

Consideration of these additional criteria is important for several reasons. First, DDTs have been identified as COPCs in portions of the Portland Harbor site. Therefore, concentrations of DDTs (i.e., sum DDD, sum DDE, sum DDT, and total DDTs) should be considered in the selection of reference sediment samples (i.e., DDT levels should not exceed conservative sediment quality guidelines). In addition, sediment sampling at the Portland Harbor site targeted the 0 to 30 cm sediment horizon. This horizon likely encompasses both the biologically-active zone (i.e., typically defined as the top 10 cm of material) and the zone of limited biological activity (i.e., deeper sediments; 10 - 30 cm). Because anoxic sediments were likely included in many of the sediment samples collected at the site, it is possible that toxicity test organisms could have responded to ammonia and/or hydrogen sulfide in a portion of the samples (i.e., these substances could have contributed to the observed toxicity). The reference sediment samples that are selected should reflect conditions in the biologically-active zone at the site, rather than conditions that benthic invertebrates at the site would not normally be exposed to. Therefore, samples selected to represent reference conditions should not have elevated levels of ammonia or hydrogen sulfide in pore water.

A5.0 Development of Clean-up Goals for Portland Harbor

It is our observation that the LRM and FPM models that have been developed to date for the Portland Harbor site are explicitly intended to support evaluation of risks to benthic invertebrates associated with exposure to contaminated sediments. That is, the toxicity thresholds developed using the models are intended to classify sediment samples into categories based on the probability that the sample will be toxic to benthic invertebrates. This is an appropriate use of the models. However, there is also a need to establish clean-up goals for the site to support efforts under the FS (e.g., PRGs). It is not clear that the existing models will provide a reliable basis for establishing site-wide clean-up goals for Portland Harbor. The models are likely to be limited in this respect for several reasons, including:

- The sampling strategy selected for the site may have resulted in interferences that complicate interpretation of the sediment toxicity data (e.g., elevated ammonia and/or hydrogen sulfide levels may occur in a portion of the samples); and,

- The site appears to have multiple concentration gradients for multiple COPCs. As a result, clear relationships between COPC concentrations and sediment toxicity may not be evident on a site-wide basis.

For this reason, an alternate approach may be required to establish clean-up goals for the site. For example, ammonia and hydrogen sulfide could be incorporated into the chemical mixture models that are developed for the site. In addition or alternatively, the site could be divided into multiple areas of interest, each of which has an apparent gradient for key COPCs and/or COPC mixtures (e.g., PCBs, DDT, PAHs, etc.). Then, area of interest-specific models could be developed for the key COPCs/COPC mixtures and the reliability of the toxicity thresholds developed using those models could be evaluated. Another option involves selection of clean-up goals for key COPCs and COPC mixtures based on the clean-up goals that have been established for sites where these contaminants are the principal COPCs (e.g., 1 ppm for total PCBs). Virtual remediation techniques could be used to evaluate residual risks to ecological receptors if such clean-up goals were adopted at the Portland Harbor site (i.e., by calculating post-remediation surface-weighted average concentrations of key COPCs/COPC mixtures by area of interest). The point is that different approaches could and possibly should be used to develop toxicity thresholds for use in the BERA and PRGs for use in the FS.

A6.0 References

ASTM (American Society for Testing and Materials). 2007. Test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. E1706-05. *In:* ASTM Annual Book of Standards, Vol. 11.06. West Conshohocken, Pennsylvania.

Boudreau, B.P. 1998. Mean mixed depth of sediments: the wherefore and the why. Limnology and Oceanography 43:524-526.

MacDonald, D.D. and P.F. Landrum. 2008. An Evaluation of the Approach for Assessing Risks to the Benthic Invertebrate Community at the Portland Harbor Superfund Site. Preliminary Draft. Prepared for United States Environmental Protection Agency, Region 10, Seattle, Washington 9 and Parametrix, Inc. Albany, Oregon. Prepared by MacDonald Environmental Sciences Ltd., Nanaimo, British Columbia and Landrum and Associates, Ann Arbor, Michigan.

MacDonald, D.D., R.L. Breton, K. Edelmann, M.S. Goldberg, C.G. Ingersoll, R.A. Lindskoog, D.B. MacDonald, D.R.J. Moore, A.V. Pawlitz, D.E. Smorong, and R.P. Thompson. 2003. Development and evaluation of preliminary remediation goals for selected contaminants of concern at the Calcasieu Estuary cooperative site, Lake Charles, Louisiana. Prepared for United States Environmental Protection Agency, Region 6. Dallas, Texas.

MacDonald, D.D., C.G. Ingersoll, D.E. Smorong, L. Fisher, C. Huntington, and G. Braun. 2005a. Development and evaluation of risk-based preliminary remediation goals for selected sediment-associated contaminants of concern in the West Branch of the Grand Calumet River. Prepared for: United States Fish and Wildlife Service. Bloomington, Indiana.

