SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 21, No. 9, pp. 1993–2005, 2002
Printed in the USA
0730-7268/02 $9.00 + .00

PREDICTING AMPHIPOD TOXICITY FROM SEDIMENT CHEMISTRY USING LOGISTIC REGRESSION MODELS

L. JAY FIELD,*† DONALD D. MACDONALD,‡ SUSAN B. NORTON,§ CHRISTOPHER G. INGERSOLL,‖ CORINNE G. SEVERN#, DAWN SMORONG,‡ and REBEKKA LINDSKOOG‡

†Coastal Protection and Restoration Division, Office of Response and Restoration, National Oceanic and Atmospheric Administration, 7600 Sand Point Way North East, Seattle, Washington 98115, USA
‡MacDonald Environmental Sciences, 4800 Island Highway N, Nanaimo, British Columbia V9T 1W6, Canada
§National Center for Environmental Assessment, U.S. Environmental Protection Agency, 1200 Pennsylvania Avenue NW, Washington, DC 20460
‖Columbia Environmental Research Center, U.S. Geological Survey, 4200 New Haven Road, Columbia, Missouri 65201
#Premier Environmental Services, 10999 Pumpkin Ridge Avenue, Las Vegas, Nevada, 89135, USA

(*Received* 3 *October* 2001; *Accepted* 3 *March* 2002)

Abstract—Individual chemical logistic regression models were developed for 37 chemicals of potential concern in contaminated sediments to predict the probability of toxicity, based on the standard 10-d survival test for the marine amphipods *Ampelisca abdita* and *Rhepoxynius abronius*. These models were derived from a large database of matching sediment chemistry and toxicity data, which 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, using either the maximum (P_{Max} model) or average (P_{Avg} model) probability predicted from the chemicals analyzed in a sample, to estimate the probability of toxicity for a sample. The average predicted probability of toxicity (from the P_{Max} model) within probability quartiles closely matched the incidence of toxicity 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. Users have a number of options for applying the logistic models, including estimating the probability of observing acute toxicity to estuarine and marine amphipods in 10-d toxicity tests at any given chemical concentration or estimating the chemical concentrations that correspond to specific probabilities of observing sediment toxicity.

Keywords—Sediment toxicity Sediment guidelines Logistic regression

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 [1,2]. 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 can accumulate to elevated levels in sediments [3]. As such, sediments can serve both as reservoirs and as potential sources of contaminants to the water column. In addition, sediment-associated contaminants can adversely affect sediment-dwelling organisms by causing direct toxicity or altering benthic invertebrate community structure [4]. Furthermore, contaminated sediments can adversely affect fish and wildlife species, either through direct exposure or through bioaccumulation in the food web. While carefully designed monitoring programs can detect effects on sediment communities, fish, and wildlife, the concentrations of chemicals in sediments can provide useful information for evaluating risks to sediment-dwelling organisms, wildlife, or human health from releases of toxic or bioaccumulative substances into the environment.

A variety of indicators provide information on the status of marine and estuarine sediments relative to ecological receptors [3]. These indicators include sediment chemistry, sediment toxicity, benthic invertebrate community status, and bioaccumulation assessments. While the results of sediment toxicity tests and benthic invertebrate community assessments can be used directly to evaluate or infer effects on sediment-dwelling organisms, effective interpretation of sediment chemistry data requires tools that link chemical concentrations to the potential for observing adverse biological effects [5]. Numerical sediment quality guidelines (SQGs) are commonly used in this capacity [6–8].

The logistic regression model (LRM) approach described in this paper is similar to other empirical approaches for deriving SQGs because it relies on matching field-collected sediment chemistry and biological effects (e.g., sediment toxicity or benthic invertebrate community structure effects) data. In contrast to other approaches to developing SQGs, however, the LRM approach does not develop threshold values. Instead, the approach develops models that enable users to select the probability of observing sediment toxicity that corresponds to their specific objectives or to estimate the probability of observing effects at a particular chemical concentration [5].

* To whom correspondence may be addressed (jay.field@noaa.gov).

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the National Oceanic and Atmospheric Administration, the U.S. Environmental Protection Agency, or the U.S. Geological Survey.

Table 1. Number of samples and percentage of samples toxic summarized by amphipod species and data source. 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 (SIG and MSD). NA = no data

	Ampelisca abdita			*Rhepoxynius abronius*		
		Percentage toxic			Percentage toxic	
Data source	No. of samples	SIG only	SIG and MSD	No. of samples	SIG only	SIG and MSD
EMAP[a]	1,203	22.2	9.5	NA	NA	NA
NST[b]	649	23.7	15.6	NA	NA	NA
MLML[c]	43	11.6	7.0	465	72.7	52.3
SEDQUAL[d]	NA	NA	NA	594	63.5	34.0
BEDS[e]	117	41.0	30.8	152	36.8	32.2
Total	2,012	23.6	12.6	1,211	63.7	40.8

[a] EMAP = U.S. Environmental Protection Agency Estuarine Monitoring and Assessment Project.
[b] NST = National Oceanic and Atmospheric Administration Status and Trends program.
[c] MLML = Moss Landing Marine Laboratory, California, USA.
[d] SEDQUAL = Washington State Department of Ecology, Sediment Quality Database.
[e] BEDS = MacDonald Environmental Sciences, Biological Effects Database for Sediments.

Logistic regression models require a large database of matching sediment chemistry and toxicity data that includes a broad range of concentrations. Since the preliminary logistic models were developed [5], the underlying sediment toxicity database has been substantially expanded from 1,200 to 3,200 samples. The primary objectives of this paper are to describe the development of individual chemical logistic regression models for an expanded list of analytes, based on the standard marine and estuarine amphipod 10-d lethality toxicity test endpoint, and to combine these individual models into a single model for predicting toxicity in field-collected sediment samples. In addition, the paper illustrates the applications of the individual logistic models for evaluating sediment quality guidelines and the use of the combined models to predict toxicity for an independent data set.

METHODS

Data acquisition and evaluation

This investigation compiled synoptically collected sediment chemistry and sediment toxicity data from throughout North America. The primary sources of these data included the National Oceanic and Atmospheric Administration's (NOAA) National Status and Trends program (NST), U.S. Environmental Protection Agency's (U.S. EPA) Environmental Monitoring and Assessment Program (EMAP), Moss Landing Marine Laboratory (MLML, which compiled data for the state of California), State of Washington Department of Ecology's Puget Sound Database (SEDQUAL), and MacDonald Environmental Sciences' Biological Effects Database for Sediments (BEDS; Table 1). Many geographic areas along the Atlantic, Gulf, and Pacific coasts are represented in the database. Although the database includes information on a variety of toxicity endpoints, only data from the American Society for Testing Materials (ASTM) standard 10-d amphipod survival toxicity tests with *Ampelisca abdita* and *Rhepoxynius abronius* were used in the analyses discussed in this paper.

All the candidate data sets considered for inclusion in the database were critically evaluated. Acceptance criteria applied 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 [5–7,9]. In the case of the data sets from the NST, EMAP, and SEDQUAL sources, the standard protocols established under each program were evaluated, and individual studies were generally examined to identify possible deviations from these protocols. All the data that met the acceptance criteria were incorporated into the project database. Toxicity data were not included in the database if negative control survival was less than 85%. Data that were compiled in the database were verified against the original data source to ensure that project data quality objectives were met. For data that were acquired electronically (the majority of the data), a minimum of 10% of the data were compared to the source files. All data that were obtained from hard-copy materials (reports and journal articles) were compared to the source documents.

Data treatment and analysis

To support subsequent data analysis, the total concentrations of PCB was calculated for each sediment sample represented in the database. In certain studies, only total concentrations of one or more of these substances were reported; in these cases, the reported values were used directly. The concentrations of total PCBs were determined using procedures that depended on the data reported in the original study. If the concentrations of Aroclors (e.g., Aroclor 1242, Aroclor 1248) were reported, 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 NST [10]. 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 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.

