Applied Geochemistry 57 (2015) 73–84

Contents lists available at ScienceDirect

Applied Geochemistry

journal homepage: www.elsevier.com/locate/apgeochem

Review

Characterizing toxicity of metal-contaminated sediments from mining areas

John M. Besser *, William G. Brumbaugh, Christopher G. Ingersoll

U.S. Geological Survey, Columbia Environmental Research Center, Columbia, MO, United States

ARTICLE INFO

Article history:
Available online 21 June 2014

ABSTRACT

This paper reviews methods for testing the toxicity of metals associated with freshwater sediments, linking toxic effects with metal exposure and bioavailability, and developing sediment quality guidelines. The most broadly applicable approach for characterizing metal toxicity is whole-sediment toxicity testing, which attempts to simulate natural exposure conditions in the laboratory. Standard methods for whole-sediment testing can be adapted to test a wide variety of taxa. Chronic sediment tests that characterize effects on multiple endpoints (e.g., survival, growth, and reproduction) can be highly sensitive indicators of adverse effects on resident invertebrate taxa. Methods for testing of aqueous phases (pore water, overlying water, or elutriates) are used less frequently. Analysis of sediment toxicity data focuses on statistical comparisons between responses in sediments from the study area and responses in one or more uncontaminated reference sediments. For large or complex study areas, a greater number of reference sediments is recommended to reliably define the normal range of responses in uncontaminated sediments – the 'reference envelope'. Data on metal concentrations and effects on test organisms across a gradient of contamination may allow development of concentration-response models, which estimate metal concentrations associated with specified levels of toxic effects (e.g. 20% effect concentration or EC20). Comparisons of toxic effects in laboratory tests with measures of impacts on resident benthic invertebrate communities can help document causal relationships between metal contamination and biological effects. Total or total-recoverable metal concentrations in sediments are the most common measure of metal contamination in sediments, but metal concentrations in labile sediment fractions (e.g., determined as part of selective sediment extraction protocols) may better represent metal bioavailability. Metals released by the weak-acid extraction of acid-volatile sulfide (AVS), termed simultaneously-extracted metals (SEM), are widely used to estimate the 'potentially-bioavailable' fraction of metals that is not bound to sulfides (i.e., SEM-AVS). Metal concentrations in pore water are widely considered to be direct measures of metal bioavailability, and predictions of toxicity based on pore-water metal concentrations may be further improved by modeling interactions of metals with other pore-water constituents using Biotic Ligand Models. Data from sediment toxicity tests and metal analyses has provided the basis for development of sediment quality guidelines, which estimate thresholds for toxicity of metals in sediments. Empirical guidelines such as Probable Effects Concentrations or (PECs) are based on associations between sediment metal concentrations and occurrence of toxic effects in large datasets. PECs do not model bioavailable metals, but they can be used to estimate the toxicity of metal mixtures using by calculation of probable effect quotients (PEQ = sediment metal concentration/PEC). In contrast, mechanistic guidelines, such as Equilibrium Partitioning Sediment Benchmarks (ESBs) attempt to predict both bioavailability and mixture toxicity. Application of these simple bioavailability models requires more extensive chemical characterization of sediments or pore water, compared to empirical guidelines, but may provide more reliable estimates of metal toxicity across a wide range of sediment types.

* Corresponding author. Tel.: +1 5738761818.
E-mail address: jbesser@usgs.gov (J.M. Besser).

http://dx.doi.org/10.1016/j.apgeochem.2014.05.021
0883-2927/Published by Elsevier Ltd.

Contents

1. Introduction

In receiving waters with circumneutral pH, cationic metals derived from mining, ore processing, and smelting activities tend to have low solubility and strong affinity for solid phases. These properties lead to loss of metals from the aqueous phase via precipitation and sorption, thereby reducing metal toxicity to fish and other aquatic organisms in the water column. However, metals associated with precipitates and suspended particles may accumulate to high concentrations in bed sediments, leading to elevated metal exposure and potential toxic effects on benthic organisms, principally benthic macroinvertebrates. Toxic effects of metal-contaminated sediments on benthic (sediment-dwelling) invertebrates can result in reduced invertebrate productivity and/or loss of metal-sensitive taxa (e.g., Canfield et al., 1994). In addition, accumulation of high levels of metals in metal-tolerant invertebrates may pose risks of dietary toxicity to fish or other predators (Cain et al., 2004). Because these effects of metal-contaminated sediments on biota may extend far downstream of metal sources, characterizing the severity and areal extent of the effects can be an important component of regulatory monitoring, site assessments, and ecological restoration. Field studies of impacts on benthic communities can document these impacts, but they often cannot establish causal relationships between metal exposure (via water, diet, and sediment) and observed community impacts. In contrast, controlled toxicity and bioaccumulation studies conducted in the laboratory can provide unambiguous evidence linking metal exposure with adverse effects on benthic organisms.

Sediment toxicity tests can provide scientific evidence for the occurrence of adverse effects on benthic organisms to meet regulatory and legal requirements and to support management decisions. In the U.S., sediment toxicity testing is often used to characterize metal exposure and effects at contaminated sites under the CERCLA (Superfund) program and to document injury to aquatic resources under the Natural Resource Damage Assessment and Remediation (NRDAR) program. Sediment toxicity testing may also be required by regulations intended to prevent ecological effects from effluents from mining or ore-processing facilities or from dredging and disposal of contaminated sediments in navigable waterways. Toxicity testing of sediments – both contaminated sediments collected in the field and sediments spiked with metals in the laboratory – is also an essential part of the process of deriving and validating sediment quality guidelines (e.g., MacDonald et al., 2000; USEPA, 2005; Vangheluwe et al., 2013).

The objective of this paper is to review methodologies for (1) testing the toxicity of metals associated with freshwater sediments, (2) characterizing associations of toxicity with measures of metal bioavailability; and (3) using these associations as the basis for sediment quality guidelines. Discussion of these topics will be supported by examples of laboratory and field studies of sediment toxicity in areas affected by past and present mining, ore-processing, and smelting activities.

2. Sediment toxicity testing

2.1. Types of test methods

Whole-sediment testing is the most widely-used and versatile approach for characterizing the bioavailability and toxicity of sediment-associated contaminants.

Standard methods for whole-sediment toxicity testing have been published by ASTM (2010a), USEPA (2000), and OECD (2004, 2007). Testing of whole sediment (solid-phase plus pore water) is intended to provide a relatively complete and realistic simulation of the exposure regime experienced by benthic organisms inhabiting surficial sediments. In contrast, 'whole-sediment' tests with water column organisms (e.g., the cladocerans, *Daphnia magna* and *Ceriodaphnia dubia*; ASTM, 2010a) primarily reflect exposure to contaminants in overlying water.

Whole sediment tests are usually conducted with composites of multiple grab samples of surficial sediments, although tests can also be conducted with intact sediment cores (or core sections) or with fine sediment fractions separated from a matrix of coarse substrata. Core sampling allows toxicity testing and chemical analyses of sediment from defined depth strata (e.g. Besser et al., 1996; Ingersoll et al., 2002), but requires substantially greater sampling effort. Sampling in locations where coarse sediments make up a large or variable proportion of bed sediments may require wet sieving with site water (e.g., sieving to <2 mm particle diameter) to ensure collection of a more homogeneous fine sediment fraction, which is more compatible with toxicity test methods and which tends to contain greater concentrations of biologically-available metals (Besser et al., 2009a). For testing with some small invertebrates, further sieving may be necessary to obtain fine sediments for testing (e.g., sieving to <0.25 mm for testing with juvenile mussels; Ingersoll et al., 2008). Although both sampling and sieving inevitably disrupts equilibria between sediments and pore water, the impact of these changes can be reduced by using site water for sieving and by allowing sieved sediments to re-equilibrate in test chambers for at least a week before testing.

Whole-sediment toxicity tests can be conducted with water collected from the site of sediment collection (e.g., Besser and Leib, 2007), but tests are usually conducted with well-characterized uncontaminated test waters, in order to focus on metal exposure via sediment and pore water. Test waters may be well water, filtered pond water, diluted or fortified waters, or formulated (reconstituted) water prepared by adding salts to de-ionized water (e.g., USEPA, 2000; ASTM et al., 2010b). It may be desirable to select or modify the characteristics of test waters (e.g., pH, conductivity/salinity, hardness, and pH/alkalinity) to be similar to water quality at the sediment collection site. An important consideration for whole-sediment tests is the volume and/or replacement rate of overlying water during tests. Dilution or flushing of dissolved constituents in overlying water is usually necessary to maintain desired ranges of water quality (e.g., pH, dissolved oxygen), to remove potentially toxic metabolic wastes (e.g., ammonia), and to avoid accumulation of unrealistic concentrations of metals. In static tests (i.e., with no replacement of overlying water, these issues may be adequately addressed by increasing the water:sediment ratio (e.g., Borgmann et al., 2001), but most laboratories maintain overlying water quality by conducting tests with manual or automated replacement of overlying water (e.g., Zumwalt et al., 1994). Water-replacement test systems may use a proportional diluter (designed for water-only toxicity tests) to deliver test water, which is split equally among replicate exposure chambers (e.g. Supplemental Fig. S1). Standard test methods recommend minimum rates of replacement of overlying water (e.g. 2 volume-additions/day) (USEPA, 2000; ASTM, 2010a), but higher rates may be necessary to maintain overlying water quality (e.g., Brumbaugh et al., 2013).

Toxicity tests conducted with pore water isolated from sediment can complement whole-sediment testing. Methods for pore-water testing are typically based on standard methods for acute water-only toxicity testing (ASTM et al., 2010b) or for effluent toxicity testing (USEPA, 2002). Pore-water testing requires isolation of large volumes of pore water, either in the field or the laboratory, using methods such as push-point sampling, pressure/squeezing, or centrifugation (USEPA, 2001; ASTM et al., 2010c). Results of pore-water toxicity tests are of interest because exposure to pore-water metals is considered to be a dominant route of metal exposure for sediment-dwelling organisms (Ankley et al., 1996). Another desirable aspect of pore-water testing is the ability to conduct tests with organisms that live in the water column (e.g., planktonic invertebrates and fish), which may have extensive records of aquatic toxicity data and may be highly sensitive to metal toxicity. However, results of pore-water tests cannot be considered equivalent to evidence of the toxicity of whole sediment, due to the physicochemical changes that occur in pore waters that are no longer in contact with metal-rich sediment particles.