MacDonald, D.D., C.G. Ingersoll, A.D. Porter, S.B Black, C. Miller, Y.K. Muirhead. 2005b. Development and evaluation of preliminary remediation goals for aquatic receptors in the Indiana Harbor Area of Concern. Technical Report. Prepared for: United States Fish and Wildlife Service. Bloomington, Indiana and Indiana Department of Environmental Management. Indianapolis, Indiana.

Wenning, R.J., G.E. Batley, C.G. Ingersoll, and D.W. Moore. 2002. Use of sediment quality guidelines and related tools for the assessment of contaminated sediments. SETAC Press. Pensacola, Florida.

White, D.S. and M.F Miller. 2008. Benthic invertebrate activity in lakes: linking present and historical bioturbation patterns. Aquatic Biology 2:269-277.

USEPA (United States Environmental Protection Agency). 2004. The incidence and severity of sediment contamination in surface waters of the United States. National sediment quality survey: Second edition (updated). EPA 823-R-02-013. Office of Research and Development. Washington, District of Columbia.

Table A1. Reliability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca,* and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass).[1]

COPC/COPC Mixture	Toxicity Test Endpoint Used to Derive STT	n	T_{10}	T_{20}	Incidence of Toxicity						
					$<T_{10}$	$\geq T_{10}$	Correct Classification Rate for T_{10}	T_{10}-T_{20}	$\leq T_{20}$	$>T_{20}$	Correct Classification Rate for T_{20}
Basis for T_{10}/T_{20}: 28-d *H. azteca* Survival											
Cadmium	Amphipod 28-d S	76	11.1	17.3	**11% (5 of 45)**	**61% (19 of 31)**	78%	60% (6 of 10)	20% (11 of 55)	**62% (13 of 21)**	75%
Lead	Amphipod 28-d S	76	150	219	**9% (4 of 45)**	**65% (20 of 31)**	**80%**	33% (3 of 9)	**13% (7 of 54)**	**77% (17 of 22)**	**84%**
Zinc	Amphipod 28-d S	76	2083	2949	**13% (7 of 53)**	**74% (17 of 23)**	**83%**	ND	**13% (7 of 53)**	**74% (17 of 23)**	**83%**
ΣSEM-AVS	Amphipod 28-d S	76	7.82	13.7	**11% (5 of 44)**	**59% (19 of 32)**	76%	29% (2 of 7)	**14% (7 of 51)**	**68% (17 of 25)**	**80%**
Mean PEC-Q	Amphipod 28-d S	76	0.556	0.732	**11% (5 of 47)**	**66% (19 of 29)**	**80%**	83% (5 of 6)	**19% (10 of 53)**	**61% (14 of 23)**	75%
Mean PEC-Q_{METAL}	Amphipod 28-d S	76	1.11	1.78	**9% (4 of 45)**	**65% (20 of 31)**	**80%**	67% (6 of 9)	**19% (10 of 54)**	**64% (14 of 22)**	76%
Basis for T_{10}/T_{20}: 28-d *L. siliquoidea* Survival											
Copper	Mussel 28-d S	48	116	141	37% (17 of 46)	**100% (2 of 2)**	65%	ND	37% (17 of 46)	**100% (2 of 2)**	65%
Zinc	Mussel 28-d S	48	20600	23700	38% (17 of 45)	**67% (2 of 3)**	63%	ND	38% (17 of 45)	**67% (2 of 3)**	63%
ΣSEM-AVS	Mussel 28-d S	48	38.5	64.1	28% (11 of 40)	**100% (8 of 8)**	77%	100% (5 of 5)	36% (16 of 45)	**100% (3 of 3)**	67%
Mean PEC-Q_{METALS}	Mussel 28-d S	48	6.03	10.7	33% (14 of 42)	**83% (5 of 6)**	69%	100% (2 of 2)	36% (16 of 44)	**75% (3 of 4)**	65%
Mean PEC-$Q_{METALS(OC)}$	Mussel 28-d S	48	482	621	33% (14 of 43)	**100% (5 of 5)**	71%	100% (3 of 3)	37% (17 of 46)	**100% (2 of 2)**	65%
Basis for T_{10}/T_{20}: 28-d *L. siliquoidea* Biomass											
Copper	Mussel 28-d B	48	33.4	47.4	**5% (2 of 41)**	71% (5 of 7)	**92%**	75% (3 of 4)	**11% (5 of 45)**	67% (2 of 3)	**88%**
Lead	Mussel 28-d B	48	1085	1351	**7% (3 of 41)**	57% (4 of 7)	**88%**	33% (1 of 3)	**9% (4 of 44)**	75% (3 of 4)	**90%**
ΣSEM-AVS	Mussel 28-d B	48	41.7	52.8	**7% (3 of 42)**	67% (4 of 6)	**90%**	0% (0 of 2)	**7% (3 of 44)**	**100% (4 of 4)**	**94%**
Mean PEC-Q_{METALS}	Mussel 28-d B	48	7.57	10.3	**9% (4 of 44)**	75% (3 of 4)	**90%**	ND	**9% (4 of 44)**	75% (3 of 4)	**90%**
Mean PEC-$Q_{METALS(OC)}$	Mussel 28-d B	48	449	490	**5% (2 of 42)**	**83% (5 of 6)**	**94%**	50% (1 of 2)	**7% (3 of 44)**	**100% (4 of 4)**	**94%**

-d S = -day survival; -d B = -day biomass; n = number of samples.
COPC = chemical of potential concern; PEC-Q = probable effect concentration-qotients; SEM-AVS = simultaneously extracted metals minus acid volatile sulfides; ND = No data.

