Review of

Sediment Quality Objectives for Enclosed Bays and Estuaries of California

By

Dominic M. Di Toro
Edward C. Davis Professor of Civil and Environmental Engineering
Department of Civil and Environmental Engineering
356 DuPont Hall
University of Delaware
Newark, DE 19716

January 3, 2008

This is a review of the report

Draft Staff Report
Water Quality Control Plan for Enclosed Bays and Estuaries
Part 1. Sediment Quality
State Water Resources Control Board
California Environmental Protection Agency
September 27, 2007

The review is in two parts. The first responds directly to the questions posed in Attachment 2. The second is an Appendix that presents a more complete discussion of the issues and a preliminary analysis of the sediment toxicity data to illustrate the application of mechanistic criteria.

1. Are benthic invertebrates important ecologically relevant receptors to protect from direct exposure to toxic pollutants in sediments within the bays and estuaries of California?

Yes, and the rationale for protecting benthic invertebrates are presented very well in the report.

2. Are multiple lines of evidence appropriate to assess the potential risk to benthic invertebrates from toxic pollutants in sediments within the bays and estuaries of California?

Clearly multiple lines of evidence are required to assess the potential risk to benthic invertebrates from toxic pollutants in sediments. This is the case both within the bays and estuaries of California and for other sites, e.g. streams, rivers and lakes. The report presents the rationale and appropriate citations to the literature supporting this position.

3. Individual lines of Evidence
    a. Are proposed sediment toxicity indicators appropriate for assessing both the potential risk of exposure from toxic pollutants and the biological effects in benthic invertebrates within the bays and estuaries of California?

The analysis of the available toxicity tests and the methodology presented in the report for converting toxicity tests for use in judging the level of toxicity appears to be sound. I find the rejection of the Ampelisca abdita test a little strange since the test is employed widely, but a rationale is presented.

b. Are proposed sediment chemistry indicators appropriate for assessing both the potential risk of exposure from toxic pollutants to benthic invertebrates within the bays and estuaries of California?

The sediment chemistry indicator developed in the report is incomplete. As the report states, there are two general methods available for assessing the potential for toxicity in sediments: empirical and mechanistic. The report embraces the empirical method and dismisses the mechanistic method in a few sentences. In Section 5.5.3.2 "What chemistry indicators should be used?" the reasons are given

> "Mechanistic SQGs based on equilibrium partitioning were not included for several reasons. Data for some of the key parameters needed to apply the mechanistic guidelines (e.g. sediment acid volatile sulfides and simultaneously extracted metals) were not available. In addition chemistry data were not available for all the potential toxicants in the samples, which limited the predictive ability of the guidelines for organics. Previous analyses using Southern California data showed that these limitations significantly affected mechanistic SQG performance; application of a partial suite of mechanistic SQGs for organics resulted in poor predictive ability (Vidal and Bay 2005)."

However *both* empirical and mechanistic methods are incomplete. Neither method can predict with more than a modest degree of certainty the outcome of a toxicity test on a sediment from the field that is contaminated with many, and possibly unknown and unmeasured contaminants. Fig. 1 presents the results of the analysis from "Comparative Sediment Quality Guideline Performance For Predicting Sediment Toxicity In Southern California, USA" by Vidal and Bay 2005.

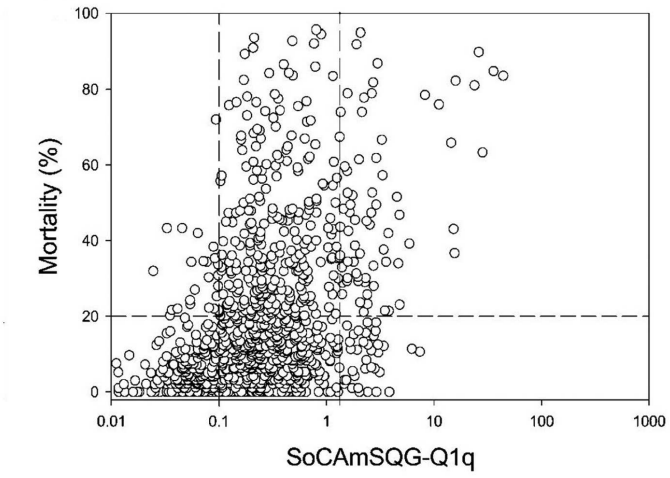

Figure 1

The empirical criteria predicts the lack of toxicity reasonably well (SoCAmSQG-Q1 < 0,1) but fails to discriminate between toxic and non toxic sediments at the same value of SoCAmSQG-

Q1q for the bulk of the data in the range of 0.1 to 1.0. The mechanistic criteria as evaluated by Vidal and Bay appeared to have no predictive ability at all in this data set.

I have prepared an appendix attached to this review that discusses these issues in more detail. It illustrates the applicability of mechanistic criteria to the available data to demonstrate their utility, even if the necessary data for a complete and rigorous application are not available. As demonstrated in the appendix, the role of mechanistic criteria is not to predict toxicity. For the reasons given above and as presented in more detail in the appendix, the role of mechanistic criteria is to determine if the observed toxicity can be explained by known modes of bioavailability and toxic mechanisms.

The results can be used to judge whether the chemical cause of the toxicity for particular sediment is likely to be metals, PAHs and other narcotics, or the pesticides that have been measured. The alternative is that none of these classes of chemicals appear to be the cause of the observed toxicity and the situation is quite uncertain. If the later is the case, then the result of the best professional judgment assessment of the situation would change to be very uncertain, regardless of the level of chemical contamination. Also, in my opinion, more information about the toxic sediment should be collected so that a more secure decision can be made.

Therefore, both mechanistic and empirical criteria should be used to judge the extent of toxicity that is likely due to chemicals, and if the chemical data are consistent with known measures of bioavailability and modes of chemical toxicity. Ignoring mechanistic criteria is not employing the best available science to support regulatory judgments. Mechanistic criteria have been developed and validated from very large datasets. A comprehensive review with citations to the primary literature is available (Di Toro et al., 2005). They are based on quantitative mechanistic models that have been published over the years in the peer reviewed literature, are highly cited, and have been tested by numerous independent investigators. They provide a framework for understanding chemical causes of sediment toxicity, and can be used to discriminate between two important cases: (1) we understand the chemical cause of the observed toxicity; (2) we do not, at our present level of understanding. Empirical criteria cannot provide this important additional information.

c. Are proposed benthic community indicators appropriate for assessing the biological effects through benthic community condition within the bays and estuaries of California?

The report presents the rationale and methodology for selecting the benthic community indicators and they appear to be sound.

