

Empirical and Theoretical Shortcomings of Sediment Quality Guidelines

Thomas P. O'Connor

National Status and Trends Program, NOAA N/SCI1,
1305 East West Highway, Silver Spring, MD 20910

Abstract—In the United States, the EPA EMAP-Estuaries Program and the NOAA Bioeffects surveys provide large data sets with which to test proposed relationships between sediment chemistry and toxicity. While there is an increased likelihood that sediment will be toxic as its concentration of chemical contaminants increases, the connection is too weak to allow chemical data to be used as a sole basis for management decisions. Additionally, guidelines based equilibrium partitioning have two questionable assumptions; one that thermodynamic equilibrium exists and, two, that pore water (also called interstitial water) is a medium of exposure.

Introduction

Sediment quality guidelines (SQGs), proposed methods for attaching biological significance to chemical data, fall into two categories; those basing estimates of toxicity on water quality criteria and assumptions about equilibrium and pore water, and those based on compilations of co-occurrences between biological characteristics and bulk chemical concentrations. Using a large dataset compiled from NOAA and EPA surveys that simultaneously measured sediment chemistry and toxicity data O'Connor and Paul (2000) showed that none of the SQGs are sufficiently reliable predictors of toxicity to be use for sediment management. That analysis is repeated here and expanded to include recent modifications of the co-occurrence methods and a newly developed equilibrium-based SQG for polycyclic aromatic hydrocarbons(PAHs)

Co-occurrence SQGs

Long and Morgan (1990) introduced SQGs based on exhaustive compilations of data from reports where bulk chemical concentrations in sediment were quantified along with a measure of biological response. The responses could be from bioassays of any kind or field evidence that benthic communities were to some extent abnormal. Using only cases with a biological effect each chemical was ranked by concentration. The 10th percentile among those concentrations was designated the Effects Range Low (ERL) and the 50th percentile deemed the Effects Range Median (ERM). The SQG then rested on a claim that sediments with bulk concentrations below the ERL were not likely to be toxic while bulk concentrations above the ERM signaled a likelihood of toxicity. It was never demonstrated that any particular chemical out of many measured in any particular sample was responsible for the biological effect and that sample. By 1995 Long et al. (1995) had published a list of ERLs and ERMs for 9 trace elements, 12 polycyclic aromatic

hydrocarbons (PAHs), p,p'DDE, and for the groups; Total DDTs, Total PCBs, Low molecular weight PAHs, High molecular weight PAHs, and Total PAHs (Table 1). A similar set of SQGs based on co-occurrence has been developed by MacDonald et al. (1996). In their case the SQGs are TELs (threshold effect levels) and PELs (probable effect levels). The differences between TELs and ERLs and between PELs and ERMs are small, so the validity of SQGs based on co-occurrence is tested here only with ERLs and ERMs.

Originally, an exceedance of any one ERM was considered a likely indication that a sample was toxic. However, upon examination of large data sets Long and MacDonald (1998) categorized the likelihood of toxicity as low, medium, or high if 1 to 5, 6 to 10, or more than 10 chemical concentrations, respectively, exceeded an ERM. As an alternative to multiple exceedances they introduced the ERM Quotient (ERMQ), which is the mean among ratios of chemical concentrations to ERMs. An ERMQ >1 would imply a medium likelihood of toxicity.

Equilibrium partitioning SQGs

The co-occurrence methods have no theoretical underpinning and lack any basis for attaching observed effects with particular chemicals. Alternative SQGs, on the other hand, are based on the assumption that pore water is the medium of exposure and that chemical quality of pore water can be calculated from bulk chemical characteristics by assuming thermodynamic equilibrium between sediment and pore water (Table 2). The SQG is then the long established Water Quality Criterion applied to calculated pore water quality. There are SQGs resting on these assumptions for neutral organic compounds EPA (1993) and for trace metals with less soluble sulfides than ferrous sulfide (Di Toro et al., 1992). In the first instance pore water concentrations are calculated on the basis of organic carbon-normalized bulk concentrations, in the second, metals cannot reach toxic levels in pore water if bulk sediment contains sufficient acid-volatile sulfide (AVS) to theoretically precipitate all metals as sulfides.

Two expansions of equilibrium partitioning theory for neutral organics have been derived for PAHs in which pore water can be deemed toxic if the incremental toxicities of individual compounds add up to a total toxicity greater than one. The first of these (Swartz et al., 1995) is based on measured or estimated LC50 concentrations for 13 PAH compounds. The second (Di Toro et al., 2000) assumes that toxicity is a narcotic effect so that all PAHs are equally toxic on a molar basis and the critical concentration is that of sum of concentrations of 34 individual compounds.