[1]Bolded results indicate that the toxicity threshold met the individual evaluation criteria for the T_{10}-value, T_{20}-value, or correct classification rate; shaded results indicate toxicity thresholds that meet all three criteria.

Table A2. Predictive ability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca*, and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass).[1]

COPC/COPC Mixture	Toxicity Test Endpoint Used to Derive STT	n	T_{10}	T_{20}	Incidence of Toxicity						
					$<T_{10}$	$\geq T_{10}$	Correct Classification Rate for T_{10}	T_{10}-T_{20}	$\leq T_{20}$	$>T_{20}$	Correct Classification Rate for T_{20}
Basis for T_{10}/T_{20}: 28-d *H. azteca* Survival											
Cadmium	Amphipod 28-d S	76	11.1	17.3	**11% (5 of 45)**	**61% (19 of 31)**	78%	60% (6 of 10)	20% (11 of 55)	**62% (13 of 21)**	75%
Cadmium	Amphipod 28-d B	76	11.1	17.3	**7% (3 of 45)**	42% (13 of 31)	72%	20% (2 of 10)	**9% (5 of 55)**	**52% (11 of 21)**	**80%**
Cadmium	Mussel 28-d S	48	11.1	17.3	**18% (5 of 28)**	**70% (14 of 20)**	77%	75% (3 of 4)	25% (8 of 32)	**69% (11 of 16)**	73%
Cadmium	Mussel 28-d B	48	11.1	17.3	**4% (1 of 28)**	30% (6 of 20)	69%	25% (1 of 4)	**6% (2 of 32)**	31% (5 of 16)	73%
Cadmium	Midge 10-d S	76	11.1	17.3	22% (10 of 45)	45% (14 of 31)	64%	40% (4 of 10)	25% (14 of 55)	48% (10 of 21)	67%
Cadmium	Midge 10-d B	76	11.1	17.3	**11% (5 of 45)**	**52% (16 of 31)**	74%	30% (3 of 10)	**15% (8 of 55)**	**62% (13 of 21)**	79%
Lead	Amphipod 28-d S	76	150	219	**9% (4 of 45)**	**65% (20 of 31)**	**80%**	33% (3 of 9)	**13% (7 of 54)**	**77% (17 of 22)**	**84%**
Lead	Amphipod 28-d B	76	150	219	**2% (1 of 45)**	48% (15 of 31)	78%	33% (3 of 9)	**7% (4 of 54)**	**55% (12 of 22)**	**82%**
Lead	Mussel 28-d S	48	150	219	22% (6 of 27)	**62% (13 of 21)**	71%	40% (2 of 5)	25% (8 of 32)	**69% (11 of 16)**	73%
Lead	Mussel 28-d B	48	150	219	**4% (1 of 27)**	29% (6 of 21)	67%	0% (0 of 5)	**3% (1 of 32)**	38% (6 of 16)	77%
Lead	Midge 10-d S	76	150	219	20% (9 of 45)	48% (15 of 31)	67%	33% (3 of 9)	22% (12 of 54)	**55% (12 of 22)**	71%
Lead	Midge 10-d B	76	150	219	**13% (6 of 45)**	48% (15 of 31)	71%	33% (3 of 9)	**17% (9 of 54)**	**55% (12 of 22)**	75%
Mean PEC-Q	Amphipod 28-d S	76	0.556	0.732	**11% (5 of 47)**	**66% (19 of 29)**	**80%**	83% (5 of 6)	**19% (10 of 53)**	**61% (14 of 23)**	75%
Mean PEC-Q	Amphipod 28-d B	76	0.556	0.732	**4% (2 of 47)**	48% (14 of 29)	78%	50% (3 of 6)	**9% (5 of 53)**	48% (11 of 23)	78%
Mean PEC-Q	Mussel 28-d S	48	0.556	0.732	21% (6 of 28)	**65% (13 of 20)**	73%	0% (0 of 2)	20% (6 of 30)	**72% (13 of 18)**	77%
Mean PEC-Q	Mussel 28-d B	48	0.556	0.732	**4% (1 of 28)**	30% (6 of 20)	69%	0% (0 of 2)	**3% (1 of 30)**	33% (6 of 18)	73%
Mean PEC-Q	Midge 10-d S	76	0.556	0.732	23% (11 of 47)	45% (13 of 29)	64%	50% (3 of 6)	26% (14 of 53)	43% (10 of 23)	64%
Mean PEC-Q	Midge 10-d B	76	0.556	0.732	**15% (7 of 47)**	48% (14 of 29)	71%	17% (1 of 6)	**15% (8 of 53)**	**57% (13 of 23)**	76%
Mean $PEC\text{-}Q_{METALS}$	Amphipod 28-d S	76	1.11	1.78	**9% (4 of 45)**	**65% (20 of 31)**	**80%**	67% (6 of 9)	**19% (10 of 54)**	**64% (14 of 22)**	76%
Mean $PEC\text{-}Q_{METALS}$	Amphipod 28-d B	76	1.11	1.78	**4% (2 of 45)**	45% (14 of 31)	75%	33% (3 of 9)	**9% (5 of 54)**	50% (11 of 22)	79%
Mean $PEC\text{-}Q_{METALS}$	Mussel 28-d S	48	1.11	1.78	**19% (5 of 27)**	**67% (14 of 21)**	75%	25% (1 of 4)	**19% (6 of 31)**	**76% (13 of 17)**	79%
Mean $PEC\text{-}Q_{METALS}$	Mussel 28-d B	48	1.11	1.78	**4% (1 of 27)**	29% (6 of 21)	67%	0% (0 of 4)	**3% (1 of 31)**	35% (6 of 17)	75%
Mean $PEC\text{-}Q_{METALS}$	Midge 10-d S	76	1.11	1.78	22% (10 of 45)	45% (14 of 31)	64%	44% (4 of 9)	26% (14 of 54)	45% (10 of 22)	66%
Mean $PEC\text{-}Q_{METALS}$	Midge 10-d B	76	1.11	1.78	**11% (5 of 45)**	**52% (16 of 31)**	74%	33% (3 of 9)	**15% (8 of 54)**	**59% (13 of 22)**	78%