Elutriate testing is a specialized method that characterizes the potential for release of toxic substances from resuspended sediments. Elutriate tests are conducted with water-column organisms in the aqueous fraction of a water:sediment suspension (USEPA and USACE, 1998; ASTM et al., 2010c) and are intended to simulate toxic effects in the water-column, not in intact sediments. Elutriate testing is most commonly used for assessment of environmental effects of dredging for navigation, but the method can also be applied to dredging for remediation of contaminated sites or to simulate resuspension of sediments during high-flow events. Aqueous-phase tests can be conducted with solutions prepared by leaching of mine wastes, which can be used to evaluate the potential for release of bioavailable metals (e.g. Smith et al., 2000). The toxicity of the aqueous phase of a sediment-water suspension can also be tested using the luminescent bacterium, *Vibrio fischeri*, (Microtox® Strategic Diagnostics, Inc., Newark, DE, USA) (Environment Canada, 2002). This test is amenable for rapid (10-min) toxicity screening of large numbers of soil or sediment samples, but it is not highly sensitive to metal toxicity nor predictive of toxicity in invertebrates or vertebrates (e.g., Kemble et al., 1994).

More specialized methods are required to document causal relationships between metals and sediment toxicity, or to estimate the concentrations of individual metals required to cause toxicity. Toxicity Identification Evaluation methods (TIE; USEPA, 2007a) document toxicity caused by metals relative to other classes of contaminants by a series of manipulations intended to selectively remove or neutralize their toxic action. TIE procedures developed for surface water – which can be applied to pore water – selectively remove the toxicity of cationic metals by using pH manipulation and complexing agents such as EDTA or sodium thiosulfate (Hockett and Mount, 1996). USEPA (2007a) has also published guidance for TIE procedures that can be used to manipulate metal toxicity in whole sediments (e.g. by addition of cation exchange resins, AVS, or base metals).

Sediment spiking studies seek to quantify toxic effects of known metal additions (ASTM, 2010a; USEPA, 2001). In principle, the toxicity of sediments spiked with known concentrations of a metal or metal mixture (added as highly-dissociable salts that allow the metal cations to interact with sediment binding sites) can provide a basis for estimating the contribution of metal(s) to toxicity of sediments collected in the field (e.g., Milani et al., 2003) or for establishing sediment quality guidelines for individual metals (Besser et al., 2013; Vangheluwe et al., 2013). The environmental realism of toxicity tests with metal-spiked sediments depends on many factors, including the metal-binding characteristics of the sediment(s), the methods used to avoid artifacts of metal spiking (e.g., pH control), and the duration of the post-spiking equilibration period (Brumbaugh et al., 2013; Hutchins et al., 2007, 2008).

2.2. Test organisms

Selection of appropriate test organisms is an important consideration in sediment toxicity testing. ASTM (2010a) and USEPA (2000) identified factors relevant to selection of standard sediment test organisms, including physiological characteristics (sensitivity to contaminants, tolerance to a range of sediment characteristics), ecological relevance (geographic distribution, degree of contact with sediment, representativeness of local/regional benthic communities), and practical considerations (amenability to laboratory culture, performance in inter-laboratory comparisons). Relatively few freshwater invertebrates met enough of these criteria to become standard sediment test organisms, most notably the amphipod, *Hyalella azteca* (Crustacea) and two species of midges (Diptera), *Chironomus dilutus* (formerly *C. tentans*) and *Chironomus riparius* (ASTM, 2010a; USEPA, 2000; OECD, 2004) (Supplemental Fig. S2). These species are readily available from ecotoxicology laboratories or aquaculture supply houses, and self-sustaining cultures are relatively easy to maintain. These three species have been used for a large proportion of whole-sediment toxicity tests conducted in North America in the past two decades, including numerous tests with metal-contaminated sediments from mining areas (e.g., Ingersoll et al., 1996, 2002; Besser et al., 2008; Seal et al., 2010). Each of these standard test organisms have been shown to respond reliably in sediments with a wide range of contaminant types and physicochemical characteristics. Although *Hyalella* is generally more sensitive to metal toxicity than *Chironomus* (e.g., Phipps et al., 1995; Besser et al., 2013) midge larvae have proved to be equally sensitive or more sensitive to particular metals or metal mixtures (e.g., Besser et al., 2008; Seal et al., 2010). Because differences in sensitivity to a particular metal mixture among taxa and among sediment types are usually not predicable (without evidence from previous studies), most sediment toxicity studies should include tests with at least two unrelated taxa to increase the likelihood of detecting meaningful differences among sites.

Depending on the nature of the study area and study objectives, it may be desirable to conduct sediment toxicity tests with a broader range of test organisms (Supplemental Fig. S2). Recent sediment toxicity tests with eight species of benthic invertebrates spiked with nickel demonstrated a wide range of sensitivity among species (Fig. 1; Besser et al., 2013). Testing with alternative test species can provide valuable toxicity information for taxa that are ecologically important, that occur locally, or that are of special conservation concern. ASTM (2010a) presents guidance for conducting toxicity tests with several alternative invertebrate species. The oligochaete worms, *Lumbriculus variegatus* and *Tubifex tubifex*, represent a group that often dominates invertebrate communities in fine, organic-rich sediments. Sediment testing has also been conducted with amphipods other than *Hyalella*, including both widespread taxa (e.g., *Gammarus* spp.) and taxa of regional importance (e.g., *Diporeia* spp. in the Great Lakes; Giesy and Hoke, 1989; Burton et al., 1996). The burrowing mayflies of the genus *Hexagenia* represents a taxonomic group (Ephemeroptera) that is highly responsive to degradation of water and sediment quality (Giesy et al., 1990; Cain et al., 2004). Guidance for standardized sediment testing has not yet been published for freshwater mussels (Unionoida), a group with a high percentage of species that are endangered or threatened (Williams et al., 1993). However, guidance is available for conducting water-only toxicity tests with mussels (ASTM et al., 2010d), and recent studies have shown that juvenile fatmucket mussels (*Lampsilis siliquoidea*) are adversely affected by exposure to fine sediments from lead- and zinc-mining districts in Missouri (Ingersoll et al., 2008). Toxicity tests with sediments from the Big River in Missouri's Old Lead Belt demonstrated toxic effects on juvenile mussels in sediments collected downstream of the mining area, consistent with elevated metal concentrations and

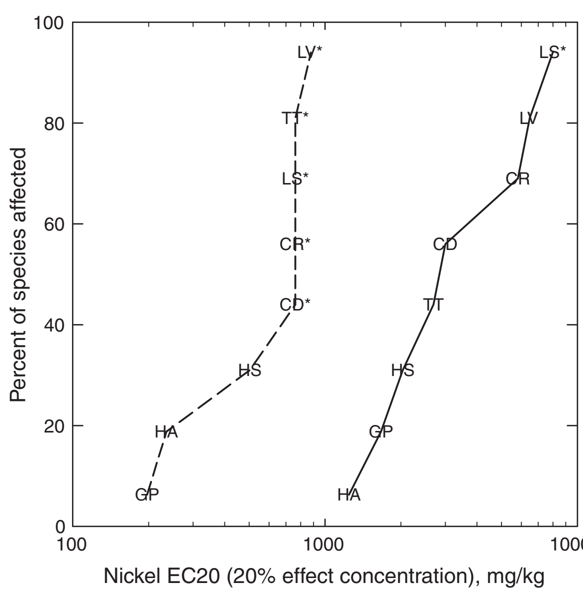

Fig. 1. Sensitivity of benthic invertebrate species to nickel-spiked sediments. Distributions of 20% effect concentrations (EC20s) for nickel toxicity to eight species in two sediments: Spring River, MO (left/dashed line) and West Bearskin Lake, MN (right/solid line). GP = *Gammarus* (amphipod), HA = *Hyalella* (amphipod), HS = *Hexagenia* (mayfly), CD = *Chironomus dilutus* (midge), CR = *Chironomus riparius* (midge), LV = *Lumbriculus* (oligochaete), TT = *Tubifex* (oligochaete), LS = *Lampsilis* (mussel). Asterisks indicate no effect at highest concentration tested (Besser et al., 2013).

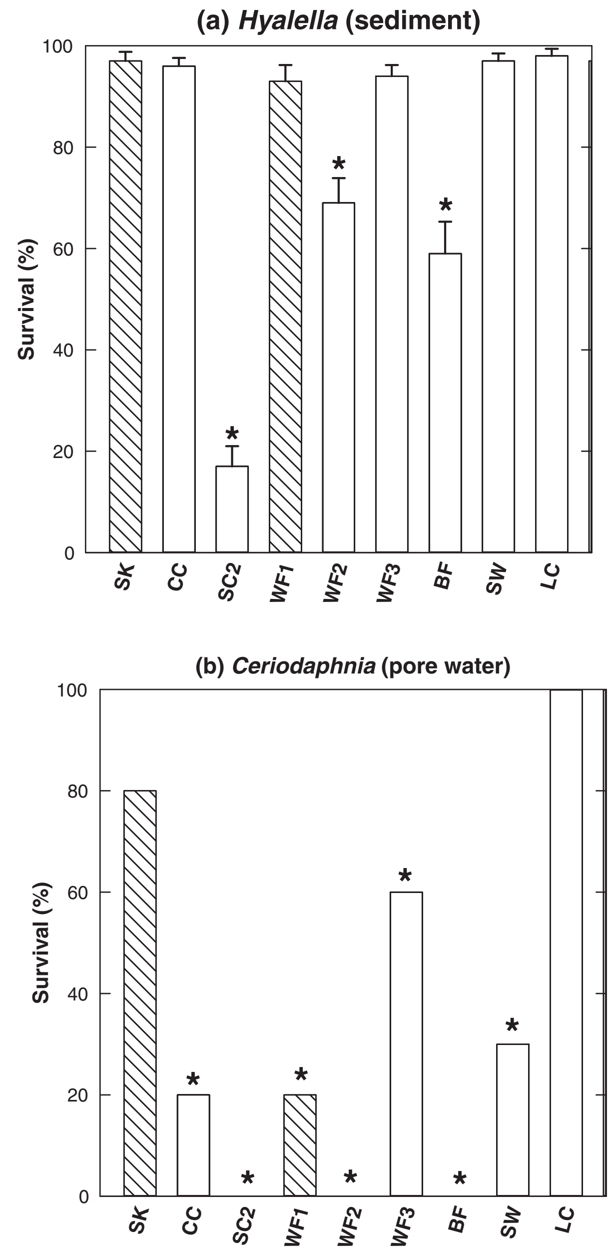

Fig. 2. Toxicity of sediment and pore waters from the Viburnum Trend mining district, MO: (a) whole-sediment tests with *Hyalella*; (b) pore-water tests with *Ceriodaphnia*. Bars are means (with standard errors); hatched bars are reference sites. Asterisks indicates significant difference from reference (or control for *Ceriodaphnia*) (Besser et al., 2009a).

greatest impacts on a wild mussel community that includes endangered species (Besser et al., 2009b).