Various methods have been used to designate individual sediment samples as toxic or nontoxic. In this study, individual

sediment samples were designated as toxic if the sample was statistically significant compared to the negative control and survival was less than 90% (the upper limit of the response level based on acceptable negative control response in 10-d marine amphipod toxicity tests [9]). We also evaluated a second commonly used approach for designating individual sediment samples as toxic, based on statistical significance compared to a negative control and a minimum significant difference (MSD) from the control (control adjusted survival of <80% [11]). The individual chemical models that were derived using the significance-only approach described previously are used for the analyses presented in this paper. These models consistently provided slightly better fits to the matching sediment chemistry and toxicity data.

The presence of multiple contaminants, many of which may be present at very low concentrations, frequently complicates evaluating the relationship between the concentration of an individual contaminant and toxicity in field-collected sediments. Consequently, the data for samples that were identified as toxic in this investigation were further screened before being used to develop the logistic models for each individual contaminant [5]. This screening process excluded toxic samples in which the selected contaminant was unlikely to contribute substantially to the observed toxicity. Following the general screening approach used by Ingersoll et al. [12] and similar to that used by others [1,7,13], 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. All nontoxic samples were included in these analyses. 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.

Concentration-interval plots

Concentration-interval plots summarize the matching sediment chemistry and toxicity data for individual contaminants. These plots were prepared by calculating the proportion of toxic samples within discrete concentration intervals. The individual points represent the median of the sample concentrations within the interval and the fraction 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.

Logistic regression modeling

The individual chemical logistic regression models were developed from the screened data set for each chemical, according to the methods described in Field et al. [5]. The data for each chemical consist of the chemical concentration and the toxicity test result (toxic or nontoxic). The model parameters (slope, intercept) define the shape of relationship between the chemical concentration (log10) and probability of a toxic result. In its simplest form, the logistic model can be described using the following equation:

$$p = \frac{\exp(B_0 + B_1(x))}{1 + \exp(B_0 + B_1(x))}$$

where p = probability of observing a toxic effect, B_0 = intercept parameter, B_1 = slope parameter, and x = chemical concentration or log chemical concentration.

This logistic model was applied to the complete screened data (not the concentration-interval summarized data) for a number of substances to develop relationships between the sediment chemistry and the sediment toxicity data. For each of these substances, the intercept (B_0), slope (B_1), and chi-square statistic (-2 log likelihood) were determined. The data for each chemical were modeled independently rather than building a model that analyzed the concentrations of multiple chemicals simultaneously. Thus, only a single concentration variable (x) was used in each individual chemical model. All the logistic regression analyses were conducted using the Statistical Analyses System (SAS®) Institute's logistic procedure [14].

After estimating the model parameters, the model was inverted to estimate the concentrations that yield a certain response probability [5]. The notation T_p (e.g., T50) is used to denote the concentration that would give a response of p percent according to the model (e.g., the T50 represents the chemical concentration at which 50% of the samples would be predicted to be toxic). Confidence intervals for these effect concentrations were derived to describe the uncertainty associated with fitting the model.

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, B_1, was significantly different from zero. For all the models generated, the probability (p value) associated with the slope parameter was <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. 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. Models that had a normalized chi-square value of <0.15 were considered a poor fit and were not used [5].

Multichemical models

Individual chemical logistic regression models were combined to provide a single probability of observing toxicity using two approaches: the maximum probability model (P_{Max}), which was derived from the individual chemical model with the highest probability for a sample, and the average probability model (P_{Avg}), which was derived from the arithmetic mean of the probabilities for all the chemicals with models measured for a sample. The results for both approaches were plotted using the interval approach described earlier, with the difference that the x-axis for these plots represented the median of either the maximum or the mean probability (instead of the concentration) for each interval and included a minimum number of 50 samples per interval. The relationship between the maximum (or mean) probability of toxicity from the individual chemical models and the proportion of toxic samples was described by a least-squares binomial regression model of the

interval data. These binomial models were used to estimate the probability of toxicity for individual samples.

Model evaluation

We used three approaches to evaluate models with acceptable goodness-of-fit values (normalized chi-square value of <0.15). First, we evaluated the internal reliability of the chemical-specific models by comparing the model predictions of the probability of the toxicity to all the information contained in the project database on the toxicity of contaminated sediments to marine amphipods (the data that were screened out of the logistic model development process were included in the reliability evaluation). A similar approach was used to evaluate the reliability of the P_{Max} model. Second, we compared spiked sediment bioassay median lethal concentration (LC50) values from the literature with individual model results. Third, we calculated the probability of observing toxicity for an independent data set (one not used to derive the models) using the P_{Max} model. Model results were compared with toxicity test outcomes by comparing the proportion of toxicity test results observed with that predicted from the models.

RESULTS

Database composition

The database is composed of matching sediment chemistry and toxicity data from the Atlantic, Gulf, and Pacific coasts of North America. The database includes data from 10-d toxicity tests with two species of amphipods (*R. abronius* and *A. abdita*), for which survival is the endpoint that was measured (Table 1). Overall, roughly 39% of the 3,223 sediment samples in the database with matching chemistry and toxicity were toxic to amphipods (percentage survival was $<90\%$ and significantly different from the negative control). For *A. abdita*, 24% of the 2,012 samples were toxic in 10-d tests (Table 1). A higher proportion of the samples (64% of 1,211 samples) tested with *R. abronius* were identified as toxic (Table 1). Using the MSD approach to classifying samples as toxic (percentage control-adjusted survival was $<80\%$ and significantly different from the negative control), 12.6% of the *A. abdita* samples and 40.8% of the *R. abronius* samples were classified as toxic.

The sediment toxicity database includes information on the concentrations of over 300 chemicals of potential concern at contaminated sediment sites. For many of these chemicals, the assembled data span a broad range of chemical concentrations. Table 2 presents the distributions of the chemistry data (10th, 50th, and 90th percentiles) for metals, PAHs, PCBs, and several organochlorine pesticides. These data show that the 10th- to 90th-percentile concentrations of the individual contaminants typically span two to three orders of magnitude, with ranges often spanning four to six orders of magnitude. Percentage total organic carbon in test sediments averaged 1.92% (standard deviation = 2.05, n = 3,117) and ranged from 0.01 to 29.4%.

Logistic regression models for individual chemicals

We derived logistic regression models for individual chemicals to evaluate the relationships between chemical concentrations and sediment toxicity. Data to generate acceptable logistic models were available for 37 substances, including 10 trace metals, 22 individual PAHs, total PCBs, and four organochlorine pesticides (Table 3). In this paper, all the logistic models were generated using dry weight–normalized chemical concentration data because previous analyses indicated that such models fit the amphipod toxicity data as well or better than the organic carbon–normalized models for nonpolar organic contaminants [5]. The data for arsenic, chromium, nickel, and *p,p′*-DDE (dichlorodiphenyldichloroethylene) provided relatively poorer fits with the logistic model. Nevertheless, the normalized chi-square statistic exceeded 0.15 for 37 substances for which logistic models were generated, indicating that the models provide good fits of the amphipod toxicity data.

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. For example, the plots for lead and mercury confirm that logistic models provide good fits of the underlying amphipod toxicity data (Fig. 1). Importantly, the range of concentrations represented in the database appears to span the effects range, as demonstrated by the low incidence of toxicity (0%) at the lowest concentrations and the high incidence of toxicity (90–100%) at the highest chemical concentrations. Similar results were obtained for fluoranthene and phenanthrene; however, the incidence of toxicity tended to be somewhat lower (roughly 90%) at the highest concentrations of these substances.

While 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 corresponded to the 20, 50, and 80% proportion of toxic samples for amphipod survival were determined and designated as the T20, T50, and T80 values, respectively (Table 4). These values provide a framework for evaluating the reliability of the individual models.

