United States
Environmental Protection
Agency

Predicting Toxicity to Amphipods From Sediment Chemistry

United States
Environmental Protection Agency/ORD
National Center for
Environmental Assessment
Washington, DC 20460

Official Business
Penalty for Private Use
$300

EPA/600/R-04/030
March 2005

Please make all necessary changes on the below label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports CHECK HERE ☐; detach, or copy this cover, and return to the address in the upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

EPA/600/R-04/030
March 2005

Predicting Toxicity to Amphipods From Sediment Chemistry

National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC 20460

DISCLAIMER

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

The contribution of contaminated sediments to effects on sediment-dwelling organisms (including plants and invertebrates), aquatic-dependent wildlife (amphibians, reptiles, fish, birds, and mammals), and human health has become more apparent in recent years. Sediments can serve both as reservoirs and as potential sources of contaminants to the water column and can adversely affect sediment-dwelling organisms by causing direct toxicity or altering benthic invertebrate community structure. Although the results of sediment toxicity tests and benthic invertebrate community assessments can be used directly to evaluate or infer effects on resident sediment-dwelling organisms, effective interpretation of sediment chemistry data requires tools that link chemical concentrations to the potential for observing adverse biological effects.

This report describes the development of logistic regression models that quantify relationships between the concentrations of contaminants in field-collected sediments and the classification of samples as toxic on the basis of tests using two species of marine amphipods, *Rhepoxynius abronius* and *Ampelisca abdita.* Individual chemical logistic regression models were developed for 37 chemicals of potential concern in contaminated sediments to predict the probability that a sample would be classified as toxic. These models were derived from a large database of matching sediment chemistry and toxicity data that includes contaminant gradients from a variety of habitats in coastal North America. Chemical concentrations corresponding to a 20, 50, and 80% probability of observing sediment toxicity (T20, T50, and T80 values) were calculated to illustrate the potential for deriving application-specific sediment effect concentrations and to provide probability ranges for evaluating the reliability of the models.

The individual chemical regression models were combined into a single model to estimate the probability of toxicity on the basis of the mixture of chemicals present in a sample. The average predicted probability of toxicity closely matched the observed proportion of toxic samples within the same ranges, demonstrating the overall reliability of the P_Max model for the database that was used to derive the model. The magnitude of the toxic effect (decreased survival) in the amphipod test increased as the predicted probability of toxicity increased.

The logistic models have a number of applications, including estimating the probability of observing acute toxicity in estuarine and marine amphipods in 10-day toxicity tests based on sediment chemistry. The models can also be used to estimate the chemical concentrations that correspond to specific probabilities of observing sediment toxicity. Most importantly, the models provide a framework for site-specific and regional assessments and for evaluating other saltwater and freshwater endpoints.

Preferred citation:

U.S. EPA (Environmental Protection Agency). (2005) Predicting toxicity to amphipods from sediment chemistry. National Center for Environmental Assessment, Washington, DC; EPA/600/R-04/030. Available from: National Technical Information Service, Springfield, VA, and online at http://www.epa.gov/ncea.

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF ABBREVIATIONS AND ACRONYMS

AET	Apparent effect threshold
ASTM	American Society for Testing Materials
BEDS	Biological Effects Database for Sediments (MacDonald Environmental Sciences)
DDD	5-dichlorodiphenyldichloroethane
DDE	Dichlorodiphenyldichloroethylene
DDT	Dichlorodiphenyltrichloroethane
EMAP	Environmental Monitoring and Assessment Program (U.S. EPA)
ERL	Effect range low
ERM	Effect range median
LC50	Median lethal concentration
LRM	Logistic regression model
MLML	Moss Landing Marine Laboratory (California)
MSD	Minimum significant difference (classification of samples as toxic)
NSTP	National Status and Trends Program (of the National Oceanic and Atmospheric Administration)
P_Max	The maximum probability of observing toxicity, taken from the set of probabilities calculated for each individual chemical in a sample
P_Max model	A regression model that predicts the probability that a sample will be toxic to amphipods on the basis of the maximum probability of observing toxicity calculated for all individual chemicals in a sample
P_Avg	The mean probability of observing toxicity based on the set of probabilities calculated for each individual chemical in a sample
P_Avg model	A regression model that predicts the probability that a sample will be toxic to amphipods based on the probability of observing toxicity averaged over all of the individual chemicals in a sample
P_Prod	The product of the probabilities of surviving exposure to all individual chemicals in the sample
PAH(s)	Polycyclic aromatic hydrocarbon(s)
PCB(s)	Polychlorinated biphenyl(s)
PEL	Probable effect level
SEDQUAL	Sediment Quality Information System (State of Washington Department of Ecology's Puget Sound database)
Sig Only	Significance only (classification of samples as toxic)
SQGs	Sediment quality guidelines
TEL	Threshold effect level
TOC	Total organic carbon
Tp	The concentration that corresponds to a toxic response of "p" percent according to the single chemical logistic models; for example, the T50 is the concentration of a chemical that corresponds to the probability that 50% of the samples would be toxic

PREFACE

The U.S. EPA's National Center for Environmental Assessment developed this report jointly with the National Oceanic and Atmospheric Administration, Coastal Protection and Restoration Division, with substantial contributions from the U.S. Geological Survey. The report is intended for risk assessors, field biologists, and research scientists interested in the development and application of methods for evaluating the ecological risks associated with chemicals in sediments.

Effective interpretation of sediment chemistry data requires tools that link chemical concentrations to the potential for observing adverse biological effects. This report describes the development of logistic regression models that quantify relationships between the concentrations of sediment-associated contaminants and toxicity in two species of marine amphipods. The models were developed using a large database of matching whole-sediment chemistry and toxicity data that contains data published up until 2000. Amphipod toxicity tests were used as a surrogate for valued ecological attributes that are more difficult to test and measure, including structure and function of benthic communities, population viability of wildlife that depend on benthos, and ecosystem processes such as organic matter decomposition and water filtration. Because amphipod sediment toxicity tests are conducted using documented, standardized methods, they are particularly amenable to analyses that combine results across studies, such as those conducted in this project.

The logistic regression model (LRM) approach described in this report is similar to other empirical approaches for deriving sediment quality guidelines in its reliance on matching field-collected sediment chemistry and biological effects data. In contrast to other approaches to developing sediment quality guidelines, however, the LRM approach does not identify threshold values. Instead, it develops models that enable users to select the probability of observing toxicity that corresponds to the users' specific objectives or to estimate the probability of observing effects at a particular chemical concentration. The models provide a nationwide framework that can be used to evaluate site-specific data, guide data collection efforts, and compare ecological risks across sites and regions.

AUTHORS, CONTRIBUTORS, AND REVIEWERS

The National Center for Environmental Assessment (NCEA) of the U.S. Environmental Protection Agency's (EPA's) Office of Research and Development was responsible for the publication of this document. It was prepared by NCEA and the National Oceanic and Atmospheric Administration, Coastal Protection and Restoration Division, under Interagency Agreements DW13937819-01-2 and DW1393980401-0.

Authors

L. Jay Field
National Oceanic and Atmospheric Administration
Office of Response and Restoration
Coastal Protection and Restoration Division
7600 Sand Point Way, NE
Seattle, WA 98115

Susan B. Norton
U.S. Environmental Protection Agency
National Center for Environmental Assessment
1200 Pennsylvania Ave. NW (8623-N)
Washington, DC 20460

Donald D. MacDonald
MacDonald Environmental Sciences Ltd.
4800 Island Highway, N.
Nanaimo, British Columbia V9T 1W6

Corinne G. Severn
Premier Environmental Services
10999 Pumpkin Ridge Avenue
Las Vegas, Nevada 89135

Christopher G. Ingersoll
U.S. Geological Survey
Columbia Environmental Research Center
4200 New Haven Road
Columbia, MO 65201

Contributors

Dawn Smorong, MacDonald Environmental Sciences Ltd., Nanaimo, British Columbia
Rebekka Lindskoog, MacDonald Environmental Sciences Ltd., Nanaimo, British Columbia
Lorraine Read, TerraStat Consulting Group, Seattle, WA
Carolyn Hong, EVS Consultants, Seattle, WA

Reviewers

David Mount
U.S. Environmental Protection Agency
National Health and Ecological Effects Research Laboratory
6201 Congdon Boulevard
Duluth, MN 55804

Keith Sappington
U.S. Environmental Protection Agency
National Center for Environmental Assessment (8623-N)
1200 Pennsylvania Ave. NW
Washington, DC 20460

Scott Ireland
U.S. Environmental Protection Agency
USEPA REGION 5
77 West Jackson Boulevard (G-17J)
Chicago, IL 60604-3507

Donald M. MacDonald
National Oceanic and Atmospheric Administration
Office of Response and Restoration
Coastal Protection and Restoration Division
7600 Sand Point Way, NE
Seattle, WA 98115

John H. Gentile
Harwell, Gentile & Associates, LC
98 Moss Lane
Brewster, MA 02631

Susan Kane Driscoll
Menzie-Cura and Associates
Winchester, MA 01890

James Shine
Harvard School of Public Health
Department of Environmental Health
Boston, MA 02215

Acknowledgments

The authors gratefully acknowledge the intellectual contributions and support from Ed Long, Rusty Fairey, Dave Mount, Scott Ireland, Jim Keating, Jon Harcum, Rick Swartz, Scott Carr, Pam Haverland, Judy Crane, Sherri Smith, Linda Porebski, Gail Sloane, Steve Bay, Tom Gries, Ron Gouguet, Peter Landrum, Eric Smith, Paul Pinsky, and Pat Cirone.

This project would not be possible without contributions of data from many individuals and programs. Data providers: NSTP from Ed Long, California database from Rusty Fairey, Tom Gries (special thanks to Tom for providing the replicate data for the Puget Sound studies), and Phil Crocker. EMAP data were provided by D. Keith and M. Hughes (Virginian Province), J. Macauley and L. Harwell (Lousianian Province), J. Hyland and T. Snoots (Carolinian Province), and D. Adams (NY/NJ REMAP).

Supplementary funding for data compilation was provided by EPA's Office of Water and the National Research Council of Canada.

1. EXECUTIVE SUMMARY

1.1. SUMMARY

The contribution of contaminated sediments to effects on sediment-dwelling organisms (including plants and invertebrates), aquatic-dependent wildlife (amphibians, reptiles, fish, birds, and mammals), and human health has become more apparent in recent years. Sediments can serve both as reservoirs and as potential sources of contaminants to the water column and can adversely affect sediment-dwelling organisms by causing direct toxicity or altering benthic invertebrate community structure. Although the results of sediment toxicity tests and benthic invertebrate community assessments can be used directly to evaluate or infer effects on resident sediment-dwelling organisms, effective interpretation of sediment chemistry data requires tools that link chemical concentrations to the potential for observing adverse biological effects.

This report describes the development of logistic regression models that quantify relationships between the concentrations of sediment-associated contaminants and toxicity to two commonly tested species of marine amphipods, *Rhepoxynius abronius* and *Ampelisca abdita*.

Amphipod toxicity tests are used as a surrogate for valued ecological attributes that are more difficult to test and measure, including the structure and function of benthic communities, population viability of wildlife that depends on benthos, and ecosystem functions such as organic matter decomposition and water filtration. Because amphipod sediment toxicity tests are conducted using documented, standardized methods (ASTM, 2002a, b), they are particularly amenable to analyses that combine results across studies, such as those conducted in this project.

This report describes logistic regression models for 37 individual chemicals. The results of these individual models are then combined into a single explanatory variable for estimating the proportion of toxic samples expected in field-collected sediment samples. In addition, the report illustrates the applications of the individual logistic models for evaluating sediment quality guidelines and the use of the multiple-chemical models to predict toxicity for other locations and endpoints. The report is intended for risk assessors, field biologists, and research scientists interested in the development and application of methods for evaluating the ecological risks associated with chemicals in sediments.

1.1.1. Individual Chemical Models

Logistic regression models for 37 chemicals were developed using a large database of matching whole-sediment chemistry and toxicity data that encompass many different contaminant gradients from a wide variety of habitats in coastal North America. Logistic regression uses a categorical (e.g., yes/no) variable as the dependent variable. Each sample was designated as toxic or not toxic on the basis of a statistical comparison of the number of amphipods that survived in the test sample relative to the negative control sample. The toxicity classification was the dependent variable for the models, and the chemical concentration in the field-collected sample was the explanatory variable. The models combined results from tests that used either marine amphipod, *R. abronius* or *A. abdita.*

The chemical-specific models provide a basis for estimating the proportion of samples expected to be toxic at different chemical concentrations. In contrast to other approaches to evaluating the potential for observing toxicity on the basis of sediment chemistry (e.g., Long et al., 1995; MacDonald et al., 1996, 2000), the logistic regression modeling approach does not rely on specific effects thresholds. Instead, users can use the models to select sediment concentrations (Tp values) that most directly meet the needs of their specific application. For example, the models can be used to estimate concentrations for individual contaminants that are likely to be associated with a relatively low incidence of sediment toxicity (e.g., 10, 15, or 20%). Such point estimates of minimal-effect concentrations might be used in a screening assessment to identify sediments that are relatively uncontaminated and have a low probability of sediment toxicity. Similarly, contaminant concentrations for which there is a high probability of observing adverse effects could be estimated. These higher point estimates could be used to identify sediments that are highly likely to be toxic to amphipods and have a greater magnitude of effect (i.e., higher percent mortality).

The Tp values can be used in much the same way as other sediment guidelines, except that the Tp value provides a specific probability of observing toxicity and is associated with an estimate of variance based on the fit of the model. The logistic regression models do not represent dose-response relationships for individual chemicals; rather, they should be considered to be indicators of toxicity based on field-collected sediment chemical mixtures.

The logistic regression approach was used to evaluate several issues that form the basis for our recommendations for using the models. We used the single-chemical models to evaluate two approaches for designating samples as toxic: (1) less than 90% survival that was

significantly different from negative control samples (Sig Only), and (2) control-normalized survival less than 80% that was significantly different from negative control samples (minimum significant difference [MSD]) (based on analyses by Thursby et al., 1997). The Sig Only approach had a greater tendency to underestimate the toxicity observed at low concentrations; however, this discrepancy may be explained by the presence of other chemicals in the sample. The MSD approach had a greater tendency to overestimate the toxicity observed at higher concentrations. We selected the Sig Only approach for further exploration and development.

We also evaluated two approaches for normalizing sediment chemistry: dry weight and organic carbon. We selected the dry weight normalization approach because the models had higher goodness-of-fit statistics than the organic carbon-normalized sediment chemistry models, and they had smaller differences between observed and predicted toxicity.

The presence of multiple contaminants, many of which may be present at very low concentrations, complicates the evaluation of relationships between individual contaminants and toxicity in field-collected samples. A data screening procedure was used to exclude samples for which the selected chemical would not serve as a good indicator of observed toxicity. We used the single-chemical models to evaluate three alternative screening criteria: (1) include all samples in the model data set for an individual chemical (unscreened), (2) exclude toxic samples that were less than or equal to the mean of nontoxic samples from the same study (1X screening), and (3) exclude toxic samples that were less than or equal to two times the mean of nontoxic samples from the same study (2X screening).

We selected the 1X screening approach for further exploration and development. The models from the unscreened alternative had much lower goodness-of-fit statistics and appeared to show a weaker relationship between chemistry and toxicity than was observed with the other screening alternatives. The 2X screening approach yielded models with slightly higher goodness-of-fit statistics, but the 1X screening approach performed slightly better at concentrations above the T80 value. The 1X approach screened out fewer samples in the model derivation, which may prove important in the future development of models for less frequently measured chemicals.

1.1.2. Multiple-Chemical Models

Because the individual chemical models were derived from field-collected sediments that included mixtures of contaminants, to some extent each individual chemical model represents the

overall toxicity of the mixtures. However, an individual model would be expected to underestimate the probability of observing toxicity in sediments contaminated with multiple chemicals. The results of the individual models were combined to better estimate the probability that a sediment sample would be toxic, based on the mixture of chemicals present in the sample. Two approaches for combining the individual chemical model results into a single explanatory variable representing the chemical mixture—the P_Max model and the P_Avg model—accurately predicted the frequency of toxicity to amphipods observed in the database:

- P_Max is the maximum probability of observing toxicity, taken from the set of probabilities calculated for each individual chemical in the sample, and
- P_Avg is the mean probability of observing toxicity, based on the set of probabilities calculated for each individual chemical in the sample.

The multiple-chemical models were used to evaluate several additional issues, including the relationship between the probability of observing a toxic effect and the magnitude of toxicity, the identification of chemicals most influential in model performance, the performance of the models in predicting toxicity of the two amphipod species, and the performance of the models in predicting toxicity observed in regional data sets or in individual studies.

The magnitude of the effect (decreased survival) in the amphipod test increased as the probability of toxicity increased, demonstrating that samples that are estimated to have the highest probability of toxicity are also likely to be associated with high mortality.

For approximately 70% of the samples, individual chemical regression models for metals produced the maximum probability used in the P_Max model. This should not be construed to imply that metals were causing toxicity in these samples, only that metals appear to be a good indicator of toxicity in field-collected samples. Indeed, removing metals (or other entire chemical classes) from the suite of individual chemical models used to generate the P_Max model resulted in only minor changes in the model and model fit.

Models were developed by combining data from tests that used *R. abronius* or *A. abdita* in order to encompass more areas of the country and a broader range of sediment chemistry. The P_Max model was used to examine differences in model performance for the two species. The observed toxicity was frequently less than predicted for *A. abdita* and greater than predicted for *R. abronius*. Nevertheless, the observed proportion of toxicity in the data for both species was

strongly related to the nationwide model, affording confidence that the combined-species P_Max model provides a common framework that is applicable to both species.

We examined the performance of models in predicting observed toxicity for individual studies within the database. On a study-by-study basis, there was mixed agreement between the frequency of observed toxicity with that predicted by the P_Max model. The mixed performance suggests that the nationwide models should not be applied to individual studies without first evaluating their performance with matching site-specific toxicity and chemistry data. However, the nationwide P_Max model provided a useful, common basis for evaluating toxicity test results for individual sites included in the database. Application of the model to regional subsets of the database used to derive the model demonstrated significant relationships between the P_Max model predictions and both observed proportion toxicity and percent control-adjusted survival. There was also a strong relationship between predicted toxicity and observed toxicity in the Calcasieu Estuary, an independent data set not included in the original derivation of the models.

1.2. RECOMMENDATIONS

As a starting point for most evaluations, we recommend using the P_Max model, which uses the highest predicted probability from any of the individual chemical models as the explanatory variable. We recommend using the P_Max model based on data from all studies (i.e., the nationwide model), on both marine amphipod species, and on 37 chemical-specific models. For the chemical-specific models, we recommend the models that classified samples as toxic on the basis of less than 90% survival that was significantly different from negative control samples (Sig Only), and that screened the data set by excluding toxic samples that were less than or equal to the mean of nontoxic samples from the same study. The bases for these recommendations are summarized briefly below.

We recommend using the P_Max model, which is based on the highest predicted probability from any of the individual chemical models, because it explained slightly more variation in the data set than did the P_Avg model. However, the two models provide slightly different insights into sediment toxicity. P_Avg may better reflect the overall degree of contamination and is less susceptible to overestimating the probability of toxicity at sites with high concentrations of one chemical. In some cases, it may be valuable to use both models to take advantage of the different perspectives that they provide.

We recommend using the nationwide model that combines data for both species of marine amphipods. Combining data across studies and species represents the fullest range of chemical concentrations and environmental conditions. In addition, the nationwide combined model provides a common basis for comparing site- and species-specific results.

We recommend basing the models on the Sig Only classification so that more subtle changes can be retained, particularly at lower concentrations. It may be valuable to compare the Sig Only nationwide model with site-specific data classified using the MSD approach when test variability obscures the relationship between chemistry and response at lower concentrations.

Finally, the chemical-specific models were greatly improved by using a screened data set that excluded toxic samples that were less than or equal to the mean of nontoxic samples from the same study. Using a more stringent criterion, such as excluding toxic samples that were less than or equal to two times the mean of nontoxic samples from the same study, resulted in improved goodness of fit in the models for most chemicals. However, the more stringent criterion excluded an average of 70% of the toxic samples, which may limit future development of models for other endpoints, regions, or chemicals that have fewer total samples. We concluded that the small improvements in model fit did not outweigh the associated reduction in sample size.

1.3. APPLICATIONS

The chemical-specific models provide a basis for estimating the probability that a sample will be toxic for 37 individual contaminants over a wide range of contaminant concentrations. In addition, they are useful for evaluating the degree of risk associated with commonly used sediment quality guidelines (SQGs). The probabilities of toxicity associated with SQG threshold values are generally consistent with their narrative intent. However, logistic regression models have several advantages over current guideline approaches: (a) they present risk on a continuous quantitative scale rather than by defining discrete categories based on threshold values, (b) the continuous estimates of risk allow users to match the degree of risk with their objectives, and (c) they express risk on a common scale of 0 to 1 across all chemicals. The individual chemicals models would be expected to underestimate the probability of observing toxicity in samples that are contaminated with many chemicals. For this purpose, we recommend using the multiple-chemical models that combine the individual model results into a single explanatory value for estimating the probability that a sample will be toxic.

The multiple-chemical models provide a useful basis for conducting screening-level assessments that require classifying or prioritizing samples on the basis of sediment chemistry. Because the models do not consider potential differences in bioavailability or exposure, the probability of toxicity may be over- or underestimated for some locations. Before applying the models to a particular site, we recommend first evaluating how well the models fit the local situation by collecting a test set of matching sediment chemistry and toxicity test data. The logistic regression models can be used to design effective test sampling programs, and they can also suggest issues that require further investigation (e.g., bioavailability). They can be very useful for classifying samples into broad categories of concern on the basis of sediment chemistry. The models should not be considered a complete substitute for direct effects assessment (e.g., toxicity tests).

We evaluated the relationship between model predictions and the results of other toxicity endpoints, including those commonly used in freshwater systems. The P_Max model predictions appear to be useful for predicting sea urchin response for *Arbacia punctulata*, based on development or fertilization tests, but not for *Strongylocentrotus purpuratus*. The models may also be useful for predicting the response of freshwater amphipods, particularly the 28-day *Hyallela azteca* growth and survival endpoint. There is the potential for developing endpoint-specific models as more data are acquired.

2. INTRODUCTION

The contribution of contaminated sediments to effects on sediment-dwelling organisms (including plants and invertebrates), aquatic-dependent wildlife (amphibians, reptiles, fish, birds, and mammals), and human health has become more apparent in recent years (Long and Morgan, 1991; U.S. EPA, 1997). Many toxic contaminants, such as metals, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), chlorophenols, and pesticides, are found in only trace amounts in water, but they can accumulate to elevated levels in sediments (Ingersoll et al., 1997). Therefore, sediments can serve both as reservoirs and as potential sources of contaminants to the water column.

Contaminants associated with sediments can adversely affect resident sediment-dwelling organisms by causing direct toxicity or by altering benthic invertebrate community structure (Chapman, 1989). Furthermore, contaminated sediments can adversely affect fish and wildlife species, either through direct exposure or through bioaccumulation in the food web.

A variety of approaches are used to evaluate the hazard posed by contaminated sediments to ecological receptors (Ingersoll et al., 1997). These approaches include sediment chemistry measurements, *ex situ* toxicity tests, benthic invertebrate community surveys, sediment toxicity identification and evaluation procedures, and bioaccumulation assessments.

The results of sediment toxicity tests and benthic invertebrate community assessments can be used directly to evaluate or infer effects on resident sediment-dwelling organisms. However, effective interpretation of sediment chemistry data requires tools that link chemical concentrations to the potential for observing adverse biological effects. Sediment chemistry values that have been linked to biological effects can provide an efficient means for evaluating the risks of sediment contamination when biological tests or surveys are unavailable.

The equilibrium partitioning approach links sediment chemistry values with biological effects by combining the results of controlled laboratory tests using manipulated concentrations of chemicals with theory on the factors controlling bioavailability (Di Toro et al., 1991, 2000). Numerical sediment quality guidelines (SQGs) empirically link biological effects with sediment chemistry by combining the results of toxicity tests using field-collected samples with the concentrations of chemicals in those same samples (Long et al., 1995; MacDonald et al., 1996, 2000).

The logistic regression model (LRM) approach described in this report is similar to other empirical approaches for deriving SQGs because it relies on matching field-collected sediment chemistry and biological effects data (e.g., sediment toxicity or benthic invertebrate community structure effects). In contrast to other approaches to developing SQGs, however, the LRM approach does not develop threshold values. Instead, it develops models that capture the relationship between sediment chemistry and the probability of observing a toxic response. By representing the relationship in continuous form, users can select the probability of observing sediment toxicity that corresponds to their specific objectives. The relationships can also be used to estimate the probability of observing effects, given the mixture of chemicals at a particular location (Field et al., 1999, 2002).

The primary objectives of this report are to describe the development of individual chemical LRMs, based on the standard marine and estuarine amphipod 10-day lethality toxicity test endpoint (Chapter 4), and to combine these individual models into a single model for predicting toxicity in field-collected sediment samples (Chapter 5). In addition, the report illustrates the applications of the individual logistic models for evaluating SQGs and the use of the combined models to predict toxicity for other sites and endpoints (Chapter 6).

The development of LRMs requires a large database of matching sediment chemistry and toxicity data that includes a broad range of concentrations. Chapter 3 describes the development of the SEDTOX02 database, which contains more than 3200 samples with matching sediment chemistry and toxicity test results.

This report contains detailed descriptions of the database and model development. It is intended for risk assessors, field biologists, and research scientists interested in a deeper understanding of the different data analysis and treatment options considered during model development and the strengths, limitations, and recommended application of the final models. Our major findings have been published in two journal articles (Field et al., 1999, 2002). We also refer readers to a companion effort (Smith et al., 2003) that fit multiple regression models using many chemical concentrations measured in the sample as explanatory variables simultaneously. These models require chemistry results for all of the chemicals used in each model. The LRMs presented here were developed for application to a greater variety of sediment chemistry results. This flexibility was used in the U.S. Environmental Protection Agency's (EPA's) National Sediment Quality Survey (U.S. EPA, 2004), which used the

approach described in this report to classify sediments into three tiers of probability of adverse effects on the basis of a wide variety of sediment chemistry results.

3. DATABASE DEVELOPMENT

3.1. INTRODUCTION

This investigation compiled synoptically collected sediment chemistry and sediment toxicity data from throughout North America into the SEDTOX02 database (NOAA, 2004), http://www.response.restoration.noaa.gov/cpr/sediment/sed_tox.html. The database is divided into separate marine and freshwater databases of identical structure. The primary sources of the estuarine and marine data included the National Oceanic and Atmospheric Administration's National Status and Trends Program (NSTP), EPA's Environmental Monitoring and Assessment Program (EMAP), Moss Landing Marine Laboratory (MLML) (which compiled data for the state of California), the State of Washington Department of Ecology's Puget Sound Database (SEDQUAL), and MacDonald Environmental Sciences' Biological Effects Database for Sediments (BEDS). Appendices A1 and A2 contain the references for the marine and the freshwater SEDTOX02 databases, respectively.

Many geographic areas along the Atlantic, Gulf, and Pacific coasts are represented in the database, and it includes information on several marine and freshwater toxicity endpoints. However, this report focuses on analyses using data from the EPA and American Society for Testing Materials standard 10-day amphipod survival toxicity tests with *Ampelisca abdita* and *Rhepoxynius abronius* (U.S. EPA, 1994a, b; ASTM, 2002a). The database for this project was developed in Microsoft FoxPro; the relational database structure is provided in Appendix B.

3.2. COMPILATION OF MATCHING SEDIMENT CHEMISTRY AND TOXICITY DATA

All of the candidate data sets considered for inclusion in the database were critically evaluated. Application of acceptance criteria (see appendices C and D) to individual studies provided a basis for determining whether experimental designs and measurement endpoints, sample collection and handling procedures, toxicity testing protocols and environmental conditions, control responses, and analytical methods were consistent with established procedures (Long et al., 1995; MacDonald et al., 1996; Field et al., 1999; ASTM, 2002a, b). In the case of the data sets from NSTP, EMAP, SEDQUAL, and MLML sources, the standard protocols established under each program were evaluated, and individual studies were generally examined to identify possible deviations from these protocols. All of the data that met the

acceptance criteria were incorporated into the project database. Samples were excluded from the database if the survival in the associated negative control sample was less than 85% (expressed as a mean of the negative control replicates).

Data sets that met the screening criteria were compiled in spreadsheets. To facilitate data entry, a template was designed to standardize the format of the matching sediment chemistry and toxicity data. Each data file included the following fields: location of the investigation (country, area, and site); date of sediment collection; sampling and sample handling protocols or procedures; species and life stage tested; test type (e.g., static porewater); type of test water (e.g., saltwater, freshwater); source of control and reference sediments; endpoint measured; method used to determine whether the sample was toxic or nontoxic; study citation; and additional explanatory comments.

The results of the toxicity tests conducted and the concentrations of all chemical analytes measured were compiled in the spreadsheet that was created for each study. These latter data were compiled on a sample-by-sample basis, including the control, reference, and test samples. All chemical concentrations were entered as normalized to dry weight, with concentrations below analytical detection limits reported at the detection limit value and a below-detection-limit data qualifier added. Toxic or nontoxic descriptors were also assigned to each endpoint for each sediment sample in the database as the first step in toxicity classification. The objective was to standardize the toxicity classification within the database wherever possible. For the marine amphipods, statistical significance derived from a comparison of test samples to the negative control was preferred over other methods for determining significance (e.g., comparison to field reference).

Data originating from NSTP, EMAP, and MLML were received with statistical significance already determined. For Puget Sound data obtained from SEDQUAL, statistical comparisons (one tailed t-test, $\alpha = 0.05$) between the appropriate negative control and the test samples were conducted on the replicate data. Other studies (received from BEDS) were evaluated on an individual study basis. For these studies, a sediment sample was considered toxic if the original investigator conducted suitable statistical analyses and reported that the sample was significantly toxic when compared with the negative control or appropriate reference site. If statistical significance was not determined by the investigator but sufficient information on the replicate sample results or on the standard deviations of the results was provided, then a

modified Student's T-test was conducted to determine the statistical significance of the results for each sample.

If no statistical analysis could be performed, then a sediment sample was considered to be toxic if the measured response was substantially different from the negative control or appropriate reference response. A 20% or greater difference was considered to be substantial in this context, generally reflecting the results of power analyses conducted on the results of numerous toxicity tests (Thursby et al., 1997; Long et al., 1998; Carr and Biedenbach, 1999).

3.3. DATA AUDITING

To ensure the overall integrity of the database, a data verification and auditing plan was developed and implemented. This plan consisted of three main elements.

First, the candidate data set was reviewed to identify potentially erroneous data. Specifically, individual data sets were reviewed to identify improbable or impossible results (e.g., extremely high or low chemical concentrations, dissolved oxygen concentrations exceeding saturation levels, survival of >100%). If anomalous data were identified in this initial review (an infrequent event), the principal investigator on the study was contacted to either verify the reported results or provide the correct data. The primary data source was subsequently corrected to reflect the input provided by the principal investigator.

The second phase of the data auditing process was data verification. For data that were acquired electronically (the majority of the data), a minimum of 10% were compared with the electronic source files. The few candidate data sets obtained in hard copy format (i.e., not in electronic data files), had substantial potential for data transcription errors. For this reason, all of the hand-entered data that were compiled in spreadsheets were fully verified against the original data source prior to importing the data into the database. Two individuals working cooperatively conducted data verification. Any errors or omissions identified were corrected, and the data corrections were subsequently verified in a similar manner.

The third phase of the data auditing process was designed to determine whether any data had been corrupted during the data translation process (i.e., transferring the data from the Excel spreadsheets into the electronic database). To confirm that the data translation subroutines were functioning appropriately, the data for several studies were exported into a spreadsheet format that resembled the spreadsheets that had been constructed initially. The information contained in these recompiled spreadsheets was then verified against the original data source. After the entire

database had been compiled in the relational database format, data screening procedures (e.g., identification of orphan records, general confirmation of the relational database structure, and extreme value checks) were applied to identify potential errors and further ensure the internal consistency of the data in the database.

3.4. DATA TREATMENT TO SUPPORT MODEL DEVELOPMENT

3.4.1. Calculation of Total PCBs

The total concentration of PCBs was calculated for each sediment sample represented in the database. The procedure used to calculate total PCBs depended on how the data were reported in the original study. If only total PCBs were reported, these values were used directly. If the concentrations of PCBs were reported as individual Aroclors (e.g., Aroclor 1242, Aroclor 1248), then the concentrations of the individual Aroclors were summed to determine the concentration of total PCBs. When the concentrations of individual congeners were reported, these values were summed to determine the total PCB concentration. If fewer than 20 congeners were reported, the sum of the congeners was multiplied by 2, following the approach used by NSTP (NOAA, 1989). If both Aroclors and congeners were measured, total PCBs were based on the congener concentrations.

In calculating the total PCB concentration, below-detection-limit values were treated as zero values. If all of the individual chemicals to be summed were below detection or if the detection limit of any one nondetected chemical exceeded the sum of detected values, the highest detection limit of the chemical constituents for the sample was used as the total value and qualified as a below-detection-limit value.

3.4.2. Classification of Toxic Samples

Standardizing the classification of toxic samples in the database was an important step, because the studies included in the database used several methods to designate individual sediment samples as toxic or nontoxic. We used two approaches for identifying a consistent response level across studies and applied them to all studies that used *R. abronius* and *A. abdita*. The first approach, referred to as the "significance only" (Sig Only) approach, classified samples as toxic if the sample was statistically different ($p \leq 0.05$) when compared with the negative control and absolute survival was less than 90%. The criterion of 90% was used to preclude

classifying samples as toxic because of low variability in the negative control and was based on the minimum acceptable mean survival for negative control response in 10-day marine amphipod toxicity tests (ASTM, 2002a). The application of this criterion changed the classification of 119 samples from toxic to nontoxic. The second approach, referred to as the minimum significant difference (MSD) approach, classified samples as toxic if the sample was significantly different ($p \leq 0.05$) when compared with the negative control and the difference in survival between the test sample and control was at least 20%, that is, the test sample had a control-adjusted survival of less than 80%. The difference of 20% corresponded to a power ($1 - \beta$) of 0.9, based on analysis of *A. abdita* data; in this study 90% of the tests could distinguish a difference of 20% at a statistical significance level (α) of 0.05 (Thursby et al., 1997).

3.4.3. Data Screening for Model Development

The presence of multiple contaminants, many of which may be present at very low concentrations, complicates the evaluation of relationships between individual contaminants and toxicity in field-collected sediments. Consequently, the data for samples that were identified as toxic in this investigation were further screened before they were used to develop the logistic models for each individual contaminant (Field et al., 1999). The objective of the screening process was to exclude toxic samples for which the chemical under consideration would not serve as a good indicator of observed toxicity.

Although there are many possible approaches to screening, for simplicity, we evaluated two approaches (1X and 2X) that followed the general screening approach used by Ingersoll et al. (1996) and were similar to approaches used by others (Long and Morgan, 1991; Long and MacDonald, 1992; MacDonald et al., 1996). In both approaches, the concentration of the selected chemical in each toxic sample was compared with the mean of the concentration of that substance in the nontoxic samples collected in the same study and geographic area.

For the 1X screening approach, if the concentration of a chemical in an individual toxic sample was less than or equal to the mean concentration of that chemical in the nontoxic samples from that study area, it was considered unlikely that the observed toxicity could be attributed to that chemical. Therefore, these toxic samples were not included in the "screened" data set used for developing the logistic model for that chemical. For the 2X screening approach, the "screened" data set used to develop the logistic model for a particular chemical excluded toxic

samples less than or equal to two times the mean concentration of that chemical in the nontoxic samples from that study area. For the development of the organic-carbon normalized models, we applied the same procedures to the organic-carbon normalized chemical concentrations. Because all of the screening approaches are based on the chemical concentrations in the nontoxic samples, an important underlying assumption is that the factors influencing bioavailability are similar for both nontoxic and toxic samples.

All nontoxic samples were included in the screened data sets produced using both approaches. Samples from reference stations were treated the same as other samples and included in the analysis. The data for chemical concentrations that were less than the reported detection limit were not used to develop the logistic models.

3.5. DATABASE CONTENTS

The final SEDTOX02 database includes matching sediment chemistry and toxicity data for both marine and freshwater systems. For convenience, in the data verification and analysis steps, the database was separated into marine and freshwater databases with identical structures.

The marine database includes matching sediment chemistry and toxicity data from the Atlantic, Gulf, and Pacific coasts of North America. Data from 10-day toxicity tests with two species of amphipods (*Rhepoxynius abronius* and *Ampelisca abdita*), for which survival is the measured endpoint, represent the largest component of the database (Table 1). The use of species differed by location; in general, *R. abronius* dominated the West Coast studies, whereas the studies conducted in the East used *A. abdita*. Most of the data originated from large programs (EMAP, NSTP, SEDQUAL, MLML), which used standardized methods of chemical analyses and toxicity tests. Overall, 1257 (39%) of the 3223 sediment samples in the database that had matching chemistry and toxicity were toxic to amphipods (i.e., survival was <90% and significantly different from that of the negative control). For *A. abdita*, 24% of the 2012 samples were toxic in 10-day tests (Table 1). A higher proportion of the samples tested with *R. abronius* (i.e., 64% of 1211 samples) were identified as toxic (Table 1). Using the MSD approach to classifying samples as toxic (i.e., control-adjusted survival was <80% and significantly different from that of the negative control), 12.6% of the *A. abdita* samples and 40.8% of the *R. abronius* samples were classified as toxic.

The database includes information on the concentrations of more than 300 chemicals of potential concern at contaminated sediment sites. More than 90% of the samples were analyzed

for at least 10 of the 37 chemicals for which we developed models (Chapter 4). More than 70% of the samples were analyzed for at least 20 of the modeled chemicals. For many of these chemicals, the assembled data span a broad range of chemical concentrations. Table 2 presents the distributions of the chemistry data (10^{th}, 50^{th}, and 90^{th} percentiles) for samples with matching amphipod toxicity data for metals, PAHs, PCBs, and several organochlorine pesticides. These data show that the 10^{th} to 90^{th} percentile concentrations of the individual contaminants typically span two to three orders of magnitude, with ranges often spanning four to six orders of magnitude.

The covariation among chemical concentrations was found to be substantial in a companion analysis (Smith et al., 2003). Principal component analysis was conducted on a reduced data set (n = 2219) that contained samples with complete data on 22 metals and PAHs. Two principal components explained 83% of the variation in the chemical data. When rotated (with varimax rotation), the first factor explained 50% of the variation and was highly correlated with the PAHs. The second factor explained 33% of the variation and was highly correlated with the metals.

The percent total organic carbon (TOC) in test sediments averaged 1.92% (standard deviation = 2.05, n = 3117) and ranged from 0.01 to 29.4%. Based on visual inspection, there was no clear pattern in TOC differences among the different studies (Figure 1). Because only 629 samples had results for acid volatile sulfides (AVS) and simultaneously extracted metals, AVS normalization methods (Hansen et al., 1996) were not used.

The marine database also includes data on 100% porewater embryological development and fertilization endpoints for two species of sea urchin (*Arbacia punctulata* and *Strongylocentrotus purpuratus)* (Table 3). The MSD values for *S. purpuratus* 96-hour embryological development and 1-hour fertilization endpoints were 78% and 88%, respectively, corresponding to a statistical significance (α) level of 0.05 and a power ($1 - \beta$) of 0.9 (Phillips et al., 2001). We used the MSD values derived by Carr and Biedenbach (1999) for *A. punctulata* 48-hour embryological development and 1-hour fertilization endpoints of 83.6% and 84.5%, respectively. These values corresponded to a statistical significance (α) level of 0.05 and a slightly more stringent power ($1 - \beta$) of 0.95. Using either the Sig Only or the MSD approach, a high percentage (>60%) of the 782 samples were toxic for the embryological development endpoint for both species combined, and there was little difference between species in percent of

toxic samples. For the fertilization endpoint, about 40% of the 612 *A. punctulata* samples and 60% of the 212 *S. purpuratus* samples were toxic. For both species and endpoints, the results showed little difference between the Sig Only and MSD classifications in percent of toxic samples.

The distribution of the bulk sediment chemistry data for the urchin tests (Tables 4 and 5) show 10^{th} to 90^{th} percentile concentrations of the individual contaminants ranging from one to three orders of magnitude for most chemicals. The ranges were typically larger for the organic compounds than for metals. Percent TOC in test sediments averaged 2.1% (2.4% for *A. punctulata* and 1.8% for *S. purpuratus*) and ranged from 0.02 to 15.8% for the embryological development test endpoint. Percent TOC values were similar for the fertilization endpoint.

The freshwater database included data from several frequently tested toxicity test endpoints. Samples classified as toxic by the original investigator were taken at face value because less information was available to independently evaluate statistical significance (e.g., replicate data were lacking) and no analyses had been conducted to identify MSD values for the freshwater endpoints. The growth endpoints were treated as growth and survival, so if either growth or survival had a toxic result, the sample was classified as toxic. Approximately 20% of the 585 short-term survival test samples (10–14-day tests were grouped together for analysis) were toxic for the freshwater midge species (*Chironomus tentans* and *Chironomus riparius*), and almost 40% of the 10-day growth samples were toxic (Table 6). For the freshwater amphipod *Hyalella azteca*, 24% of 567 samples were toxic in the 10–14-day survival test, and almost 40% of the 125 samples were toxic in the 28-day growth and survival test.

The chemistry associated with the freshwater toxicity endpoints exhibited ranges of one to three orders of magnitude (Tables 7, 8, and 9). In general, the chemical concentrations in samples from the freshwater database were higher than those in the samples from the marine database. The freshwater short-term toxicity tests also had higher TOC concentrations, with average concentrations >4% and ranges from <0.1 to >50%. The *H. azteca* 28-day growth and survival test samples averaged 3% TOC (standard deviation = 2.1) and ranged from 0.1 to 11.6%.

4. LOGISTIC REGRESSION MODELS FOR INDIVIDUAL CHEMICALS

4.1. INTRODUCTION

This chapter presents logistic regression models (LRMs) that predict the probability of toxicity on the basis of individual chemical concentrations. The toxicity endpoint modeled was the 10-day survival test conducted using two species of marine amphipods (*A. abdita* and *R. abronius*). Our objective for this analysis was to develop single-chemical models that could serve as screening-level concentration-response relationships for individual chemicals. These relationships could then be used to identify concentrations of individual chemicals that represent different degrees of risk. The relationships could also be combined into multiple-chemical models (Chapter 5).

The individual chemical models provide an opportunity to explore and compare different methods of toxicity classification and chemistry normalization. Specifically, we evaluated two approaches for designating samples as toxic: (1) less than 90% survival that was significantly different from negative control samples (Sig Only), and (2) control-adjusted survival less than 80% that was significantly different from negative control samples (MSD) (based on the analysis by Thursby et al., 1997). We evaluated two approaches for normalizing sediment chemistry: dry-weight normalized and organic carbon normalized. Models were evaluated using several approaches: goodness-of-fit statistics, visual examination of plots of the model and the underlying data, and comparison of the predicted probabilities with the proportion of toxic samples observed within ranges of predicted probability for the entire data set.

We also used the models to evaluate the implications of several data treatment options. We evaluated three alternative screening criteria used to decide whether a given sample was included in the model data set for an individual chemical: (1) include all samples, (2) exclude toxic samples with concentrations that were less than or equal to the mean concentration of nontoxic samples from the same study (1X screening), and (3) exclude toxic samples with concentrations that were less than or equal to two times the mean concentration of nontoxic samples from the same study (2X screening). Finally, we evaluated the implications of developing models that combined data for both amphipod species.

4.2. METHODS

4.2.1. Logistic Regression Modeling

Statistical models fit to the marine amphipod data in SEDTOX02 describe relationships between the probability of a toxic outcome in the amphipod tests and concentrations of the chemicals of interest. Exploratory plots of the data generated distributions that resembled typical sigmoidal dose-response curves. The shape of these curves indicated that it might be appropriate to model these relationships using an LRM. Logistic regression is typically applied to dose-response data, such as that generated by spiked-sediment bioassays, or laboratory tests with a binary outcome (Morgan, 1992).

The individual chemical LRMs were developed from the screened data set for each chemical. The data screening procedures used in this study were intended to identify the chemicals that serve as useful indicators of the toxic response observed in individual sediment samples. The screening procedures also transformed the underlying data into a form that is more consistent with the sigmoidal form of LRMs.

The individual chemical LRMs used the dichotomous toxicity test result (toxic or nontoxic) as the dependent variable and the chemical concentration as the explanatory variable. The model parameters (slope, intercept) define the shape of relationship between the chemical concentration (Log10) and the probability of a toxic result. In its simplest form, the logistic model can be described using the following equation:

$$p = \frac{exp[B0 + B1(x)]}{1 + exp[B0 + B1(x)]}$$

where:

p = probability of observing a toxic effect,

B0 = intercept parameter,

B1 = slope parameter, and

x = chemical concentration or log chemical concentration.

This logistic model was applied to the complete screened data for a number of substances to develop relationships between the sediment chemistry and the toxicity test results. For each substance modeled, the intercept (B0), slope (B1), and chi-square statistic (–2 log likelihood)

were determined. The data for each chemical were modeled independently. Thus, there was only a single concentration variable (x) in each individual chemical model. However, for each model, it is possible to consider the addition of various covariates, such as test species or endpoints. If such a covariate is considered, then a separate intercept term, a separate slope term, or both a distinct intercept and a distinct slope term can be fit for each level of the covariate. All of the logistic regression analyses were conducted using the SAS Institute's logistic procedure (SAS Institute, 1990). The slope and intercept parameters for the model were estimated using the maximum-likelihood approach.

The chi-square statistic provides useful information for interpreting the results of the logistic modeling. Specifically, the chi-square statistic was used to determine whether the slope parameter, B1, was significantly different from zero. For all of the models generated, the probability (*p* value) associated with the slope parameter was less than 0.0001; therefore, the null hypothesis (slope = 0) can be rejected. Additionally, the chi-square statistic can be used to assess how well the model fits the data. For data sets with similar sample sizes, a larger chi-square statistic indicates a better fit of the model to the data. Note, however, that for a similar fit, the chi-square statistic increases with sample size and thus cannot be used to compare the fit of data sets that are not roughly the same size. Normalizing the chi-square statistic to the sample size (N) provides a goodness-of-fit measure that could be applied across all the data sets. There are no established criteria for considering a normalized chi-square statistic to be good. For the purposes of this report, models that had a normalized chi-square value of greater than 0.15 were considered to be a good fit. (The use of a stronger criterion is explored in Chapter 6.)

After the parameters are estimated, the model can be inverted to estimate the concentrations that yield a certain response probability. The notation Tp (e.g., T50) is used to denote the concentration that would give a toxic response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic). Confidence intervals for these effect concentrations that describe the uncertainty associated with fitting the model were derived using the delta method. The delta method is based on a truncated Taylor series expansion that uses the variance-covariance matrix derived from the maximum-likelihood fit and the derivative of the function of interest (in this report, T20, T50, and T80 were used as examples) with respect to each parameter (Morgan, 1992).

4.2.2. Concentration Interval Plots

Concentration interval plots were used to visualize the relationship between the matching sediment chemistry and toxicity data for individual contaminants. The plots were prepared by calculating the proportion of toxic samples within discrete concentration intervals. The individual points represent the median of the chemical concentrations in the samples within the interval and the proportion of the samples classified as toxic within the interval. Each point on the plots represents a minimum of 15 individual samples (a greater number of samples was included in the interval if more than one sample had the same concentration). The range represented by each concentration interval was determined from an ascending list of unique sample concentrations for each contaminant, with the number of intervals determined by the total number of unique sample concentrations for the selected contaminant. The purpose of the plots was only to help visualize the general relationship between chemical concentrations and the probability of observing toxicity. The goodness of fit of each model was evaluated using the normalized chi-square statistics discussed above.

4.3. INDIVIDUAL CHEMICAL LRM RESULTS

4.3.1. Model Results

This section presents LRM results for individual chemicals. We describe models that used the two toxicity classification approaches and the two chemistry normalization methods. All of these models combined data for the two amphipod species and used a screening criterion that excluded toxic samples with concentrations less than or equal to the mean of nontoxic samples from the same study (1X screening approach).

4.3.1.1. *Models Based on Sig Only Toxicity Classification*

Acceptable logistic models were generated for 37 substances for the Sig Only approach, including 10 trace metals, 22 individual PAHs, total PCBs, and 4 organochlorine pesticides (Table 10). All of these models used dry-weight chemical concentrations. The slopes for all models were positive, indicating that increased chemical concentrations were associated with increased probability of toxicity. Although the models for all 37 substances had normalized chi-square statistics exceeding our criterion of 0.15, the models for arsenic, nickel, and p,p'-DDE had normalized chi-square values of 0.17, 0.18, and 0.16, respectively, indicating relatively poorer fits.

Concentration interval plots provide additional information for evaluating the relationships between chemical concentration and the probability of observing sediment toxicity in the screened data set used to derive the model (Figure 2). For example, the plots for lead, mercury, and zinc confirm that logistic models provide good fits of the underlying amphipod toxicity data. Importantly, the range of concentrations represented in the database appears to span the effects range, as demonstrated by the low proportion of toxic samples (0%) observed at the lowest concentrations and the high proportion of toxic samples (90 to 100%) observed at the highest chemical concentrations. Similar results were obtained for many of the organic compounds (e.g., fluoranthene and phenanthrene); however, the observed proportion of toxic samples tended to be somewhat lower (roughly 90%) at the highest concentrations of these substances. The plot for p,p'-DDE shows both high variability and the presence of several outliers, consistent with its relatively low normalized chi-square value.

Although the logistic models provide effective tools for estimating the probability of observing sediment toxicity at various chemical concentrations, point estimates of sediment effect concentrations are also useful for assessing sediment quality conditions. As an example, the chemical concentrations that correspond to the 20, 50, and 80% proportion of toxic samples for amphipod survival were determined and designated as Tp values: T20, T50, and T80, respectively (Table 11).

The reliability of the chemical-specific logistic models was evaluated by comparing the probability of toxicity predicted by the models to the proportion of samples actually observed to be toxic. This comparison differs from the concentration interval plots in that the data screened out of the logistic model development process were included in the reliability evaluation. Predicted versus observed values were compared for four ranges of chemical concentrations defined by the Tp values (i.e., <T20, >T20–T50, >T50–T80, and >T80). The percent of samples within each concentration range that were toxic was determined (Table 12). The logistic models and associated point estimates were considered reliable if the observed proportion of toxic samples was consistent with the predicted probability of toxicity.

The results of this evaluation indicate that the logistic models and associated point estimates of sediment effect concentrations generally provide a reliable basis for estimating the observed proportion of toxic samples in the project database. The models underestimated the proportion of toxic samples at concentrations below the T20 value for all 37 chemicals, although 35 of the 37 chemicals were within 10% of the top of the range. The underestimation of toxicity

below the T20 value may be a consequence of the screening procedure, and is discussed further in Section 4.4.1. Between the T20 and T50 values, the proportion of toxic samples observed for most of the chemicals (30 of 37) was within the predicted range of 20 to 50%. The proportion of toxic samples observed between the T50 and T80 values was within the predicted range of 50 to 80% toxicity for all 37 chemicals. Above the T80 value, the proportion of toxicity was equal to or exceeded 80% for 22 chemicals. Arsenic and p,p'-DDE had a substantially lower proportion of toxic samples than predicted. The models for these chemicals would be expected to overestimate toxicity for high concentrations.

Among the logistic models for the various classes of contaminants, those for PAHs were the most reliable. For 16 of 22 PAHs, the actual proportion of toxic samples was correctly predicted within three of the four concentration ranges defined by the Tp values; however, a higher-than-predicted proportion of toxic samples was observed above the T20 values for all PAHs (Table 12). Among the logistic models for the trace metals, those for chromium, copper, lead, mercury, and zinc were the most reliable, as indicated by the level of agreement between the predicted and observed proportion of toxic samples to amphipods. Likewise, the logistic model for total PCBs provided an accurate basis for predicting toxicity to amphipods in the database. A somewhat lower level of reliability was observed for the organochlorine pesticide models.

4.3.1.2. *Models Based on MSD Toxicity Classification*

Acceptable logistic models were generated for 33 substances for the MSD approach, including 7 trace metals, 22 individual PAHs, total PCBs, and 3 organochlorine pesticides (Table 13). All of these models used dry-weight chemical concentrations. The slopes for all models were positive, indicating that increased chemical concentrations were associated with increased probability of toxicity. The models for eight chemicals—chromium, silver, 1-methylnaphthalene, 2-methylnaphthalene, 2-6 dimethylnapthalene, biphenyl, naphthalene, and perylene—had among the lowest normalized chi-square values, indicating relatively poorer fits. Models for antimony, arsenic, nickel, and p,p'-DDE had normalized chi-square values of less than 0.15.

Concentration interval plots for the MSD toxicity classification approach are shown in Figure 3. The proportion of toxic samples is often less than 80% at the highest concentrations in

the database (e.g., zinc and 1-methylphenanthrene). This truncation at the higher concentrations reflects the more stringent criteria for classifying samples as toxic.

The T20, T50, and T80 values are shown in Table 14. The percent of toxic samples within the ranges defined by these Tp values is shown in Table 15. The agreement between the models and the observed proportion of toxic samples was very good below the T80 value. At concentrations below the T20 value, the observed proportion of toxic samples was within the predicted range for all of the chemicals except p,p'-DDT. The observed proportion of toxic samples was within the predicted ranges at concentrations between the T20 and T50 values for all of the 33 models. The proportion of toxic samples observed between the T50 and T80 values was within the predicted range of 50 to 80% toxicity for 31 of the 33 chemicals. However, above the T80 value, the proportion of toxic samples exceeded 80% for only 10 chemicals. Above the T80 value, the models overestimated the proportion of toxic samples observed for 16 chemicals, and insufficient data were available to evaluate seven models.

4.3.1.3. *Models Based on Organic Carbon*-Normalized Chemical Concentrations

Acceptable logistic models based on organic carbon-normalized chemical concentrations were developed for 25 organic chemicals, including 21 individual PAHs, total PCBs, and 3 organochlorine pesticides (Table 16). The number of samples available for developing these models was slightly reduced from the total because a small percent of the samples was not analyzed for organic carbon. (Models based on acid volatile sulfides and simultaneously extracted metals were not pursued because of the low number of samples available with these measurements). All organic carbon-normalized models used the Sig Only classification of toxic samples. The models for four chemicals—1-methylnaphthalene, 2-methylnaphthalene, biphenyl, and naphthalene—had the lowest acceptable normalized chi-square values, indicating relatively poorer fits.

