University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Valery Forbes Publications | Papers in the Biological Sciences

1-1-2002

Uncertainties in Sediment Quality Weight-of-Evidence (WOE) Assessments

Graeme E. Batley
Centre for Advanced Analytical Chemistry, CSIRO Energy Technology, Bangor, Australia, graeme.batley@csiro.au

G. Allen Burton
Institute for Environmental Quality, Wright State University, allen.burton@wright.edu

Peter M. Chapman
EVS Environment Consultants, North Vancouver, BC

Valery E. Forbes
University of Nebraska-Lincoln, vforbes3@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/biosciforbes

Part of the Pharmacology, Toxicology and Environmental Health Commons

Batley, Graeme E.; Burton, G. Allen; Chapman, Peter M.; and Forbes, Valery E., "Uncertainties in Sediment Quality Weight-of-Evidence (WOE) Assessments" (2002). *Valery Forbes Publications.* Paper 31.
http://digitalcommons.unl.edu/biosciforbes/31

This Article is brought to you for free and open access by the Papers in the Biological Sciences at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Valery Forbes Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in *Human and Ecological Risk Assessment* 8:7 (2002), pp. 1517-1547.
Copyright © 2002 ASP/Taylor & Francis. Used by permission.

Uncertainties in Sediment Quality Weight-of-Evidence (WOE) Assessments

Graeme E. Batley,[1] G. Allen Burton,[2]
Peter M. Chapman,[3] and Valery E. Forbes [4]

1. Centre for Advanced Analytical Chemistry, CSIRO Energy Technology, Bangor, Australia (corresponding author; graeme.batley@csiro.au)
2. Institute for Environmental Quality, Wright State University, Dayton, Ohio, USA
3. EVS Environment Consultants, 195 Pemberton Avenue, North Vancouver, British Columbia, Canada
4. Department of Life Sciences and Chemistry, Roskilde University, Roskilde, Denmark (affiliation as of 2010: School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, Nebraska, USA; vforbes3@unl.edu)

Abstract
Uncertainties in sediment quality assessments are discussed in five categories: (1) sediment sampling, transport and storage; (2) sediment chemistry; (3) ecotoxicology; (4) benthic community structure; and (5) data uncertainties and QA/QC. Three major exposure routes are considered: whole sediments, and waters in sediment pores and at the sediment-water interface. If these uncertainties are not recognized and addressed in the assessment process, then erroneous conclusions may result. Recommendations are provided for addressing the identified uncertainties in each of the key areas. The purpose of this paper is to improve the reporting of sediment quality assessments.

Keywords: risk assessment, weight-of-evidence, contamination, toxicity, benthos

Introduction

Any meaningful assessment of sediment quality needs to involve consideration of multiple lines of evidence, typically from sediment chemistry, ecotoxicology, and benthic ecology. Implicit in any such evaluation will be the assumption that the quality of the data is acceptable. There are, however, a number of uncertainties and limitations associated with measurements in each of these lines of evidence some of which have been the subject of previous publications (ASTM 1994; Burton 1991; 1995a,b; Burton et al. 2000; Environment Canada 1994; Carr et al. 2001; US EPA 2002). While many of the researchers involved in developing the protocols for particular chemical, ecotoxicological or ecological studies may be aware of and indeed may have documented these limitations, these are often lost in subsequent adoption by users.

This paper discusses such uncertainties and limitations relative to weight-of-evidence (WOE) sediment quality assessments for whole sediment, pore water, and water at the sediment water interface (SWI). Specific issues are discussed, and recommended resolutions proposed.

Sediment Sampling, Transport and Storage

Prior to any measurements, the first potential source of errors in sediment assessment can arise during sampling of sediment, pore, and sediment-water interface (SWI) waters, in particular the possibility for changes in chemical speciation (Batley 1999) and bioavailability (Burton 1991). Sediment and pore or SWI water composition is spatially and in some cases (Chapman and Wang 2001), temporally variable. Sampling issues provide the largest source of uncertainty in sediment data sets (Crumbling et al. 2001). The choice of appropriate reference sites is also critical; multiple reference sites are preferred. The sediments in both reference and exposed sites should have the same gross physical and chemical characteristics, including such parameters as grain size, particulate organic carbon, and the same overlying water characteristics in terms especially of pH, hardness and salinity. Biological parameters (i.e., resident biota, particularly the benthos) should also be broadly similar in terms of the distribution of major taxa (e.g., family level) and biomass.

Numerous publications have documented the artifacts that may result from different sampling collection, handling and manipulation procedures, along with differences in chemistry and toxicity resulting from use of different methods (e.g., dredge sample vs. core sample) (e.g., citations in ASTM 1994; Burton 1991; Diamond et al. 2002; Environment Canada 1994; US EPA 2002). However, there is little agreement on the preferred methods to use or clear guidance on the relationship between methods and possible artifacts. Given the varying heterogeneity of sediments, no one method can be universally applied to all sediment assessments. Rather, it is important that sediment investigators have a clear understanding of the unique advantages and limitations of the various methods and possible artifacts that may result from their use.

Sediment Heterogeneity

Sediments are notoriously both physically and chemically heterogeneous (Burton 1995b; Watzin et al. 1997). From a management and cost perspective, there are sampling issues involved in describing the extent of contaminant distributions both laterally and vertically (the latter being more an issue in sediments being dredged and disposed of at sea). There are also important heterogeneity issues in determining the toxic impacts of a contaminated area. In addition to the well-known vertical and horizontal gradients that exist in sediments from diagenesis and contamination dilution, respectively, are patchiness issues. In particular, stream systems are extremely patchy from a physical, chemical and biological perspective (Minshall 1988). Recent studies (Shuttleworth and Davison 1999) have provided evidence for the presence of metal microniches in sediments, and similar microniches of bacterial activity have also been detected (Apte et al. unpublished results; Stemmer et al. 1990ab).

Spatial averaging may be appropriate for the study being conducted, but if the concern is to determine "hot spots" then more detailed sampling is needed. However, if the concern is to determine broad-scale or relative sediment contamination, then less extensive sampling may be required. Any spatial averaging must be appropriate to the study purpose and to the chemical and biological measurements. Without knowing the variability of the site, it is difficult to determine the ecological significance of contamination and set effective clean-up goals.

Sediment and Pore Water Sampling Depths

In depositional areas, sediment deposition typically occurs at rates between 2 and 10 mm/year. For accurate chemical characterization of contaminant distributions, sediments are often sampled at depth intervals of 1 or 2 cm. However, for toxicity testing, sediments are usually collected as cores or grab samples, where sediments are homogenized over depths of several to tens of centimeters. Oxic sediments are confined to the uppermost layer, which may be 2 cm or greater in sandy sediments, but typically as little as the top 2 mm or less in organic-rich, silty sediments.

If it is reasoned that most of the biological activity is in the top one to few cm, then sampling only this fraction for toxicity testing is appropriate, unless the investigation is a dredging issue. There are, however, organisms (primarily in marine systems) that can burrow to depths well below this, and toxic effects at depth may also be a concern. In this case, toxicity testing on the surface layer only may not reflect the chemistry of the whole system.

The depth of sampling should be governed by the realities of the resident infauna. Ideally, tested depths should reflect the habitat of the majority of the infauna, and the depth of collections for benthic community structure analyses. If, however, programmatic considerations take precedence (e.g., sampling only the top 2 cm to assess relatively recent contamination), then the final assessment must note and discuss any differences between depth sampled and infaunal habitat depths. For example, is it likely that infauna burrowing below 2 cm are affected by surficial sediment contamination?

Pore waters are often a focus of study because they offer greater analytical simplicity and are regarded as a dominant exposure route for benthic invertebrates. However, the extreme vertical heterogeneity of pore water chemistry over distances of centimeters, the complex physico-chemical interrelationship between pore waters and surrounding solids, and the often crude and varied sampling methods make accurate characterizations a challenge. Pore waters near the SWI may be oxic or anoxic, depending on the depositional nature of the sediments and the hydrodynamics of the site. Potentially available metal concentrations near this boundary can be elevated in oxic environments. Near the redox boundary, reduced iron and manganese in the pore waters can also show high concentration spikes (Lasier et al. 2000), while in deeper sediments metals may be moderated by the presence of acid-volatile sulfide (AVS).

Sampling depth must be based on both study objectives and on biological and physico-chemical realities. Within these constraints, depth-integrated pore water and sediments should generally be collected for testing and analysis. The same depths

should be sampled for chemical analyses, toxicity testing, and benthic infaunal community structure analyses.

Pore Water Sampling Methods

Sampling will, by definition, perturb the chemical form of the pore waters. Alteration of sediment chemistry during sampling will affect the processes of toxicant mobilization, and subsequent bioavailability via toxicant exposure/uptake, particularly for metal contaminants. These chemical changes may, or may not, result in significant changes in toxicity; however, typically this is not known *a priori*. A major pore water sampling concern is anoxic sediment oxidation. This will involve primarily the oxidation of particulate and dissolved Fe(II), Mn(II), and sulfide. Although pH dependent, these processes are rapid at the pH of natural waters. The rate of oxidation of particulate metal sulfides is slow, unlike iron monosulfide, and so these metals are unlikely to make a significant contribution to changes in toxicity of oxidized pore waters. There is, however, the possibility that pore water pH will be reduced as a consequence of hydrogen ion release during FeS oxidation, and this may aid metal mobilization where carbonate buffering is inadequate (Simpson et al. 1998; 2000a). Oxidized iron (and manganese) may reprecipitate as hydrous iron oxide, carrying down soluble metals. This process may occur in the pore or SWI waters.

Methods for sampling pore water are reviewed by ASTM (1994), Carr et al. (2001), Environment Canada (1994), Doe et al. (2002), and the U.S. Environmental Protection Agency (US EPA 2002). Separation procedures include centrifugation, suction, pore water dialysis and squeezing. Variously, these have the potential to change chemical equilibria during separation, to lose dissolved gases or volatiles, and to differ in their ability to sample from a discrete depth in the sediment. The most common and recommended laboratory practice is to centrifuge the sample. The pore water collected in this way may have a different chemical composition, depending on the depth of sediment sampled, given that the sampling technique may be combining oxic and anoxic sediments from different depths. Suction is an imprecise technique that works best in coarser sediments. It has the possibility to draw in pore waters from an ill-defined area around the extract point and could entrain overlying waters. Squeezing may be more effective than centrifugation of sandy sediments, but both may contribute more to degassing of the sediments and ion concentration shifts. Losses of volatiles are possibly more likely with squeezing. If the pore waters are permitted to oxidize, metals will certainly be lost from solution by coprecipitation. *In situ* pore water collection reduces the likelihood for oxidation/sampling artifacts (discussed below).

Oxidation must be minimized; sampling under a nitrogen gas atmosphere is recommended. For instance, oxidation of anoxic sediments during centrifugation should be avoided, by using an inert gas blanket (typically nitrogen). At this point in time, ASTM, Environment Canada and the US EPA have recommended centrifugation as the "best" method to extract pore water. Others, however, suggest that the optimal methods vary with the site and project objectives, as all methods have advantages and disadvantages (Carr et al. 2001; Chapman et al. 2002).

Sampling designed to collect only SWI pore waters will not be without problems due to pore water oxidation, as typically these samples will be anoxic or suboxic in parts, and subject to chemical changes.