Selection of test organisms for pore-water toxicity tests has been based primarily on known sensitivity to contaminants and limitations of testing with small sample volumes. The most common pore-water test organisms are cladocerans ("water fleas"), especially *Ceriodaphnia dubia*. *Ceriodaphnia* tests are attractive because a full life cycle test can be conducted with a small volume (approx. 1 L) of pore water in about 1 week, using a standard method developed for effluent toxicity testing (USEPA, 2002). Although *Ceriodaphnia* is sensitive to toxicity of metals, interpretation of pore water tests with water-column organism may be difficult because they may not tolerate elevated levels of constituents such as iron, manganese, and ammonia that occur frequently in

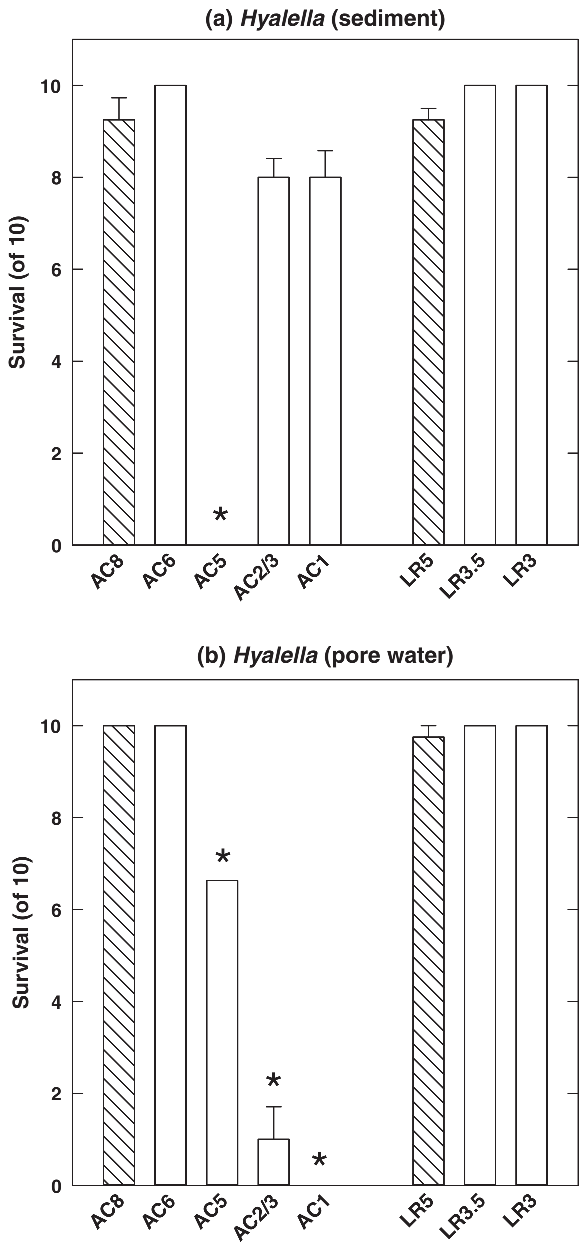

Fig. 3. Toxicity of sediments and pore waters from streams draining the Palmerton zinc smelter site, PA, to *Hyalella*: (a) sediment tests; (b) pore-water tests. Bars are means (with standard errors); hatched bars are reference sites on Aquashicola Creek (AC) and Lehigh River (LR). Asterisk = significant difference from reference (Besser et al., 2010).

pore water and may be tolerated by sediment-dwelling taxa. For example, a study of sediments and pore water from streams of the Viburnum Trend mining district found limited toxicity to *Hyalella* in whole-sediment tests, but severe toxicity of pore water to *Ceriodaphnia* at many sites (Besser et al., 2009a; Fig. 2). Although the greater response of *Ceriodaphnia* to pore waters from mining-impacted sites was consistent with its greater sensitivity to metals, this response was confounded by toxicity at one or more sites with no known mining impacts, possibly due to ammonia toxicity. Another standard effluent test method, a 7-day exposure with larval fathead minnows (*Pimephales promelas*; USEPA, 2002) is rarely used for pore-water testing because it requires greater volumes of pore water. Other fish taxa that have eggs and larvae that occur in benthic habitats (e.g., trout, sculpins, and sturgeon) may be appropriate test species for pore-water or sediment toxicity testing.

Some benthic taxa used for whole-sediment testing are also adaptable to pore-water testing. *Hyalella* performs well in water-only tests if it is provided with an inert substratum such as silica sand or nylon mesh. Thus, toxic effects on *Hyalella* can be directly compared between whole-sediment tests and pore-water (or surface-water) tests. In a recent study of streams draining a zinc smelter site in Palmerton, Pennsylvania (Besser et al., 2010), toxicity to *Hyalella* was greater in tests with pore water (and stream water; data not shown), compared to whole-sediment tests (Fig. 3), suggesting that impacts on benthic communities at this site were caused primarily by dissolved metals entering the stream via contaminated ground water, rather than by metal-contaminated sediments.

2.3. Acute and chronic tests and endpoints

Sediment and pore-water toxicity tests can vary in duration and complexity, depending on study objectives and resources. Acute (short-term) tests typically rely on the endpoint of lethality and are most useful for screening large numbers of samples. Acute lethality tests with pore-water or elutriates typically last 2–4 days and lethality is the endpoint most commonly reported in 10-d whole-sediment tests (Ingersoll et al., 1998; USEPA, 2000; ASTM, 2010a). Chronic (long-term) sediment toxicity tests, which typically last 4 weeks or longer, allow measurement of one or more sublethal endpoints, which are often much more sensitive than survival (Ingersoll et al., 1998). Growth in length or weight is often a sensitive endpoint that can reflect biochemical and physiological stresses and/or behavioral changes (e.g., avoidance or reduced feeding) caused by toxic substances. Total biomass of survivors from replicate groups (e.g., total ash-free dry weight) tends to be a sensitive endpoint because it integrates effects on both survival and growth, and it serves as a surrogate for biomass production by invertebrate populations in the field.

Reproduction is often a highly-sensitive endpoint, but reproduction tests require longer and more complex test protocols and labor-intensive sample processing. Sublethal endpoints tend to have greater variability than lethality (Ingersoll et al., 1998), due to such factors as differences in sex ratio of test organisms (may influence reproduction); differences in organic matter content or nutrient richness (may influence growth and reproduction); and measurement error during weighing of low-mass samples. For example, *Hyalella* reproduction showed both greater sensitivity and greater variation, compared to survival, in sediments from Missouri's Viburnum Trend mining district (Besser et al., 2009a). Interpretation of reproduction data was hindered by variation between sediment types (river vs. reservoir), variation among sites within each sediment type (including variation among uncontaminated reference sites), and variation among experimental replicates.

USEPA (2000) and ASTM et al. (2010e) developed a method for a chronic whole sediment bioaccumulation test with *Lumbriculus*, a species that can typically survive exposure to high concentrations of metals and other contaminants. The bioaccumulation endpoint does not provide direct evidence of toxicity, but it can provide information about metal bioavailability and potential toxicity of metal-contaminated invertebrate diets to fish or other predators (e.g., Woodward et al., 1995; Mount et al., 2006).

2.4. Control and reference sediments

Characterizing the occurrence and severity of sediment toxicity requires comparing responses of test organisms in sediments from a study area (test sediments) to responses in uncontaminated sediments – either control or reference sediments – that have comparable physicochemical characteristics. Control sediments are

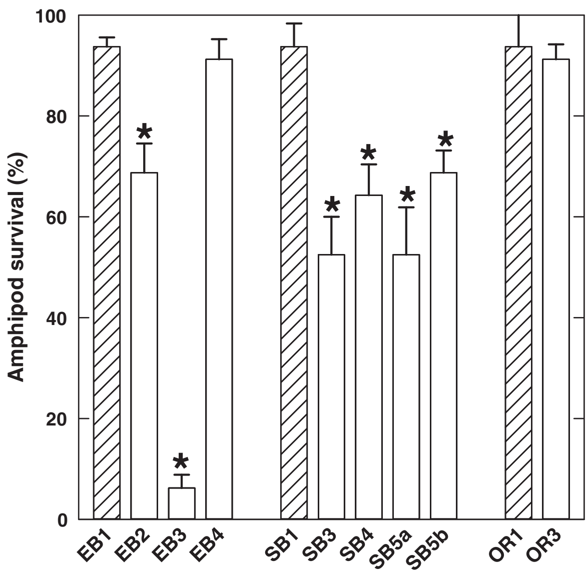

Fig. 4. Toxicity of sediment from streams draining the Ely Mine site, Vermont, to *Hyalella*. Bars are means (with standard errors); hatched bars are upstream reference sites in Ely Brook (EB), Schoolhouse Brook (SB), and Ompompanoosuc River (OR). Asterisk = significant difference from reference site. (Seal et al., 2010).

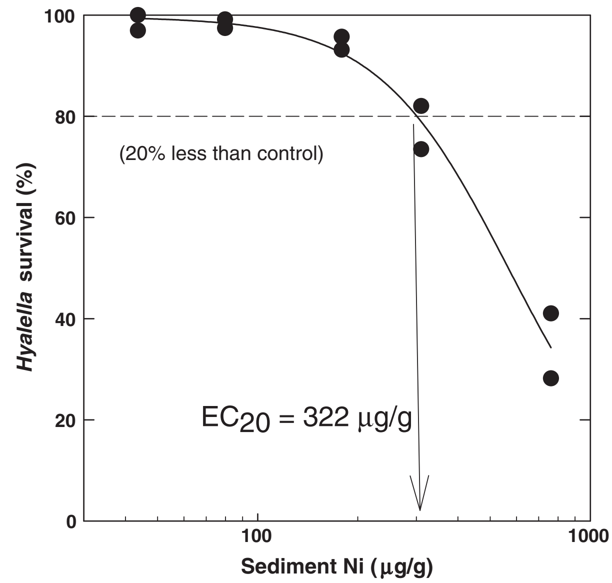

Fig. 5. Concentration-response curve for toxicity to of nickel to *Hyalella* in spiked sediments from Spring River, MO. Solid line is logistic regression; EC20 = concentration causing 20% effect. (Besser et al., 2013).