Table A2. Predictive ability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca,* and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass).[1]

COPC/COPC Mixture	Toxicity Test Endpoint Used to Derive STT	n	T_{10}	T_{20}	Incidence of Toxicity						
					$<T_{10}$	$\geq T_{10}$	Correct Classification Rate for T_{10}	T_{10}-T_{20}	$\leq T_{20}$	$>T_{20}$	Correct Classification Rate for T_{20}
Basis for T_{10}/T_{20}: 28-d *H. azteca* Survival (cont.)											
ΣSEM-AVS	Amphipod 28-d S	76	7.82	13.7	**11% (5 of 44)**	**59% (19 of 32)**	76%	29% (2 of 7)	**14% (7 of 51)**	**68% (17 of 25)**	**80%**
ΣSEM-AVS	Amphipod 28-d B	76	7.82	13.7	**7% (3 of 44)**	41% (13 of 32)	71%	14% (1 of 7)	**8% (4 of 51)**	48% (12 of 25)	78%
ΣSEM-AVS	Mussel 28-d S	48	7.82	13.7	**17% (5 of 29)**	**74% (14 of 19)**	79%	0% (0 of 2)	**16% (5 of 31)**	**82% (14 of 17)**	**83%**
ΣSEM-AVS	Mussel 28-d B	48	7.82	13.7	**3% (1 of 29)**	32% (6 of 19)	71%	0% (0 of 2)	**3% (1 of 31)**	35% (6 of 17)	75%
ΣSEM-AVS	Midge 10-d S	76	7.82	13.7	20% (9 of 44)	47% (15 of 32)	66%	14% (1 of 7)	20% (10 of 51)	**56% (14 of 25)**	72%
ΣSEM-AVS	Midge 10-d B	76	7.82	13.7	**9% (4 of 44)**	**53% (17 of 32)**	75%	43% (3 of 7)	**14% (7 of 51)**	**56% (14 of 25)**	76%
Zinc	Amphipod 28-d S	76	2083	2949	**13% (7 of 53)**	**74% (17 of 23)**	**83%**	ND	**13% (7 of 53)**	**74% (17 of 23)**	**83%**
Zinc	Amphipod 28-d B	76	2083	2949	**6% (3 of 53)**	**57% (13 of 23)**	**83%**	ND	**6% (3 of 53)**	**57% (13 of 23)**	**83%**
Zinc	Mussel 28-d S	48	2083	2949	22% (7 of 32)	**75% (12 of 16)**	77%	ND	22% (7 of 32)	**75% (12 of 16)**	77%
Zinc	Mussel 28-d B	48	2083	2949	**3% (1 of 32)**	38% (6 of 16)	77%	ND	**3% (1 of 32)**	38% (6 of 16)	77%
Zinc	Midge 10-d S	76	2083	2949	21% (11 of 53)	**57% (13 of 23)**	72%	ND	21% (11 of 53)	**57% (13 of 23)**	72%
Zinc	Midge 10-d B	76	2083	2949	**13% (7 of 53)**	**61% (14 of 23)**	79%	ND	**13% (7 of 53)**	**61% (14 of 23)**	79%
Basis for T_{10}/T_{20}: 28-d *L. siliquoidea* Survival											
Copper	Amphipod 28-d S	75	116	141	29% (21 of 73)	**100% (2 of 2)**	72%	ND	29% (21 of 73)	**100% (2 of 2)**	72%
Copper	Amphipod 28-d B	75	116	141	**18% (13 of 73)**	**100% (2 of 2)**	**83%**	ND	**18% (13 of 73)**	**100% (2 of 2)**	**83%**
Copper	Mussel 28-d S	48	116	141	37% (17 of 46)	**100% (2 of 2)**	65%	ND	37% (17 of 46)	**100% (2 of 2)**	65%
Copper	Mussel 28-d B	48	116	141	**11% (5 of 46)**	**100% (2 of 2)**	**90%**	ND	**11% (5 of 46)**	**100% (2 of 2)**	**90%**
Copper	Midge 10-d S	75	116	141	29% (21 of 73)	**100% (2 of 2)**	72%	ND	29% (21 of 73)	**100% (2 of 2)**	72%
Copper	Midge 10-d B	75	116	141	26% (19 of 73)	**100% (2 of 2)**	75%	ND	26% (19 of 73)	**100% (2 of 2)**	75%
Mean PEC-Q_{METALS}	Amphipod 28-d S	75	6.03	10.7	23% (15 of 66)	**89% (8 of 9)**	79%	100% (4 of 4)	27% (19 of 70)	**80% (4 of 5)**	73%
Mean PEC-Q_{METALS}	Amphipod 28-d B	75	6.03	10.7	**11% (7 of 66)**	**89% (8 of 9)**	**89%**	100% (4 of 4)	**16% (11 of 70)**	**80% (4 of 5)**	**84%**
Mean PEC-Q_{METALS}	Mussel 28-d S	48	6.03	10.7	33% (14 of 42)	**83% (5 of 6)**	69%	100% (2 of 2)	36% (16 of 44)	**75% (3 of 4)**	65%
Mean PEC-Q_{METALS}	Mussel 28-d B	48	6.03	10.7	**7% (3 of 42)**	**67% (4 of 6)**	**90%**	50% (1 of 2)	**9% (4 of 44)**	**75% (3 of 4)**	**90%**