Concentration interval plots for the organic carbon-normalization approach are shown in Figure 4. Contrary to the expectation that normalizing sediment chemistry for nonpolar organic chemicals would reduce the variability in the concentration response relationships, the concentration interval plots do not show less variability than the Sig Only dry-weight normalization plots shown in Figure 2.

The T20, T50, and T80 values based on the organic-carbon normalized models are shown in Table 17. The percent of toxic samples within the ranges defined by these Tp values is shown

in Table 18. The models underestimated the incidence of amphipod toxicity at concentrations below the T20 value for all 25 chemicals, although 21 of the 25 were within 10% of the top of the range. Between the T20 and T50 values, the proportion of toxic samples observed for 8 of the chemicals was within the predicted range of 20 to 50%, and 18 were slightly above the predicted range. The proportion of toxic samples observed between the T50 and T80 values was within the predicted range of 50 to 80% toxicity for all chemicals except biphenyl. Above the T80 value, the proportion of toxicity exceeded 80% for fluorene only, and insufficient data were available to evaluate 7 chemicals. The models for the other 17 chemicals would be expected to overestimate toxicity for high concentrations.

4.3.2. Model Comparisons

We compared the performance of the models to provide additional insights into the strengths and limitations of the different options and to explore the implications of those differences. The models were compared on the basis of goodness of fit and the degree of agreement between the predicted proportion of toxic samples and the proportion observed in the database. The implications of different modeling options were further evaluated by comparing the resulting Tp values.

4.3.2.1. *Goodness-of-Fit Comparisons*

We compared normalized chi-square statistics across all the different chemical models for the different modeling alternatives. We also evaluated the number of chemicals whose models exceeded the 0.15 normalized chi-square criterion.

The Sig Only approach generated acceptable models for 4 more chemicals than did the MSD approach. Models for antimony, arsenic, nickel, and p,p'-DDE had normalized chi-square values of less than 0.15 for the MSD toxicity classification. Regression models based on toxicity classification using the Sig Only approach consistently had higher normalized chi-square values than did those using the MSD approach, with the exception of benzo(b)fluoranthene (Figure 5). The organic carbon-normalized approach generated acceptable models for 25 of the 27 organic chemicals that had acceptable models using dry-weight normalization. Regression models based on dry-weight-normalized concentrations had higher normalized chi-square values than did those using organic carbon-normalized concentrations for 25 of the 27 chemicals (Figure 6).

4.3.2.2. *Reliability Comparisons*

Tables 19, 20, and 21 show the difference between the observed proportion of toxic samples and the average predicted proportion of toxic samples within the four ranges defined by the Tp values for the Sig Only, MSD, and organic carbon-normalization approaches, respectively. Using antimony as an example, of the 1041 samples a with predicted probability of toxicity of <T20, 315 were classified as toxic, corresponding to an observed proportion toxic of 30.3%. The 1041 samples had an average predicted probability of 11%. The difference between the observed and predicted values, 19.2%, is shown in the upper left cell of Table 19. Large differences between observed and predicted values indicate poor model performance. Positive numbers indicate that the observed toxicity is greater than that predicted by the models, whereas negative numbers mean that the observed toxicity is less than that predicted by the models. Differences greater than 20% in either direction are highlighted to facilitate comparisons among the three tables.

Comparing the observed and predicted values for the Sig Only and MSD approaches (Tables 19 and 20, respectively), the MSD differences are smaller for the predictions less than the T50 values. Differences are comparable for the T50–T80 range. For concentrations above the T80, the Sig Only models had fewer large differences between observed and predicted than did the MSD models. On average, the magnitude of the differences at concentrations above the T80 was smaller for the Sig Only models than for the MSD models (absolute average differences of 9.1 and 15.3%, respectively). The Tp values for the Sig Only and MSD approaches show a high degree of correlation (Figure 7). The difference between the Tp values is largest for the T20 values (factor of 3.8) and smallest for the T80 values (factor of 1.7). The organic carbon-normalized models showed many large differences between observed and predicted toxicity for concentrations less than the T50 values and greater than the T80 values (Table 21).

4.3.3. Summary of Toxicity Classification Evaluations

We fit LRMs to marine amphipod data in the SEDTOX02 database using two approaches for classifying samples as toxic and two approaches for normalizing sediment chemistry. The models for nonpolar organic chemicals based on organic carbon-normalized sediment chemistry had lower goodness-of-fit statistics than did the dry-weight-normalized models and larger differences between observed and predicted toxicity.

The two approaches for classifying toxicity had different strengths and limitations. A greater number of models had acceptable goodness-of-fit statistics with the Sig Only approach than with the MSD approach. Models fit using the Sig Only classification approach had slightly better goodness-of-fit statistics than did those fit using the MSD approach. The differences between observed and predicted proportions of toxic samples were smaller for the MSD approach at concentrations less than the T50 value. The Sig Only approach had a greater tendency to underestimate observed toxicity at low concentrations. In contrast, at concentrations greater than the T80 value, differences between observed and predicted proportions of toxic samples were smaller for the Sig Only approach. The MSD approach had a greater tendency to overestimate toxicity at these higher concentrations.

It is not unexpected that individual chemical models would predict a proportion of toxic samples lower than that observed at low concentrations. Chemicals in the sample other than the one being modeled may be responsible for the observed toxicity. Predictions at high concentrations have greater significance for evaluating risk at contaminated sites. For these reasons, we selected the Sig Only approach for further evaluation and development.

4.4. EVALUATION OF ALTERNATIVE SCREENING APPROACHES

As discussed in Section 3.4.3, the data for each contaminant were screened prior to applying the logistic model or plotting the data. The data screening procedure was designed to exclude samples where the chemical under consideration would not serve as a good indicator of observed toxicity. The standard screening approach (1X screening) that we used as the basis for comparisons was described in the methods section (Section 4.2). In this approach, the concentration of the selected chemical in each toxic sample was compared with the mean of the concentration of that substance in the nontoxic samples collected in the same study and geographic area. If the concentration of a chemical in an individual toxic sample was less than or equal to the mean concentration of that chemical in the nontoxic samples from that study area, it was considered unlikely that the observed toxicity could be attributed to that chemical. Therefore, these toxic samples were not included in the screened data set used for developing the logistic model for that chemical.

Although there are many possible ways to identify samples for screening (e.g., various statistical criteria), we confined our evaluation to simple approaches similar to those used previously (Ingersoll et al., 1996; Long and Morgan, 1991; Long and MacDonald, 1992;

MacDonald et al., 1996). This section compares the standard method with two alternatives: (1) using all of the data in developing single-chemical models (unscreened), and (2) eliminating toxic samples having concentrations less than or equal to twice the mean concentration of that chemical in the nontoxic samples from that study area (2X screening).

The models developed using the two alternative screening approaches were first compared with our standard approach on the basis of concentration interval plots and goodness of fit. The 2X screening alternative was further evaluated by comparing the degree of agreement between the predicted proportion of toxic samples with the proportion observed in the database and the resulting Tp values.

4.4.1. Unscreened Versus 1X Screening

The unscreened data set includes all of the data for a chemical, whereas the 1X screening removed 41.3 to 59.5 % (average of 48.5%) of the toxic samples from the derivation of the models. The effects of including all data for a chemical are illustrated in Figure 8 using lead and phenanthrene as representative examples. Each plot shows the LRMs and the concentration interval data for the respective screening approach.

Visual inspection of these plots revealed that inclusion of the data for those toxic samples in which the chemical of concern is a poor indicator of the observed response tended to scatter and skew the data distributions, particularly at lower concentrations. The plots of the unscreened data for lead and phenanthrene show very few intervals with <20% effects. Importantly, the unscreened data for phenanthrene showed only a weak relationship between chemistry and toxicity for concentrations <1000 mg/kg dry weight, and the incidence of effects is approximately 20 to 60% below that concentration. However, there were few differences between the distributions of the screened and unscreened data at higher concentrations.

Models fit with the unscreened data had much lower normalized chi-square values (Figure 9). The models for only four chemicals—copper, fluorene, dieldrin, and DDT—exceeded the normalized chi-square criterion of 0.15 that we used to identify models with fits sufficient to support the calculation of the Tp values. The 1X screening approach yielded a greater number of acceptable models describing the relationship between contaminant concentrations and biological effects.

4.4.2. 2X Screening Versus 1X Screening

The 2X screening approach removed 63.8 to 84.2% (average of 70.4%) of the toxic samples from the derivation of the model. The effects of the 2X screening alternative are illustrated in Figure 10 using lead and fluoranthrene as representative examples. Each plot shows the LRM and the concentration interval data for the respective screening approach.

The concentration interval plots show that the 2X screening approach reduced variability somewhat. The reduced variability was also reflected in slightly higher normalized chi-square values for the models fit using the 2X approach for most chemicals (Figure 11). In addition, the concentration interval plots indicate that the logistic regression curves for the 2X screening approach were shifted to the right at lower concentrations.

Another way to examine the differences between the models developed from the two screening approaches is to compare the Tp values. Figure 12 shows the relationship between the Tp value concentrations calculated using the different screening approaches. The greatest difference was seen in the T20 values, which is consistent with the patterns observed in concentration interval plots. The difference in Tp values decreased with concentration; the T80 values were almost identical.

Comparing differences between the observed and predicted proportions of toxic samples in the 1X and the 2X screening approaches (Tables 19 and 22, respectively), the differences in the 2X screening approach were smaller for predictions below the T50 values. Differences were comparable for the T50–T80 value range. For concentrations above the T80 value, the differences in the 1X screening approach were slightly smaller than those in the 2X approach (absolute average 9.1% difference compared to 10.1%).

4.4.3. Summary of Screening Approach Evaluations

We evaluated alternative screening approaches used to exclude samples where the chemical under consideration would not serve as a good indicator of observed toxicity. The models from the unscreened alternative showed a weaker relationship between chemistry and toxicity than was observed with the other screening alternatives. The models generated using the 2X screening approach had better goodness-of-fit statistics. The 2X approach had lower differences between observed and predicted toxicity at concentrations below the T50 value; however, it had a slightly greater tendency to overestimate toxicity at concentrations above the

T80 value. In addition, the 2X approach excluded approximately 22% more toxic samples from the model derivation.

The 2X screening approach has some advantages over the 1X screening approach, particularly at concentrations less than the T20 value. However, the 1X screening approach performed slightly better at concentrations above the T80 value and screened out considerably fewer toxic samples in the model derivation. We selected one modeling approach to manage the many permutations associated with further model development and exploration. We retained the 1X model because of the better performance at high Tp values and retention of more samples.

4.5. COMPARISON OF TOXICITY TEST ENDPOINTS

We used the LRM approach to investigate the implications of using a model that combines the response of the two amphipod species. We explored more complex models that included different slopes or intercepts for the two species. We compared the Tp values that would result from individual species models with those of the combined models. In addition, we compared the toxicity observed for each species with predictions based on the combined model.

4.5.1. Statistical Comparisons of Species-Specific LRMs

Using the LRM approach, it is possible to fit a separate slope, an intercept, or both for each of the two amphipod species. If the two species responded with different sensitivities to the chemicals, a separate slope or intercept would be statistically significant. We tested the significance of separate slope, intercept terms using a sequential chi-square comparison (Neter et al., 1996). The first test compared a common slope, common intercept model (Model A) with a common slope, different intercept model (Model B). The results of this analysis indicated that an additional intercept term was statistically significant ($\alpha = 0.05$, 1 degree of freedom) for all of the chemicals considered (Table 23). The second test compared Model B with a different slope, different intercept model (Model C). The distinct slope term was not significant ($\alpha = 0.05$, 1 degree of freedom) for 20 of the 37 chemicals. For these chemicals, parallel models with a different intercept for each species were preferred.

4.5.2. Comparison of Tp values for Separate Amphipod Models

We developed separate logistic models for each amphipod species and calculated Tp values for comparison with the combined amphipod models (Tables 24 and 25). The results showed that the T20, T50, and T80 values for *A. abdita* were higher than those of the combined model by a factor of approximately 2, indicating that *A. abdita* is slightly less responsive than the combined model would predict. In contrast, the T20, T50, and T80 values for *R. abronius* were lower than those of the combined model by factors of 3.9 for the T20 value and 3.1 for the T80 value, indicating that *R. abronius* is more responsive than the combined models would predict.

4.5.3. Reliability Comparisons

We compared the observed proportion of toxic samples for the two species in the unscreened data set with predicted values from the combined model within the four ranges defined by the combined model Tp values (i.e., <T20, T20–T50, T50–T80, and >T80). On average, the difference between observed and predicted proportion of toxic samples for *A. abdita* was less than 10% for all concentration ranges (Table 26). The observed proportion of toxic samples using *R. abronius* was much greater than predicted below the T50 value. Above the T50 value, the difference between observed and predicted proportions of toxic samples using *R. abronius* averaged 11%.

4.5.4. Summary of Species Comparisons

The results of the species-specific analyses suggest that there are substantial differences in the chemical-specific models for the two species. The model comparison indicated that parallel models with different intercepts was the preferred model for more than half of the chemicals. Comparing the Tp values derived from separate species models with those derived from the combined model suggests that *R. abronius* has a greater response than *A. abdita* at similar concentrations. Still, on average, the T50 values differed by approximately a factor of 2 for *A. abdita* and a factor of 3 for *R. abronius*. The combined model consistently and substantially underpredicted the observed proportion of toxic samples for *R. abronius* below the T50 value. Above the T50 value, differences between the observed proportion of toxic samples and that predicted using the combined individual chemicals models were minimal for most of the *A. abdita* data and for *R. abronius* data.

The observed species differences have several possible explanations: unmeasured chemicals or factors other than chemistry may have influenced *R. abronius* results at low chemical concentrations, or there may be inherent differences in sensitivity between the two species. However, the greatest difference between the species-specific observations and the combined models are at lower concentrations for *R. abronius*. As discussed above, individual chemical models may underestimate observed toxicity at low concentrations because chemicals in the sample other than the one being modeled may be responsible for the observed toxicity. Therefore, the species-specific differences could also be explained if the database for *R. abronius* contains a disproportionate amount of data from areas with a high degree of contamination from multiple chemicals. This possibility is difficult to investigate with single-chemical models, but the issue is revisited in Chapter 5.

4.6. SENSITIVITY OF MODELS TO ERRORS IN UNDERLYING DATA

We did not conduct a formal analysis to evaluate the sensitivity of the models to potential errors in the underlying database. However, after initial model development and evaluation, we discovered an error in PCB concentration units for 15 samples. These samples had erroneously high concentrations, and, in addition, a relatively small number of other samples in the database had similarly high PCB concentrations. This situation provided an opportunity to evaluate the degree of change in the models resulting from errors in a small number of highly influential values.

The correction in concentrations for these 15 samples changed the LRM for PCBs, particularly at high concentrations (Figure 13). It also improved the model fit: the normalized chi-square value changed from 0.24 to 0.27. As expected, the Tp values changed most at high concentrations; the T80 was reduced by 42% (Table 27).

4.7. INDIVIDUAL CHEMICAL MODELS AND SPIKED-SEDIMENT BIOASSAY MEDIAN LETHAL CONCENTRATION (LC50*)* VALUES

Dose-response data from laboratory spiked-sediment bioassays provide additional perspective on the concentrations of individual chemicals that can be considered to cause toxicity. However, results from spiked-sediment bioassays are not immediately comparable with predictions from the LRMs. Toxicity from spiked-sediment bioassays is more confidently attributed to the chemical added to the sample. In contrast, the response to any individual

chemical measured in field-collected sediments may be confounded by the presence of other chemicals.

In addition, the magnitude of response reported in spiked-sediment bioassays is much greater than the magnitude of response required to classify a sample as toxic in this study. Most of the studies in the literature on spiked-sediment toxicity report LC50s. An LC50 value represents the concentration corresponding to 50% survival of test organisms. In this study, many samples with much higher test survival were classified as toxic. The control-normalized survival averaged across all samples and chemicals decreases with increasing probability of toxicity defined by the model Tp values (Figure 14). The relationship is such that an average survival of 50% (LC50) corresponds to concentrations exceeding T50 values.

Reported LC50 values for 10-day, spiked-sediment toxicity tests conducted with marine amphipods were compared with the probability of toxicity estimated from the individual chemical models (Table 28). Using the logistic models, the probability of toxicity at the reported LC50 values ranged from 0.54 for zinc to 0.97 for mercury, with most estimates falling within the 0.8 to 0.9 range. This is consistent with the average percent survival observed at high probability of toxicity, as shown in Figure 14.

4.8. SUMMARY AND CONCLUSIONS

This section presents LRM results for models that predict the probability of toxicity on the basis of individual chemical concentrations. The toxicity endpoint modeled was the 10-day survival test conducted using two species of marine amphipods (*A. abdita* and *R. abronius*).

The LRM approach was used to explore and compare different methods of toxicity classification and chemistry normalization. We evaluated two approaches for designating samples as toxic: (1) less than 90% survival that was significantly different from negative control samples (Sig Only), and (2) control-normalized survival less than 80% that was significantly different from negative control samples (MSD) (based on the analysis by Thursby et al., 1997). We evaluated two approaches for normalizing sediment chemistry: dry-weight normalized and organic carbon normalized.

Models were evaluated using several approaches: goodness-of-fit statistics, visual examination of plots of the model and the underlying data, and reliability (predicted vs. observed proportion of toxic samples within ranges of probability) for the entire data set. The MSD approach had a greater tendency to overestimate the proportion of toxic samples observed at

higher concentrations. Although the Sig Only approach had a greater tendency to underpredict observed toxicity at low concentrations, this discrepancy may be explained by the presence of other chemicals in the sample. The models based on organic carbon-normalized sediment chemistry had lower goodness-of-fit statistics than did the dry-weight-normalized models, and they also had larger differences between observed and predicted toxicity. We selected the Sig Only approach with dry-weight normalization for further evaluation and development.

We also used the LRM approach to evaluate the implications of several data treatment options. We evaluated three screening criteria used to decide whether a given sample was included in the model data set for an individual chemical: (1) include all samples (unscreened), (2) exclude toxic samples with concentrations that were less than or equal to the mean concentration of nontoxic samples from the same study (1X screening), and (3) exclude toxic samples with concentrations that were less than or equal to two times the mean concentration of nontoxic samples from the same study (2X screening). The models from the unscreened alternative showed a weaker relationship between chemistry and toxicity than was observed with the other screening alternatives. The 2X screening approach has some advantages over the 1X screening approach, particularly at concentrations less than the T20 value. However, the 1X screening approach performed slightly better at concentrations above the T80 value and screened out fewer samples in the model derivation, which may prove important in the future development of models for less-frequently measured chemicals. We concluded that the small improvements in model fit did not outweigh the associated reduction in sample size, and retained the 1X approach for further development and exploration.

Finally, we evaluated the implications of developing models that combined data for the two amphipod species. Agreement between the observed proportion of toxic samples and that predicted using the combined individual chemicals models was good for both species above the T50 values. Below the T50 value, the combined models performed well for *A. abdita* but consistently and substantially underpredicted the observed proportion of toxic samples for *R. abronius*. The discrepancies may be explained by inherent differences in sensitivity between the two species, by a greater responsiveness of *R. abronius* to nonchemical factors, or by the database for *R. abronius* containing a disproportionate amount of data from areas with a high degree of contamination from multiple chemicals. These issues will be investigated further using multiple-chemical models.

The single-chemical models presented in this chapter can be used to develop screening-level concentration-response relationships for individual chemicals. These relationships could then be used to identify concentrations of individual chemicals that correspond to different degrees of risk, depending on the objectives of the user. This application is similar to the current use of SQGs; the use of the single-chemical models in evaluating SQGs is discussed in Chapter 7. The risks of a toxic response posed by the mixture of chemicals present in a particular sample are best evaluated using a multiple-chemical approach, which is discussed in the next chapter.

5. MULTIPLE-CHEMICAL MODELS

5.1. INTRODUCTION

One of the major challenges in assessing the ecological risk associated with exposure to contaminated sediments is the presence of chemical mixtures. Field-collected sediments, as a rule, contain complex mixtures of chemicals and other factors that influence toxic response. Because the individual models described in Chapter 4 were derived from field-collected sediments rather than from laboratory dose-response studies, to some extent the individual models incorporate the overall toxicity of the mixture of chemicals in the samples. We sought to improve the estimates of response by combining the information contained in the individual chemical models into a single estimate of toxic response. The multiple-chemical model could then be applied to predict whether new samples with known sediment chemistry would be expected to produce a toxic response in an amphipod test.

This chapter describes the approach used to develop and evaluate multiple-chemical models. Sections 5.2 and 5.3 discuss model development and results, respectively. The models were used to evaluate several issues (Section 5.4): the relationship between the probability of observing a toxic effect and the magnitude of toxicity, the identification of chemicals that most influence model performance, the effect of reducing the number of individual chemical models used in developing the multiple-chemical models, the performance of the models in predicting toxicity of the two amphipod species, and the performance of the models in predicting toxicity observed in individual studies. Finally, the performance of the models was evaluated using an independent data set.

5.2. MODEL DEVELOPMENT

Most evaluations of the effects of mixtures on aquatic toxicity endpoints such as survival and growth have focused on two empirical models of noninteractive joint action: concentration addition and response addition (Broderius, 1991). Concentration addition, which is also referred to as “simple similar action,” assumes that contaminants act independently but by a similar mode of action. Toxic unit models, which are a specialized case of concentration addition, have been applied to the assessment of the toxicity of PAH mixtures in sediment (Swartz et al., 1995; Di Toro et al., 2000; Lee et al., 2001), but they are unlikely to be applicable to complex mixtures of contaminants commonly found in the environment that have different modes of toxic action. In

response addition, or independent action, which is expected to apply to cases where contaminants have a different mode of action, toxicity would be predicted only when one or more contaminants exceeds its toxicity threshold.

We explored three alternative approaches for combining the individual chemical models into a single explanatory variable that could be used to estimate the probability of observing toxicity in a given sample:

- the maximum probability of observing toxicity for a sample, taken from the set of probabilities calculated for each individual chemical in the sample (P_Max),
- the mean probability of observing toxicity for a sample, based on the set of probabilities calculated for each individual chemical in the sample (P_Avg), and
- the product of the probabilities of surviving exposure to all individual chemicals in the sample (P_Prod), calculated by multiplying the values of one minus the probability of observing toxicity for each chemical in the sample and then subtracting the resulting product from one.

All three approaches can be considered similar to response-addition models. A model based on P_Max would predict toxicity on the basis of the individual chemical model with the highest probability of toxicity. This approach would be expected to be most effective in predicting toxicity when the degree of correlation between responses to individual chemicals is high. The P_Prod and P_Avg approaches include all chemicals in the estimate of toxicity. Models using these explanatory variables would be expected to be most effective when the degree of correlation between responses to individual chemicals is intermediate and low, respectively (U.S. EPA, 2000).

A possible disadvantage of P_Avg as an explanatory variable is that it gives more weight to chemical classes that have many individual chemicals that tend to co-occur. For example, P_Avg incorporates the output from models for 22 individual PAHs. Because individual PAHs are likely to co-occur in environmental samples, P_Avg may be influenced more by the concentrations of PAHs than by the concentrations of other chemicals. In addition, the P_Avg approach may appear to reduce the influence of chemicals associated with high probabilities of toxicity by averaging them with chemicals having low probabilities of toxicity.

P_Max, P_Avg, and P_Prod were all developed using the simple chemical parameter estimates shown in Table 10, except for PCBs. As discussed in Section 4.5, we discovered an

error in PCB units for 15 samples after the models had been fit and evaluated. This change resulted in a minor change in the PCB model parameter estimates. However, because the effects of the correction on the multiple-chemical models were extremely small (as discussed further in Section 5.4.4), the multiple-chemical models were not changed.

Preliminary plots of P_Avg and P_Max versus the percent of samples that were toxic in sequential intervals of increasing contamination appeared nearly linear, so a multiple linear regression modeling approach was pursued. The dependent variable for the models was the frequency of toxicity observed within the sequential intervals of increasing contamination (i.e., probability intervals). The probability intervals were defined by first sorting the samples by the predicted probability (i.e., P_Avg and P_Max) and then combining samples into groups containing 50 unique predicted probability values. The median predicted probability associated with each interval was used as the explanatory variable. Because the dependent and explanatory variables both depend on how samples are binned, we evaluated the sensitivity of the models to different binning approaches (Appendix E). The results of this analysis indicated that, above a minimum number of regression samples (i.e., the number of bins) and samples within the bins, the regression model coefficients showed little change with variations in binning approach.

We initially considered a fourth alternative, the sum of the probabilities of observing toxicity for each individual chemical in the sample (P_Sum model). This approach was eliminated from further consideration because the resulting range of predicted probabilities had no upper bound, making the results very sensitive to the number of chemicals measured in a particular sample.

We developed the multiple-chemical models using all samples with matching chemistry and toxicity (i.e., no additional data screening procedures were employed). To minimize the potential impact of samples with partial chemistry, only samples with measured values for at least 10 chemicals were included in the data set used to derive the multiple-chemical models. We evaluated the alternative approaches by examining goodness of fit (as indicated by R-squared values), concentration interval plots (described in Section 4.2.2), and the agreement between the observed proportion of toxic samples and that predicted using the multiple-chemical models (described in Section 4.3.2.2).

A multivariate approach to the problem uses all chemical data simultaneously to predict the probability of a toxic test result was explored in a companion effort (Smith et al., 2003). The multivariate approach requires complete data for all chemicals included in the model. The single

explanatory variable approach developed in this chapter is potentially applicable to a wider array of chemistry data.

5.3. MODEL RESULTS

Probability density functions of P_Avg, P_Max, and P_Prod values are shown in Figure 15. The probabilities generated using the P_Avg and P_Max approaches were fairly evenly distributed across the range of probabilities. The distribution of P_Prod values was so heavily skewed toward higher values that P_Prod would have limited ability to discriminate among moderately to highly contaminated samples. We eliminated this model from further consideration.

The regression models generated using P_Avg and P_Max are shown in Figure 16. A quadratic term was significant for the P_Max model, indicating that the relationship is curvilinear. Both models explained a large amount of the variation in the observed frequency of toxic samples in the probability intervals (R-squared values were 0.89 and 0.93 for P_Avg and P_Max models, respectively). Regression diagnostic plots (e.g., Cooks distances, residual distributions [not shown]) revealed no overly influential values or severely nonnormal residual distributions.

We compared the differences between the frequency of observed toxicity with that predicted by the P_Avg and P_Max models within quartiles of the predicted probability (Figure 17). All samples in the database were used in the comparison (including samples with fewer than 10 chemicals measured). When applied to individual samples, the P_Avg model could yield predictions of probability of toxicity greater than 1, so the predicted probability was capped at 1. The mean predicted probability of toxicity within probability quartiles closely matched the observed proportion of toxic samples within the same probability quartiles, demonstrating the overall reliability of both the P_Max and the P_Avg models within the database that was used to derive the model.

The P_Max and P_Avg models explained 92% and 88%, respectively, of the variation in the frequencies of toxicity observed in the probability intervals derived from the entire database (Figure 18). The P_Max model adjusts for the difference between the maximum probability from the individual chemical models and the observed proportion of toxic samples within the same probability interval. For example, for a maximum probability of 1 from the individual chemical models (x-axis), the observed proportion of toxic samples (and the predicted

probability from the P_Max model) is 0.84 (Figure 16). The data used to derive the P_Avg model show the opposite situation, where the mean probability value from all of the individual chemical models is somewhat lower than the corresponding observed proportion of toxic samples. Thus, mean probabilities of 0.5 and 0.75 correspond to proportions of toxic samples of 0.7 and 0.9, respectively; the P_Avg adjusts the final estimate of toxicity accordingly (Figure 16).

5.4. USING THE MULTIPLE-CHEMICAL MODEL AS AN ANALYTICAL FRAMEWORK

The multiple-chemical models provide a consistent analytical framework that can be used to evaluate several issues that are relevant to evaluating sediment toxicity. In the following sections, we demonstrate how the models can be used to

- evaluate the relationship between the probability of observing a toxic effect and the magnitude of toxicity,
- identify the chemicals that serve as the most (or least) effective surrogates for toxicity,
- evaluate the effect of reducing the number of individual chemical models combined into the multiple-chemical model,
- evaluate the performance of the models in predicting toxicity of the two amphipod species, and
- evaluate the performance of the models in predicting toxicity observed in specific studies.

We used the P_Max model in the evaluations discussed in the remainder of this report because it explained a slightly higher amount of variation, it is less influenced by the number of chemicals analyzed in a sample, and we could reduce the number of analytical permutations.

5.4.1. Relationship Between Probability of Observing a Toxic Effect and Magnitude of Toxicity

The magnitude of the effect (decreased survival) in the amphipod test increased as the probability of toxicity increased (Figures 19 and 20). Toxic samples with a probability of

toxicity less than or equal to 0.25 had an average control-adjusted survival of greater than 75%, whereas samples with a probability of toxicity greater than 0.75 had an average control-adjusted survival of less than 50%. Figure 20 shows the strong relationship between the probability of toxicity predicted by the P_Max model and control-adjusted survival. This demonstrates that samples that are estimated to have the highest probability of toxicity are also likely to cause a high magnitude of mortality.

5.4.2. Chemicals that Serve as the Most Effective Surrogates for Toxicity

The P_Max model is based on the individual chemical that has the highest probability of toxicity. For approximately 70% of the samples, individual chemical regression models for metals produced the maximum probability used in the P_Max model (Table 29). This should not be construed to imply that metals were causing toxicity in these samples. It does indicate, however, that metals appear to be a good predictor of toxicity in field-collected samples where mixtures of contaminants are likely to be present.

The influence of different chemical classes on the model was explored by excluding the individual chemical models for PAHs, metals, and pesticides and PCBs. The models produced by excluding the different chemical classes are shown in Figures 21, 22, and 23, (PAHs, metals, and pesticides and PCBs, respectively). The difference between the models was most evident when metals were excluded, which is not surprising, as 70% of the samples had a metal associated with the maximum probability for that sample. Still, there were only minor changes in the models. This is consistent with the interpretation that the chemicals associated with the P_Max values are serving as surrogates for the overall degree of contamination present in a sample.

5.4.3. Effect of High Levels of Several Chemicals

The number of chemicals in a sample that have a high probability of toxicity according to the individual chemical models (e.g., $p>0.75$) makes a difference in how well the model predictions match the observed proportion of the samples that are toxic (Figure 24). As shown, when only one chemical in a sample has a probability of toxicity greater than 0.75, the P_Max model tends to overestimate the incidence of observed toxicity. The P_Max model slightly underestimated the frequency of toxicity observed in samples that contained two or more chemicals with a probability of toxicity greater than 0.75. The degree of underestimation,

however, remained about the same even for samples containing many chemicals with probabilities greater than 0.75.

5.4.4. Sensitivity of P_Max Model to Individual Chemical Models

The P_Max model described above was developed by including all 37 chemicals with chemical-specific models with normalized chi-square values greater than 0.15. We investigated model sensitivity to the inclusion criterion. In particular, we hypothesized that a more stringent inclusion criterion (i.e., excluding chemicals with poorer fits) would improve the multiple-chemical models. The P_Max model fit using a normalized chi-square criterion of 0.27, which represented the average normalized chi-square value for acceptable models, included 19 chemicals, contained only the linear term, and explained slightly more variation ($R^2 = 0.94$, Figure 25) than the original P_Max models (Figure 16).

5.4.4.1. *Effect of PCB Model Correction*

We examined the influence of the updated PCB model on the P_Max model. As discussed in Section 4.6, the correction in PCB units for 15 samples resulted in a model sufficiently different to cause us to update the PCB model parameters reported in Table 1. However, the effect of updating the PCB model on the P_Max model results was extremely small. The model parameters changed slightly, to 0.11 (intercept), 0.34 (linear term), and 0.39 (quadratic term) from 0.11, 0.33, and 0.4, respectively. The maximum difference in predicted probability of toxicity for an individual sample was 0.0025. Because these differences were so small, we elected to continue using the original P_Max model in the evaluations discussed in this report.

5.4.5. Predictions of P_Max Model for Individual Species

The P_Max models also provide a framework for re-examining the species-specific differences in model performance. Figures 26 and 27 plot the probability of toxicity predicted from the nationwide P_Max model against three species-specific variables: (1) the proportion of observed toxicity based on the Sig Only classification, (2) control-adjusted survival, and (3) the proportion of toxicity based on MSD classification. There was a strong relationship between the predicted probability of toxicity and the three variables. The relationship between predicted and observed proportion toxicity was slightly stronger using the MSD classification for both species.

The relationship between predicted and observed toxicity was very different for the two species. In general, *R. abronius* showed a higher response rate than did *A. abdita* at similar predicted probabilities. This difference is consistent with that discussed for the individual chemical models in Section 4.5 and may be explained by inherent differences in sensitivity between the two species or a greater sensitivity of *R. abronius* to nonchemical factors, or the database for *R. abronius* may contain a disproportionate amount of data from areas with a high degree of contamination from multiple chemicals.

As shown in Figures 26 and 27, most of the *A. abdita* data fall between predicted probabilities of 0 and 0.5, whereas the *R. abronius* data fall mostly between 0.25 and 0.75. Although we cannot completely resolve the cause of differences in species responses, differences in sediment chemistry cannot be discounted as a contributing factor. Although the variability in the modeled response of *R. abronius* was higher than that in the *A. abdita* response, the model still served as an effective indicator of the observed proportion of toxic samples. Combining data from tests that used either species enables the development of models that encompass more areas of the country and a broader range of sediment chemistry. The strong relationships of the individual species' responses with the nationwide model affords confidence that the combined-species model provides a common framework that is applicable to both species.

5.4.6. Predictions of P_Max Model for Individual Studies

The multiple-chemical models are a function of the covariation among chemicals on a nationwide basis. Accordingly, the nationwide model may not accurately predict observed toxicity at a particular site if the chemical mixture is very different from the average across the nation or if site-specific factors greatly influence the degree of toxicity (e.g., by increasing or decreasing chemical bioavailability).

To examine the application of the nationwide model to individual studies, we first examined the performance of models in predicting observed toxicity for individual studies within the database that are represented by more than 20 samples (Table 30). We compared the differences between the frequency of observed toxicity with that predicted by the P_Max models within quartiles of the predicted probability. Of the 39 studies with sufficient data to evaluate, 14 had observed frequency of toxicity within 20% of that predicted in all quartiles having data. Nine studies had observed frequency of toxicity that differed by more than 20% in all quartiles having data. The remainder of the studies had mixed performances.

We conducted a more in-depth evaluation of the performance of the models for predicting toxicity in specific studies by comparing predicted and observed toxicity within smaller intervals. To have sufficient data for this evaluation, we combined data from several studies within a common geographic region (Table 31). Figures 28 through 32 plot the probability of toxicity predicted from the nationwide P_Max model against three site-specific variables: (1) the proportion of observed toxicity based on the Sig Only classification, (2) control-adjusted survival, and (3) the proportion of toxicity based on MSD classification.

Figures 28 through 32 show that the nationwide P_Max model provides a useful basis for evaluating toxicity test results for individual regions included in the database. There were significant relationships between the P_Max model predictions and observed proportion of toxicity, but the relationships were not one-to-one, indicating that regions differed from the nationwide model by intercept or slope or both (left-hand graphs in Figures 28 through 32). The differences were not consistent across the regions. The probability of toxicity was strongly related to percent control-adjusted survival for all regions except California (center graphs in Figures 28 through 32). For the California data (right-hand graph in Figure 32), comparing the P_Max model (which was derived using the Sig Only toxicity classification) with observations classified using the MSD approach greatly improved the strength of the relationship. For the other regions, using the MSD approach only slightly clarified the relationship, mostly at low probability of toxicity.

In theory, model predictions also could be compared with observations from studies with fewer samples (e.g., from a smaller area). Comparisons are most useful if the observed data represent a large range of predicted probabilities of toxicity (left-hand graph in Figure 33). Studies with observations from a small range of predicted probabilities may show no apparent relationship (right-hand graph in Figure 33). A consistently high proportion of toxic samples (left-hand graph in Figure 34) may show a stronger relationship with the nationwide model by classifying toxic samples using an MSD approach (right-hand graph in Figure 34).

5.5. APPLICATION OF THE MODELS TO AN INDEPENDENT DATA SET

Application of the models to independent data (data not used in model derivation) was an important step in evaluating the models. The P_Max model was applied to a small independent data set consisting of three studies from the Calcasieu Estuary (Louisiana) that had matching sediment chemistry and toxicity data for *A. abdita* (Redmond et al., 1996; unpublished data set

provided electronically by P. Crocker, U.S. EPA, Region 6, Dallas, TX; MacDonald et al., 2001). The Calcasieu Estuary is a highly contaminated industrial waterbody that is included on the National Priorities List of hazardous waste sites. The Calcasieu Estuary data were not included in the database used to derive the models. The data set contains 170 matched chemistry and toxicity test results using *A. abdita* and represents a wide range of contaminant concentrations.

There was a strong linear relationship between the predicted probability of toxicity and the proportion of toxicity observed, based on both the Sig Only and the MSD classifications (left-hand and right-hand graphs in Figure 35). In general, the frequency of toxicity observed in the Calcasieu Estuary samples was greater than that predicted by the P_Max model. The predicted probability of toxicity showed a strong relationship with control-adjusted survival (center graph in Figure 35).

5.6. SUMMARY AND CONCLUSIONS

We developed multiple-chemical models to combine the individual chemical models into a single prediction of the toxicity of sediment samples with known sediment chemistry. Two approaches for combining the individual chemical models (the P_Max and P_Avg) produced models that accurately predicted the frequency of toxicity to amphipods observed in the database:

- P_Max is the maximum probability of observing toxicity for a sample, taken from the set of probabilities calculated for each individual chemical in the sample, and
- P_Avg is the mean probability of observing toxicity for a sample, based on the set of probabilities calculated for each individual chemical in the sample.

The P_Max model explained 92% of the variation in the frequency of toxic samples observed in probability intervals constructed using the entire data set. The R-squared value for models using the P_Avg value was slightly lower.

We used the P_Max model to evaluate several issues: the relationship between the probability of observing a toxic effect and the magnitude of toxicity, the identification of chemicals most influential in model performance, the effect of reducing the number of individual chemicals models used in developing the multiple-chemical models, the performance of the

models in predicting toxicity of the two amphipod species, and the performance of the models in predicting toxicity observed in individual studies and regional areas.

The magnitude of the effect (decreased survival) in the amphipod test increased as the probability of toxicity increased, demonstrating that samples that are estimated to have the highest probability of toxicity are also likely to be extremely toxic.

For approximately 70% of the samples, individual chemical regression models for metals produced the maximum probability used in the P_Max model. This should not be construed to imply that metals were causing toxicity in these samples, only that metals appear to be a good predictor of toxicity in field-collected samples. Indeed, removing metals (or other entire chemical classes) from the suite of individual chemical models used to generate the P_Max model resulted in only minor changes in the model and model fit.

The P_Max model also provides a useful basis for comparing the species-specific differences in model performance. As observed with the individual chemical models, the observed toxicity was almost always less than predicted for *A. abdita* and greater than predicted for *R. abronius*. However, the distribution of the data differs substantially for the two species across the probability of toxicity predicted by the P_Max model. Nevertheless, the observed proportion of toxicity observed in both species was strongly related to the nationwide model, affording confidence that the combined-species P_Max model provides a common framework applicable to both species.

Although the P_Max model reliably predicted toxicity for the entire database used to derive the models, it may not accurately predict observed toxicity at a particular site if the chemical mixture is very different from the distribution of mixtures in the nationwide database or if site-specific factors influence the degree of toxicity (e.g., bioavailability). To address this issue, we examined the performance of models in predicting observed toxicity for individual studies within the database. On a study-by-study basis, there was mixed agreement between the frequency of observed toxicity and that predicted by the P_Max model. The mixed performance emphasizes the importance of calibrating the national models with site-specific data (discussed further in Section 6.4). Still, the nationwide P_Max model provided a useful, common basis for evaluating toxicity test results for individual sites included in the database.

Application of the model to regional subsets of the database used to derive the model demonstrated significant relationships between the P_Max model predictions and both observed proportion toxicity and percent control-adjusted survival. There was also a strong relationship

between predicted toxicity and observed toxicity in the Calcasieu Estuary, a data set not included in the original derivation of the models.

6. APPLICATIONS OF MODELS

6.1. INTRODUCTION

This chapter discusses the application of the single-chemical and the multiple-chemical LRMs to problems that are frequently encountered by risk assessors. Section 6.2 discusses the relevance of the models to evaluating other toxicity endpoints and test systems, including those commonly used in freshwater systems. Section 6.3 discusses the models in the context of other empirical approaches for evaluating risks associated with individual chemicals found in sediments. Finally, in section 6.4, we discuss how these models might best be used to evaluate the risks of sediment contamination at specific sites or regions.

6.2. APPLICATION OF THE P_MAX MODEL TO DATA FOR OTHER ENDPOINTS

The SEDTOX02 database contains matched sediment chemistry and toxicity test data for endpoints other than the two marine amphipods. The data available for two sea urchins (*Strongylocentrotus purpuratus* and *Arbacia punctulata*) are presented in Tables 3 through 5. Data for the freshwater amphipod (*Hyalella azteca*) and midges (*Chironomus tentans* and *Chironomus riparius*) are shown in Tables 6 through 9. These data proved to be insufficient to support models for many chemicals of concern for these test endpoints; instead, they were used to evaluate the application of the marine amphipod models to these other endpoints.

Models for the marine amphipods were applied to the sea urchin embryological development and fertilization data sets. The data for the two urchin species were combined for each of the development (*A. punctulata* 48h and *S. purpuratus* 96h) and fertilization (1h) endpoints. The relationship between the average predicted probability of toxicity from the P_Max model and the proportion of toxicity observed within probability intervals is shown in Figure 36. Note that there is a very weak relationship between the model and both urchin response variables. However, if the data for the two urchin species are evaluated separately, *A. punctulata* shows a strong relationship for the development endpoint (Figure 37) and a weaker relationship for the fertilization endpoint (Figure 38). The results for *S. purpuratus* demonstrated no relationship between the predicted probability of toxicity and either the development or the fertilization endpoint (Figures 39 and 40).

Models for the marine amphipods were applied to freshwater data sets for three freshwater endpoints, *H. azteca* 10–14-day survival, *Chironomus* spp. 10–14-day survival, and

H. azteca 28-day growth and survival. The proportion of toxicity observed using *H. azteca* (Figure 41) and *Chironomus* (Figure 42) 10–14-day survival tests was less than that predicted by the P_Max model. Below the mean predicted probability of 0.75, observed proportion of toxic samples was consistently low (<25%). This relationship suggested a threshold, so we used a spline model to describe the relationships (General Additive Model with 2 degrees of freedom [Insightful Corp., 2001]). The spline model explained 51% and 64% of the variability seen in the *H. azteca* and *Chironomus* data, respectively. One possible explanation for this type of threshold relationship is that the marine amphipod toxicity tests respond at lower concentrations of chemicals in sediments than do the freshwater short-term toxicity tests.

The *H. azteca* 28-day growth and survival endpoint showed a much stronger relationship between the predicted probability based on the P_Max model and the observed proportion toxic in a relatively small data set (n = 126) (Figure 43). There is a strong relationship ($R^2 = 0.95$) between the probability of toxicity based on the P_Max model versus the *H. azteca* 28-day growth and survival endpoint (Figure 43).

In summary, the results of comparing other endpoints with the P_Max model predictions suggest that the marine amphipod models are useful for predicting sea urchin response on the basis of embryological development or fertilization tests for *A. punctulata* but not for *S. purpuratus*. The marine amphipod models also have utility for predicting the response of freshwater amphipods, particularly the 28-day *H. azteca* growth and survival endpoint. There is the potential for developing endpoint-specific models as more data are acquired.

6.3. USING LRMs TO EVALUATE EXISTING EMPIRICAL GUIDELINES

Hazardous waste site evaluations often involve the collection of substantial quantities of sediment chemistry data, and these data are frequently used to support screening-level ecological risk assessments. To evaluate such data, sediment assessors often use numerical sediment quality guidelines (SQGs) such as threshold effect levels (TELs) and probable effect levels (PELs), effect range low (ERL) and effect range median (ERM), and apparent effect thresholds (AETs) (Gries and Waldow, 1996; Long and Morgan, 1991; Long and MacDonald, 1992; Smith et al., 1996; Ingersoll et al., 1996, 2001, 2002; MacDonald et al., 1996, 2001).

Although derivation methods for the different SQGs are well described and are consistent for all the chemicals within a given type of SQG, there is no straightforward method that enables the user to either evaluate the degree to which individual SQGs meet their objectives or compare

the risk levels associated with different SQGs. The logistic model approach provides a way to put the individual SQG values into perspective by estimating the probability of observing toxicity to amphipods at the chemical concentrations defined by the SQGs. Examples are shown in Table 32 for three commonly used sets of SQGs that represent a range of threshold values: TELs and PELs, ERLs and ERMs, and AETs. ERLs and TELs represent chemical concentrations below which toxicity would be expected to occur infrequently (<25%) (Long and Morgan, 1991; MacDonald et al., 1996), whereas effects are expected to be frequently observed at concentrations exceeding PEL and ERM concentrations. In contrast, endpoint-specific AET values represent concentrations above which toxicity is always expected for that endpoint.

The results are generally consistent with the narrative intent of the SQGs for most of the chemicals for which SQGs had been derived. The highest probability of observing toxicity to amphipods was noted for the amphipod AETs, with an estimated proportion of toxic samples ranging from 45 to 99% and a median value of 90%. The predicted probability of observing toxicity was lower for the PEL and ERM values, with median values of 55% and 72%, respectively. At concentrations corresponding to the TELs, predicted probabilities of observing sediment toxicity ranged from 10 to 41% (depending on the chemical under consideration), with the probability of toxicity below 25% for 24 of the 27 chemicals considered (Table 32). The probability of observing sediment toxicity was a little higher at the ERL concentrations (ranging from 11 to 47%), with a median value of 33%. The probabilities may be higher than expected for ERLs and TELs because they are calculated for individual chemicals. In practice, these guidelines are most appropriately applied jointly; that is, a sediment sample has a low probability of causing toxicity if all chemicals are below the ERL or TEL.

The LRMs help users select the sediment effect concentrations that most directly meet the needs of their specific application. T10, T15, or T20 values could be calculated and used to identify concentrations for individual contaminants that are likely to be associated with a relatively low incidence of sediment toxicity (10, 15, or 20%, respectively). Such point estimates of minimal-effect concentrations might be used in a screening assessment to identify sediments that are relatively uncontaminated and have a low probability of sediment toxicity. As discussed above, this evaluation would best be conducted by evaluating all chemicals simultaneously. For example, of the samples having all chemicals at concentrations below their respective T20 value, 19% were toxic, based on the Sig Only classification.

Similarly, contaminant concentrations for which there is a high probability of observing adverse effects could be estimated by calculating T70, T80, or T90 values. These higher point estimates could be used to identify sediments that are highly likely to be toxic to amphipods and to have a greater magnitude of effect (i.e., higher percent mortality). The Tp values can be used in much the same way as other sediment guidelines, with the difference that the Tp value is associated with a specific probability of observing toxicity and can include confidence bounds on the sediment concentrations associated with a given Tp value.

Individual chemical SQGs are useful for identifying thresholds below which sediment toxicity is unlikely to be observed and above which sediment toxicity is likely to occur. However, it is difficult to determine the extent to which risk increases with the magnitude of exceedance of an SQG. In addition, individual SQGs must be combined to address risks associated with mixtures. To address both of these issues, practitioners have applied hazard quotient approaches, which sum the ratios of the measured concentration of each contaminant to its corresponding toxicity threshold. Several investigators have applied mean SQG quotients to evaluate mixtures of contaminants in field-collected sediment samples (Long et al., 1998; MacDonald et al., 2000; Fairey et al., 2001, Ingersoll et al., 2001). Such evaluations are based on an assumption that concentration-response relationships for each chemical are similar. The logistic regression modeling approach avoids this assumption by fitting concentration-response relationships separately for each chemical and then standardizing the response variables to values between 0 and 1.

Another approach to evaluating risks associated with sediment chemistry is the equilibrium partitioning (EqP) approach. EqP links sediment chemistry values with biological effects by combining the results of controlled laboratory tests using manipulated concentrations of chemicals with theory on the factors controlling bioavailability (Di Toro et al., 1991, 2000). EqP values represent concentrations of individual chemicals below which effects would not be expected to occur. Individual EqP values cannot be directly compared with SQGs derived using field-collected data (including Tp values) because the EqP values reflect the toxicity attributable to individual chemicals, whereas the SQGs reflect the toxicity of the mixture. For example, the Tp values associated with EqP values for individual PAHs are very high (Table 33), a result that is consistent with a high degree of covariation among PAHs in sediments.

The EqP approach can be used to estimate the toxicity associated with a mixture of PAHs using a toxic unit approach (Di Toro et al., 2000). Again, any comparison with empirical

approaches is imperfect because the latter reflect the contribution of chemicals other than PAHs to toxicity. The PAH toxic units were calculated on the basis of EPA's Final Chronic Value following the methods described in EPA's National Sediment Quality Survey (U.S. EPA, 2004). Most of the samples in the database had PAH toxic units less than 1. Both the P_Max values and the proportion of toxic samples in the database increased with increasing PAH toxic units (Table 34). Below a PAH toxic unit of 1, 37% of the samples were toxic. At PAH toxic units of 10 and greater, 81% of the samples were toxic.

In summary, the LRMs provide a useful framework for evaluating the degree of risk associated with commonly used SQGs. Logistic models have several advantages over current SQGs approaches: (a) they present risk on a continuous quantitative scale rather than by defining discrete categories based on threshold values, (b) the continuous estimates of risk allow users to match the degree of risk with their objectives, (c) they express risk on a common scale of 0 to 1 across all chemicals, and (d) they provide a more direct avenue for assessing risk of multiple chemicals.

6.4. APPLICATION OF MODELS TO EVALUATIONS OF SITE-SPECIFIC OR REGIONAL DATA

The models described in this report were derived from a large database of matching sediment chemistry and toxicity that included data from many different coastal areas of North America and many different chemical gradients. The models make predictions based on the central tendency of the relationship between sediment chemistry and a toxic outcome across this broad gradient; therefore, the models can be used to support screening-level assessments that roughly rank or prioritize samples on the basis of sediment chemistry, particularly when concentrations of chemicals vary over a wide range. For example, the National Sediment Quality Survey (U.S. EPA, 2004) used these models to help classify locations into three tiers reflecting the probability of adverse effects.

However, because the models do not consider potential differences in bioavailability, test species, or site-specific mixtures of chemicals, the probability of toxicity may be over- or underestimated for some locations. As discussed in Section 5.4.6, the differences between quantitative model predictions and site- or region-specific observations can be substantial. Although the derivation of site-specific models may be desirable, data from an individual site are rarely sufficient to support model derivation. Rather than deriving site- or regional-specific

models with a small data set, we recommend using site-specific data to determine how well the nationwide models fit the local situation.

The evaluation of the independent data set from the Calcasieu Estuary provides an example. By comparing the proportion of toxic samples or the mean control-adjusted survival with the mean predicted probability of toxicity within discrete probability ranges (e.g., probability quartiles, as shown in Table 30, or probability intervals if sufficient data are available [Figure 35]), the performance of the models with data from the site can be evaluated. In the best case, the observed proportion of toxic samples will closely match that predicted by the model. In this case, the models can be applied with confidence to samples having only sediment chemistry. If the relationship between observed and predicted proportions of toxic samples is strong but different, the relationship can still be used to predict site-specific toxicity. In addition, a strong relationship between the percent survival and the predicted probability of toxicity provides important information on the magnitude of the response. If the relationship between predicted and observed proportions of toxic samples or percent survival is weak, then the models are likely to be less useful for the site.

Although there are no universal criteria for determining whether a model fit is acceptable, standard regression techniques and diagnostic plots can be used to identify whether specific values are overly influencing the relationship or the assumptions of linear regression are not met (Neter et al., 1996).

To best compare model predictions with site-specific observations, we recommend collecting matching chemistry and toxicity testing results from samples that have a wide range of chemical gradients and predicted probabilities of toxicity. It is particularly important to collect sufficient data from areas of high and low probability of toxicity to better define the relationships. Therefore, LRMs have great utility in helping to design sampling programs.

The models can also suggest issues that require further investigation. If the models predict a higher proportion of toxic samples than the proportion observed (false positives), then issues related to bioavailability may be investigated further. The individual chemical models could be used to determine whether specific chemical models are associated with the high false positive rate. If toxicity occurs at a much higher frequency than predicted (false negatives), then it may be important to consider chemicals not accounted for (e.g., no models available) or issues related to the sediment matrix (e.g., grain size effects).

6.5. SUMMARY AND CONCLUSIONS

In this chapter we discussed the application of the single-chemical and multiple-chemical LRMs to problems that are frequently encountered by risk assessors. We evaluated the relationship between model predictions and the results of other toxicity endpoints, including those commonly used in freshwater systems. The P_Max model predictions appear to be useful for predicting sea urchin response for *A. punctulata*, based on development or fertilization tests, but not *S. purpuratus*. The models are also useful for predicting the response of freshwater amphipods, particularly the 28-day *H. azteca* growth and survival endpoint. There is the potential for developing endpoint-specific models as more data are acquired.

The LRMs provide a useful framework for evaluating the degree of risk associated with commonly used SQGs. The probabilities of toxicity associated with SQG threshold values are generally consistent with their narrative intent. However, LRMs have several advantages over current guideline approaches. They present risk on a continuous quantitative scale rather than by defining discrete categories based on threshold values. The continuous estimates of risk allow users to match the degree of risk with their objectives, and risk is expressed on a standardized scale of 0 to 1 across all chemicals. The logistic models provide a direct yet flexible avenue for assessing risk of multiple chemicals and provide the basis for more quantitatively evaluating the reliability and fit of these models to observations.

Finally, we discussed how these models might best be used to evaluate the risks of sediment contamination at specific sites or regions. The LRM approach can be used to conduct screening-level assessments that roughly classify or prioritize samples on the basis of sediment chemistry. Because the models do not consider potential differences in bioavailability or exposure, the probability of toxicity may be over- or underestimated for some locations. For applications that require a greater degree of accuracy (e.g., remediation decisions), we recommend first evaluating how well the models fit the local situation by collecting a test set of matching sediment chemistry and toxicity test data. The LRMs can be used to design effective test sampling programs, and they can also suggest issues requiring further investigation (e.g., bioavailability). The LRMs may be most useful for classifying samples into broad categories of concern on the basis of sediment chemistry.