uncontaminated sediments that have been extensively characterized and are known to provide an adequate substratum for test organisms. They may be field-collected aquatic sediments, wetted soils, or formulated sediments prepared in the laboratory from commercially-available components. Control sediments are rarely used for statistical comparisons with test sediments because of the difficulty of matching physicochemical characteristics test sediments. Control sediments are typically used to document adequate performance (survival, growth) of a particular batch of test organisms in the chosen exposure system, based on specifications of standard methods (e.g., USEPA, 2000; ASTM, 2010a) or laboratory control charts. Reference sediments are field-collected sediments – ideally collected from multiple sites in or near the study area – that are documented to be substantially free of the contaminants of interest and that closely reflect the range of characteristics of sediments (e.g., particle size, organic matter, mineralogy) and study sites (e.g., water quality, stream order, nutrient enrichment) from the study area. The performance of test organisms in one or more reference sediments typically provides the basis for determining whether test responses represent toxic effects (i.e., fall below the normal range for uncontaminated sediments in the study area). The range of responses in reference sites is sometimes termed the 'reference envelope' (e.g., Hunt et al., 2001). In a simple field study, such as a small, homogeneous drainage with known contaminant sources, a single upstream reference site may be adequate. In a study of the Ely Mine site in Vermont (Seal et al., 2010), where metal contamination from a localized source entered a small headwater stream, then flowed into a larger stream and a small river, the toxicity of test sediments from each stream segment was determined by comparisons to a single upstream reference site in each stream segment (Fig. 4). This simple sampling design allowed us to document attenuation of contaminant impacts despite downstream changes in sediment and stream characteristics. For larger and more complex study areas, multiple reference sites are needed to adequately define a reference envelope. In the Old Lead Belt study (Besser et al., 2009b), where metal contamination was dispersed from headwaters to mouth along a 162-km reach of the Big River, multiple reference sites were selected in the Big River and nearby drainages to represent the range of stream gradient, particle size, and nutrient richness in the Big River study sites. A similar approach was taken in the Viburnum Trend study (Besser et al., 2009a), where three reference sites were selected to represent headwater, mid-river, and reservoir study sites within the Black River watershed. Both studies demonstrated variation of test endpoints among reference sites, which was more pronounced for growth and reproduction than for survival. Larger, more complex studies typically require larger numbers of reference sites to adequately define the reference envelope. A study of the toxicity of sediments from the Tri-State (Missouri/Kansas/Oklahoma) mining district included 15 reference sites, more than 20% of the total number of sites studied (Ingersoll et al., 2008).

2.5. Analysis of toxicity data

Statistical approaches for comparisons with control or reference sediments vary depending on the complexity of the study design and the objectives of the study. Conventional statistical analyses of data from toxicity tests with field-collected sediments (USEPA, 2000; ASTM, 2010a) rely on analysis of variance, combined with pairwise comparisons of means between test sites and means from a single reference or control sample (e.g., Dunnett's test) or multiple comparisons of means among study and reference sites (e.g., Tukey's test). In studies with large numbers of reference samples, study sites may be designated as toxic if they fall below a specified quantile of the distribution of responses in the reference samples (e.g., the 10th percentile; Hunt et al., 2001; Ingersoll et al., 2008). In the Big River study (Besser et al., 2009b), which had only four reference sites, all sample means below the range of reference means were designated as toxic.

For some sediment tests, responses of test organisms can be further analyzed using concentration-response models, which can characterize the responses of toxicity endpoints to a gradient of chemical concentrations. For the survival/lethality endpoint, which follows a binomial distribution (i.e., live or dead), the classic approach is to fit the data to a log-normal probability distribution. Variables that follow continuous distributions, including most sublethal endpoints, are typically analyzed by non-linear regression (e.g., logistic regression). Concentration-response models use

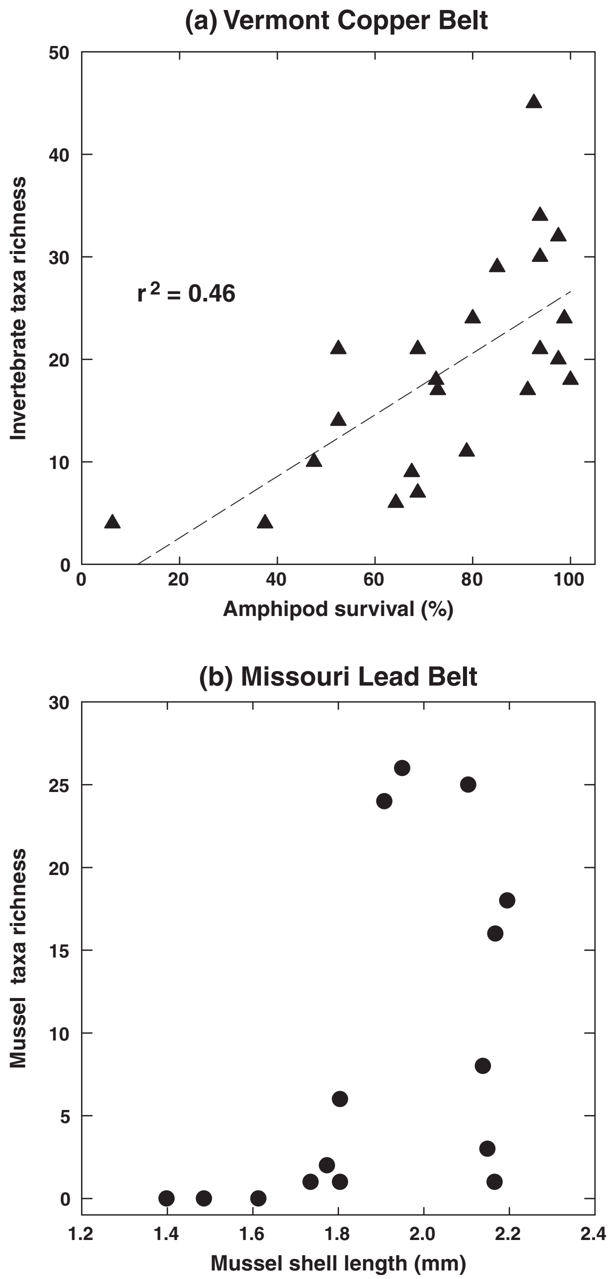

Fig. 6. Correspondence of sediment toxicity endpoints (laboratory data) and characteristics of benthic communities (field data): (a) amphipod survival and number of invertebrate taxa in Vermont Copper Belt; (b) mussel shell length and mussel taxa richness in Old Lead Belt, MO. Dashed line is linear regression. (Seal et al., 2010; Piatak et al., 2013; Besser et al., 2009b).

information from the full range of toxic responses to level of chemical exposure associated with a specified increment toxic effect, such as the median lethal concentration or LC50 (concentration causing 50% mortality) or the 20% effect concentration or EC_{20}. These models are often used to characterize effects of a single chemical in water or in spiked sediments (e.g., Fig. 5), but they can also be used to characterize toxicity along a gradient of metal contamination in field-collected sediments, provided sediment characteristics and the mixture of metals present are similar among sites. In such cases, models based on concentrations of a predominant toxic metal or on an index of toxicity hazards for the metal mixture (discussed below), may allow estimation of site-specific thresholds for toxicity of metals in sediment or pore water (e.g., MacDonald et al., 2009).

Comparisons of toxicity data, chemistry data, and data from field surveys of benthic communities in the field, sometimes called the 'sediment quality triad' (e.g., Canfield et al., 1994), provide a greater weight of evidence that can document causal relationships between chemical contamination and biological effects. Collection and analysis of benthic community data can be expensive and time-consuming and community responses can be highly variable – often due to influences other than current levels of sediment contamination, such as alteration of physical habitat or historical impacts (e.g., slow re-colonization). Despite these limitations, recent studies in the Vermont Copper Belt (Seal et al., 2010) and the Old Lead Belt (Besser et al., 2009b) showed strong associations between laboratory toxicity data and field data on invertebrate community characteristics (Fig. 6a and b). Both of these studies found consistent, severe impacts on invertebrate taxa richness at sites with sediments that were toxic in laboratory tests. In both examples, taxa richness varied widely at sites where sediments exhibited little or no toxicity, suggesting that some impacts occurring in the field may not be adequately reflected in laboratory tests. For example, low taxa richness of mussels at sites with little or no sediment toxicity (Fig. 6b) may reflect long-term toxic effects on these long-lived organisms that are not evident during 28-day laboratory tests.

3. Characterization of metal exposure and bioavailability

3.1. Chemical characterization of sediment

Characterizing relationships between toxicity and metal concentrations requires appropriate methods for sampling and analysis of metals and other constituents in sediment and pore water. Analyses of metal concentrations solubilized from sediments by different chemical extractants can provide a wide range of information about metal bioavailability. Methods for determination of true total metal concentrations in sediments, which include dissolution of residual metals (i.e., metals released during dissolution of crystalline silicate minerals using hydrofluoric acid; Filgueiras et al., 2002) are of limited use for characterizing bioavailability. Measurements of 'total recoverable' metals (e.g., microwave digestion with nitric acid, hydrochloric acid and/or hydrogen peroxide) do not recover residual crystalline metals and often provide a more useful conservative (high) estimate of potentially bioavailable metals (Brumbaugh et al., 2007). More elaborate protocols for selective sediment extractions of metals (Tessier et al., 1984; Filgueiras et al., 2002), which are intended to quantify metal fractions associated with different chemical phases (e.g., exchangeable, bound to Fe/Mn oxides, bound to sulfides/organic matter, and residual) may also provide useful information about metal bioavailability. In a study of the toxicity of sediments from Lake Roosevelt, a reservoir on the Upper Columbia River in Washington, total-recoverable metal concentrations had a weak relationship with toxic effects on midge larvae (Besser et al., 2008). However, sediments with greater proportions of metals in labile sediment fractions, determined by selective sediment extractions (Paulson and Cox, 2007), caused reduced growth of midge larvae, despite relatively low total-recoverable metal concentrations. Although selective sediment extraction protocols are rarely used in conjunction with sediment toxicity testing, it has long been recognized that metal concentrations determined from weak acid extracts (e.g., 1 N HCl) have stronger correlations with metal bioavailability than do total metal concentrations, although the strength of these correlations varies among sediment types (Luoma, 1989).