Table A2. Predictive ability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca,* and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass).[1]

COPC/COPC Mixture	Toxicity Test Endpoint Used to Derive STT	n	T_{10}	T_{20}	Incidence of Toxicity						
					$<T_{10}$	$\geq T_{10}$	Correct Classification Rate for T_{10}	T_{10}-T_{20}	$\leq T_{20}$	$>T_{20}$	Correct Classification Rate for T_{20}
Basis for T_{10}/T_{20}: 28-d *H. azteca* Survival (cont.)											
Mean PEC-Q_{METALS}	Midge 10-d S	75	6.03	10.7	26% (17 of 66)	**67% (6 of 9)**	73%	50% (2 of 4)	27% (19 of 70)	**80% (4 of 5)**	73%
Mean PEC-Q_{METALS}	Midge 10-d B	75	6.03	10.7	21% (14 of 66)	**78% (7 of 9)**	79%	75% (3 of 4)	24% (17 of 70)	**80% (4 of 5)**	76%
Mean PEC-$Q_{METALS.(O}$	Amphipod 28-d S	75	482	621	21% (14 of 66)	**100% (9 of 9)**	**81%**	100% (4 of 4)	26% (18 of 70)	**100% (5 of 5)**	76%
Mean PEC-$Q_{METALS.(O}$	Amphipod 28-d B	75	482	621	**9% (6 of 66)**	**100% (9 of 9)**	**92%**	100% (4 of 4)	**14% (10 of 70)**	**100% (5 of 5)**	**87%**
Mean PEC-$Q_{METALS.(O}$	Mussel 28-d S	48	482	621	33% (14 of 43)	**100% (5 of 5)**	71%	100% (3 of 3)	37% (17 of 46)	**100% (2 of 2)**	65%
Mean PEC-$Q_{METALS.(O}$	Mussel 28-d B	48	482	621	**5% (2 of 43)**	**100% (5 of 5)**	**96%**	100% (3 of 3)	**11% (5 of 46)**	**100% (2 of 2)**	**90%**
Mean PEC-$Q_{METALS.(O}$	Midge 10-d S	75	482	621	23% (15 of 66)	**89% (8 of 9)**	79%	100% (4 of 4)	27% (19 of 70)	**80% (4 of 5)**	73%
Mean PEC-$Q_{METALS.(O}$	Midge 10-d B	75	482	621	**18% (12 of 66)**	**100% (9 of 9)**	**84%**	100% (4 of 4)	23% (16 of 70)	**100% (5 of 5)**	79%
ΣSEM-AVS	Amphipod 28-d S	76	38.5	64.1	26% (18 of 68)	**75% (6 of 8)**	74%	60% (3 of 5)	29% (21 of 73)	**100% (3 of 3)**	72%
ΣSEM-AVS	Amphipod 28-d B	76	38.5	64.1	**16% (11 of 68)**	**63% (5 of 8)**	**82%**	40% (2 of 5)	**18% (13 of 73)**	**100% (3 of 3)**	**83%**
ΣSEM-AVS	Mussel 28-d S	48	38.5	64.1	28% (11 of 40)	**100% (8 of 8)**	77%	100% (5 of 5)	36% (16 of 45)	**100% (3 of 3)**	67%
ΣSEM-AVS	Mussel 28-d B	48	38.5	64.1	**5% (2 of 40)**	**63% (5 of 8)**	**90%**	40% (2 of 5)	**9% (4 of 45)**	**100% (3 of 3)**	**92%**
ΣSEM-AVS	Midge 10-d S	76	38.5	64.1	29% (20 of 68)	50% (4 of 8)	68%	20% (1 of 5)	29% (21 of 73)	**100% (3 of 3)**	72%
ΣSEM-AVS	Midge 10-d B	76	38.5	64.1	25% (17 of 68)	50% (4 of 8)	72%	20% (1 of 5)	25% (18 of 73)	**100% (3 of 3)**	76%
Zinc	Amphipod 28-d S	75	20600	23700	28% (20 of 71)	**75% (3 of 4)**	72%	100% (1 of 1)	29% (21 of 72)	**67% (2 of 3)**	71%
Zinc	Amphipod 28-d B	75	20600	23700	**17% (12 of 71)**	**75% (3 of 4)**	**83%**	100% (1 of 1)	**18% (13 of 72)**	**67% (2 of 3)**	**81%**
Zinc	Mussel 28-d S	48	20600	23700	38% (17 of 45)	**67% (2 of 3)**	63%	ND	38% (17 of 45)	**67% (2 of 3)**	63%
Zinc	Mussel 28-d B	48	20600	23700	**11% (5 of 45)**	**67% (2 of 3)**	**88%**	ND	**11% (5 of 45)**	**67% (2 of 3)**	**88%**
Zinc	Midge 10-d S	75	20600	23700	28% (20 of 71)	**75% (3 of 4)**	72%	100% (1 of 1)	29% (21 of 72)	**67% (2 of 3)**	71%
Zinc	Midge 10-d B	75	20600	23700	25% (18 of 71)	**75% (3 of 4)**	75%	100% (1 of 1)	26% (19 of 72)	**67% (2 of 3)**	73%