4. Is the integration framework appropriate for determining if a station meets the narrative objective?

The integration framework – the quantification of best professional judgment (BPJ) – is to be commended. It produces a specific outcome for the data to be evaluated. The test of the method by experts on a small dataset is a nice demonstration of its utility in quantifying BPJ and making it applicable to specific sediment.

I would suggest one further test. Evaluate the entire dataset for which the necessary triad information is available. What proportion of the tested sediments is in which level of concern? There are a number of arbitrary cutoff levels in the framework, and it is important to know if these choices trigger many highly toxic sediments. A criterion that is too restrictive and triggers too many false positives is not a useful regulatory tool.

5. Is the implementation of the narrative SGO appropriate given the limitations of the individual tools and potential uncertainty associated with sediment quality assessment?

I would strongly recommend the inclusion of the results of an analysis of the data using mechanistic criteria for the purposes of determining the probable cause(s) of toxicity, or whither the cause is unknown. An example application is included in the appendix to this review.

1. Are there any additional scientific issues that are part of the scientific basis of the proposed rule not described above?

I would recommend that a report be prepared that documents the calculations that lead to the LRM in the report so that the analysis can be reproduced, including the analysis leading to Table 2 from Direct Effects Calculation

In order to apply mechanistic criteria without the approximations used in the appendix, certain data are required. Although the historical data may not include the appropriate measurements, all future data collection should include at least: SEM and AVS for a proper assessment of metal toxicity; a complete suite of PAHs including alkylated PAHs and sediment organic carbon to evaluate PAH toxicity. Not requiring such data is not consistent with using the best science.

Section 5.7.4. The staff recommendation is to apply the narrative SQGs to NPDES permits as receiving water limits. Unless mechanistic criteria can successfully identify the chemical causes of the toxicity it is not possible to establish receiving water limits. As discussed above it is the universally agreed that empirical criteria cannot be used to identify the chemical causes of toxicity.

2. Taken as a whole is the scientific portion of the proposed rule based upon sound scientific knowledge methods and practices?

With the exception of the exclusion of mechanistic criteria for judging the possible chemical causes of toxicity – and this is a glaring problem – the implementation is based on sound scientific knowledge methods and practices.

References

Di Toro D. M., Berry W. J., Burgess R. M., Mount D. R., O'Connor T. P., and Swartz R. C. (2005) The Predictive Ability of Sediment Quality Guidelines Derived Using Equilibrium Partitioning. In *Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments* (ed. R. J. Wenning and C. G. Ingersoll). SETAC Press.

Vidal D. E. and Bay S. M. (2005) Comparative Sediment Quality Guideline Performance For Predicting Sediment Toxicity In Southern California, USA. *Environ Tox. Chem.* **24**(12), 3173–3182.

Appendix 1

Empirical and Mechanistic Criteria

To put my review in context, I will quote from the paper "Comparative Sediment Quality Guideline Performance For Predicting Sediment Toxicity In Southern California, USA" (Vidal and Bay, 2005), cited in the report (p76), which examines these issues. First, the nature of the two methods:

> "Sediment quality guidelines can be classified in two main categories based on the approach used to derive their values: empirical and mechanistic. Empirical SQG approaches are based on the statistical analysis of large databases of synoptic sediment chemistry and toxicity data to identify chemical concentrations associated with various levels of biological effects. Examples of this type of SQG include the effects range–low and effects range–median (ERM) values, which are concentrations corresponding to the 10th and 50th percentiles of the distribution observed in toxic samples, respectively [2]. Variations in chemical speciation and bioavailability are not directly addressed in empirical SQGs; such effects are indirectly incorporated into these guidelines through the use of a database containing samples from diverse locations and sediment types. Empirical SQGs have two major practical advantages: they can be calculated for a large number of contaminants, and only routine chemical analysis data are needed for their application. "
>
> "The second principal type of SQG approach includes values based on mechanistic models that incorporate factors that affect the bioavailability of chemicals in the sediment. Mechanistic SQGs may incorporate the effects of sediment organic carbon or sulfides (for metals) on the equilibrium partitioning of contaminants and also use laboratory dose–response models to account for the effects of multiple contaminants [3–5]. Sediment quality guidelines based on equilibrium partitioning (EqP) for organics have been developed for selected pesticides and organics [6–8]. The EqP for organics theory assumes that nonionic chemicals in sediment partition between the organic carbon present in the sediment as well as in the interstitial (pore) water and the benthic organisms living on the sediment. At equilibrium, if a concentration is known in one of the phases (e.g., sediment), then the other ones can be predicted [6]. By accounting for variations in bioavailability and mixture effects, mechanistic SQGs have a greater ability relative to empirical SQGs to determine the specific contaminants responsible for toxicity. Mechanistic SQGs often require more extensive chemical data, and published values are not available for many contaminants, relative to empirical SQGs."

This is a correct characterization of the current understanding of the nature and appropriate use of the two methods. The report embraces the empirical methods and dismisses the mechanistic methods in a few sentences.

> "5.5.3.2 What chemistry indicators should be used? … Mechanistic SQGs based on equilibrium partitioning were not included for several reasons. Data for some of the key parameters needed to apply the mechanistic guidelines (e.g. sediment acid volatile sulfides and simultaneously extracted metals) were not available. In addition chemistry data were not available for all the potential toxicants in the samples, which limited the predictive ability of the guidelines for organics. Previous analyses using Southern California data showed that these limitations significantly affected mechanistic SQG performance; application of a partial suite of mechanistic SQGs for organics resulted in poor predictive ability (Vidal and Bay 2005)."

I regard this dismissal as premature and potentially dangerous. There has been much discussion in the literature and at meetings about the appropriate uses of empirical and mechanistic guides (Wenning and Ingersoll, 2005). The empirical guidelines suggested in this report are based on fitting a logistic probability model to large sets of amphipod mortality data sets collected in California. An equation is developed for each measured potential toxicant in the sediment. Then these probabilities are combined to make predictions of results of these The limitations of such a procedure are well known. To quote from Vidal and Bay, 2005

> "The results of these analyses showed that exceedances of individual empirical chemical guidelines are unreliable indicators of toxicity and do not necessarily indicate the cause of toxicity. For example, the mean SQGQ1q and mean ERMq had similar nontoxicity efficiency and specificity values, yet the mean SQGQ1q uses only nine chemicals in comparison to the 24 used for the mean ERMq. The presence of many contaminants in a sediment sample and the high degree of correlation among them indicates that most empirical SQG values should not be used in isolation but rather be used in combination to provide an overall indication of the potential for adverse effects (e.g., likely to be toxic or nontoxic). The exceedance of an individual empirical SQG value is not an indication that a chemical is toxic to organisms. Other studies have also suggested caution in the use of individual chemical SQG values when assessing sediment quality [14,16]. "

The Regression Model

The California regression model is based on the log logistic equation (page 13 of Appendix A and page 2 of Direct Effects Calculation)

$$p = \exp(b_0 + b_1 \log_{10}(c))/(1+ \exp(b_0 + b_1 \log_{10}(c)) \quad (1)$$

It can be shown that this equation is equivalent to the more intuitive formulation

$$p = 1/(1 + (EC50/c)^{\beta}) \quad (2)$$

where

$$\beta = b_1 /\ln(10) \quad (3)$$

$$EC50 = \exp(-b_0/ \beta) \quad (4)$$

The EC50 is the concentration at which a 50% mortality is predicted and β is the usual slope parameter.