Organic contaminants in pore waters present different problems. A confounding factor in the sampling of pore waters containing hydrophobic organics having high sediment:water partition coefficients is their preferential adsorption to particles of any kind and to surfaces. This can mean that many of these compounds may not be in true solution as expected, but attached to colloidal particles. Recent studies (Batley et al. unpublished results) using solvent-filled dialysis samplers for organochlorines in waters, indicate a decrease in the concentration of sampled compounds in the presence of iron oxide/humic acid colloids. Colloidal association has the potential to affect contaminant bioavailability; however, such speciation of organics in waters or pore waters is rarely considered in either chemical or biological studies.

Obtaining Sufficient Pore Water

For toxicity testing, obtaining sufficient pore water can be a major limitation (ASTM 1994; Chapman et al. 2002; Environment Canada 1994; US EPA 2002). For many tests this means extracting pore waters from kilogram quantities of sediments, and here centrifugation is probably the best procedure. *In situ* dialysis (peeper) sampling offers the best prospects for obtaining a sample whose chemistry is unchanged (Teasdale et al. 1995). A limitation with traditionally used peepers is the volume of sample that can be collected, typically 7 to 15 mL. In some studies, pore waters are sampled using large peepers buried in the test sediment (Hansen et al. 1996; Burton 1992; Sarda and Burton 1995). The time needed for peeper equilibration varies from hours to weeks, depending on their surface to volume ratio (Teasdale et al. 1995). Pore waters sampled by peepers in systems where the redox boundary is changing (e.g., tidal systems), may artificially concentrate iron (and metals) in the boundary layer chambers. This is because iron, precipitated in an oxic cycle, is slow to reduce and be solubilized when the pore waters go anoxic, and more will precipitate during subsequent oxic cycles (Teasdale et al. unpublished results).

Appropriate *in situ* dialysis (peepers) is currently the best method for obtaining samples with minimally changed chemistry. However, centrifugation is presently the only reasonable alternative for collecting large volumes of pore waters relatively rapidly. For organic contaminants, obtaining sufficient pore water to be able to analytically detect environmental concentrations is a major problem that has limited studies of these compounds.

Sediment Storage

It is not always possible to test sediments immediately, however, protracted storage has the potential to alter sediment chemistry. There are many different opinions as to the length of time that sediments can be stored without significant chemical or toxicological changes before testing (Burton 1995b). However, there are no sureties. Field collected sediments should be tested as soon as possible, and storage times (and conditions) reported.

Sediments should be stored cool, in the dark, and to minimize oxidation, samples should be stored in filled, sealed containers with oxygen excluded. It is not recommended that oxic sediments should be treated differently. Few sediments will be exclusively oxic. Most are what is classified as suboxic, and introducing oxygen can be equally detrimental. Although anoxia may develop in suboxic sediments due to bacterial processes, these will be slowed at low temperatures, as will oxygen ingress.

Freezing is a common storage practice, however this can alter sediment structure, and rupture bacterial cells releasing their contents. For example, an order of magnitude increase in pore water selenium in pore waters from frozen compared with unfrozen sediments has been noted (Jung and Batley 2002). While this may not be a significant issue in many cases, its importance should be assessed, and if significant (or unknown), freezing should be avoided.

The container type will be important. Polyethylene and polypropylene containers are not impervious to oxygen, so are not recommended for prolonged storage of unfrozen sediments. Other plastics may contain dissolved oxygen (Teasdale et al. 1998), which will oxidize sediments that are in intimate contact.

Pore Water Handling and Storage

Recent studies (Simpson and Batley 2002) have shown that during about 30 min storage of estuarine pore waters at pH 7.7, some 63% of iron (II) is oxidized at 20°C, and even at 4°C there is almost 50% oxidation. In highly metal-contaminated estuarine pore waters, adsorptive losses of zinc, lead and copper to precipitating oxidized iron can be as high as 50% at 4 or 20°C over 24 h (Simpson and Batley 2002). However, federal agency guidance states two to eight weeks storage is acceptable, with testing as soon as possible recommended (Environment Canada 1994; US EPA 2002).

For sediment pore waters, early extraction is recommended, regardless of whether metals or organics are being studied. If pore waters are permitted to oxidize, and metals are precipitated, the tests and analyses on these pore waters will no longer be representative of the conditions in the field sediments.

Filtration of pore waters adds a further step that might enhance losses of metals and organics and perturb pore water chemistry, although for chemical measurements, knowledge of the truly dissolved form is highly desirable. In highly metal-contaminated waters, a few sediment particles have the potential to dramatically increase the measured 'dissolved' metal, as will sorbed organics increase 'soluble' organics. Extra care should be taken to filter pore waters under nitrogen and for analysis, acidify samples soon after filtration. Unfortunately, "dissolved" non-polar organic compounds tend to sorb to filters. Centrifugation might be an easier way to remove particles in samples where this is necessary for toxicity testing (this is particularly needed for algal testing).

Ideally, toxicity testing should be commenced immediately following extraction, to minimize physico-chemical changes, which are more likely in extracted pore waters than when whole sediments are present. Pore and SWI waters should ideally be stored in the dark at 4°C. The possibility of chemical changes to pore and SWI waters as noted above must be considered; oxidation must be minimized to the extent possible. The recommendations from the recent SETAC Technical Workshop (Carr et al. 2001) are to develop a

standard procedure to allow these reactions to go to completion, yet this begs the question of the relevance of this sample compared to what the organisms would naturally be exposed to. In the case of pore water-only toxicity testing, given the uncertainties, it may be better to undertake high-quality chemical investigations first on the freshly sampled and then on the fully oxidized pore waters, and predict the toxicity in the unaltered pore water from tests on the oxidized sample and known organism concentration.

Sediment Chemistry

While chemistry cannot be used alone as a line-of-evidence, its proper characterization is of paramount importance for arriving at accurate assessment conclusions on exposure-effect relationships, stressor interactions, fate predictions, and subsequent management decisions.

The same contaminant stressors should be measured in whole sediments, pore and SWI waters. In the case of pore and SWI waters, the following additional measurements are also desirable: pH, dissolved organic carbon, salinity, and other modifiers of contaminant bioavailability. In the case of whole sediments, modifiers that should be measured include grain size, total organic carbon, and AVS and simultaneously extracted metals (SEM).

Metals Bioavailability

The chemical characterization of sediments initially generally involves the measurement of contaminant stressor concentrations for assessing compliance with sediment quality guidelines, or determining contaminant presence/distribution. This is usually supported by measurements of grain size, redox potential, and key matrix elements and binding phases such as aluminum, iron, manganese and total organic carbon. Since not all of a contaminant's total concentration is bioavailable, it makes more sense to attempt to measure the bioavailable fraction. Measurements of total metal concentrations have little bearing on bioavailability. Chemical analysis methods exist that provide varying degrees of improvement over total metals analyses, although these have not yet progressed to the point that the bioavailable fraction can be specifically determined.

For metals, a range of extraction procedures has been advocated, including the use of EDTA, dilute acid, or acid containing pepsin as a digestive enzyme (Maher et al. 1992). The cold, dilute (1M) hydrochloric acid extraction that is used for AVS and SEM measurements (see below) is recommended (Simpson et al. 1998). One of us (GEB, unpublished results) has observed that in many silty sediments, where the more bioavailable, anthropogenic metals are deposited from the water column, 50 to 100% of zinc, lead, and copper are extractable with the dilute acid. Many investigators have found relationships between weak extractability of inorganic compounds and bioavailability (e.g., Ying et al. 1992).

Determination of AVS/SEM

Estimating whether metals in sediments may have adverse biological effects generally involves a determination of AVS and SEM. AVS (mainly as FeS) in sediments can serve as

a control on metal solubility and bioavailability in pore waters such that until SEM exceeds AVS, certain metals will not be bioavailable nor acutely toxic (Ankley et al. 1994; 1996). There are issues related to the behavior of non-acid-soluble sulfides such as those of Cu, Co, and Ni, in taking into account the excess of AVS (Simpson et al. 1998). If these do not release sulfide, and do not also release metals, then there is no impact on the AVS-SEM ratio from SEM-Cu (or Ni or Co) and any excess of AVS will still cope with pore water copper, reducing its toxicity. If however these metal sulfides are otherwise oxidized and the metals released, then the AVS-SEM value will be overprotective, that is, there will be a possible prediction of an excess of SEM over AVS. Similarly, coatings of sulfides on metal oxides in some field sediments could generate the same prediction of potential toxicity, when in fact there would not be any (Simpson et al. 1998; 2000a). Note that where field sediments are disturbed (e.g., during dredging), FeS will be easily oxidized (in contrast to other heavy metal sulfides), with a potential to increase toxicity.

It has always been stipulated that an excess of AVS over SEM can be used as a predictor of an absence of toxicity, and that where AVS is deficient there may be toxicity, but dependent on solution speciation. Because the uncertainties related to AVS considerations are type I errors, that is, falsely showing an excess of SEM over AVS, their use is still recommended in assessing metal bioavailability is sediments. There have however been other concerns raised over issues such as the spatial and temporal variability of AVS, its applicability in oxic and dynamic environments, and the fact that field validation has been limited to only a few metals.

AVS/SEM determinations can predict non-toxicity but cannot predict toxicity. One reason is that such determinations do not account for the additional binding capacity provided by organic carbon. The other reason is that SEM is a simplistic representation of the combined effects (speciation, bioavailability, toxicity) of the individual metals in any given metal mixture.

Organic Carbon Normalization

It is generally assumed that most organic contaminants will be bioavailable to sediment-ingesting organisms. Where pore water is the exposure route, the soluble concentration will be that which partitions into pore waters. Sediment quality values for organics are usually generated from water-only toxicity data, then applied using measured pore water organic concentrations and the assumption of equilibrium partitioning to calculate the sediment concentrations (e.g., Crommentuijn et al. 2002). Because binding of organics is preferentially to organic carbon, normalization to organic carbon is generally adopted so that the more organic carbon in the sediment, the less it is likely to partition to pore waters, so the more can be tolerated. Guidelines are often referred to 1% organic carbon. There are clearly upper and lower bounds within which this normalization is considered appropriate. It has been suggested that normalization applies to carbon contents >0.2% (Di Toro et al. 1991), while other studies have indicated an upper bound of 25 to 30% organic carbon (Kadeg et al. 1986).

It should be noted that organic carbon normalization is most appropriate for non-polar organics. We recommend that normalization be applied over the range 0.5 to 10% organic carbon. For very high organic carbon concentrations, the carbon may be as oil or tar

rather than metal binding substrates such as humic materials, and normalization fails. At lower organic carbon values, other physical and chemical factors influence the partitioning process (Kadeg et al. 1986), the normalized sediment quality values become extremely low, and given the uncertainties in the commonly used unnormalized values (for many effects data carbon is not reported), these are difficult to justify. For less than 0.5% carbon, considering the carbon as 0.5% is probably acceptable.