Reliability of individual chemical logistic models

The reliability of the chemical-specific logistic models was evaluated by comparing the model predictions of the probability of the toxicity to all the information contained in the project database on the toxicity of contaminated sediments to amphipods. For example, the data screened out of the logistic model development process were included in the reliability evaluation. In this evaluation, the T values derived for each substance were used to define four ranges of chemical concentrations ($\leq$T20, $>$T20–T50, $>$T50–T80, and $>$T80), and the percentage of samples that were toxic within each concentration range was determined (Table 5). The logistic models and associated point estimates were considered reliable if the observed incidence of toxicity was consistent with the predicted incidence of toxicity. In this evaluation, chemical concentrations below the T20 value were predicted to be associated with a low incidence of toxicity ($<20\%$). Similarly, a high incidence of toxicity ($>80\%$) was expected when chemical concentrations exceeded the T80 values. Moderately low (20–50%) and moderately high (50–80%) incidences of toxicity were expected at concentrations between the T20 and T50 values and between the T50 and T80 values, respectively.

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 incidence of sediment toxicity in the project database (Table 5). The largest number of samples for each chemical had concentrations in the range below the T20 value, and the number of samples decreased within each of the subsequent ranges.

Table 2. Distribution of the sediment chemistry concentrations for the sediment samples with matching toxicity data

Chemical	No. of samples	10th percentile	50th percentile	90th percentile
Metals (mg/kg dry wt)				
Antimony	2,173	0.17	0.67	2.9
Arsenic	2,844	2.2	7.4	19
Cadmium	2,958	0.05	0.3	1.9
Chromium	2,827	9.1	50	130
Copper	3,091	2.6	26	160
Lead	3,010	5.5	23	130
Mercury	2,788	0.02	0.11	0.79
Nickel	2,916	2.4	19	44
Silver	2,552	0.03	0.21	1.9
Zinc	3,013	16	89	300
Polycyclic aromatic hydrocarbons (μg/kg dry wt)				
1-Methylnaphthalene	1,677	0.50	6.5	60
1-Methylphenanthrene	1,697	0.30	11	130
2,6-Dimethylnaphthalene	1,505	0.40	6.3	67
2-Methylnaphthalene	2,077	0.77	12	130
Acenaphthene	1,795	0.20	7.0	130
Acenaphthylene	1,747	0.18	7.0	120
Anthracene	2,268	0.47	20	410
Benz[*a*]anthracene	2,574	1.2	40	760
Benzo[*a*]pyrene	2,526	1.2	53	910
Benzo[*b*]fluoranthene	1,645	0.86	48	1,000
Benzo[*ghi*]perylene	2,210	1.1	46	550
Benzo[*k*]fluoranthene	1,691	0.50	26	620
Biphenyl	1,507	0.47	6.8	54
Chrysene	2,650	1.7	53	1,000
Dibenz[*a,h*]anthracene	1,886	0.20	14	170
Fluoranthene	2,734	3.0	81	1,400
Fluorene	2,011	0.35	11	160
Indeno[1,2,3-*cd*]pyrene	2,212	0.94	47	600
Naphthalene	2,201	1.8	16	220
Perylene	2,174	1.5	36	370
Phenanthrene	2,688	1.7	44	660
Pyrene	2,768	3.2	87	1,500
Polychlorinated biphenyls (PCBs; μg/kg dry wt)				
Total PCBs	1,989	2.0	39	640
Organochlorine pesticides (μg/kg dry wt)				
Dieldrin	770	0.04	0.79	5.1
p,p′-DDD[a]	1,672	0.084	1.8	20
p,p′-DDE[b]	1,899	0.080	2.2	59
p,p′-DDT[c]	1,176	0.058	1.2	18

[a] DDD = dichlorodiphenyldichloroethane.
[b] DDE = dichlorodiphenyldichloroethylene.
[c] DDT = dichlorodiphenyltrichloroethane.

The incidence of amphipod toxicity was underestimated for concentrations below the T20 value for all 37 chemicals, although 35 of 37 were within 10% of the top of the range. Between the T20 and T50 values, the incidence of toxicity for most (30 of 37) of the chemicals was within the predicted range of 20 to 50% toxicity. The incidence of toxicity between the T50 and T80 values was within the predicted range 50 to 80% toxicity for all 37 chemicals. The observed incidence of toxicity was below the T80 value for 11 chemicals, but most were within 10% of the predicted range. The exceptions included arsenic and *p,p′*-DDE. The models for these chemicals would be most likely to overestimate toxicity.

Among the various classes of contaminants, the logistic models for PAHs were the most reliable. For 16 of 22 PAHs, the actual incidence of toxicity to amphipods was correctly predicted within three of the four concentration ranges defined by the T values; however, a higher-than-predicted incidence of toxicity was observed above the T20 values for all PAHs (Table 5). Among the trace metals, the logistic models for chromium, copper, lead, mercury, and zinc were the most reliable, as indicated by the level of agreement between the predicted and observed incidence of toxicity 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.

Individual chemical models and spiked-sediment bioassay LC50 values

Dose–response data from laboratory-spiked sediment bioassays provide perspective on the concentrations of individual chemicals that can be considered to cause toxicity. 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 were classified as toxic with much higher test survival.

Table 3. Logistic regression model parameters, normalized chi-square values, and number of samples in the screened database for individual chemicals

Chemical	No. of samples	Intercept (B_0)	Slope (B_1)	Chi-square value/*N*
Metals (mg/kg dry wt)				
Antimony	1,718	−0.90	2.41	0.25
Arsenic	2,336	−4.14	3.17	0.17
Cadmium	2,413	−0.34	2.51	0.31
Chromium	2,399	−6.44	3.00	0.20
Copper	2,580	−5.79	2.93	0.38
Lead	2,481	−5.45	2.77	0.27
Mercury	2,296	0.80	2.55	0.32
Nickel	2,450	−4.61	2.77	0.18
Silver	2,103	−0.11	1.97	0.25
Zinc	2,516	−7.98	3.34	0.28
Polycyclic aromatic hydrocarbons (μg/kg dry wt)				
1-Methylnaphthalene	1,368	−4.14	2.10	0.24
1-Methylphenanthrene	1,401	−3.59	1.75	0.28
2,6-Dimethylnaphthalene	1,249	−4.05	1.90	0.20
2-Methylnaphthalene	1,704	−3.76	1.78	0.25
Acenaphthene	1,424	−3.62	1.75	0.33
Acenaphthylene	1,447	−2.96	1.38	0.23
Anthracene	1,823	−3.66	1.49	0.29
Benz[*a*]anthracene	2,099	−4.20	1.58	0.30
Benzo[*a*]pyrene	2,053	−4.30	1.58	0.30
Benzo[*b*]fluoranthene	1,348	−4.54	1.49	0.27
Benzo[*ghi*]perylene	1,818	−4.28	1.59	0.25
Benzo[*k*]fluoranthene	1,376	−4.28	1.57	0.29
Biphenyl	1,226	−4.11	2.21	0.26
Chrysene	2,126	−4.32	1.54	0.29
Dibenz[*a,h*]anthracene	1,546	−3.63	1.77	0.33
Fluoranthene	2,189	−4.46	1.48	0.26
Fluorene	1,668	−3.71	1.81	0.32
Indeno[1,2,3-*cd*]pyrene	1,837	−4.37	1.62	0.27
Naphthalene	1,816	−3.78	1.62	0.24
Perylene	1,823	−4.68	1.76	0.22
Phenanthrene	2,173	−4.46	1.68	0.30
Pyrene	2,240	−4.71	1.59	0.29
Polychlorinated biphenyls (PCBs; μg/kg dry wt)				
Total PCBs	1,617	−3.46	1.35	0.27
Organochlorine pesticides (μg/kg dry wt)				
Dieldrin	633	−1.17	2.56	0.35
p,p′-DDD[a]	1,360	−1.90	1.49	0.27
p,p′-DDE[b]	1,552	−1.84	0.91	0.16
p,p′-DDT[c]	931	−1.77	1.68	0.34

[a] DDD = dichlorodiphenyldichloroethane.
[b] DDE = dichlorodiphenyldichloroethylene.
[c] DDT = dichlorodiphenyltrichloroethane.