The example in the Direct Effects Calculation can be used to check these equations.
For cadmium: c = 0.15 mg/kg, b_0 = 0.2894, b_1 = 3.1764 and p = 0.09. Using the above equations: β = 1.38, EC50 = 0.81 mg/kg and p = 0.09 as before. Note that the EC50 is approximately 1 mg Cd/kg by visual inspection of Fig. 2 in Direct Effects Calculation, which is consistent with EC50 = 0.81 mg/kg calculated above.The parameters for the other chemicals are listed below

Table 1
(Table 2 from Direct Effects Calculation and EC50 and β)

	units	b_0	b_1	β	EC50
Cd	mg/kg	0.2894	3.1764	1.38	0.81
Cu	mg/kg	-5.5931	2.5885	1.12	144.79
Pb	mg/kg	-4.7228	2.8404	1.23	46.00
Hg	mg/kg	-0.0618	2.6837	1.17	1.05
Zn	mg/kg	-5.1337	2.4205	1.05	132.11
HPAH	ug/kg	-8.1922	1.9995	0.87	12506.17
LPAH	ug/kg	-6.8071	1.8827	0.82	4126.72
Alpha Chlordane	ug/kg	-3.408	4.457	1.94	5.82
Dieldrin	ug/kg	-1.8344	2.589	1.12	5.11
Trans Nonachlor	ug/kg	-4.259	5.3135	2.31	6.33
Total PCBs	ug/kg	-4.4144	1.4837	0.64	944.64
4-4-DDT	ug/kg	-3.5531	3.2621	1.42	12.28

The Basis for the Model

The model parameters (b_0 and b_1, or equivalently EC50 and β) are based on regression fits to the toxicity and chemical data set assembled for this purpose. The report, appendices, and supplementary information do not contain the data and procedures from which these parameters were derived. In an attempt to understand the procedure in more detail, I have attempted to reproduce the fitting procedure. The Access database StatewideSQO_11_17_06.mdb is available on the web. I retrieved the Eohaustorius estuarius (EE) mortality data and the corresponding chemistry. It was not clear what data was used in generating the report values and I did not have the time completely understand this very large database. I restricted the retrieval to "SP" (survival percentage) and "SD_RESULT" (not replicates etc.) which seemed reasonable choices. One of my recommendations is that a report be prepared that documents the calculations that lead to the LRM in the report so that the analysis can be reproduced. Nevertheless the results of this analysis are very instructive.

This analysis will focus on cadmium as an illustration. The Cd data are presented below in Fig. 1.

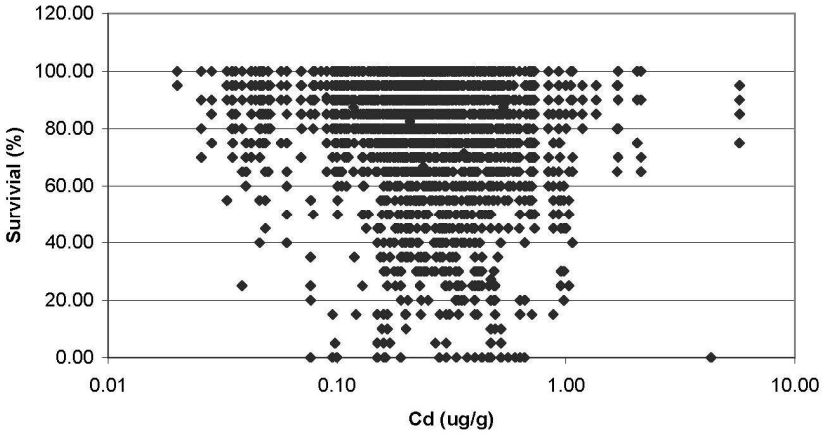

Figure 1

The other metals and PAH data are shown in figures 11 and 12 in the Figure appendix. The data all share a common feature. At low concentrations there is mostly >80% survival indicating no toxicity. At higher concentrations, some samples are not toxic (>80% survival) and others are highly toxic (0% survival). Note that these two extremes can occur *at the same cadmium concentration!* This is the central problem in understanding the toxicity of chemicals in field collected sediments with multiple contaminants. The difficulty is that it is not clear that Cd is causing toxicity in any of these sediments since bioavailability is not accounted for in empirical criteria. It is mechanistic criteria that strive to causally relate a chemical concentration to a toxic response.

This idea behind logistic regression models is to see if it is more probable that as the Cd concentration increases, the survival percentage increases. Fig. 3 presents the results of a fit of the logistic regression equation (2) to the data. The logistic equation using the parameters in Table 1 is also shown. A fit to the data produces an almost flat relationship, indicating that there is virtually no relationship between percent survival and Cd concentration. Yet the logistic equation using the Table 1 parameters seem to indicate a strong relationship.

The reason is, I think, that the data are prescreened before the logistic equation is fit. The procedure is described in Field et al., 2002.

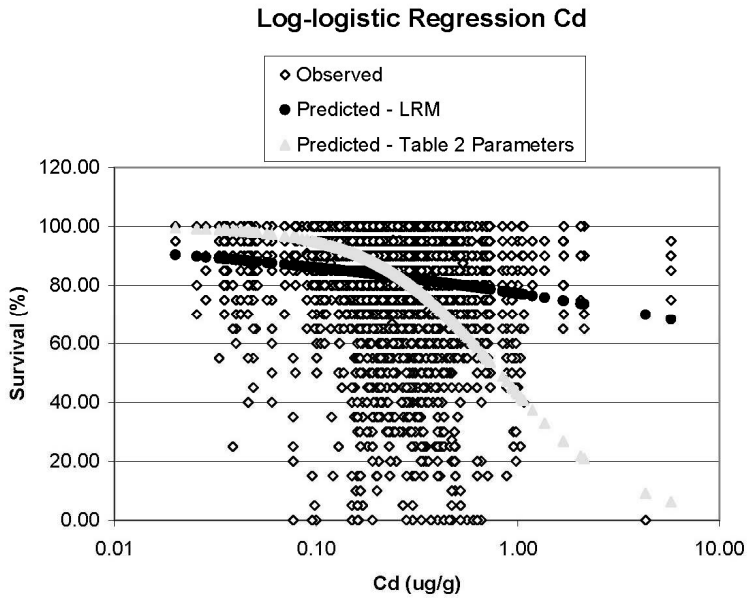

Figure 2

"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."