Sediment Quality Guidelines

Sediment quality guidelines (SQGs) comprise key uncertainties; for instance, correlations involving the presence or absence of species are equated with causation (von Stackelberg and Menzie 2002), and there is no established guidance for their usage in risk management (Nord 2001; Chapman et al. 2001). The most commonly applied SQGs are those derived from US effects (and in some cases also no effects) databases and from equilibrium partitioning considerations (Long et al. 1995; MacDonald et al. 1996). These were largely data for marine sediments, but have been expanded to include freshwater sediments (Smith et al. 1996). The US database is now extensive, but in its original concept it was a tool for ranking contaminated sediments. It includes contaminant concentrations and the associated toxicity effects based on a limited range of biological endpoints and test species. From the ranked data set, a lower 10th (or 15th) percentile and median concentration were calculated for each contaminant to derive numbers that were indicative of a low and high probability, respectively, of biological effects. These values were respectively termed effects range-low (ERL) and effects range-median (ERM) (Long et al. 1995). Subsequently, a data set that included a consideration of no-effects data was also developed (MacDonald et al. 1996), with the respective numbers being a threshold effects level (TEL) and a probable effects level (PEL). Agreement between the two sets of values was reasonable. A large effort was devoted to demonstrating how effective the upper numbers (ERMs and PELs) were in predicting biological effects. Not surprisingly the upper numbers were found to be good predictors and much better than the lower numbers, but it is important to emphasize that these SQGs do not infer cause-effect relationships. The critical question, however, was how useful is a number that is predictive of toxicity. Recent applications have stressed the value of using the lower numbers as a trigger for further investigation, along the lines of a weight-of-evidence approach (e.g., ANZECC/ARMCANZ 2000a). Further development of sediment quality values would ideally involve a much larger species range, such that statistical distribution of no effect concentrations (NECs) that is used to derive water quality guidelines, might be applied also to sediments, in order to protect a given percentage of species (Aldenberg and Slob 1993). This is more defensible than a guideline that is based on a lower 10th or 15th percentile of effects on a limited number of species, as is currently the case. However, whether this can ever be achieved is debatable.

Presently, SQG numbers should only be used for screening; their value for this purpose will be directly related to how site-specific their derivation has been. For instance, applying US-derived SQGs to other areas of the globe will result in increased uncertainties compared to using locally derived SQGs (Chapman et al. 1999; 2001).

Toxicity Testing

The ability of toxicity testing to provide reliable inputs into WOE sediment quality assessments will also require that areas of uncertainty in the testing protocols are adequately addressed. These include recognizing the effects on test responses of the physical characteristics of sediment particles, and of experimental manipulations such as sediment sieving or sediment spiking, on test responses. Interpretation of test results must also consider the type of the test organism, its exposure routes, and the relevance of laboratory tests to the field.

Grain Size Effects

Grain size effects have an important impact on some benthic fauna. Burrowing species appear to have an optimum sediment particle size that enables ease of burrowing. Experiences with burrowing shrimp (*Callianassa* spp.), for example (Batley and Peterson, unpublished data), indicate that sand/silt mixtures are preferred. In the finer grained sediments devoid of sand, the organisms were unable to burrow and appeared to die of exhaustion in the attempt. Similarly, the amphipod *Rhepoxynius abronius* is adversely affected if sediments are too fine (DeWitt et al. 1988). On the other hand, large particles are equally unattractive.

Ideally, control, reference, and test sediments should have similar grain sizes. At the very least, grain sizes must not adversely affect the test organisms. Likewise deposit feeders can only ingest a certain maximum particle size, and many of them in fact preferentially select fine (i.e., surface and therefore potentially contaminant-rich) particles.

Sieving

For some toxicity testing, removal of coarse particles has been recommended, together with homogenizing of the sample. This will clearly perturb the chemistry of both the sediment and pore waters, more so if this is done under oxic conditions; however, redox equilibria will reestablish as bacterial activity continues.

Sieving is not encouraged (Environment Canada 1994; US EPA 2002). However, if sieving is necessary to remove coarse particles and/or resident organisms, it should preferably be done in a glove box under nitrogen to minimize sediment chemistry changes. Sieved sediments should be allowed to age after disturbance before any testing, to allow redox equilibria to reestablish (Simpson and Batley 2002).

Sediment Spiking

Tests may be done with contaminated sediments collected from the field or with clean sediments that have been spiked with known concentrations of defined contaminants. The former approach is commonly applied in site-specific, retrospective risk assessments; the latter approach is used to assist in defining sediment quality values or to investigate bioavailability and subsequent toxicity of sediment-associated contaminants, e.g., pesticides (OECD 2001).

Various methods for spiking are used. The most common involve either spiking of the overlying water in a sediment-water system after the test organisms have been added (US EPA 2002), or spiking the sediment and subsequently adding the spiked sediment to test beakers followed by the addition of overlying water and test organisms (ASTM 1994). Which design is more appropriate for a given situation depends on the kind of scenario being simulated. For example, spiked water tests are often used to simulate pesticide spray drift events where the initial peak concentrations of pesticides in pore water are of concern.

Sediment spiking is frequently used to provide samples with a range of testable concentrations. However, these concentrations may far exceed contaminant concentrations typically found in impacted sediments. Lee et al. (2000) have pointed out that spiked sediment tests often involve abnormally elevated pore water concentrations, whereas at contaminant concentrations normally observed in heavily impacted sites, the pore water exposure pathway may not be dominant. Performing tests at environmentally realistic concentrations can help reduce uncertainties associated with concentration-dependent changes in partitioning and bioavailability.

Acid-volatile iron monosulfide is a major buffer for metal release from sediments. Metal additions will only result in detectable pore water concentrations once this capacity is exceeded. These additions can result in exchange not only with iron in FeS, but also with metals in other less stable sulfides (Simpson et al. 2000b). For example, for metal additions in excess of the binding capacity of FeS, copper additions displaced cadmium, and cadmium displaced zinc, consistent with the stability order Cu > Cd > Zn > Fe. This could result in cases where toxicity resulting from displacement of one metal might be mistakenly thought to be caused by another that is being added.

There is also an operational problem associated with the proper equilibration (aging) of organic or inorganic spike additions so that their form is similar to that of the contaminants in the natural sediment. A number of studies have shown the effects of "aging" on sequestration of compounds to the sediment (Alexander et al. 1997). For example, added metal spikes may take some time to reach the same crystalline form as is naturally present, and different forms are likely to have differing reactivities. One of us (GEB) allows metal additions to age for at least 14 days before sediments are used in further studies (Simpson et al. manuscript in preparation). For comparison, Berry et al. (1996) reported that after 9 days, pore water concentrations in their spiked sediments had decreased to a constant low value. In the later work of Bat and Raffaelli (1998), where spikes were mixed with sediments over four hours, the supernatant decanted and clean seawater added to equilibrate to test conditions, overlying water metal concentrations in the high spike cases reached several hundred μg/L. Not surprisingly these sediments were highly toxic to *Corophium volutator*. In reality, surface water concentrations of such magnitude are never found, although pore water concentrations may in the worst cases reach such levels, largely due to pH changes. Federal guidance suggests equilibration periods of at least two weeks (Environment Canada 1994; US EPA 2002).

Studies by Costa et al. (1996, 1998), using the amphipod, raise further concerns regarding spiked sediments. They reported a 10-day LC50 of 6.8 mg Cu/kg in copper-spiked

sandy sediments, although for water-only exposure the 96-h LC50 was 300 μg/L, and suggested that this represented multiple exposure routes as the pore water concentrations were low (around 20 μg/L). If the fine fraction in the sediment was increased from 0.5% to 25 and 50%, the LC50 increased, respectively, to 56.7 and 160 mg/kg, respectively (Costa et al. 1996). This reflects the increase in binding phases such as hydrous iron and manganese oxides (and to a lesser extent organic matter) in the fine fraction. No matching pore water toxicity data were supplied, and there were scant details on sediment and pore water chemistry, especially pH. The need to fully document chemistry in such studies cannot be too highly emphasized. Without this, the relative sensitivities of different species cannot be compared. In this example, further studies would be required to determine whether the result was an artifact of the spiking process.

Spiking of sediments or waters with organic compounds is also a problem, because to get a sufficient concentration of a hydrophobic organic in solution to spike either waters or sediments, it is necessary to improve its solubility by addition of an organic solvent (e.g., Murdoch et al. 1997; Bettinetti et al. 2002). Even with dilution, the presence of the solvent will affect the partitioning of the organic from sediment to water, thus altering the sample from its natural state. Recent studies on jet fuel toxicity in sediments found differences depended on whether a carrier solvent was used and if a subsequent volatilization period was allowed (Grasman and Burton, unpublished results). A wide variety of spiking methods have been reported in the literature, with no clearly superior approach recognized (ASTM 1994; OECD 2001; US EPA 2002).

Spiked sediment tests are also appropriate for investigating the potential for recolonization by benthos of previously contaminated sediments in the field. These studies have been conducted in freshwater and marine systems and provide a more realistic exposure (Morrisey et al. 1996; Hare and Shooner 1995).

Sediment spiking studies need to be done with care recognizing the need for adequate equilibration times and the reality of laboratory artifacts including possible metal displacement effects. Final contaminant concentrations (pore and SWI waters as well as whole sediment) need to be environmentally realistic, and different exposure routes adequately assessed. It is critical to fully document physico-chemical conditions prior to, during and after sediment spiking, for all exposure routes, if possible.

Physical and Chemical Properties of the Sediment and Overlying Water

As discussed previously, the physical and chemical properties of sediment and overlying water can have a critical influence on the behavior of chemicals that, in turn, can have a major effect on bioavailability, toxicokinetics and toxicity. If the aim is to determine the toxicity of field sediments from specific sites, then the problem is to ensure that differences in toxicity among sites are due to contamination and not to other properties of the sediment. The best approach in this instance may be to measure all of the relevant sediment properties and use multivariate analyses to separate the effects of the contaminants from effects of other potentially confounding variables. If the aim is to compare the relative toxicities of different chemicals to a species or to compare the relative toxicity of the same chemical to various species, it may be appropriate to use an artificial sediment with

defined and repeatable chemical and physical properties (Ribeiro et al. 1999). Artificial sediments have the advantage that specific sediment properties can be manipulated individually to explore their importance to contaminant partitioning, bioavailability and toxicity. The approach of mixing sand, clay, humus, etc. to create an artificial sediment may be useful for risk assessment of new chemicals to ensure comparability in the same way that standard artificial water, media, and test species are used, but the relationship to natural sediments may be questionable. An alternative approach might be to use uncontaminated natural sediments, but care should be taken that sufficient time is given for the equilibration and aging of added contaminants with the sediment phases to better mimic the field situation.

In toxicity testing, it is important in any test to be able to separate pore from SWI water exposure. It is unusual in the field to have high concentrations of toxicants in the overlying water arising from sediment release, and more usual to have higher concentrations in pore waters in association with low concentrations in the overlying water.