Empirical test of SQGs

The 2475-sample data set (O'Connor and Paul, 2000 and references therein) allows an empirical test of how well the proposed SQGs conform to sediment toxicity or, in the case of the ERLs and AVS methods, to non-toxicity. Samples were judged toxic if less

than 80% of test amphipods did not survive a ten-day exposure. This test was introduced by Swartz et al. (1979) because surveys from a variety of locations showed that amphipod species decreased in abundance near sources of contamination. Through wide application, it has become the *de facto* standard of sediment toxicity. The AVS guideline weighs AVS against the sum of concentrations of metals that are simultaneously extracted (SEM) with AVS in 1M HCl. The metal concentrations in our large dataset are based on total extractions (TotM) in concentrated HF. Since TotM must equal or (usually) exceed SEM, any sample in which AVS exceeds TotM is also one in which AVS exceeds SEM.

The results, in Table 3, reveal that generally less than half of the samples with chemical concentrations above any of the SQGs were actually toxic. A few did better than 50%, but at the price of excluding many samples. For example SQGs requiring 6 or more ERM exceedances or an ERMQ >1 were about 60% accurate in identifying toxic samples, but that increase in accuracy caused the SQG to apply to ten times fewer samples than the simple single-ERM exceedance SQG. Similarly the SQG based on calculated pore water toxicity of 13 PAHs was 53% accurate but applied only to 17 samples, and the SQG based on acute toxicity from 34 PAHs was 58% accurate but applied to only 12 samples.

The two SQGs that predict non-toxicity appear to have worked well, only 5% and 11% of the samples without an ERL exceedance or where AVS exceeded metal concentrations, respectively, were toxic. This success, however, may be an artifact because fully 84% of the samples were non-toxic. A rigorous test of these SQGs would examine their ability to identify non-toxic samples from a set of primarily toxic samples. Also, it is worth noting that the median TotM concentrations among samples with and without excess AVS were 1.63 and 1.65 $\mu\text{M/g}$, respectively, or essentially identical. The corresponding median AVS concentrations were very different, 6.44 and 0.43 $\mu\text{M/g}$, respectively. So, despite the fact that the objective of SQGs is to estimate toxicity from chemical contamination, this guideline is very sensitive to AVS and has little to do with metal contamination.

In the entire 2475 sample dataset only 388 (16%) were toxic. So, while the SQGs that predict toxicity did so with generally less than 50% accuracy, they did much better than a random choice would have done. This fact favors the use of SQGs as guidelines to identify samples requiring further examination, such as performing toxicity tests or quantifying the indigenous benthic community. The demonstrated 50% accuracy also signals, however, that SQGs, by themselves, should not be used by environmental managers as a basis for any decisions.

Theoretical limitations to chemical predictors

The SQGs based on equilibrium partitioning theory rest on two questionable assumptions; one that pore water is in equilibrium with sediment and, two that benthic organisms expose themselves to pore water.

Equilibrium has been demonstrated (McGroody and Farrington, 1995) not to apply to PAHs that are in sediments as particles of soot, and Di Toro et al. (2000) acknowledge that SQGs cannot be applied to sediments with that component, especially common in sediments with high PAH concentrations. Even for other neutral organic compounds there are no examples from the field where concentrations in the bulk sediment and pore water phases are in equilibrium. The only actual measurements of pore water concentrations brought to bear by the U. S. EPA (1993) on the issue of demonstrating equilibrium are laboratory data with spiked systems.

Systems in thermodynamic equilibrium are not only at steady-state (no temporal gradients) but also have no spatial gradients. There are countless examples in the literature of very sharp vertical gradients in chemical concentrations in sediments. In cases where sediment cores can be dated, these vertical gradients provide chronologies of changes in chemical flux to seafloor. Though measured much less frequently, there are also gradients in pore water concentrations. For example, Luther et al. (1998) have measured sharp concentration differences in pore water over distance of a few mm.

Even more to the point, however, are the facts that Luther et al. (1998) and Zeibis et al. (1996) have managed to measure oxygen concentrations within the burrows of benthic organisms. Infaunal benthic organisms demonstrably ventilate their burrows. While surrounding sediments and pore waters are anoxic, organisms ensure that the water they live in is aerobic. So, benthic organisms are not living in pore water. Even if pore water is in thermodynamic equilibrium with bulk sediment, it is not a medium of exposure (or toxicity) to benthic life.