An example of the importance of pre-screening the data is shown in Fig. 4 from Field et al, 1999. Before screening, there is virtually no relationship between probability of toxicity and phenanthrene concentration. After prescreening, there is a very nice relationship. Thus the role of pre-screening is critical to the development of LRMs.

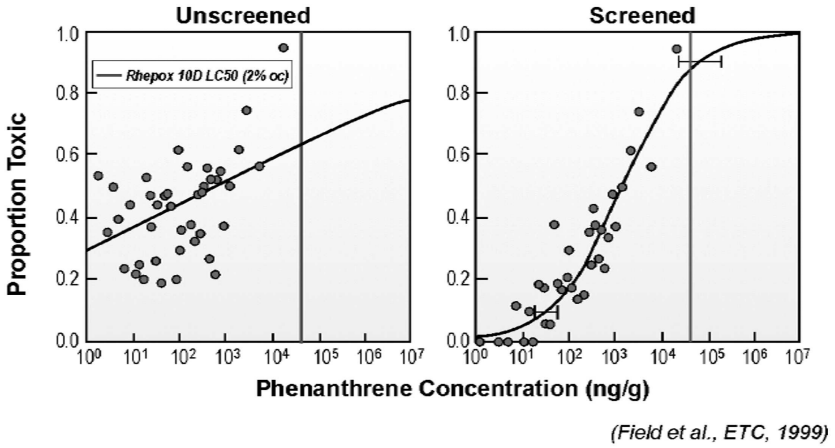

Figure 3

Following this procedure, the median concentration of all nontoxic (survival >80%) samples was found (Cd = 0.26 mg/kg). Then all toxic samples (survival >80%) for which Cd < 0.26 mg/kg were removed. The result is shown in Fig. 5. Since the samples that exhibited toxicity at low Cd concentrations (the samples in the lower left quadrant) have been removed, there is now a relationship between toxicity and Cd concentration. A fit of equation (2) to the screened data is now closer to the result using the Table 1 parameters. Since the methodology used to derive the results in the report are not available, it is not possible to understand why there is still a discrepancy. Nevertheless, it is clear that the pre-screening of the data is a critical part of the analysis.

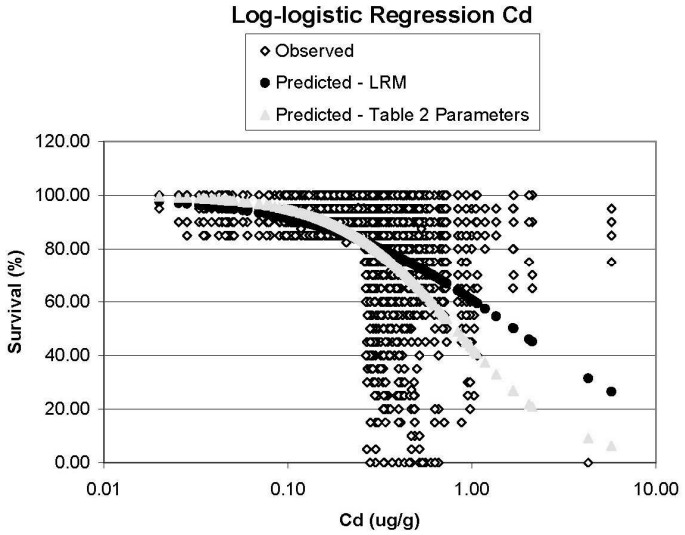

Figure 4

Critique of the Logistic Regression Model (LRM)

Consider the situation when the logistic regression model (LRM) is applied to a new sediment sample. The probability of survival that is computed from the Cd concentration uses the curve derived from the data in Fig. 5. But applying that curve presupposes that the new data comes from the prescreened data set, i.e. it is known a priori that whatever toxicity the sample might exhibit is not due to Cd if the Cd concentration is low. But there is no way of actually knowing that is the case for the new sample at hand. It is, rather, an assumption upon which the method is based. Also note that this result is not peculiar to cadmium. All the toxicity-chemistry data share the same general pattern, and all are pre-screened to produce the LRM.

Another interesting feature of the LRM is that the EC50s for the metals, which are derived from the screened dataset, are comparable to the median concentrations of the metals in the entire dataset. Fig. 6 presents the ratio of the EC50 (Table 1) to the median concentrations computed from the entire data set and also for the non-toxic samples. The ratio ranges from 1 to 4, indicating that the EC50 used in the LRM is a measure of the general level of contamination of the sediments in the dataset. Also the β's are roughly the same. This suggests that for the metals at least, the LRMs are modeling the extent of contamination. They predict low toxicity if the level of metal concentration is well below the median concentration in the datasets.

This is not an unreasonable way to predict *lack* of toxicity for relatively clean, i.e. uncontaminated, sediments. However, it is not much of a guide for predicting the actual toxicity if the level of contamination is larger. The reason the logistic model "fits the data" is that the troublesome data – those showing toxicity at low concentrations – are removed by the pre-screening procedure.