The 'equilibrium-partitioning' approach to characterizing metal bioavailability is based on simple models of the distribution of metals between sediment particles and pore water metals, which are assumed to be the most bioavailable fraction (Ankley et al., 1996; USEPA, 2005). The strongest binding component for cationic metals in many anoxic sediments is acid-volatile sulfide (AVS), which is operationally defined as the sulfide fraction released by a 1 N HCl extraction at room temperature (USEPA, 1991; Brumbaugh and Arms, 1996). The metal fraction released during the AVS extraction is termed simultaneously extracted metals (SEM). Because nickel, zinc, cadmium, lead, copper, and silver form highly-insoluble sulfides, equilibrium models predict that these metals should bind strongly to available AVS, leaving very low metal concentrations in pore water. Thus, sediments should be non-toxic if summed concentrations of these SEM metals (ΣSEM; subsequently referred to as SEM) that are less than or equal to the concentration of AVS (SEM-AVS $\leqslant 0$), on a molar basis (DiToro et al., 1992; Ankley et al., 1996). Conversely, if SEM-AVS is positive, the fraction of SEM metals in excess of the AVS binding capacity is considered to be 'potentially bioavailable', although these metals may be bound to other sediment components such as organic matter (Ankley et al., 1996). Several studies have demonstrated the ability of the SEM-AVS approach to explain variation in metal bioavailability among sediments (DiToro et al., 1992; USEPA, 2005). A recent study of the toxicity of nickel-spiked sediments (Besser et al., 2013) found that toxicity values (EC20s) for *Hyalella* in eight sediments with a wide range of AVS concentrations were more variable when EC20s were expressed as total-recoverable Ni than when expressed as SEM-AVS. Toxicity values expressed as pore-water EC20s were even less variable, consistent with the hypothesis that pore-water metals are a more direct measure of metal bioavailability.

3.2. Sampling and characterization of pore water

A wide variety of methods have been used to obtain samples of pore water for analysis and toxicity testing, include methods based on suction (e.g., push-point or airstone), pressure, centrifugation, and passive diffusion samplers or peepers (USEPA, 2001; ASTM et al., 2010c). Pore-water samples obtained by various methods have generally been found to be comparable (e.g., Carignan et al., 1985; Schults et al., 1992), especially if the method includes a filtration step with 0.45 μm or smaller pore diameter. Some methods (push-point, pressure, and centrifugation) are useful for collection of relatively large volumes of pore water for toxicity testing or for analyses that require larger sample volumes. Small-volume peeper samplers can be deployed directly in toxicity test chambers, allowing in-situ measurement of pore-water metal concentrations in the field and during toxicity tests (Brumbaugh et al., 2007, 2013).

Estimates of metal bioavailability in pore water may be further refined by use of specialized samplers or modeling. The 'diffusion gradient-thin film' sampler (DGT; Zhang et al.; 1995) measures fluxes of labile metals into an ion-exchange resin embedded in an immobilized membrane, and the sampling matrix can be cut into sections to allow characterizations of metal fluxes in small-scale (ca. 1 cm) vertical gradients in the sediment column. Fluxes estimated by DGT can be used to estimate ambient metal concentrations (Brumbaugh et al., 2007). Geochemical modeling of metal speciation in pore waters allows further characterization of bioavailable metal species (e.g., free metal ions and bioavailable complexes). Centrifuged and filtered pore water samples, DGT samples, and geochemical models were used to examine relationships between metal bioavailability of a metal mixture and toxicity in Viburnum Trend sediments (Brumbaugh et al., 2007; Besser et al., 2009b). Centrifuged pore water samples indicated similar concentrations of 'dissolved' lead and nickel in three Viburnum Trend pore waters, but a greater fraction of pore-water Ni was associated with DGT samplers (Fig. 7). Concentrations of both dissolved Ni (and modeled Ni^{2+}), but not dissolved Pb or Pb^{2+}, in pore water were strongly associated with toxic effects in whole-sediment tests, suggesting that DGT samplers were useful indicators of bioavailable metal fractions (Brumbaugh et al., 2007, 2013; Besser et al., 2009a).

Biotic ligand models (BLM; DiToro et al., 2001; Niyogi and Wood, 2004) further characterize the toxicity of aqueous metals in terms of binding of metal ions to 'biotic ligands' (metal-binding sites) on gills or other body surfaces. BLM models characterize metal binding at biotic ligand sites as affected by binding of competing cations (e.g., H^+, Ca^{2+}, Mg^{2+}) and by binding of metal ions to competing aqueous ligands (e.g., DOC and carbonates). The BLM approach has been successful for modeling acute metal toxicity in surface waters (e.g., Fig. 8; Wang et al., 2009), but it has been little used to date in sediment toxicity. One limitation to the use of BLMs for modeling toxicity of metals in pore water is that pore waters often contain high concentrations of dissolved constituents, which may fall outside the validated range of BLM models developed for surface waters.

4. Sediment quality guidelines

4.1. Empirical sediment quality guidelines

Sediment quality guidelines (SQGs; Wenning et al., 2005) are intended to estimate concentrations of metals or other contaminants in sediments that will not cause toxic effects on benthic invertebrates. Efforts to develop SQGs have utilized broad two approaches: (1) empirical guidelines, which are based on trends in large databases of sediment chemistry and biological effects data, and (2) mechanistic guidelines, which are based on equilibrium-partitioning models.

Empirical sediment guidelines are based on associations between the frequency of adverse biological effects (in field and laboratory studies) and measured concentrations of metals and

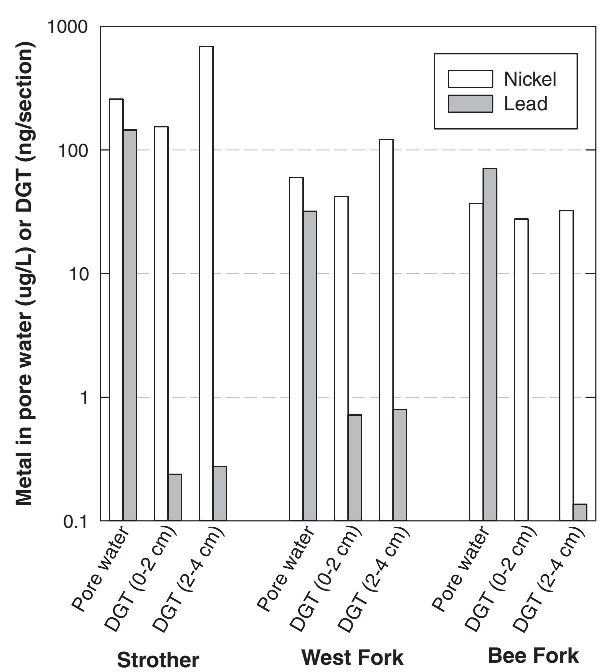

Fig. 7. Comparison of methods for nickel and lead from pore waters of sediments from three streams in the Viburnum Trend, MO. White bars = nickel, gray bars = lead. 'Pore water' indicates centrifuged/filtered samples; 'DGT' indicates sampling from two depth by diffusion-gradient thin film samplers (Brumbaugh et al., 2007).

other contaminants (Batley et al., 2005). The most widely used SQGs were developed by MacDonald et al. (2000), based on a 'consensus' of SQG values developed by several different methodologies. They defined the Threshold Effects Concentrations (TEC) as the concentration below which biological effects are not expected and the Probable Effects Concentration (PEC) as the concentration above which biological effects are expected to occur frequently. The primary advantage of PECs (and other empirical SQGs) is that they are supported by large databases of sediment toxicity (or alteration of invertebrate communities in community surveys) paired with measured sediment chemical concentrations. The main limitations of these databases are that they typically rely on measurements of total metal concentrations in sediments, rather than bioavailable fractions, and that observed toxic effects cannot be attributed to a single contaminant. Thus, the PEC for a single metal represents some combination of its inherent toxic potency, its 'average' bioavailability, and its 'average' contribution to the toxicity of mixtures of metals and other contaminants at many sites.

PECs have been widely used for screening sediments based on measured total metal concentrations. The simplest approach is to identify which metals exceed PECs and therefore may pose greater potential for toxicity. The predictive ability of PECs may be improved by quantifying these exceedances in the form of hazard quotients or 'PEC quotients' (here termed Probable Effect Quotients or PEQs), which are defined as the measured concentration of a metal divided by its PEC. PEQs allow comparisons of toxicity hazards among different metals and among sediments. For example, both sediment toxicity to freshwater mussels and reduced taxa richness of mussel communities occurred in sediments from the Old Lead Belt with PEQs for zinc greater than 1.0 (Fig. 9). If it is assumed that toxicity of metals is roughly additive, PEQs for multiple metals can be added (or averaged) to allow comparisons of cumulative toxicity hazards of metal mixtures among sites. These comparisons are probably most informative in local or regional studies, where controls on bioavailability and relative abundance of co-occurring metals are more likely to be consistent among sediments. In an analysis of a large dataset of freshwater sediment toxicity tests with *Hyalella* (623 tests) and *Chironomus* (599 tests), Ingersoll et al. (2001) reported increasing frequencies of toxic effects in sediments when mean metal PEQs exceeded 0.1 for amphipods or 0.5 for midges. Recent studies have also reported

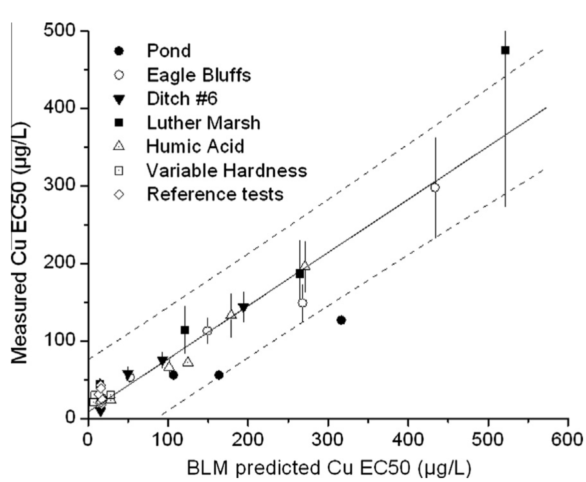

Fig. 8. Measured toxicity of aqueous copper to mussels and toxicity predicted by biotic ligand model (BLM) in various test waters. EC50 = concentration causing 50% reduction in shell length. Points represent results of single tests or means of multiple tests (with standard errors). Solid line indicates equal values; dashed lines indicate differences of a factor of two (Wang et al., 2009).