7. CONCLUSIONS AND RECOMMENDATIONS

7.1. CONCLUSIONS

A large database of matching sediment chemistry and toxicity data was carefully evaluated and assembled. The database encompasses many different contaminant gradients from a wide variety of habitats in coastal North America. Using this database, LRMs were developed for 37 individual chemicals that describe relationships between the concentrations of sediment-associated contaminants and acute toxicity to two commonly tested species of marine amphipods: *Rhepoxynius abronius* and *Ampelisca abdita*.

The chemical-specific models that were derived in this investigation provide a basis for estimating the proportion of samples expected to be toxic over a wide range of contaminant concentrations for 37 individual contaminants. As such, these models help users select the sediment effect concentrations that most directly meet the needs of their specific application. For example, T10, T15, or T20 values could be calculated and used to identify concentrations for individual contaminants that are likely to be associated with a relatively low incidence of sediment toxicity (10, 15, or 20%, respectively). Such point estimates of minimal-effect concentrations might be used in a screening assessment to identify sediments that are relatively uncontaminated and have a low probability of sediment toxicity. Similarly, contaminant concentrations for which there is a high probability of observing adverse effects could be estimated by calculating T70, T80, or T90 values. These higher point estimates could be used to identify sediments that are highly likely to be toxic to amphipods and have a greater magnitude of effect (i.e., higher percent mortality).

The Tp values can be used in much the same way as other sediment guidelines, with the difference that the Tp value is associated with a specific probability of observing toxicity and an estimate of variance based on the fit of the model. Although the LRMs do not represent dose-response relationships for individual chemicals, they can be considered indicators of toxicity based on field-collected sediment chemical mixtures.

Because the individual models were derived from field-collected sediments that include mixtures of contaminants rather than from individual dose-response relationships, to some extent they incorporate the overall toxicity of the mixtures. Combining the individual model results into a single explanatory variable provided a way to estimate the probability that a particular

sample will be toxic. Two combined explanatory values (P_Max and P_Avg) accurately predicted the frequency of toxicity to amphipods observed in the database:

- P_Max is the maximum probability of observing toxicity, taken from the set of probabilities calculated for each individual chemical in the sample, and
- P_Avg is the mean probability of observing toxicity, based on the set of probabilities calculated for each individual chemical in the sample.

The logistic regression approach served as a framework for evaluating several issues that form the basis for our recommendations for using the models. We used the single-chemical models to evaluate two approaches for designating samples as toxic: (1) less than 90% survival that was significantly different from negative control samples (Sig Only), and (2) control-normalized survival less than 80% that was significantly different from negative control samples (MSD) (based on analyses by Thursby et al., 1997). Although the Sig Only approach had a greater tendency to underestimate the toxicity observed at low concentrations, this discrepancy may be explained by the presence of other chemicals in the sample. The MSD approach had a greater tendency to overestimate the toxicity observed at higher concentrations.

We also evaluated two approaches for normalizing sediment chemistry: dry-weight normalized and organic carbon normalized. The models based on organic carbon-normalized sediment chemistry had lower goodness-of-fit statistics than the dry-weight-normalized models and larger differences between observed and predicted proportions of toxic samples.

We used the single-chemical models to evaluate three screening criteria used to decide whether a given sample was included in the model data set for an individual chemical: (1) include all samples (unscreened), (2) exclude toxic samples that were less than or equal to the mean of nontoxic samples from the same study (1X screening), and (3) exclude toxic samples that were less than or equal to two times the mean of nontoxic samples from the same study (2X screening). The models developed using 1X screening showed much stronger relationships between chemistry and toxicity than was observed when all samples were included. The 1X screening approach performed slightly better than the 2X screening approach at concentrations above the T80 value. In addition, the 1X approach screened out considerably fewer toxic samples in the model derivation than did the 2X approach, which may prove important in the future development of models for less frequently measured chemicals. We selected the 1X

screening approach to limit the many permutations with further model development and exploration.

The multiple-chemical models were used as a framework for evaluating several additional issues, including the relationship between the probability of observing a toxic effect and the magnitude of toxicity, the identification of chemicals most influential in model performance, the performance of the models in predicting toxicity of the two amphipod species, and the performance of the models in predicting toxicity observed in regional data sets or in individual studies.

The magnitude of the effect (decreased survival) in the amphipod test increased as the probability of toxicity increased, demonstrating that samples that are estimated to have the highest probability of toxicity are also likely to be extremely toxic.

For approximately 70% of the samples, individual chemical regression models for metals produced the maximum probability used in the P_Max model. This should not be construed to imply that metals were causing toxicity in these samples, only that metals appear to be a good predictor of toxicity in field-collected samples. Indeed, removing metals (or other entire chemical classes) from the suite of individual chemical models used to generate the P_Max model resulted in only minor changes in the model and model fit.

Models were developed by combining data from tests that used two species of marine amphipod (*R. abronius* or *A. abdita*) in order to encompass more areas of the country and a broader range of sediment chemistry. The P_Max model provided a useful basis for examining differences in model performance for the two species. The observed toxicity was frequently less than predicted for *A. abdita* and greater than predicted for *R. abronius*. Nevertheless, the observed proportion of toxicity in the data for both species was strongly related to the nationwide model, affording confidence that the combined-species P_Max model provides a common framework applicable to both species.

We examined the performance of models in predicting observed toxicity for individual studies within the database. On a study-by-study basis, there was mixed agreement between the frequency of observed toxicity and that predicted by the P_Max model. The mixed performance suggests that the national models should not be applied to individual studies without first evaluating their performance with matching site-specific toxicity and chemistry data. However, the nationwide P_Max model provided a useful common basis for evaluating toxicity test results for individual sites included in the database. Application of the model to regional subsets of the

database used to derive the model demonstrated significant relationships between the P_Max model predictions and both the observed proportion of toxic samples and percent control-adjusted survival. There was also a strong relationship between the predicted and observed proportion of toxic samples in the Calcasieu Estuary, a data set not included in the original derivation of the models.

7.2. RECOMMENDATIONS

Our analyses and model comparisons resulted in several recommendations. As a starting point for most evaluations, we recommend using the P_Max model, which is derived from the highest predicted probability from any of the individual chemical models. We recommend using the model based on data from all studies (i.e., the nationwide model), on both marine amphipod species, and on 37 chemical-specific models. For the chemical-specific models, we recommend the models that classified samples as toxic on the basis of less than 90% survival that was significantly different from negative control samples (Sig Only) and that screened the data set by excluding toxic samples that were less than or equal to the mean of nontoxic samples from the same study (1X screening). The bases for these recommendations are summarized briefly below.

The P_Max model, which is based on the highest predicted probability from any of the individual chemical models, explained slightly more variation in the data set than did the P_Avg model. The two models provide slightly different insights into sediment toxicity. P_Avg may better reflect the overall degree of contamination, and it is less susceptible to overestimating the probability of toxicity at sites with high concentrations of one chemical. P_Max more accurately predicted toxicity at sites having high concentrations of more than one chemical.

We recommend using the nationwide model that combines data for both species of marine amphipods. Combining data across studies and species represents the fullest range of chemical concentrations and environmental conditions. In addition, the nationwide combined model provides a common framework for comparing site- and species-specific results.

We recommend basing the models on the Sig Only classification so that more subtle changes can be retained, particularly at lower concentrations. It may be valuable to compare the Sig Only nationwide model to site-specific data classified using the MSD approach when test variability obscures the relationship between chemistry and response at lower concentrations.

Finally, the chemical-specific models were greatly improved by using a screened data set that excluded toxic samples that were less than or equal to the mean of nontoxic samples from

the same study. Using a more stringent criterion, such as excluding toxic samples that were less than or equal to two times the mean of nontoxic samples from the same study, resulted in additional improvements for most chemicals. However, the more stringent criterion excluded an average of 70% of the toxic samples, which may limit future development of models for other endpoints, regions, or chemicals that have fewer total samples. We concluded that the small improvements in model fit did not outweigh the associated reduction in sample size.

7.3. APPLICATIONS

The chemical-specific models provide a basis for estimating the probability that a sample will be toxic for 37 individual contaminants over a wide range of contaminant concentrations. In addition, they provide a useful framework for evaluating the degree of risk associated with commonly used SQGs. The probabilities of toxicity associated with SQG threshold values are generally consistent with their narrative intent. However, LRMs have several advantages over current guideline approaches: they present risk on a continuous quantitative scale rather than by defining discrete categories based on threshold values, the continuous estimates of risk allow users to match the degree of risk with their objectives, and they express risk on a common scale of 0 to 1 across all chemicals. The individual chemical models would be expected to underestimate the probability of observing toxicity in samples that are contaminated with many chemicals. For this purpose, we recommend using the multiple-chemical models that combine the individual model results into a single explanatory value for estimating the probability that a sample will be toxic.

The multiple-chemical models provide a useful framework for conducting screening-level assessments that require classifying or prioritizing samples on the basis of sediment chemistry. Because the models do not consider potential differences in bioavailability or exposure, the probability of toxicity may be over- or underestimated for some locations. Before applying the models to a particular site, we recommend first evaluating how well the models fit the local situation by collecting a test set of matching sediment chemistry and toxicity test data. The LRMs can be used to design effective test sampling programs, and they can also suggest issues that require further investigation (e.g., bioavailability). The LRMs should not be considered a complete substitute for direct-effects assessment (e.g., toxicity tests).

We evaluated the relationship between model predictions and the results of other toxicity endpoints, including those commonly used in freshwater systems. The P_Max model predictions

appear to be useful for predicting sea urchin response for *A. punctulata*, based on development or fertilization tests, but not for *S. purpuratus*. The models may also be useful for predicting the response of freshwater amphipods, particularly the 28-day *H. azteca* growth and survival endpoint. There is the potential for developing endpoint-specific models as more data are acquired.

7.4. FUTURE DIRECTIONS

The results of this study suggest many promising avenues for future work. First and foremost, this research provides evidence of the value of combining and standardizing information from many different studies. Future efforts directed at encouraging investigators to add data to the SEDTOX02 database will enable the investigation of additional species, test endpoints, and chemicals. Providing funds and mechanisms to share the database with the scientific community will advance our knowledge of the effects of sediment chemicals on aquatic organisms.

We hope that the release of the SEDTOX02 database will prompt continuing refinement and exploration of modeling approaches linking toxicity test results with sediment chemistry. Possibilities include optimizing the screening approach, evaluating the minimum number of chemicals that need to be included for acceptable performance of the multiple-chemical models, exploring alternative multivariate and logistic modeling approaches, and investigating the reasons for variable model performance in individual studies.

Additional guidance on applying these models to site-specific assessments would increase their use and foster consistent application. Guidance development would be aided by testing the models with additional independent data sets and developing case studies that illustrate their application.

We hope that this study will prompt the collection of more data sets that match toxicity test results with sediment chemistry. Additional data from freshwater systems are especially needed. The most useful data would be broad-scale surveys that include contaminated sites. Sediments should be analyzed for the full range of chemical classes and tested using high-quality, consistent toxicity test methods.

The results of this study suggest that additional work is needed to characterize the bioavailability of chemicals to sediment-dwelling organisms. Contrary to expectation, normalizing chemical concentrations to TOC did not improve model fits. As additional data are

acquired, we may be able to evaluate the effectiveness of acid volatile sulfides as a normalizing factor for metal concentrations.

Finally, the toxicity endpoint modeled in this project serves as a surrogate for valued ecological attributes that are more difficult to test and measure. These include the structure and function of benthic communities, population viability of wildlife that depends on the benthos, and ecosystem functions such as organic matter decomposition and water filtration. Additional research is needed to improve the linkages between these valued endpoints, toxicity test results, and sediment chemistry.

Table 1. Number of samples and percent toxic samples summarized by marine amphipod species and data source[a]

Data source	*Ampelisca abdita*			*Rhepoxynius abronius*		
	Number of samples	Percent toxic		Number of samples	Percent toxic	
		Sig Only	MSD		Sig Only	MSD
EMAP	1203	22.2	9.5	NA	NA	NA
NSTP	649	23.7	15.6	NA	NA	NA
MLML	43	11.6	7	465	72.7	52.3
SEDQUAL	NA	NA	NA	594	63.5	34
BEDS	117	41	30.8	152	36.8	32.2
Total	2012	23.6	12.6	1211	63.7	40.8

[a]Samples were classified as toxic if significantly different from control and less than 90% survival (Sig Only) and if significantly different from control and less than 80% control-adjusted survival (MSD).

BEDS = Biological Effects Database for Sediments, MacDonald Environmental Sciences
EMAP = U.S. Environmental Protection Agency Estuarine Monitoring and Assessment Project
MLML = Moss Landing Marine Laboratory (California)
NSTP = National Oceanic and Atmospheric Administration Status and Trends Program
SEDQUAL = Sediment Quality Information System, Washington State Department of Ecology

NA = no data

Table 2. Distribution of chemical concentrations in sediment samples with matching toxicity data for marine amphipods

Chemical	Number of samples	Percentile		
		10th	50th	90th
Metals (mg/kg dry wt.)				
Antimony	2173	0.2	0.7	2.9
Arsenic	2844	2.2	7.4	19
Cadmium	2958	0.05	0.3	1.9
Chromium	2827	9.1	50	130
Copper	3091	2.6	26	160
Lead	3010	5.5	23	130
Mercury	2788	0.02	0.1	0.8
Nickel	2916	2.4	19	44
Silver	2552	0.03	0.2	1.9
Zinc	3013	16	89	300
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
1-Methylnaphthalene	1677	0.5	6.5	60
1-Methylphenanthrene	1697	0.3	11	130
2,6-Dimethylnaphthalene	1505	0.4	6.3	67
2-Methylnaphthalene	2077	0.8	12	130
Acenaphthene	1795	0.2	7	130
Acenaphthylene	1747	0.2	7	120
Anthracene	2268	0.5	20	410
Benz(a)anthracene	2574	1.2	40	760
Benzo(a)pyrene	2526	1.2	53	910
Benzo(b)fluoranthene	1645	0.9	48	1000
Benzo(g,h,I)perylene	2210	1.1	46	550
Benzo(k)fluoranthene	1691	0.5	26	620
Biphenyl	1507	0.5	6.8	54
Chrysene	2650	1.7	53	1000
Dibenz(a,h)anthracene	1886	0.2	14	170
Fluoranthene	2734	3	81	1400
Fluorene	2011	0.4	11	160
Indeno(1,2,3-c,d)pyrene	2212	0.9	47	600
Naphthalene	2201	1.8	16	220
Perylene	2174	1.5	36	370
Phenanthrene	2688	1.7	44	660
Pyrene	2768	3.2	87	1500
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	1989	2	39	640
Organochlorine pesticides (µg/kg dry wt.)				
Dieldrin	770	0.04	0.8	5.1
p,p'-DDD	1672	0.08	1.8	20
p,p'-DDE	1899	0.08	2.2	59
p,p'-DDT	1176	0.06	1.2	18

Table 3. Number of samples and percent toxic samples summarized by sea urchin species and test endpoint

Urchin species	Development			Fertilization		
	Number of samples	Percent toxic		Number of samples	Percent toxic	
		Sig Only	MSD		Sig Only	MSD
A. punctulata	472	65	64.6	612	40.5	37.1
S. purpuratus	310	61.6	58.1	212	62.3	61.8
Total	782	63.7	62	824	46.1	43.5

Table 4. Distribution of chemical concentrations in sediment samples with matching sea urchin fertilization toxicity data

Chemical	Number of samples	Percentile		
		10th	50th	90th
Metals (mg/kg dry wt.)				
Arsenic	786	2.1	7.6	18
Cadmium	724	0.03	0.2	1.5
Chromium, total	788	5.9	43	110
Copper	788	2	21	150
Lead	781	3.4	23	120
Mercury	709	0.02	0.1	0.6
Nickel	768	1.3	14	28
Silver	622	0.03	0.3	1.5
Zinc	789	9.4	82	290
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
2-Methylnaphthalene	572	0.9	9.3	64
Acenaphthene	492	0.3	5.7	110
Acenaphthylene	533	0.3	6.5	69
Anthracene	642	0.7	16	340
Benz(a)anthracene	718	1.5	33	840
Benzo(a)pyrene	723	1.5	51	1030
Chrysene	749	2	46	1160
Dibenz(a,h)anthracene	595	0.3	12	150
Fluoranthene	770	3.9	81	1660
Fluorene	574	0.5	5.1	110
Naphthalene	604	2	9.8	84
Phenanthrene	736	2	27	600
Pyrene	754	4.7	86	1660
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	622	2	36	290
Organochlorine pesticides (µg/kg dry wt.)				
DDT, total of six isomers	590	0.6	6.9	82
p,p'- DDE	566	0.2	3.2	40

Table 5. Distribution of chemical concentrations in sediment samples with matching sea urchin developmental toxicity data

Chemical	Number of samples	Percentile		
		10th	50th	90th
Metals (mg/kg dry wt.)				
Arsenic	680	2.5	8.4	18
Cadmium	635	0.04	0.2	1.3
Chromium, total	689	6.1	46	140
Copper	689	2.2	25	150
Lead	682	3.2	23	120
Mercury	703	0.02	0.2	0.7
Nickel	671	1.2	15	31
Silver	573	0.03	0.3	1.6
Zinc	694	9.6	94	280
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
2-Methylnaphthalene	533	0.9	9.5	79
Acenaphthene	478	0.3	5.3	91
Acenaphthylene	525	0.3	6.7	60
Anthracene	588	0.6	15	320
Benz(a)anthracene	653	1.3	36	840
Benzo(a)pyrene	655	1.2	54	980
Chrysene	668	1.7	50	1100
Dibenz(a,h)anthracene	578	0.3	13	180
Fluoranthene	683	3.1	81	1450
Fluorene	537	0.4	5.3	100
Naphthalene	576	1.9	9.5	84
Phenanthrene	664	1.6	30	530
Pyrene	662	3	86	1460
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	637	2.2	33	300
Organochlorine pesticides (µg/kg dry wt.)				
DDTS, total of six isomers	528	0.8	7.8	79
p,p'-DDE	520	0.2	3.4	43

Table 6. Endpoint, number of samples, and percent toxic samples for *Chironomus* spp. (*C. tentans* and *C. riparius*) and *Hyalella azteca*[a]

Test species	Endpoint	Number of samples	Percent toxic
Chironomus spp.	10–14-day survival	585	19.8
	10-day growth and survival	286	37.8
H. azteca	10–14-day survival	567	24.2
	28-day survival	125	19.2
	28-day growth and survival	125	38.4

[a]Samples were classified as toxic by the original investigator.

Table 7. Distribution of chemical concentrations in sediment samples with matching toxicity data for the *H. azteca* 10–14-day survival test

Chemical	Number of samples	Percentile		
		10th	50th	90th
Metals (mg/kg dry wt.)				
Arsenic	352	1.7	5.3	27.3
Cadmium	279	0.1	0.8	8.2
Chromium, total	350	9.5	31	111
Copper	396	8	32.7	207
Lead	325	8.4	35.8	362
Mercury	390	0.04	0.28	74.5
Nickel	299	6.1	15	50
Silver	146	0.04	0.2	4.2
Zinc	361	29.5	140	661
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
2-Methylnaphthalene	125	2.7	66.2	21,015
Acenaphthene	248	2.7	80	18,000
Acenaphthylene	201	2	53	1,828
Anthracene	309	8	172	12,000
Benz(a)anthracene	357	16	440	10,000
Benzo(a)pyrene	359	19	467	7,200
Chrysene	376	24	565	11,000
Dibenz(a,h)anthracene	199	3.7	101	1,156
Fluoranthene	406	37.7	950	16,802
Fluorene	276	6.65	111	14,000
Naphthalene	292	9.8	167	12,000
Phenanthrene	378	18	520	15,000
Pyrene	403	40	840	19,691
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	171	8.5	109	1,490
Organochlorine pesticides (µg/kg dry wt.)				
DDT, total of six isomers	110	2.2	17	143
p,p'-DDE	117	2	10	56

Table 8. Distribution of chemical concentrations in sediment samples with matching toxicity data for the *C. tentans* or *C. riparius* 10–14-day survival test

Chemical	Number of samples	Percentile		
		10th	50th	90th
Metals (mg/kg dry wt.)				
Arsenic	392	2.2	7.5	28
Cadmium	266	0.3	1.4	5.9
Chromium, total	399	9	26	140
Copper	413	7	28	130
Lead	406	12	32	110
Mercury	388	0.04	0.2	1
Nickel	380	7.1	16	61
Silver	44	0.05	0.4	1.3
Zinc	403	33	93	540
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
2-Methylnaphthalene	50	9.6	140	16,500
Acenaphthene	162	3	41	1,100
Acenaphthylene	155	4	49	1,230
Anthracene	205	14	120	2,460
Benz(a)anthracene	231	23	350	4,940
Benzo(a)pyrene	251	36	360	4,950
Chrysene	266	35	440	4,100
Dibenz(a,h)anthracene	157	6.9	120	920
Fluoranthene	275	59	640	7,510
Fluorene	185	8.4	69	1,800
Naphthalene	184	10	130	10,000
Phenanthrene	251	35	390	4,600
Pyrene	275	48	620	5,800
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	108	42	390	4,000
Organochlorine pesticides (µg/kg dry wt.)				
DDT, total of six isomers	10	18	76	3,570
p,p'-DDE	21	0.9	25	84

Table 9. Distribution of chemical concentrations in sediment samples with matching toxicity data for the *H. azteca* 28-day growth and survival test

Chemical	Number of samples	Percentile		
		10th	50th	90th
Metals (mg/kg dry wt.)				
Arsenic	57	3.9	24	93
Cadmium	71	0.5	2.4	12
Chromium, total	72	27	61	150
Copper	72	22	86	410
Lead	72	19	91	200
Mercury	68	0.1	0.3	0.8
Nickel	56	8.3	18	50
Silver	39	0.1	1	1
Zinc	72	94	270	940
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
2-Methylnaphthalene	54	16	150	1280
Acenaphthene	26	29	56	250
Acenaphthylene	44	10	46	280
Anthracene	48	10	120	1640
Benz(a)anthracene	88	12	99	690
Benzo(a)pyrene	75	20	190	1230
Chrysene	102	20	120	690
Dibenz(a,h)anthracene	22	10	48	110
Fluoranthene	103	26	190	1450
Fluorene	41	38	77	1530
Naphthalene	69	15	44	1080
Phenanthrene	92	16	95	1400
Pyrene	103	21	170	1480
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	45	250	2740	7700
Organochlorine pesticides (µg/kg dry wt.)				
DDT, total of six isomers	41	0.3	1.9	88
p,p'-DDE	45	0.1	1	84

Table 10. Normalized chi-square values and number of samples for individual chemicals based on Sig Only classification of toxic samples for the screened marine amphipod database (logistic regression model parameters are shown for models having normalized chi-square values greater than 0.15)

Chemical	Number of samples	Intercept (B_0)	Slope (B_1)	Chi-square value/N
Metals (mg/kg dry wt.)				
Antimony	1718	–0.9	2.41	0.25
Arsenic	2336	–4.14	3.17	0.17
Cadmium	2413	–0.34	2.51	0.31
Chromium	2399	–6.44	3	0.2
Copper	2580	–5.79	2.93	0.38
Lead	2481	–5.45	2.77	0.27
Mercury	2296	0.8	2.55	0.32
Nickel	2450	–4.61	2.77	0.18
Selenium	1655			0.07
Silver	2103	–0.11	1.97	0.25
Zinc	2516	–7.98	3.34	0.28
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
1-Methylnaphthalene	1368	–4.14	2.1	0.24
1-Methylphenanthrene	1401	–3.59	1.75	0.28
2,6-Dimethylnaphthalene	1249	–4.05	1.9	0.2
2-Methylnaphthalene	1704	–3.76	1.78	0.25
Acenaphthene	1424	–3.62	1.75	0.33
Acenaphthylene	1447	–2.96	1.38	0.23
Anthracene	1823	–3.66	1.49	0.29
Benz(a)anthracene	2099	–4.2	1.58	0.3
Benzo(a)pyrene	2053	–4.3	1.58	0.3
Benzo(b)fluoranthene	1348	–4.54	1.49	0.27
Benzo(g,h,i)perylene	1818	–4.28	1.59	0.25
Benzo(k)fluoranthene	1376	–4.28	1.57	0.29
Biphenyl	1226	–4.11	2.21	0.26
Chrysene	2126	–4.32	1.54	0.29
Dibenz(a,h)anthracene	1546	–3.63	1.77	0.33
Fluoranthene	2189	–4.46	1.48	0.26
Fluorene	1668	–3.71	1.81	0.32
Indeno(1,2,3-c,d)pyrene	1837	–4.37	1.62	0.27
Naphthalene	1816	–3.78	1.62	0.24
Perylene	1823	–4.68	1.76	0.22
Phenanthrene	2173	–4.46	1.68	0.3
Pyrene	2240	–4.71	1.59	0.29
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	1617	–3.46	1.35	0.27
Organochlorine pesticides (µg/kg dry wt.)				
Dieldrin	633	–1.17	2.56	0.35
Gamma-hexachlorocyclohexane	534			0.098
p,p'-DDD	1360	–1.9	1.49	0.27
p,p'-DDE	1552	–1.84	0.91	0.16
p,p'-DDT	931	–1.77	1.68	0.34

Table 11. Logistic model point estimates of T20, T50, and T80 values (95% confidence interval) for individual chemicals based on Sig Only classification of toxic samples for the screened marine amphipod database[a]

Chemical	T20	T50	T80
Metals (mg/kg dry wt.)			
Antimony	0.63 (0.55–0.72)	2.4 (2–2.8)	8.9 (6.6–12)
Arsenic	7.4 (6.8–8.1)	20 (18–23)	56 (45–69)
Cadmium	0.38 (0.34–0.43)	1.4 (1.2–1.5)	4.9 (4–6)
Chromium, total	49 (44–53)	140 (130–160)	410 (330–510)
Copper	32 (29–35)	94 (86–100)	280 (240–330)
Lead	30 (27–33)	94 (84–100)	300 (240–360)
Mercury	0.14 (0.12–0.15)	0.48 (0.43–0.54)	1.7 (1.4–2.1)
Nickel	15 (13–16)	47 (42–52)	150 (120–190)
Silver	0.23 (0.19–0.26)	1.1 (0.98–1.3)	5.8 (4.4–7.6)
Zinc	94 (87–100)	240 (220–270)	640 (540–750)
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)			
1-Methylnaphthalene	21 (17–25)	94 (73–120)	430 (280–670)
1-Methylphenanthrene	18 (15–23)	110 (88–140)	700 (450–1070)
2,6-Dimethylnaphthalene	25 (20–31)	130 (96–180)	710 (410–1230)
2-Methylnaphthalene	21 (18–26)	130 (100–160)	770 (510–1140)
Acenaphthene	19 (15–24)	120 (90–150)	710 (470–1090)
Acenaphthylene	14 (11–18)	140 (100–190)	1420 (800–2520)
Anthracene	34 (27–42)	290 (230–370)	2490 (1630–3790)
Benz(a)anthracene	61 (50–75)	470 (380–570)	3530 (2490–5020)
Benzo(a)pyrene	69 (57–85)	520 (430–630)	3910 (2750–5550)
Benzo(b)fluoranthene	130 (100–170)	1110 (810–1510)	9410 (5530–16000)
Benzo(g,h,i)perylene	67 (54–82)	500 (390–630)	3710 (2440–5630)
Benzo(k)fluoranthene	70 (55–90)	540 (410–710)	4120 (2540–6680)
Biphenyl	17 (14–21)	73 (57–93)	310 (210–470)
Chrysene	82 (67–99)	650 (530–800)	5190 (3600–7480)
Dibenz(a,h)anthracene	19 (15–23)	110 (92–140)	690 (480–990)
Fluoranthene	120 (98–150)	1030 (830–1280)	8950 (6070–13200)
Fluorene	19 (16–24)	110 (92–140)	660 (460–950)
Indeno(1,2,3-c,d)pyrene	68 (56–84)	490 (390–610)	3480 (2350–5160)
Naphthalene	30 (25–37)	220 (170–280)	1570 (1020–2410)
Perylene	74 (62–89)	450 (360–570)	2770 (1820–4210)
Phenanthrene	68 (57–81)	460 (380–550)	3060 (2190–4260)
Pyrene	120 (100–150)	930 (770–1130)	6980 (4940–9860)
Polychlorinated biphenyls (µg/kg dry wt.)			
PCBs, total	35 (27–44)	370 (280–480)	3930 (2410–6390)
Organochlorine pesticides (µg/kg dry wt.)			
Dieldrin	0.83 (0.65–1)	2.9 (2.3–3.6)	10 (6.9–15)
p,p'-DDD	2.2 (1.7–2.8)	19 (14–25)	160 (95–270)
p,p'-DDE	3.1 (2.2–4.4)	100 (61–180)	3410 (1280–9120)
p,p'-DDT	1.7 (1.3–2.2)	11 (8.3–15)	76 (45–130)

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

Table 12. Percent of toxic samples within ranges defined by Sig Only logistic model T20, T50, and T80 values and number of samples used to derive the logistic model for each chemical for the marine amphipod database[a]

Chemical	<T20	T20–T50	T50–T80	>T80	Number of samples
Metals (mg/kg dry wt.)					
Antimony	30.3	48.5	67.9	82.5	2173
Arsenic	30	43.8	56.3	69.7	2844
Cadmium	27.6	50.9	62.7	78.7	2958
Chromium	24.5	42.7	55.7	80	2827
Copper	22.1	50.6	64.9	85	3091
Lead	28.5	45	60.6	90	3010
Mercury	25.5	49	66.1	79.4	2788
Nickel	25.3	44.7	60.4	NA	2916
Silver	25.6	52.5	60.7	73.7	2552
Zinc	23.6	47.3	67.8	71.3	3013
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)					
1-Methylnaphthalene	23.7	48.3	60.3	75	1677
1-Methylphenanthrene	24.7	45.8	65.5	80	1697
2,6-Dimethylnaphthalene	23.5	42.2	57.4	NA	1505
2-Methylnaphthalene	25.4	47.5	61.1	88	2077
Acenaphthene	25.2	50.3	67.7	91.4	1795
Acenaphthylene	24	44.7	67.6	NA	1747
Anthracene	26.5	48.8	66.9	77.1	2268
Benz(a)anthracene	28.5	45.3	65	82.9	2574
Benzo(a)pyrene	27.7	48.5	64.2	83.8	2526
Benzo(b)fluoranthene	24	46.3	67.4	NA	1645
Benzo(g,h,i)perylene	25.3	46.6	63.6	86.7	2210
Benzo(k)fluoranthene	25.5	44.2	68.3	93.3	1691
Biphenyl	25.2	46.3	56.7	83.3	1507
Chrysene	28.7	47.8	64.8	86.1	2650
Dibenz(a,h)anthracene	23.9	49	65.3	85.7	1886
Fluoranthene	28.4	47.1	64.9	87.9	2734
Fluorene	22.1	48.4	68.2	87.2	2011
Indeno(1,2,3-c,d)pyrene	25.6	44.5	64.9	90.5	2212
Naphthalene	26.4	43.8	64.1	89.7	2201
Perylene	26.9	39.4	59.8	NA	2174
Phenanthrene	28.2	47.3	64.1	85.7	2688
Pyrene	28.3	46.2	65	87.2	2768
Polychlorinated biphenyls (µg/kg dry wt.)					
PCBs, total	26.8	46.3	72.7	81.5	1989
Organochlorine pesticides (µg/kg dry wt.)					
Dieldrin	20.2	53.8	66.7	78.8	770
p,p'-DDD	25.9	49.4	64.7	80.5	1672
p,p'-DDE	22.5	53.4	54.9	57.6	1899
p,p'-DDT	25.6	56.5	66.7	76.7	1176

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

NA = fewer than 10 samples

Table 13. Normalized chi-square values and number of samples for individual chemicals based on MSD classification of toxic samples for the screened marine amphipod database (logistic regression model parameters are shown for models having normalized chi-square values greater than 0.15)

Chemical	Number of samples	Intercept (B_0)	Slope (B_1)	Chi-square value/N
Metals (mg/kg dry wt.)				
Antimony	1905			0.14
Arsenic	2551			0.11
Cadmium	2641	–1.17	2.27	0.2
Chromium, total	2609	–7.47	3.21	0.16
Copper	2803	–6.98	3.06	0.29
Lead	2718	–6.29	2.79	0.21
Mercury	2472	–0.06	2.68	0.25
Nickel	2653			0.13
Selenium	1822			0.05
Silver	2294	–0.85	2.03	0.19
Zinc	2742	–9.26	3.54	0.22
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
1-Methylnaphthalene	1509	–4.94	2.24	0.19
1-Methylphenanthrene	1523	–4.89	2.02	0.23
2,6-Dimethylnaphthalene	1366	–5	2.15	0.17
2-Methylnaphthalene	1857	–4.37	1.72	0.17
Acenaphthene	1596	–4.29	1.7	0.23
Acenaphthylene	1570	–4.4	1.74	0.2
Anthracene	2000	–4.8	1.61	0.22
Benz(a)anthracene	2264	–5.96	1.88	0.23
Benzo(a)pyrene	2212	–6.15	1.94	0.24
Benzo(b)fluoranthene	1451	–7.5	2.3	0.28
Benzo(g,h,i)perylene	1939	–6.55	2.12	0.22
Benzo(k)fluoranthene	1490	–6.63	2.18	0.27
Biphenyl	1353	–4.53	2.1	0.18
Chrysene	2307	–5.97	1.78	0.22
Dibenz(a,h)anthracene	1677	–5.23	2.19	0.28
Fluoranthene	2373	–6.65	1.91	0.23
Fluorene	1817	–4.76	1.89	0.23
Indeno(1,2,3-c,d)pyrene	1962	–6.59	2.15	0.23
Naphthalene	1971	–4.82	1.72	0.18
Perylene	1955	–6.29	2.15	0.19
Phenanthrene	2335	–6.13	1.92	0.22
Pyrene	2404	–7.01	2.03	0.24
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	1766	–4.41	1.48	0.24
Organochlorine pesticides (µg/kg dry wt.)				
Dieldrin	682	–1.83	2.59	0.28
Gamma-Hexachlorocyclohexane	573			0.08
p,p'-DDD	1480	–2.59	1.6	0.24
p,p'-DDE	1682			0.13
p,p'-DDT	1009	–2.51	1.64	0.27

Table 14. Logistic model point estimates of T20, T50, and T80 values (95% confidence interval) for individual chemicals based on MSD classification of toxic samples for the screened marine amphipod database[a]

Chemical	T20	T50	T80
Metals (mg/kg dry wt.)			
Cadmium	0.8 (0.71–0.91)	3.3 (2.7–4)	13 (9.8–18)
Chromium, total	78 (72–85)	210 (180–240)	570 (440–730)
Copper	67 (61–74)	190 (170–210)	540 (440–660)
Lead	57 (52–63)	180 (160–210)	560 (440–720)
Mercury	0.32 (0.29–0.36)	1.1 (0.91–1.2)	3.5 (2.7–4.4)
Silver	0.55 (0.47–0.63)	2.6 (2.1–3.2)	13 (8.9–18)
Zinc	170 (150–180)	410 (370–460)	1,020 (840–1,230)
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)			
1-Methylnaphthalene	39 (32–47)	160 (120–220)	670 (410–1,090)
1-Methylphenanthrene	54 (44–66)	260 (200–350)	1,270 (790–2,040)
2,6-Dimethylnaphthalene	48 (39–61)	210 (150–310)	950 (530–1,700)
2-Methylnaphthalene	55 (45–67)	350 (250–490)	2,260 (1,300–3,930)
Acenaphthene	51 (40–64)	330 (230–460)	2,150 (1,220–3,770)
Acenaphthylene	54 (43–68)	330 (230–480)	2,080 (1,160–3,740)
Anthracene	130 (100–160)	940 (690–1280)	6,770 (4,040–11,300)
Benz(a)anthracene	270 (230–320)	1,460 (1,150–1,860)	7,980 (5,350–11,900)
Benzo(a)pyrene	290 (240–340)	1,510 (1,200–1,890)	7,840 (5,350–11,500)
Benzo(b)fluoranthene	460 (370–560)	1,820 (1,420–2,340)	7,300 (4,860–11,000)
Benzo(g,h,i)perylene	270 (230–330)	1,240 (960–1,590)	5,580 (3,670–8,480)
Benzo(k)fluoranthene	250 (200–310)	1,080 (830–1,410)	4,660 (3,060–7,120)
Biphenyl	31 (25–38)	140 (100–200)	650 (380–1,130)
Chrysene	370 (310–440)	2,200 (1,690–2,860)	13,200 (8,530–20,300)
Dibenz(a,h)anthracene	57 (48–68)	250 (200–310)	1,070 (730–1,550)
Fluoranthene	570 (480–680)	3,050 (2,390–3,900)	16,200 (10,900–24,300)
Fluorene	61 (50–74)	330 (250–430)	1,770 (1,110–2,810)
Indeno(1,2,3-c,d)pyrene	260 (220–310)	1,170 (920–1,480)	5,150 (3,480–7,630)
Naphthalene	100 (81–120)	640 (460–890)	4,130 (2,430–7,020)
Perylene	190 (160–230)	840 (650–1,090)	3,710 (2,420–5,680)
Phenanthrene	290 (250–350)	1,540 (1,200–1,980)	8,100 (5,370–12,200)
Pyrene	590 (500–700)	2,850 (2,270–3,570)	13,800 (9,480–19,900)
Polychlorinated biphenyls (µg/kg dry wt.)			
PCBs, total	110 (88–140)	940 (690–1,290)	8,120 (4,770–13,800)
Organochlorine pesticides (µg/kg dry wt.)			
Dieldrin	1.5 (1.2–1.9)	5.1 (3.9–6.6)	18 (11–27)
p,p'-DDD	5.6 (4.5–7.1)	41 (29–58)	300 (170–530)
p,p'-DDT	4.9 (3.6–6.5)	34 (23–52)	240 (120–470)

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

Table 15. Percent of toxic samples within ranges defined by logistic model T20, T50, and T80 values and number of samples in the database used to derive the logistic model for each chemical based on MSD classification of toxic samples for the marine amphipod database[a]

Chemical	<T20	T20–T50	T50–T80	>T80	Number of samples
Metals (mg/kg dry wt.*)*					
Cadmium	17.4	39.9	61.5	69.2	2958
Chromium, total	15.4	37.7	54.1	75.9	2827
Copper	14.2	40.1	66.5	76.6	3091
Lead	16.6	37.8	63.9	61.5	3010
Mercury	16.7	44.4	56.4	80.8	2788
Silver	16.8	43.5	46.5	NA	2552
Zinc	14.6	39.9	65.8	58.3	3013
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)					
1-Methylnaphthalene	14.8	39.6	50	NA	1677
1-Methylphenanthrene	15.5	40.7	63.6	50	1697
2,6-Dimethylnaphthalene	14.5	33.3	51.7	NA	1505
2-Methylnaphthalene	17.1	37	55.6	90	2077
Acenaphthene	15.9	40.1	56.3	75	1795
Acenaphthylene	15.2	37.7	63.8	NA	1747
Anthracene	16.9	39.5	62.2	78.6	2268
Benz(a)anthracene	17.5	39.4	62.6	73.3	2574
Benzo(a)pyrene	18	40.4	63	81.8	2526
Benzo(b)fluoranthene	15.9	41.7	61.9	86.7	1645
Benzo(g,h,i)perylene	17.4	37.5	64.9	NA	2210
Benzo(k)fluoranthene	15.8	42	62.1	84.6	1691
Benzofluoranthenes, total	23.3	37.9	63.6	NA	617
Biphenyl	16.3	36.5	50	NA	1507
Chrysene	18.3	40.3	58.2	72.7	2650
Dibenz(a,h)anthracene	16.5	38.4	69.7	68.4	1886
Fluoranthene	18.1	41.8	62.9	64.7	2734
Fluorene	14.8	38.4	63.2	69.6	2011
Indeno(1,2,3-c,d)pyrene	16.5	36.9	65.7	83.3	2212
Naphthalene	16.5	37.5	52.5	92.9	2201
Perylene	16.5	35.8	48.6	NA	2174
Phenanthrene	18.4	40	63.4	69.2	2688
Pyrene	18.1	42.6	57.5	72.7	2768
Polychlorinated biphenyls (µg/kg dry wt.)					
PCBs, total	18.6	44.2	56.2	69.2	1989
Organochlorine pesticides (µg/kg dry wt.)					
Dieldrin	17	45.5	60.9	72.7	770
p,p'-DDD	16.9	44.5	58	74.3	1672
p,p'-DDT	20.2	48.5	50.9	83.9	1176

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

NA = fewer than 10 samples

Table 16. Normalized chi-square values and number of samples for individual chemicals for organic carbon-normalized concentrations for the screened marine amphipod database (logistic regression model parameters are shown for models having normalized chi-square values greater than 0.15)

Chemical (µg/kg organic carbon)	Number of samples	Intercept (B_0)	Slope (B_1)	Chi-square value/N
Polycyclic aromatic hydrocarbons				
1-Methylnaphthalene	1291	–7	1.79	0.17
1-Methylphenanthrene	1308	–7.36	1.88	0.27
2,6-Dimethylnaphthalene	1185			0.15
2-Methylnaphthalene	1573	–6.82	1.67	0.19
Acenaphthene	1329	–6.3	1.58	0.25
Acenaphthylene	1361	–5.55	1.4	0.2
Anthracene	1676	–6.86	1.57	0.26
Benz(a)anthracene	1921	–8.25	1.81	0.29
Benzo(a)pyrene	1897	–8.16	1.79	0.29
Benzo(b)fluoranthene	1245	–8.66	1.81	0.27
Benzo(g,h,i)perylene	1664	–7.73	1.68	0.22
Benzo(k)fluoranthene	1274	–7.92	1.75	0.27
Biphenyl	1157	–6.54	1.71	0.17
Chrysene	1974	–8.2	1.76	0.28
Dibenz(a,h)anthracene	1411	–7.3	1.86	0.3
Fluoranthene	2021	–8.15	1.67	0.25
Fluorene	1519	–7.31	1.84	0.28
Indeno(1,2,3-c,d)pyrene	1685	–8.07	1.77	0.24
Naphthalene	1684	–6.27	1.44	0.16
Perylene	1709	–8.9	1.99	0.21
Phenanthrene	2005	–8.3	1.82	0.27
Pyrene	2092	–8.48	1.76	0.27
Polychlorinated biphenyls				
PCBs, total	1530	–6.03	1.38	0.22
Organochlorine pesticides				
Dieldrin	569	–4.33	1.85	0.24
Gamma-Hexachlorocyclohexane	523			0.07
p,p'-DDD	1261	–5.34	1.72	0.28
p,p'-DDE	1425			0.15
p,p'-DDT	854	–4.51	1.52	0.28

Table 17. Logistic model point estimates of T20, T50, and T80 organic carbon-normalized concentrations (95% confidence interval) for individual chemicals based on Sig Only classification of toxic samples in the screened marine amphipod database[a]

Chemical (µg/kg organic carbon)	T20	T50	T80
Polycyclic aromatic hydrocarbons			
1-Methylnaphthalene	1,390 (1,120–1,730)	8,310 (5,870–11,800)	49,600 (27500–89,600)
1-Methylphenanthrene	1,510 (1,220–1,850)	8,230 (6,430–10,500)	45,000 (29,200–69,200)
2-Methylnaphthalene	1,800 (1,480–2,200)	12,200 (9,110–16,400)	82,600 (49,500–138,000)
Acenaphthene	1,260 (990–1,610)	9,470 (6,900–13,000)	71,100 (41,600–122,000)
Acenaphthylene	920 (710–1,190)	8,960 (6,320–12,700)	87,100 (46,900–162,000)
Anthracene	3,010 (2,430–3,740)	22,900 (17,600–29,700)	174,000 (111,000–273,000)
Benz(a)anthracene	6,060 (5,070–7,250)	35,200 (29,000–42,700)	204,000 (146,000–287,000)
Benzo(a)pyrene	6,230 (5,190–7,480)	37,200 (30,700–45,200)	223,000 (159,000–313,000)
Benzo(b)fluoranthene	10,200 (8,150–12,900)	59,500 (45,200–78,300)	346,000 (217,000–550,000)
Benzo(g,h,i)perylene	6,000 (4,920–7,320)	40,200 (31,100–5,1900)	269,000 (171,000–424,000)
Benzo(k)fluoranthene	5,490 (4,360–6,910)	34,100 (25,800–45,000)	212,000 (132,000–338,000)
Biphenyl	1,020 (810–1,290)	6,580 (4,570–9,470)	42,400 (22,500–79,800)
Chrysene	7,380 (6,160–8,830)	45,200 (37,100–55,000)	277,000 (196,000–390,000)
Dibenz(a,h)anthracene	1,490 (1,210–1,830)	8,250 (6,610–10,300)	45,700 (31,000–67,400)
Fluoranthene	11,200 (9,370–13,500)	76,000 (61,100–94,500)	514,000 (350,000–753,000)
Fluorene	1,640 (1,350–2,000)	9,300 (7,320–11,800)	52,700 (35,000–79,200)
Indeno(1,2,3-c,d)pyrene	6,050 (5,000–7,330)	36,900 (29,300–46,500)	225,000 (150,000–338,000)
Naphthalene	2,470 (1,990–3,070)	22,700 (16,200–31,700)	208,000 (114,000–378,000)
Perylene	6,000 (5,080–7,070)	29,800 (23,800–37,400)	149,000 (99,400–222,000)
Phenanthrene	6,220 (5,250–7,360)	35,800 (29,400–43,800)	207,000 (146,000–293,000)
Pyrene	10,500 (8,870–12,500)	64,300 (53,000–77,900)	392,000 (280,000–551,000)
Polychlorinated biphenyls			
PCBs, total	2,290 (1,790–2,930)	23,100 (17,500–30,600)	233,000 (138,000–394,000)
Organochlorine pesticides			
Dieldrin	39 (29–53)	220 (150–310)	1,230 (650–2,320)
p,p'-DDD	200 (160–240)	1240 (950–1,630)	7,910 (4,930–12,700)
p,p'-DDT	110 (85–160)	940 (650–1,380)	7,740 (4,040–14,900)

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

Table 18. Percent of toxic samples within ranges defined by logistic model T20, T50, and T80 values for organic carbon-normalized concentrations and number of samples in the database used to derive the logistic model for each chemical based on Sig Only classification of toxic samples for the marine amphipod database[a]

Chemical (µg/kg organic carbon)	<T20	T20–T50	T50–T80	>T80	Number of samples
Polycyclic aromatic hydrocarbons					
1-Methylnaphthalene	26.3	47	55.3	NA	1612
1-Methylphenanthrene	27	48.7	65.5	60	1641
2-Methylnaphthalene	29.8	48.2	57.8	62.5	2008
Acenaphthene	27.2	54.7	68.2	76.2	1743
Acenaphthylene	24.7	51.1	58.3	NA	1684
Anthracene	29.8	52.6	64.2	72	2189
Benz(a)anthracene	31	51.8	61.4	75	2489
Benzo(a)pyrene	29.3	53.7	61.2	78.3	2443
Benzo(b)fluoranthene	24.9	52.7	65	NA	1571
Benzo(g,h,i)perylene	27.4	52.1	61.4	NA	2135
Benzo(k)fluoranthene	26.4	50.9	67.1	78.6	1621
Biphenyl	28	47.1	43.6	NA	1453
Chrysene	30.2	53.4	61.2	50	2564
Dibenz(a,h)anthracene	24.8	54.7	68.3	77.4	1815
Fluoranthene	30.1	53.2	61.9	78.3	2644
Fluorene	26	52.2	66.9	82.1	1936
Indeno(1,2,3-c,d)pyrene	27.8	49.7	62.2	78.6	2135
Naphthalene	29.5	48	56.9	NA	2123
Perylene	28.4	45	50	72.7	2100
Phenanthrene	30.5	51.3	63.6	NA	2606
Pyrene	29.5	51.9	61.5	74.4	2678
Polychlorinated biphenyls					
PCBs, total	27.2	51.5	64.3	77.8	1940
Organochlorine pesticides					
Dieldrin	21.8	58.3	74.7	46.2	729
p,p'-DDD	27.8	54.9	61.2	70.5	1623
p,p'-DDT	28.6	59.2	61.4	78.8	1131

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

NA = fewer than 10 samples

Table 19. Differences between percent predicted toxic samples and percent observed toxic samples (observed minus predicted) within ranges defined by logistic model T20, T50, and T80 values based on Sig Only classification of toxic samples for the marine amphipod database [a,b]

Chemical	<T20	T20–T50	T50–T80	>T80
Metals (mg/kg dry wt.)				
Antimony	19.2	17.4	6.1	–9.9
Arsenic	19.6	12.5	–3.6	–19.7
Cadmium	19.2	17.3	0.9	–12
Chromium, total	15.7	11.8	–3.4	–7.7
Copper	15.3	17.7	2	–2.7
Lead	19.1	14.2	–3.3	3.9
Mercury	17.5	16.2	2.4	–7.2
Nickel	16.5	11.4	–1.7	–34.2
Silver	15.7	19.4	–2.7	–11.6
Zinc	15.1	14.7	5.8	–16.2
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
1–Methylnaphthalene	17.6	16.3	–2.7	–11.5
1–Methylphenanthrene	17.5	12.8	4.3	–7.8
2,6–Dimethylnaphthalene	17.3	10.9	–2	
2–Methylnaphthalene	17.4	15.3	–0.3	–0.5
Acenaphthene	19.6	16.8	5	2.1
Acenaphthylene	16.4	11.5	7.5	–12
Anthracene	19.1	15.7	5.3	–10.1
Benz(a)anthracene	21	12	2.7	–3.2
Benzo(a)pyrene	20.1	14.5	1.1	–1.5
Benzo(b)fluoranthene	17.6	13.4	4.2	11
Benzo(g,h,i)perylene	17.7	13.4	1.6	0.8
Benzo(k)fluoranthene	18.9	11.7	5.8	5.2
Biphenyl	19.2	14.3	–4.3	–7
Chrysene	20.9	14.1	1.9	–0.2
Dibenz(a,h)anthracene	17.4	15.9	2.7	–0.9
Fluoranthene	20.5	14.3	3.1	0.9
Fluorene	15.7	14.5	5.7	–2.3
Indeno(1,2,3–c,d)pyrene	18.2	11.2	2.8	4
Naphthalene	17.9	11.6	2.3	0.9
Perylene	19.2	7.5	–0.3	–42.9
Phenanthrene	20.5	13.5	2.3	–2.8
Pyrene	20.6	13.2	2.7	–0.2
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	17.1	14.1	11.9	–9.5
Organochlorine pesticides (µg/kg dry wt.)				
Dieldrin	12.9	21.7	4.2	–7.8
p,p'–DDD	16.9	16.3	2.7	–11.2
p,p'–DDE	11.3	22	–5.1	–29.8
p,p'–DDT	17.3	25.1	2	–16.4

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).
[b]Differences greater than 20% are shaded. Blank cell indicates fewer than six samples.

Table 20. Differences between percent predicted toxic samples and percent observed toxic samples (observed minus predicted) within ranges defined by logistic model T20, T50, and T80 values based on MSD classification of toxic samples for the marine amphipod database[a,b]

Chemical	<T20	T20–T50	T50–T80	>T80
Metals (mg/kg dry wt.)				
Cadmium	10.5	8.8	0.3	–21.2
Chromium, total	7	7.9	–8.6	–11.2
Copper	8.8	7.5	3.4	–11.2
Lead	9.8	5.8	2.4	–28.1
Mercury	10.8	13.2	–5.9	–6.3
Silver	9.6	11.4	–12.2	
Zinc	8	9.1	2.9	–30.9
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
1–Methylnaphthalene	9.9	9.5	–11.2	
1–Methylphenanthrene	10.1	9.5	2.7	–38.6
2,6–Dimethylnaphthalene	9.9	1.4	–7	
2–Methylnaphthalene	10.5	6	–4.3	2.5
Acenaphthene	10.7	10.1	–3.7	–14.3
Acenaphthylene	10	6.3	3.4	–20.4
Anthracene	11.6	8.5	1.9	–9.5
Benz(a)anthracene	12.6	7.3	1.3	–14.1
Benzo(a)pyrene	12.8	8.4	0.8	–4.5
Benzo(b)fluoranthene	12.2	9	–2.6	0
Benzo(g,h,i)perylene	12.5	6.8	5.2	32
Benzo(k)fluoranthene	12	9.4	0.1	–5.7
Biphenyl	10.6	5.5	–9.9	–13.1
Chrysene	13.1	9.3	–1.4	–15.3
Dibenz(a,h)anthracene	11.5	6.6	6.4	–19.8
Fluoranthene	13.1	10.5	2.1	–24.4
Fluorene	9.4	7.8	2	–20.8
Indeno(1,2,3–c,d)pyrene	11.7	6	4.2	–4.8
Naphthalene	10.6	6	–8.5	3.2
Perylene	11.5	3.9	–8.5	
Phenanthrene	13.3	9.2	2.6	–19.7
Pyrene	13.3	10.4	–3.4	–15
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	10.1	13	–6.6	–19.4
Organochlorine pesticides (µg/kg dry wt.)				
Dieldrin	10.3	12.5	–1.9	–13
p,p'–DDD	9.7	14.4	–3.3	–16.6
p,p'–DDT	13.3	18	–9.2	–8.2

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).
[b]Differences greater than 20% are shaded. Blank cell indicates fewer than six samples.