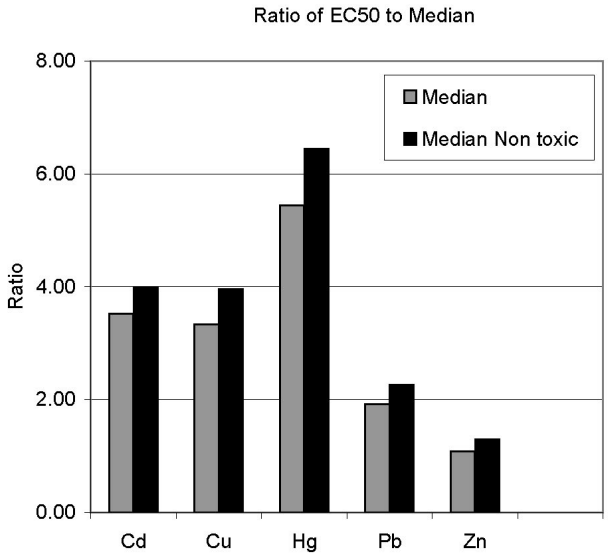

Figure 5

Use of Mechanistic Models

Mechanistic models attempt to relate toxicity to the bioavailable fraction of the chemicals in sediments. The most well developed of there are for mixtures of metals (Ankley et al., 1996, USEPA, 2005) and mixtures of PAHs (Di Toro & McGrath, 2000, USEPA, 2003). They use the Equilibrium Partition Model (Di Toro et al., 1990) as the general framework and apply toxicity mixture and partitioning models to predict the toxicity of single chemicals and chemical mixtures. The models have been validated using spiked sediments (Berry et al., 1996) for which the toxic chemical(s) are not in doubt. Additionally field datasets have been employed that are heavily contaminated with either metals (Hansen et al., 1996) or PAHs (Di Toro & McGrath, 2000) for which the chemicals causing the toxicity can be reasonably assumed to be known.

It has been found that for the large dataset employed for establishing the empirical criteria in this report, the mechanistic criteria do not appear to be as predictive as the empirical criteria. Fig. 7 presents the results of the analysis from Vidal and Bay 2005.

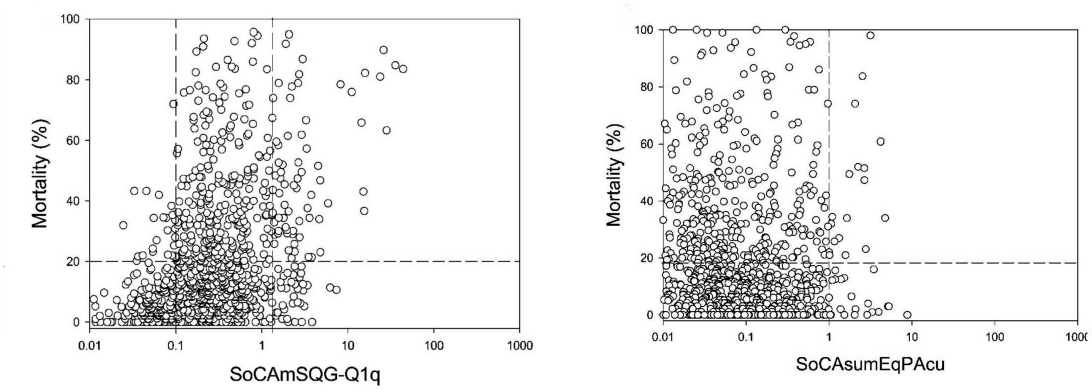

Figure 6 from Vidal and Bay 2005

The SoCAmSQG-Q1q model has very much the same pattern of predictive power as the individual datasets (Fig. 5). For a low level of contamination there is only a control level of mortality. At higher levels of contamination there are both toxic and non-toxic sediments at the same level of contamination (the x-axis). By contrast the EqP comparison shows no discrimination.

There are a number of possible reasons for the failure of the EqP based predictions Certainly one important problem is the lack of the appropriate measure of the critical metal binding parameter acid volatile sulfides (AVS) (Di Toro et al., 1990, 1992) in the majority of sediments in the dataset. The second is the lack of measurements for all the significant PAHs that may be present (McGrath & Di Toro 2000). Finally, and the most vexing problem, is the lack of measurements for other compounds that may be causing toxicity. Nevertheless, the EqP models can be very useful in understanding the possible causes of toxicity.

SEM-AVS Model of Metal Toxicity

For metal toxicity, it has been shown that if the molar sum of the metal concentrations that is simultaneously extracted (∑SEM) with the AVS is less than the AVS concentration, i.e. ∑ SEM – AVS < 0 no toxicity is expected. This has been demonstrated using acute and chronic laboratory spiked and field deployed spiked sediments (Di Toro et al., 2005). SEM data are not available but the molar sum of the total extracted metals (Total Metal = Cd + Cu + Ni + Pb + Zn) are available and inferences can be drawn from these concentrations. Fig.8 presents the data.

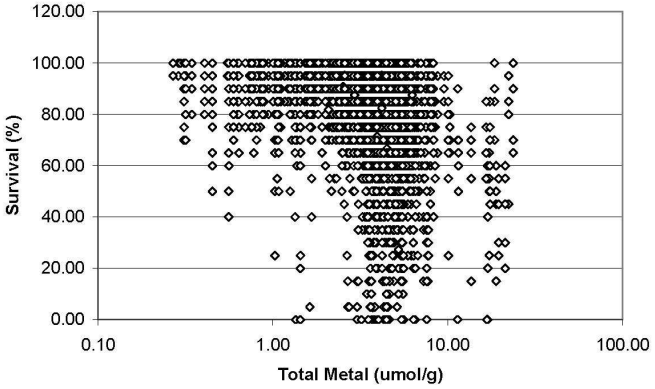

Figure 7

The pattern of the data is not much different from either Cd (fig 2) or the other metals (Figs. 11-12). However the difference is that this distribution can be interpreted in terms concentration of AVS in sediments. For example, little mortality is seen for total metal concentrations < 2 umol/g. If the AVS in all the sediment samples were at least 2 umol/g, not a large amount of AVS for muddy sediments, then the lack of toxicity due to metals would be expected. If AVS concentrations were available for all the data, then metal toxicity could be unambiguously ruled out for those sediment for which Total Metal – AVS < 0, since this would guarantee that $\sum$ SEM – AVS < 0.

There is a small amount of AVS data in the database for which Total Metal – AVS can be calculated and compared to observed mortality. These are shown in Fig. 9. Most of the toxic sediments have AVS concentrations greater than Total Metal, i.e. Total Metal –AVS < 0. Since Total metal > $\sum$ SEM, the data would plot further toward the negative values if $\sum$ SEM were available. This would indicate that in these sediments AVS is greater than $\sum$ SEM and it is unlikely that metals are causing toxicity in this subset of the database. The point is that a judgment can be made about the likely cause of toxicity in these sediments that is not possible using the empirical criteria.