Table A2. Predictive ability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca,* and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass).[1]

COPC/COPC Mixture	Toxicity Test Endpoint Used to Derive STT	n	T_{10}	T_{20}	Incidence of Toxicity						
					$<T_{10}$	$\geq T_{10}$	Correct Classification Rate for T_{10}	T_{10}-T_{20}	$\leq T_{20}$	$>T_{20}$	Correct Classification Rate for T_{20}
Basis for T_{10}/T_{20}: 28-d *L. siliquoidea* Biomass											
Copper	Amphipod 28-d S	75	33.4	47.4	**19% (12 of 63)**	**92% (11 of 12)**	**83%**	100% (5 of 5)	25% (17 of 68)	**86% (6 of 7)**	76%
Copper	Amphipod 28-d B	75	33.4	47.4	**10% (6 of 63)**	**75% (9 of 12)**	**88%**	80% (4 of 5)	**15% (10 of 68)**	**71% (5 of 7)**	**84%**
Copper	Mussel 28-d S	48	33.4	47.4	32% (13 of 41)	**86% (6 of 7)**	71%	100% (4 of 4)	38% (17 of 45)	**67% (2 of 3)**	63%
Copper	Mussel 28-d B	48	33.4	47.4	**5% (2 of 41)**	**71% (5 of 7)**	**92%**	75% (3 of 4)	**11% (5 of 45)**	**67% (2 of 3)**	**88%**
Copper	Midge 10-d S	75	33.4	47.4	25% (16 of 63)	**58% (7 of 12)**	72%	60% (3 of 5)	28% (19 of 68)	**57% (4 of 7)**	71%
Copper	Midge 10-d B	75	33.4	47.4	21% (13 of 63)	**67% (8 of 12)**	77%	60% (3 of 5)	24% (16 of 68)	**71% (5 of 7)**	76%
Lead	Amphipod 28-d S	75	1085	1351	24% (16 of 66)	**78% (7 of 9)**	76%	75% (3 of 4)	27% (19 of 70)	**80% (4 of 5)**	73%
Lead	Amphipod 28-d B	75	1085	1351	**12% (8 of 66)**	**78% (7 of 9)**	**87%**	75% (3 of 4)	**16% (11 of 70)**	**80% (4 of 5)**	**84%**
Lead	Mussel 28-d S	48	1085	1351	34% (14 of 41)	**71% (5 of 7)**	67%	67% (2 of 3)	36% (16 of 44)	**75% (3 of 4)**	65%
Lead	Mussel 28-d B	48	1085	1351	**7% (3 of 41)**	**57% (4 of 7)**	**88%**	33% (1 of 3)	**9% (4 of 44)**	**75% (3 of 4)**	**90%**
Lead	Midge 10-d S	75	1085	1351	26% (17 of 66)	**67% (6 of 9)**	73%	50% (2 of 4)	27% (19 of 70)	**80% (4 of 5)**	73%
Lead	Midge 10-d B	75	1085	1351	21% (14 of 66)	**78% (7 of 9)**	79%	50% (2 of 4)	23% (16 of 70)	**100% (5 of 5)**	79%
Mean PEC-Q_{METALS}	Amphipod 28-d S	75	7.57	10.3	25% (17 of 68)	**86% (6 of 7)**	76%	100% (2 of 2)	27% (19 of 70)	**80% (4 of 5)**	73%
Mean PEC-Q_{METALS}	Amphipod 28-d B	75	7.57	10.3	**13% (9 of 68)**	**86% (6 of 7)**	**87%**	100% (2 of 2)	**16% (11 of 70)**	**80% (4 of 5)**	**84%**
Mean PEC-Q_{METALS}	Mussel 28-d S	48	7.57	10.3	36% (16 of 44)	**75% (3 of 4)**	65%	ND	36% (16 of 44)	**75% (3 of 4)**	65%
Mean PEC-Q_{METALS}	Mussel 28-d B	48	7.57	10.3	**9% (4 of 44)**	**75% (3 of 4)**	**90%**	ND	**9% (4 of 44)**	**75% (3 of 4)**	**90%**
Mean PEC-Q_{METALS}	Midge 10-d S	75	7.57	10.3	26% (18 of 68)	**71% (5 of 7)**	73%	50% (1 of 2)	27% (19 of 70)	**80% (4 of 5)**	73%
Mean PEC-Q_{METALS}	Midge 10-d B	75	7.57	10.3	22% (15 of 68)	**86% (6 of 7)**	79%	100% (2 of 2)	24% (17 of 70)	**80% (4 of 5)**	76%
Mean PEC-$Q_{METALS(OC}$	Amphipod 28-d S	75	449	490	20% (13 of 65)	**100% (10 of 10)**	**83%**	100% (2 of 2)	22% (15 of 67)	**100% (8 of 8)**	**80%**
Mean PEC-$Q_{METALS(OC}$	Amphipod 28-d B	75	449	490	**8% (5 of 65)**	**100% (10 of 10)**	**93%**	100% (2 of 2)	**10% (7 of 67)**	**100% (8 of 8)**	**91%**
Mean PEC-$Q_{METALS(OC}$	Mussel 28-d S	48	449	490	31% (13 of 42)	**100% (6 of 6)**	73%	100% (2 of 2)	34% (15 of 44)	**100% (4 of 4)**	69%
Mean PEC-$Q_{METALS(OC}$	Mussel 28-d B	48	449	490	**5% (2 of 42)**	**83% (5 of 6)**	**94%**	50% (1 of 2)	**7% (3 of 44)**	**100% (4 of 4)**	**94%**