Thus, for comparisons with the individual chemical logistic regression models, LC20 values would be more consistent with the method used in this study to classify samples as toxic, but these values were not generally available. Consequently, reported LC50 values for 10-d spiked-sediment toxicity test conducted with marine amphipods were compared to the probability of toxicity estimated from the individual chemical models (Table 6). 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 of the probability of toxicity within the range of 0.8 to 0.9.

Multichemical models

The logistic regression models for individual chemicals were combined in two ways to estimate the probability of toxicity for the mixture of contaminants in a given sample:

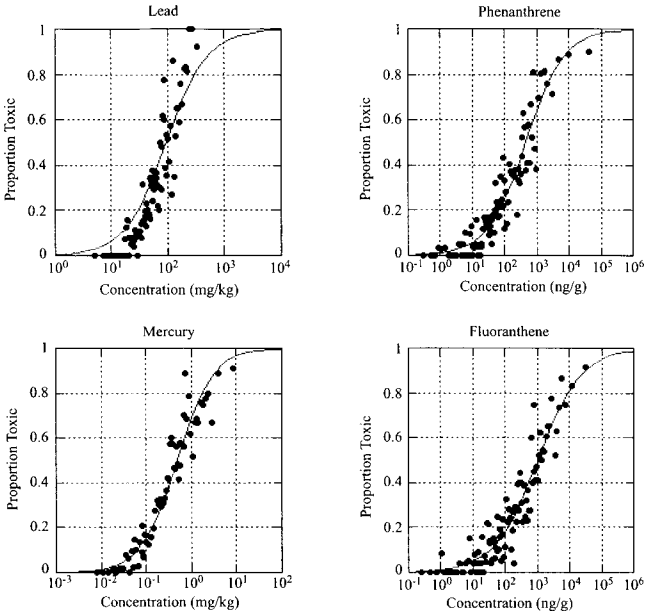

Fig. 1. Logistic regression models and proportion of samples toxic in concentration intervals in the screened database for lead (mg/kg), mercury (mg/kg), phenanthrene (ng/g), and fluoranthene (ng/g). The individual points represent the median of the sample concentrations within the interval and the fraction of the samples toxic within the interval.

using either the maximum (P_{Max} model) or the mean (P_{Avg} model) probabilities from the individual models (Fig. 2). The combined models are derived from the probability-interval plots, which summarize all the data in the database. In order to minimize the potential impact of samples with partial chemistry, only samples with at least 10 chemicals analyzed were included in the data set used to derive the multichemical models. Unlike the methods used for deriving the individual chemical models, all the samples with matching chemistry and toxicity were included in the evaluation (no additional data screening procedures were employed). The multichemical models are derived from the probability-interval plots, which summarize all of the data in the database. The parameter estimates shown in Table 3 were used to develop the multichemical models, except for PCBs. A correction in PCB units for 15 samples resulted in a minor change in the PCB model. However, because the effects of the correction on the multichemical models were extremely small (the maximum differences in predicted probability of toxicity were 0.0025 for the P_{Max} model and 0.029 for the P_{Avg} model), the multichemical models were not changed.

The results of this evaluation show that both combined models accurately predict toxicity to amphipods. Above a maximum probability of about 0.2, the maximum probability is somewhat higher than the corresponding proportion toxic (Fig. 2). The P_{Max} model calibrates the maximum probability to account for the difference between the maximum probability from the individual chemical models and the observed proportion toxic for samples within the same probability interval. For example, for a maximum probability of 1.0 from the individual chemical models (x-axis), the observed proportion toxic (and the predicted probability from the P_{Max} model) is 0.84. The data used to derive the P_{Avg} model show the opposite situation, where mean probability values correspond to a higher proportion toxic (underestimate toxicity). Thus, using the

Table 4. Logistic model point estimates of T20, T50, and T80 concentrations and 95% confidence intervals (CI) for individual chemicals. The notation T_p (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)

	T20			T50			T80		
Chemical	Lower CI	T value	Upper CI	Lower CI	T value	Upper CI	Lower CI	T value	Upper CI
Metals (mg/kg dry wt)									
Antimony	0.55	0.63	0.72	2.0	2.4	2.8	6.6	8.9	12
Arsenic	6.8	7.4	8.1	18	20	23	45	56	69
Cadmium	0.34	0.38	0.43	1.2	1.4	1.5	4.0	4.9	6.0
Chromium	44	49	53	126	141	158	329	410	510
Copper	29	32	35	86	94	103	239	280	328
Lead	27	30	33	84	94	104	244	297	360
Mercury	0.12	0.14	0.15	0.43	0.48	0.54	1.4	1.7	2.1
Nickel	13	15	16	42	47	52	118	147	185
Silver	0.19	0.23	0.26	0.98	1.1	1.3	4.4	5.8	7.6
Zinc	87	94	102	224	245	267	542	636	746
Polycyclic aromatic hydrocarbons (μg/kg dry wt)									
1-Methylnaphthalene	17	21	25	73	94	122	281	433	669
1-Methylphenanthrene	15	18	23	88	112	143	454	696	1,067
2,6-Dimethylnaphthalene	20	25	31	96	133	185	413	713	1,231
2-Methylnaphthalene	18	21	26	102	128	161	514	767	1,145
Acenaphthene	15	19	24	90	116	148	469	714	1,085
Acenaphthylene	11	14	18	102	140	194	799	1,418	2,517
Anthracene	27	34	42	228	290	369	1,630	2,486	3,792
Benz[*a*]anthracene	50	61	75	382	466	567	2,491	3,535	5,017
Benzo[*a*]pyrene	57	69	85	428	520	633	2,754	3,908	5,546
Benzo[*b*]fluoranthene	100	130	169	814	1,107	1,506	5,525	9,413	16,035
Benzo[*ghi*]perylene	54	67	82	395	497	625	2,444	3,710	5,631
Benzo[*k*]fluoranthene	55	70	90	405	537	713	2,541	4,121	6,685
Biphenyl	14	17	21	57	73	93	206	310	466
Chrysene	67	82	99	529	650	799	3,595	5,186	7,479
Dibenz[*a,h*]anthracene	15	19	23	92	113	139	475	685	988
Fluoranthene	98	119	146	832	1,034	1,284	6,066	8,952	13,212
Fluorene	16	19	24	92	114	140	465	665	951
Indeno[1,2,3-*cd*]pyrene	56	68	84	393	488	607	2,350	3,482	5,159
Naphthalene	25	30	37	170	217	278	1,022	1,569	2,409
Perylene	62	74	89	358	453	572	1,819	2,767	4,209
Phenanthrene	57	68	81	377	455	550	2,191	3,056	4,263
Pyrene	103	125	150	768	932	1,132	4,942	6,982	9,865
Polychlorinated biphenyls (PCBs; μg/kg dry wt)									
Total PCBs	27	35	44	282	368	481	2,412	3,926	6,393
Organochlorine pesticides (μg/kg dry wt)									
Dieldrin	0.65	0.83	1.0	2.3	2.9	3.6	6.9	10	15
p,p′-DDD[a]	1.7	2.2	2.8	14	19	25	95	159	267
p,p′-DDE[b]	2.2	3.1	4.4	61	103	176	1,278	3,414	9,119
p,p′-DDT[c]	1.3	1.7	2.2	8.3	11	15	45	76	129

[a] DDD = dichlorodiphenyldichloroethane.
[b] DDE = dichlorodiphenyldichloroethylene.
[c] DDT = dichlorodiphenyltrichloroethane.

P_{Avg} model, mean probabilities of 0.5 and 0.75 correspond to proportion toxic of 0.7 and 0.9, respectively.