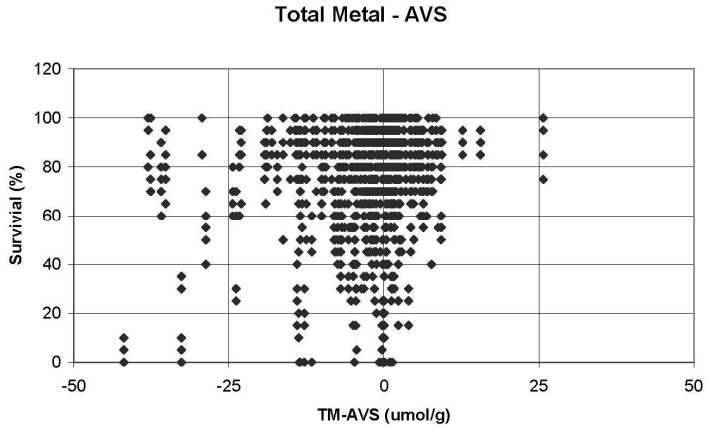

Figure 8

Target Lipid Model of PAH Mixture Toxicity

An EqP model has been developed for mixtures of PAHs is sediments, based on the Target Lipid Model of Narcosis Toxicity (Di Toro et al., 2000). The Criteria corresponding to chronic effects and 10 day Rhepoxynius abronius survival are listed in Table 2. The average and standard deviation of criteria for low (LPAH) and high (HPAH) molecular weight PAH sums as well as total PAH are listed. The toxicity of a mixture is found by summing the toxic units – the ratio of the concentrations to the criteria in Table 2 -- comparing the results to one toxic unit for 50% effect. To a good approximation, the same result is obtained be evaluated by summing the organic carbon normalized molar concentrations of PAHs and comparing the sum to the average criteria. The reason is that the organic carbon normalized sediment criteria for the individual PAHs do not vary very much. For example, the criteria vary from 16.18 to 21.96 umol/gOC for the R. abronius LC50s. An explanation based on the equations for toxic units is available (Di Toro & McGrath, 2000).

Table 2

PAH Sediment Criteria for Chronic Effects and 10 day Rhepoxynius abronius Survival

Chemical	CAS number	MW (g/mol)	Log Kow	Chronic EC50 (umol/gOC)	R. abronius LC50 (umol/gOC)
Acenaphthylene	208968	152.2	3.22	5.03	16.18
Naphthalene	91203	128.19	3.36	5.09	16.38
1-Methylnaphthalene	90120	142.2	3.84	5.31	17.08
2-Methylnaphthalene	91576	142.2	3.86	5.32	17.11

Acenaphthene	83329	154.21	4.01	5.39	17.34
Fluorene	86737	166.2	4.21	5.48	17.64
2,6-Dimethylnaphthalene	581420	156.23	4.37	5.56	17.89
Anthracene	120127	178.2	4.53	5.64	18.15
Phenanthrene	85018	178.2	4.57	5.66	18.21
2,3,5-Trimethylnaphthalene	2245387	170.26	4.86	5.8	18.68
LPAH				5.43(0.25)	17.5(0.80)
Pyrene	129000	202.26	4.92	5.83	18.78
1-Methylphenanthrene	832699	192.26	5.04	5.89	18.97
Fluoranthene	206440	202.26	5.08	5.92	19.04
Benzo[a]anthracene	56553	228.29	5.67	6.23	20.05
Chrysene	218019	228.29	5.71	6.25	20.12
Benzo[a]pyrene	50328	252.31	6.11	6.47	20.84
Perylene	198550	252.31	6.14	6.49	20.89
Benzo[e]pyrene	192972	252.32	6.14	6.49	20.89
Benzo[b]fluoranthene	205992	252.32	6.27	6.56	21.13
Benzo[k]fluoranthene	207089	252.32	6.29	6.58	21.17
Benzo[ghi]perylene	191242	276.34	6.51	6.7	21.58
Dibenz[a,h]anthracene	53703	278.35	6.71	6.82	21.96
Indeno[1,2,3-cd]pyrene	193395	276.34	6.72	6.83	21.98
HPAH				6.39(0.34)	20.6(1.1)
TPAH				5.97 (0.57)	19.2(1.84)

The total PAH data in units of umol/gOC is presented in Fig. 10. It is computed from the low (LPAH) and high (HPAH) molecular weight PAH data using average molecular weights for these classes, and the organic carbon concentration of the sediment, which is in the database.

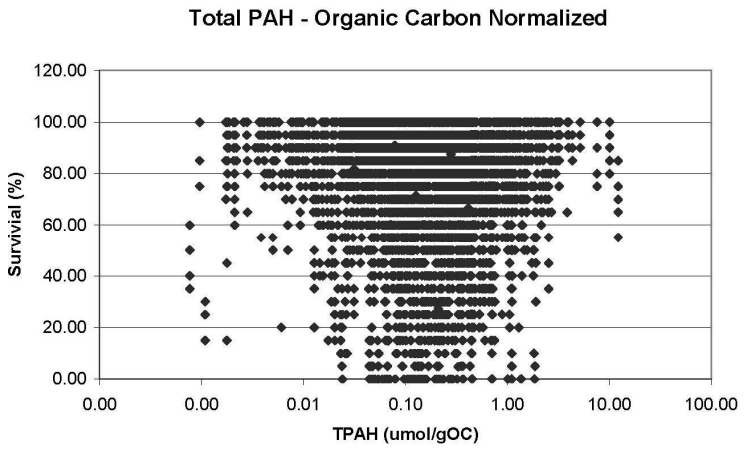

Figure 9

The data has the same shape as the previous chemicals, but as with the metals data, the appropriate toxicity threshold is known. For the 10 day Rhepoxynius abronius survival the average LC50 is 19.2 umol/gOC (TPAH in Table 2). No sediment in Fig. 10 appears to exceed this threshold so it appears that PAHs are not the cause of the toxicity in any of these samples.

There is a significant problem, however, with applying this logic to these data. The criteria apply to the sum of *all* PAHs. But the available data are for only the PAHs in bold face type in Table 2. In particular the alkylated PAHs, which are primarily associated with petroleum contamination, can be a large component of the TPAH and these are not being adequately measured. For these data there is only one representative component 2-methylnaphthalene. Thus it is possible that the total PAH concentration in the sediments could be larger.

The conclusion of this analysis is either that PAHs are not the cause of toxicity in these sediments, or there is large fraction of PAHs that are not being measured, that are contributing to toxicity.

Summary of Empirical and Mechanistic Model Applications

The purpose of this appendix is to examine the utility of empirical and mechanistic models in the evaluation of toxicity of sediment samples. The empirical models estimate the probability of observing toxicity based on the level of contamination. When the sediments have low levels of most contaminants, they predict that the sediment will not be toxic. This conclusion is almost forced by the pre-screening procedure. As levels increase the prediction is that toxicity becomes more likely. But it should be clear from the above analysis that the *cause(s)* of the toxicity cannot be judged from empirical criteria. They are simply responding to the increasing level of overall contamination. The higher the overall level of contamination, the more like it is that toxicity will be found.