Test Species: Exposure Routes, Sensitivity, Residency

Testing as part of a weight-of-evidence investigation should concentrate on species that inhabit sediments. Such testing should also cover the range of uptake routes: sediment ingestion, pore water or surface water uptake and food uptake. Sediment quality databases contain toxicological data from a wide range of species, including effects on shrimp, oysters and other bivalves, *Daphnia* spp., amphipods and fish, oligochaetes and polychaetes, with testing involving whole sediments and pore waters (Long et al. 1995). More recent investigations have also considered SWI waters (Anderson et al. 2001; Greenberg et al. 2000). Whole sediment testing is a highly desirable approach, because of the diversity of uptake pathways. Pore and SWI water testing, on the other hand, only measure exposure by the water route. Appropriate test species for whole sediment testing might include burrowing amphipods, bivalves that feed in the sediments rather than overlying water, polychaetes, and even benthic algae. Using overlying water species bioassays for pore water testing, for example, sea urchin fertilization, might be appropriate in providing an early warning, but inappropriate in defining guidelines. Microtox® tests (Carr et al. 1996) or other tests that use solvent extracts of sediments are inappropriate. The latter are presumed to be attempting to extract the equivalent of lipid-soluble organics from a sediment, but the test species is not one that responds to sediment-bound organics, rather it measures water-only exposure. The effects measured will not, however, be related to water-soluble bioavailable organics. Correlations between effects with this test and uptake by natural benthic sediment ingestors would be necessary, otherwise the bioassay is merely being used as a surrogate of other (presumably) less sensitive species.

Test species should be appropriately sensitive, however caution should be used when basing cause-and-effect conclusions only on correlations. If species are used that are insensitive to a target toxicant (e.g., metals), it is possible that low concentrations of other contaminants present (e.g., PAHs) to which the organism is particularly sensitive, might trigger the toxic response, that could be mistakenly attributed to metal toxicity. Indeed, a

lot of information on species sensitivities is still lacking. Within each species class, there is likely to be a range of natural sensitivities to sediment or pore or SWI water contaminants, as there is for species in the overlying waters (McPherson and Chapman 2000). The burrowing estuarine amphipod *Corophium volutator* has been shown to be relatively insensitive to heavy metals in pore waters. Bat et al. (1998) found LC50 values of 21, 9.8, and 9.0 mg/L for Cu, Zn, and Cd in seawater. By contrast, the freshwater amphipod *Hyalella azteca* has an LC50 value for cadmium of only 2.8 μg/L, while for Cu and Zn the values were 31 and 73 μg/L, respectively, dependent on pH and hardness (these values were for 40 mg/L $CaCO_3$) (Phipps et al. 1995). The latter would seem an ideal test organism for metal testing; however, it is debatable whether the former is of any value as a test species, if pore water concentrations in the mg/L concentrations are required for a measurable effect. *Ampelisca abdita*, another saltwater amphipod species, gave 10-day LC50s of 36 μg Cd/L, 21 μg Cu/L, and 342 μg Zn/L, making it far more acceptable (Berry et al. 1996).

A Corophium amphipod species, being tested in Australia for estuarine sediments could be used in high-salinity waters (35‰) only if the salinity was increased gradually over several days from the naturally low salinity in sediments where the species was collected. This species then exhibited a higher sensitivity to metal contaminants than other amphipods found naturally in high salinity waters (Hyne et al. unpublished results). Since the salinity stress may have contributed to the increased sensitivity even though the mechanisms of toxicity for salinity and cadmium are probably different (Pechenik et al. 2001), the relevance of these test results is doubtful. This raises the issue of whether or not tests should use a nonresident sensitive species (as a bioavailability measure) to indicate a potential management issue, or should a possibly less sensitive species typical of the site be used. There is also the issue of whether or not tests should use such species if they had adapted to contaminants at the particular site. Adaptation can and does occur, but can vary considerably, both in time and space (Groenendijk et al. 1999). The use of contaminant-adapted species in sediment bioassays is not recommended.

An estuarine sediment toxicity test using benthic algal species has been developed (Adams and Stauber 2004). These species attach themselves to surface sediment particles, but their exposure route is via water (pore or overlying). Discriminating between pore water and overlying water exposure is difficult, but it can be reasoned that, at best, they will be responding to the pore waters effluxing from the sediment. As such, however, they may be exposed to the highest pore water contaminant concentrations as these are often found in the uppermost, oxic sediments. Benthic algae are an important part of the food chain, and as part of a suite of tests, might provide useful and different ecotoxicological information to that from other benthic test species.

Test species should ideally include taxa present in the reference areas. Exposure routes should be complete (e.g., not just pore water). Test species need to be appropriately sensitive, collectively, to the range of contaminants expected in the sediments, although this is not necessarily known *a priori* in all cases. "Appropriately sensitive" would not initially include adapted species, but could include such species during later analyses when the site-specific significance of more sensitive species tests is being assessed. Test species should also include a range of taxa (invertebrates and possibly vertebrates or plants as appropriate to the site and situation). The importance of ecological relevance in the choice

of test species, a knowledge of exposure pathways, and their relative sensitivities to contaminant concentrations by these pathways, is discussed in detail by Chapman (2002). For instance, Croteau et al. (2001) have shown differences in cadmium accumulation among species of *Chaoborus* in freshwater lakes related to differences in both rates of cadmium assimilation and feeding habits. In particular, it is important to include animals that live buried in and feed on sediments, for example, many annelids, since they can ingest and absorb very high concentrations of contaminants during feeding (Selck et al. 1998).

Field versus Laboratory Responses

An uncertainty common to all toxicity tests is the relationship between the field and laboratory. This is outside of any consideration of adaptation and/or acclimation of laboratory test species. De Witt et al. (1999) showed in tests with estuarine amphipods (*Chaetocorophium* cf. *lucasi*) exposed to cadmium-spiked sediments, that the laboratory sensitivity was equal to or greater than the sensitivity *in situ*. The reasons for this included the more realistic and environmentally variable conditions in the field, coupled with the impacts of differing levels of biological organization and larger spatial scales. Similarly, Kraaij et al. (2001) concluded that increased toxicity to PAHs in the laboratory compared to the field was related to increased bioavailability of PAHs in laboratory toxicity tests. However, Long et al. (2001) in a review of amphipod sediment toxicity test results and corresponding benthic infauna diversity and abundance, concluded that "ecologically relevant losses in the abundance and diversity of the benthic infauna frequently corresponded with reduced amphipod survival in the laboratory tests." However, in contrast, Kater et al. (2001), comparing laboratory and *in situ* responses of the amphipod *Corophium volutator*, concluded that "in most cases, the *in situ* response was significantly higher compared to the laboratory response."

As noted by Chapman (1995a,b), laboratory sediment toxicity tests cannot be directly compared to field responses because there is no certainty that field responses can or must always result from mechanisms, effects, and/or phenomena observed in the laboratory. As further noted by Chapman (2000) laboratory toxicity tests, compared with the field, may be either under- or overprotective, depending on a wide variety of factors. Expanding the realism of laboratory tests by conducting them in the field, for instance, using enclosures, creates other complications. Enclosure experiments may create artifacts related to predator:prey relationships, hydrodynamic effects, and variability in sediment chemistry.

Thus, laboratory (or field) toxicity test results, although they can be compared to resident benthic community structure data, provide a separate line of evidence. They do not require validation based on such data, and they should not be used alone for decision-making. Causation in toxicity test results is best determined by procedures such as determining contaminant body residues (CBRs) (see below), or by toxicity identification and evaluation (TIE) (Hunt et al. 2001; Ho et al. 2002).

Relevance of Test Exposure Scenarios to the Field

The timescale of exposure in ecotoxicological tests generally follows one of two protocols: exposure to high toxicant concentrations for short periods of time (i.e., so-called acute

tests of generally between 24 and 96 h), or exposure to lower concentrations for longer periods of time (i.e., so-called chronic tests that should extend for a substantial fraction of an organism's lifespan). In acute tests, the overlying water is either not changed (static test designs) or may be continuously (flow-through) or intermittently (static-renewal) exchanged with fresh water. Clearly test duration as well as exchange of overlying water can influence partitioning of contaminants between sediment, pore water, and overlying water compartments (Simpson and Batley 2002) and will influence the extent to which exposure is constant or variable during the course of the test. It is not always obvious which test design most closely simulates exposure in the field. However, rarely are exposures constant in the field, and studies have shown that pulse exposures may be more or less toxic. In addition, laboratory exposures do not allow for interactions of critical factors, such as suspended solids, temperature, and natural sunlight. One approach for dealing with this source of uncertainty is to express toxicity in relation to organism body burden (the CBR concept), rather than to sediment (or water) concentration (Lotufo 1998; Jarvinen and Ankley 1999). However, the CBR approach has not been field validated. CBR investigations involve determining the body burden of contaminants related to toxic effects, then using this information in a predictive mode (Borgmann and Norwood 1997; Borgmann et al. 2001).

Behavior of Test Organisms

The ingestion of contaminated sediment is an important route of uptake for some benthic organisms (Kaag et al. 1997; Selck et al. 1998; Bendell-Young 1999; Lee et al. 2000)). If this is an important potential route of exposure that must be investigated, it is essential that test animals are feeding actively during sediment toxicity testing. Ideally simultaneous measurements of feeding rates would also be conducted. However, this raises the question of whether clean or contaminated food should be added during the bioassays. Some benthic organisms exhibit different feeding methods, depending on food availability and other environmental conditions, which can have important consequences for contaminant uptake by different routes of exposure. For example, the brittlestar, *Amphiura filiformis*, may either filter feed or deposit feed on surface sediment depending on food availability in these two compartments and on whether or not there is a water current (Buchanan 1964). Such possible changes in feeding behavior need to be taken into account when designing and interpreting toxicity tests.

Extrapolation from Tested Species to all Relevant Species

Ecotoxicity tests for assessing sediment quality are generally conducted on a very small number of standard test species for which test protocols have been developed. There are two primary approaches typically used to extrapolate such test results to assess risk for all of the untested species in the systems of interest. The first approach involves dividing the measured test result by an application factor to account for uncertainty. If test results are available for several species, usually the application factor is applied to the most sensitive. Typically the size of the application factor decreases as the number of species tested increases and the closeness of the test design to the field situation increases. An al-

ternative approach is increasingly being used when toxicity data are available for more species (i.e., > four: van Leeuwen and Hermens 1995). This approach involves fitting the toxicity data to a parametric (Aldenberg and Slob 1993; Wagner and Løkke 1991; Shao 2000; Warne 2002) or nonparametric (Klaine et al. 1996) frequency distribution and estimating the percent species affected at relevant exposure concentrations. There are uncertainties involved in both of these approaches, largely related to how the test values reflect the actual sensitivity of intact communities.

The parametric approach has been applied in Australia to the derivation of water quality guidelines for toxicants and appears to work well provided there are sufficient data points (five or more species in four 4 or more taxonomic groups) and if the data are not forced to fit a log-logistic distribution (ANZECC/ARMCANZ 2000a). Guideline application however involves consideration of community level impacts and ecotoxicology. Extrapolating the parametric approach to sediment toxicity testing is not yet possible, because of the limited number of species that have been tested.

Applying a fixed factor to account for species differences in sensitivity to chemicals is uncertain because, even if an appropriate average factor can be determined, interspecies variation in sensitivity is chemical dependent. Attempts to improve this approach by defining different factors for different classes of chemicals (Roex et al. 2000) may help to some extent but do not remove the uncertainty entirely (Forbes and Calow 2002a). Although species sensitivity distribution methods appear to be a more scientifically based approach in theory, problems with their application may result in effects thresholds that are no more accurate predictors of likely effects of chemicals on field populations than the fixed factor approach (Forbes and Calow 2002b).