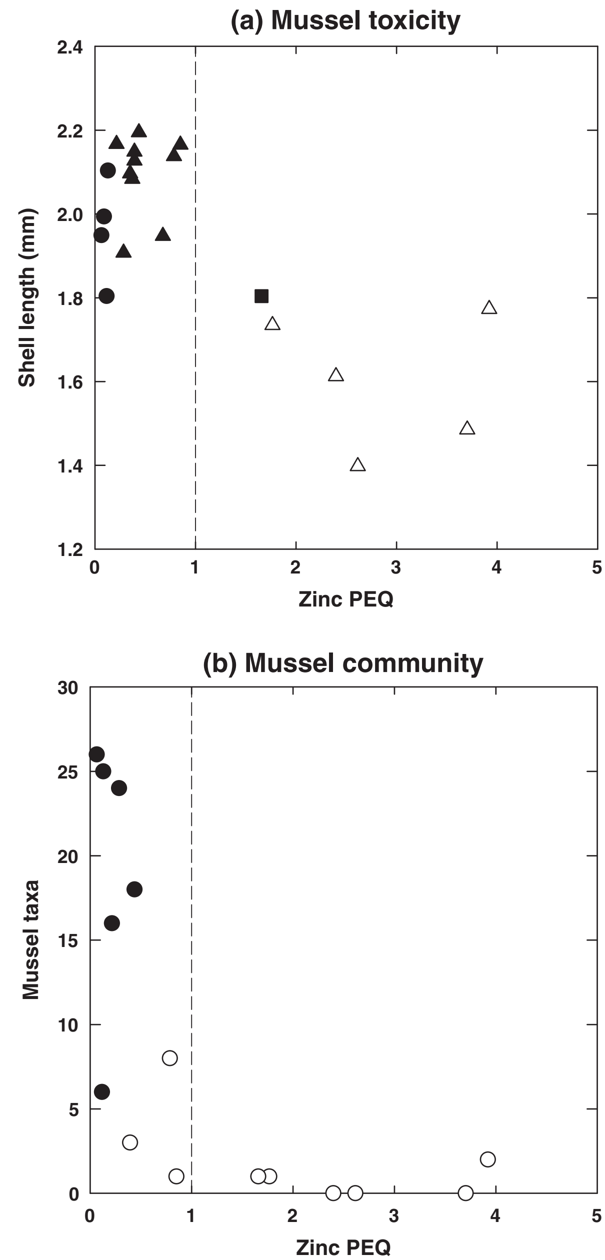

Fig. 9. Relationships between sediment zinc concentrations, expressed as probable effect quotients (PEQs), and adverse effects on freshwater mussels: (a) growth in laboratory toxicity tests; (b) taxa richness of resident mussel communities. Hollow symbols indicate reductions relative to reference conditions. (Besser et al., 2009b).

strong relationships between frequency of toxicity to *Hyalella* and summed PEQs for Pb, Zn and Cd in sediments from the Tri-State Mining District (Fig. 10a; MacDonald et al., 2009).

4.2. Equilibrium sediment benchmarks

Equilibrium-partitioning Sediment Benchmarks (ESBs; USEPA, 2005) were developed to provide guidance for interpreting toxicity hazards from mixtures of nickel, zinc, copper, silver, cadmium, and lead in pore-water or sediment. The pore-water ESB is based on a hazard-quotient model of the toxicity of metal mixtures, with the dissolved concentration of each metal in pore water divided by its chronic water quality criterion to derive 'toxic units' (TU; also called criteria units). Toxic units for multiple metals are summed to estimate of the toxicity risk of the mixture, with no

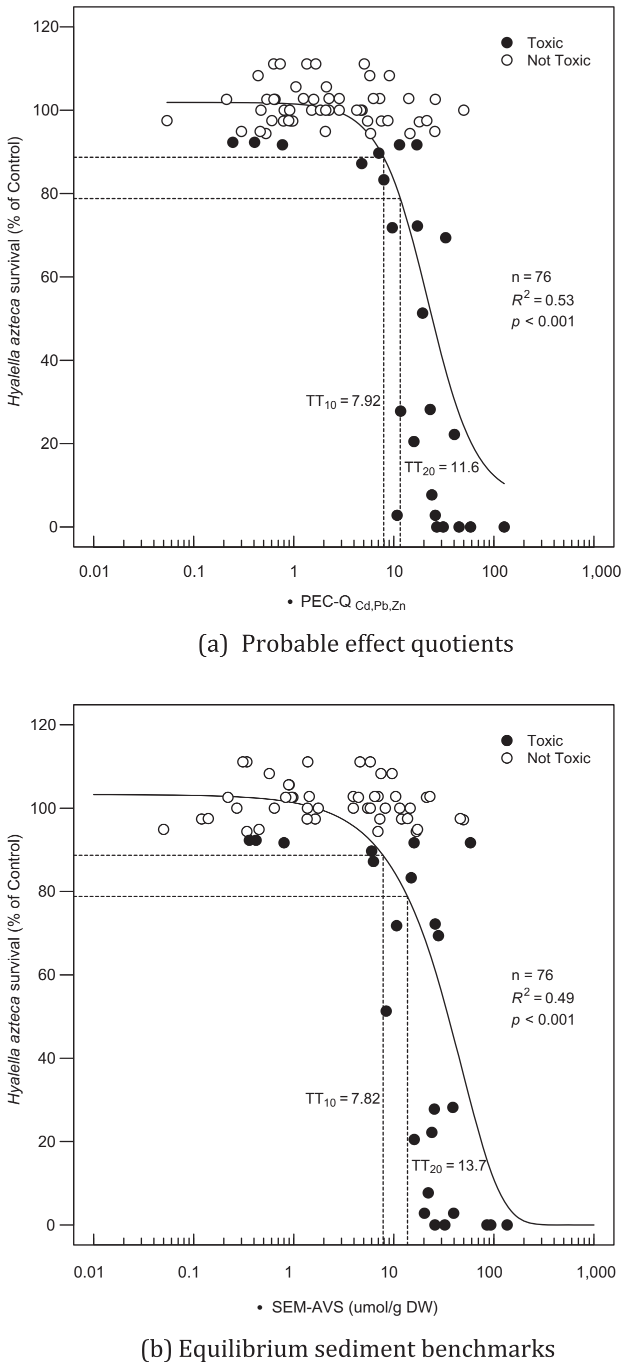

Fig. 10. Relationships of *Hyalella* survival with sediment quality guidelines in sediments from the Tri-States mining district: (a) summed PEC quotients; (b) equilibrium-partitioning sediment benchmarks. Curves are logistic regressions; dashed lines indicate thresholds for 10% and 20% reductions (T_{10} and T_{20}) (MacDonald et al., 2009).

toxicity predicted if the sum of TUs is less than 1.0, and increasing toxicity hazards as sums exceed 1.0. This approach assumes that toxicity of metals is additive, although both greater than additive and less than additive toxicity have been reported in toxicity tests with metal mixtures (Spehar and Fiandt, 1986). Predictions of toxicity based on pore water toxic units should be relatively reliable because water quality criteria for metals are based on the extensive literature on the toxicity of metals to many aquatic taxa and because metal criteria explicitly consider the influence of pore water characteristics on metal toxicity (e.g., hardness-toxicity regressions or biotic ligand models; USEPA, 2007b, 2009).

Sediment ESBs (USEPA, 2005) are intended to predict the distribution of metals between sediment particles (low toxicity) and pore water (high toxicity). The simplest forms of the sediment ESBs predict (1) that no metal toxicity should occur if the molar concentration of AVS is greater than the summed molar concentration of SEM, and (2) that positive values for SEM-AVS estimate the potentially bioavailable fraction of sediment metals. The third form of the sediment ESB refined the estimate of bioavailable metals by normalizing SEM-AVS to the organic carbon fraction of the sediment ($f$oc), expressed as (SEM-AVS)/$f$oc. Using this formulation, USEPA (2005) analyzed a large database of metal-contaminated sediments (metal-spiked and field-collected) and estimated a 95% upper boundary for non-toxic sediments of 130 μmol/g oc and a 95% lower boundary for toxic sediments of 3000 μmol/g oc. Sediments with metal concentrations between these two values are classified as having ‘uncertain toxicity’.

ESB models attempt to predict the toxicity of metal-contaminated sediments by directly measuring metal exposure (e.g., dissolved metals in pore water) or by applying simple models of metal bioavailability (e.g., binding by AVS and organic matter) on a site-specific basis. Also, both sediment and pore-water ESBs explicitly model the contribution of metal mixtures to sediment toxicity. In contrast, empirical SQGs are not site-specific and SQGs for individual metals are presumably influenced by effects of co-occurring metals (and other toxicants). However, characterization of toxicity hazards using ESB models requires both additional measurements of metal concentrations in sediment fractions other than total metals (i.e., SEM or pore water metals) and additional analyses of physicochemical characteristics of sediment (AVS and TOC) or pore water (e.g., hardness), compared to empirical SQGs.

The greater sophistication and complexity of sediment ESBs does not always translate into more accurate predictions of metal toxicity. In some studies of metal-contaminated habitats, where metal mixtures and controls on bioavailability are relatively constant among locations, sediment ESBs do not noticeably improve predictions of metal toxicity over PEQs. For example, toxicity of sediments from the Tri-State mining district to *Hyalella* was predicted equally well by summed PEQs (Fig. 10a) and by ESBs for sediment (Fig. 10b) and pore water (MacDonald et al., 2009). However, ESB models can prove more robust for explaining differences in bioavailability across a wide range of conditions. For example, sediments from the Vermont Copper belt included sediments from acidic sites that were not toxic to invertebrates despite summed PEQ values as high as 100. However, the ESB index values for these sediments were less than 3000 μmol/g oc (i.e, in the uncertain-toxicity range), because a high percentage of metals in these sediments were not recovered in the SEM fraction (Seal et al., 2010). In studies of nickel-spiked sediments, normalization of sediment nickel concentrations to SEM-AVS reduced or eliminated differences in nickel toxicity values between sediments with wide differences of AVS and TOC for several species of invertebrates, but had lesser effect on toxicity values for burrowing mayflies (*Hexagenia* sp.) (Besser et al., 2013). In sediments with high AVS concentrations, toxicity to *Hexagenia* occurred in some sediments with negative values of SEM-AVS, contrary to predictions of the ESB model. In this case, the apparent failure of the AVS-binding hypothesis may represent our inability to measure the localized release of toxic metal concentrations in the microhabitat inhabited by a benthic organism.

4.3. European environmental quality standards

Regulations guiding development of European environmental quality standards for metals incorporate elements of both empirical and equilibrium-partitioning approaches. Toxicity tests with spiked sediments and a wide variety of test organisms are used to estimate a probable no-effect concentration (PNEC), which is intended to protect sensitive species of benthic organisms under 'reasonable worst case' or high-bioavailability conditions (European Commission, 2011). The PNEC may then be adjusted to reflect site-specific conditions, using bioavailability models based on additional testing of spiked sediments with a broad range of sediment characteristics (Bodar et al., 2005; Besser et al., 2013; Vangheluwe et al., 2013).

5. Conclusions

Sediment toxicity tests are an invaluable tool for characterizing biological impacts of metal-contaminated sediments associated with mining, ore-processing, and smelting. The toxicity of metals and aquatic sediments reflects both the geochemistry of sediment and pore water and the biology and behavior of benthic organisms. Metal bioavailability is often controlled by strong metal-binding phases such as acid-volatile sulfides and organic matter, which greatly influence metal concentrations in pore water. The toxicity of aqueous metals can be reliably modeled in terms of binding of dissolved metals to 'biotic ligands' on the body surface of aquatic organisms. This framework for understanding bioavailability has increased the sophistication of sampling and analytical methods and provided a robust technical basis for sediment quality guidelines. Despite this strong technical basis, chemical analyses and chemistry-based sediment quality guidelines sometimes do not accurately predict toxic effects of sediments in the laboratory or in the field.