The mechanistic criteria can make predictions about which classes of chemicals are possibly involved in the observed toxicity. If the AVS exceeds the total metal concentration, metal toxicity is almost surely not present. If the organic carbon molar sum of the PAHs in the sediment, including the alkylated compounds, is less than the appropriate LC50 for the species being tested, e.g. 19.2 umol/gOC for *Rhepoxynius abronius* survival, then PAHs are almost surely not the cause of toxicity.

If neither metals, nor PAHs are the causes of toxicity, and similar screening calculations can be made for other measured constituents, this information can be included in the next step in the investigation. At least, we know we either know or do not know the causes of toxicity. If the causes are known, we can proceed with confidence. If the cause is unknown, than a completely different approach is warranted. This is crucial information to making judgments about whether sediments are toxic due to chemical contamination, and whether the information at hand is consistent with known chemical modes of toxicity in sediments.

Figure Appendix

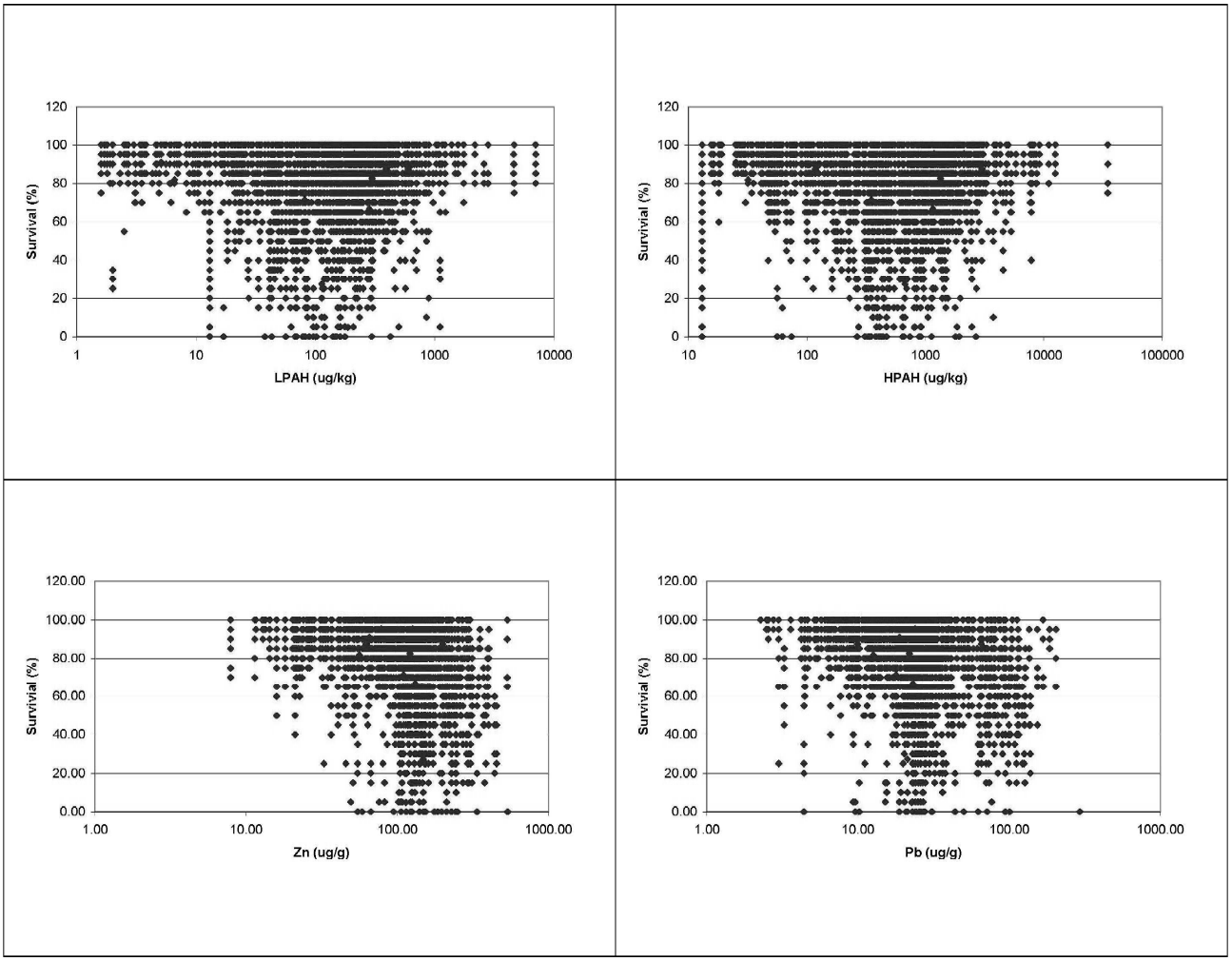

Figure 11

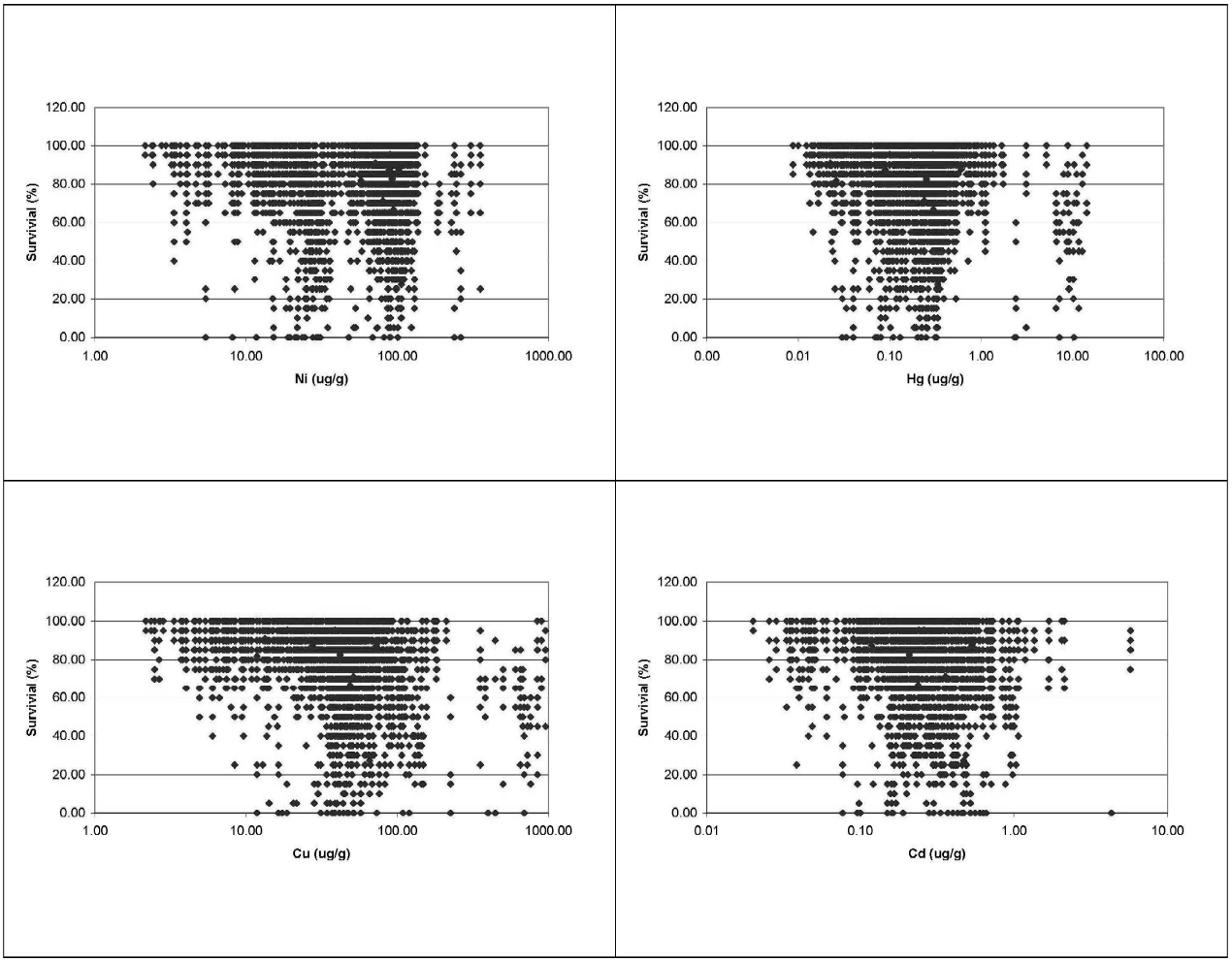
Survivial (%)
120.00
100.00
80.00
60.00
40.00
20.00
0.00
1.00
10.00
100.00
1000.00
Ni (ug/g)
Survivial (%)
0.00
0.01
0.10
1.00
10.00
100.00
Hg (ug/g)
Survivial (%)
1.00
10.00
100.00
1000.00
Cu (ug/g)
Survivial (%)
0.01
0.10
1.00
10.00
Cd (ug/g)

Figure 12

References

Ankley G. T., Di Toro D. M., Hansen D. J., and Berry W. J. (1996) Technical basis and proposal for deriving sediment quality criteria for metals. Environ. Tox. Chem. 15(12), 2056-2066.

Berry W. J., Hansen D. J., Mahony J. D., Robson D. L., Di Toro D. M., Shipley B. P., Rogers B., Corbin J. M., and Boothman W. S. (1996) Predicting the toxicity of metals-spiked laboratory sediments using acid volatile sulfide and interstitial water normalizations. Environ. Toxicol. Chem. 15(12), 2067-2079.

Di Toro D. M., Mahony J. D., Hansen D. J., Scott K. J., Carlson A. R., and Ankley G. T. (1992) Acid volatile sulfide predicts the acute toxicity of cadmium and nickel in sediments. Environ. Sci. Tech. 26(1), 96-101.

Di Toro D. M., Mahony J. D., Hansen D. J., Scott K. J., Hicks M. B., Mayr S. M., and Redmond M. S. (1990) Toxicity of Cadmium in Sediments: The Role of Acid Volatile Sulfide. Environ. Toxicol. Chem. 9, 1487-1502.