Extrapolation from Measured Test Endpoints to the Likelihood of Population-Level Impacts

Responses measured in laboratory toxicity tests are typically individual-level survival, reproduction and/or growth, yet sediment quality should ideally be defined in terms of its effect on the likelihood of population persistence of aquatic organisms (or their consumers). The extent to which individual-level responses are translated into population-level impacts depends (at the very least) on the life cycle of the tested species, on which individual-level responses are altered, and on the demographic starting point of the population (i.e., growing, shrinking, stable) (Calow et al. 1997). Such extrapolations are confounded by unexpected differences in the nature and magnitude of responses at the individual compared to the ecosystem-level (e.g., delayed response due to compensating factors followed by dramatic population declines. Analyses to date suggest that impacts of chemicals on individual performance are often greater than or equal to impacts on population growth rate, however, further work is needed in this area (Forbes and Calow 1999; Maltby et al. 2001). Population models can provide powerful tools to explore the likely linkages between individual and population-level effects and to enable predictions concerning factors that are likely to exacerbate or ameliorate effects of chemicals on population dynamics (Forbes et al. 2001).

Benthic Community Structure

Spatial and Temporal Sampling Scales

A common challenge in designing benthic community sampling programs is determining the optimum partitioning of samples between number of stations and number of replicates per station. Denmark has recently undertaken an extensive review of its national benthic monitoring program (Nørrevang, personal communication). An important conclusion from this analysis was that, for the purpose of general monitoring, taking fewer samples at more stations appears to be more effective at detecting changes in community structure than taking more replicate samples at fewer stations. However, the ideal sampling scheme may vary depending on whether the sampling is performed for routine monitoring purposes or whether it is directed toward a specific point source. For the latter, it may be more effective to sample more intensively at fewer stations.

For situations in which time trends of changes in the benthic community are of concern, the temporal scale of sampling is critical. Seasonal changes in natural population dynamics of benthic species need to be accounted for, either by sampling several times during the year, or (at the very least) sampling at the same time every year. If sampling is to be done only once annually, it should be done well outside of the main reproductive period of the relevant organisms, since slight differences in timing of reproduction among years can lead to large uncertainties if sampling is performed close to the reproductive period. When analyzing temporal changes in community structure it is important to determine whether there have been any changes in sampling techniques or equipment used (Pearson et al. 1985; Rosenberg et al. 1987).

Separating and Identifying Organisms

Variations between benthic community field and laboratory methods are common and sometimes extreme (e.g., samplers and sieve sizes, operator-specific sample collection, subsampling by different methods or none, the use or not of different types of magnification during sorting, and different levels of taxonomic identification) and can affect the determination of environmental alteration or impairment of the benthos (Carter and Resh 2001).

Collecting benthic infaunal organisms from sediments for community structure analyses basically involves retrieving whole sediments then sieving them to remove finer materials and make it easier to pick out organisms in the laboratory. Sieving is generally, but not always, done on site.

The choice of sieve size is critically important, depending on the organisms present and the focus of any investigation. Typically either 1.0- or 0.5-mm sieves are used. These sieve sizes are adequate for larger infauna, but not for smaller infauna which may represent a significant portion of the individuals and species present, and which may also be important prey for higher organisms such as fish (e.g., oligochaete worms, nematodes, harpacticoid copepods). These sieve sizes are also not adequate for juvenile stages of many large infauna (e.g., clams such as *Macoma*). Moreover, sieving effort may not be uniform depending on the length of hours worked by a sampling crew (towards the end of a long day techniques may become more "sloppy"). Whether or not smaller organisms are retained on larger sieves is a function of the amount of sieving done, the presence of or-

ganic matter onto which those organisms may attach, and the orientation of the organisms (worms in vertical aspect may pass through sieves, but generally not in horizontal aspect). Sieving preserved organisms tends to result in lower retention rates because rigid and unmoving organisms are more likely to pass through the holes in the sieve, thus initial sieving is usually done with live organisms with any subsequent sieving involving smaller sieve sizes.

Choice of sieve sizes should be dictated by the study objectives (e.g., collecting only adult organisms, or also collecting juveniles), and by the key species present. Thus, an initial reconnaissance to determine species present coupled with conceptual diagrams of key species relationships, should be considered rather than simply using "set" sieve sizes that may be inappropriate.

Species Identifications

Species are determined in two different and sometimes mutually exclusive ways. A species is traditionally defined as related organisms potentially capable of interbreeding. However, for most sediment infauna (and other invertebrates), the potential for interbreeding is never assessed. Instead, species are typically defined based on any published designation and are entirely subjective (Shaw 1969).

Some "species" have been found not to be distinct species while others have been found to be groupings of "sibling species." For instance, Chapman and Brinkhurst (1987) changed "species" of oligochaetes described on the basis of their external hairs and setae into other "species" by changing water quality conditions (hardness, salinity, pH). The polychaete worm *Capitella capitata*, comprises a variety of "sibling species" (Grassle and Grassle 1976) that have been found to vary in both life-history features and physiological tolerance (Linke-Gamenick et al. 2000a,b).

Thus, species lists compiled for benthic community structure assessments may not be accurate based on the true definition of a "species." The reliability of species determinations for key and keystone species remains to be tested; in the meantime, conducting assessments based on both species and higher level taxonomic classifications is an alternative, as is the option of examining functional rather than structural units. However, functional groupings are not necessarily easy to determine (Mermillod-Blondin et al. 2001).

Many benthic invertebrates can be difficult to identify to species level, and doing so may require significant effort in time and expertise. Interestingly, several studies have indicated that community structure measures based on higher taxonomic levels than species (i.e., family level), may be more effective at detecting change than measures based on individual species (Gray et al. 1988; Hewlett 2000); however, others have shown genus-level identifications to be superior (Hawkins and Vinson 2000). The optimal taxonomic level may depend to some extent on the variation in sensitivity within the respective taxonomic levels, the taxonomic richness of the benthic system under study, and the geographic scale of coverage.

Choice of Metric for Expressing Community Structure

A number of metrics or indices have been developed by ecologists for quantifying community structure and changes thereof (US EPA 1999). Older methods include the Shan-

non-Wiener diversity index, Fisher's a, Simpson's index of concentration, and other less well-known indices (Rosenzweig 1995). For benthic communities, the Shannon-Wiener index and ABC curves of abundance/biomass (Warwick 1986) have been among the most commonly used univariate approaches though a wide variety of others have been developed for use in marine pollution studies (Clark 2001). However, such aggregate numerical indices are limited in that they do not incorporate qualitative features of the species included in them. They need to be interpreted with great care (French and Lindley 2000) and can provide misleading results where pollution is due to complex chemical mixtures (Washington, 1984). In addition to these traditional methods, a number of multivariate methods have been developed to examine the relationship between community structure and sediment variables. These appear to be more sensitive than the univariate methods at detecting effects of contaminants on benthic sediment communities (Gray et al. 1990), particularly since simplistic methods such as diversity indices can give misleading results (Washington 1984).

Considering Exposure Routes

As for toxicity test species, it is important to determine whether key species used for benthic community structure evaluations are exposed to pore waters or to SWI via burrow irrigation. The possibility of sediment ingestion as an exposure route also should be evaluated for such species as discussed earlier.

For example, some burrowing species are effectively exposed to SWI waters rather than pore waters because of effective burrow irrigation (Warren et al. 1998). Such species are likely to give a poor measure of sediment toxicity, unless there is significant efflux of contaminants from the sediment. This can occur in enclosed laboratory test systems, providing an unrealistic measure of sediment toxicity compared to real-world conditions. In many sites, upwelling and downwelling conditions may exist. The upwelling and downwelling of groundwater or surface water through sediments alters benthic exposure, reducing exposure to static pore water and whole sediments, and elevating exposure to the respective compartment (i.e., groundwater or surface water), so that benthic organisms are primarily exposed to either deeper groundwaters or overlying surface waters (Greenberg et al. 2000). Thus, matching up chemical analyses with the correct compartment is essential to be able to match exposure with effects, Without documenting this exposure (such as with mini-piezometers) it is impossible to accurately compare chemistry of water, sediments or pore waters with biological effects.

Other sediment species (e.g., most chironomids) live in tubes lined with transparent fibrous salivary secretions (Jonasson 1972). Such species have limited contact with pore waters outside their tubes. Some bivalve species (e.g., the cockle *Cerastoderma edule*) inhabit sediments and feed from the overlying water (Absil et al. 1996). Other bivalve species (e.g., the clam *Macoma balthica*) are facultative feeders and can change their food source if required (Stecko and Bendell-Young 2000). What the food source is, for example, algae or other meiofauna in the water column or in or on the sediments, and the relative importance of this versus pore water or sediment particles as a source of toxicants, is critical to detecting biological impacts with test species. Experiments using radiotracers (e.g., Wang

and Fisher 1999) have demonstrated the relative importance of these pathways and more such studies are needed for key species.

As for toxicity test species, for key (e.g., dominant or keystone) species used in benthic community structure assessments, basic information is required regarding the pathways by which they accumulate toxicants, and the relative sensitivities of these species to contaminant concentrations by these pathways. This needs to be done site-specifically, but generic information should also be compiled and published for general use.

Stressor Identification

Stressors may be physical, chemical and in some instances biological. They may be toxic (e.g., non-essential metals), nontoxic (e.g., nutrients), or both (e.g., essential metals). Stressors may also act in a direct or indirect manner. Indirect stressors (e.g., pH, organic carbon, food) modify the effects of potentially direct stressors (e.g., metals). Physical stressors include light, temperature, flow, and substrate type. The latter are considered as natural or habitat stressors. Note for ecological studies these habitat stressors will be of particular concern and their spatial (and temporal) variability will need to be characterized. Grab sampling on one occasion and using single site composites for physicochemical, toxicity, or benthological characterization introduces significant levels of uncertainty.

The effects of all stressors must be considered on both spatial and temporal scales. Deconvoluting their interactions is important as part of cause-and-effect determinations. These will generally also involve determining the contributions of individual toxicants present in mixtures of different contaminants within sediments. Laboratory studies will generally be necessary for such determinations, using methods such as the US EPA's Toxicity Identification Evaluation approach using either pore waters or whole sediments (Ho et al. 2002; US EPA 1991). Such determinations for water column organisms (e.g., Barata and Baird 2000) need to be applied to sediments. *In situ*-based approaches for separating stressors have also been used to a limited extent, allowing separation of stress from suspended solids, photo-induced toxicity, metals, organics and ammonia (Ireland et al. 1996; Sasson-Brickson et al. 1991; Burton and Moore 1999; Burton et al. 2002). These field approaches have been shown to more accurately detect toxicity and stressors than laboratory-based methods.

Confounding Variables

Correlations between benthic community structure and sediment contaminant concentrations may indicate cause-effect relationships but may also occur as a result of correlations between contaminants and other sediment properties to which the benthic species are responding. Scatterplot matrices plotting any of the measured sediment attributes against each other and against community structure measures can provide a rapid assessment of whether confounding correlations may be important. Multivariate approaches have an advantage in this regard, in allowing a large number of variables to be condensed to the most significant variable, but, as discussed by Forbes and Calow (2002c), it is not appropriate to ascribe causation on the basis of correlation alone; other information (e.g., from

controlled exposures) is necessary. If important stressors, or interacting variables, are not properly characterized and linked correctly with receptor exposures, then correlations may be misleading and type I and II errors are possible.