A major advantage of the P_{Max} model is that it is less sensitive to the number of chemicals measured in each sample. For example, the P_{Avg} model incorporates the output from models for 22 individual PAHs. Since individual PAHs are likely to co-occur in environmental samples, the P_{Avg} may be influenced more by the concentrations of PAHs than by the concentrations of other chemicals. Although both the P_{Max} and P_{Avg} models provide a good fit to the data, we will discuss only the P_{Max} model in the remainder of this paper.

The average predicted probability of toxicity (from the P_{Max} model) within probability quartiles closely matches the incidence of toxicity within the same probability quartiles (Fig. 3), demonstrating the overall reliability of the P_{Max} model within the database that was used to derive the model. In addition, the magnitude of the effect (decreased survival) in the amphipod test increases as the probability of toxicity increases (Fig. 4). Toxic samples with a probability of toxicity less than or equal to 0.25 have an average control-adjusted survival of greater than 75%, while samples with a probability of toxicity greater than 0.75 have an average control-adjusted survival of less than 50%. This demonstrates that samples that are estimated to have the highest probability of toxicity are also likely to be extremely toxic.

The number of chemicals in a sample that have a high probability of toxicity (e.g., $p > 0.75$) makes a difference in how well the model predictions match the observed percentage of the samples that are toxic (Fig. 5). 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 probability of toxicity. The difference between the predicted probability of toxicity and incidence of toxicity can be considered to be a measure of the false-positive rate. As the number of chem-

Table 5. The percentage of toxic samples within ranges defined by logistic model T20, T50, and T80 values and the number of samples in the database used to derive the logistic model for each chemical

Chemical	<T20	T20–T50	T50–T80	>T80	No. of samples
Metals (mg/kg dry wt)					
Antimony	30.3	48.5	67.9	82.5	2,173
Arsenic	30.0	43.8	56.3	69.7	2,844
Cadmium	27.6	50.9	62.7	78.7	2,958
Chromium	24.5	42.7	55.7	80.0	2,827
Copper	22.1	50.6	64.9	85.0	3,091
Lead	28.5	45.0	60.6	90.0	3,010
Mercury	25.5	49.0	66.1	79.4	2,788
Nickel	25.3	44.7	60.4	NA[a]	2,916
Silver	25.6	52.5	60.7	73.7	2,552
Zinc	23.6	47.3	67.8	71.3	3,013
Polycyclic aromatic hydrocarbons (μg/kg dry wt)					
1-Methylnaphthalene	23.7	48.3	60.3	75.0	1,677
1-Methylphenanthrene	24.7	45.8	65.5	80.0	1,697
2,6-Dimethylnaphthalene	23.5	42.2	57.4	NA	1,505
2-Methylnaphthalene	25.4	47.5	61.1	88.0	2,077
Acenaphthene	25.2	50.3	67.7	91.4	1,795
Acenaphthylene	24.0	44.7	67.6	NA	1,747
Anthracene	26.5	48.8	66.9	77.1	2,268
Benz[*a*]anthracene	28.5	45.3	65.0	82.9	2,574
Benzo[*a*]pyrene	27.7	48.5	64.2	83.8	2,526
Benzo[*b*]fluoranthene	24.0	46.3	67.4	NA	1,645
Benzo[*ghi*]perylene	25.3	46.6	63.6	86.7	2,210
Benzo[*k*]fluoranthene	25.5	44.2	68.3	93.3	1,691
Benzofluoranthenes, total	25.2	46.3	56.7	83.3	1,507
Biphenyl	28.7	47.8	64.8	86.1	2,650
Chrysene	23.9	49.0	65.3	85.7	1,886
Dibenz[*a,h*]anthracene	28.4	47.1	64.9	87.9	2,734
Fluoranthene	22.1	48.4	68.2	87.2	2,011
Fluorene	25.6	44.5	64.9	90.5	2,212
Indeno[1,2,3-*cd*]pyrene	26.4	43.8	64.1	89.7	2,201
Naphthalene	26.9	39.4	59.8	NA	2,174
Perylene	28.2	47.3	64.1	85.7	2,688
Phenanthrene	28.3	46.2	65.0	87.2	2,768
Pyrene	23.7	48.3	60.3	75.0	1,677
Polychlorinated biphenyls (PCBs; μg/kg dry wt)					
Total PCBs	26.8	46.3	72.7	81.5	1,989
Organochlorine pesticides (μg/kg dry wt)					
Dieldrin	20.2	53.8	66.7	78.8	770
p,p′-DDD[b]	25.9	49.4	64.7	80.5	1,672
p,p′-DDE[c]	22.5	53.4	54.9	57.6	1,899
p,p′-DDT[d]	25.6	56.5	66.7	76.7	1,176

[a] NA = fewer than 10 samples.
[b] DDD = dichlorodiphenyldichloroethane.
[c] DDE = dichlorodiphenyldichloroethylene.
[d] DDT = dichlorodiphenyltrichloroethane.

Table 6. Estimated probability of toxicity from chemical-specific logistic regression models for median lethal concentration (LC50) values (dry wt) reported from 10-d spiked sediment amphipod toxicity tests

Chemical	LC50	Probability of toxicity	Source
Cadmium (mg/kg)	9.81	0.90	Mearns et al. [24]
	8.8–10.0	0.88–0.90	Kemp et al. [25]
	8.2–11.5	0.88–0.91	Robinson et al. [26]
	6.9	0.85	Swartz et al. [27]
Mercury (mg/kg)	13.1	0.97	Swartz et al. [17]
Zinc (mg/kg)	276	0.54	Swartz et al. [17]
Fluoranthene (mg/kg)	4.2	0.71	Swartz et al. [17]
	3.3–10.5	0.68–0.82	Swartz et al. [28]
Phenanthrene (mg/kg)	3.68	0.82	Swartz et al. [29]
Total PCBs[a] (mg/kg)	8.8	0.87	Swartz et al. [17]
p,p′-DDT[b] (ng/g)	11.2–125	0.50–0.85	Word et al. [30]

[a] PCBs = polychlorinated biphenyls.
[b] DDT = dichlorodiphenyltrichloroethane.

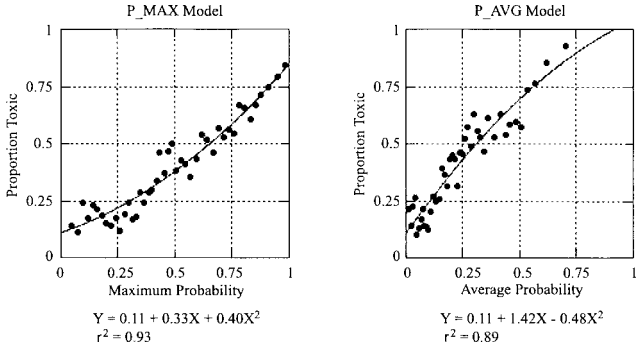

Fig. 2. Maximum (P_{Max}) and average (P_{Avg}) probability models (curves) and probability interval plots (points) where each point represents the median sample probability of a minimum of 50 individual samples within the interval and the fraction of the samples toxic within the interval.

icals in a sample with a high probability of toxicity increases, the false-positive rate decreases.

Application of the models to independent data (data not used in model derivation) is an important step in evaluation of the models. The P_{Max} model was applied to a small independent data set ($n = 65$) consisting of two studies from the Calcasieu Estuary in Louisiana, USA, that had matching sediment chemistry and toxicity data for *A. abdita* [16], unpublished data set provided electronically by P. Crocker, U.S. EPA, Region 6, Dallas, TX, USA]. These data were not included in the database used to derive the models. Although the data set has a limited number of samples, data from a wide range of contaminant concentrations are represented. The average predicted probability of toxicity within probability quartiles was within 25% of the measured incidence of toxicity for each quartile, indicating that the P_{Max} model generally was able to successfully classify the samples as toxic or not toxic (Fig. 6). With the exception of the third quartile ($0.5 < p < 0.75$), the P_{Max} model underestimated the toxicity observed in Calcasieu Estuary samples.