Table 21. Differences between mean percent predicted toxic samples and percent observed toxic samples (observed minus predicted) within ranges defined by logistic model T20, T50, and T80 values for organic carbon-normalized concentrations based on Sig Only classification of toxic samples for the marine amphipod database[a,b]

Chemical (µg/kg organic carbon)	<T20	T20–T50	T50–T80	>T80
Polycyclic aromatic hydrocarbons				
1–Methylnaphthalene	18.7	16.4	–2.8	–43.5
1–Methylphenanthrene	19.7	17	2.7	–27.1
2–Methylnaphthalene	20.9	16.6	–1	–24.5
Acenaphthene	20.2	22.6	7	–12.5
Acenaphthylene	16.2	19.2	–1.2	–34.4
Anthracene	22.2	20	2.1	–14.9
Benz(a)anthracene	23.9	18.9	–1	–11.5
Benzo(a)pyrene	22.2	20.8	–1.6	–6.7
Benzo(b)fluoranthene	18.9	20.1	1.1	–3.7
Benzo(g,h,i)perylene	19.8	20.3	1	
Benzo(k)fluoranthene	19.8	19.1	5.2	–7.3
Biphenyl	19.6	15.7	–15.3	–35.4
Chrysene	22.9	20.8	–0.9	–8.8
Dibenz(a,h)anthracene	18.2	22.8	5.8	–6.1
Fluoranthene	22.1	21.3	0.4	–4.9
Fluorene	19.1	19.8	5.4	–11.6
Indeno(1,2,3–c,d)pyrene	20.5	18	1.4	–20.2
Naphthalene	19.4	16.1	–2.9	–13.3
Perylene	20.5	12.5	–8.9	
Phenanthrene	22.7	19	1.6	–14.6
Pyrene	21.7	19.5	–0.9	–14
Polychlorinated biphenyls				
PCBs, total	17.1	19.1	4.5	–10.9
Organochlorine pesticides				
Dieldrin	11.7	25.3	13.5	–41.5
p,p'–DDD	19.7	22.6	0.3	–20.4
p,p'–DDT	19.6	28.6	–1.6	–12.1

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).
[b]Differences greater than 20% are shaded. Blank cell indicates fewer than six samples.

Table 22. Differences between mean percent predicted toxic samples and percent observed toxic samples (observed minus predicted) within ranges defined by logistic model T20, T50, and T80 values for each chemical using a screening factor of 2X the mean of nontoxic samples and Sig Only classification of toxic samples for the marine amphipod database[a,b]

Chemical	<T20	T20–T50	T50–T80	>T80
Metals (mg/kg dry wt.)				
Antimony	15.7	14.6	6.1	–11.8
Arsenic	16.3	9.9	–5.1	–22.7
Cadmium	16.3	14.7	–2.6	–13.1
Chromium, total	12	9.5	–4	–7.7
Copper	12.1	14.6	0.8	–4.1
Lead	16	11.4	–4.8	2.6
Mercury	14.2	13.4	0.5	–7.2
Nickel	12.7	9	–3.4	–34.2
Silver	12.2	17.4	–4.5	–11.6
Zinc	11.9	11.8	4.1	–16.2
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
1–Methylnaphthalene	14.4	14.1	–2.7	–19.8
1–Methylphenanthrene	14	10	3.7	–7.8
2,6–Dimethylnaphthalene	14.4	9.8	–5	
2–Methylnaphthalene	14.5	12.6	–3	–4.5
Acenaphthene	16.5	13.5	3.2	2.1
Acenaphthylene	13.3	8.3	6.2	–12
Anthracene	16.4	12.2	4.5	–10.1
Benz(a)anthracene	18	9	1.6	–3.2
Benzo(a)pyrene	17.1	11.7	–0.3	–1.5
Benzo(b)fluoranthene	14.5	12.1	4.2	11
Benzo(g,h,I)perylene	14.5	10.9	0.8	0.8
Benzo(k)fluoranthene	15.6	10.3	5.8	5.2
Biphenyl	16.1	12.3	–7.4	–7
Chrysene	18	10.9	1.3	–0.2
Dibenz(a,h)anthracene	14.6	12.9	1.5	–0.9
Fluoranthene	17.4	11.5	1.9	0.9
Fluorene	12.9	10.5	4.8	–2.3
Indeno(1,2,3–c,d)pyrene	15.1	8.6	2.3	4
Naphthalene	15.1	8.6	–0.7	–2.5
Perylene	16.4	5.4	–2.1	–42.9
Phenanthrene	17.6	10.6	0.5	–2.8
Pyrene	17.7	10.1	1.9	–0.2
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, total	14.2	11.9	10.6	–21.3
Organochlorine pesticides (µg/kg dry wt.)				
Dieldrin	10.6	19.4	1.7	–10.8
p,p'–DDD	14.2	14.3	2.7	–11.2
p,p'–DDE	8.1	20.5	–6.3	–29.8
p,p'–DDT	13.2	22.7	1.1	–16.4

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).
[b]Differences greater than 20% are shaded. Blank cell indicates fewer than six samples.

Table 23. Statistical comparisons of the logistic regression models for *A. abdita* and *R. abronius* using the chi-square statistic (–2 log likelihood)

Chemical	Common slope, common intercept	Common slope, separate intercept	Separate slope, separate intercept	Preferred model
Metals (mg/kg dry wt.)				
Antimony	548.39	506.3	430.25	a
Arsenic	582.09	579.8	404.54	b
Cadmium	899.71	872.55	754.76	a
Chromium, total	645.99	618.91	468.43	a
Copper	1062.54	1061.94	987.62	b
Lead	928.02	911.34	679.52	a
Mercury	890.64	890.64	734.98	b
Nickel	614.01	586.86	441.04	a
Silver	753.84	753.56	530.67	b
Zinc	913.32	894.37	702.21	a
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)				
1–Methylnaphthalene	357.46	350.53	327.48	a
1–Methylphenanthrene	487.05	486.26	397.74	b
2,6–Dimethylnaphthalene	291.27	290.06	251.51	b
2–Methylnaphthalene	559.79	559.72	426.77	b
Acenaphthene	496.32	487.75	475.79	a
Acenaphthylene	381.16	380.26	332.24	b
Anthracene	599.92	599.88	526.43	b
Benz(a)anthracene	800.12	799.74	624.63	b
Benzo(a)pyrene	779.46	779.43	614.03	b
Benzo(b)fluoranthene	415.05	410.87	358.83	a
Benzo(g,h,i)perylene	614.51	611.4	454.16	b
Benzo(k)fluoranthene	438.18	428.92	393.22	a
Biphenyl	352.24	346.47	322.79	a
Chrysene	775.03	775	608.87	b
Dibenz(a,h)anthracene	550.44	550.43	503.33	b
Fluoranthene	760.93	760.37	576.24	b
Fluorene	588.33	586.25	538.94	b
Indeno(1,2,3–c,d)pyrene	619.06	617.96	494.56	b
Naphthalene	553.13	547.8	426.97	a
Perylene	535.85	526.83	398.32	a
Phenanthrene	830.04	824.66	647.82	a
Pyrene	842.54	841.84	642.07	b
Polychlorinated biphenyls (µg/kg dry wt.)				
PCBs, Total	457.11	448.19	389.03	a
Organochlorine pesticides (µg/kg dry wt.)				
Dieldrin	230.52	229.97	224.24	b
p,p'–DDD	443.41	412.79	364.34	a
p,p'–DDE	385.15	272.17	251.13	a
p,p'–DDT	340.44	323.15	312.15	a

a = Separate slope, separate intercept
b = Common slope, separate intercept

Table 24. Ratio of Tp values from species-specific *A. abdita* models to corresponding Tp values from the combined amphipod models (only models with normalized chi-square >0.15 are included)[a]

Chemical	T20	T50	T80
Metals			
Antimony	1.46	1.04	0.74
Cadmium	1.71	1.4	1.15
Chromium, total	1.48	1.19	0.96
Copper	1.42	1.49	1.56
Lead	1.82	1.49	1.21
Mercury	1.77	1.92	2.09
Silver	2.63	2.66	2.68
Zinc	1.59	1.37	1.18
Polycyclic aromatic hydrocarbons			
1–Methylnaphthalene	1.21	1.29	1.38
1–Methylphenanthrene	1.96	2.09	2.23
2,6–Dimethylnaphthalene	1.32	1.26	1.21
2–Methylnaphthalene	2.11	2.09	2.07
Acenaphthene	1.24	1.72	2.4
Acenaphthylene	1.74	2.25	2.9
Anthracene	2.06	2.6	3.28
Benz(a)anthracene	2.82	2.8	2.78
Benzo(a)pyrene	2.71	2.91	3.13
Benzo(b)fluoranthene	1.92	2.01	2.11
Benzo(g,h,i)perylene	2.48	2.46	2.44
Benzo(k)fluoranthene	1.58	1.42	1.27
Biphenyl	1.21	1.44	1.72
Chrysene	2.85	3.25	3.71
Dibenz(a,h)anthracene	1.64	1.97	2.37
Fluoranthene	3.13	3.21	3.29
Fluorene	1.53	2.05	2.74
Indeno(1,2,3–c,d)pyrene	2.13	2.16	2.19
Naphthalene	2.05	1.83	1.63
Perylene	1.86	1.42	1.08
Phenanthrene	2.75	2.44	2.17
Pyrene	3.03	3.04	3.05
Polychlorinated biphenyls			
PCBs, total	1.9	2.6	3.56
Organochlorine pesticides			
Dieldrin	1.12	1.12	1.12
p,p'–DDD	1.83	1.3	0.92
p,p'–DDE	1.46	0.33	0.08
p,p'–DDT	1.36	1.07	0.85
MEAN	1.91	1.91	1.98

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

Table 25. Ratio of Tp values from the combined amphipod models to corresponding Tp values from the species-specific *R. abronius* models (only models with normalized chi–square >0.15 are included)[a]

Chemical	T20	T50	T80
Metals			
Cadmium	3.42	1.82	0.97
Copper	1.73	1.5	1.3
Lead	3.2	2.34	1.71
Mercury	2.32	2.13	1.97
Silver	3.37	3	2.67
Zinc	2.67	1.85	1.29
Polycyclic aromatic hydrocarbons			
1–Methylnaphthalene	1.68	3.39	6.82
1–Methylphenanthrene	6.14	4.12	2.77
2,6–Dimethylnaphthalene	3.5	5.88	9.89
2–Methylnaphthalene	3.97	4.3	4.66
Acenaphthene	1.05	1.69	2.73
Acenaphthylene	3.18	3.72	4.36
Anthracene	3.44	2.89	2.43
Benz(a)anthracene	4.48	3.88	3.36
Benzo(a)pyrene	4.24	3.77	3.35
Biphenyl	1.58	2.63	4.36
Chrysene	4.54	3.79	3.17
Dibenz(a,h)anthracene	2.72	2.2	1.78
Fluoranthene	5.1	4.06	3.22
Fluorene	2.07	2.27	2.49
Indeno(1,2,3–c,d)pyrene	5.83	4.19	3.01
Naphthalene	8.52	4.57	2.45
Perylene	8.93	4.9	2.69
Phenanthrene	5.48	3.49	2.22
Pyrene	4.77	3.89	3.17
Organochlorine pesticides			
Dieldrin	2.24	1.68	1.26
MEAN	3.85	3.23	3.08

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

Table 26. Differences between mean percent predicted toxic samples and percent observed toxic samples (observed minus predicted) for *A. abdita* and *R. abronius* within ranges defined by T20, T50, and T80 values[a,b]

	A. abdita				*R. abronius*			
Chemical	**<T20**	**T20–T50**	**T50–T80**	**>T80**	**<T20**	**T20–T50**	**T50–T80**	**>T80**
Metals (mg/kg dry wt.)								
Antimony	9	–1	–3	–19	54	34	8	–6
Arsenic	12	–10	–30		41	37	17	–29
Cadmium	8	–9	–7	1	47	35	6	–14
Chromium, total	7	–5	–7	3	41	34	5	–17
Copper	10	–9	–11	–7	37	31	15	–2
Lead	7	–6	–15	6	47	31	13	4
Mercury	9	–3	–17	–14	43	35	19	4
Nickel	9	–5	–8	–28	46	28	4	–14
Silver	7	–9	–16	–10	40	40	15	–5
Zinc	8	–7	–7	–13	39	34	20	–6
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)								
1–Methylnaphthalene	12	10	–6	–1	56	41		
1–Methylphenanthrene	10	1	–11	–12	51	34	21	9
2,6–Dimethylnaphthalene	12	1	–1		57	44		
2–Methylnaphthalene	10	–3	–13	–3	45	40	25	0
Acenaphthene	15	2	–7	–8	45	35	14	13
Acenaphthylene	11	3	2	–33	42	32	20	
Anthracene	11	–1	–4	–18	46	35	13	2
Benz(a)anthracene	9	–9	–9	–6	47	34	15	6
Benzo(a)pyrene	10	–8	–11	–1	47	39	13	1
Benzo(b)fluoranthene	9	–1	2	–1	53	38	7	8
Benzo(g,h,i)perylene	9	–6	–3	–7	52	40	7	9
Benzo(k)fluoranthene	10	3	0	10	55	34	12	3
Biphenyl	13	6	–17	–17	59	46	22	15
Chrysene	8	–7	–7	–9	49	37	11	6
Dibenz(a,h)anthracene	11	–4	–2	–1	48	36	6	2
Fluoranthene	9	–6	–7	–17	48	35	13	12
Fluorene	11	0	–4	–13	44	34	15	4
Indeno(1,2,3–c,d)pyrene	9	–8	–2	0	49	40	9	4
Naphthalene	8	0	–3	–8	50	33	11	6
Perylene	8	0	–8	–15	53	32	11	
Phenanthrene	9	–2	–15	–8	48	31	18	–1
Pyrene	9	–7	–11	–7	47	35	14	6
Polychlorinated biphenyls (µg/kg dry wt.)								
PCBs, total	8	–3	1	–16	52	29	24	4
Organochlorine pesticides (µg/kg dry wt.)								
Dieldrin	9	14	4	–26	48	44	15	6
p,p'–DDD	8	0	2	–1	55	34	8	–16
p,p'–DDE	6	8	34		59	35	–2	–35
p,p'–DDT	13	5	–5	1	50	45	12	–21

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

[b]Percent predicted toxicity and Tp values were calculated using the combined species logistic model using the Sig Only classification of toxicity and the screened marine amphipod database. Differences greater than 20% are shaded. Blank cell indicates fewer than six samples.

Table 27. Changes in logistic model point estimates of T20, T50, and T80 concentrations for PCBs based on Sig Only classification of toxic samples based on corrected PCB model for the screened marine amphipod database[a]

Estimated	T20	T50	T80
Original	32.5	468	6750
Corrected	34.5	368	3930

[a]The notation Tp (e.g., T50) is used to denote the concentration that would give a response of "p" percent according to the model (e.g., the probability that 50% of the samples would be toxic).

Table 28. Estimated probability of toxicity from marine amphipod chemical-specific logistic regression models for LC50 values (dry wt.) reported from 10–day spiked sediment amphipod toxicity tests

Chemical	LC50	Probability of toxicity	Source
Cadmium (mg/kg)	9.81	0.9	Mearns et al., 1986
	8.8–10	0.88–0.9	Kemp et al., 1986
	8.2–11.5	0.88–0.91	Robinson et al., 1988
	6.9	0.85	Swartz et al., 1985
Mercury (mg/kg)	13.1	0.97	Swartz et al., 1988
Zinc (mg/kg)	276	0.54	Swartz et al., 1988
Fluoranthene (mg/kg)	4.2	0.71	Swartz et al., 1988
	3.3–10.5	0.68–0.82	Swartz et al., 1987
Phenanthrene (mg/kg)	3.68	0.82	Swartz et al., 1989
Total PCBs (µg/kg)	8.8	0.87	Swartz et al., 1988
p,p'–DDT (µg/g)	11.2–125	0.5–0.85	Word et al., 1987

PCBs = polychlorinated biphenyls

Table 29. Number and percent of samples by chemical class that represented the maximum probability of toxicity used in the P_Max model derived from the marine amphipod database

Chemical class	Number	Percent
Metals	2234	69.3
PAHs	596	18.5
Pesticides–PCBs	393	12.2

PAHs = polycyclic aromatic hydrocarbons
PCBs = polychlorinated biphenyls

Table 30. Differences between mean percent predicted toxic samples and percent observed toxic samples (observed minus predicted) by probability quartile for individual studies with at least 20 samples[a]

Study	Number of samples	Probability quartile			
		<25	25–50	50–75	>75
A. abdita					
New Bedford Harbor Monitoring, 1993	77		–20	–18	19
NSTP Hudson–Raritan Phase I, 1991	34			–10	4
NSTP Long Island Sound, 1994	63		35	30	
NSTP Boston Harbor, 1994	30		–14	–25	
REMAP–Hudson/Raritan Bay, 1993	41		–34	–20	
REMAP–Long Island Sound, 1993	43	–7	–13		
REMAP–Hudson/Raritan Bay, 1994	42	–6	–15	–4	
REMAP–Long Island Sound, 1994	42	–14	–25	–48	
EMAP–Delaware Bay, 1990	42	18	–5		
EMAP–Chesapeake Bay, 1990	61	3	–19	–6	
EMAP–Chesapeake Bay, 1991	62	9	12		
EMAP–Chesapeake Bay, 1992	59	–17	–16	–56	
EMAP–Chesapeake Bay, 1993	62	–9	–31		
NSTP Charleston Harbor, 1993	79	–18	–36	–55	
NSTP Savannah River, 1994	60	–6	–30	–27	
NSTP Biscayne Bay, Phase I, 1995	105	8	–1	18	
NSTP Biscayne Bay, Phase II, 1996	120	–15	–33		
EMAP Virginia and N. Carolina, 1994	50	–11	–31		
EMAP Virginia and N. Carolina, 1995	50	–16	–32		
NSTP Choctawhatchee, 1994	21	–16	–27		
NSTP St. Andrews Bay, 1993	31	–19	–40		
NSTP Tampa Bay Phase II, 1992	45	–23	–35	–58	
R. abronius					
Port of Tacoma, Blair Waterway	21	14	37		
Commencement Bay Remedial Invest.	50		16	9	
Elliott Bay sediment survey, 1985	97		–15	–5	7
Puget Sound Eight–Bay survey, 1985	48		31	22	
Everett Harbor, 1985	29		–10	7	
Port of Tacoma Remedial Investigation	79		53	28	21
Puget Sound Ambient Monitoring, 1989	50	–8	–6		
Puget Sound Ambient Monitoring, 1990	50	49	64		
Puget Sound Ambient Monitoring, 1991	47	–5	–29		
BPTCP, 1992 Q3, LA	58		3	2	
BPTCP, Screening, 1992 Q4, San Diego	23		61	36	
BPTCP, Screening, 1993 Q2–3, San Diego	78		15	–8	
BPTCP, Screening, 1994 Q1, LA	45		60	35	
BPTCP, Screening, 1994 Q1, Santa Ana	24		43		
BPTCP, Screening, 1994 Q1, San Diego	93	67	48	23	
EMAP So.CA, 1994 Q3, San Diego	25	33	41		
Palos Verdes shelf and Santa Monica Bay (Swartz et al., 1991)	31		–39		8

[a]Differences greater than 20% are shaded. Blank cell indicates fewer than six samples.

Table 31. List of studies (primary data source) combined for analysis of broader geographic areas

Geographic area	Number of samples
Hudson-Raritan/Long Island Sound (NSTP and Regional EMAP)	280
Virginian Province (EMAP)	489
Southeastern US (EMAP Carolinian and NSTP)	636
Puget Sound (SEDQUAL)	594
California (MLML)	508

Table 32. Percent of samples predicted to be toxic to amphipods at the chemical concentrations defined by sediment quality guidelines

Chemical	ERL	ERM	TEL	PEL	AET
Metals					
Antimony	NA	NA	NA	NA	99
Arsenic	22	85	20	73	99
Cadmium	46	89	32	77	93
Chromium (total)	33	78	22	54	94
Copper	21	79	11	54	97
Lead	30	73	20	55	96
Mercury	22	60	19	60	85
Nickel	28	53	22	48	92
Silver	47	73	41	59	81
Zinc	33	68	27	54	98
Polycyclic aromatic hydrocarbons					
2–Methylnaphthalene	39	78	19	59	89
Acenaphthene	18	75	10	45	90
Acenaphthylene	33	71	13	49	79
Anthracene	31	70	24	47	92
Benz(a)anthracene	40	70	22	57	84
Benzo(a)pyrene	47	68	23	57	79
Benzo(g,h,I)perylene	NA	NA	NA	NA	78
Chrysene	41	73	23	54	91
Dibenz(*a*,*h*)anthracene	39	66	10	53	90
Fluoranthene	41	74	19	56	90
Fluorene	20	77	21	55	94
Indeno(1,2,3–c,d)pyrene	NA	NA	NA	NA	83
Naphthalene	45	83	22	60	84
Phenanthrene	39	70	23	53	94
Pyrene	44	67	22	57	88
Polychlorinated biphenyls					
PCBs, total	16	40	16	40	78
Organochlorine pesticides					
Dieldrin	NA	NA	18	61	55
p,p'–DDD	NA	NA	15	36	69
p,p'–DDE	18	37	17	62	45
p,p'–DDT	NA	NA	16	35	91

AET = Apparent effect threshold for amphipod survival (Gries and Waldo, 1996)
ERL = Effect–range low (Long and MacDonald, 1992)
ERM = Effect–range median (Long and MacDonald, 1992)
PEL = Probable effect level (MacDonald et al., 1996)
TEL = Threshold effect level (MacDonald et al., 1996)

NA = No guideline value available

Table 33. Percent of samples predicted to be toxic to amphipods at the chemical concentrations defined by the Final Chronic Value for individual polycyclic aromatic hydrocarbons

Chemical	Final chronic value	
	ppb dry weight, assuming 1% OC	Percent predicted toxic
Acenaphthene	4,910	95
Acenaphthylene	4,520	89
Anthracene	5,940	88
Benz(a)anthracene	8,410	88
Benzo(a)pyrene	9,650	88
Benzo(b)fluoranthene	9,790	80
Benzo(k)fluoranthene	9,810	88
Benzo(g,h,i)perylene	10,950	89
Chrysene	8,440	85
Dibenz(a,h)anthracene	11,230	97
Fluoranthene	7,070	77
Fluorene	5,380	95
Indeno(1,2,3-c,d)pyrene	11,150	90
Naphthalene	3,850	88
Perylene	9,670	91
Phenanthrene	5,960	87
Pyrene	6,970	80

Table 34. Percent of samples predicted and observed to be toxic to amphipods within ranges defined by toxic units for polycyclic aromatic hydrocarbons (PAHs)

PAH toxic unit		Percent toxic		Number of samples
Range	Average	Mean predicted	Observed	
<1	0.09	36	37.8	2823
1–2	1.41	57	54.4	182
2–3	2.49	62	58.1	86
3–5	3.86	63	61.7	60
5–10	6.61	68	66.7	42
10–100	20.41	77	81	21

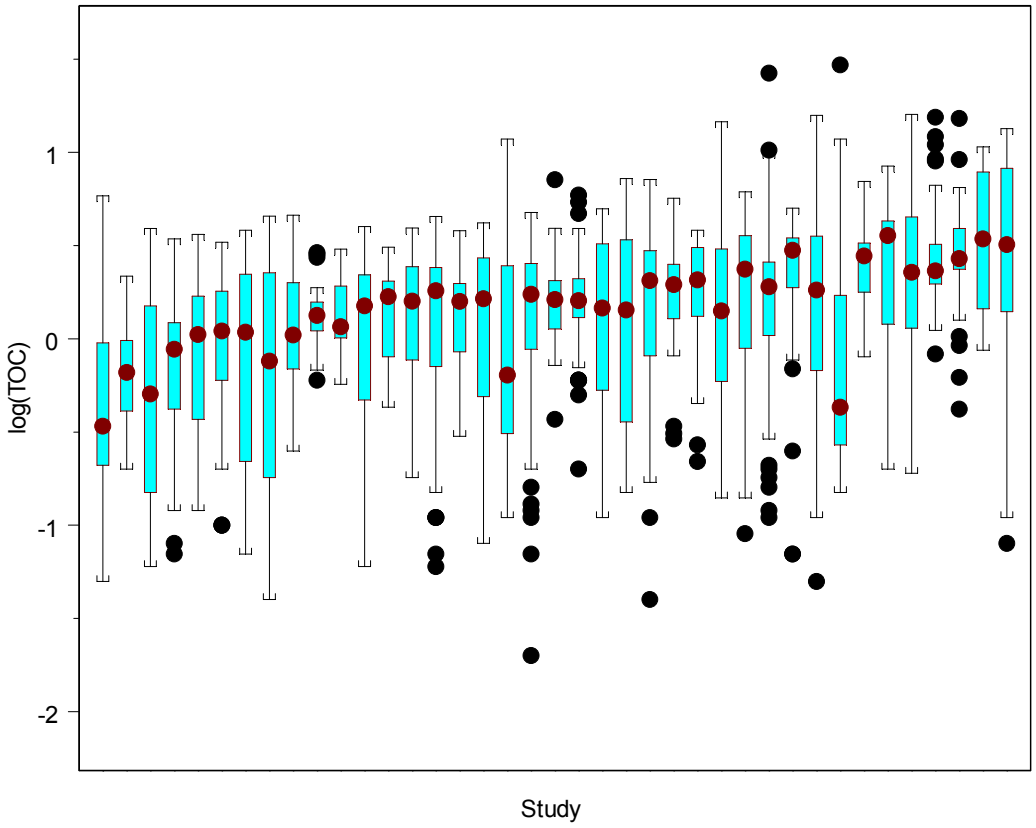


Figure 1. Box plots (plotted as per Tukey, 1977) summarizing the distribution of total organic carbon (TOC) (log 10) values for each study with greater than 20 samples from the marine amphipod database. Studies are ordered from left to right by mean TOC. The top and bottom of each rectangular box correspond to the upper and lower quartiles of the data, respectively; the dot within each box corresponds to the median of the data.

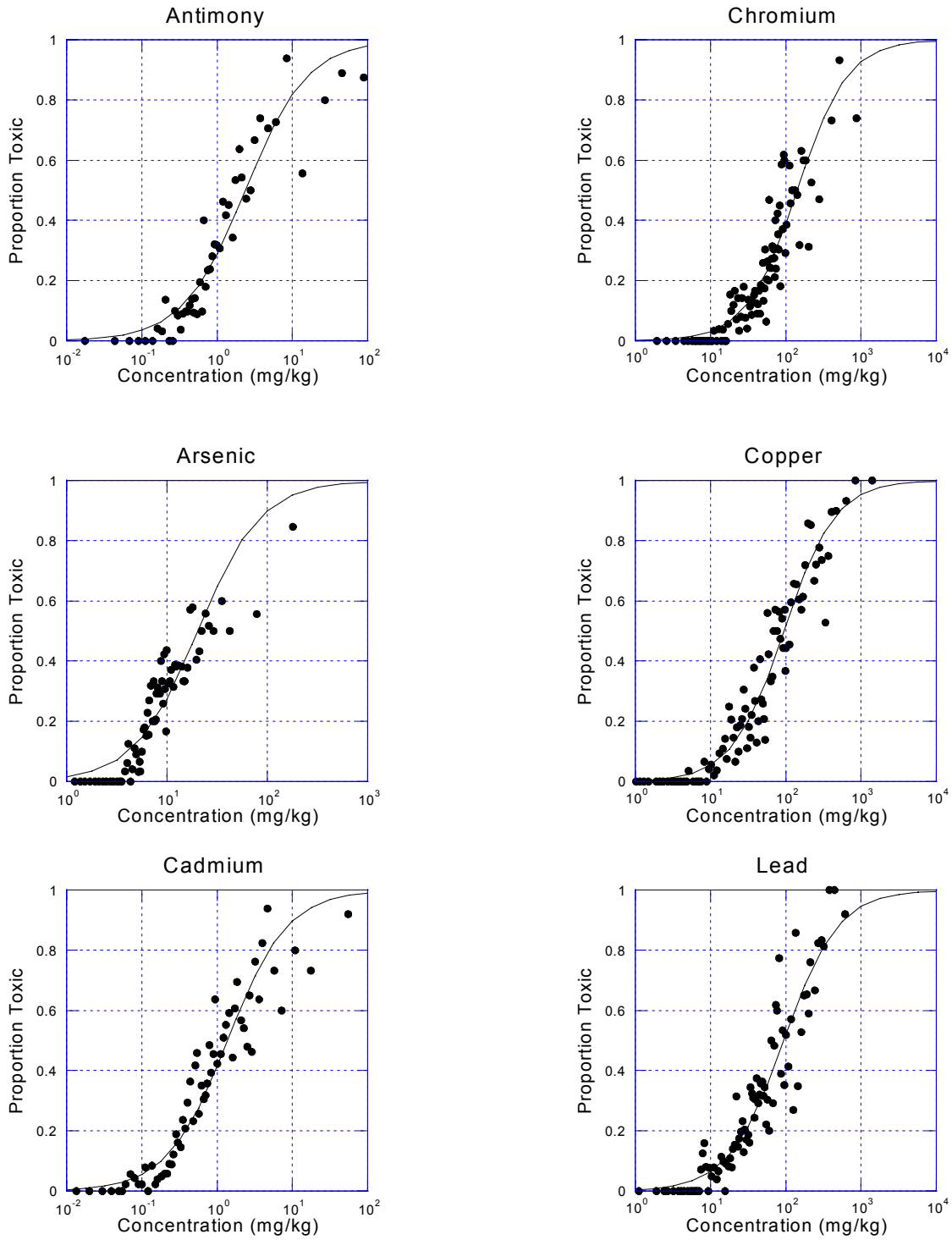


Figure 2. Logistic regression models and proportion of toxic samples in concentration intervals in screened marine amphipod database for 37 chemicals based on Sig Only classification of toxic samples. The individual points correspond to the median of the sample concentrations within the interval and the proportion of the samples that are toxic within the interval.

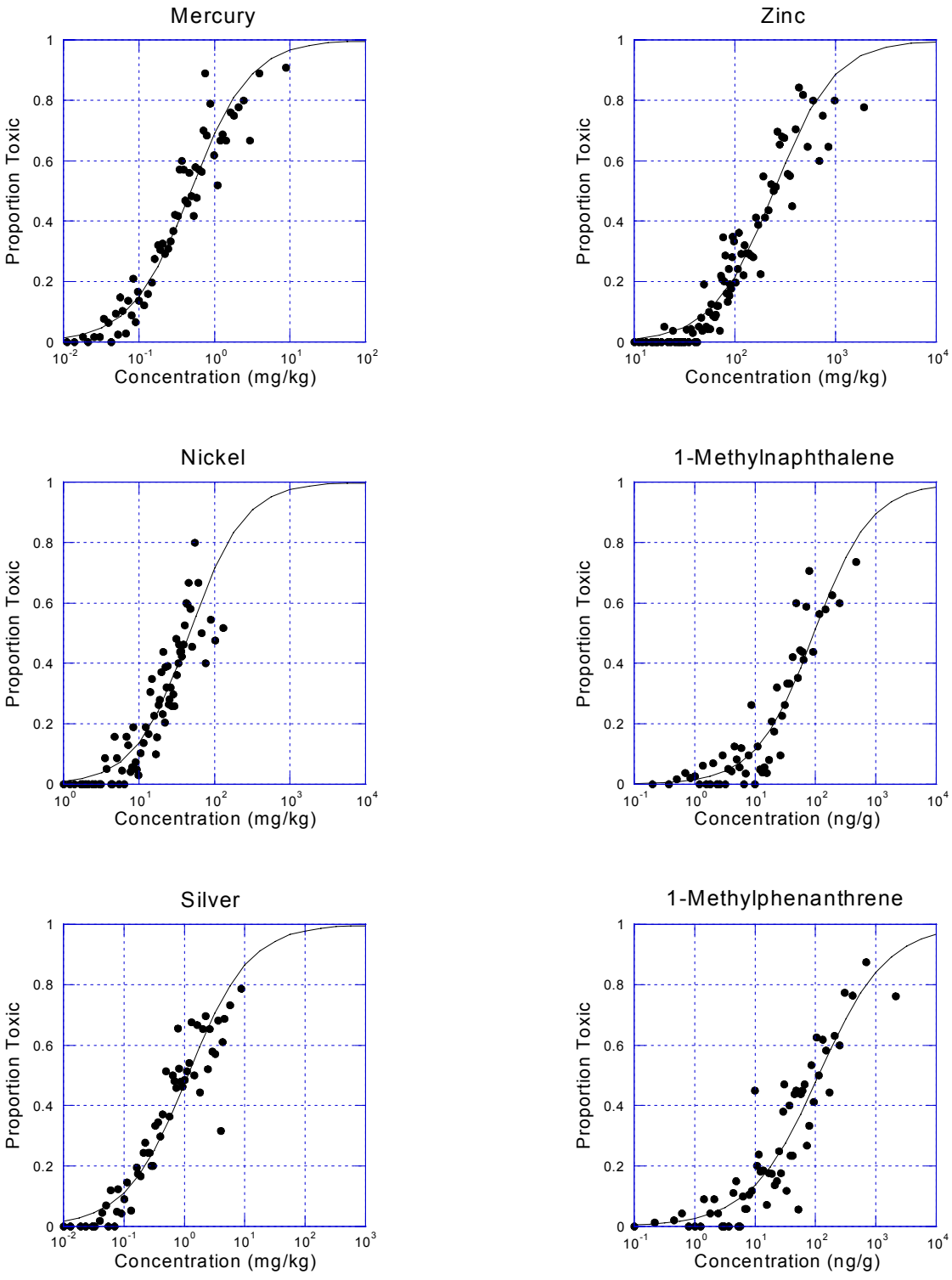


Figure 2. (continued)

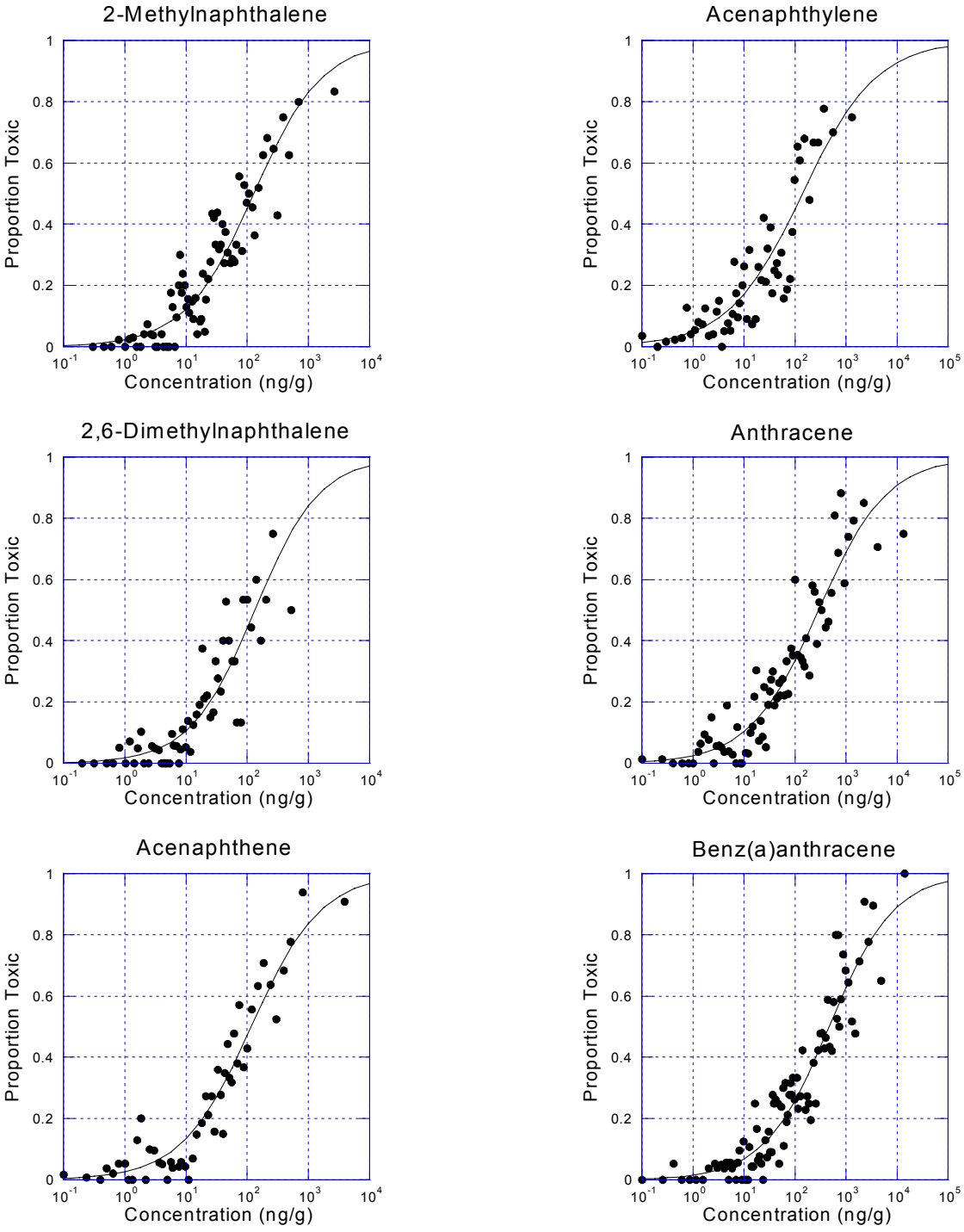


Figure 2. (continued)

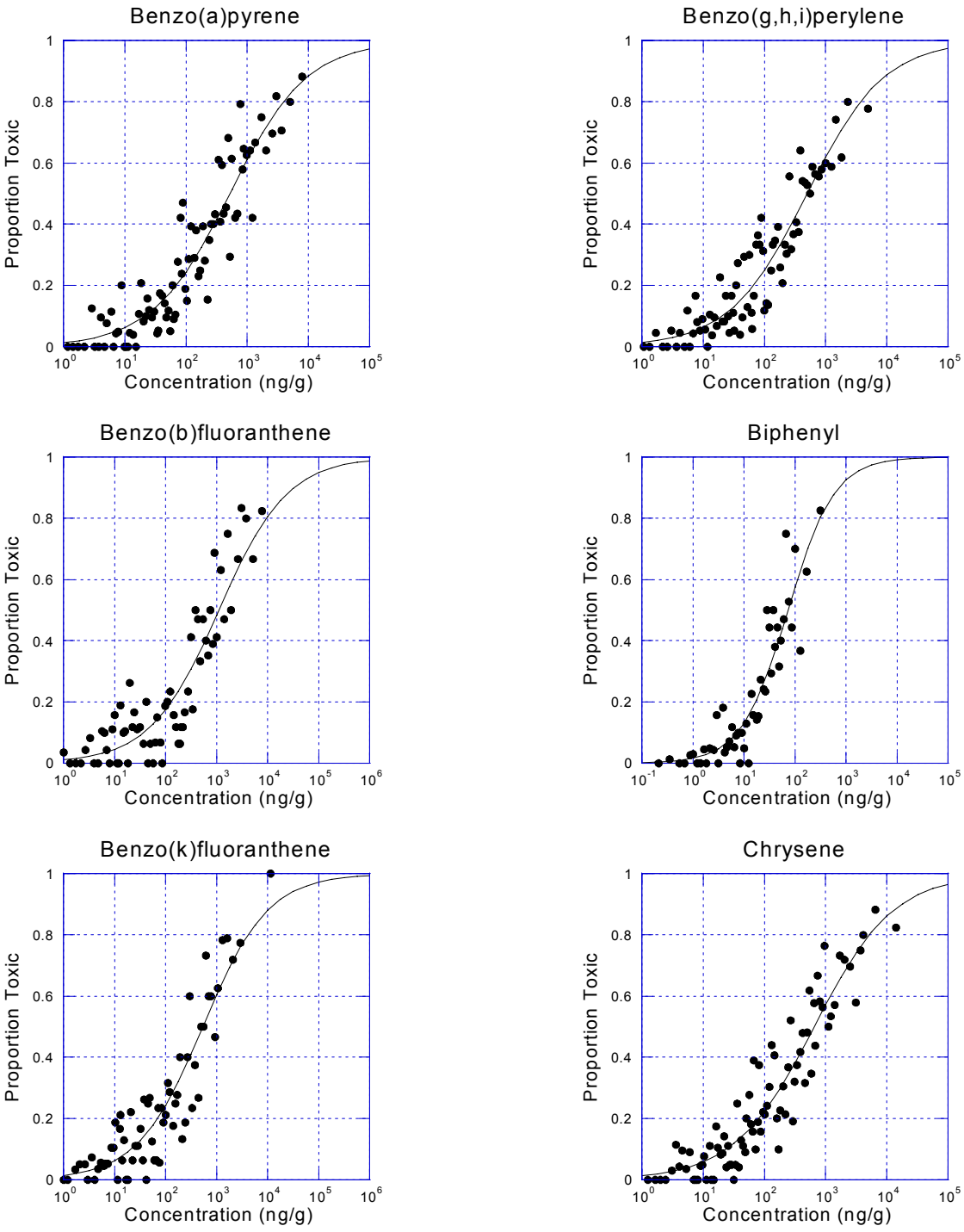


Figure 2. (continued)

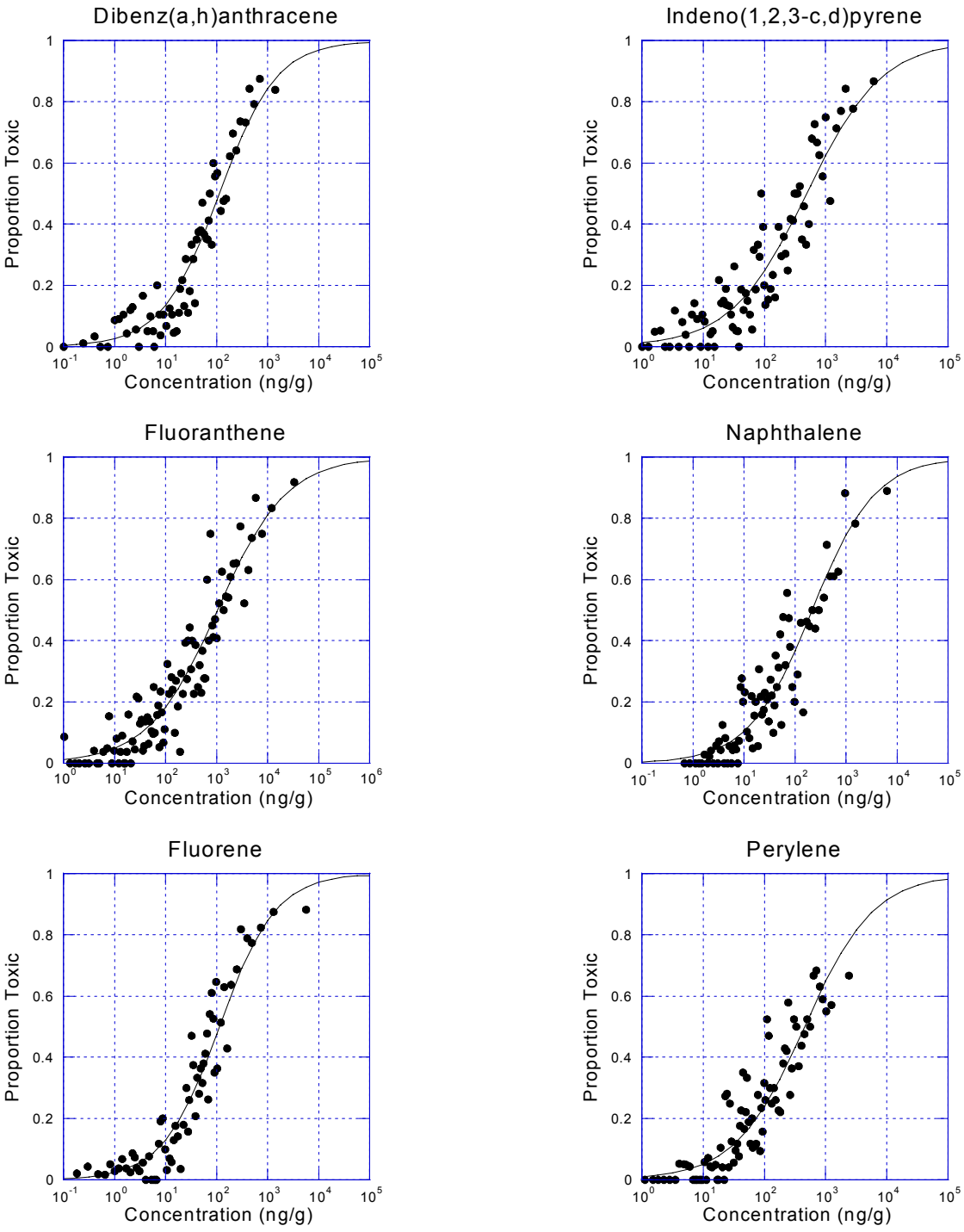


Figure 2. (continued)

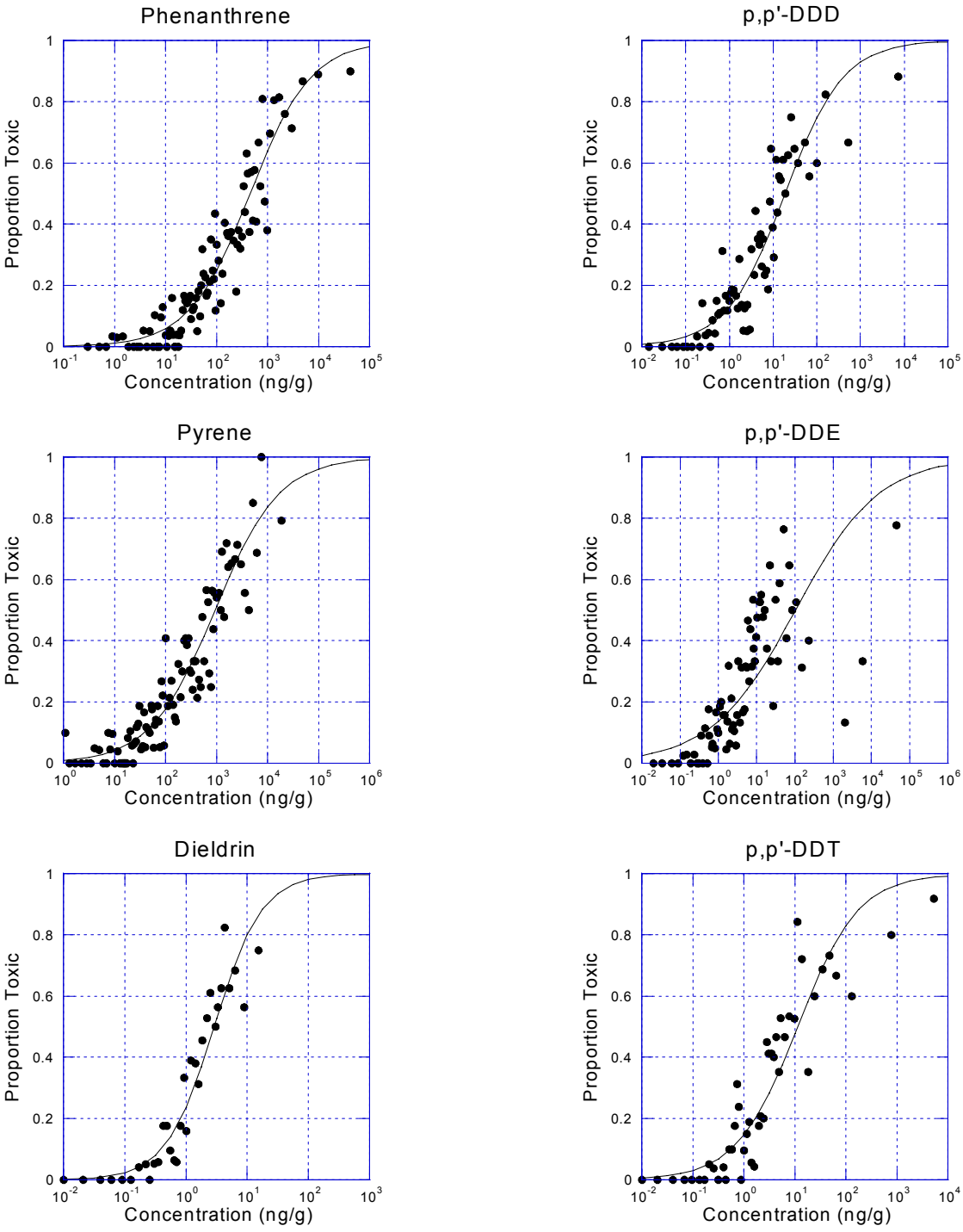


Figure 2. (continued)

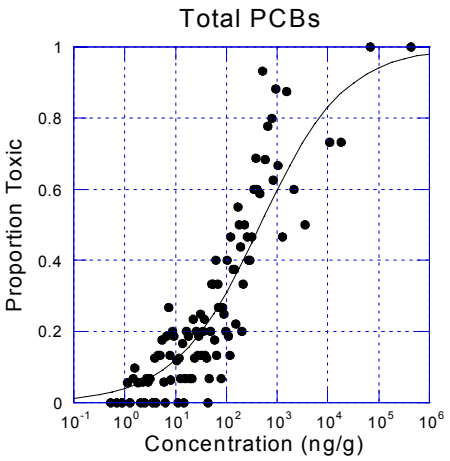


Figure 2. (continued).

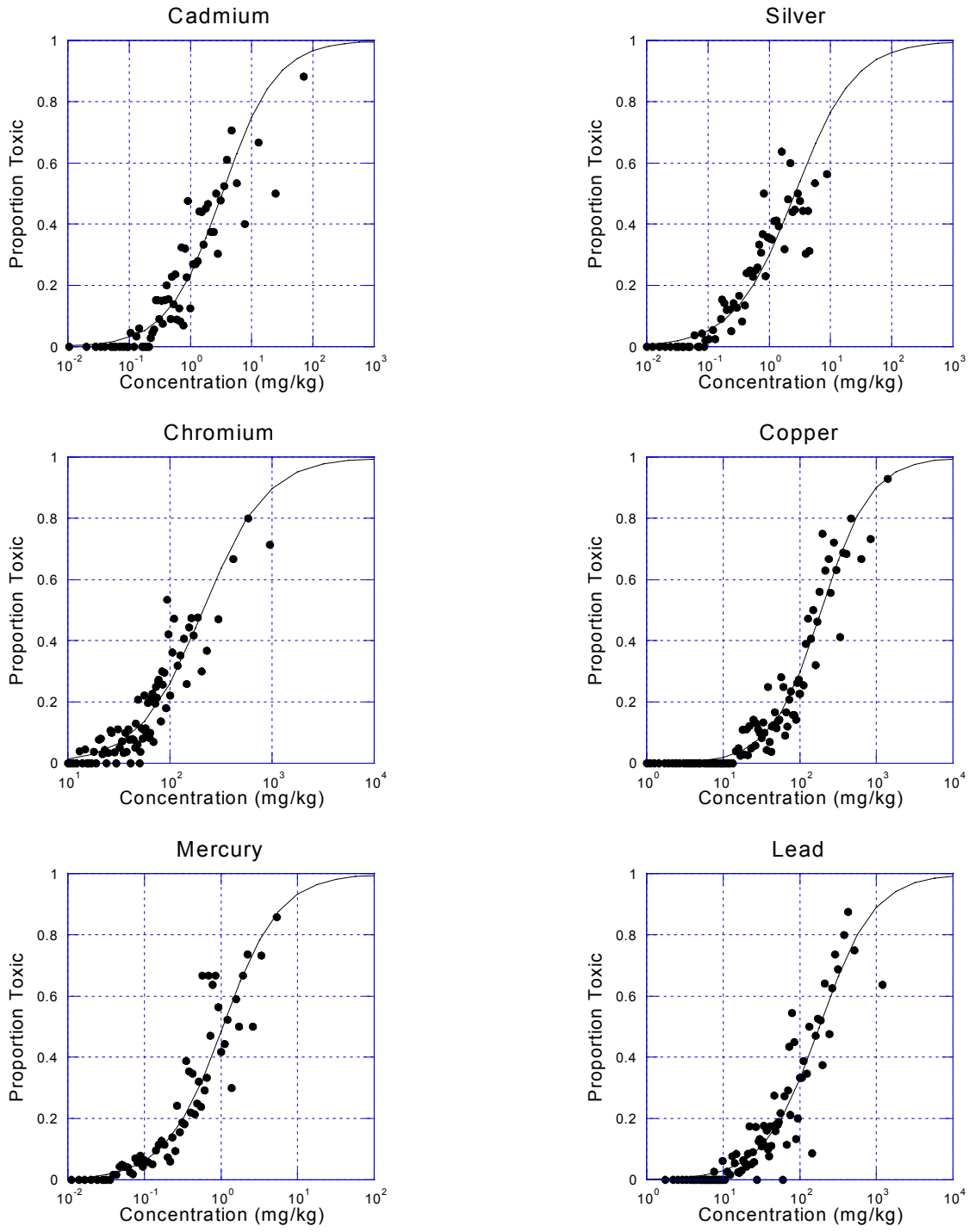


Figure 3. Logistic regression models and proportion of toxic samples in concentration intervals in the screened marine amphipod database for 33 chemicals based on MSD classification of toxic samples. The individual points correspond to the median of the sample concentrations within the interval and the proportion of the samples that are toxic within the interval.