In contrast, direct assessment of toxicity using whole-sediment toxicity tests can reliably quantify the responses of benthic invertebrates to metal-contaminated sediments. Although laboratory toxicity tests cannot reproduce the complexity of the natural environment, modern test methods allow us to manipulate several variables affecting metal bioavailability and toxicity in sediments, including test organism, life stage, diet, water quality, and water replacement rates. Spiked-sediment tests give us the additional opportunity to choose the type of sediment environment to be tested and the specific metal or metals of interest. Depending on study objectives and experimental design, toxicity data from sediment tests can be used to document the extent and severity of toxic effects of metal contaminated sediments, relative to uncontaminated reference sites, or to model quantitative relationships between metal exposure concentrations and toxic effects. When combined with careful chemical characterization of metal exposure and appropriate field validation, sediment toxicity testing provides a powerful tool to document cause-and-effect relationships between sources of metal contamination and adverse effects on resident aquatic biota.

Acknowledgements

We thank the scientific and technical staff of the USGS Columbia Environmental Research Center, who made many of the studies described here possible. We also greatly appreciate the scientific and financial contributions of our collaborators and clients in USGS, US Fish and Wildlife Service, US Environmental Protection Agency, Nickel Producers Environmental Research Association, Macdonald Environmental Sciences Limited, and many others. Travis Schmidt and Andrew Todd provided helpful comments on the manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.apgeochem.2014.05.021.

References

Ankley, G.T., DiToro, D.M., Hansen, D.J., Berry, W.J., 1996. Technical basis and proposal for deriving sediment quality criteria for metals. Environ. Toxicol. Chem. 15, 2056–2066.

ASTM, 2010a. Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates, E1706-05: West Conshohocken PA, ASTM International.

ASTM, 2010b. Standard guide for conducting acute toxicity tests on test materials with fishes, macroinvertebrates, and amphibians, E 729-96: West Conshohocken, PA, ASTM International, pp. 21.

ASTM, 2010c. Standard guide for storage, characterization, and manipulation of sediments for toxicological testing and for selection of samplers used to collect benthic invertebrates, E1391-03: West Conshohocken PA, ASTM International. pp. 93.

ASTM, 2010d. Standard guide for conducting laboratory toxicity tests with freshwater mussels, E2455-06: West Conshohocken, PA, ASTM International, pp. 53.

ASTM, 2010e. Test method for measuring the bioaccumulation of sediment-associated contaminants by freshwater invertebrates, E1688-05: West Conshohocken PA, ASTM International, pp. 43.

Batley, G.E., Stahl, R.G., Babut, M.P., Bott, T.L., Clark, J.R., Field, L.J., Ho, K.T., Mount, D.R., Swartz, R.C., Tessier, A., 2005. Scientific underpinnings of sediment quality guidelines. In: Wenning, R.J., Batley, G.E., Ingersoll, C., Moore, D.W. (Eds.), Use of Sediment Quality Guidelines and Related Tools for the Assessment of Contaminated Sediments: Pensacola, FL, Society of Environmental Toxicology and Chemistry, pp. 39–120.

Besser, J.M., Brumbaugh, W.G., Ivey, C.D., Ingersoll, C., Moran, P.W., 2008. Biological and chemical characterization of metal bioavailability in sediments from Lake Roosevelt, Columbia River, Washington USA. Arch. Environ. Contam. Toxicol. 54, 557–570.

Besser, J.M., Brumbaugh, W.G., Allert, A.L., Poulton, B.C., Schmitt, C.J., Ingersoll, C., 2009a. Ecological impacts of lead mining on Ozark streams: Toxicity of sediment and pore water. Ecotoxicol. Environ. Safety 72, 516–526.

Besser, J.M., Brumbaugh, W.G., Hardesty, D., Hughes, J., Ingersoll, C.G., Roberts, A., Weber, J., Mosby, D., Hundley, J., McMurray, S., Faiman, S., 2009b. Laboratory and field assessment of impacts of sediment metal contamination on freshwater mussels in a southeast Missouri lead-mining district. Book of Abstracts, Annual meeting of -SETAC-North America, November 2009. Society of Environmental Toxicology and Chemistry, Pensacola FL. pp. 49.

Besser, J.M., Brumbaugh, W.G., Hardesty, D., Ingersoll, C.G., 2010. Toxicity and metal contamination of water and sediment from the Palmerton Zinc site, Palmerton, Pennsylvania. Administrative Report CERC-2008-NRDAR-04. U.S. Geological Survey, Columbia MO. pp. 48.

Besser, J.M., Brumbaugh, W.G., Ingersoll, C.G., Ivey, C.D., Kunz, J.L., Kemble, N.E., Schlekat, C.E., Garman, E.R., 2013. Chronic toxicity of nickel-spiked freshwater sediments: Variation in toxicity among eight invertebrate taxa and eight sediments. Environ. Toxicol. Chem. 32, 2495–2506.

Besser, J.M., Ingersoll, C.G., Giesy, J.P., 1996. Effects of spatial and temporal variability of acid-volatile sulfide on the bioavailability of copper and zinc in freshwater sediments. Environ. Toxicol. Chem. 15, 286–293.

Besser, J.M., Leib, K.J., 2007. Toxicity of metals in water and sediment to aquatic biota. In: Church, S.E., von Guerard, P., Finger, S.E., (Eds.), Professional Paper 1951-E20, Integrated Investigations of Environmental Effects of Historical Mining in Upper Animas Watershed, San Juan County, Colorado: Denver, Colorado, U.S. Geological Survey, pp. 837–850.

Bodar, C.W.M., Pronk, M.E.J., Sijm, D.T.H.M., 2005. The European Union risk assessment on zinc and zinc compounds: the process and the facts. Integr. Environ. Assess Manage. 1, 301–319.

Borgmann, U., Neron, R., Norwood, W., 2001. Quantification of bioavailable nickel in sediments and toxic thresholds to Hyalella azteca. Environ. Pollut. 111, 189–198.

Brumbaugh, W.G., Arms, J.W., 1996. Quality control considerations for the determination of acid-volatile sulfide and simultaneously extracted metals in sediments. Environ. Toxicol. Chem. 15, 282–285.

Brumbaugh, W.G., Besser, J.M., Ingersoll, C.G., May, T.W., Ivey, C.D., Schlekat, C.E., Garman, E.R., 2013. Preparation and characterization of nickel-spiked freshwater sediments for toxicity tests: toward more environmentally realistic nickel partitioning. Environ. Toxicol. Chem. 32, 2482–2494.

Brumbaugh, W.G., May, T.W., Besser, J.M., Allert, A.L., Schmitt, C.J., 2007. Assessment of elemental concentrations in streams of the New Lead Belt in southeastern Missouri, 2002-2005, Science Investigations Report 2007-5057: Columbia, Missouri, U.S. Geological Survey, pp. 57.

Burton Jr, A., Ingersoll, C.G., Burnett, L.A.C., Henry, M., Hinman, M.L., Klaine, S.J., Landrum, P.F., Ross, P., Tuchman, M., 1996. A comparison of sediment toxicity test methods at three Great Lake areas of concern. J. Great Lakes Res. 22, 495–511.

Cain, D.J., Luoma, S.N., Wallace, W.G., 2004. Linking metal bioaccumulation of aquatic insects to their distribution patterns in a mining-impacted river. Environ. Toxicol. Chem. 23, 1463–1473.

Canfield, T.J., Kemble, N.E., Brumbaugh, W.G., Dwyer, F.J., Ingersoll, C.G., Fairchild, J.F., 1994. Use of benthic invertebrate community structure and the sediment quality triad to evaluate metal-contaminated sediment in the upper Clark Fork River, Montana. Environ. Toxicol. Chem. 13, 1999–2012.

Carignan, R., Rapin, F., Tessier, A., 1985. Sediment porewater sampling for metal analysis: a comparison of techniques. Geochim. Cosmochim. Acta 49, 2493–2497.

DiToro, D.M., Allen, H.E., Bergman, H.L., Meyer, J.S., Paquin, P.R., Santore, R.C., 2001. Biotic ligand model of the acute toxicity of metals. 1. Technical basis. Environ. Toxicol. Chem. 20, 2383–2396.

DiToro, D.M., Mahony, J.D., Hansen, D.J., Scott, K.J., Carlson, A.R., Ankley, G.T., 1992. Acid volatile sulfide predicts the acute toxicity of cadmium and nickel in sediments. Environ. Sci. Technol. 26, 96–101.

Environment Canada, 2002 Biological Test Method: Reference Method for Determining the Toxicity of Sediment Using Luminescent Bacteria in a Solid-Phase Test, EPS1/RM/42 Ottawa CA, Environment Canada.

European Commission 2011. Technical Guidance For Deriving Environmental Quality Standards. Guidance Document No. 27. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Technical Report 2011–055. Luxembourg.

Filgueiras, A., Lavilla, I., Bendicho, C., 2002. Chemical sequential extraction for metal partitioning in environmental solid samples: J. Environ. Monit 4, 823–857.

Giesy, J.P., Hoke, R.A., 1989. Freshwater sediment toxicity bioassessment: rationale for species selection and test design. J. Great Lakes Res. 15, 539–569.

Giesy, J.P., Rosiu, C.J., Graney, R.L., Henry, M.G., 1990. Benthic invertebrate bioassays with toxic sediment and pore water. Environ. Toxicol. Chem. 9, 233–248.

Hockett, J.R., Mount, D.R., 1996. Use of metal chelating agents to differentiate among sources of acute aquatic toxicity. Environ. Toxicol. Chem. 15, 1687–1693.

Hunt, J.W., Anderson, B.S., Phillips, B.M., Newman, J., Tjeerdema, R.S., Fairey, R., Puckett, H.M., Stephenson, M., Smith, R.W., Wilson, C.J., 2001. Evaluation and use of sediment toxicity reference sites for statistical comparisons in regional assessments. Environ. Toxicol. Chem. 20, 1266–1275.

Hutchins, C.M., Teasdale, P.R., Lee, J., Simpson, S.L., 2007. The effect of manipulating sediment pH on the porewater chemistry of copper-and zinc-spiked sediments. Chemosphere 69, 1089–1099.