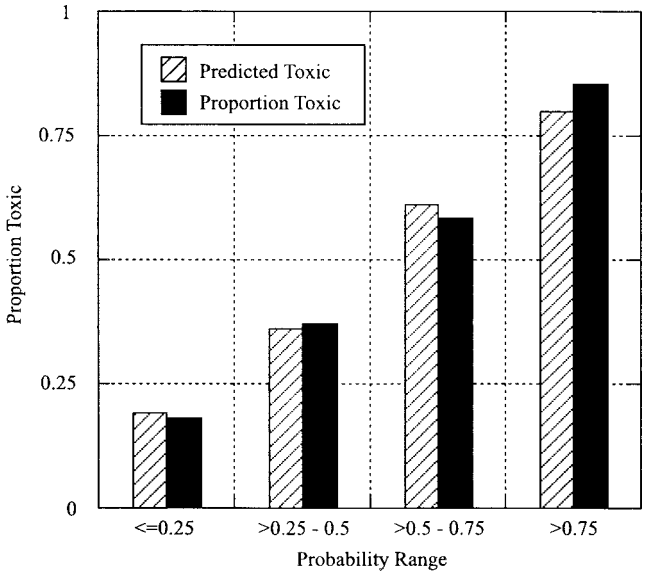

Fig. 3. Average predicted and proportion toxic within probability quartiles for P_{Max} model.

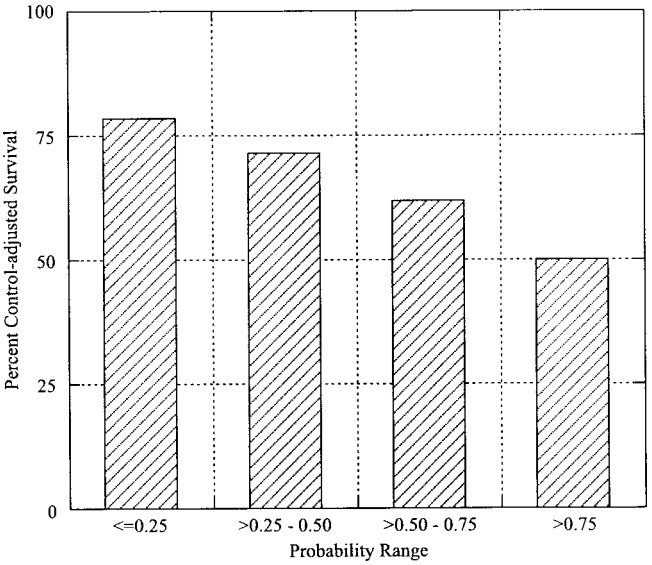

Fig. 4. Percentage control-adjusted survival for toxic samples within probability quartiles for P_{Max} model.

DISCUSSION

Evaluation of the logistic regression modeling approach

The results of this investigation indicate that logistic regression models provide an effective basis for describing relationships between the concentrations of sediment-associated contaminants and toxicity to two species of marine amphipods. The chemical-specific models that were derived in this investigation provide a basis for estimating the proportion of samples expected to be toxic for 37 individual contaminants over a wide range of contaminant concentrations. As such, these models help users select sediment effect concentrations that most directly meet the needs of their specific application. For

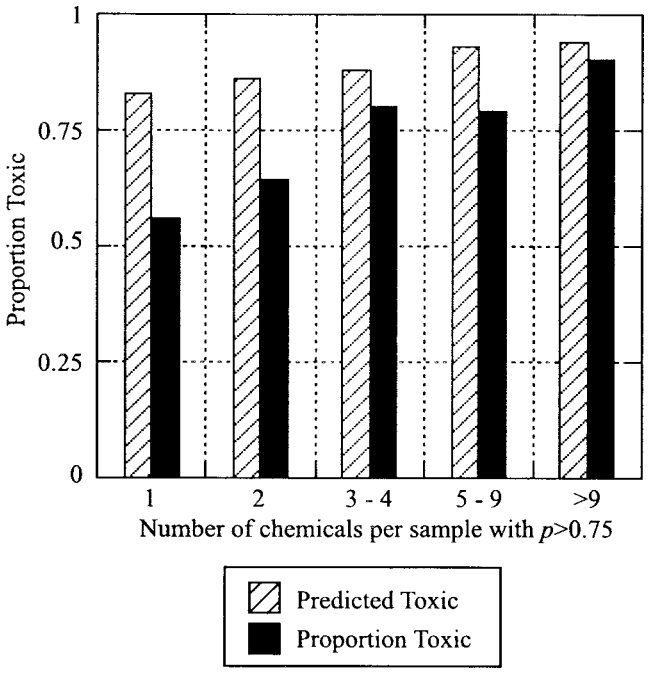

Fig. 5. Average predicted and proportion toxic by number of chemicals per sample with probability of toxicity from individual chemical models >0.75.

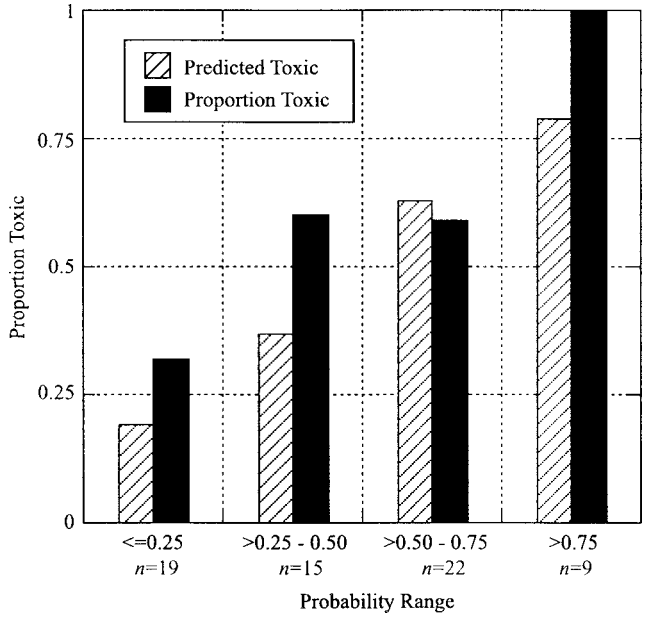

Fig. 6. Average predicted and proportion toxic within probability quartiles for an independent data set from the Calcasieu Estuary (LA, USA) (n = 65).

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 (higher percentage mortality). The T values can be used in much the same way as other sediment guidelines, with the difference that the T value is associated with a specific probability of observing toxicity and an estimate of variance based on the fit of the model.

The individual models derived in this study have lower T50 values for all the seven chemicals that were modeled in our earlier study [5]. The most likely explanations for this difference include standardization of the approach used to classify samples as toxic and the large increase in the size of the database. In the previous study, toxicity was classified by the original investigators. For example, in the earlier study, samples from the Puget Sound database were classified as toxic if significantly different from a field reference and less than 75% survival; in the present study, these samples were classified as toxic if significantly different from the negative control and less than 90% survival. This change resulted in a greater number of samples classified as toxic in data from Puget Sound than in the earlier study.

The P_{Max} model is based on the individual chemical with the highest probability of toxicity. For approximately 70% of the samples, one of the 10 metals for which individual regression models had been developed had the maximum probability used in the P_{Max} model. This should not be construed to imply that metals are causing toxicity in these samples. It indicates only that metals appear to be a good predictor of toxicity in field-collected samples where mixtures of contaminants are likely to be present. It is not possible to determine from the models alone whether the metals or the other chemicals considered in this evaluation make a substantial contribution to the observed toxicity in any individual sample.

Comparison of the model results to spiked-sediment bioassays reported in the literature indicates that LC50 values are equivalent to T80 to T90 values for several chemicals. This is consistent with the observation that control-adjusted survival is approximately 50% in samples with a high probability of toxicity. It may be more appropriate to use LC15 or LC20 values for comparisons to the chemical-specific models since the models are based on whether samples are classified as toxic, which does not require 50% mortality. Unfortunately, the LC15 or LC20 values are rarely reported in the literature.