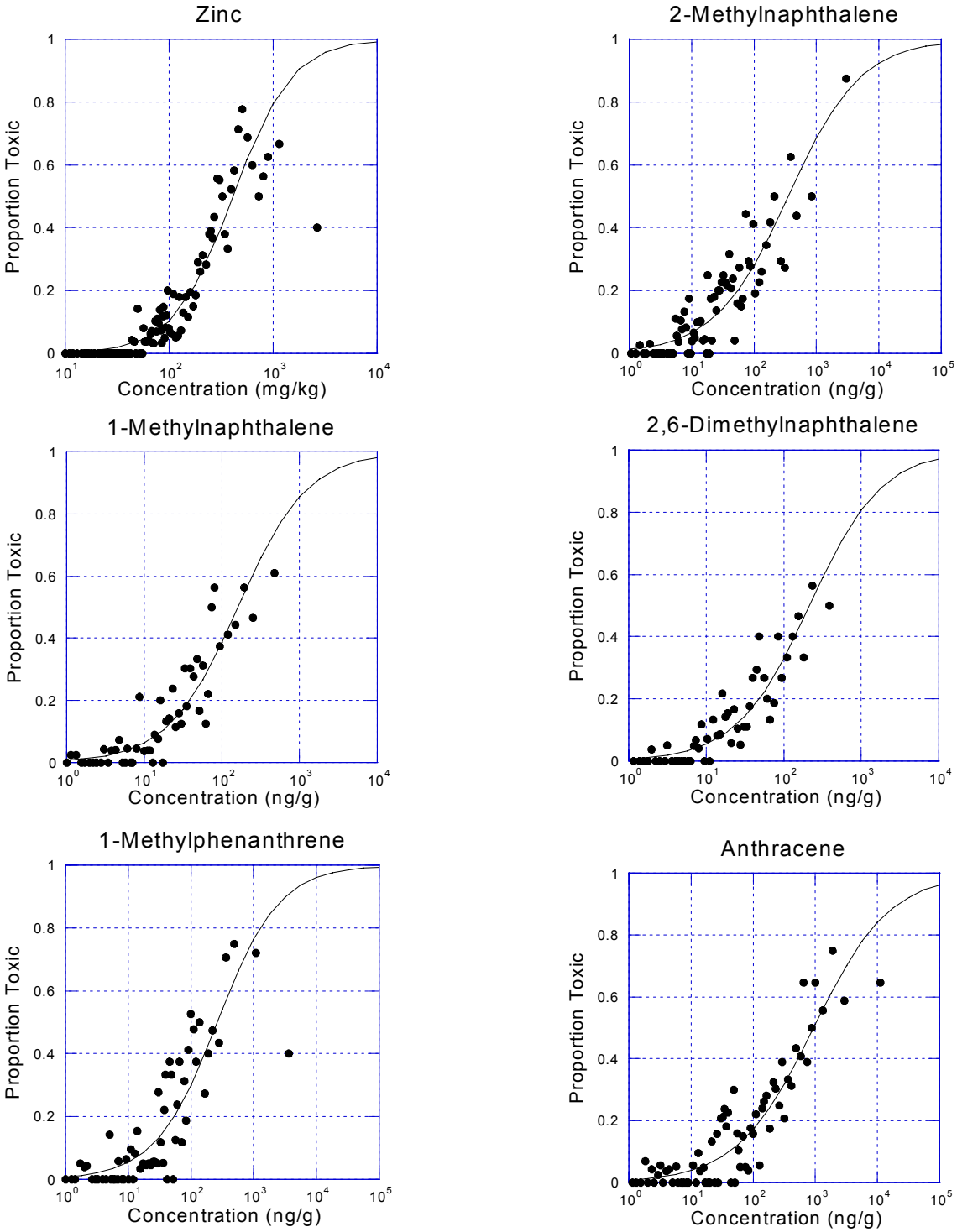


Figure 3. (continued)

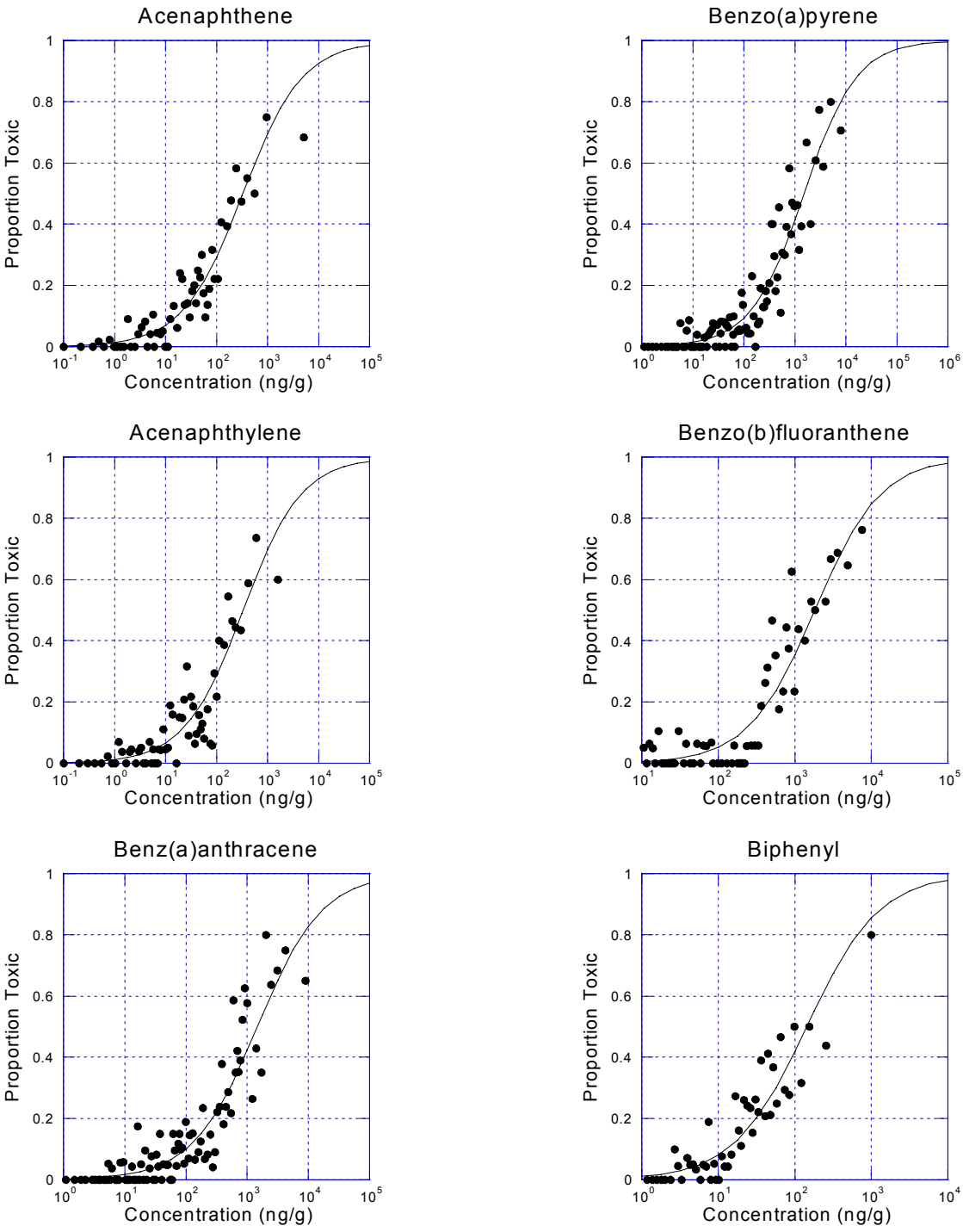


Figure 3. (continued)

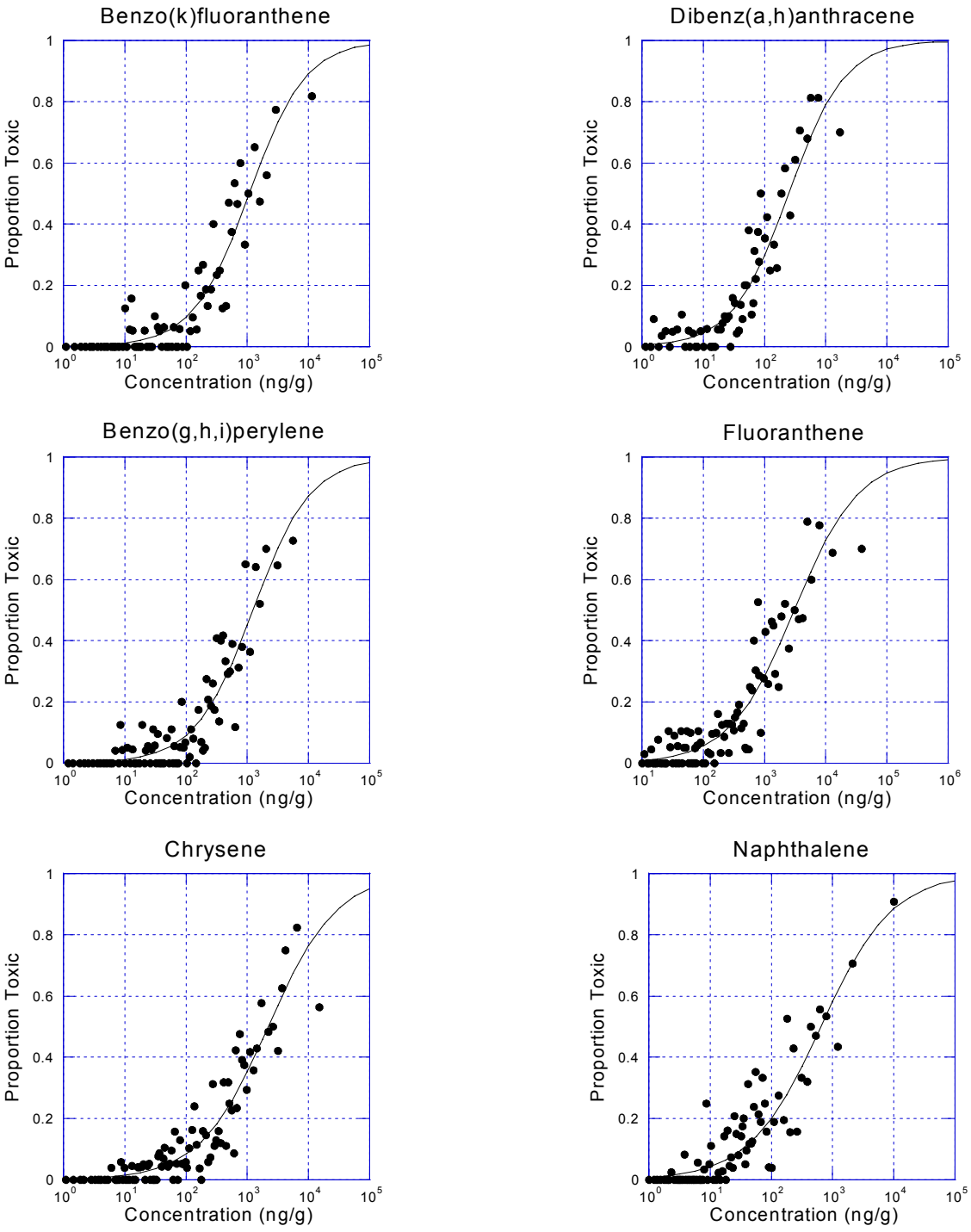


Figure 3. (continued)

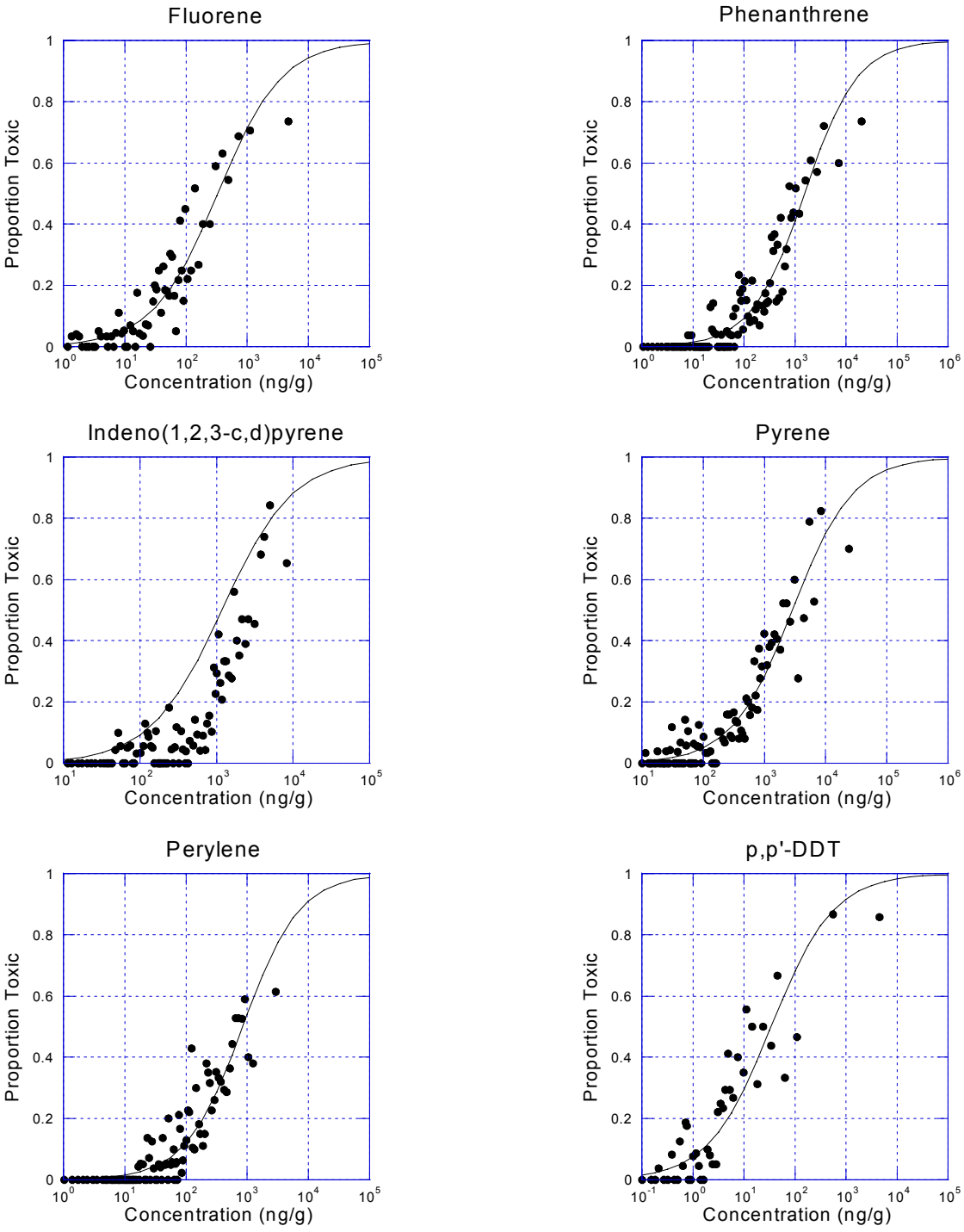
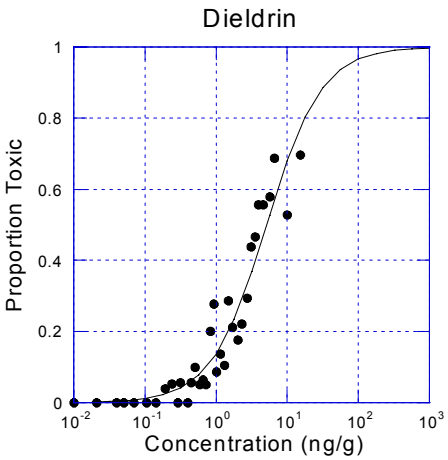
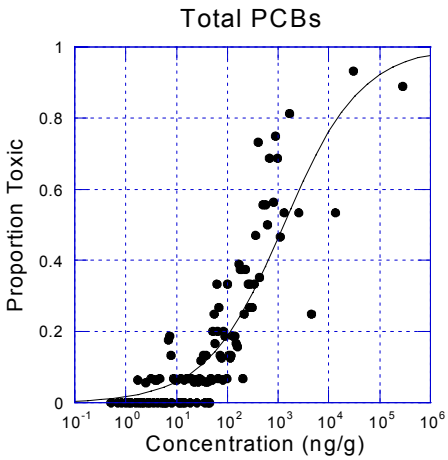


Figure 3. (continued)





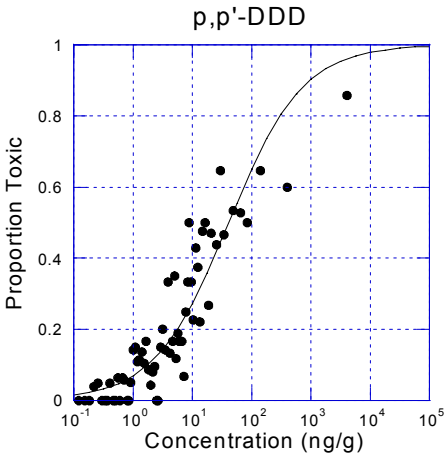


Figure 3. (continued).

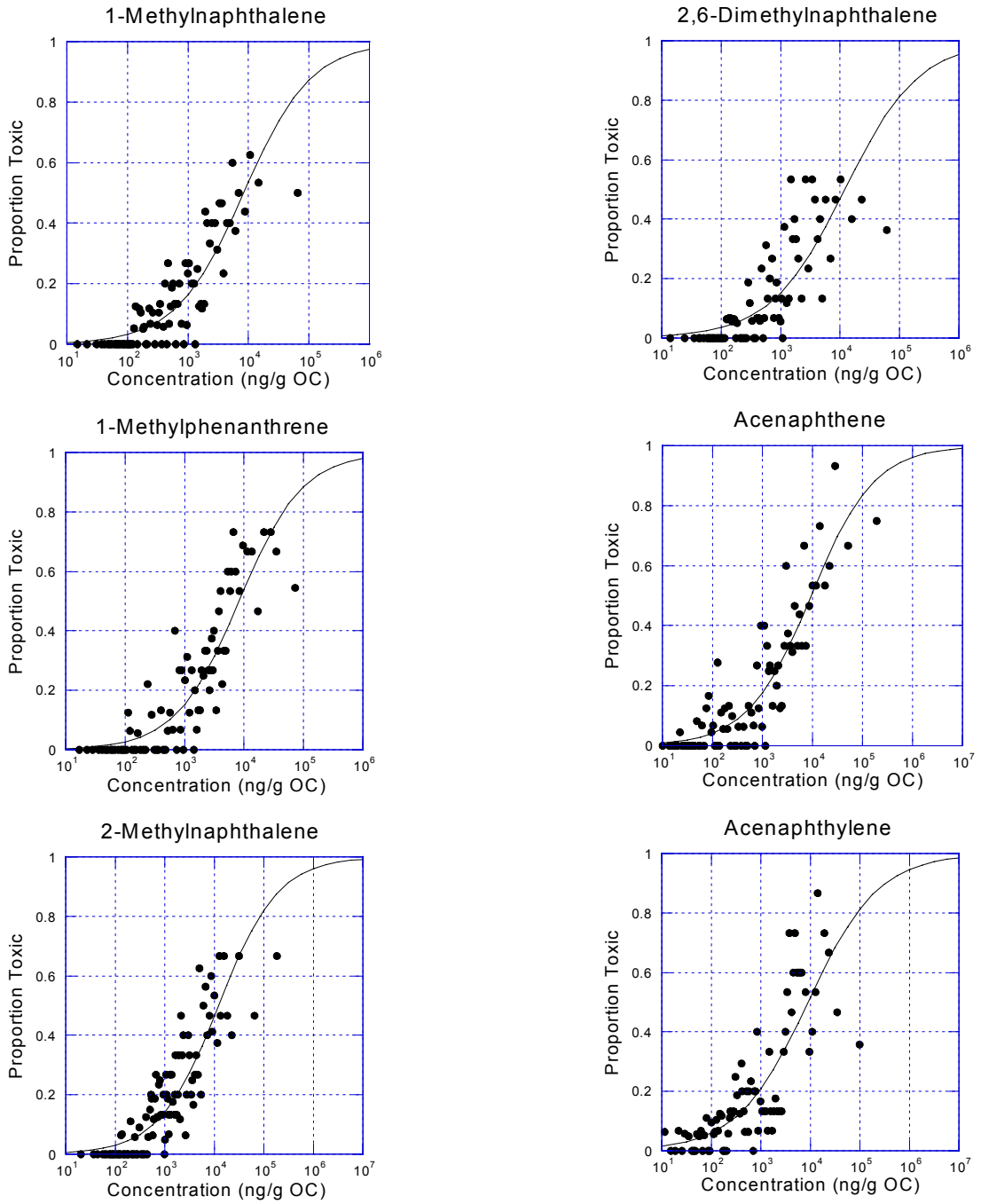


Figure 4. Logistic regression models and proportion of toxic samples in organic carbon-normalized concentration intervals in the screened marine amphipod database for 25 chemicals based on Sig Only classification of toxic samples. The individual points correspond to the median of the sample concentrations within the interval and the proportion of the samples toxic within the interval.

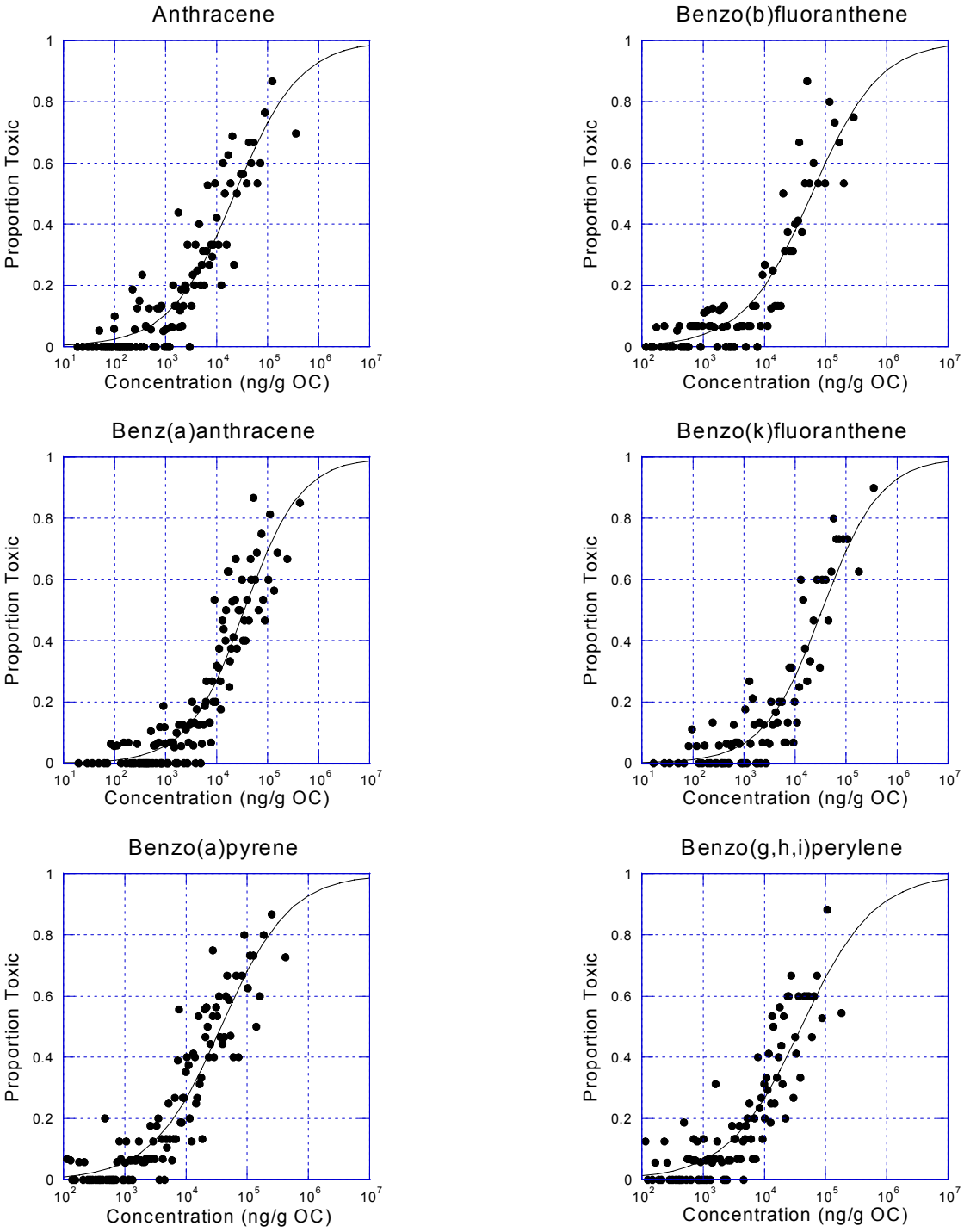


Figure 4. (continued)

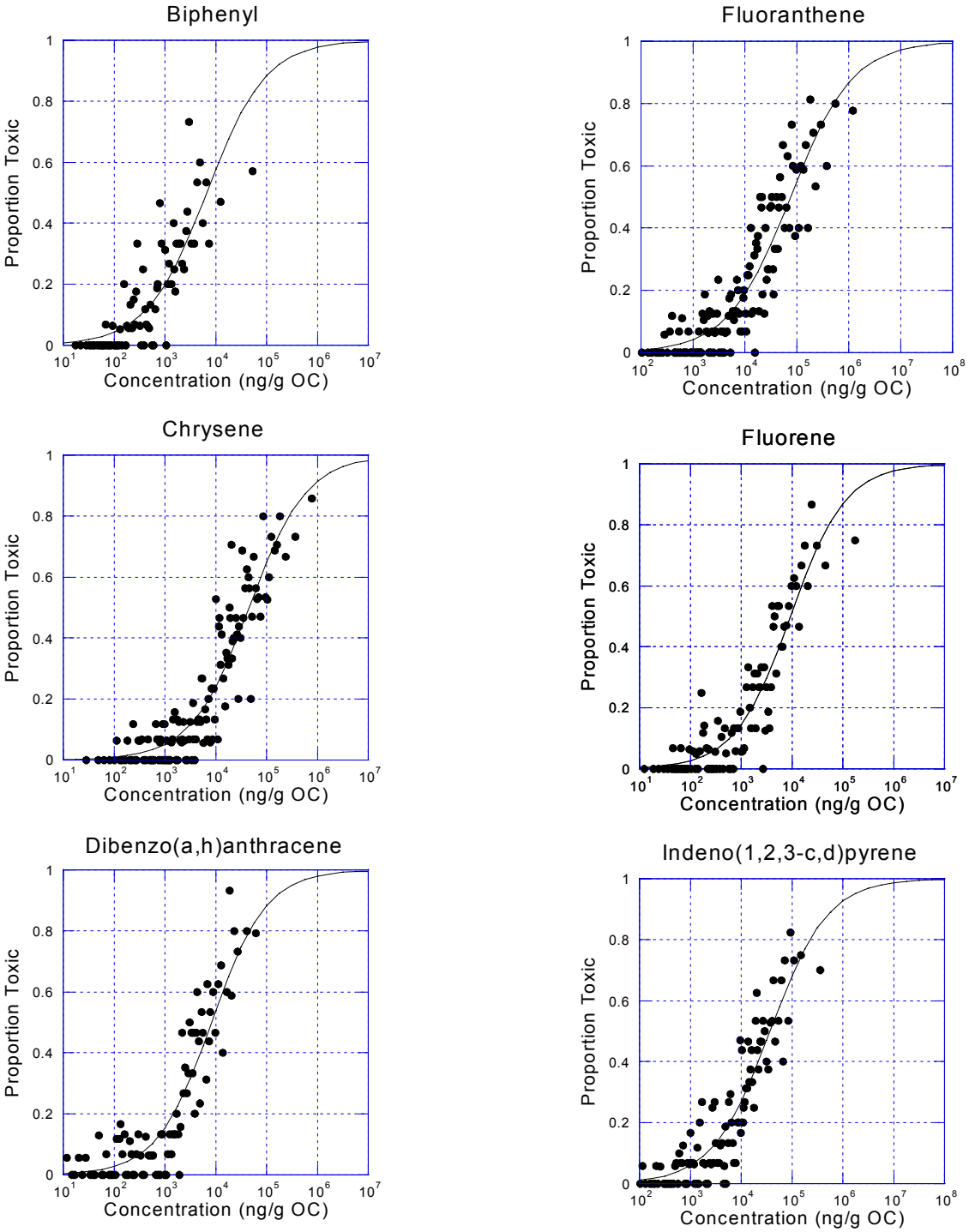


Figure 4. (continued)

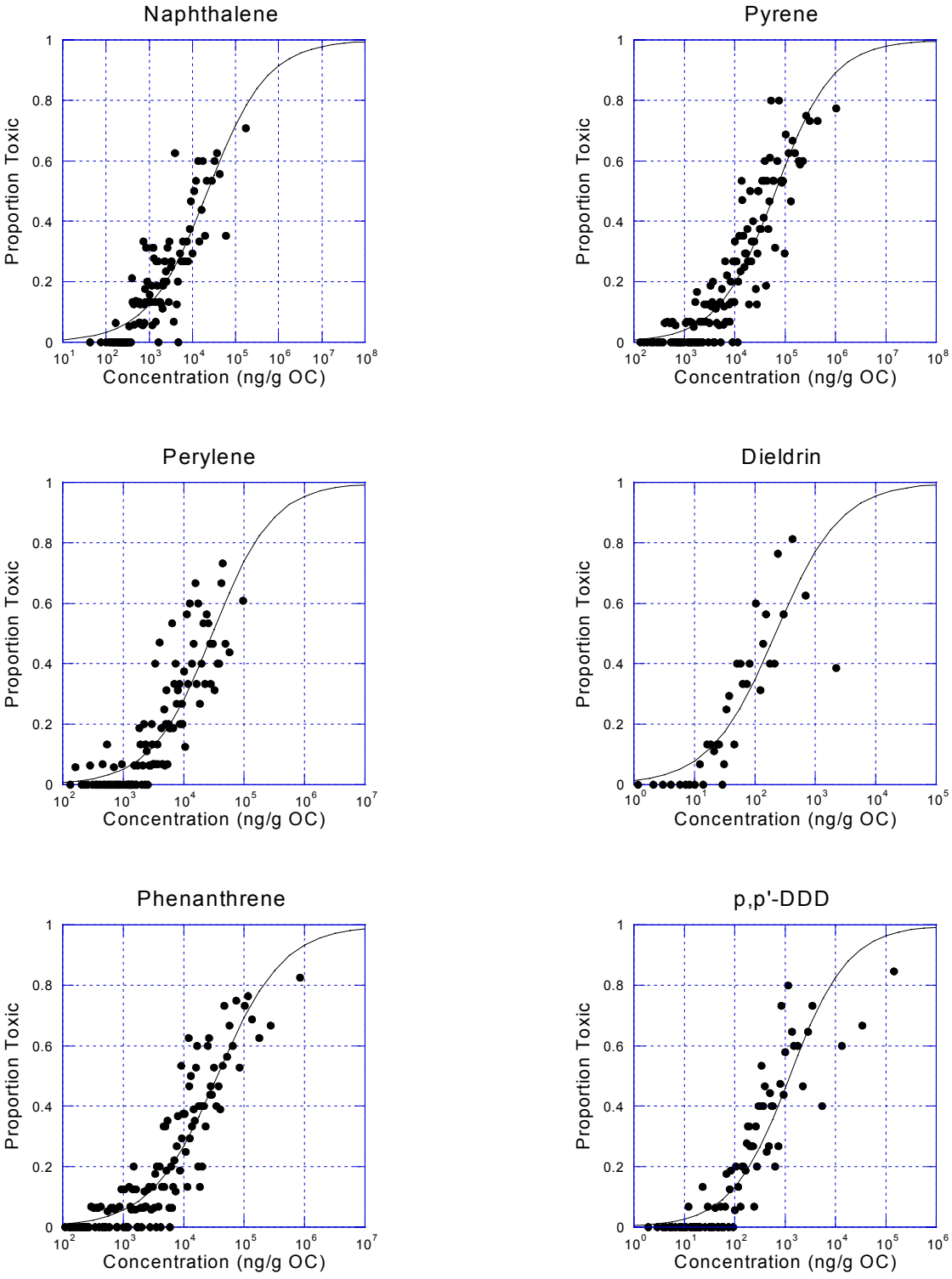
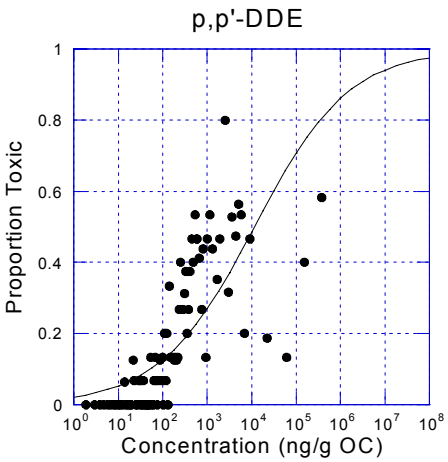
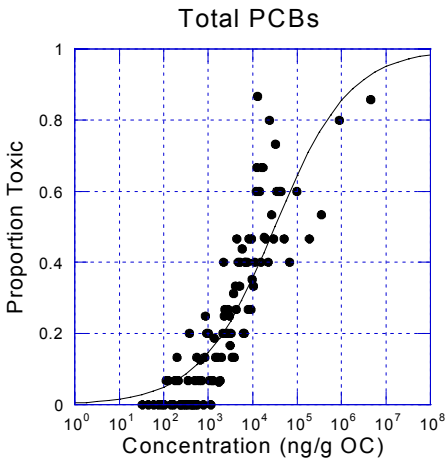


Figure 4. (continued)





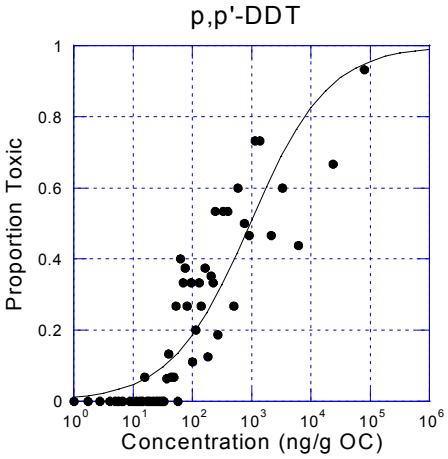


Figure 4. (continued).

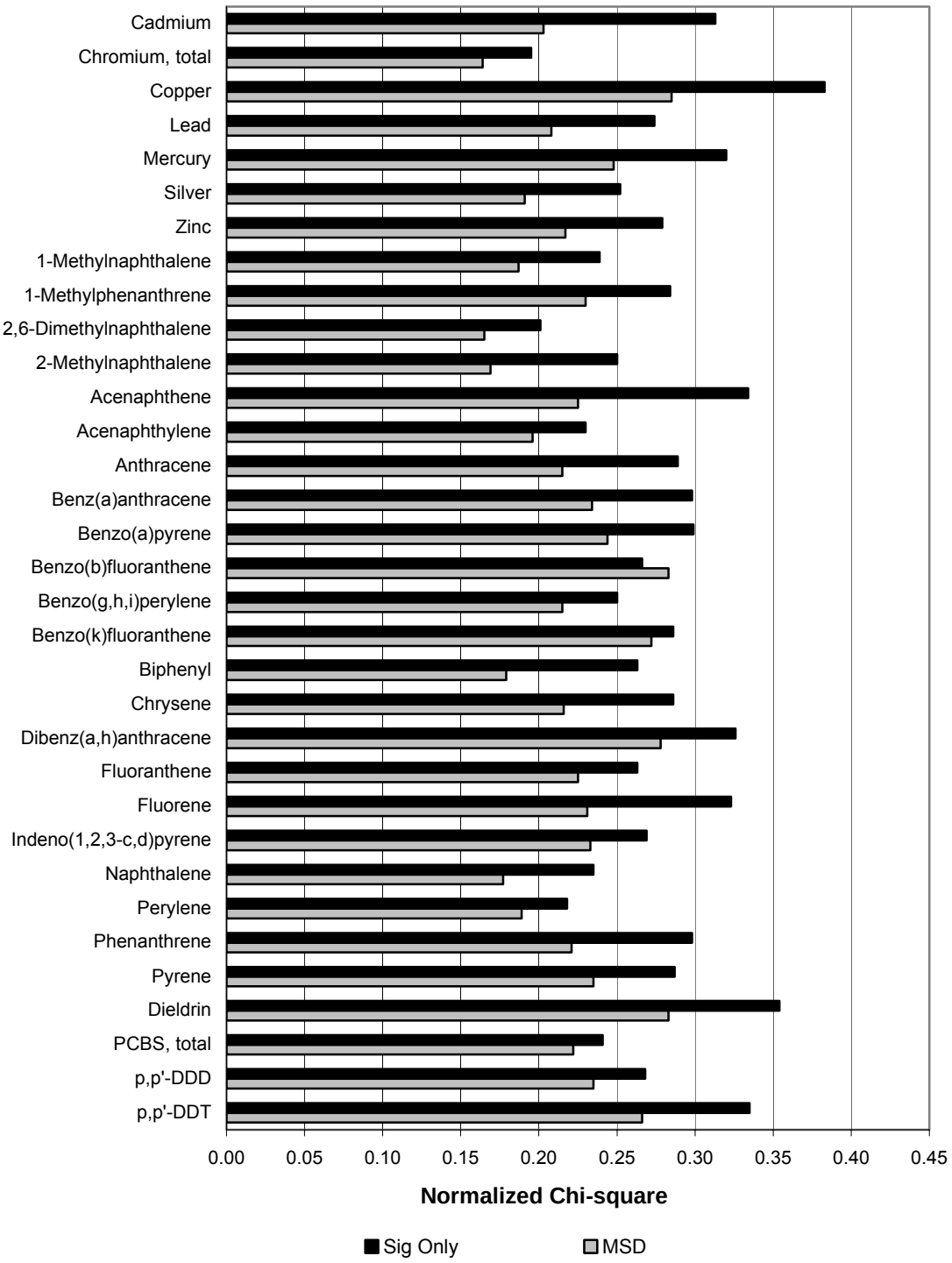


Figure 5. Comparison of logistic model goodness of fit for the marine amphipod survival endpoint with different toxicity classifications: Sig Only versus MSD.

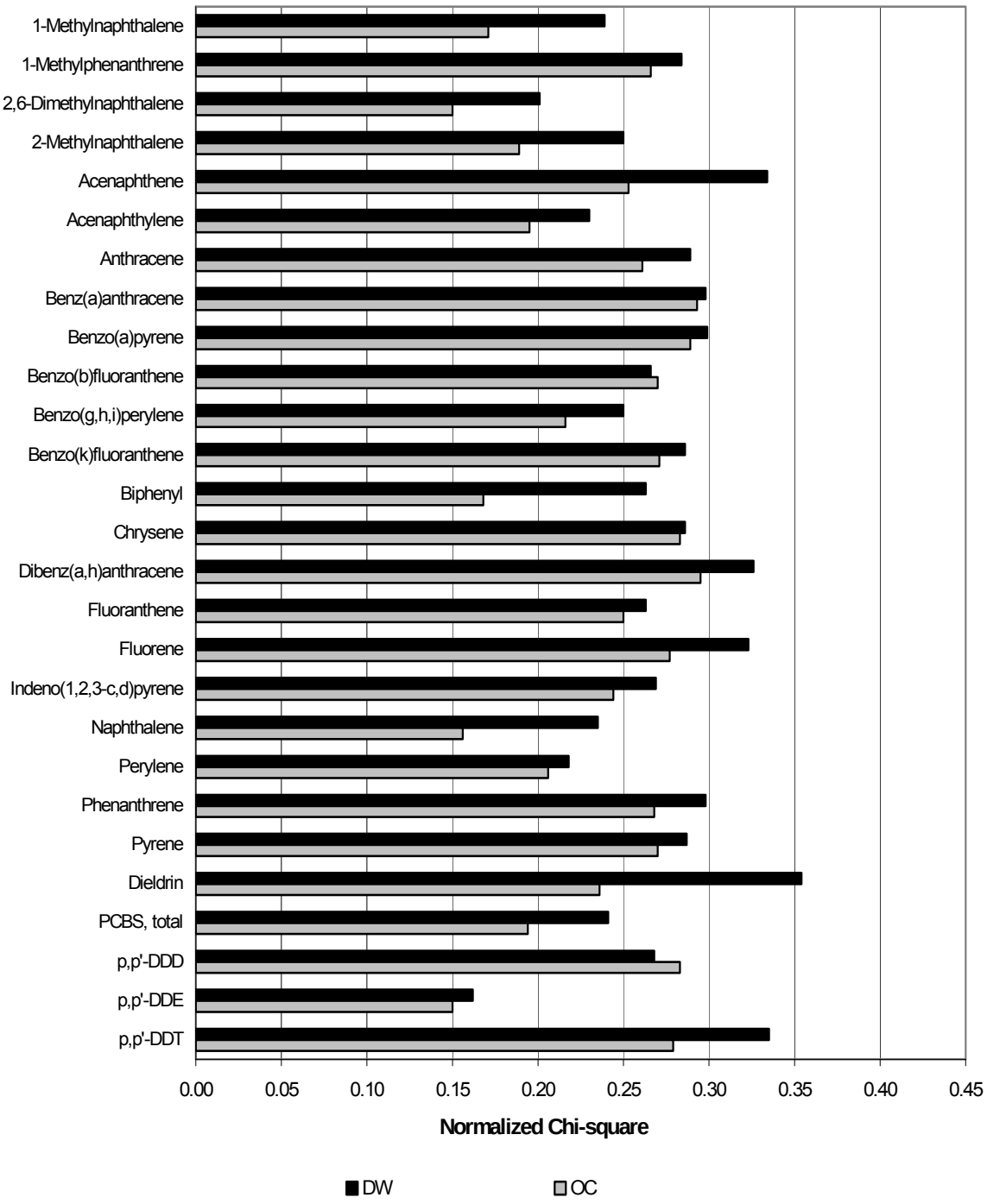


Figure 6. Comparison of logistic model goodness of fit for the marine amphipod endpoint survival with different approaches to the expression of chemical concentrations: dry weight (DW) versus organic carbon-normalized (OC) concentrations. Both approaches use the Sig Only classification of toxic samples.

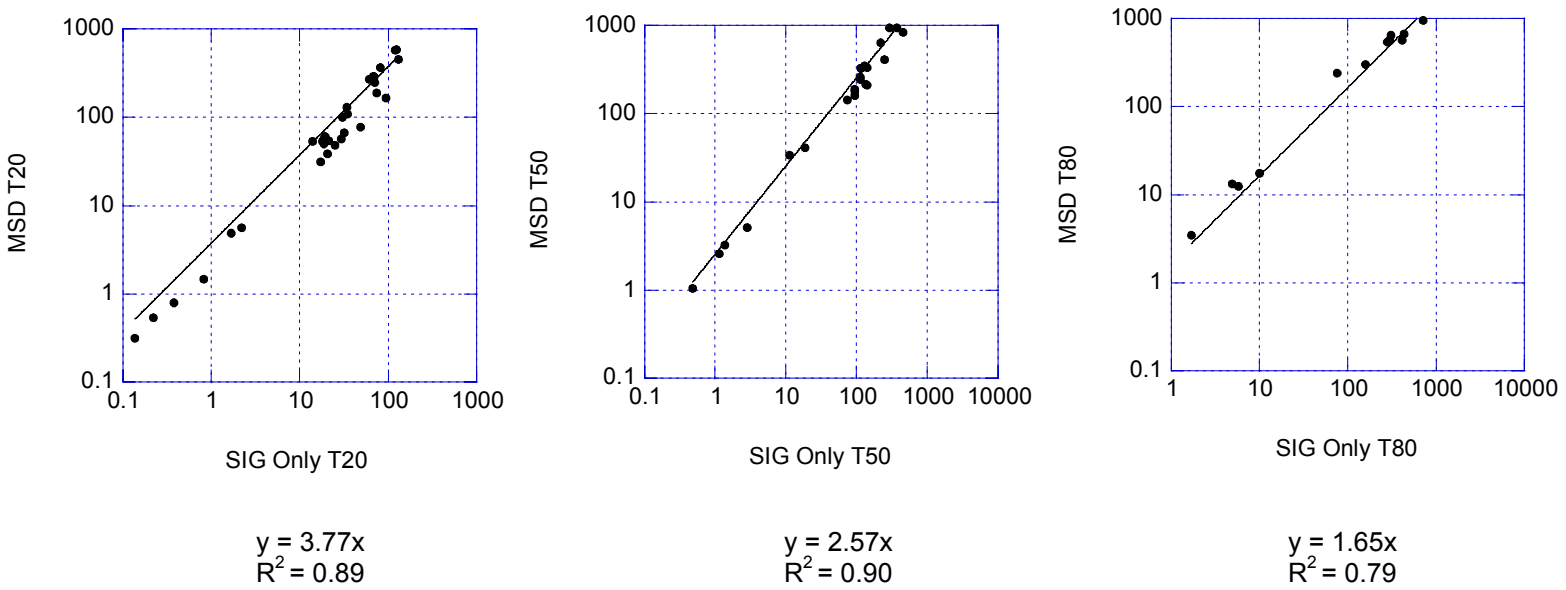


Figure 7. Comparison of logistic regression model Tp values (T20, T50, T80) for Sig Only and MSD toxicity classification approaches for the marine amphipod database.

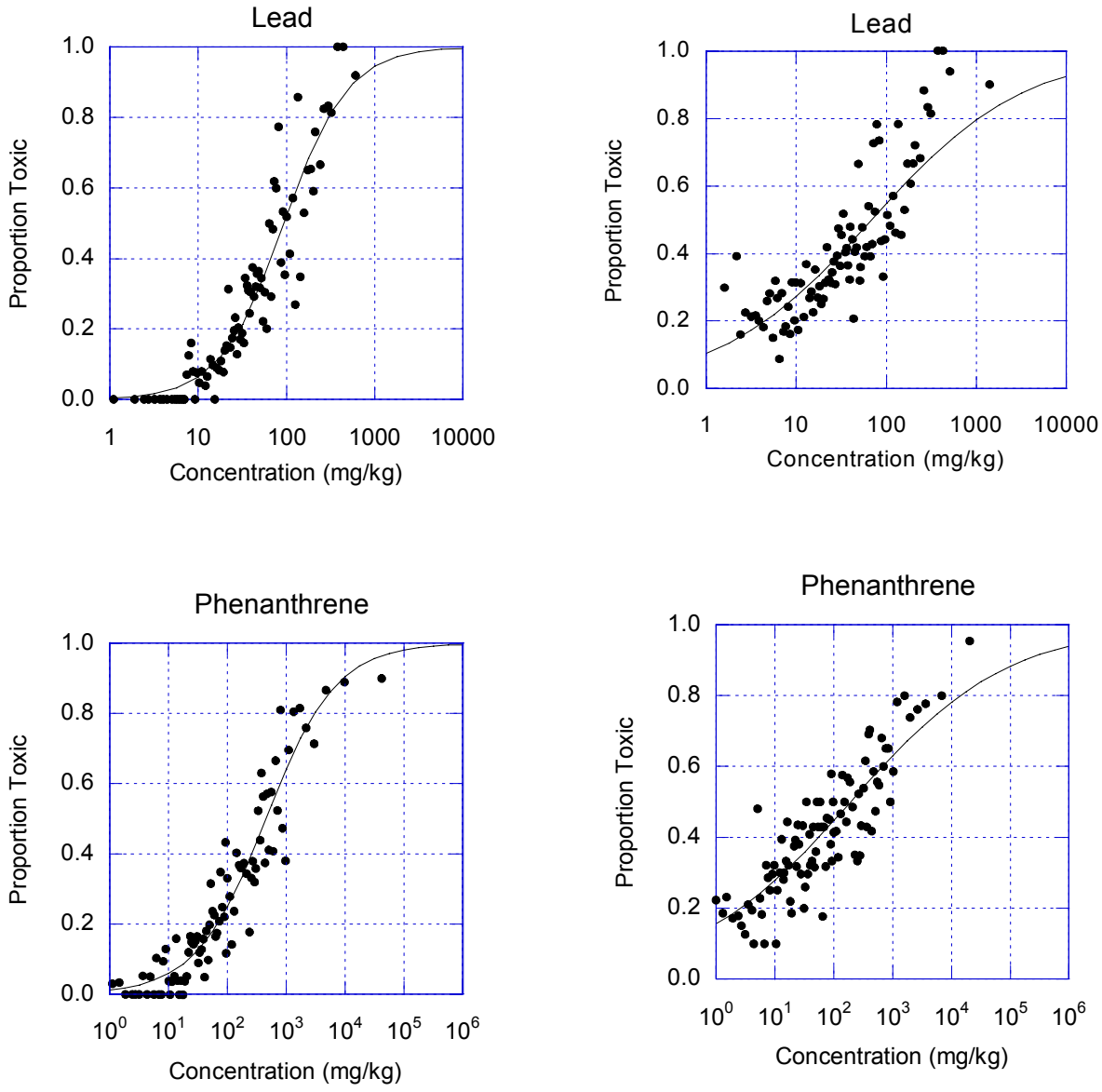


Figure 8. Comparison of 1X mean screened (left) and unscreened (right) logistic regression models and proportion of samples toxic in concentration intervals for lead and phenanthrene using the marine amphipod database.

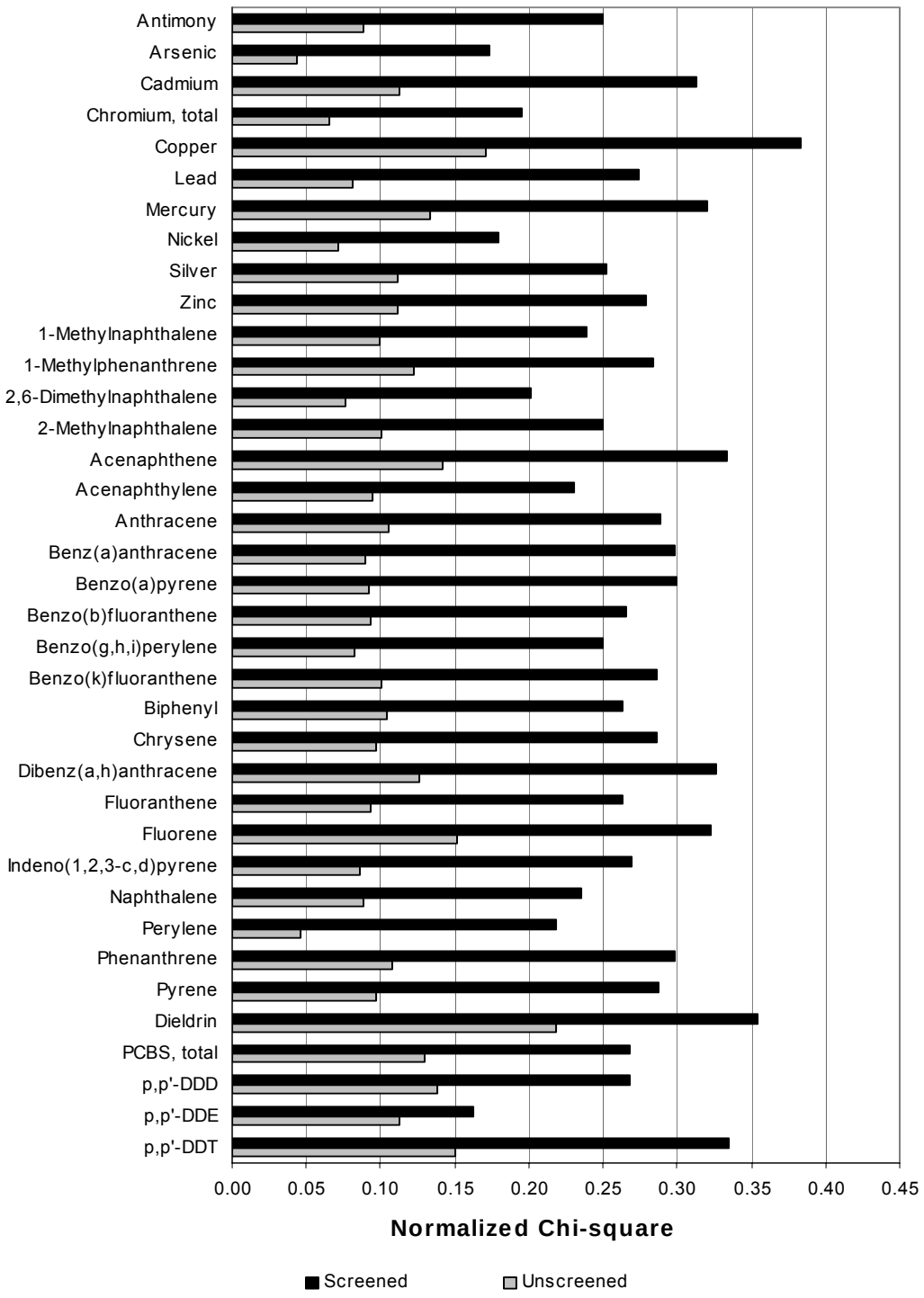


Figure 9. Comparison of logistic model goodness of fit for the marine amphipod survival endpoint using different screening methods (1X screening vs. no screening) and the Sig Only classification of toxic samples.

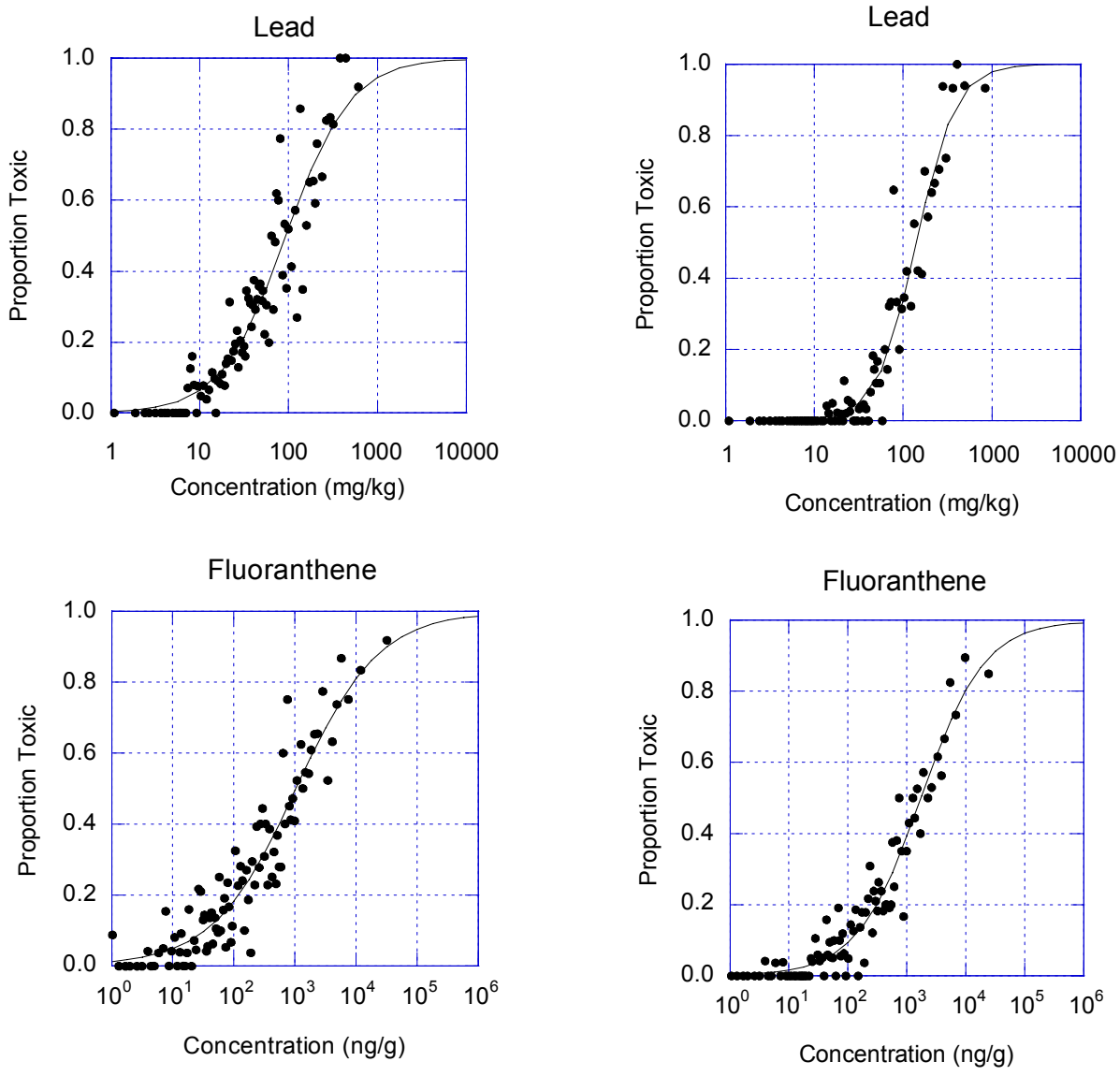


Figure 10. Logistic regression models and proportion of samples toxic in concentration intervals for the standard screening approach (1X mean) (left) and the 2X mean screening approach (right) for lead and fluoranthene using the marine amphipod database.