Table A2. Predictive ability of the sediment toxicity thresholds (STTs) that were derived based on the results of 28-day toxicity tests with the amphipod, *Hyalella azteca*, and the mussel, *Lampsilis siliquoidea* (Endpoints: survival and biomass).[1]

COPC/COPC Mixture	Toxicity Test Endpoint Used to Derive STT	n	T_{10}	T_{20}	Incidence of Toxicity						
					$<T_{10}$	$\geq T_{10}$	Correct Classification Rate for T_{10}	T_{10}-T_{20}	$\leq T_{20}$	$>T_{20}$	Correct Classification Rate for T_{20}
Basis for T_{10}/T_{20}: 28-d *L. siliquoidea* Biomass (cont.)											
Mean PEC-$Q_{METALS(OC}$	Midge 10-d S	75	449	490	23% (15 of 65)	**80% (8 of 10)**	77%	50% (1 of 2)	24% (16 of 67)	**88% (7 of 8)**	77%
Mean PEC-$Q_{METALS(OC}$	Midge 10-d B	75	449	490	**18% (12 of 65)**	**90% (9 of 10)**	**83%**	50% (1 of 2)	**19% (13 of 67)**	**100% (8 of 8)**	**83%**
ΣSEM-AVS	Amphipod 28-d S	76	41.7	52.8	29% (20 of 70)	**67% (4 of 6)**	71%	0% (0 of 2)	28% (20 of 72)	**100% (4 of 4)**	74%
ΣSEM-AVS	Amphipod 28-d B	76	41.7	52.8	**19% (13 of 70)**	50% (3 of 6)	79%	0% (0 of 2)	**18% (13 of 72)**	**75% (3 of 4)**	**82%**
ΣSEM-AVS	Mussel 28-d S	48	41.7	52.8	31% (13 of 42)	**100% (6 of 6)**	73%	100% (2 of 2)	34% (15 of 44)	**100% (4 of 4)**	69%
ΣSEM-AVS	Mussel 28-d B	48	41.7	52.8	**7% (3 of 42)**	**67% (4 of 6)**	**90%**	0% (0 of 2)	**7% (3 of 44)**	**100% (4 of 4)**	**94%**
ΣSEM-AVS	Midge 10-d S	76	41.7	52.8	30% (21 of 70)	50% (3 of 6)	68%	0% (0 of 2)	29% (21 of 72)	**75% (3 of 4)**	71%
ΣSEM-AVS	Midge 10-d B	76	41.7	52.8	26% (18 of 70)	50% (3 of 6)	72%	0% (0 of 2)	25% (18 of 72)	**75% (3 of 4)**	75%

-d S = -day survival; -d B = -day biomass; n = number of samples.
COPC = chemical of potential concern; PEC-Q = probable effect concentration-qotients; SEM-AVS = simultaneously extracted metals minus acid volatile sulfides; ND = No data; OC = organic carbon.
[1]Bolded results indicate that the toxicity threshold met the individual evaluation criteria for the T_{10}-value, T_{20}-value, or correct classification rate.