Estimating the probability of toxicity for sediment quality guidelines

The individual chemical logistic models can be used to estimate the probability of observing toxicity to amphipods at the chemical concentrations that are defined by the SQGs. Examples are shown in Table 7 for three commonly used sets of SQGs that represent a range of threshold values: threshold effect levels (TEL), probable effect levels (PEL [7]; effect range low [ERL] and effect range median [ERM] [13]), and apparent effect thresholds (AET) for marine amphipods [15]. Both ERLs and TELs represent chemical concentrations below which toxicity would occur infrequently (<25%) [1,7], while 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. For the TELs, the probability of observing sediment toxicity at these concentrations ranged from 10 to 41% (depending on the chemical under consideration), with the probability of toxicity below 25% for 24 of 27 chemicals considered (Table 7). The probability of observing sediment toxicity was generally higher at the ERL concentrations (ranging from 11–47%), with a median value of 33%. The predicted incidence of toxicity was higher for the PEL and ERM values, with median values of 55 and 72%, respectively. The highest probability of observing toxicity to amphipods was noted for the amphipod AETs, with the estimated proportion of the toxic samples ranging from 45 to 99% and the median value of 90%.

Although derivation methods for the different SQGs are well described and are consistent for all the chemicals within a given type of SQG, no straightforward method exists that enables the user to either evaluate the degree to which individual SQGs meet their objectives or evaluate the reliability of individual SQGs. The logistic model approach provides a way to put the individual SQG values into perspective by estimating the probability of toxicity to amphipods. In addition, the goodness of fit for each model provides an objective measure of the quality of the models for individual chemicals.

Multichemical models

One of the major challenges in assessing the ecological risk associated with exposure to contaminated sediments is the presence of chemical mixtures. Swartz et al. [17] demonstrated that mixtures of two to four contaminants produced greater toxicity to a marine amphipod than the individual chemicals

Table 7. Estimated proportion of samples toxic to amphipods at the chemical concentrations defined by sediment quality guidelines (SQG)

Chemical	ERL[a]	ERM[b]	TEL[c]	PEL[d]	AET[e]
Metals					
Antimony	NA[f]	NA	NA	NA	0.99
Arsenic	0.22	0.85	0.20	0.73	0.99
Cadmium	0.46	0.89	0.32	0.77	0.93
Chromium (total)	0.33	0.78	0.22	0.54	0.94
Copper	0.21	0.79	0.11	0.54	0.97
Lead	0.30	0.73	0.20	0.55	0.96
Mercury	0.22	0.60	0.19	0.60	0.85
Nickel	0.28	0.53	0.22	0.48	0.92
Silver	0.47	0.73	0.41	0.59	0.81
Zinc	0.33	0.68	0.27	0.54	0.98
Polycyclic aromatic hydrocarbons					
2-Methylnaphthalene	0.39	0.78	0.19	0.59	0.89
Acenaphthene	0.18	0.75	0.10	0.45	0.90
Acenaphthylene	0.33	0.71	0.13	0.49	0.79
Anthracene	0.31	0.70	0.24	0.47	0.92
Benz[*a*]anthracene	0.40	0.70	0.22	0.57	0.84
Benzo[*a*]pyrene	0.47	0.68	0.23	0.57	0.79
Benzo[*ghi*]perylene	NA	NA	NA	NA	0.78
Chrysene	0.41	0.73	0.23	0.54	0.91
Dibenz[*a*,*h*]anthracene	0.39	0.66	0.10	0.53	0.90
Fluoranthene	0.41	0.74	0.19	0.56	0.90
Fluorene	0.20	0.77	0.21	0.55	0.94
Indeno[1,2,3-*cd*]pyrene	NA	NA	NA	NA	0.83
Naphthalene	0.45	0.83	0.22	0.60	0.84
Phenanthrene	0.39	0.70	0.23	0.53	0.94
Pyrene	0.44	0.67	0.22	0.57	0.88
Polychlorinated biphenyls (PCBs)					
Total PCBs	0.16	0.40	0.16	0.40	0.78
Organochlorine pesticides					
Dieldrin	NA	NA	0.18	0.61	0.55
p,*p*′-DDD[g]	NA	NA	0.15	0.36	0.69
p,*p*′-DDE[h]	0.18	0.37	0.17	0.62	0.45
p,*p*′-DDT[i]	NA	NA	0.16	0.35	0.91

[a] ERL = effect range low [1].
[b] ERM = effect range median [1].
[c] TEL = threshold effect level [7].
[d] PEL = probable effect level [7].
[e] AET = apparent effect threshold for amphipod survival [15].
[f] NA = no SQG value available.
[g] DDD = dichlorodiphenyldichloroethane.
[h] DDE = dichlorodiphenyldichloroethylene.
[i] DDT = dichlorodiphenyltrichloroethane.

alone. 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 [18]. Concentration addition, which is also referred to as simple similar action, assumes that contaminants act independently but by 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 [19–21] but are unlikely to be applicable to complex mixtures of contaminants commonly found in the environment that may represent different modes of action. Response addition, or independent action, is expected to apply to cases where contaminants have a different mode of action and toxicity would be predicted only when one or more contaminants exceeds their toxicity threshold.

The P_{Max} model can be considered to be similar to a response addition model, where toxicity is predicted on the basis of the individual chemical model that has the highest probability of toxicity. However, because the individual models themselves were derived from field-collected sediments that include mixtures of contaminants rather than individual dose–response relationships, to some extent the individual models incorporate the overall toxicity of the mixtures.

The multichemical models provide an estimate of toxicity for individual samples based on the output from the individual chemical models. The P_{Max} model, which is based on the highest predicted probability from any of the individual chemical models, is less sensitive to the number of chemicals analyzed than a model based on the mean value (P_{Avg} model). However, because the predicted probability of toxicity is based on a single chemical, a greater potential may exist for false positives because of the application of less reliable individual models for some chemicals or unusual conditions. The individual models for nickel, *p*,*p*′-DDE, and a few other chemicals were shown (Table 5) to have a lower incidence of toxicity at concentrations exceeding T80 values. Thus, the probability of toxicity could be overestimated for any sample where the chemical with the maximum probability value had a high rate of false positives. In addition, a greater tendency exists for false positives when high probability predictions are based on only one or two chemicals having a high probability of toxicity. These results are similar to the analysis presented by Long et al. [22] that demonstrated an increasing incidence of toxicity as the number of ERMs or PELs exceeded in individual samples increased. This supports the concept that empirical approaches, such as the one described in this paper, are not defining dose–response relationships for individual chemicals but serve as indicators of toxicity based on chemical mixtures.

Application of models to evaluations of site-specific or regional data

Hazardous waste-site evaluations often involve the collection of substantial quantities of sediment chemistry data. For example, information on the magnitude and areal extent of sediment contamination is frequently collected to support screening-level ecological risk assessments. In the past, sediment assessors have used numerical SQGs (e.g., ERLs, ERMs, TELs, PELs, AETs, and/or others) to evaluate such data. While such SQGs are useful for identifying thresholds below which sediment toxicity is unlikely to be observed and above which sediment toxicity is likely to occur, it is difficult to determine the extent to which risk increases with the magnitude of exceedance of the SQGs. Calculation of hazard quotients (HQ, the ratio of the measured concentration of a contaminant in sediments to the corresponding toxicity threshold) for each chemical using the SQGs can provide additional information for assessing risk to sediment-dwelling organisms. Several investigators have applied mean SQG quotients to evaluate mixtures of contaminants in field-collected sediment samples [8,22,23]. However, such evaluations are based on an assumption that concentration–response relationships for each chemical are similar. The logistic regression models approach avoids this assumption and provides a way to apply separate concentration response relationships for each chemical.