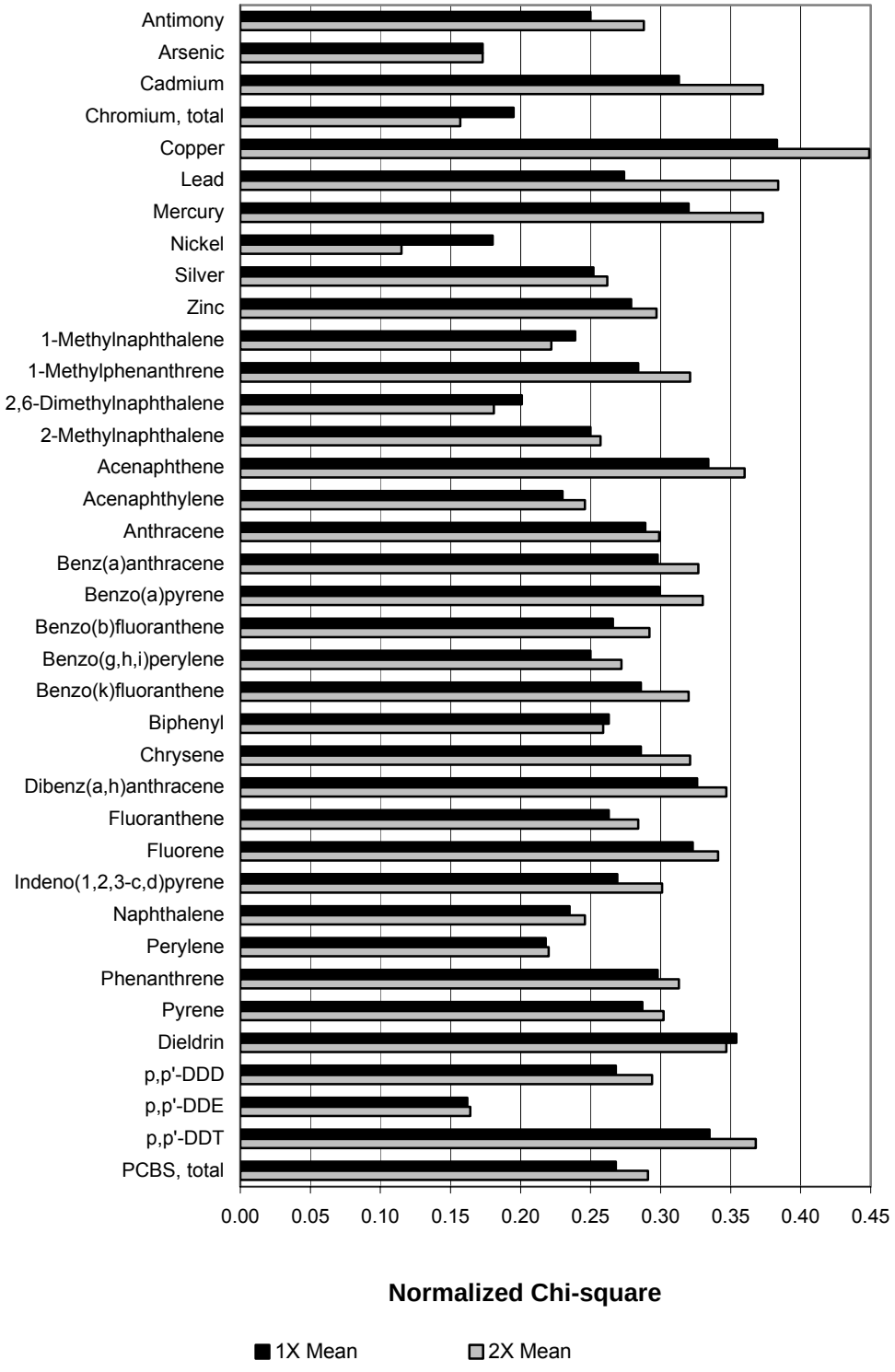


Figure 11. Comparison of logistic model goodness of fit for the marine amphipod survival endpoint using different screening methods (1X mean vs. 2X mean) and Sig Only classification of toxic samples.

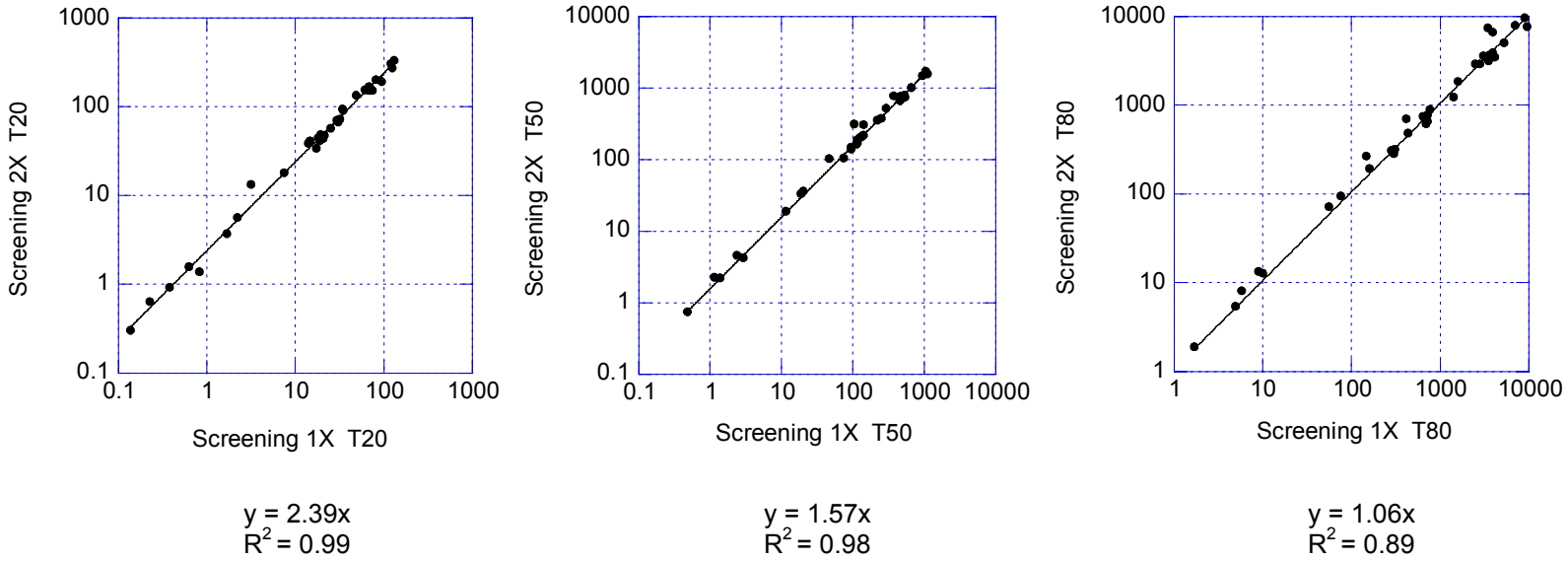


Figure 12. Ratio of logistic regression model Tp values (T20, T50, T80) for the standard screening approach (1X mean) to the 2X mean screening approach for the marine amphipod database.

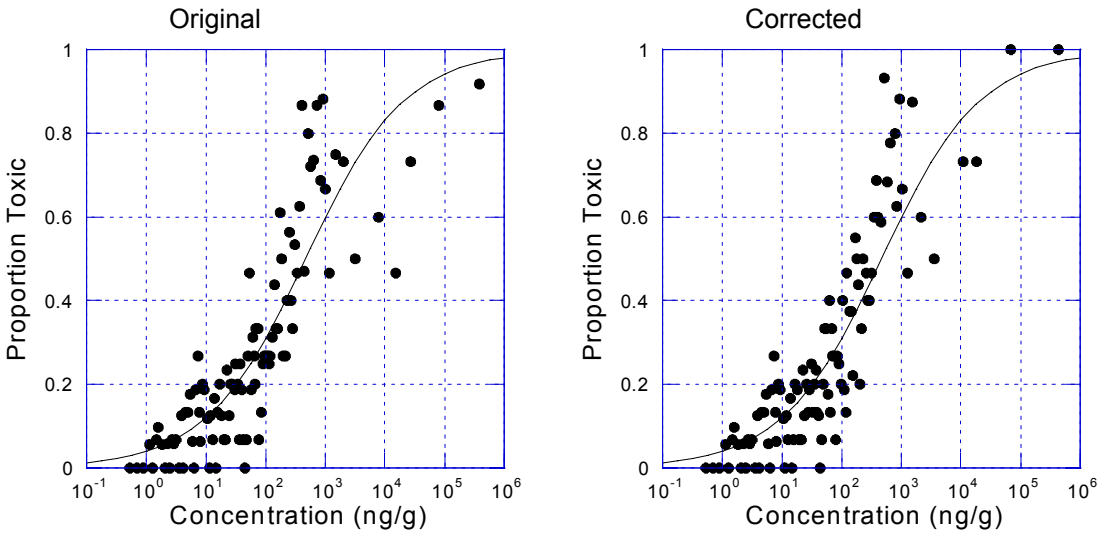


Figure 13. Logistic regression models and concentration interval plots showing the effect of correction in PCB units for 15 samples in the marine amphipod database.

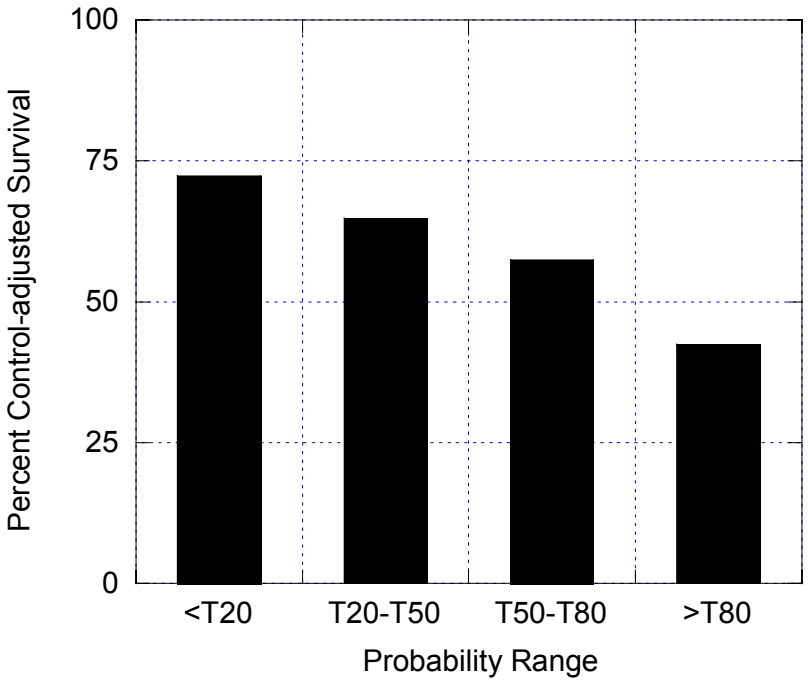


Figure 14. Mean percent control-adjusted marine amphipod survival for toxic samples within intervals defined by the Tp values for all individual chemical models.

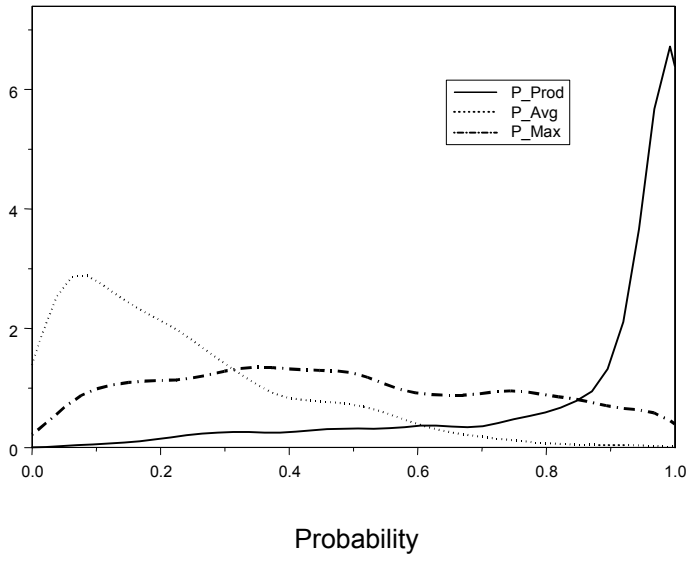


Figure 15. Probability density functions for P_Avg, P_Max, and P_Prod. The probability that a variable will have a value with a small interval around x can be approximated by multiplying the value of y at x by the width of the interval.

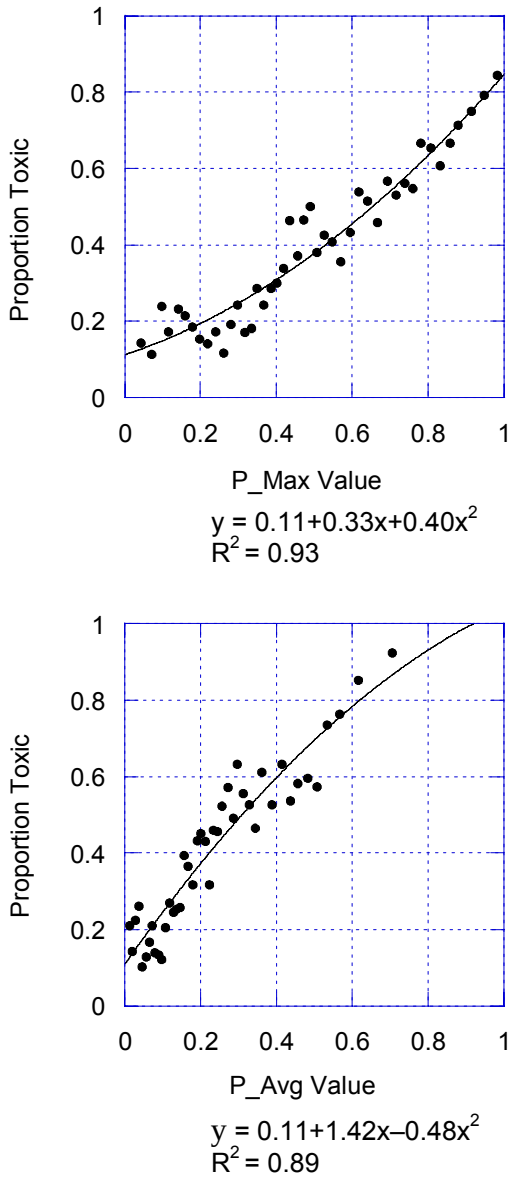


Figure 16. Proportion of observed toxic samples versus the P_Max and P_Avg value probability intervals based on Sig Only classification of toxic samples in the marine amphipod database. The regression lines are the P_Max and P_Avg models. Each point represents the median sample probability of a minimum of 50 individual samples within the interval and the proportion of the samples that were toxic within the interval. Only samples having measurements for 10 more modeled chemicals were used (n = 2856).

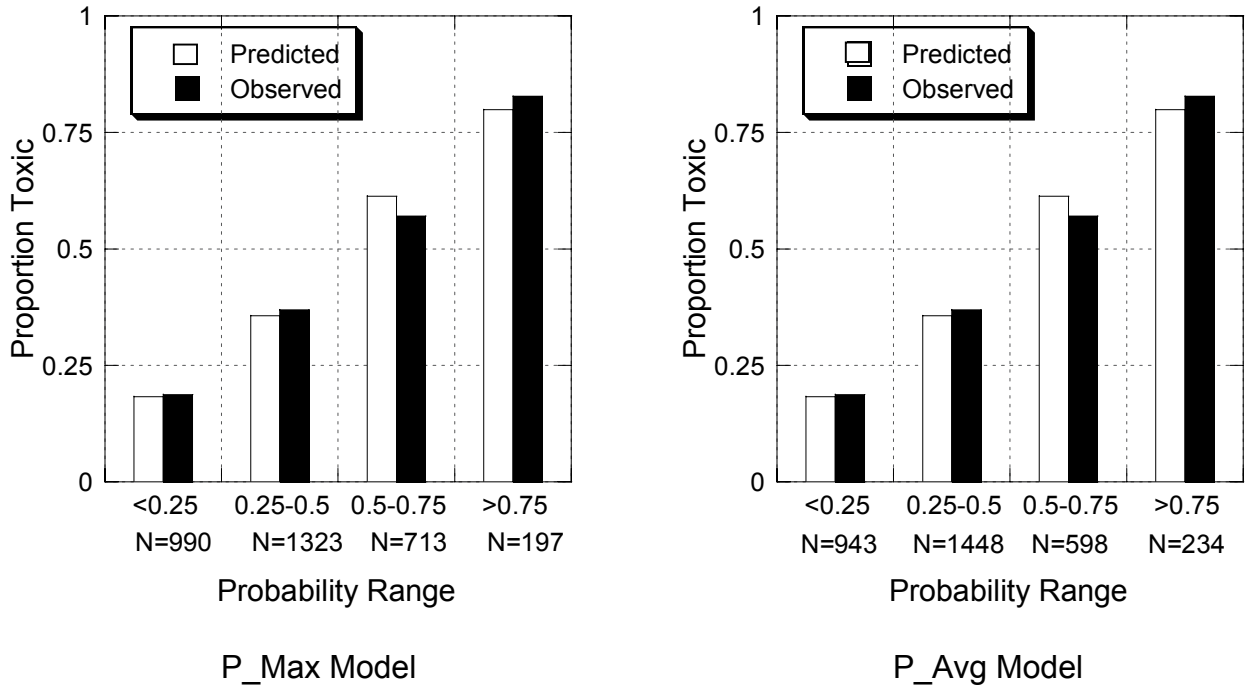


Figure 17. Mean predicted and observed proportion of toxic samples within probability quartiles for P_Max and P_Avg models derived from the marine amphipod database.

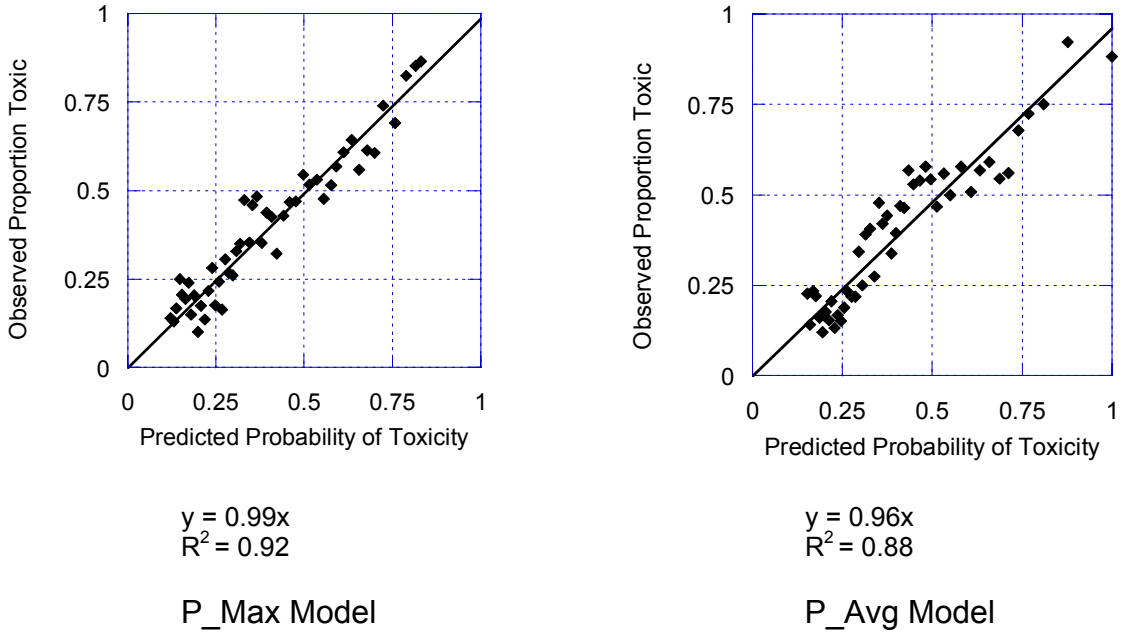


Figure 18. Observed proportion of toxic samples versus the probability of toxicity predicted using P_Max and P_Avg models derived from the marine amphipod database. Each point represents the median sample probability of minimum of 50 individual samples within the interval and the mean percent survival within the interval (n = 3223).

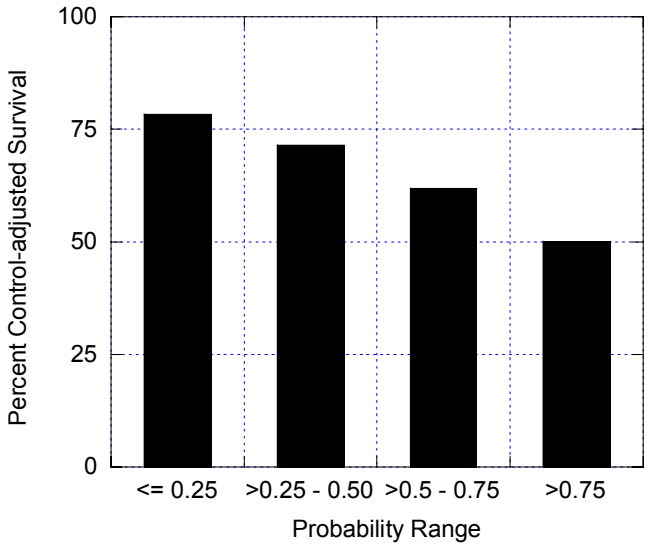


Figure 19. Mean percent control–adjusted survival for toxic samples within probability quartile intervals for P_Max model derived from the marine amphipod database.

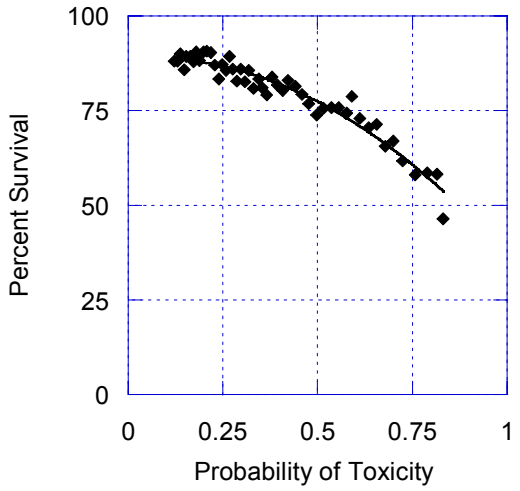


Figure 20. Mean control-adjusted percent survival for samples within probability intervals for P_Max model (R^2 = 0.94) derived from the marine amphipod database. Each point represents the median sample probability of a minimum of 50 individual samples within the interval and the mean percent control-adjusted survival within the interval (n = 3223).

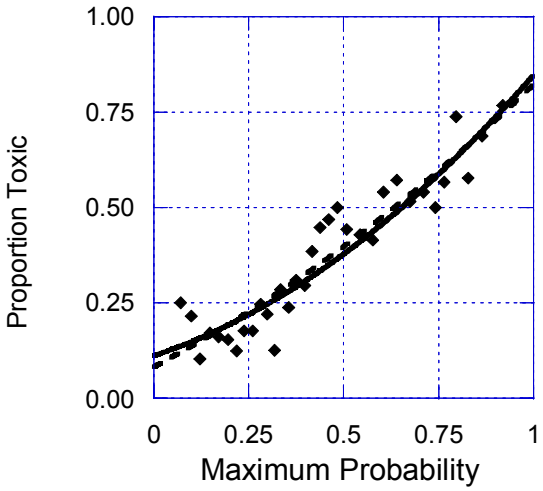


Figure 21. Comparison of the original marine amphipod P_Max model with the model derived from data excluding polycyclic aromatic hydrocarbon chemistry (R^2 = 0.88). Data points represent the proportion of toxic samples within unique probability intervals (minimum of 50 samples per interval) and the dotted line is the model derived from the interval plot. The solid line is the original P_Max model.

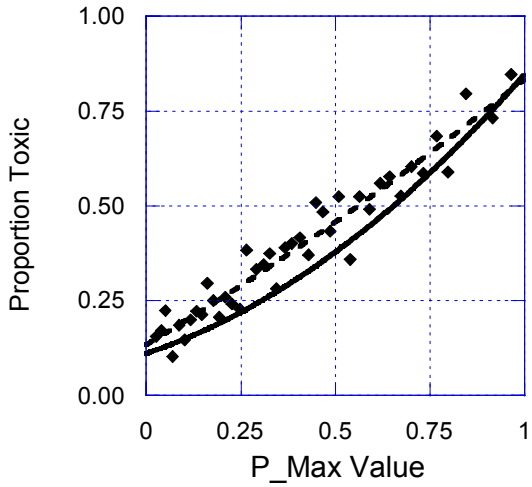


Figure 22. Comparison of the original marine amphipod P_Max model with model derived from data excluding metals chemistry (R^2 = 0.93). Data points represent the proportion of toxic samples within unique probability intervals (minimum of 50 samples per interval) and the dotted line is model derived from the interval plot. The solid line is the original P_Max model.

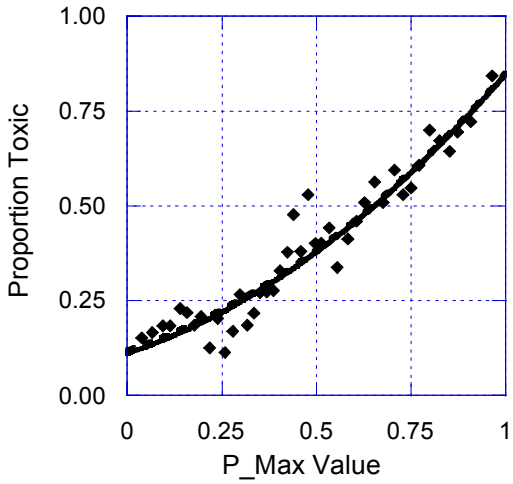


Figure 23. Comparison of the original marine amphipod P_Max model with model derived from data excluding pesticides and PCB chemistry (R^2 = 0.92). Data points represent the proportion of toxic samples within unique probability intervals (minimum of 50 samples per interval) and the dotted line is model derived from the interval plot. The solid line is the original P_Max model.

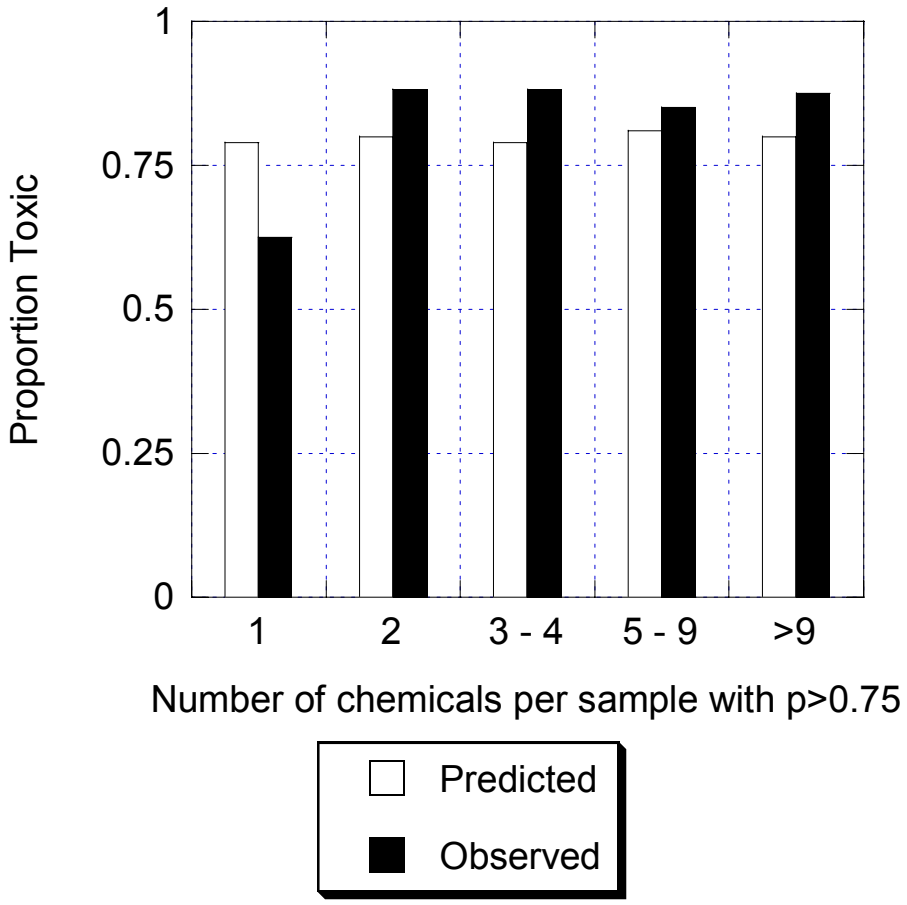


Figure 24. Mean predicted and observed proportion of toxic samples for samples with a predicted probability of >0.75 from the P_Max model in the marine amphipod database.

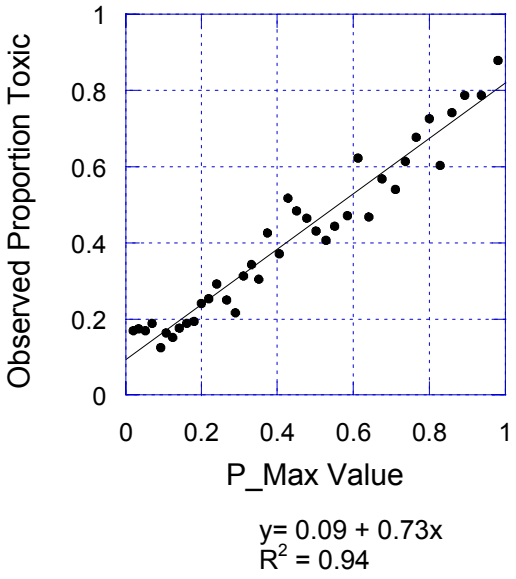


Figure 25. Observed proportion of toxic samples versus P_Max values including only those chemicals with normalized chi-square values exceeding 0.27. Data points represent the proportion toxic within unique probability intervals (minimum of 50 samples per interval) and the line is the P_Max model derived from the interval plot.

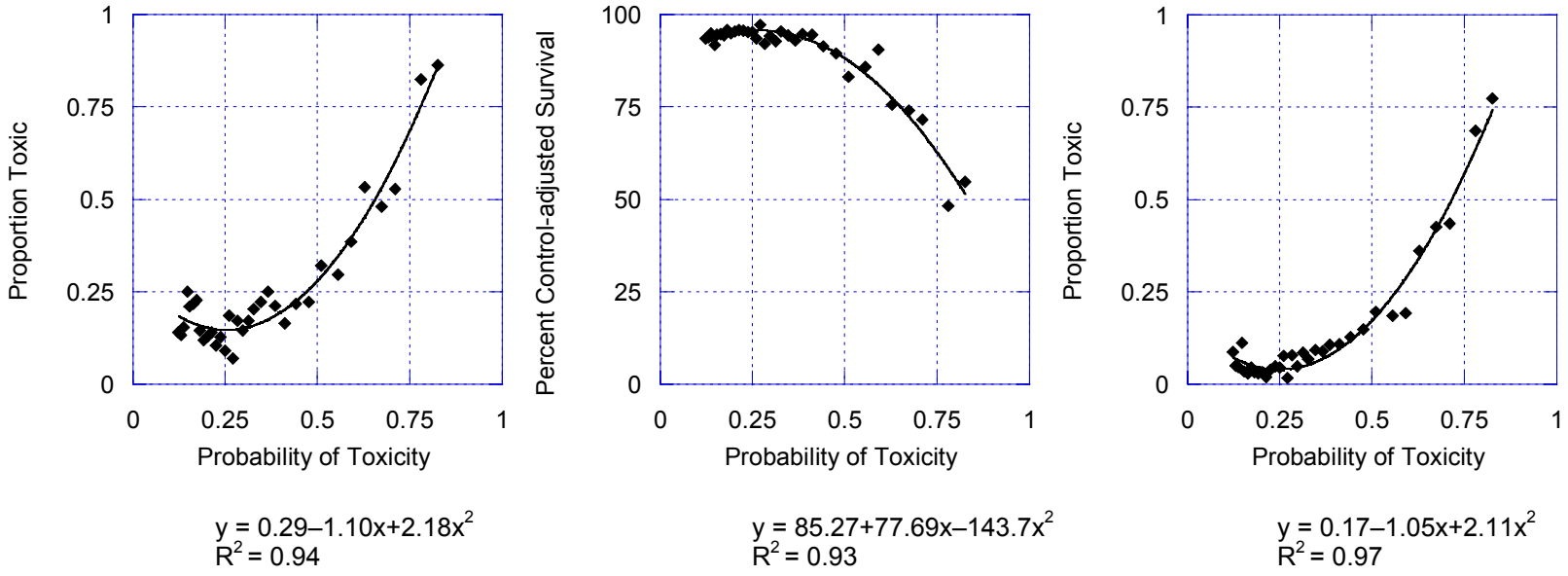


Figure 26. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data for *A. abdita*: proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 12 individual samples within the interval (n = 2022).

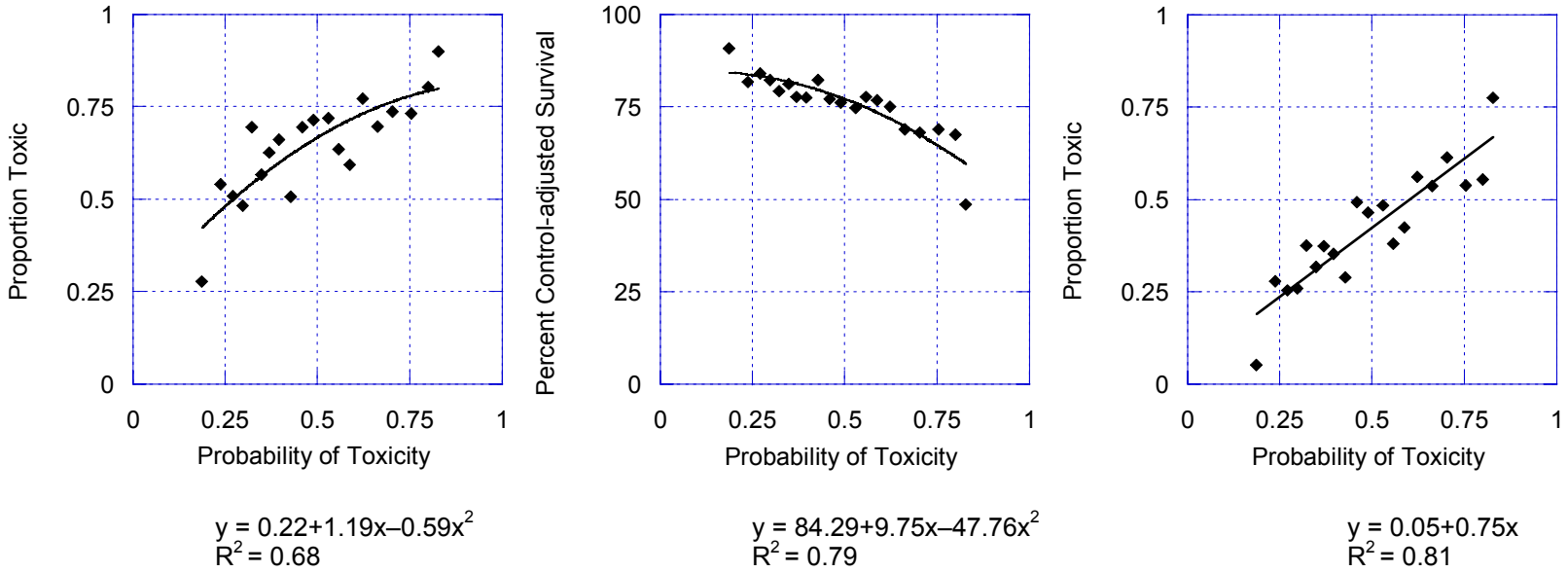


Figure 27. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data for *R. abronius*: proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 12 individual samples within the interval (n = 1211).

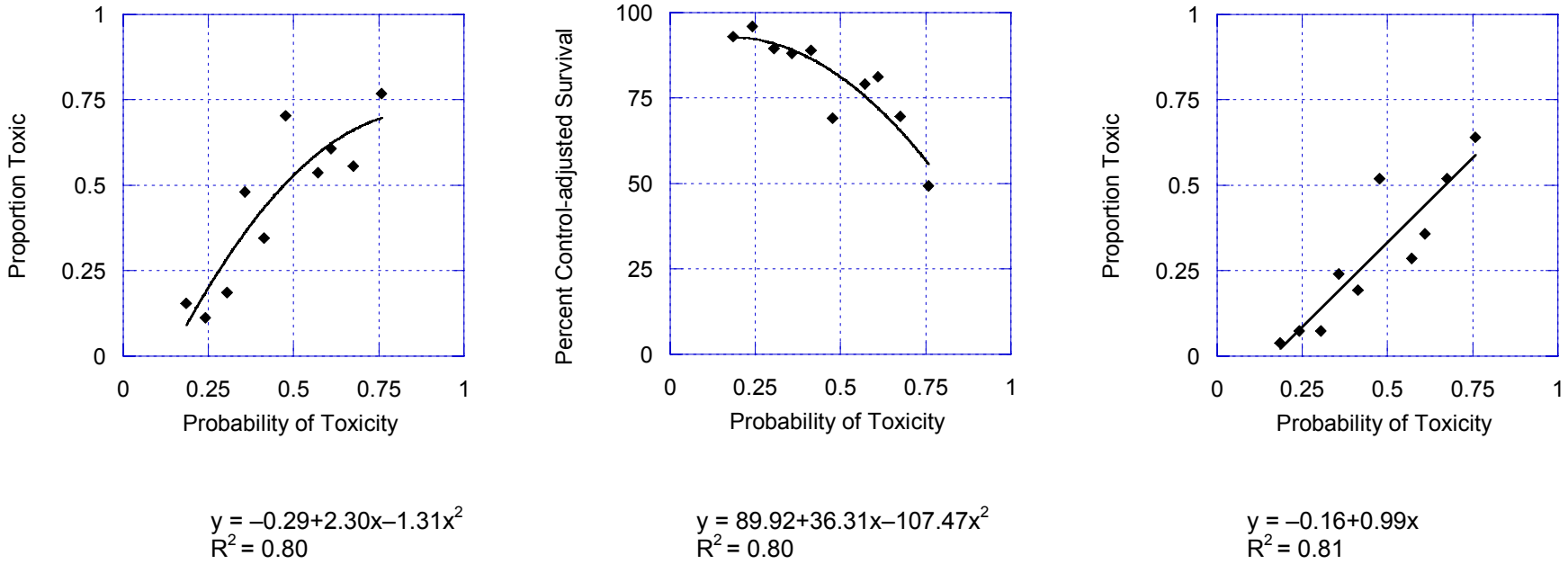


Figure 28. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data from Hudson-Raritan/Long Island NSTP and Regional EMAP (*A. abdita*): proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval (n = 280).

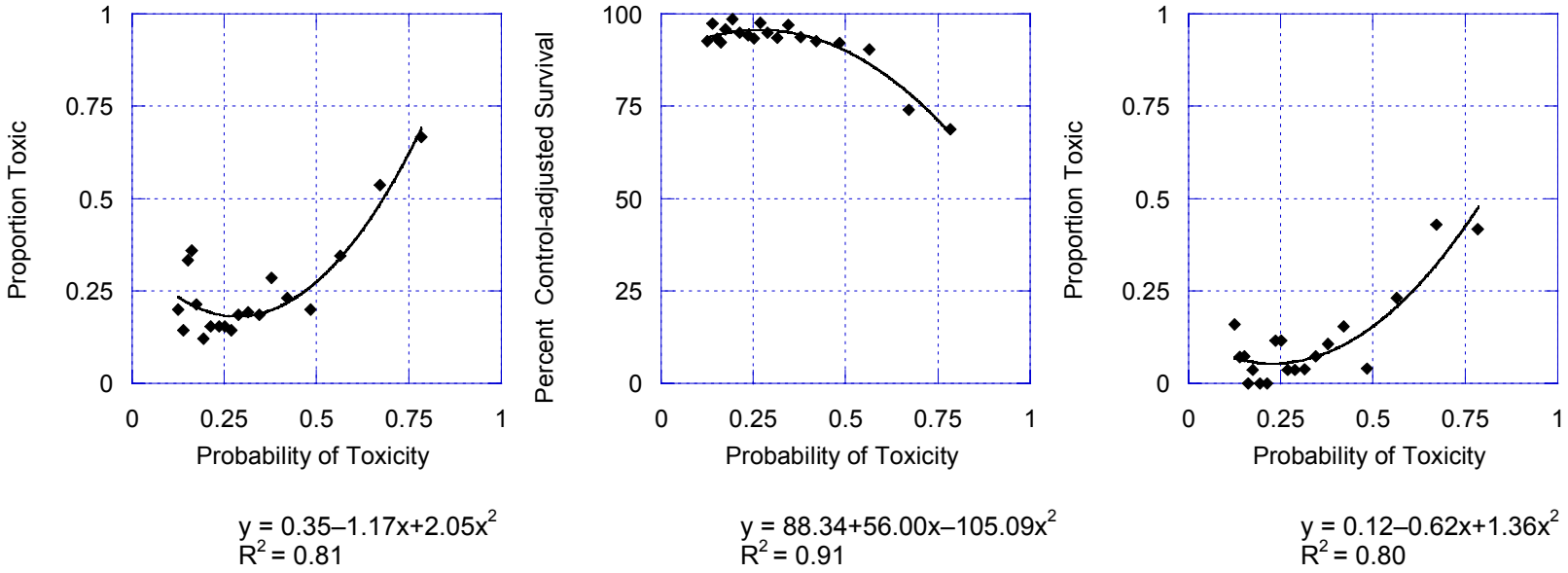


Figure 29. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data from Virginian Province EMAP (*A. abdita*): proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval (n = 489).

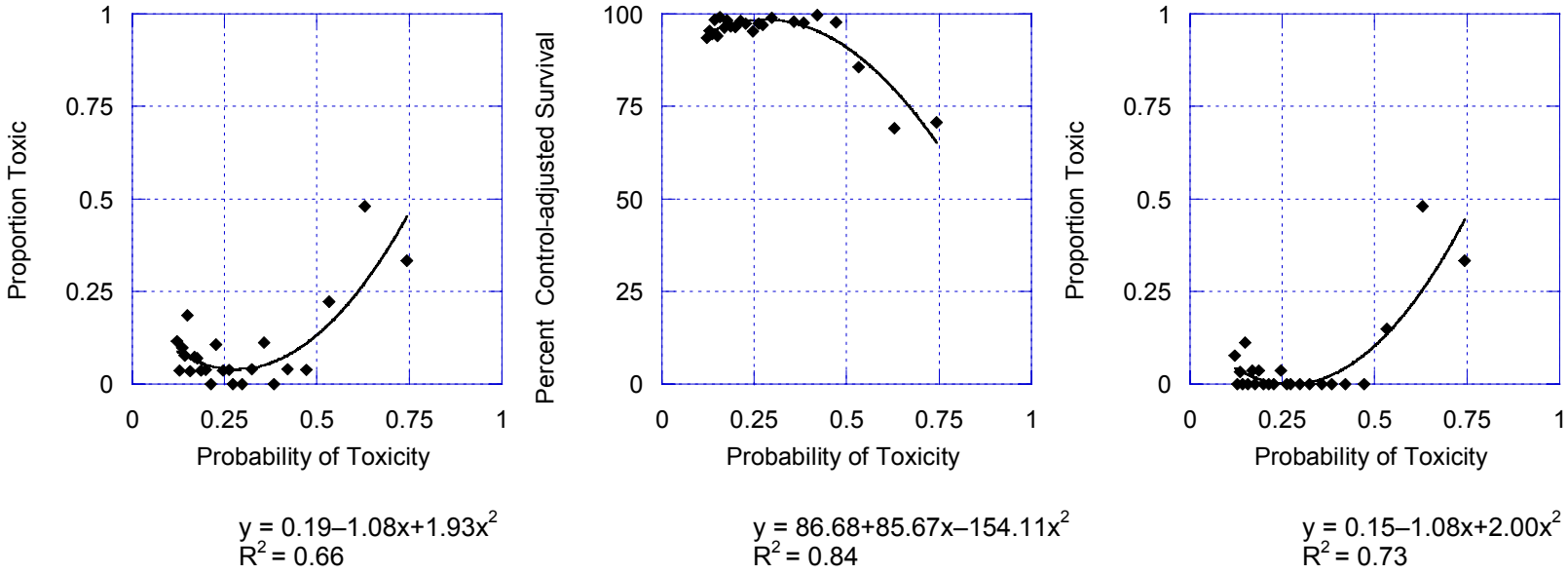


Figure 30. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data from NSTP and Carolinian EMAP from the southeastern U.S. (*A. abdita*): proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval (n = 636).

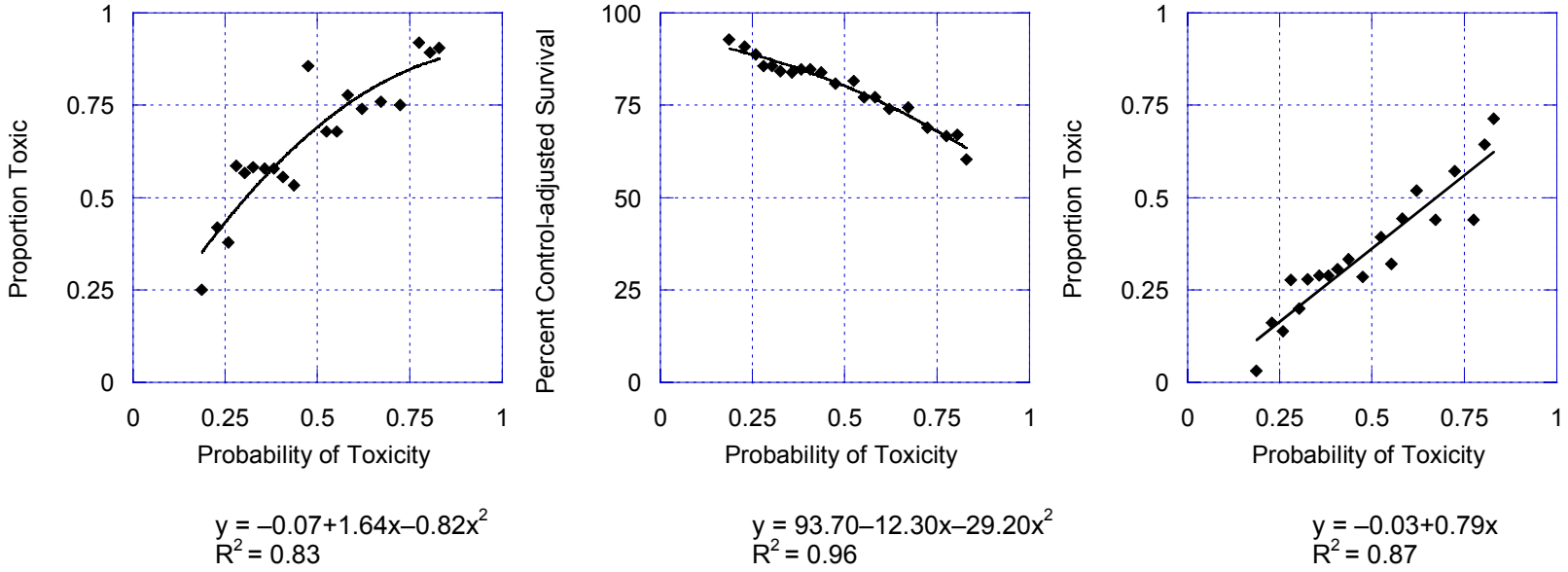


Figure 31. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data from Puget Sound, WA (*R. abronius*): proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval (n = 594).

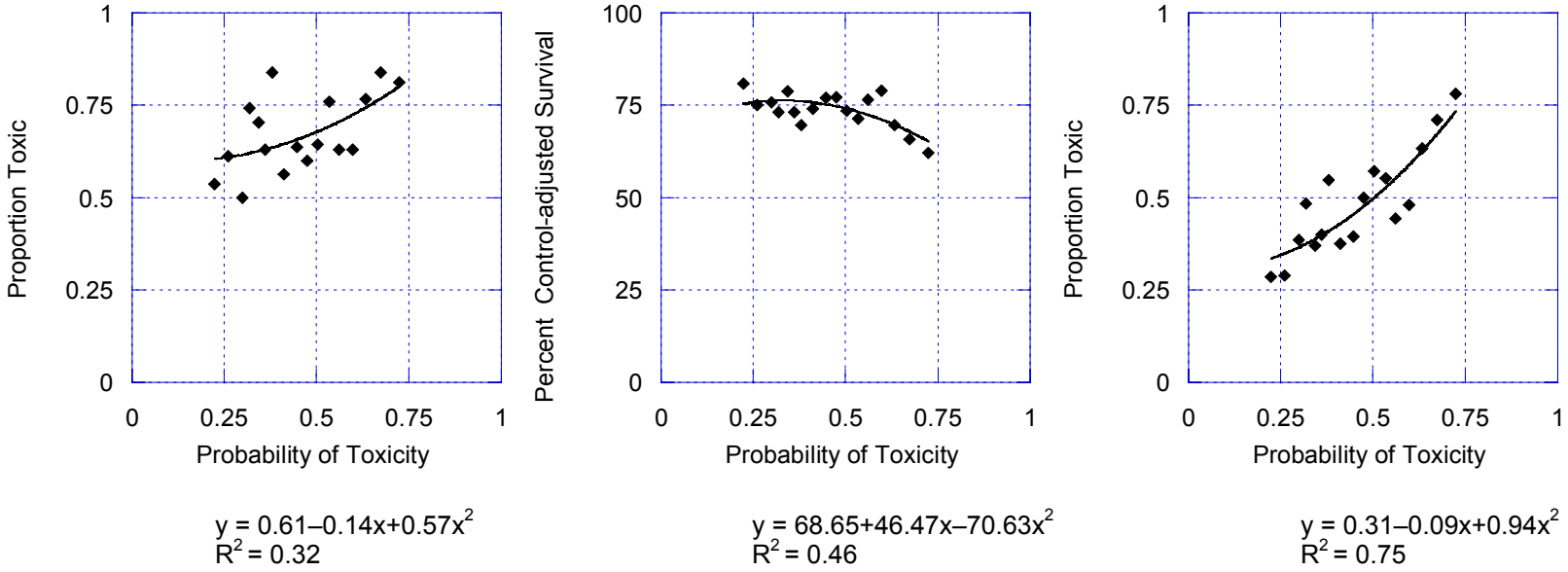


Figure 32. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data from California (*R. abronius* and *A. abdita*): proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval (n = 508).

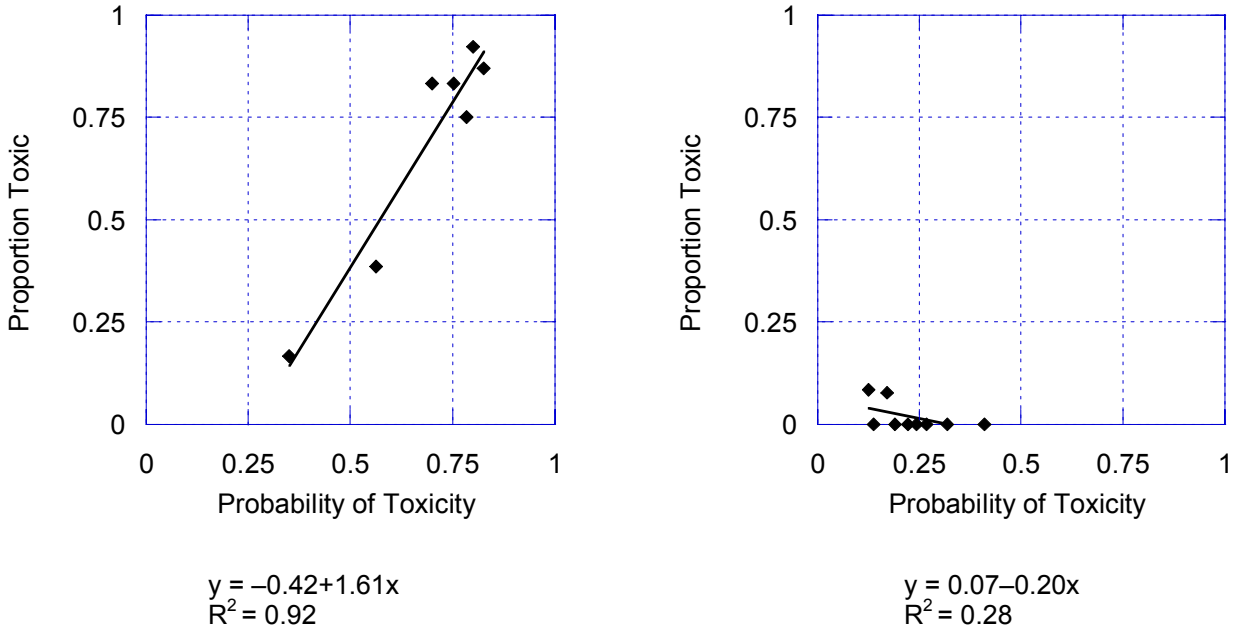


Figure 33. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with proportion of observed toxicity based on Sig Only classification from individual studies: Elliott Bay, Puget Sound, WA (*R. abronius*, n = 97) (left) and Biscayne Bay, FL (*A. abdita*, n = 120) (right). Each point represents the median sample probability of a minimum of 12 individual samples within the interval.