Interpreting Changes in Structure

Ultimately, the largest source of uncertainty in benthic community assessments is likely to be in their ecological interpretation. There are fundamental biological questions (e.g., relationships between community structure and ecosystem processes) for which we usually only have partial understanding, but that require decisions to be made as to what it is about benthic communities that we want to protect. From purely an ecological perspective, can arguments be made for weighting all species or taxa equally, or should more weight be given to some taxa (e.g., functional keystones (Levinton 1995))? In terms of the key processes in benthic sediments, can we identify that degree of change in community structure that is 'unacceptable'? In other words, at what point should management action be triggered to protect the ecology of the system? Given that complete ecological understanding of these issues is lacking, a practical approach may be to consider any statistically detectable impact on community structure as ecologically undesirable. There are still discussions as to whether statistics are the criterion when biological/environmental evidence of an effect is available (e.g., expert judgment) (Johnson 1999; Paine 2002). Naturally, doing so requires full consideration of the power of the analysis and the type I and type II errors involved.

Data Uncertainties and Quality Assurance/Quality Control

Data Uncertainties

Despite the problems associated with the generation of reliable chemical, ecotoxicological, and benthic data discussed above, inherent uncertainties often fail to be reflected in the published results. While explicitly recognizing uncertainties does not make a difference to the big picture, it acknowledges that the author is at least conscious of the uncertainties, which is not always the case. Detailed recommendations for managing data uncertainties are provided by Crumbling et al. (2001).

Similar problems apply to sediment quality values (SQGs). The ERL/ERM and TEL/PEL values are frequently quoted with too many significant figures. Indeed for many reasons, including those touched on above, the precision of these SQGs is inherently poor. Smith et al. (1996) in deriving preliminary TELs and PELs for freshwater sediments, considered that agreement with other reported values was 'good' when the values were 'within a factor of three of at least two other assessment values'. Environment Canada, nevertheless, has published interim guidelines for zinc in freshwater and marine waters (online at http://ceqg-rcqe.ccme.ca/download/en/247) of 123 and 124 mg Zn/kg, respectively, clearly not reflecting the uncertainties implied above. The significant figures on many of their numbers are not defensible (cf. their lead guideline in marine waters of 30.2 mg/kg and freshwater of 35.0 mg/kg).

Quality Assurance/Quality Control (QA/QC)

QA/QC should be an essential component of chemical, ecotoxicological and benthic community structure measurements, especially with respect to uncertainty, but is frequently overlooked. Protocols for QA/QC are well established (OECD 1998; ANZECC/ARMCANZ 2000b), and apply to sampling and sample handling, as well as to chemical and biological analyses. For example, in sampling, equipment and container blanks will be important, as will duplicate samples to reveal the magnitude of errors. For chemical analyses, the use of certified reference sediments is highly desirable. In the case of toxicity testing, the use of positive controls (reference toxicants) is recommended, but the use of spiked samples, as used in chemical analysis, is of dubious value because of the difficulties in knowing the chemical form and bioavailability of the spike, as discussed earlier. Nevertheless, there remain situations where adequate quality controls are difficult to devise and apply and are frequently omitted (e.g., particularly in toxicity testing). Problems arise when dealing with multiple stressors. These may be physical (grain size) or chemical (multiple toxicants). The former should be dealt with by appropriate controls, while the latter may require more imaginative approaches.

Conclusions

The above discussions have highlighted a number of potential uncertainties in experimental design, measurement, and interpretation that can have serious implications for sediment quality assessments. These are summarized in the checklist in Table 1 together with suggested solutions. Note that the recommended solutions should not be considered absolute; in many cases there may need to be a compromise between what is desirable and what is achievable. However, use and adaptation of this generic checklist will allow for better sediment quality assessments than are generally found in technical reports or even in the published literature.

Acknowledgments — This manuscript is a product of the Weight-of-Evidence Workshop held in Madrid, Spain, in May 2001. The Madrid workshop was supported by the International Lead Zinc Research Organization, the International Copper Association, the Nickel Producers Environmental Research Organization, the Cobalt Development Institute, and Rio Tinto Borax. The Society of Environmental Toxicology and Chemistry provided logistical support.

Table 1. Checklist for evaluating data uncertainties/limitations

Parameter	Sources of Uncertainty	Recommendations
Sampling, Transport and Storage	Choice of reference sites	Ensure physico-chemical and biological characteristics similar between reference and exposed sites; use multiple reference sites
	Sediment heterogeneity	Spatial averaging appropriate to study purposes and chemical and biological measurements
	Sediment and pore water sampling depths	Depth-integrated samples; depth based on study objectives, physic-chemical and particularly biological realities; same depths for chemistry, toxicity and biology; note that AVS is depth dependent, so sample homogenizing over large depths gives misleading results
	Pore water sampling methods	No "best" method, minimize oxidation, sample under nitrogen gas atmosphere
	Obtaining sufficient pore water	*In situ* peepers best for minimally changed chemistry; centrifugation best for rapidly generating large volumes
	Sample storage	Store cool in the dark, excluding oxygen; test as soon as possible, report storage time and conditions
	Changes prior to testing/ analysis	Take all reasonable precautions; recognize possibility of such changes, some of which can be predicted from knowledge of sediment physic-chemical characteristics
Sediment Chemistry	Appropriate measurements	Measure all contaminants of potential concern, and key modifiers (e.g. – water, pH, DOC, sediment – AVS/SEM, grain size, TOC)
	Metals bioavailability	Measure easily extractable metals (e.g. cold IM HCl), not total metals, plus parameters that affect metal bioavailability (e.g. AVS)
	AVS/SEM comparisons	A useful qualifier of metal bioavailability, but possibility of type I errors for some metals; note possible depth dependence
	Organics bioavailability	Consider organic carbon normalization; quantify bioavailability directly by measuring bioaccumulation in biota
	Organic carbon normalization	Recommended that organic carbon normalization be applied to non-polar organics only over the range 0.5-10% organic carbon
	Sediment quality guidelines	Use only for screening and not to infer cause of toxicity
Ecotoxicology	Grain size effects	Similar grain sizes for references and test sediments; grain sizes must not adversely affect test organisms
	Sieving – removal of coarse particles	Not encouraged; if necessary, do under nitrogen and allow redox equilibria to re-establish
	Test species; exposure routes, sensitivity, residency	Appropriately sensitive range of ecologically important taxa including species from reference areas; exposure routes complete
	Sediment spiking	Adequate equilibration times; recognize laboratory artifacts; environmentally realistic concentrations; complete exposure routes; full physic-chemical documentation
	Laboratory vs. field	Separate lines of evidence; one does not validate the other. Do not use alone for decision-making
	Cause-and-effect	Correlative (e.g. gradient) analyses coupled with TIE or CBR determinations

Table 1. Checklist for evaluating data uncertainties/limitations *(continued)*

Parameter	Sources of Uncertainty	Recommendations
Benthic Community Structure	Spatial and temporal scales	Tailor to purpose; ensure common sampling protocols in comparing temporal data
	Sieve sizes	Based on study objectives and key species present/species relationships
	Species identifications	Species designations may not be reliable; conduct concurrent higher level taxonomic community structure assessments and/or consider functional groupings
	Exposure routes	Must be known for at least key (dominant) and keystone species, along with relative sensitivities via these routes
	Stressor identification	Physico-chemical and biotic stressors muct be characterized, including interactions (spatial and temporal)
Data Uncertainties and QA/QC	Data—significant figures	Should be considered; generally no more than two significant figures are warranted or defensible
	Data analysis	Estimate Type II errors for chosen design
	QA/QC	Ensure that QA/QC are adequately addressed

References

Absil MCP, Berntssen M and Gerringa LJA. 1996. The influence of sediment, food and organic ligands on copper uptake by sediment-dwelling bivalves. *Aquat Toxicol* 34: 13-29.

Adams MS and Stauber JL. 2004. Development of a whole-sediment toxicity test using a benthic marine microalga. *Environ Toxicol Chem* 23: 1,957-1,968.

Aldenberg T and Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotox Environ Safety* 25: 48-63.

Alexander M, Hatzinger PB and Kelsey JW, 1997. Sequestration and realistic risk from toxic chemicals remaining after bioremediation. *Ann NY Acad Sci* 829: 1-5.

Anderson BS, Hunt JW, Phillips BM, et al. 2001. Influence of sample manipulation on contaminant flux and toxicity at the sediment-water interface. *Mar Environ Res* 51: 191-211.

Ankley G, Thomas N, Di Toro D, et al. 1994. Assessing the bioavailability of metals in sediments: A proposed approach. *Environ Manag* 18: 331-337.

Ankley GT, Di Toro DM, Hansen DJ, et al. 1996. Technical basis and proposal for deriving sediment quality values for metals. *Environ Toxicol Chem* 15: 2,056-2,066.

ANZECC/ARMCANZ. 2000a. *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*. Australian and New Zealand Environment and Conservation Council, Ministry for the Environment, Wellington, New Zealand and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia.

ANZECC/ARMCANZ. 2000b. *Australian Guidelines for Water Quality Monitoring and Reporting*. Australian and New Zealand Environment and Conservation Council, Ministry for the Environment, Wellington, New Zealand and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia.

ASTM (American Society for Testing and Materials). 1994. *Standard Guide for Collection, Storage, Characterization, and Manipulation of Sediments for Toxicological Testing*. E 1391-94. Philadelphia, Pennsylvania, USA.

Barata C and Baird DJ. 2000. Determining the ecotoxicological mode of action of chemicals from measurements made on individuals: results from instar-based tests with *Daphnia magna* Straus. *Aquat Toxicol* 48: 195-209.

Bat L and Raffaelli D. 1998. Sediment toxicity testing: a bioassay approach using the amphipod *Corophium volutator* and the polychaete *Arenicola marina*. *J Exp Mar Biol Ecol* 226: 217-239.

Bat L, Raffaelli D and Marr IL. 1998. The accumulation of copper, zinc and cadmium by the amphipod *Corophium volutator*. *J Exp Mar Biol Ecol* 223: 167-184.

Batley GE. 1999. Quality assurance in environmental monitoring. *Mar Pollut Bull* 39: 23-31.

Bendell-Young LI. 1999. Application of a kinetic model of bioaccumulation across a pH and salinity gradient for the prediction of cadmium uptake by the sediment dwelling Chironomidae. *Environ Sci Technol* 33: 1,501-1,508.

Berry WJ, Hansen DJ, Mahoney JD, et al. 1996. Predicting the toxicity of metal-spiked laboratory using acid-volatile sulfide and interstitial water normalizations. *Environ Toxicol Chem* 15: 2,067-2,079.

Bettinetti R, Cuccato D, Galassi S, et al. 2002. Toxicity of 4-nonylphenol in spiked sediment to three populations of *Chironomus riparius*. *Chemosphere* 46: 201-207.

Borgmann U and Norwood P. 1997. Identification of the toxic agent in metal-contaminated sediments from Manitouwadge Lake, Ontario, using toxicity-accumulation relationships in *Hyalella azteca*. *Can J Fish Aquat Sci* 56: 1,494-1,503.