Di Toro D. M., McGrath J. A., and Hansen D. J. (2000) Technical basis for narcotic chemicals and polycyclic aromatic hydrocarbon criteria. I. Water and tissue. Environ.Toxicol.Chem. 19, 1951-1970.

Di Toro D. M., McGrath J. M., Hansen D. J., Berry W. J., Paquin P. R., Mathew R., Wu K. B., and Santore R. C. (2005) Predicting Sediment Metal Toxicity Using a Sediment Biotic Ligand Model: Methodology and Initial Application. Environ Tox. Chem. 24(10), 2410-2427.

Di Toro D. M., Zarba C. S., Hansen D. J., Berry W. J., Swartz R. C., Cowan C. E., Pavlou S. P., Allen H. E., Thomas N. A., and Paquin P. R. (1991) Technical Basis for the Equilibrium Partitioning Method for Establishing Sediment Quality Criteria. Environ. Toxicol. Chem. 11(12), 1541-1583.

Field L., Macdonald D., Norton S., Ingersoll C., Severn C., Smorong D., and Lindskoog R. (2002) Predicting Amphipod Toxicity From Sediment Chemistry Using Logistic Regression Models. Environ. Toxicol. Chem. 21(9), 1993-2005.

Field L., Macdonald D., Norton S., Ingersoll C., Severn C., Smorong D., and Lindskoog R. (2002) Predicting Amphipod Toxicity From Sediment Chemistry Using Logistic Regression Models. Environ. Toxicol. Chem. 21(9), 1993-2005.

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.

Hansen D. J., Berry W. J., Mahony J. D., Boothman W. S., Di Toro D. M., Robson D. L., Ankley G. T., Ma D., Yan Q., and Pesch C. E. (1996) Predicting the toxicity of metals-contaminated field sediments using interstitial concentrations of metal and acid volatile sulfide normalizations. Environ. Toxicol. Chem. 15(12), 2080-2094.

McGrath J. A. and Di Toro D. M. (2000) Distinguishing between petrogenic and pyrogenic soruces of PAHs in field collected sediments. 20th Annual Meeting, Society of Environmental Toxicology and Chemistry, 181.

USEPA. (2003) Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms: PAH Mixtures. U.S Environmental Protection Agency.

USEPA. (2005) Equilibrium Partitioning Sediment Benchmarks (ESBs) for the Protection of Benthic Organisms: Metal Mixtures (Cadmium, Copper, Lead, Nickel, Silver and Zinc) [Draft]. U.S Environmental Protection Agency.

Vidal D. E. and Bay S. M. (2005) Comparative Sediment Quality Guideline Performance For Predicting Sediment Toxicity In Southern California, USA. Environ Tox. Chem. 24(12), 3173–3182.

Wenning, R.J., Ingersoll, C.G. (2005) Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments SETAC Press, Pensacola, FL