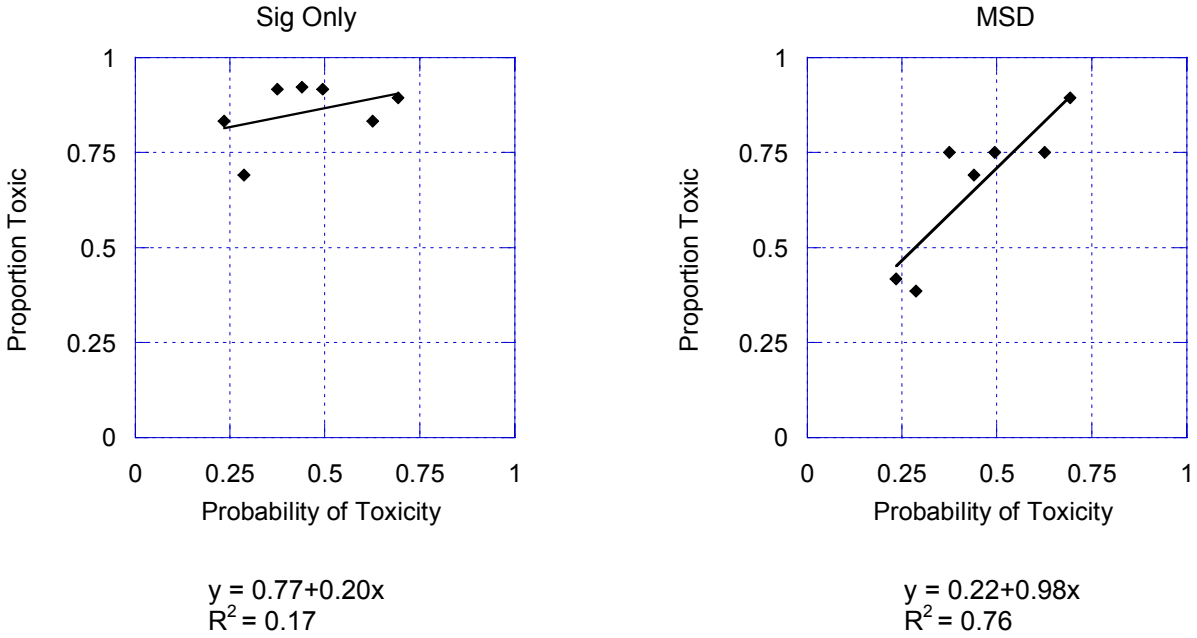


Figure 34. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with proportion of observed toxicity from an individual study from San Diego Bay, CA (*R. abronius*, n = 93); toxicity based on Sig Only classification and MSD classification. Each point represents the median sample probability of a minimum of 12 individual samples within the interval.

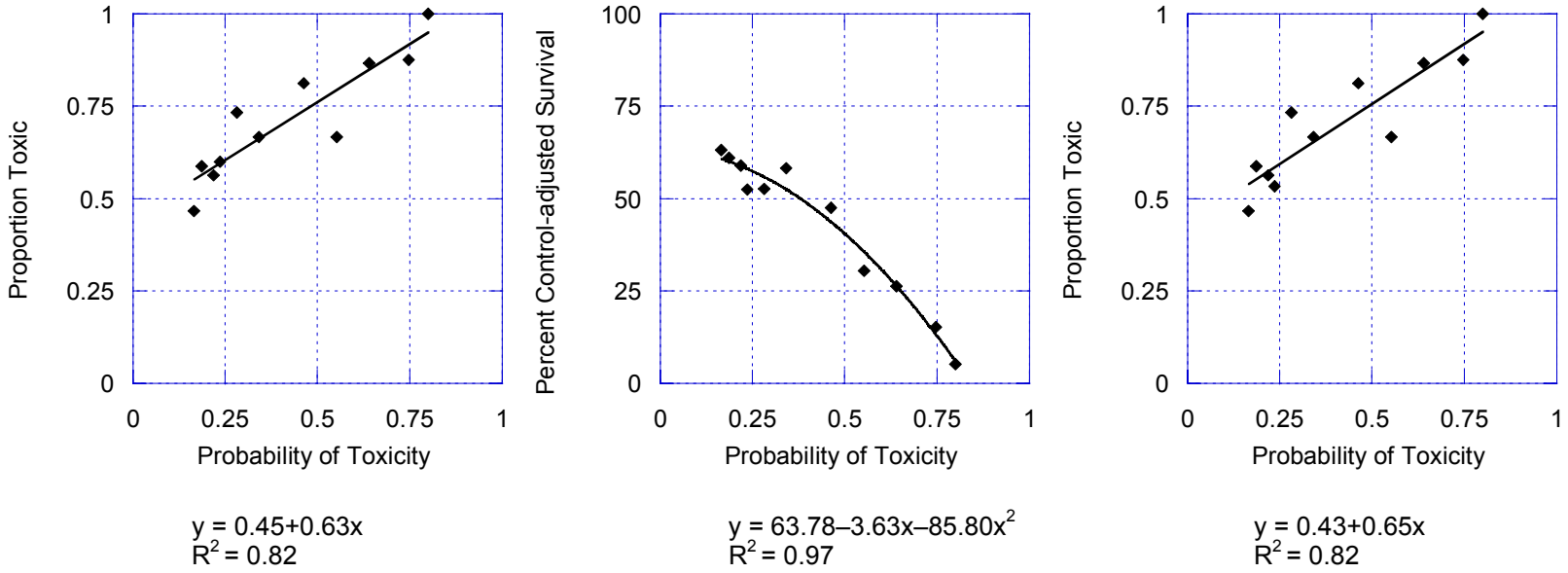


Figure 35. Median predicted probability of toxicity within probability intervals for the marine amphipod P_Max model compared with data from the Calcasieu Estuary (*A. abdita*): proportion of observed toxicity based on Sig Only classification (left), control-adjusted survival (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 12 individual samples within the interval (n = 170).

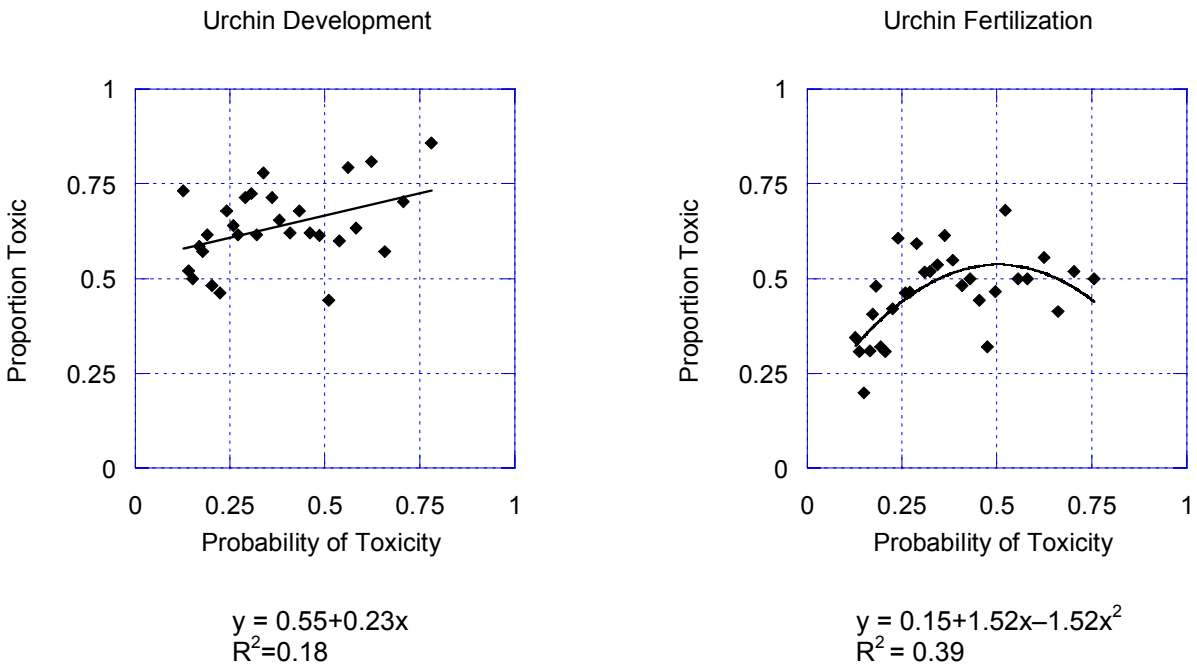


Figure 36. Average predicted probability and proportion toxic within probability intervals for the marine amphipod P_Max model applied to the sea urchin development and fertilization endpoints. Each point represents the median sample probability of a minimum of 25 individual samples within the interval and the proportion of the toxic samples within the interval (n = 782 and 824 for development and fertilization, respectively).

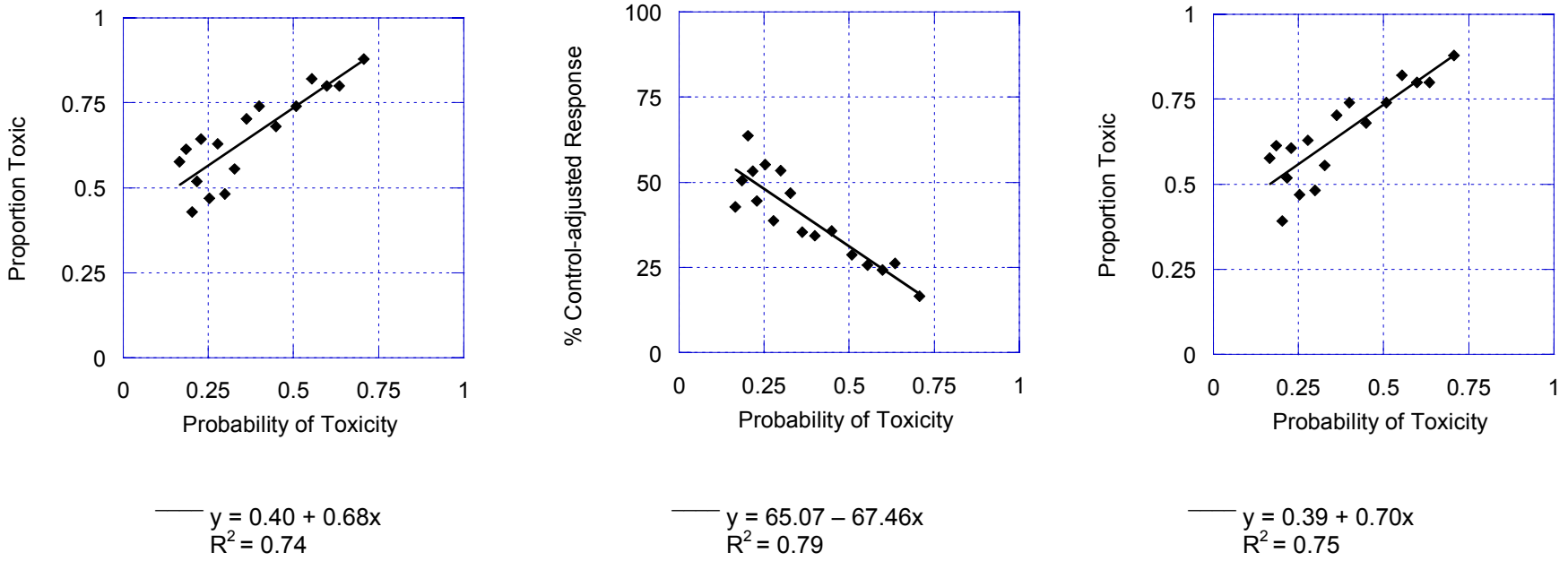


Figure 37. Median predicted probability and proportion toxic within probability intervals for the marine amphipod P_Max model applied to sea urchin (*A. punctulata*) development endpoint: proportion of observed toxicity based on Sig Only classification (left), control-adjusted response (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval and the proportion of the toxic samples within the interval (n = 472).

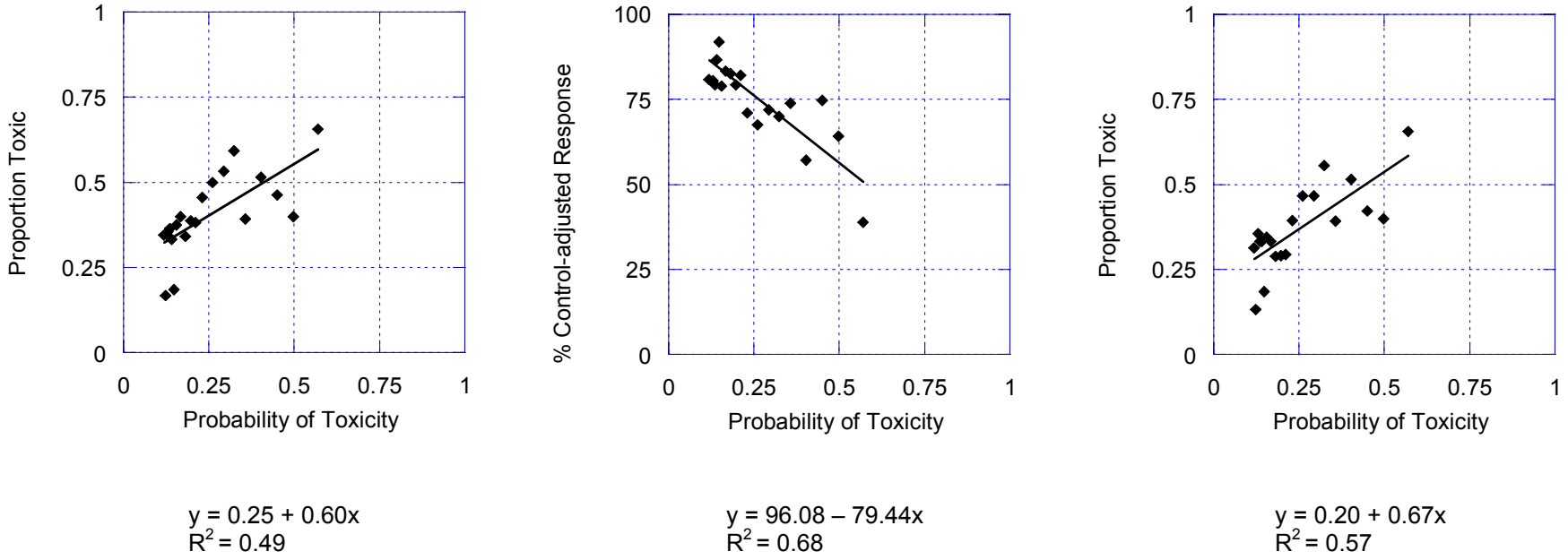


Figure 38. Median predicted probability and proportion toxic within probability intervals for the marine amphipod P_Max model applied to sea urchin (*A. punctulata*) fertilization endpoint: proportion of observed toxicity based on Sig Only classification (left), control-adjusted response (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval and the proportion of the toxic samples within the interval (n = 612).

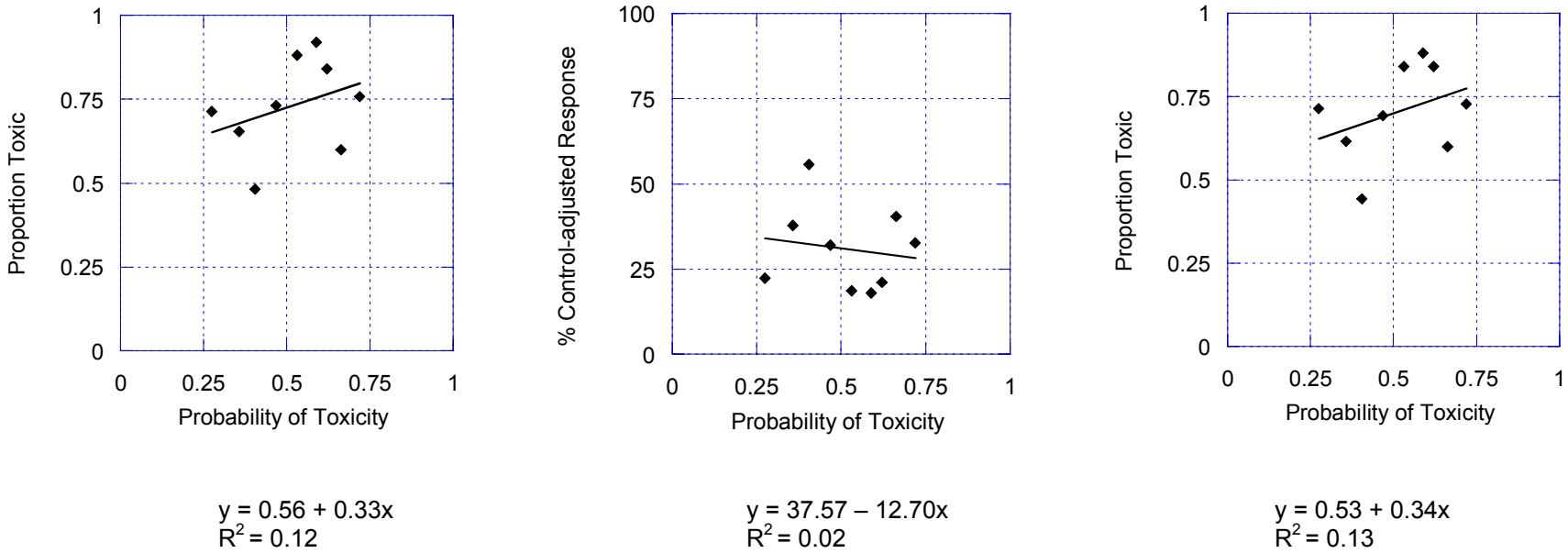


Figure 39. Median predicted probability and proportion toxic within probability intervals for the marine amphipod P_Max model applied to sea urchin (*S. purpuratus*) development endpoint: proportion of observed toxicity based on Sig Only classification (left), control-adjusted response (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval and the proportion of the toxic samples within the interval (n = 310).

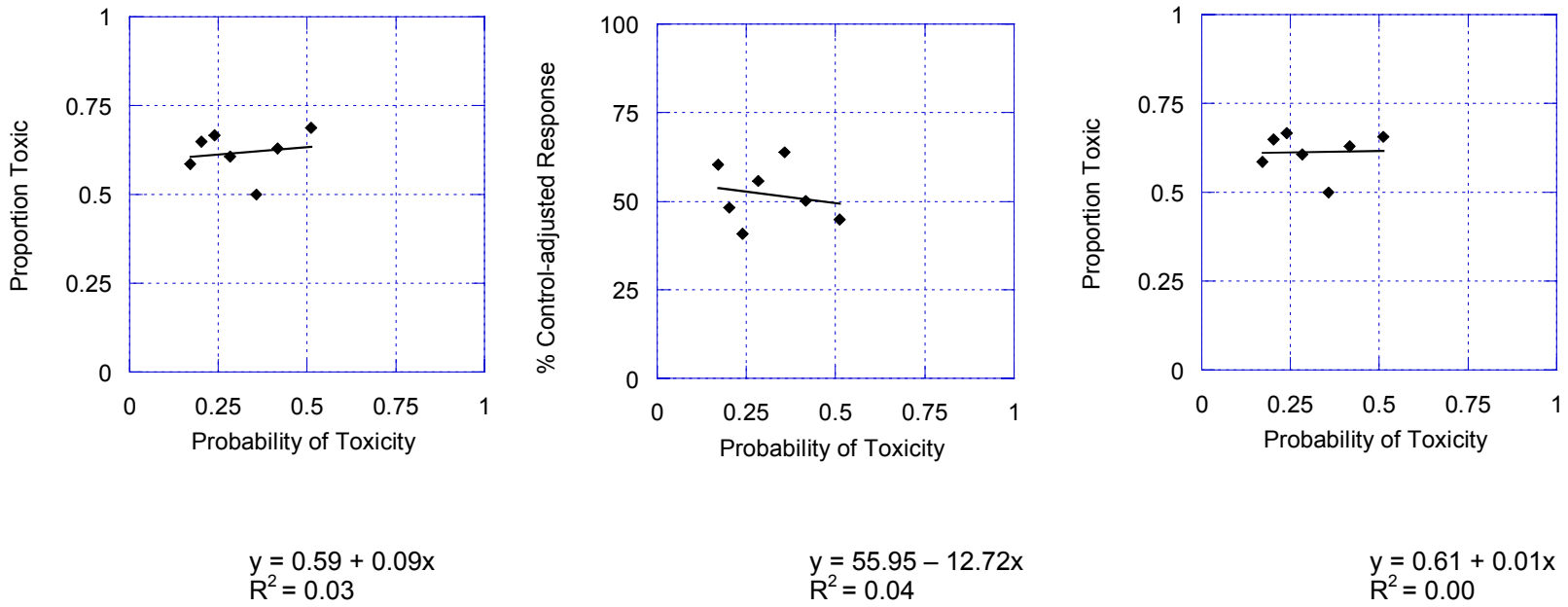


Figure 40. Median predicted probability and proportion toxic within probability intervals for the marine amphipod P_Max model applied to sea urchin (*S. purpuratus*) fertilization endpoint: proportion of observed toxicity based on Sig Only classification (left), control-adjusted response (center), and proportion of observed toxicity based on MSD classification (right). Each point represents the median sample probability of a minimum of 25 individual samples within the interval and the proportion of the toxic samples within the interval (n = 212).

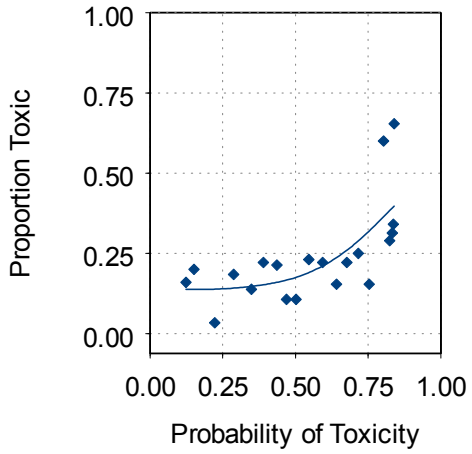


Figure 41. Proportion of samples that were toxic based on the *H. azteca* 10–14-day survival endpoint versus median probability of toxicity predicted using the marine amphipod P_Max model. Each point represents the median sample probability of a minimum of 25 individual samples within the interval and the proportion of toxic samples within the same interval (n = 567).

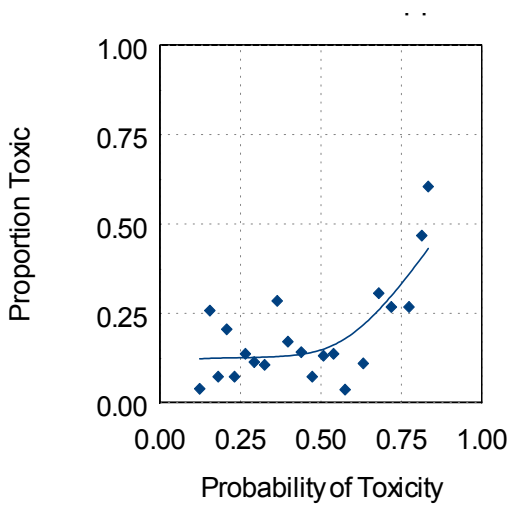


Figure 42. Proportion of samples that were toxic based on the *C. tentans* or *C. riparius* 10–14-day survival endpoint versus median probability of toxicity predicted using the marine amphipod P_Max model. Each point represents the median sample probability of a minimum of 25 individual samples within the interval and the proportion of toxic samples observed within the same interval (n = 585).

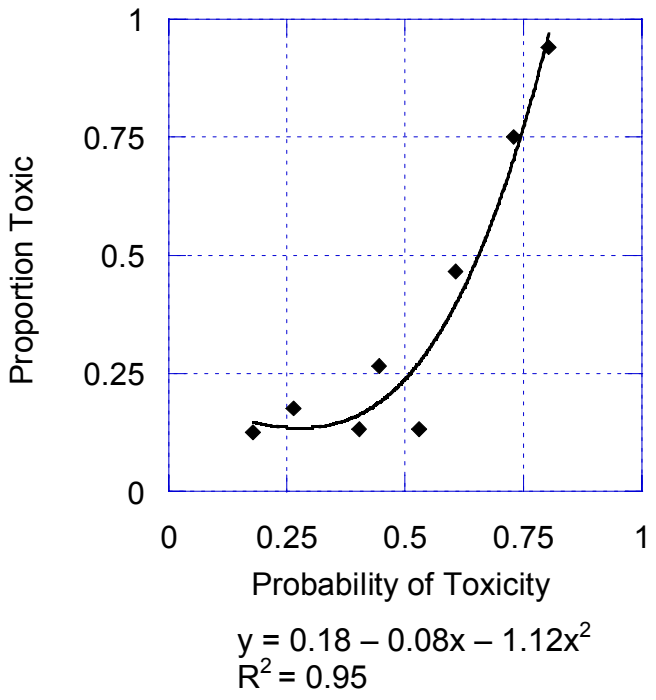


Figure 43. Proportion of samples that were toxic based on the *H. azteca* 28-day growth and survival endpoint versus the median probability of toxicity predicted using the marine amphipod P_Max model. Each point represents the median sample probability of a minimum of 15 individual samples within the interval and the proportion of the toxic samples observed within the same interval (n = 126).

REFERENCES

ASTM (American Society for Testing and Materials). (2002a) Standard guide for conducting 10-day, static sediment toxicity tests with marine and estuarine amphipods. In: Annual book of standards. Vol. 11.05. E 1367-99; pp. 693–719.

ASTM. (2002b) Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. E1706-00B. In: Annual book of standards. Vol. 11.05. E 1367-99; pp. 1125–1241.

Broderius, SJ. (1991) Modeling the joint toxicity of xenobiotics to aquatic organisms: basic concepts and approaches. In: Aquatic toxicology and risk assessment. Vol. 14. Mayes, MA; Barron, MG, eds. Philadelphia, PA: American Society for Testing and Materials; pp. 107–127.

Carr, RS; Biedenbach, JM. (1999) Use of power analysis to develop detectable significance criteria for sea urchin toxicity tests. Aquat Ecosys Health Manage 2:413–418.
Chapman, PM. (1989) Current approaches to developing sediment quality criteria. Environ Toxicol Chem 8:589–599.

Di Toro, DM; Zarba, CS; Hansen, DJ; et al. (1991) Technical basis for establishing sediment quality criteria for nonionic organic chemicals using equilibrium partitioning. Environ Toxicol Chem 10:1541–1583.

Di Toro, DM; McGrath, JA; Hansen, DJ. (2000) Technical basis for narcotic chemicals and polycyclic aromatic hydrocarbon criteria: II. mixtures and sediments. Environ Toxicol Chem 19:1971–1982.

Fairey, R; Long, ER; Roberts, CA; et al. (2001) An evaluation of methods for calculating mean sediment quality guideline quotients as indicators of contamination and acute toxicity to amphipods by chemical mixtures. Environ Toxicol Chem 20:2276–2286.

Field, LJ; MacDonald, DD; Norton, SB; et al. (1999) Evaluating sediment chemistry and toxicity data using logistic regression modeling. Environ Toxicol Chem 18:1311–1322.

Field, LJ; MacDonald, DD; Norton, SB; et al. (2002) Predicting amphipod toxicity from sediment chemistry using logistic regression models. Environ Toxicol Chem 21(9):1993–2005.

Gries, TH; Waldow, KH. (1996) Progress re-evaluating Puget Sound apparent effects thresholds (AETs). Vol. I. 1994 amphipod and echinoderm larval AETs. Puget Sound dredged disposal analysis (PSDDA). U.S. Army Corps of Engineers, Seattle District, U.S. Environmental Protection Agency, Region 10, State of Washington Departments of Ecology and Natural Resources, Seattle, WA.

Hansen, DF; Berry, WJ; Mahony, JD; et al. (1996) Predicting the toxicity of metal-contaminated field sediments using interstitial concentration of metals and acid-volatile sulfide normalizations. Environ Toxicol Chem 15:2080–2094.

Ingersoll, CG; Haverland, PS; Brunson, EL; et al. (1996) Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironomus riparius*. J Great Lakes Res 22:602–623.

Ingersoll, CG; Dillon, T; Biddinger, RG, eds. (1997) Ecological risk assessments of contaminated sediment. Pensacola, FL: SETAC Press.

Ingersoll, CG; MacDonald, DD; Wang, N; et al. (2001) Predictions of sediment toxicity using consensus-based freshwater sediment quality guidelines. Arch Environ Contam Toxicol 41:8–21.

Ingersoll, CG; MacDonald, DD; Brumbaugh, WG; et al. (2002) Toxicity assessment of sediments from the Grand Calumet River and Indiana Harbor Canal in northwestern Indiana. Arch Environ Contam Toxicol 43:156–167.

Insightful Corp. (2001) S-Plus 6 for windows guide to statistics. Seattle, WA.

Kemp, PF; Swartz, RC; Lamberson, JO. (1986) Response of the photocephalid amphipod, *Rhepoxynius abronius*, to a small oil spill in Yaquina Bay, Oregon. Estuaries 9:340–347.

Lee, JH; Landrum, PF; Field, LJ; et al. (2001) Application of a ∑PAH model and a logistic regression model to sediment toxicity data based on a species-specific water-only LC50 toxic unit for *Hyalella azteca.* Environ Toxicol Chem 20:2102–2113.

Long, ER; MacDonald, DD. (1992) National Status and Trends Program approach. In: Sediment classification methods compendium. EPA 823-R-92-006. Washington, DC: U.S. Environmental Protection Agency.

Long, ER; Morgan, LG. (1990) The potential for biological effects of sediment-sorbed contaminants tested in the National Status and Trends Program. NOAA Tech Memo NOS OMA 52. National Oceanic and Atmospheric Administration, Seattle, WA.

Long, ER; Morgan, LG. (1991) The potential for biological effects of sediment-sorbed contaminants tested in the National Status and Trends Program. NOAA Tech Memo NOS OMA 52. National Oceanic and Atmospheric Administration, Seattle, WA.

Long, ER; MacDonald, DD; Smith, SL; et al. (1995) Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. Environ Manage 19:81–97.

Long, ER; Field, LJ; MacDonald, DD. (1998) Predicting toxicity in marine sediments with numerical sediment quality guidelines. Environ Toxicol Chem 17:714–727.

MacDonald, DD; Carr, RS; Calder, FD; et al. (1996) Development and evaluation of sediment quality guidelines for Florida coastal waters. Ecotoxicology 5:253–278.

MacDonald, DD; Ingersoll, CG; Berger, TA. (2000) Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39:20–31.

MacDonald, DD; Moore, DR; Pawlisz, A; et al. (2001) Calcasieu Estuary remedial investigation/feasibility study (RI/FS): baseline ecological risk assessment (BERA). Vol. 2.

Mearns, AJ; Swartz, RC; Cummins, JM; et al. (1986) Inter-laboratory comparison of a sediment toxicity test using the marine amphipod, *Rhepoxynius abronius*. Mar Environ Res 19:13–37.

Morgan, BJT. (1992) Analysis of quantal response data. London, UK: Chapman and Hall.

Neter, J; Kutner, MH; Nachtscheim, CJ; et al. (1996) Applied linear statistical models. 4th ed. Chicago, IL: Irwin.

NOAA (National Oceanic and Atmospheric Administration). (1989) Standard analytical procedures of the NOAA Analytical Facility. 2nd ed. NOAA technical memorandum NMFS F/NWC-92, 1985-86. National Status and Trends Program, U.S. Department of Commerce, Rockville, MD.

NOAA (National Oceanic and Atmospheric Administration). (2004) Predicting sediment toxicity. Available online at http://www.response.restoration.noaa.gov/cpr/sediment/sed_tox.html.

Phillips, BM; Hunt, JW; Anderson, BS; et al. (2001) Statistical significance of sediment toxicity test results: threshold values derived by the detectable significance approach. Environ Toxicol Chem 20:371–373.

Redmond, MS; Crocker, PA; McKenna, KM; et al. (1996) Sediment toxicity testing with the amphipod *Ampleisca abdita* in Calcasieu Estuary. Arch Environ Contam Toxicol 30:53–61.

Robinson, AM; Lamberson, JO; Cole, FA; et al. (1988) Effects of culture conditions on the sensitivity of a phoxocephalid amphipod, *Rhepoxynius abronius*, to cadmium in sediment. Environ Toxicol Chem 7:953–959.

SAS Institute. (1990) SAS/STAT user's guide, version 6, 4th ed. Vols. 1 and 2. Cary, NC.
Smith, SL; MacDonald, DD; Keenleyside, KA; et al. (1996) A preliminary evaluation of sediment quality assessment values for freshwater ecosystems. J Great Lakes Res 22:624–638.

Smith, E; Robinson, T; Field, LJ; et al. (2003) Predicting sediment toxicity using logistic regression: a concentration–addition approach. Environ Toxicol Chem 22:565–575.

Swartz, RC; Ditsworth, GR; Schults, DW; et al. (1985) Sediment toxicity to a marine infaunal amphipod: cadmium and its interactions with sewage sludge. Mar Environ Res 18:133–153.

Swartz, RC; Schults, DW; DeWitt, TH; et al. (1987) Toxicity of fluoranthene in sediment to marine amphipods: a test of the equilibrium partitioning approach to sediment quality criteria. 89th Annual Meeting, Society for Environmental Toxicology and Chemistry. Pensacola, FL.

Swartz, RC; Kemp, PF; Schults, DW; et al. (1988) Effects of mixtures of sediment contaminants on the marine infaunal amphipod, *Rhepoxyinius abronius*. Environ Toxicol Chem 7:1013–1020.

Swartz, RC; Kemp, PF; Schults, DW; et al. (1989) Acute toxicity of sediment from Eagle Harbor, Washington, to the marine infaunal amphipod, *Rhepoxyinius abronius*. Environ Toxicol Chem 8:215–222.

Swartz, RC; Schults, DW; Ozretich, RJ; et al. (1995) ∑PAH: a model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments. Environ Toxicol Chem 14:1977–1987.

Thursby, GB; Heltshe, J; Scott, KJ. (1997) Revised approach to toxicity test acceptability criteria using a statistical performance assessment. Environ Toxicol Chem 16:1322–1329.

Tukey, JW. (1977) Exploratory data analysis. Reading, MA: Addison-Wesley.

U.S. EPA (Environmental Protection Agency). (1994a) Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. 1st ed. EPA/600/R-94/024. Duluth, MN.

U.S. EPA. (1994b) Methods for measuring the toxicity of sediment-associated contaminants with estuarine and marine amphipods. EPA-600/R-94/025. Narragansett, RI.

U.S. EPA. (1997) The incidence and severity of sediment contamination in surface waters of the United States. Vol. 1: National sediment quality survey. EPA 823-R-97-006. Office of Science and Technology, Washington, DC.

U.S. EPA. (2000) Supplementary guidance for conducting health risk assessment of chemical mixtures. EPA 630-R-00-002. Office of Research and Development, Washington, DC.

U.S. EPA. (2004) The incidence and severity of sediment contamination in surface waters of the United States. National sediment quality survey. Second edition. EPA-823-R-04-007. Office of Science and Technology, Washington, DC.

Word, JQ; Ward, JA; Franklin, LM; et al. (1987) Evaluation of the equilibrium partitioning theory for estimating the toxicity of the nonpolar organic compound DDT to the sediment dwelling amphipod *Rhepoxynius abronius*. Battelle Washington Environmental Program Office. Report prepared for Criteria and Standards Division, U.S. Environmental Protection Agency.

APPENDIX A1: REFERENCES FOR MARINE SEDTOX02 DATABASE

Adams, DA; O'Conner, JS; Weisberg, SB. (1998) Sediment quality of the NY/NJ Harbor system: an investigation under the Regional Environmental Monitoring and Assessment Program (R-EMAP). Final report. U.S. EPA, Region 2, Division of Environmental Sciences & Assessment EPA/902-R-98-001.

Analytical. (1988) Volatile organics, pesticides, PCBs, acid, base neutral analyses for PSDDA phase i baseline monitoring (Hardcopies and Lotus worksheet files). Received from ARI on July 5, 1988.

Anderson, JW; Bay, SM; Thompson, BE. (1988) Characteristics and effects of contaminated sediments from Southern California. California State Water Resources Control Board, Sacramento, CA.

Anderson, B; Hunt, J; Tudor, S; et al. (1997) Chemistry, toxicity and benthic community conditions in sediments of the Southern California bays and estuaries. California State Water Resources Control Board, Sacramento, CA.

Anderson, B; Hunt, J; Phillips, B; et al. (1998) Sediment chemistry, toxicity, and benthic community conditions in selected water bodies of the Los Angeles Region. Final Report. California State Water Resources Control Board, Sacramento, CA.

Bay, S; Greentein, D; Brown, J; et al. (1994) Investigation of toxicity in Palos Verdes sediments. Final report 1994. Prepared by Southern California Coastal Water Resources Project, Westminster CA. Prepared for Santa Monica Bay Restoration Project.

Carr, RS. (1992) Survey of Galveston Bay bottom sediments and benthic communities. Draft final report. Prepared by U.S. Fish and Wildlife Service, Corpus Christi State University. Prepared for Galveston Bay National Estuary Program and U.S. Environmental Protection Agency (USEPA), Region 6.

Casillas, E; Weber, D; Haley, C; et al. (1992) Comparison of growth and mortality in juvenile sand dollars (*Dendraster excentricus*) as indicators of contaminated marine sediments. Environ Toxicol Chem 11:559–569

Chapman, PM; Dexter, RN; Long, ER. (1987) Synoptic measures of sediment contamination, toxicity and infaunal community composition (the Sediment Quality Triad) in San Francisco Bay. Marine Ecol Progr Ser 37.

Chapman, PM; Power, EA; Dexter, RN; et al. (1991) Evaluation of effects associated with an oil platform, using Sediment Quality Triad. Environ Toxicol Chem, Vol. 10.

Crocker; et al. (1991) Water quality, ambient toxicity and biological investigations in the Houston Ship Channel and tidal San Jacinto River. U.S. EPA.

Downing, J; Fairey, R; Roberts, C; et al. (1998) Chemical and biological measures of sediment quality in the Central Coast Region. Final report. California State Water Resources Control Board, Sacramento, CA.

EA Mid-Atlantic, EA Engineering, Science, and Technology, Inc. (1994) Results of dredged material testing of samples from the proposed Bayou Casotte Turning Basin, Pascagoula, Mississippi. U.S. Army Engineer District, Mobile, Alabama.

Engle, V; Harwell, L; Heitmuller, T. (1999) Catalog documentation EMAP-estuaries program level database Louisianian Province 1991-1994 sediment chemistry data & sediment toxicity data. U.S. EPA and U.S. Geological Survey.

Fairey, R; Bretz, C; Lamerdin, S; et al. (1996) Chemistry, toxicity and benthic community conditions in sediments of the San Diego bay region. California State Water Resources Control Board, Sacramento, CA.

Fairey, R; Downing, J; Roberts, C; et al. (1998) Chemistry, toxicity and benthic community conditions in sediments of the San Diego bay region: final addendum report. California State Water Resources Control Board, Sacramento, CA.

Ferraro, SP; Swartz, RC; Cole, FA; et al. (1991) Temporal changes in the benthos along a pollution gradient: Discriminating the effects of natural phenomena from sewage-industrial wastewater effects. Estuar Coast Shelf Sci 33(4):383–407.

Flegal, AR; Risebrough, RW; Anderson, B; et al. (1996) San Francisco estuary pilot regional monitoring program: Sediment studies. San Francisco Bay Regional Water Quality Control Board, State Water Resources Control Board.

Hall, LW, Jr; Ziegenfuss, MC; Fischer, SA; et al. (1992) A pilot study for ambient toxicity testing in Chesapeake Bay. Year 2 report. Prepared for U.S. Environmental Protection Agency. CBP/TRS 82/92, Contract No. 68-WO-00-43. Printed by U.S. EPA.

Hart Crowser & Associates Inc. (1986) Results of chemical and biological analyses of sediment samples, Everett Marina, and Jetty Island Pier, Everett, Washington. Prepared by Hart Crowser and Associates, Inc.

Hunt, JB; Anderson, B; Phillips, B; et al. (1997) Evaluation of sediment toxicity tests and reference sites in San Francisco Bay. California State Water Resources Control Board, Sacramento, CA.

Hunt, JB; Anderson, B; Phillips, B; et al. (1998) Sediment quality and biological effects in San Francisco Bay. Bay Protection and Toxic Cleanup Program. Final technical report. California State Water Resources Control Board, Sacramento, CA.

Hyland, JL; Herrlinger, TJ; Snoots, TR; et al. (1996) Environmental quality of estuaries of the Carolinian Province: 1994. Annual statistical summary for the 1994 EMAP-Estuaries Demonstration Project in the Carol. NOAA Technical Memorandum NOS ORCA 97, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Hyland, JL; Balthis, L; Hackney, CT; et al. (1998) Environmental quality of estuaries of the Carolinian Province: 1995. Annual statistical summary for the 1995 EMAP - Estuaries demonstration project in the Car. NOAA Technical Memorandum NOS ORCA 123, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Jacobi, M; Fairey, R; Roberts, C; et al. (1998) Chemical and biological measures of sediment quality and tissue bioaccumulation in the North Coast Region. Final Report. California State Water Resources Control Board, Sacramento, CA.

Lamberson, JO; Redmond, MS; Jones, JKP; et al. (1993) Development of a 10-day amphipod sediment toxicity test for Hawaii. USEPA, Environmental Research Lab - Narragansett. Contract No 68-CO-0051.

Long, ER; Buchman, MF; Bay, SM; et al. (1990) Comparative evaluation of five toxicity tests with sediments form San Francisco Bay and Tomales Bay, California. Environ Toxicol Chem 9:1193–1214.

Long, ER; Wolfe, DA; Carr, RS; et al. (1994) Magnitude and extent of sediment toxicity in Tampa Bay, Florida. NOAA Technical Memorandum NOS ORCA 78. National Oceanic and Atmospheric Administration, Silver Spring, MD.

Long, ER; Wolfe, DA; Scott, KJ; et al. (1995) Magnitude and extent of sediment toxicity in the Hudson-Raritan Estuary. NOAA Technical Memorandum NOS ORCA 88, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Long, ER; Sloane, GM; Carr, RS; et al. (1996) Sediment toxicity in Boston Harbor: Magnitude, extent, and relationships with chemical toxicants. NOAA Technical Memorandum NOS ORCA 96. National Oceanic and Atmospheric Administration, Silver Spring, MD.

Long, ER; Sloane, GM; Carr, RS; et al. (1997) Magnitude and extent of sediment toxicity in four bays for the Florida panhandle: Pensacola, Choctawatchee, St. Andrew and Apalachicola. National Oceanic and Atmospheric Administration, Silver Spring, MD.

Long, E; Scott, G; Kucklick, J; et al. (1998) Magnitude and extent of sediment toxicity in selected estuaries of South Carolina and Georgia. NOAA Technical Memorandum NOS ORCA 128, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Long, E; Sloane, G; Scott, G; et al. (1999) Magnitude and extent of chemical contamination and toxicity in sediments of Biscayne Bay and vicinity-final draft. NOAA Technical Memorandum NOS ORCA 141, National Oceanic and Atmospheric Administration, Silver Spring, MD.

McGee, BL; Schlekat, CE; Reinharz, E. (1993) Assessing sublethal levels of sediment contamination using the estuarine amphipod *Leptocheirus plumulosus*. Environ Toxicol Chem 12:577–587.

McLeay, D; Yee, S; Doe, K; et al. (1991) Final report: Phase II and phase III studies by EC Laboratories of 10-day tests for sediment toxicity using marine or estuarine infaunal amphipods. Prepared by McLeay Associates Ltd. for Environment Canada.

Murdoch, MH; Chapman, PM; Johns, DM; et al. (1996) Chronic effects of organochlorine exposure in sediment to the marine polychaete *Neanthes arenaceodentata*. Prepared by EVS Environmental Consultants, North Vancouver, BC.

Nelson, WG; Bergen, BJ; Benyi, SJ; et al. (1996) New Bedford Harbor long-term monitoring assessment report: Baseline sampling. Research Report United States Environmental Protection Agency.

PTI Environmental Services (1989) PSDDA baseline survey of phase ii disposal sites. Bellevue, WA.

PTI Environmental Services. (1990) Reconnaissance survey of reference area sediments in shallow waters of Carr inlet. Technical memorandum. Prepared for U.S. Environmental Protection Agency, Region X, Seattle, WA.

PTI Environmental Services. (1991) Reconnaissance survey of chemical contamination and biological effects in southern Puget Sound. Prepared for U.S. Environmental Protection Agency, Region X, Seattle, WA.

PTI Environmental Services; Tetra Tech, Inc. (1988) Everett Harbor Action Program: analysis of toxic problem areas. Final report. Prepared for U.S. Environmental Protection Agency, Region X, Office of Puget Sound, Seattle, WA.

PTI Environmental Services; Tetra Tech, Inc. (1988) Puget Sound Estuary Program Elliott Bay Action Program: Analysis of toxic problem areas. Final report. Prepared for U.S. Environmental Protection Agency, Region X, Office of Puget Sound, Seattle, WA.

Pacific Northwest Laboratory, Battelle Marine Research Laboratory. (1986) Reconnaissance survey of eight bays in Puget Sound. Final report - Volumes I and II. Prepared for U.S. Environmental Protection Agency, Region X, Seattle, WA.

Pastorok, RA; Becker, DS. (1990) Comparative sensitivity of sediment toxicity bioassays at three Superfund sites in Puget Sound. ASTM STP 1096.

Phillips, B; Anderson, B; Hunt, J; et al. (1998) Sediment chemistry, toxicity, and benthic community conditions in selected water bodies of the Santa Ana Region. Final report. California State Water Resources Control Board, Sacramento, CA.

Rice, CA; Plesha, PD; Casillas, E; et al. (1995) Growth and survival of three marine invertebrate species in sediments from the Hudson-Raritan Estuary, New York. Environ Toxicol Chem 14(11):1931–1940

Striplin, PL; Sparks-McConkey, PJ; Davis, DA; et al. (1992) Puget Sound ambient monitoring program, marine sediment monitoring program, annual report 1990. Washington State Department of Ecology, Olympia, WA.

Strobel, CJ; Buffum, HW; Benyi, SJ; et al. (1995) Statistical summary. EMAP-Estuaries. Virginian Province - 1990 to 1993. U.S. Environmental Protection Agency, Narragansett, RI. EPA/620-R-94/-26.

Swartz, RC; Schults, DW; Ditsworth, GR; et al. (1985) Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall. Validation and predictability of laboratory methods for assessing the fate and effects of contaminants in aquatic ecosystems, ASTM STP 865.

Swartz, RC; Cole, FA; Schults, DW; et al. (1986) Ecological changes in the Southern California Bight near a large sewage outfall: benthic conditions in 1980 and 1983. Marine Ecol Progr Ser. U.S. Environmental Protection Agency.

Swartz, RC; Kemp, PF; Schults, DW; et al. (1988) Acute toxicity of sediment from Eagle Harbor, Washington, to the infaunal amphipod *Rhepoxynius abronius*. Environ Toxicol Chem, Vol. 8.

Swartz, RC; Schults, DW; Lamberson, JO; et al. (1991) Vertical profiles of toxicity, organic carbon, and chemical contaminants in sediment cores from the Palos Verdes shelf and Santa Monica Bay, California. Marine Environ Res 31(3):215–225

Tay, Kok-Leng; Doe, KG; Wade, SJ; et al. (1990) Sediment assessment in Halifax Harbour. In: Aquatic toxicity workshop proceedings: 17th 1991, Vol. 2.

Tetra Tech, Inc. (1990) Puget Sound ambient monitoring program 1989 marine sediment monitoring final report. Bellevue, WA.

Tetra Tech, Inc. (1988a) Commencement Bay nearshore/tideflats feasibility study, appendices, public review draft. Bellevue, WA.

Tetra Tech, Inc. (1988b) Commencement Bay nearshore/tideflats feasibility study. Vol. 1, public review draft. Bellevue, WA.

Ward, JA; Pinza, MR; Barrows, ME; et al. (1992) Ecological evaluation of proposed dredged material from Wilmington Harbor, North Carolina. Draft report. Prepared by Pacific Northwest Laboratory. Prepared for the U.S. Army Corps of Engineers with the U.S. Department of Energy. Contract DE-AC06-76RLO 1830.

Washington State Department of Ecology. (1994) Puget Sound ambient monitoring program, marine sediment monitoring program, annual report 1991. Olympia, WA. Publication #94-93.

Weyerhaeuser Analytical Laboratory. (1984) Blair/Milwaukee waterways [URS001]: metals and conventional analyses. Tacoma, WA.

Windom, H. (1995) An evaluation of ambient sediment quality in the Brunswick Harbor and Savannah River entrance channels, Georgia. Excerpts from an unpublished report. Chemical data. Skidway Institute of Oceanography, Savannah, GA.

Wolfe, DA; Bricker, SB; Long, ER; et al. (1994) Biological effects of toxic contaminants in sediments from Long Island Sound and environs. NOAA Technical Memorandum NOS ORCA 80, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Word, JQ; Ward, JA; Squires, AL. (1990) Results of chemical, toxicological, and bioaccumulation evaluations of dioxins, furans, and guiaicol/organic acids in sediments from the Grays Harbor/Chahalis R. Prepared by Battelle Memorial Institute, Battelle/Marine Sciences Laboratory. Prepared for the U.S. Army Corps of Engineers and the U.S. Department of Energy.

APPENDIX A2: REFERENCES FOR FRESHWATER SEDTOX02 DATABASE

Abraham, WJ; Preddice, T; Kuzia, EJ; et al. (1993) Lower Genesee River Project phase I report: summary 1992 results. New York State Department of Environmental Conservation.

Anderson, B; Hunt, J; Phillips, B; et al. (1998) Sediment chemistry, toxicity, and benthic community conditions in selected water bodies of the Los Angeles Region. Final report. California State Water Resources Control Board.

Ankley, GT; Leonard, E; Glass, G; et al. (1990) Chemical and biological assessment of sediments from Oswego Harbor, 1990: final report. New York State Department of Environmental Conservation.

Ankley, GT; Collyard, SA; Monson, PD; et al. (1994) Influence of ultraviolet light on the toxicity of sediments contaminated with polycyclic aromatic hydrocarbons. Environ Toxicol Chem 13(11).

Applied Biology, Inc. (1982) Chemical, physical and bioassay analysis of sediment samples, Erie Harbor, Erie, Pennsylvania.

Aqua Tech Environmental Consultants, Inc. (1985) The analyses of sediments from Rochester Harbor. Prepared for U.S. Army Corps of Engineers, Buffalo.

Aqua Tech Environmental Consultants, Inc. (1986) The analyses of sediments from Cleveland Harbor, Cleveland, Ohio. Prepared for U.S. Corps of Engineers, Buffalo.

Aquatec Biological Sciences. (1997) Results of *Chironomus tentans* survival and growth toxicity tests for the 1997 Lake Ontario sediment and biota survey BTRs 1275&1279. RECRA LabNet, New York State Department of Environmental Conservation, Aquatec Biological Sciences.

AScI Corporation/AScI-Duluth Environmental Testing Division. (1996) Results of chronic toxicity testing with whole sediment samples from IT Corporation-Interlake Project. International Technology Corp.

Bedard, D; Petro, S. (1995) Laboratory sediment bioassay report on Wheatley Harbour sediments 1992. Ontario Ministry of Environment and Energy.

Bedard, D; Petro, S. (1997) Laboratory sediment bioassay report on St. Mary's River sediments 1992 & 1995. ISBN 0-7778-6674-9; Standards Development Branch, Ontario Ministry of Environment and Energy.

Bedard, D; Petro, S. (1997) Laboratory sediment bioassay report on Upper St. Clair River sediments in the vicinity of industrial point sources 1994 & 1995. St. Clair River remedial action plan.

Bedard, D; Petro, S. (1998) Laboratory sediment bioassay report on Lyons Creek sediments. ISBN -7778-7195-5; Ontario, Ministry of the Environment.

Bennett, J; Cubbage, J. (1992) Copper in sediments from Steilacoom Lake, Pierce County, Washington. Washington State Department of Ecology. Environmental Investigations and Laboratory Services Program, Olympia, WA.

Bennett, J; Cubbage, J. (1992) Effects of polycyclic aromatic hydrocarbons (PAHs) in sediments from Lake Washington on freshwater bioassay organisms and benthic macroinvertebrates. New York State Department of Environmental Conservation.

Borgmann, U; Norwood, WP. (1997) Identification of the toxic agent in metal-contaminated sediments from Manitouwadge Lake, Ontario, using toxicity-accumulation relationships in *Hyalella azteca*. Can J Fish Aqua Science 54.

Bortleson, G. (1994) Sediment-quality assessment of Franklin D. Roosevelt Lake and the upstream reach of the Columbia River, Washington, 1992. Personal Comm. Gil Bortleson, USGS, Tacoma, WA.

Bowser, M. (1992) Sediment sampling and testing, Erie Harbor, Pennsylvania. Prepared for U.S. Army Corps of Engineers, Buffalo District.

Brady, MD. (1994) Baseline sediment characterization Columbia Aluminum Company, Goldendale, Washington. ENSR Consulting and Engineering.

Burton, D. (1994) Sediment sampling and analysis plan - West Branch Grand Calumet River. U.S. EPA, Department of Justice.

Call, DJ; Balcer, MD; Brooke, LT; et al. (1991) Sediment quality evaluation in the Lower Fox River and Southern Green Bay of Lake Michigan. Center for Lake Superior Environmental Studies. University of Wisconsin-Superior. Superior, WI.

Carter, JA; Mroz, RE; Tay, KL; et al. (1998) An evaluation of the use of soil and sediment bioassays in the assessment of three contaminated sites in Atlantic Canada. Water Qual Res J Canada.

Crane, JL; Schubauer-Berigan, M; Schmude, K. (1997) Sediment assessment of hotspot areas in the Duluth/Superior Harbor. Minnesota Pollution Control Agency.

Cubbage, J. (1992) Survey of contaminants in sediments in Lake Union and Adjoining Waters (Salmon Bay, Lake Washington Ship Canal, and Portage Bay). Urban Bay Action Team, Northwest Regional Office, Washington State Department of Ecology.

Dickson, KL; Waller, WT; Kennedy, JH; et al. (1989) A water quality and ecological survey of the Trinity River. University of North Texas and University of Texas. Prepared for City of Dallas Water Utilities.

Downing, J; Fairey, R; Roberts, C; et al. (1998) Chemical and biological measures of sediment quality in the Central Coast Region: final report. California State Water Resources Control Board.

Earth Tech Inc. (1998) Phase II, Flat River sediment investigation. Washington State Department of Ecology, Water Body No. WA-08-9350 (Segment No. 04-08-001).