Hutchins, C.M., Teasdale, P.R., Lee, S.Y., Simpson, S.L., 2008. Cu and Zn concentration gradients created by dilution of pH neutral metal-spiked marine sediment: a comparison of sediment geochemistry with direct methods of metal addition. Environ. Sci. Technol. 42, 2912–2918.

Ingersoll, C.G., Besser, J.M., Brumbaugh, W.G., Ivey, C.D., Kemble, N.E., Kunz, J.L., May, T.W., Wang, N., MacDonald, D.D., Smorong, D.E., 2008. Sediment chemistry, toxicity, and bioaccumulation data report for the US Environmental Protection Agency – Department of the Interior sampling of metal-contaminated sediment in the Tri-state Mining District in Missouri, Oklahoma, and Kansas; MacDonald Environmental Sciences Ltd., Nanaimo, BC.

Ingersoll, C. G., Brunson, E. L., Dwyer, F. J., Hardesty, D. K., and Kemble, N. E., 1998, Use of sublethal endpoints in sediment toxicity tests with the amphipod Hyalella azteca: Environ Toxicol Technol, v. 17, p. 1508–1523.

Ingersoll, C.G., Haverland, P.S., Brunson, E.L., Canfield, T.J., Dwyer, F.J., Henke, C.E., Kemble, N.E., Mount, D.R., 1996. Calculation and evaluation of sediment effect concentrations for the amphipod Hyallela azteca and the midge Chironomus riparius. J. Great Lakes Res. 22, 602–623.

Ingersoll, C.G., MacDonald, D.D., Brumbaugh, W.G., Johnson, B.T., Kemble, N.E., Kunz, J.L., May, T.W., Wang, N., Smith, J.R., Sparks, D.W., 2002. Toxicity assessment of sediments from the Grand Calumet River and Indiana Harbor Canal in northwestern Indiana, USA. Arch. Environ. Contam. Toxicol. 43, 156–167.

Ingersoll, C.G., MacDonald, D.D., Wang, N., Crane, J.L., Field, L.J., Haverland, P.S., Kemble, N.E., Lindskoog, R.A., Severn, C., Smorong, D.E., 2001. Predictions of sediment toxicity using consensus-based freshwater sediment quality guidelines. Arch. Environ. Contam. Toxicol. 41, 8–21.

Kemble, N.E., Brumbaugh, W.G., Brunson, E.L., Dwyer, F.J., Ingersoll, C.G., Monda, D.P., Woodward, D.F., 1994. Toxicity of metal-contaminated sediment from the upper Clark Fork River, Montana, to aquatic invertebrates and fish in laboratory exposures. Environ. Toxicol. Chem. 13, 1985–1997.

Luoma, S.N., 1989. Can we determine the biological availability of sediment-bound trace elements? Hydrobiologia 176 (177), 379–396.

MacDonald, D.D., Ingersoll, C.G., Berger, T.A., 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch. Environ. Contamin. Toxicol. 39, 20–31.

MacDonald, D.D., Smorong, D.E., Ingersoll, C.G., Besser, J.M., Brumbaugh, W.G., Kemble, N.E., May, T.W., Ivey, C.D., Irving, S., O'Hare, M., 2009. Development and Evaluation of Sediment and Pore-Water Toxicity Thresholds to Support Sediment Quality Assessments in the Tri-State Mining District (TSMD), Missouri, Oklahoma, and Kansas. Final Technical Report to U.S. Environmental Protection Agency, Dallas TX and Kansas City KS, and U.S. Fish and Wildlife Service, Columbia MO; MacDonald Environmental Sciences Ltd., Nanaimo, British Columbia.

Milani, D., Reynoldson, T.B., Borgmann, U., Kolasa, J., 2003. The relative sensitivity of four benthic invertebrates to metals in spiked-sediment exposures and application to contaminated field sediment. Environ. Toxicol. Chem. 22, 845–854.

Mount, D.R., Highland, T.L., Mattson, V.R., Dawson, T.D., Lott, K.G., Ingersoll, C.G., 2006. Use of the oligochaete, Lumbriculus variegatus, as a prey organisms for toxicant exposure of fish through the diet. Environ. Toxicol. Chem. 25, 195–202.

Niyogi, S., Wood, C.M., 2004. Biotic ligand model, a flexible tool for developing site-specific water quality guidelines for metals. Environ. Sci. Technol. 38, 6177–6192.

OECD, 2004. Test No. 218: Sediment-Water Chironomid Toxicity Using Spiked Sediment: Geneva, Organization for Economic Cooperation and Development.

OECD, 2007. Test No. 225: Sediment-Water Chironomid Toxicity Using Spiked Sediment: Geneva, Organization for Economic Cooperation and Development.

Paulson, A.J., Cox, S.E., 2007. Release of elements to natural water from Lake Roosevelt sediments. Environ. Toxicol. Chem. 12, 2550–2559.

Phipps, G.L., Mattson, V.R., Ankley, G.T., 1995. Relative sensitivity of three freshwater benthic macroinvertebrates to ten contaminants. Arch. Environ. Contam. Toxicol. 28, 281–286.

Piatak, N.M., Argue, D.M., Seal, R.R., Kiah, R.G., Besser, J.M., Coles, J.F., Hammarstrom, JM, Levitan, DM, Deacon, J.R., Ingersoll, C.G., 2013. Aquatic Assessment of the Pike Hill Copper Mine Superfund Site, Corinth, Vermont, Scientific Investigations Report 2012-5288. Reston VA: U.S. Geological Survey.

Schults, D.W., Ferraro, S.P., Smith, L.W., Roberts, F.A., Poindexter, C.K., 1992. A comparison of methods for collecting interstitial water for trace organic compounds and metals analyses. Arch. Environ. Contam. Toxicol.

Seal, R.R., Kiah, R.G., Piatak, N.M., Besser, J.M., Coles, J.F., Hammarstrom, J.M., Argur, D.M., Levitan, D.M., Deacon, J.R., Ingersoll, C., 2010. Aquatic Assessment of the Ely Copper Mine Superfund Site, Vershire, Vermont. Scientific Investigations Report 2010–5084, U.S. Geological Survey, Reston VA, pp. 131.

Smith, K., Briggs, P., Campbell, D., Castle, C., Desborough, G., Eppinger III, R., Fitterman, D., Hageman, P., Leinz, R., and Meeker, G., 2000. Tools for the rapid screening and characterization of historical metal-mining waste dumps: 2000 Billings Land Reclamation Symposium, pp. 20–24.

Spehar, R.L., Fiandt, J.T., 1986. Acute and chronic effects of water quality criteria-based metal mixtures on three aquatic species. Environ. Toxicol. Chem. 5, 917–931.

Tessier, A., Campbell, T.G.C., Auclair, J.C., 1984. Relationships between the partitioning of trace metals in sediments and their accumulation in the tissues of the freshwater mollusk Ellitio complanata in a mining area. Can. J. Fish. Aquat. Sci. 41, 1463–1472.

USEPA, 1991. Draft analytical method for determination of acid volatile sulfide and selected simultaneously extractable metals in sediment. Draft report EPA-821-R-91-100: Washington DC, U.S. Environmental Protection Agency.

USEPA, 2000, Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates, Report EPA/600/R-99/064: Duluth, Minnesota U.S. Environmental Protection Agency, pp. 133.

USEPA, 2001. Methods for collection, storage and manipulation of sediments for chemical and toxicological analyses, Technical Manual EPA-823-B-01-002: Washington DC, U.S. Environmental Protection Agency.

USEPA, 2002, Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms, Fourth Edition, EPA-821-R-02-013: Washington, DC, U.S. Environmental Protection Agency, 335 p.

USEPA, 2005. Procedures for the derivation of equilibrium partitioning sediment benchmarks (ESBs) for the protection of benthic organisms: Metal mixtures (cadmium copper, lead nickel, silver, and zinc). EPA-600-R-Q2-011: Washington DC, U.S. Environmental Protection Agency.

USEPA, 2007a. Sediment Toxicity Identification Evaluation (TIE) Phases I, II, and III Guidance Document. EPA-600-R-07-080, Washington DC, U.S. Environmental Protection Agency, pp. 134.

USEPA, 2007b. Aquatic life ambient freshwater quality criteria – copper, 2007 revision, EPA-822-R-07-001: Washington DC, U.S. Environmental Protection Agency, pp. 208.

USEPA, 2009, National Recommended Water Quality Criteria: Washington DC, U.S. Environmental Protection Agency, pp. 22.

USEPA, USACE, 1998. Evaluation of dredged material proposed for discharge in waters of the U.S. – Testing manual. EPA-823-B-98-004.: Washington DC, U.S. Environmental Protection Agency and U. S. Army Corps of Engineers.

Vangheluwe, M.L., Verdonck, F.A., Besser, J.M., Brumbaugh, W.G., Ingersoll, C.G., Schlekat, C.E., Garman, E.R., 2013. Improving sediment quality guidelines for nickel: development and application of predictive bioavailability models to assess chronic toxicity of nickel in freshwater sediments. Environ. Toxicol. Chem. 32, 2507–2519.

Wang, N., Mebane, C.A., Kunz, J.L., Ingersoll, C.G., May, T.W., Arnold, W., Santore, R.C., Augspurger, T., Dwyer, F.J., Barnhart, M.C., 2009. Evaluation of acute copper toxicity to juvenile freshwater mussels (fatmucket, Lampsilis siliquoidea) in natural and reconstituted waters. Environ. Toxicol. Chem. 28, 2367–2377.

Wenning, R.J., Batley, G.E., Ingersoll, C.G., Moore, D.W., 2005. Use of sediment quality guidelines and related tools for the assessment of contaminated sediments. SETAC Press, Pensacola FL.

Williams, J.D., Warren, M.L., Cummings, K.S., Harris, J.L., Neves, R.J., 1993. Conservation status of freshwater mussels of the United States and Canada. Fisheries 18, 6–22.

Woodward, D.F., Farag, A.M., Bergman, H.L., DeLonay, A.J., Little, E.E., Smith, C.E., Barrows, F.T., 1995. Metals-contaminated benthic invertebrates in the Clark Fork River, Montana: effects on age-0 brown trout and rainbow trout. Can. J. Fisheries Aquat. Sci. 52, 1994–2004.

Zhang, H., Davison, W., Miller, S., Tych, W., 1995. In-situ high resolution fluxes of Ci, Cu, Fe, and Mn and concentrations of Zn and Cd in pore water by DGT. Geochim. Cosmochim. Acta 59, 418–419.

Zumwalt, D.C., Dwyer, F.J., Greer, I.E., Ingersoll, C.G., 1994. A water-renewal system that accurately delivers small volumes of water to exposure chambers. Environ. Toxicol. Chem. 13, 1311–1314.