The models described in this paper 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 contamination gradients. Because the models require a large amount of data for their derivation, data from an individual site are rarely sufficient to derive site-specific models. Rather than deriving site- or regional-specific models, we recommend applying the models to data from the

site or region to determine how well the models fit the local data. The evaluation of the independent data set from the Calcasieu Estuary provides an example of how this could be accomplished. By comparing the percentage of toxic samples with the mean predicted probability of toxicity within discrete probability ranges (e.g., probability quartiles as shown in Fig. 6), the performance of the models with data from the site can be evaluated. If the models predict a higher percentage of toxic samples than 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 (no models available) or issues related to the sediment matrix (e.g., grain size effects).

CONCLUSIONS

A large database of matching sediment chemistry and toxicity data for marine amphipod survival, which includes many different contaminant gradients from a wide variety of habitats in coastal North America, was used to derive logistic regression models for 37 individual chemicals. The logistic regression models do not represent dose–response relationships for individual chemicals but can be considered to be indicators of toxicity based on field-collected sediment chemical mixtures. Combining the individual models into a single model, using either the maximum or the average probability predicted from the chemicals analyzed in a sample, provides a single value for estimating the probability that a sample will be toxic. These models enable users to select sediment quality guidelines that match the level of protectiveness (as measured by the probability of toxicity) appropriate for the objectives of their assessment and to estimate the uncertainty associated with the chosen level of protectiveness.

The LRM approach provides a useful framework for conducting screening-level assessments that require classifying or prioritizing samples on the basis of sediment chemistry but should not be considered as a substitute for direct effects assessment (e.g., toxicity tests). Because the models do not consider potential differences in bioavailability or exposure, the probability of toxicity may be over- or underestimated for some locations. Application of the models to additional independent data sets will provide the best measure of the ability of the models to predict toxicity in different environments and contaminant gradients.

Acknowledgement—Primary funding for this project was provided by the U.S. EPA, Office of Research and Development, National Center for Environmental Assessment. Supplementary funding for data compilation was provided by the U.S. EPA Office of Water and the National Research Council of Canada. This project would not be possible without contributions of data from many individuals and programs. Special thanks are extended to E. Long, T. Gries, and R. Fairey. The EMAP data were provided by D. Keith, M. Hughes, J. Macauley, L. Harwell, J. Hyland, T. Snoots, and D. Adams. We also thank E. Smith, J. Keating, S. Ireland, and J. Harcum. M.L. Haines, J. Brown, and C. Hong, who provided technical assistance during data compilation.

REFERENCES

1. 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, USA.
2. U.S. Environmental Protection Agency. 1997. The incidence and severity of sediment contamination in surface waters of the United States. Vol 1. National sediment quality survey. 823-R-97-006. Washington, DC.
3. Ingersoll CG, Dillon T, Biddinger RG. 1997. *Ecological Risk Assessments of Contaminated Sediment.* SETAC, Pensacola, FL, USA, p 389.
4. Chapman PM. 1989. Current approaches to developing sediment quality criteria. *Environ Toxicol Chem* 8:589–599.
5. Field LJ, MacDonald DD, Norton SB, Severn CG, Ingersoll CG. 1999. Evaluating sediment chemistry and toxicity data using logistic regression modeling. *Environ Toxicol Chem* 18:1311–1322.
6. Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manage* 19: 81–97.
7. MacDonald DD, Carr RS, Calder FD, Long ER, Ingersoll CG. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5:253–278.
8. 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.
9. American Society for Testing Materials. 1997. Standard guide for conducting solid phase, 10-day, static sediment toxicity tests with marine and estuarine infaunal amphipods. In *Annual Book of Standards,* Vol 11.05. E 1367-92. Philadephia, PA, pp 732–757.
10. Lauenstein GG, Cantillo AY, Dolvin SS. 1993. NOAA National Status and Trends Program development and methods. In Lauenstein GG, Cantillo AY, eds, *Sampling and Analytical Methods of the National Status and Trends Programs National Benthic Surveillance and Mussel Watch Projects 1984–1992,* Vol 1—Overview and Summary Methods. NOAA Technical Memorandum NOS ORCA 71. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA, p 117.
11. 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.
12. Ingersoll CG, Haverland PS, Brunson EL, Canfield TJ, Dwyer FJ, Henke CE, Kemble NE. 1996. Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironomus riparius*. *J Gt Lakes Res* 22:602–623.
13. Long ER, MacDonald DD. 1992. National Status and Trends Program approach. In *Sediment Classification Methods Compendium.* EPA 823-R-92-006. U.S. Environmental Protection Agency, Washington DC, pp 211–228.
14. SAS Institute. 1990. *SAS/STAT® User's Guide,* Ver 6, 4th ed, Vols 1 and 2. Cary, NC, USA.
15. Gries TH, Waldow KH. 1996. *Progress Re-Evaluating Puget Sound Apparent Effects Thresholds (AETs),* Vol I—1994 Amphipod and Echinoderm Larval AETs. Washington Departments of Ecology and Natural Resources, Seattle, WA, USA.
16. Redmond MS, PA Crocker, KM McKenna, EA Petrocelli, KJ Scott, CR Demas. 1996. Sediment toxicity testing with the amphipod *Ampleisca abdita* in Calcasieu Estuary. *Arch Environ Contam Toxicol* 30:53–61.
17. Swartz RC, Kemp PF, Schults DW, Lamberson JO. 1988. Effects of mixtures of sediment contaminants on the marine infaunal amphipod, *Rhepoxyinius abronius*. *Environ Toxicol Chem* 7: 1013–1020.
18. Broderius SJ. 1991. Modeling the joint toxicity of xenobiotics to aquatic organisms: Basic concepts and approaches. In Mayes MA, Barron MG, eds, *Aquatic Toxicology and Risk Assessment,* Vol 14. STP 1124. American Society for Testing and Materials, Philadelphia, PA, pp 107–127.
19. Swartz RC, Schults DW, Ozretich RJ, Lamberson JO, Cole FA, DeWitt TH, Redmond MS, Ferraro SP. 1995. ΣPAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments. *Environ Toxicol Chem* 14:1977–1987.
20. 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.
21. Lee JH, Landrum PF, Field LJ, Koh CH. 2001. Application of a ΣPAH 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.

22. Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environ Toxicol Chem* 17:714–727.
23. Fairey R, Long ER, Roberts CA, Anderson BS, Phillips BM, Hunt JW, Puckett HR, Wilson CJ. 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.
24. Mearns AJ, Swartz RC, Cummins JM, Dinnel PA, Plesha P, Chapman PM. 1986. Inter-laboratory comparison of a sediment toxicity test using the marine amphipod, *Rhepoxynius abronius*. *Mar Environ Res* 19:13–37.
25. 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.
26. Robinson AM, Lamberson JO, Cole FA, Swartz RC. 1988. Effects of culture conditions on the sensitivity of a phoxocephalis amphipod, *Rhepoxynius abronius*, to cadmium in sediment. *Environ Toxicol Chem* 7:953–959.
27. Swartz RC, Ditsworth GR, Schults DW, Lamberson JO. 1985. Sediment toxicity to a marine infaunal amphipod: Cadmium and its interactions with sewage sludge. *Mar Environ Res* 18:133–153.
28. Swartz RC, Schults DW, DeWitt TH, Ditsworth GR, Lamberson JO. 1990. Toxicity of fluoranthene in sediment to marine amphipods: A test of the equilibrium partitioning approach to sediment quality criteria. *Environ Toxicol Chem* 9:1071–1080.
29. Swartz RC, Kemp PF, Schults DW, Ditsworth GR, Ozretich RJ. 1989. Acute toxicity of sediment from Eagle Harbor, Washington to the marine infaunal amphipod, *Rhepoxyinius abronius*. *Environ Toxicol Chem* 8:215–222.
30. Word JQ, Ward JA, Franklin LM, Cullinan VI, Kiesser SL. 1987. Evaluation of the equilibrium partitioning theory for estimating the toxicity of the nonpolar organic compound DDT to the sediment dwelling amphipod *Rhepoxynius abronius*. U.S. Environmental Protection Agency, Washington, DC.