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

Buchanan JB. 1964. A comparative study of some features of the biology of *Amphiura filiformis* and *Amphiura chiajei* (Ophiuroidea) considered in relation to their distribution. *J Mar Biol Assoc* 44: 565-576.

Burton GA Jr. 1991. Assessing freshwater sediment toxicity. *Environ Toxicol Chem* 10: 1,585-1,627.

Burton GA Jr. 1992. Sediment collection and processing: factors affecting realism. *In*: Burton GA Jr (editor), *Sediment Toxicity Assessment*, p. 37-66. Lewis Publishers. Boca Raton, Florida, USA.

Burton GA Jr. 1995a. Quality assurance issues in assessing receiving waters. *In*: Herricks EE (editor), *Stormwater Runoff and Receiving Systems: Impact, Monitoring, and Assessment*, p. 275-284. Lewis Publishers, Boca Raton, Florida, USA.

Burton GA Jr. 1995b. Critical issues in sediment bioassays and toxicity testing. *J Aquat Ecosystem Health* 4: 151-156.

Burton GA Jr and Moore L. 1999. *An Assessment of Stormwater Runoff Effects in Wolf Creek, Dayton, OH. Final Report*. City of Dayton, Ohio, USA.

Burton GA Jr, Pitt R and Clark S. 2000. The role of whole effluent toxicity test methods in assessing stormwater and sediment contamination. *CRC Crit Rev Environ Sci Technol* 30: 413-447.

Burton GA Jr, Rowland CD, Greenberg MS, et al. 2002. A tiered, weight-of-evidence approach for evaluating aquatic ecosystems. *In*: Munawar M (editor), *Sediment Quality Assessment: Approaches, Insights and Technology for the 21st Century*. Ecovision World Monograph Series, Backhuys Publishers, Leiden, The Netherlands (in press).

Calow P, Sibly RM, and Forbes VE. 1997. Risk assessment on the basis of simplified life-history scenarios. *Environ Toxicol Chem* 16: 1,983-1,989.

Carr RS, Nipper M, Adams WJ, et al. 2001. *Summary of SETAC Technical Workshop, Pore Water Toxicity Testing: Biological, Chemical and Ecological Considerations with a Review of Methods and Applications, and Recommendations for Future Areas of Research*. SETAC Press, Pensacola, Florida, USA.

Carr RS, Long ER, Windom HL, et al. 1996. Sediment quality assessment studies of Tampa Bay, Florida. *Environ Toxicol Chem* 15: 1,218-1,231.

Carter JL and Resh VH. 2001. After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. *J N Am Benthol Soc* 20: 658-682.

Chapman PM. 1995a. Extrapolating laboratory toxicity results to the field. *Environ Toxicol Chem* 14: 927-930.

Chapman PM. 1995b. Do sediment toxicity results require field validation? *Environ Toxicol Chem* 4: 1,451-1,453.

Chapman PM. 2000. Whole effluent toxicity testing–usefulness, level of protection, and risk assessment. *Environ Toxicol Chem* 19: 3-13.

Chapman PM. 2002. Integrating ecology and benthic ecology: Putting the eco back into ecotoxicology. *Mar Pollut Bull* 44: 7-15.

Chapman PM and Brinkhurst RO. 1987. Hair today, gone tomorrow (induced setal changes in tubificid oligochaetes). *Hydrobiologia* 155: 45-55.

Chapman PM and Wang F. 2001. Assessing sediment contamination in estuaries. *Environ Toxicol Chem* 20: 3-22.

Chapman PM, Allard PJ and Vigers GA. 1999. Development of sediment quality values for Hong Kong Special Administrative Region: A possible model for other jurisdictions. *Mar Pollut Bull* 38: 161-169.

Chapman PM, Birge WJ, Adams WJ, et al. 2001. Sediment quality values (SQVs) – challenges and recommendations. *SETAC Globe* 2(2): 24-26.

Chapman PM, Wang F, Germano JD, et al. 2002. Pore water testing and analysis: the good, the bad, and the ugly. *Mar Pollut Bull* 44: 359-366.

Clark RB. 2001. *Marine Pollution*, 5th edition. Oxford University Press, Oxford, UK.

Costa FO, Correia AD, and Costa MH, 1996. Sensitivity of a marine amphipod to non-contaminant variables and to copper in the sediment. *Ecologie* 27: 269-276.

Costa FO, Correia AD, and Costa MH, 1998. Acute marine sediment toxicity: a potential new test with the amphipod *Gammarus locusta*. *Ecotox Environ Safety* 40: 81-87.

Crommentuijn T, Sijm D, de Bruijn J, et al. 2002. Maximum permissible and negligible concentrations for some organic substances and pesticides. *J Environ Manag* 58: 297-312.

Croteau M-N, Hare L, and Tessier A. 2001. Differences in Cd accumulation among species of the lake-dwelling biomonitor *Chaoborus*. *Can J Fish Aquat Sci* 58: 1,737-1,746.

Crumbling DM, Groenjes C, Lesnik B, et al. 2001. Managing uncertainty in environmental decisions. *Environ Sci Technol* 35: 404A-409A.

De Witt TH, Ditsworth GR, and Swartz RC. 1988. Effects of natural sediment features on survival of the Phoxocephalid amphipod, *Rhepoxynius abronius*. *Mar Environ Res* 25: 199-224.

De Witt TH, Hickey CW, Morrisey DJ, et al. 1999. Do amphipods have the same concentration response to contaminated sediment *in situ* as *in vitro*? *Environ Toxicol Chem* 18: 1,026-1,037.

Diamond J, Burton GA Jr, and Scott J. 2002. Sediment collection, manipulation and characterization for toxicity testing. *In*: Whittemore R (editor), *Sediment Contamination*. Water Environment Federation, Fairfax, Virginia, USA (in press).

Di Toro DM, Zarba CS, Hansen DJ, et al. 1991. Technical basis for establishing sediment quality for nonionic chemicals using equilibrium partitioning. *Environ Toxicol Chem* 10: 1,541-1,583.

Doe KG, Burton, GA Jr, and Ho KT. 2003. An overview of pore water toxicity testing. *In*: Carr RS, Nipper M (editors), *Porewater Toxicity Testing: Biological, Chemical and Methodological Considerations*. SETAC Press, Pensacola, Florida, USA.

Environment Canada. 1994. *Guidance Document for Preparation of Sediments for Physicochemical Characterization and Biological Testing*. Environment Protection Series, Report EPS 1/RM/29. Ottawa, Canada.

Forbes VE and Calow P. 1999. Is the per capita rate of increase a good measure of population-level effects in ecotoxicology? *Environ Toxicol Chem* 18: 1,544-1,556.

Forbes VE and Calow P. 2002a. Extrapolation in ecological risk assessment—balancing pragmatism and precaution in chemical controls legislation. *BioScience* 52: 249-257.

Forbes VE and Calow P. 2002b. Species sensitivity distributions revisited: a critical appraisal. *Human Ecol Risk Assess* 8: 1,625-1,640.

Forbes VE and Calow P. 2002c. Applying weight of evidence in retrospective ecological risk assessment when quantitative data are limited. *Human Ecol Risk Assess* 8: 1,625-1,640.

Forbes VE, Sibly RM and Calow P. 2001. Toxicant impacts on density-limited populations: a critical review of theory, practice, and results. *Ecol Appl* 11: 1,249-1,257.

French D and Lindley D. 2000. Exploring the data. In: Sparks T (editor), *Statistics in Ecotoxicology*, p. 33-68. John Wiley & Sons, Chichester, UK.

Grassle JP and Grassle JF. 1976. Sibling species in the marine pollution indicator *Capitella* (Polychaeta). *Science* 192: 567-569.

Gray JS, Aschan M, Carr MR, et al. 1988. Analysis of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment. *Mar Ecol Prog Ser* 46: 151-165.

Gray JS, Clarke KR, Warwick RM, et al. 1990. Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. *Mar Ecol Prog Ser* 66: 285-299.

Greenberg MS, Burton, GA Jr, and Duncan PB. 2000. Considering groundwater-surface water interactions in sediment toxicity assessment. *SETAC Globe* 1(2): 42-44.

Groenendijk D, van Opzeeland B, Dionisio Pires LM, et al. 1999. Fluctuating life-history parameters indicating temporal variability in metal adaptation in riverine chironomids. *Arch Environ Contam Toxicol* 37: 175-181.

Hansen DJ, Berry WJ, Mahoney JD, et al. 1996. Predicting the toxicity of metal-contaminated field sediments using interstitial concentration of metals and acid-volatile sulfide normalizations. *Environ Toxicol Chem* 15: 2,080-2,094.

Hare L and Shooner F. 1995. Do aquatic insects avoid cadmium-contaminated sediments? *Environ Toxicol Chem* 14: 1,071-1,077.

Hawkins CP and Vinson MR. 2000. Weak correspondence between landscape classifications and stream invertebrate assemblages: implications for bioassessment. *J N Amer Benthol Soc* 19: 501-517.

Hewlett R. 2000. Implications of taxonomic resolution and sample habitat for stream classification at a broad geographic scale. *J N Am Benthol Soc* 19: 352-361.

Ho KT, Burgess RM, Pelletier MC, et al. 2002. An overview of toxicant identification in toxic sediments and dredged materials. *Mar Pollut Bull* 44: 286-293.

Hunt JW, Anderson BS, Phillips BM, et al. 2001. A large-scale categorization of sites in San Francisco Bay, USA, based on the sediment quality triad, toxicity identification evaluations, and gradient studies. *Environ Toxicol Chem* 20: 1,252-1,265.

Ireland DS, Burton GA Jr, and Hess GG. 1996. *In situ* toxicity evaluations of turbidity and photoinduction of polycyclic aromatic hydrocarbons. *Environ Toxicol Chem* 15: 574-581.

Jarvinen AW and Ankley GT. 1999. *Linkage of Effects to Tissue Residues: Development of a Comprehensive Database for Aquatic Organisms Exposed to Inorganic and Organic Chemicals*. SETAC Press, Pensacola, Florida, USA.

Johnson DH. 1999. The insignificance of statistical significance testing. *J Wildlife Manag* 63: 763-772.

Jonasson PM. 1972. Ecology and production of the profundal benthos in relation to phytoplankton in Lake Esrom. *Oikos Suppl* 14: 1-48.

Jung RF and Batley GE. 2002. Artifacts in the sampling for selenium analysis sediment pore waters. *Environ Toxicol Chem* (submitted).

Kaag NHBM, Foekema EM, Scholten MCT, et al. 1997. Comparison of contaminant accumulation in three species of marine invertebrates with different feeding habits. *Environ Toxicol Chem* 16: 837-842.

Kadeg RD, Pavlou AS, and Duxbury AS. 1986. *Elaboration of Sediment Normalization Theory for Non-Polar Organic Compounds*. Battelle Pacific Northwest Labs Report No. PB95142899, prepared for US EPA Office of Water Regulation and Standards, Washington, DC, USA.

Kater BJ, Postma JF, Dubbeldam M, et al. 2001. Comparison of laboratory and *in situ* sediment bioassays using *Corophium volutator*. *Environ Toxicol Chem* 20: 1,291-1,295.

Klaine SJ, Cobb GP, Dickerson RL, et al. 1996. An ecological risk assessment for the use of the biocide, dibromonitrilopropionamide (DBNPA) in industrial cooling systems. *Environ Toxicol Chem* 15: 21-30.