EG & G, Bionomics. (1983) An assessment of the chemical and toxicological properties of dredged sediments collected from Cleveland Harbor, Ohio. Prepared for U.S. Army Engineering District, Buffalo.

ENSR Consulting and Engineering. (1994) Aluminum Company of America; Vancouver Works baseline sediment characterization. ENSR Consulting and Engineering, Document # 0225-006-405.

EVS Environment Consultants. (1997) Risk assessment of Sheboygan River, Wisconsin, 1997. Prepared for National Oceanic and Atmospheric Administration.

FishPro Engineering and Environmental Consultants. (1991) biological report on sediment and water bioassays and benthic community determination at Unimar Yard 1 dry dock facility.

Garabedian, B; Estabrook, F; Neuderfer, G; et al. (1997) Oswego River sediment study: Final report - Summary of May 23, June 28 and September 13-15, 1994 Results. Prepared for Department of the Army, Buffalo District.

Geraghty & Miller. (1997) Confirmation sampling results, Green Bay Paint sludge site, Menominee, Michigan. Geraghty & Miller Inc. Environmental Services, Michigan Department of Environmental Quality.

Hall, LW, Jr; Ziegenfuss, MC; Fischer, SA; et al. (1992) A pilot study for ambient toxicity testing in Chesapeake Bay. Year 2 report. Prepared for the U.S. Environmental Protection Agency for the Chesapeake Bay Program. Prepared by the University of Maryland System, Queensland, Maryland.

Hefty, LN. (1998) Toxicity testing using *Hyalella azteca*. Department of the Army, Buffalo District, Corps of Engineers.

Hunt, J; Anderson, B; Phillips, B; et al. (1998) Sediment quality and biological effects in San Francisco Bay: final report. California State Water Resources Control Board.

Ingersoll, CG; Buckler, DR; Crecelius, EA; et al. (1992) Draft final report : Biological assessment of contaminated Great Lakes sediment. Prepared for the US EPA GLNPO Assessment and Remediation of Contaminated Sediments (ARCS) Project.

Ingersoll, CG; Brumbaugh, WG; Farag, AM; et al. (1993) Effects of metal-contaminated sediment, water, and diet on aquatic organisms. Final report for the USEPA Milltown endangerment assessment project. NTIS, PB93-21529, NFCRC-U.W. Springfield, VA.

Ingersoll, CG; Haverland, P; Brunson, E; et al. (1996) Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironomus riparius*. Prepared for the US EPA GLNPO Assessment and Remediation of Contaminated Sediments (ARCS) Project.

Ingersoll, CG; Brunson, EL; Dwyer, FJ; et al. (1998) Sublethal endpoints in sediment tests with *Hyalella azteca*. Environ Toxicol Chem 17:1508–1593.

Jaagumagi, R; Petro S. (1992) An in-place pollutants study of the Otonabee River and Rice Lake. Queens Printer for Ontario. ISBN 0-7729-8010-1.

Jaagumagi, R; Bedard, D. (1997) Sediment and biological assessment of Canagagigue Creek at the Uniroyal Chemical Ltd. Plant, Elmira, Ontario 1995-96. ISBN 0-7778-6938-1; Environmental Monitoring and Reporting Branch and Standards Development Branch, Ontario Ministry of Environment and Energy.

Johnson, A. (1992) Review of metals, bioassay, and macroinvertebrate data from Lake Roosevelt benthic samples collected in 1989. Department of Ecology, State of Washington.

Johnson, A; Norton, D. (1988) Screening survey for chemical contaminants and toxicity in sediments at five lower Columbia River ports, September 22-24, 1987. Washington State Department of Ecology.

Johnson, A; Norton, D. (1989) Screening survey for chemical contaminants and toxicity in drainage basins at Paine Field, August 10-12, 1987. Washington State Department of Ecology.

Johnson, A; Norton D; Yake, B. (1989) An assessment of metals contamination in Lake Roosevelt. Washington State Department of Ecology. Toxics Investigations/Groundwater Monitoring Section, Olympia, WA.

Kemble, NE; Burton, EL; Canfield, TJ; et al. (1997) An assessment of sediments from the Upper Mississippi River, Final report - June 1997. EPA 823-R-97-005.

Kemble, NE; Hardesty, DK; Ingersoll, CG; et al. (2000) An evaluation of the toxicity of contaminated sediments from Waukegan Harbor, Illinois following remediation. Arch Environ Contam Toxicol 39:451–462.

Kolok, AS; Plaisance, EP; Abdelghani, A. (1998) Individual variation in the swimming performance of fishes: an overlooked source of variation in toxicity studies. Environ Toxicol Chem 17 (2).

Kosmond, LD. (1996) Lower Fox River system sediment characterization: Sediment quality triad assessment and application of sediment quality guidelines. Wisconsin Department of Natural Resources, Sediment Management and Remedial Techniques Team.

Krantzberg, G. (1994) Spatial and temporal variability in metal bioavailability and toxicity of sediment from Hamilton Harbour, Lake Ontario. Environ Toxicol Chem 13(10):1685–1698.

Landau Associates. (1993) Mill Creek and East Drain sediment sampling and analysis report, western processing, phase II (Draft Report). Landau Associates.

Leppanen, CJ. (1997) Effects of North Hollywood dump leaching on Wolf River sediment toxicity, sediment quality and benthic macroinvertebrate community. University of Memphis.

Malueg, KW; Schuytema, GS; Krawczyk, DF; et al. (1984) Laboratory sediment toxicity tests, sediment chemistry and distribution of *benthic macroinvertebrates* in sediments from the Keweenaw Waterway, Michigan. Environ Toxicol Chem 3:233–242.

Malueg, KW; Schuytema, GS; Gakstatter, JH; et al. (1984) Toxicity of sediments from three metal-contaminated areas. Environ Toxicol Chem 3:279–291.

McGee, BL; Schlekat, CE; Reinharz, E. (1993) Assessing sublethal levels of sediment contamination using the estuarine amphipod *Leptocheirus plumulosus*. Environ Toxicol Chem 12:577–587 (E).

Minnesota Pollution Control Agency. (1993) Toxicity test results of sediment samples collected in 1993 from the Interlake/Duluth Tar and USX Superfund sites.

New York State Department of Environmental Conservation (NYSDEC). (1995) Phase II - final report: Lower Genesee River study: Summary of 1992, 1993 and 1994 results. U. S. Army Corps of Engineers Buffalo District, U. S. EPA, University of Wisconsin-Superior, AScI Corporation.

Pethybridge, AH. (1981) Sediment quality testing at Ogdensburgh Harbor, New York. Technical Report #I0175-04. Prepared for U.S. Army Engineer District, Buffalo.

Phelps, HL. (1993) Sediment toxicity of the Anacostia River estuary, Washington, DC. Bull Environ Contam Toxicol 51.

Pope, RJ. (1993) Analysis of benthic invertebrates from Algoma Slip Core Samples. Water Resource Branch, The Ontario Ministry of the Environment and Energy.

Prairie, R; Mckee, P. (1992) Sediment quality assessment near two Ontario mine sites: How relevant are the provincial guidelines? Paper presented at the International Land Reclamation and Mine Drainage Conference and the Third International Conference on the Abatement of Acidic Drainage.

PTI Environmental Services. (1992) McCormick & Baxter creosoting company. Remedial investigation report. Volume IV of IV. Prepared for Oregon Department of Environmental Quality. Portland, Oregon.

Redsike, R; Fahnenstiel, G; Meier, P; et al. (1997) Preliminary investigation of the extent and effects of sediment contamination in White Lake near the Whitehall leather tannery. U.S. Environmental Protection Agency, National Oceanic and Atmospheric Administration.

REMAP 95. (1997) Ten-day toxicity tests with *Hyalella azteca* and *Chironomus tentans* on sediments from the St. Louis River area of concern 1995 R-EMAP test results batch #1-7. Minnesota Pollution Control Agency, Monitoring and Assessment Section.

REMAP 96. (1997) Ten-day toxicity tests with *Hyalella azteca* and *Chironomus tentans* on sediments from the St. Louis River area of concern 1996 R-EMAP test results batch #1-7. Minnesota Pollution Control Agency, Monitoring and Assessment Section.

Simonin, HA; Paul, E; Neuderfer, G; et al. (1996) The effects of long-term copper sulfate use on the toxicity of lake sediments. Presented at Society of Environmental Toxicology and Chemistry Annual Meeting, Washington, DC.

Smith, JA; Glowacky, RS; Crerar, PJ; et al. (1984) Analysis of sediment from Toledo Harbor - Maumee River. Aqua Tech Environmental Consultants, Inc. Department of the Army, Buffalo District, Corps of Engineers.

Smith, JA; Glowacky, RS; Crerar, PJ; et al. (1984) Analysis of sediment, Oak Orchard Harbor, Oak Orchard, New York. Aqua Tech Environmental Consultants, Inc. Prepared for Department of the Army, Buffalo District, Corps of Engineers.

Smith, JA; Glowacky, RS; Crerar, PJ; et al. (1985) Analyses of Sediments from Huron Harbor, Huron, Ohio. Prepared for Department of the Army, Buffalo District, Corps of Engineers. Prepared by Aqua Tech Environmental Consultants, Inc.

Smith, JA; Crerar, PJ; Glowacky, RS; et al. (1985) Analysis of sediment from Sandusky Harbor, Sandusky, Ohio. Aqua Tech Environmental Consultants, Inc. Prepared for Department of the Army, Buffalo District, Corps of Engineers.

Smith, JA; Crerar, PJ; Glowacky, RS; et al. (1985) The analyses of sediments from Conneaut Harbor, Conneaut, Ohio. Aqua Tech Environmental Consultants, Inc. U.S. Army Corps of Engineers, Buffalo District.

Smith, JA; Glowacky, RS; Crerar, PJ; et al. (1986) The analyses of sediments from West Harbor, Ottawa County, Ohio. Prepared for Department of the Army, Buffalo District Corps of Engineers. Prepared by Aqua Tech Environmental Consultants, Inc.

Smith, DL; Talbot, LM; Campbell, CK. (1992) Contaminated sediment bioassay (toxicity tests) study of Wisconsin Great Lakes coastal harbors and tributaries, draft.

Smith, JA; Glowacky, RS; Crerar, PJ; et al. (1996) The analyses of sediments from Fairport Harbor. Prepared for Department of the Army, Buffalo District, Corps of Engineers. Prepared by Aqua Tech Environmental Consultants, Inc.

Sprenger, MD; Kracko, K. (1992) Final report for supplemental environmental investigation, Cannelton Industries site, Sault Ste. Marie, MI. U.S. Environmental Protection Agency.

Tetra Tech. (1994) Lower Columbia River backwater reconnaissance survey Lower Columbia River bi-state program.

T.P. Associates International, Inc. (1987) The analyses of sediments from Oswego Harbor. Minnesota Pollution Control Agency.

T.P. Associates International, Inc. (1987) The analyses of sediments from Oak Orchard Harbor. Prepared for Water Quality Section, U.S. Army Engineer District, Buffalo.

T.P. Associates International, Inc. (1987) The analyses of sediments from Olcott Harbor. Prepared for Department of the Army, Buffalo District, Corps of Engineers.

T.P. Associates International, Inc. (1988) The analyses of sediments from Wilson Harbor. Prepared for U.S. Army Engineer District, Buffalo.

U. S. Army Engineer District, Chicago. (1994) Clean Water Act Section 404 (b) (1) Contaminant Determination for Burns Waterway Harbor. U. S. Army Engineer District, Chicago.

U.S. EPA (Environmental Protection Agency). (1996) Assessment of sediment in the Indiana Harbor area of concern.

U.S. EPA. (1997) Field sampling methodologies: Clinton River, MaComb County, Michigan (Stations CL9716-CL9727). U.S. EPA, Great Lakes National Program Office.

Wade, TL; Velinsky, DJ; Reinharz, E; et al. (1994) Tidal River sediments in the Washington, D.C. area. II. Distribution and sources of organic contaminants. Estuarine Research Federation.

Webster, CD; McLaughlin, JC; Wan, L; et al. (1995) A report of the whole sediment toxicity of 12 sites in the Maumee area of concern to *Hyalella azteca*. Ohio EPA.

WENCK Associates, Inc. (1995) Harbor sediment sampling documentation report. Prepared for Lakehead Pipe Line Company.

APPENDIX B: SEDIMENT TOXICITY DATABASE (SEDTOX02) STRUCTURE

The following tables describe the structure for the database management system developed for the SEDTOX02 database (NOAA, 2004). Above each table is the name of the database table and a brief description of the purpose. Following the database table name is a "key" that describes how unique records in the table are defined. The SEDTOX02 database is divided into separate databases with identical structure for marine/estuarine and freshwater data.

Site: Table to define general location
Key: Siteid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
SITENAME	Char	40	Descriptive name for site
EPAREGION	Num	2, 0	Region for site location; 11 for Canada

Study: Provides basic information regarding the study (e.g., name, contact, etc.).
Key: Siteid+Studyid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
STUDYNAME	Char	40	Short name of study
CONTACT	Char	40	Contact source for data
SEDCHEM	Logical	1	Sediment chemistry data, Y or N?
SEDTOX	Logical	1	Sediment toxicity data, Y or N?

Studynot: Primarily a table for descriptive notes regarding the study design and method for recording data in the database. Information may include how replicates are recorded and any chemical sums calculated.
Key: Siteid+Studyid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
NOTES	Memo	10	Memo field data processing notes

Studyref: Contains information regarding the document that describes the study and data.
Key: Siteid+Studyid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
YEAR	Char	4	Year of publication
AUTHORS	Char	160	Authors, if published
TITLE	Char	160	Title, if published
SOURCE	Char	160	Citation or agency, if published

Station: Listing of stations for sediment chemistry, tissue chemistry, and bioassay sample collections.
Key: Siteid+Studyid+Stationid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
STATIONID	Char	6	Station identifier
LATITUDE	Num	12, 8	Latitude in decimal degrees, NAD83
LONGITUDE	Num	13, 8	Longitude in decimal degrees, NAD83
EST_STN	Char	8	How coordinates were established

Sample: Sediment sample collection information, including station id, sample date, sample depth (in centimeters). Field sample lab replicates are treated as separate samples.
Key: Siteid+Studyid+Stationid+Sampleid+Labrep

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
STATIONID	Char	6	Station identifier
SAMPLEID	Char	2	Sample identifier
LABREP	Char	2	Lab replicate number
SAMPDATE	Char	8	Date sample collected as YYYYMMDD
SAMPTIME	Char	5	Time sample collected
UDEPTH	Num	8, 2	Top depth of sample from sed/water interface in cm
LDEPTH	Num	8, 2	Bottom depth of sample from sed/water interface in cm
TOC	Num	6, 2	Total organic carbon as percent
PCTFINES	Num	6, 2	Percent fines
UAN_PW	Num	10, 4	Unionized ammonia in porewater
H2S_PW	Num	10, 4	Hydrogen sulfide in porewater
EXSAMPID	Char	12	Investigator's sample identifier

Chem: Chemistry data associated with surface sediment samples.
Key: Siteid+Studyid+Stationid+Sampleid+Labrep+Chemcode

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
STATIONID	Char	6	Station identifier
SAMPLEID	Char	2	Sample identifier
FIELDREP	Char	2	Field replicate number
LABREP	Char	2	Lab replicate number
CHEMCODE	Char	10	Code for parameter name
CONC	Num	12, 5	Measured concentration
QUALCODE	Char	5	Assigned qualifier for concentration
UNITS	Char	6	Units of concentration for parameter
MEASBASIS	Char	2	Wet (WW) or dry weight (DW) indication
MISSINGVAL	Logical	1	Data missing, Y or N?

Biosumm: Mean results (of replicate data) for sediment bioassay.
Key: Siteid+Studyid+Stationid+Sampleid+Testid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
STATIONID	Char	6	Station identifier
SAMPLEID	Char	2	Sample identifier
TESTID	Char	12	Bioassay test code
FIELDREP	Char	2	Field replicate number
GROUP	Char	2	Sample grouping
SERIES	Char	2	Bioassay test series number
EFFECTVAL	Num	7, 2	Measured effect value
SIGEFFECT	Logical	1	Was effect significant, Y or N?
NEG	Logical	1	Negative control sample, Y or N?
REF	Logical	1	Reference sample, Y or N?
STAT	Logical	1	Used for statistical comparison, Y or N?
CTRLADJ	Num	6, 2	Control-adjusted effect value
SIG_ORIGIN	Char	2	Code for toxic sample used by original study

Bmaster: Descriptive notes associated with bioassay tests where available.
Key: Siteid+Studyid+Testid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
GROUP	Char	2	Sample grouping based on spatial or temporal
TESTID	Char	12	Bioassay test code
SPIKED	Logical	1	Sediment spiked with contaminant, Y or N?
TESTCOMM	Memo	10	Bioassay test comments

Qualify: Defines qualifiers used with chemical data

Key: Siteid+Studyid+Qualcode

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
SITEID	Char	4	Site identifier
STUDYID	Char	2	Study identifier
QUALCODE	Char	5	Assigned qualifier code for concentration
QUALIFIERS	Char	30	Qualifiers used in original study
DESCRIPT	Char	80	Description of qualifiers used in study

Chemdict: Provides a unique list of chemical names and associated chemical codes that are used in the Chem table.
Key: Chemcode

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
CHEMCODE	Char	10	Code for parameter name
CHEMNAME	Char	40	Full chemical name
CHEMCLASS	Char	8	Chemical classification
CATEGORY	Char	8	Alternate chemical classification
SUBCATGY	Char	10	Subclassification for alternate chemical class
CHEMTOTAL	Char	10	Classification used for totaling chemicals
MOLWT	Num	7, 3	Molecular weight of chemical
CASNUM	Char	24	CAS number
UNITS	Char	6	Units of concentration for chemical (sed/tiss)
WA_UNITS	Char	6	Units of concentration for chemical in water media

Testdict: Provides a unique list of bioassay tests and associated test codes that are used in the Biosumm and Bmaster tables.
Key: Testid

FIELD NAME	TYPE	WIDTH, DEC	DESCRIPTION
TESTID	Char	12	Bioassay test code
MEDIUM	Char	15	Medium tested
MEDCODE	Char	2	Code for medium tested
GROUP	Char	20	Bioassay species grouping
ALTGROUP	Char	20	Alternate bioassay species grouping
SPECIES	Char	40	Bioassay organism
SPPCODE	Char	3	Code for bioassay organism
LHS	Char	10	Life history stage of bioassay organism
LHSCODE	Char	1	Code for life history stage
ENDPOINT	Char	30	Bioassay test endpoint
ENDCODE	Char	2	Code for test endpoint
DURATION	Char	10	Duration of test
DURCODE	Char	4	Code for test duration

APPENDIX C: DATA ACQUISITION SCREENING METHODS

The SEDTOX02 database was developed to support the development and assessment of numerical sediment quality guidelines. The database is composed of matching (i.e., synoptically collected) sediment chemistry and laboratory toxicity data from freshwater, estuarine, and marine sites. The following screening criteria provided a means of evaluating candidate data sets and ensuring general consistency in the information included in the database. The evaluation criteria were based on ASTM (2002a, b), Environment Canada (1998a, b), and U.S. EPA (2000). However, the screening criteria are not necessarily recommended for applications beyond their intended purpose. Data from spiked-sediment bioassays were not included in the database.

A. Approach for Evaluating Data Set Acceptability

1. Data sets must contain synoptically collected sediment chemistry and biological effects data. That is, the sediment samples for biological and chemical testing must have been collected from the same location and at the same time.

2. Data sets may contain any number of sediment samples, provided that there is at least one nontoxic sample. Preference should be given to data sets that contain ≥5 samples.

3. Data sets that include toxicity and chemistry data generated on sediment samples from any sediment horizon (i.e., surficial sediments, cored sediments, etc.) should be considered to be acceptable provided that the sediment chemistry and toxicity data are matching (i.e., for the same sediment samples).

4. Data sets that include toxicity and chemistry data generated on composite sediment samples should be considered to be acceptable. Preference should be given to data sets that composite sediments over limited geographic areas.

5. Data sets that include data generated from dilution series of bulk sediments and/or porewater are not acceptable for incorporation into the database (however, data from the 100% dilution are acceptable).

6. It is not essential that data sets include the coordinates (i.e., longitude and latitude) of the sampling site along with the chemical and biological data. However, these data will be included as available.

7. It is not essential that data sets have a minimum range of chemical concentrations in the sediment samples (i.e., the 10-fold criteria that was used previously to assess data acceptability is no longer required).

8. Data sets should be preferentially included if full sediment chemistry has been conducted (i.e., metals, PAHs, PCBs, and pesticides). Data sets with incomplete sediment chemistry may also be included, particularly if prior knowledge indicates that certain chemicals are unlikely to occur at the site.

9. Analytical detection limits should be below the respective ERLs or TELs for each chemical analyte.

10. The chemical analytical methods used in the study must be reported and should meet minimum data quality requirements/objectives (i.e., the precision, accuracy, and detection limits must be reported; data quality may be evaluated using various protocols and best professional judgement; and the rationale for decisions regarding data acceptability must be documented).

11. Concentrations of SEM metals (e.g., Cd, Cu, Pb, Ni, Zn) may be included in the database, provided that data on the concentrations of total metals (i.e., strong acid digestion) are also available.

12. Information on the environmental conditions in the bioassay chambers should be captured in the database, including data on DO, pH, salinity, water hardness, temperature, NH_3, and H_2S.

13. Chemistry data that were generated using atypical methods (e.g., X-ray fluorescence for metals, screening methods for PAHs, etc.) are not acceptable for inclusion in the database.

14. Data from elutriate tests must not be included in the database because there is too little connection between the chemistry and the laboratory toxicity data.

15. Data sets generated using organic extracts may be included in the database (e.g., Microtox), provided they are available with other toxicity data. Microtox and Mutatox data are not being targeted for inclusion in the database because the linkage between the toxicity data and effects on sediment-dwelling organisms is tenuous.

16. Acceptable environmental conditions must be maintained throughout the toxicity test (as defined in the protocol for the toxicity test). Consequently, the temperature, pH, hardness, conductivity, salinity, and DO of the overlying water should have been measured during the test. If these variables have been measured but not reported, it is reasonable to assume that the conditions during the test **were not acceptable** and additional information should be obtained from the investigators.

17. The responses of the test organisms exposed to negative controls must be reported and must be within acceptable limits (i.e., as defined in ASTM standard methods). For toxicity tests for which a negative control sediment is not available, the selected field reference sediment must be shown to be functionally equivalent to a negative control sediment, as indicated by nontoxicity (as defined above); concentrations of measured contaminants should not exceed their respective TELs or ERLs; and the levels of particle

size distribution, pH, Eh, salinity, and TOC must be similar to those in the basin area under investigation.

B. Considerations for Prioritizing Data Sets

1. The procedures used for collecting, handling, and storing sediments should be consistent with the protocols that have been established by ASTM. Generally, higher priority for inclusion in SEDTOX should be assigned if

    (i) surficial sediments were collected and tested;

    (ii) the sediments were tested within 8 weeks of collection (some flexibility in applying this criterion is warranted, as similar bioassay responses have been observed up to 1 to 2 years after sediment collection; and,

    (iii) the sediments were not frozen prior to biological testing.

2. Data sets should be preferentially included if full sediment chemistry has been conducted (i.e., metals, PAHs, PCBs, and pesticides). Data sets with incomplete sediment chemistry may also be included, particularly if prior knowledge indicates that certain chemicals are unlikely to occur at the site.

3. Data on all test organisms and endpoints for which standard toxicity testing methods are available should be captured in the database, as available. However, higher priority should be given to data sets that include one or more of the following tests/endpoints:

    - marine amphipod (*Ampelisca* and *Rhepoxynius*) survival;
    - marine sea urchin (*Arbacia* and *Strongylocentrotus*) fertilization;
    - freshwater amphipod (*Hyalella*) survival, growth, and reproduction; and,
    - freshwater midge (*Chironomus riparius* and *C. tentans*) survival and growth.

    The data on other species and associated endpoints should be captured in the database when available, along with data on the high-priority toxicity tests.

4. Priority should be given to data sets that test surficial sediments and do not composite samples over large geographic areas.

C. Considerations Related to Preparing Data for Import

1. Sediment samples from the same study for which inconsistent chemistry data are available (i.e., metals only for some samples and complete chemistry for other samples) should be grouped separately. The portion of the data set with complete sediment chemistry should be preferentially included in the database. The portion of the data set with incomplete chemistry should be considered as a separate data set.

2. Detection limits (DLs) should be reported for all measured chemicals (i.e., for all analytes for which the concentration in one or more samples is reported to be < DL). Below-DL values must be treated as missing data if the DL has not been reported.

3. Calculations of tPAH, tHMW-PAH, or tLMW-PAH will be conducted using subroutines in the database. Previously calculated values will not be incorporated into the database. LMW-PAHs are considered to include the following two- and three-ringed substances: naphthalene, acenaphthene, acenaphthylene, fluorene, phenanthrene, and anthracene. HMW-PAHs are considered to include the following four- and five-ringed substances: fluoranthene, pyrene, benz(a)anthracene, chrysene, benzo(a)pyrene, benzofluoranthene, dibenzo (a,h)anthracene, and benzo(g,h,i)perylene.

4. The reported concentrations of tPCBs should be treated as equivalent, regardless of which method was used to determine the levels. If tPCB concentrations were not reported, the value should be calculated using an appropriate method (e.g., sum of detected Aroclors or sum of detected congeners, etc.). The method that was used in the determination should be recorded in the database.

5. The reported concentrations of tDDTs should be treated as equivalent, regardless of which method was used to determine the levels, provided that the p,p'-isomers of DDT, DDE, and DDD were measured. The method that was used in the determination should be recorded in the database.

6. The reported concentrations of tPCDDs/PCDFs should be treated as equivalent, regardless of which method was used to determine the levels, provided that the most toxic substances were measured (i.e., TCDD, HCDD, TCDF, HCDF). The method that was used in the determination should be recorded in the database.

7. The data on the levels of H_2S and NH_3 in the replicate bioassay chambers will be summarized on a per-sample basis and included in the database. For these variables, all of the measurements should be treated as equivalent, regardless of the analytical methods that were used.

8. Data from tests conducted with different types of media (i.e., porewater vs. organic extracts for Microtox) should be treated separately in the data analyses.

9. Sediment samples should be grouped in a consistent manner to facilitate data analyses, specifically:

    (i) Samples that were collected from the same area within the same year should be grouped together.

    (ii) Samples that were collected from the same area in different years should be separated into groups, based on the year that the samples were collected.

    (iii) In general, samples that were collected within a single study that was conducted during one year should be grouped together; however, it may be

necessary to create separate groupings for the samples by basin area. For example, EMAP data will be grouped by basin area using maps of appropriate scale.

10. The toxicity of sediment samples from the basin area under investigation should be determined on the basis of statistical comparisons with the negative control. For bulk sediments and porewater, negative control sediments may be obtained from a suitable reference site(s), as specified in the ASTM (2002b) standard methods (e.g., >80% survival and full chemistry) and other relevant information (i.e., contaminant concentrations < ERLs).

11. Sediment samples must be designated as toxic or nontoxic using the results of statistical analyses. Negative control data should also be provided for each batch of samples to facilitate the determination of toxicity on the basis of minimum significant differences from the negative control responses.

12. In tests that are designed to evaluate effects on growth and/or reproduction, samples will be treated as though toxic for these endpoints if a significant effect on survival was determined for that sample (i.e., even if it was not possible to measure growth or reproduction directly or even if it was possible to measure effects on survivors and no effects were observed).

References Cited

ASTM (American Society for Testing and Materials). (2002a) Standard guide for conducting 10-day, static sediment toxicity tests with marine and estuarine amphipods. E 1367-99. In: Annual book of standards. Vol. 11.05. West Conshohocken, PA, pp. 693–719.

ASTM. (2002b) Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. E1706–00B. In Annual book of standards. Vol. 11.05. West Conshohocken, PA, pp. 1125–1241.

Environment Canada. (1998a) Biological Test Method: Test for growth and survival in sediment using the freshwater amphipod *Hyalella azteca*. Environment Canada, Ottawa, Ontario. Technical report number EPS1/RM/33.

Environment Canada. (1998b) Biological Test Method: Test for growth and survival in sediment using larvae of freshwater midges (*Chironomus tentans* or *Chironomus riparius*). Environment Canada, Ottawa, Ontario, Technical report number EPS1/RM/32.

U.S. EPA (Environmental Protection Agency). (2000) Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates, second edition. EPA 600/R-99/064, Washington, DC.

APPENDIX D: DATA EVALUATION METHODS

SCREENING CRITERIA FOR BEDS/SEDTOX

CO-OCCURRENCE DATA

***Must be Present**

Reference:_____ Reference Number:_____

1.* Does data set contain matching sediment chemistry and biological effects (i.e., biological and chemical data collected from the same location at the same time)?
NO ____ UNACCEPTABLE
YES ____ Page Reference(s):

---

2. What is the location of sampling site(s)? Collection Date? Page reference for site description?

---

3. Freshwater ___ Estuarine ___ Marine ___

Salinity_____

I. SEDIMENT CHEMISTRY

4.* Is there at least one nontoxic sample?
NO ___ UNACCEPTABLE

YES ___ Number of nontoxic: ___ Number of toxic:

---

5. Was bioassay conducted on unique ___ or composite ___ samples.
Number of replicates? ___ Size of composite area?

---

6. What chemistry data have been collected? (i.e., metals, PAHs, pesticides, pH, DO, TOC)
Metals ___ PCBs ___ pH ___ TOC ___
PAHs ___ Pesticides ___ DO ___ AVS ___

---

7. Are detection limits below the respective ERLs or TELs?
NO ___ UNACCEPTABLE YES

---

8. Are total metal concentrations measured?
NO ___ SEM metals may not be included.
YES ___ SEM metals may be included.

9. Collection instrument _____ Sediment depth _____.

10.* What type of sediment was used:

Bulk Sediment ___ (Sediment with pore water)
Porewater ___ (Extract porewater from sediments and expose water column species)
Organic Extract ___ (Sediment extracted with organic solvent and expose liquid form)
Elutriate (Sediments with water, mixed, settled and exposed water column species)

OTHER TOXICITY DATA ALSO NEEDED UNACCEPTABLE

---

11. What type of toxicity test was conducted? Length of test?

Static ___ (Water, sed-no change)
Static Renewal ___ (Water, sed-some water change)
Flow-Through ___ (Water, sed-water flowing through)

12.* Are appropriate analytical procedures used to determine total concentrations of the analytes in bulk sediment samples? What method(s) was used?
(Metals: partial digestion, analysis of elutriates or extracts are unacceptable.)

---

13. Is a dilution series used?

NO ___ YES ___ UNACCEPTABLE

14.* Are measured dry weight contaminant concentrations reported? Conversion from wet weight to dry weight concentration may occur ONLY if data on moisture or TOC are provided. Nominal concentrations are unacceptable.

NO ___ UNACCEPTABLE YES ___ Page reference(s):

II. BIOEFFECTS

15.*

a Do toxicity tests employ appropriate laboratory procedures? (ASTM: E1367, E1611, E1706)

NO ___ UNACCEPTABLE YES

b Have the following been recorded during testing?

Temperature _____; pH _____; Hardness _____; Conductivity _____;
Salinity _____; DO _____; Alkalinity _____; Ammonia _____.

c Does DO Remain above 60% ___ Needed for Marine
40% ___ Needed for Fresh Water
NO ___ UNACCEPTABLE

d Temperature

e Is temperature within natural range, fluctuate less then 3 C, and have a time-weighted average within 1 C of selected temp?
YES ___ NO ___UNACCEPTABLE Range

f Do hardness, alkalinity, pH, or ammonia vary by more than 50% (for freshwater samples)?
NO ___YES ___ UNACCEPTABLE
Range: DO _____; Alk _____; pH _____; NH_3

g Have salinity levels in porewater been adjusted (for marine samples)?
NO ___YES ___ UNACCEPTABLE Range

h List procedure reference(s) or brief details:

16.* Were biological responses compared to the control ___ or reference ___ sites?
List the Control and Reference sites? Positive Control _____.
reference = uncontaminated site within the same waterbody or watershed; control = uncontaminated site outside the tested water body

---

17.*

a Have sediment samples used for biological testing been frozen?
NO ___YES ___ If yes, both biological and chemical testing must be performed after thawing sediments.

b Have sediment samples been stored for more than eight (8) weeks prior to biological testing?
NO ___ YES ___ UNACCEPTABLE
What was the holding time? _____ .

c Are appropriate procedures used for collecting, handling, and storage of sediments?
NO ___ YES ___ List procedures reference(s) or brief details:

---

18. Identify species used in toxicity testing. Identify organism sources.

---

19. What life stage were the test species at the start of the test?
(Hyalella azteca 7–14 day old; Chironomus tentans third-instar larvae; Chironomus riparius second instar or younger; Daphnia magna 5 days old; Ceriodaphnia dubia <24h old; Hexagenia spp. 3–4 months old; Tubifex tubifex adult; Diporeia spp. juveniles)

---

20. Organism acclimation time _____.

---

21. What percentage of the control survived?
Mean range _____ 70% for (Chironomus riparius, Chironomus tentans)
80% for (Hexagenia spp., Daphnia magna, Ceriodaphnia dubia, Hyalella azteca)
90% for (Diporeia spp., Tubifex tubifex, Polychaetous annelids, marine amphipods, others)

NO ___ UNACCEPTABLE

---

22. Reference Samples
Survival _____%
Conc. less than TEL and ERLs? YES ___ NO ___

Grain size ___ % sand ___ % silt ___ % clay ___

---

23. Benthic Community Analysis
a Is there a benthic community abundance analysis?
NO ___ YES ___ List taxa (e.g., amphipod, sponges,...) upon which the analysis focuses:

---

b* Do all of the sites within a sampling area have the same general characteristics (i.e., same depth of overlying water, same salinity in overlying water, etc)?
NO ___ UNACCEPTABLE YES ___ Briefly list details:

III. STATISTICAL ANALYSIS

24. Are appropriate statistical procedures reported?
NO ___ YES ___ List procedure reference(s):

Additional Notes/Comments:

APPENDIX E: SENSITIVITY OF P_MAX AND P_AVG MODELS TO DIFFERENT BINNING SCENARIOS

E.1 Introduction

Chapter 5 in the main text discusses the development of multiple chemicals to predict the frequency of toxicity observed in the SEDTOX02 marine amphipod database. The P_Max and P_Avg models are used to calibrate the probability of toxicity calculated from the individual chemical models with the observed incidence of toxicity. This approach uses a nonlinear regression equation fit to data summarized by binning individual data points according to their P_Max or P_Avg values calculated from the individual chemical models. After the data were placed into bins, the median P_Max or P_Avg value of each bin was used as the explanatory variable and the corresponding proportion of toxic samples was used as the dependent variable. Each bin was selected to contain 50 unique P_Max or P_Avg values.

This sensitivity analysis evaluated two questions:

1. How does bin size and binning approach affect the regression coefficients and R-squared value for the P_Max and P_Avg models
2. What is the effect of the median (vs. the minimum, mean, or maximum) of the P_Max value in the bin for the x-axis values? This latter analysis was conducted only for the P_Max model. Results for the P_Avg model would be expected to be similar.

E.2 Methods

Three binning scenarios were identified as reasonable approaches for setting the bin endpoints. Within each binning scenario, variable bin widths were targeted. The complete set of binning scenarios plus bin widths or sample sizes are as follows:

Scenario A

This binning approach takes the set of ordered unique P_Max or P_Avg values in the data set and sorts them so that there are *n* unique P_Max or P_Avg values in each bin. The number of samples in each bin will be greater than or equal to *n*, as there may be samples with duplicate P_Max or P_Avg values. This scenario was investigated for n = 5, 10, 15, 25, 50, 75, 85, and 100. (Note: this method with n = 50 was used in Chapter 5).

Scenario B

This binning approach takes the set of ordered P_Max or P_Avg values (duplicates included) in the data set and sorts them so that there are at least *n* P_Max or P_Avg values in each bin. If the value for the last sample of the bin is the same as the next value, then those duplicate values are included in the first bin. For example, if we are using a target bin size of n = 5 and the ordered P_Max values start 0.2, 0.233, 0.234, 0.236, 0.239, 0.239, 0.239, 0.34, then the first bin would start with 0.2 and would end with 0.239, including all three of the 0.239 values for a sample size of seven. The next bin would start with 0.34. This scenario was investigated for n = 5, 10, 15, 25, 50, 75, 85, and 100.

Scenario C

This binning approach takes the set of ordered P_Max or P_Avg values (duplicates included) and divides them into bins of equal width on the P_Max or P_Avg probability scale. In this scenario, the number of samples per bin may be highly variable, and some bins may contain no samples. This scenario was investigated for bin widths of 0.002, 0.004, 0.01, and 0.02.

The three binning scenarios described above were run on the complete data set consisting of samples that had detected values for 10 or more of the modeled chemicals (n = 2856). All analyses were conducted using the 1X screening approach and the Sig Only classification of toxicity. For each sample size within a binning scenario, the data set was binned and summarized, and the nonlinear least squares regression equation was fit using S-PLUS 2000.

E.3. Results

The regression coefficients and R-squared value are shown in the Tables E-1 through E-3 for P_Max and E-4 through E-6 for the P_Avg models. Each scenario exhibited the same patterns in the data and goodness of fit as bin size decreased and the regression sample size increased: variability around the best fit line increased (a smaller R-squared) with increasing regression sample size (i.e., the more bins and the fewer the number of data points summarized per bin). For example, Figure E-1 shows increasing variability with increasing regression sample size for the Scenario A bin approach using the P_Max values.

All iterations within each scenario resulted in very similar regression coefficients. The models tend to deviate towards the tails of the P_Max or P_Avg range; the biggest differences in the predicted probabilities of toxicity from using a different binning scenario/method will be below P_Max values of 0.2 or above 0.8 (Figure E-2).

The same patterns were apparent across binning scenarios and bin sizes, regardless of whether the median or the maximum P_Max values were used on the x-axis (Tables E-1 and E-2, Figure E-3). However, the differences between the best-fit lines for the different bin sizes varied more when the maximum P_Max values were used than when the median P_Max values were used on the x-axis. This is because the differences between median and maximum values in a bin increase as the bin sizes increase, making the regression data set more different across bin size scenarios. In the end, however, all regression lines were very similar.

Conclusions

The binning scenario and bin sample size have a small effect on the outcome of the P_Max and P_Avg models. The R-squared values vary considerably as a function of regression sample size (as would any goodness-of-fit metric) and is not a reliable measure of the accuracy of toxicity predictions. When compared across the same binning approach, the P_Max model produced slightly higher R-squared values than did the P_Avg model for most sample sizes and binning scenarios.

Table E-1. Sensitivity analysis results for P_Max models; Scenario A: unique P_Max values, vary the minimum unique sample size

Unique sample count	Number of bins[a]	Minimum sample count per bin[b]	Maximum sample count per bin[b]	Coefficients for x = median (P_Max)			Coefficients for x = max (P_Max)			
				intercept	x	x^2	R^2	intercept	x	x^2
5	440	5	20	0.09	0.394	0.356	0.5417	0.089	0.395	0.354
10	220	10	25	0.098	0.371	0.372	0.6977	0.097	0.374	0.367
15	146	15	33	0.096	0.379	0.369	0.763	0.094	0.384	0.362
25	88	25	46	0.108	0.339	0.398	0.8479	0.105	0.344	0.386
50[c]	44	52	79	0.112	0.331	0.405	0.9299	0.106	0.337	0.387
75	29	80	117	0.112	0.329	0.406	0.9344	0.101	0.352	0.362
85	25	98	181	0.109	0.349	0.382	0.9382	0.095	0.387	0.319
100	22	108	147	0.111	0.33	0.41	0.9481	0.099	0.35	0.362

[a]Number of data points used in the regression.
[b]Number of samples summarized within each data point.
[c]Approach used in Chapter 5.

Table E-2. Sensitivity analysis results for P_Max models; Scenario B: nonunique P_Max values, vary the minimum sample size

Unique sample count	Number of bins[a]	Minimum sample count per bin[b]	Maximum sample count per bin[b]	Coefficients for x = median (P_Max)				Coefficients for x = max (P_Max)			
				intercept	x	x^2	R^2	intercept	x	x^2	R^2
5	533	5	13	0.097	0.386	0.362	0.464	0.096	0.387	0.359	0.464
10	274	10	19	0.107	0.346	0.394	0.609	0.106	0.347	0.391	0.6091
15	185	15	25	0.111	0.335	0.401	0.6909	0.11	0.338	0.395	0.6908
25	113	25	29	0.108	0.343	0.397	0.8011	0.106	0.348	0.387	0.8012
50	56	50	79	0.111	0.332	0.405	0.8987	0.106	0.347	0.378	0.8994
75	37	75	145	0.108	0.347	0.39	0.9106	0.096	0.388	0.329	0.9109
85	33	85	116	0.105	0.359	0.38	0.9255	0.093	0.397	0.322	0.9257
100	28	100	149	0.105	0.364	0.373	0.9173	0.092	0.404	0.308	0.918

[a]Number of data points used in the regression.
[b]Number of samples summarized within each data point.

Table E-3. Sensitivity analysis results for P_Max models; Scenario C: equal widths on P_Max scale

Width of bin on P_Max scale	Number of bins[a]	Minimum sample count per bin[b]	Maximum sample count per bin[b]	Coefficients for x = median (P_Max)			
				intercept	x	x^2	R^2
0.002	491	0	21	0.105	0.374	0.342	0.4586
0.004	246	1	28	0.096	0.383	0.368	0.6827
0.01	99	1	47	0.101	0.349	0.404	0.8462
0.02	50	1	92	0.117	0.276	0.477	0.9198

[a]Number of data points used in the regression.
[b]Number of samples summarized within each data point.

Table E-4. Sensitivity analysis results for P_Avg models; Scenario A: unique P_Avg values, vary the minimum unique sample size

Unique sample count	Number of bins[a]	Minimum sample count per bin[b]	Maximum sample count per bin[b]	Coefficients for x = median (P_Avg)			
				intercept	x	x^2	R^2
5	430	5	12	0.11	1.428	-0.525	0.5368
10	215	10	21	0.112	1.421	-0.519	0.7096
15	143	15	32	0.11	1.443	-0.551	0.7765
25	86	25	51	0.109	1.439	-0.538	0.8245
50[c]	43	52	90	0.109	1.417	-0.488	0.8922
75	28	79	136	0.101	1.517	-0.69	0.9068
85	25	91	150	0.105	1.468	-0.59	0.9086
100	21	107	174	0.103	1.496	-0.644	0.912

[a]Number of data points used in the regression.
[b]Number of samples summarized within each data point.
[c]Approach used in Chapter 5

Table E-5. Sensitivity analysis results for P_Avg models; Scenario B: nonunique P_Avg values, vary the minimum sample size

Target sample count	Number of bins[a]	Minimum sample count per bin[b]	Maximum sample count per bin[b]	Coefficients for x = median (P_Avg)			
				intercept	x	x^2	R^2
5	541	5	9	0.104	1.444	-0.533	0.5017
10	276	10	19	0.104	1.457	-0.558	0.6607
15	186	15	28	0.103	1.466	-0.563	0.7231
25	112	25	47	0.103	1.466	-0.571	0.7951
50	56	50	92	0.1	1.513	-0.672	0.854
75	37	75	144	0.098	1.54	-0.723	0.8796
85	33	85	124	0.099	1.515	-0.662	0.9019
100	28	100	147	0.1	1.513	-0.661	0.9114

[a]Number of data points used in the regression.
[b]Number of samples summarized within each data point.

Table E-6. Sensitivity analysis results for P_Avg models. Scenario C: equal widths on P_Avg scale

Width of bin on P_Max scale	Number of bins[a]	Minimum sample count per bin[b]	Maximum sample count per bin[b]	Coefficients for x = median (P_Avg)			
				intercept	x	x^2	R^2
0.002	455	0	26	0.125	1.266	-0.273	0.5358
0.004	228	0	45	0.122	1.274	-0.235	0.7724
0.01	91	0	104	0.117	1.338	-0.351	0.8477
0.02	46	0	204	0.116	1.343	-0.378	0.8993

[a]Number of data points used in the regression.
[b]Number of samples summarized within each data point.

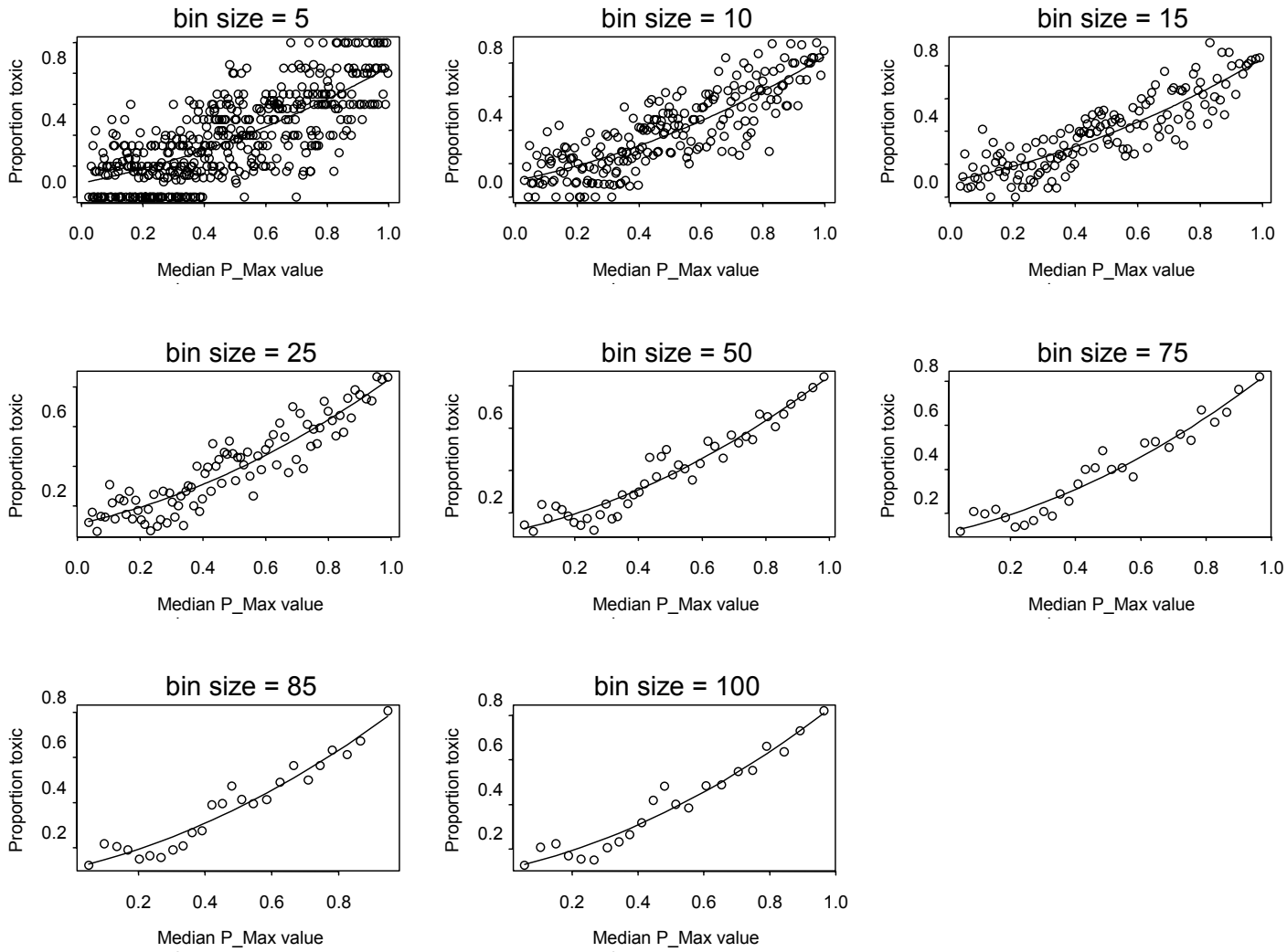


Figure E-1. Proportion of samples observed to be toxic versus the median P_Max value for various bin sizes, where the bin size is the number of unique P_Max values in each bin (Scenario A).

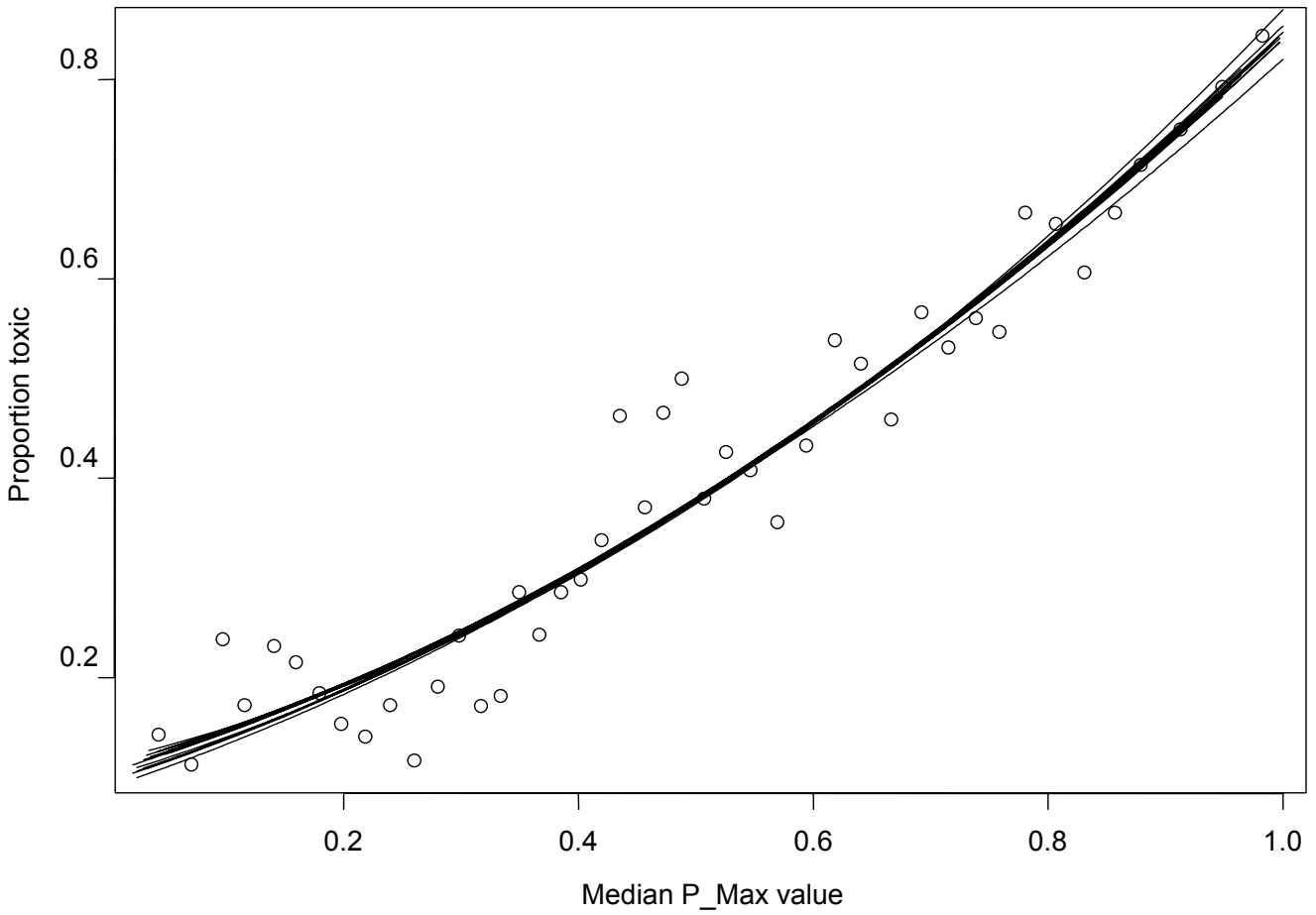


Figure E-2. Best-fit regression lines for all P_Max models. Points shown are for bin size of 50 unique P_Max values (Scenario A).

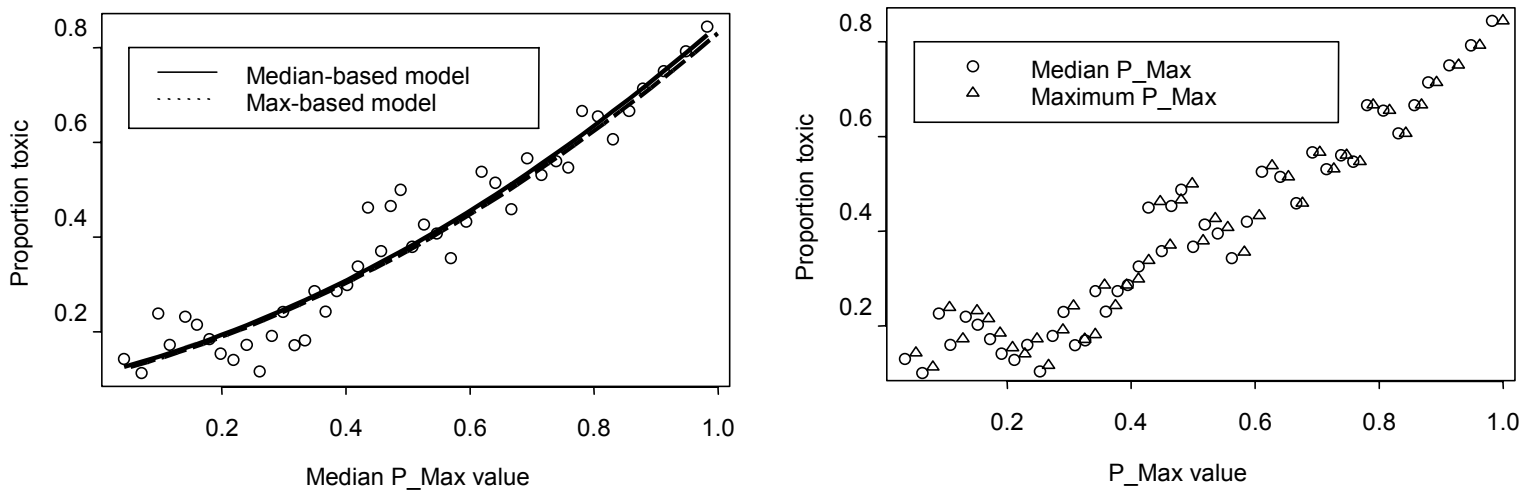


Figure E-3. Comparison of maximum and median P_Max values and resulting models. Points shown are for bin sizes of 50 unique P_Max values (Scenario A).