Conclusions

The connection between sediment chemistry and toxicity is not strong enough to be predictive. There is no obvious way to improve the situation. While the chemical data are fixed, the definition of toxicity is somewhat fluid. Requiring a more than 20% loss of amphipods, or using a less sensitive assay, or requiring a "hit" in more than one assay would all result in fewer samples being declared toxic. Then all the various SQG exceedances would still apply and their corresponding accuracies would be even lower. Using a more sensitive test would find more toxic samples and seem to increase accuracy but also increase the percentage of samples that are toxic but not predicted to be so by any SQG (already 202, or 52%, of the 388 toxic samples in Table 3 do not exceed any chemical SQG). In the extreme, one could use a series of tests and declare a sample toxic on the basis of a "hit" in any one. Long and MacDonald (1998), for example, did that with a set of samples on which three tests were run. This did result in 78% of the samples with a single ERM exceedance also being toxic, but with 70% of all the samples (regardless of chemistry) being in the toxic category. In effect a very low threshold for declaring toxicity makes predictions based on chemistry or any other basis irrelevant.

Nonetheless, the SQGs are better than a random choice for finding toxic sediments and they can serve as guidelines to identify samples requiring further examination.

References

DiToro, D. M, Mahoney, 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 . Technol .26*: 96-101

Di Toro, D. M. and McGrath, J. A. 2000. Technical basis for narcotic chemicals and polycyclic aromatic hydrocarbon criteria. II. Mixtures and sediments. *Environ. Toxicol. Chem.* 19: 1971-1982

Long, E. R. and Morgan, L. G. 1991. The potential for biological effects of sediment sorbed contaminants tested in the National Status and Trends Program. NOAA Technical Memorandum NOMA 52. National Oceanic and Atmospheric Administration. Seattle, WA

Long, E. R., MacDonald, D.D., Smith, S. L., and Calder, F. D. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manage* 19: 81-97

Long, E. D., and MacDonald, D. D. 1998. Recommended uses of empirically derived, sediment quality guidelines for marine and estuarine ecosystems. *Human and Ecol. Risk Assess.* 4: 1019-1039

Luther, G. W. III, Brendel, P. J., Lewis, B. L., Sundby, B., Lefrancois, L., Silverberg, N., and Nuzzioi, D. B. 1998. Simultaneous measurement of O₂, Mn, Fe, I, and s(-II) in marine pore waters with a solid-state voltammetric microelectrode. *Limnol. Oceanogr.* 43: 325-333

McGroody, S. S. and Farrington, J. W. 1995. Sediment pore water partitioning of polycyclic aromatic hydrocarbons in three cores from Boston Harbor, Massachusetts. *Environm. Sci. Technol.* 29: 1542-1550

MacDonald, D. D., Carr, R. S., Calder, F. D., Long, E. R., and Ingersoll, C. R. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Exotoxicology* 5: 253-278

O'Connor, T. P. and Paul, J. F. 2000. Misfit between sediment toxicity and chemistry. *Mar. Poll. Bull.* 40: 59-64

Swartz, R. C., DeBen, W. A., and Cole, F. A. 1979. A bioassay for the toxicity of sediment to marine macrobenthos. *J. Wat. Poll. Contr. Fed.* 51: 944-950

Swartz R. C., Schults D. W., Ozretich R. J., Lambertson J. O., Cole F. A., DeWitt T. H., Redmond M. S., and Ferraro S.P. 1995. Σ PAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected samples. *Environ .Sci . Technol* .14:1977-1988

U. S. Environmental Protection Agency. 1993. Technical basis for establishing sediment quality criteria for nonionic contaminants for the protection of benthic organisms by using equilibrium partitioning. EPA 822-R-93-011. U. S. EPA, Office of Science and Technology, Health and Ecological Criteria Division, Washington, DC

Zeibis, W., Forster, M., Huettel, M., and Jorgensen, B. B.. 1996. Complex burrows of the mud shrimp *Callinassa truncata* and their geochemical impact in the sea bed. *Nature* 382: 619-622

Table 1. ER-Ls and ER-Ms from Long et al. (1995b). ER-Ls are bulk sediment concentrations (dry-weight) below which sediment is unlikely to be toxic. ER-Ms are concentrations above which toxicity is probable.