Kraaij R, Ciarelli S, Tolls J, et al. 2001. Bioavailability of lab-contaminated and native polycyclic aromatic hydrocarbons to the amphipod *Corophium volutator* relates to chemical desorption. *Environ Toxicol Chem* 20: 1,716-1,724.

Lasier PJ, Winger PV, and Bogenrieder KT. 2000. Toxicity of manganese to *CerioDaphnia dubia* and *Hyalella azteca*. *Arch Environ Contam Toxicol* 38: 298-304.

Lee BG, Griscom SB, Lee JS, et al. 2000. Influences of dietary uptake and reactive sulfides on metal bioavailability from aquatic sediments. *Science* 287: 282-284.

Levinton JS 1995. Bioturbators as ecosystem engineers: control of the sediment fabric, inter-individual interactions, and material fluxes. *In*: Jones CG and Lawton JH (editors), *Linking Species and Ecosystems*, p. 29-36. Chapman and Hall, London, UK.

Linke-Gamenick I, Forbes VE, and Méndez N. 2000a. Effects of chronic fluoranthene exposure on sibling species of *Capitella* with different development modes. *Mar Ecol Prog Ser* 203: 191-203.

Linke-Gamenick I, Vismann B, and Forbes VE. 2000b. Effects of fluoranthene and ambient oxygen levels on survival and metabolism in three sibling species of *Capitella* (Polychaeta). *Mar Ecol Prog Ser* 194: 169-177.

Long ER, MacDonald DD, Smith SL, et al. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manag* 19: 81-97.

Long ER, Hong CB and Severn CG. 2001. Relationships between acute sediment toxicity in laboratory tests and abundance and diversity of benthic infauna in marine sediments: A review. *Environ Toxicol Chem* 20: 46-60.

Lotufo GR. 1998. Lethal and sublethal toxicity of sediment-associated fluoranthene to benthic copepods: application of the critical-body-residue approach. *Aquat Toxicol* 44: 17-30.

MacDonald DD, Carr RS, Calder FD, et al. 1996. Development and application of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5: 253-278.

Maher WA, Norris RH, Curran S, et al. 1992. Zinc in the sediments, water and biota of Lake Burley Griffin, Canberra. *Sci Total Environ* 125: 235-252.

Maltby L, Kedwards TJ, Forbes VE, et al. 2001. Linking individual-level responses and population-level consequences. *In*: Baird DJ and Burton GA Jr (editors), *Ecological Variability: Separating Natural from Anthropogenic Causes of Ecosystem Impairment*, p. 27-82. SETAC Press, Pensacola, Florida, USA.

McPherson CA and Chapman PM. 2000. Copper effects on potential sediment test organisms: The importance of appropriate sensitivity. *Mar Pollut Bull* 40: 656-665.

Mermillod-Blondin F, Gérino M, Degrange V, et al. 2001. Testing the functional redundancy of *Limnodrilus* and *Tubifex* (Oligochaeta, Tubificidae) in hyporheic sediments: An experimental study in microcosms. *Can J Fish Aquat Sci* 58: 1,747-1,759.

Minshall GW. 1988. Stream ecosystem theory: A global perspective. *J N Am Benthol Soc* 7: 263-288.

Morrisey DJ, Underwood AJ, and Howitt L. 1996. Effects of copper on the faunas of marine soft-sediments: An experimental field study. *Mar Biol* 125: 199-213.

Murdoch MH, Chapman PM, Norman DM, et al. 1997. Spiking sediment with organochlorine compounds for toxicity testing. *Environ Toxicol Chem* 16: 1,504-1,509.

Nord MA. 2001. Recommendations for the implementation of a national sediment quality policy in the United States. *Hum Ecol Risk Assess* 7: 641-650.

OECD (Organization for Economic Co-operation and Development). 1998. The OECD Principles of Good Laboratory Practice. *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring*, no. 1. Paris, France.

OECD (Organization for Economic Co-operation and Development). 2001. *Guideline for the Testing of Chemicals: Sediment-Water Chironomid Toxicity Test Using Spiked Sediment*. Draft new guideline 218. Paris, France.

Paine MD. 2002. Statistical significance in environmental effects monitoring (EEM) programs. *SETAC Globe* 3(1): 23-24.

Pearson TH, Josefson AB, and Rosenberg R. 1985. Petersen's benthic stations revisited. I. Is the Kattegatt becoming eutrophic? *J Exp Mar Biol Ecol* 92: 157-206.

Pechenik JA, Gleason T, Daniels D, et al. 2001. Influence of larval exposure to salinity and cadmium stress on juvenile performance of two marine invertebrates (*Capitella* sp. I and *Crepidula fornicata*). *J Exp Mar Biol Ecol* 264: 101-114.

Phipps GL, Mattson VR and Ankley GT. 1995. Relative sensitivity of three freshwater benthic macroinvertebrates to ten contaminants. *Arch Environ Contam Toxicol* 28: 281-286.

Ribeiro R, Kelly LA, Gonclaves F, et al. 1999. New artificial sediment for *Chironomus riparius* toxicity testing. *Bull Environ Contam Toxicol* 63: 691-697.

Roex EWM, Van Gestel CAM, Van Wezel AP, et al. 2000. Ratios between acute aquatic toxicity and effects on population growth rates in relation to toxicant mode of action. *Environ Toxicol Chem* 19: 685-693.

Rosenberg R, Gray JS, Josefson AB, et al. 1987. Petersen's benthic stations revisited. II. Is the Oslofjord and eastern Skagerrak enriched? *J Exp Mar Biol Ecol* 105: 219-251.

Rosenzweig ML. 1995. *Species Diversity in Space and Time*. Cambridge University Press, Cambridge, UK.

Sarda N and Burton GA Jr. 1995. Ammonia variation in sediments: spatial, temporal and method-related effects. *Environ Toxicol Chem* 14: 1,499-1,506.

Sasson-Brickson G and Burton GA Jr. 1991. *In situ* and laboratory toxicity testing with *CerioDaphnia dubia*. *Environ Toxicol Chem* 10: 201-207.

Selck H, Forbes VE and Forbes TL. 1998. Toxicity and toxicokinetics of cadmium in *Capitella* sp. I: Relative importance of water and sediment as routes of cadmium uptake. *Mar Ecol Prog Ser* 164: 167-178.

Shao Q. 2000. Estimation for hazardous concentrations based on NOEC data: An alternative approach. *Envirometrics* 11: 583-595.

Shaw AB. 1969. Adam and Eve, paleontology, and the non-objective arts. *J Paleontology* 43: 1,085-1,098.

Shuttleworth SM and Davison, W. 1999. Two-dimensional and fine structure in the concentrations of iron and manganese in sediment pore-waters. *Environ Sci Technol* 33: 4,169-4,175.

Simpson SL, Apte SC and Batley GE. 1998. Effect of short-term resuspension events on trace metal speciation in polluted anoxic sediments. *Environ Sci Technol* 32: 620-625.

Simpson SL, Apte SC and Batley GE. 2000a. Effect of short-term resuspension events on the oxidation of cadmium, lead and zinc sulfide phases in anoxic sediments. *Environ Sci Technol* 34: 4,533-4,537.

Simpson SL, Rosner J and Ellis J. 2000b. The competitive displacement reactions of Cd, Cu and Zn added to a polluted, sulfidic estuarine sediment. *Environ Toxicol Chem* 19: 1,992-1,999.

Simpson SL and Batley GE 2002. Disturbances to metal partitioning during pore water and whole-sediment toxicity testing of Fe(II)-rich estuarine sediments. *Environ Toxicol Chem* (in press).

Smith SL, MacDonald DD, Keenleyside KA, et al. 1996. A preliminary evaluation of sediment quality assessment values for freshwater ecosystems. *J Great Lakes Res* 22: 624-638.

Stecko JO and Bendell-Young LI. 2000. Uptake of 109Cd from sediments by the bivalves *Macoma balthica* and *Protothaca staminea*. *Aquat Toxicol* 47: 147-159.

Stemmer BL, Burton GA Jr and Leibfritz-Frederick S. 1990a. Effect of sediment test variables on selenium toxicity to *Daphnia magna*. *Environ Toxicol Chem* 9: 381-389.

Stemmer BL, Burton, GA Jr and Sasson-Brickson G. 1990b. Effect of sediment spatial variance and collection method on cladoceran toxicity and indigenous microbial activity determinations. *Environ Toxicol Chem* 9: 1,035-1,044.

Teasdale PR, Batley GE, Apte SC, et al. 1995. Pore water sampling with sediment peepers. *Trends Anal Chem* 14: 250-256.

Teasdale PR, Allen L, Apte SC, et al. 1998. *In situ* collection of diagenetic and induced oxyhydroxide precipitates from riverine and estuarine sediments. *Environ Technol* 19: 1,191-1,201 .

US EPA (U.S. Environmental Protection Agency). 1991. *Methods for Aquatic Toxicity Identification Evaluations – Phase I Toxicity Characterization Procedures*. EPA-600-6-91-003. Office of Water, Washington, DC, USA.

US EPA (U.S. Environmental Protection Agency). 1999. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers; Periphyton, Benthic Macroinvertebrates, and Fish*. EPA 841-B-99-002. Office of Water, Washington, DC, USA.

US EPA (U.S. Environmental Protection Agency). 2002. *Methods for Collection, Storage and Manipulation of Sediments for Chemical and Toxicological Analyses*. EPA-823-B-01-002. Office of Water, Washington, DC, USA.

Van Leeuwen CJ and Hermens JLM. 1995. *Risk Assessment of Chemicals*. Kluwer Academic, Dordrecht, The Netherlands.

Von Stackelberg K and Menzie CA. 2002. A cautionary note on the use of species presence and absence data in deriving sediment criteria. *Environ Toxicol Chem* 21: 466-472.

Wagner C and Løkke H. 1991. Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Res* 25: 1,237-1,242.

Wang WX and Fisher NS. 1999. Assimilation efficiencies of chemical contaminants in aquatic invertebrates: a synthesis. *Environ Toxicol Chem* 18: 2,034-2,045.

Warne M StJ. 2002. Derivation of the ANZECC and ARMCANZ water quality guidelines for toxicants. *Australas J Ecotoxicol* 7: 123-136.

Warren LA, Tessier A and Hare L. 1998. Modelling cadmium accumulation by benthic invertebrates *in situ*: the relative contributions of sediment and overlying water reservoirs to organism cadmium concentrations. *Limnol Oceanogr* 43: 1,442-1,454.

Warwick RM. 1986. A new method for detecting pollution effects on marine macrobenthic communities. *Mar Biol* 92: 557-562.

Washington HG. 1984. Diversity, biotic and similarity indices. A review with special relevance to aquatic ecosystems. *Wat Res* 18: 653-694.

Watzin MC, McIntosh AW, Brown EA, et al. 1997. Assessing sediment quality in heterogeneous environments: a case study of a small urban harbor in Lake Champlain, Vermont, USA. *Env Toxicol Chem* 16: 2,125-2,135.

Ying W, Batley GE and Ahsanullah M. 1992. The ability of sediment extracts to measure the bioavailability of metals to three marine invertebrates. *Sci Total Environ* 125: 67-84.