Table A3. Incidence of toxicity to *Ampelisca abdita* and *Hyalella azteca* exposed to whole-sediment samples with various mean probable effect concentration-quotient (PEC-Q) distributions.

Species Tested	Endpoint Measured	Mean PEC-Q Range	Number of Samples	Number of Toxic Samples	Proportion Toxic
*Ampelisca abdita**	10-day survival	<0.24	124	61	48.4 %
		0.24 to <0.45	16	16	100.0 %
		$\geq$ 0.45	25	23	92.0 %
*Hyalella azteca***	28-day survival	<0.24	75	14	18.7 %
		0.24 to <0.45	9	6	66.7 %
		$\geq$ 0.45	16	11	68.8 %

*Toxicity was determined based on comparisons to reference results for Phase II samples and to control results for historical sites.
**Toxicity was determined based on comparison to reference results.

Table A4. Biological conditions that occur within the three categories of risk to the benthic invertebrate community in the Calcasieu Estuary, identified using the risk designations assigned to each sample.

Benthic Metric/Toxicity Test	Endpoint Measured	Low	Indeterminate	High
		mean ± SD (n)	mean ± SD (n)	mean ± SD (n)
Sediment Toxicity				
28-d *Hyalella azteca*	% survival	91.6 ± 7.03 (54)	80.5 ± 19.5 (15)	53.6 ± 28.6 (20)
28-d *Hyalella azteca*	length (mm)	3.82 ± 0.487 (54)	3.80 ± 0.625 (15)	3.76 ± 0.555 (19)
10-d *Ampelisca abdita*	% survival	62.4 ± 17.3 (54)	43.1 ± 23.6 (15)	15.5 ± 17.6 (20)
60-m *Arbacia punctulata*	% fertilization	68.4 ± 25.8 (30)	56.2 ± 36.2 (10)	23.0 ± 29.1 (5)
Benthic Invertebrate Community Structure				
Mean total abundance (H/H)	#/35.4 cm sq.	3.94 ± 3.38 (54)	1.48 ± 1.54 (15)	1.52 ± 2.63 (20)
Mean total abundance (H/M)	#/35.4 cm sq.	3.53 ± 5.04 (54)	0.787 ± 0.955 (15)	0.420 ± 0.908 (20)
Mean total abundance (L/L)	#/35.4 cm sq.	0.300 ± 1.18 (54)	0.760 ± 2.94 (15)	0 ± 0 (20)
Mean total abundance (M/H)	#/35.4 cm sq.	0.0667 ± 0.145 (54)	0.0667 ± 0.209 (15)	0.0200 ± 0.0894 (20)
Mean total abundance (M/L)	#/35.4 cm sq.	0.633 ± 1.78 (54)	0.587 ± 2.27 (15)	0 ± 0 (20)
Mean total abundance (M/M)	#/35.4 cm sq.	0.548 ± 0.734 (54)	0.293 ± 0.506 (15)	0.0800 ± 0.151 (20)
Nonnormalized mIBI	no units	9.15 ± 8.59 (54)	6.88 ± 14.0 (15)	2.56 ± 2.07 (20)
Normalized mIBI	no units	0.495 ± 0.177 (54)	0.354 ± 0.136 (15)	0.299 ± 0.058 (20)
Pollution Indicator Spp. (H/H + H/M + M/H)	#/35.4 cm sq.	7.54 ± 7.65 (54)	2.33 ± 2.00 (15)	1.96 ± 3.03 (20)
Pollution Sensitive (L/L + M/L)	#/35.4 cm sq.	0.933 ± 2.32 (54)	1.35 ± 5.22 (15)	0 ± 0 (20)
Richness = total # sp.	# species/35.4 cm sq.	6.72 ± 4.38 (54)	3.87 ± 3.64 (15)	2.45 ± 1.93 (20)
Total Abundance	#/35.4 cm sq.	9.03 ± 8.38 (54)	4.00 ± 7.35 (15)	2.07 ± 3.14 (20)

SD = standard deviation; n = number of samples; d = day; m = minute; H = high; M = medium; L = low; sp. = species; mIBI = macroinvertebrate index of biotic integrity; cm sq. = squared centimeters.

Figure A1. Relationship between the geometric mean of the mean PEC-Q and the average survival of the freshwater amphipod, *Hyalella azteca*, in 28-d toxicity tests (data source: MacDonald *et al.* 2002; dashed lines represent 95% prediction limits).