Chem.	ERL	ERM
Ag	1 ppm	3.7 ppm
As	8.2	70
Cd	1.2	9.6
Cr	81	370
Cu	34	270
Hg	0.15	0.71
Ni	21	52
Pb	47	220
Zn	150	410
p,p'-DDE	2.2 ppb	27 ppb
Total DDT ^a	1.6	46
Total PCB	23	180
LMWPAH ^a	550	3160
HMWPAH ^a	1700	9600
Total PAH ^a	4000	45000
Acenaphthene	16	500
Acenaphthylene	44	640
Anthracene	85	1100
Flourene	19	540
2-Methylnaphthalene	70	670
Naphthalene	160	2100
Phenanthrene	240	1500
Benzo(a)pyrene	261	1600
Chrysene	384	2800
Dibenzo(a,h)anthracene	63	260
Fluoranthene	600	5100
Pyrene	670	2600

^aDDE is part of the Total DDT and individual PAH compounds are part of the Total PAH, LMWPAH (2- and 3-ring compounds), and HMWPAH (more than 3-ring compounds). However, double and triple counting is of no consequence when we are interested in knowing only whether there is at least one exceedance of ERL or ERM.

Table 2. SQGs based on equilibrium partitioning of neutral organic compounds or acid-volatile sulfide (SQGs for neutral organics do not apply if Total Organic Carbon (TOC) content of sediment is less than 0.2%, dry weight)

Individual chemicals (EPA, 1993)

Chemical	SQG*	
dieldrin	20000	ng/gTOC
endrin	760	
acenaphthene	230000	
fluoranthene	660000	
phenanthrene	240000	

*Based on Final Chronic Value (FCV) of the Water Quality Criteria and the equilibrium partitioning assumptions

Sum of 13 PAHs (Swartz et al., 1995)

$$\Sigma \text{PAH}_{iW} / \text{LC50} > 1$$

Where PAH_{iW} is the interstitial water concentration calculated on the basis of the equilibrium partitioning assumptions

Narcosis due to combined 34 PAHs (Di Toro et al., 2000)

Biological basis	SQG*($\Sigma_{34}\text{PAH}$)	SQG* $\Sigma_{23}\text{PAH}$
FAV ($C_L=19.3$)	29 uM/gTOC	18 uM/gTOC
LC50 ($C_L=12.2$)	18	11
FCV ($C_L=3.79$)	5.7	3.5

Where the C_L 's are body burdens on a lipid basis of PAHs that correspond to the Final Acute Values (FAV) and Final Chronic Values (FCV) for Water Quality Criteria and for the 10-day LC50 for an amphipod species. The SQGs are calculated on the basis of the equilibrium partitioning assumptions. SQGs are based on the sum of concentrations of 34 PAH compounds. Where only 23 PAH compounds were measured (as in most of the samples in the 2475-sample dataset used here). Di Toro et al. (2000) calculated that the median ratio of 1.64 between $\Sigma_{34}\text{PAH}$ and $\Sigma_{23}\text{PAH}$.

Acid Volatile Sulfide (Di Toro et al., 1992)

This SQG applies only to metals whose solubilities are less than that of FeS. If on a mole basis sum of concentrations of such metals, (i.e. $\Sigma \text{Zn,Cu,Ni,Pb,Cd,Hg,Ag}$) is less than the mole concentration of sulfide that can be volatilized with 1M HCl (i.e. AVS), the concentrations of all of those metals in interstitial water is below Water Quality Criteria and that sediment cannot exert toxicity due to those metals.

Table 3. Frequencies of 10-day amphipod toxicity in samples with chemical concentrations in excess of an SQG (or non-toxicity in samples with an SQG for non-toxicity).

	Total	Toxic
All Samples	2475	388 (16%)
Samples with one chemical concentration >ERM	453	186 (41%)
Samples with six or more chemical concentrations >ERM	46	29 (63%)
Samples with ERM Quotient ($1/n\sum(\text{Conc}/\text{ERM})$) >1	59	37 (63%)
Samples with one neutral organic concentration >SQG	31	13 (42%)
Samples with $\sum(\text{PAH}_{iW}/\text{LC50}) >1$	17	9 (53%)
Samples with $\sum_{23}\text{PAH} > 18 \text{ uM/gTOC}$	12	7 (58%)
Samples with $\sum_{23}\text{PAH} > 11 \text{ uM/gTOC}$	25	12 (48%)
Samples with $\sum_{23}\text{PAH} > 3.5 \text{ uM/gTOC}$	123	54 (44%)
	Total	Non-Toxic
All Samples	2475	2087 (84%)
Samples with no chemical concentration >ERL	730	697 (95%)
Samples with $\sum\text{TotM}^* < \text{AVS}$	1256	1117 (89%)

* $\sum\text{TotM}$ = sum of total Zn,Cu,Ni,Pb,Cd,Hg, and Ag extracted with HF