Journal of Aquatic Ecosystem Health **4**: 183–194, 1995.

Sediment quality assessment: status and outlook

Peter M. Chapman
EVS Environment Consultants, 195 Pemberton Avenue, North Vancouver, B.C., Canada V7P 2R4

Received October 1994; accepted in revised form 28 June 1995

Key words: sediment, aquatic, toxicity, assessment

Abstract

This paper is a synthesis of the 44 presentations at the First International Symposium on Sediment Quality Assessment (Sweden, August 1994). The paper includes Initial Premises, Sediments, Tools (with particular emphasis on bioassays), Strategies and Challenges. Major testable hypotheses are proposed as follows (ranging in complexity, recognizing differences apparent at the Symposium in level of expertise and knowledge): (1) there is no single 'perfect' method of sediment assessment, there are only 'tools in the toolbox'; (2) significant sediment pollution (contamination resulting in adverse biological effects) comes from non-anthropogenic sources; (3) artificial sediments will provide future reference comparisons; (4) knowledge of suspended sediments is required to understand bedded sediments; (5) ammonia and/or hydrogen sulfide cannot explain all sediment toxicity, in particular non-acute responses; (6) subcellular (e.g., genetic) responses are a research tool, not yet appropriate for monitoring or assessment; (7) although the effects of sediment storage cannot be predicted, non-toxic and highly toxic sediments are less affected by prolonged storage than are moderately toxic sediments; (8) sediment ingestion is a more important route of exposure than pore water for some organisms; (9) water column organisms and aqueous exposures should not be used for whole-organism sediment tests; (10) validation of sediment bioassays is not always simple or possible. Two major conclusions are: (i) generalizations are not [yet] possible regarding sediment quality; (ii) correctly assessing sediment quality is primarily a function of the correct reference comparison.

1. Introduction

This paper provides my synthesis of work presented at and discussions held during the First International Symposium on Sediment Quality Assessment: Rationale, Challenges and Strategies (Göteborg, Sweden, August 22–25, 1994). It summarizes 15 poster presentations and 39 platform presentations. As such the paper is not intended to be comprehensive, but rather to be provocative (i.e., to stimulate new thought and hypotheses). The format does not exactly follow that of the Symposium, which was divided into six separate platform sessions (metals, organic chemicals, bioassays, quality assessment, toxicity identification and remedial action, and international cooperation and harmonization), and a poster session. But it does synthesize all information I consider the most important (others may disagree). Detailed information on the presentations referred to herein is provided in the Symposium's Final Program and Abstracts (AEHMS, 1994 – available from the editor of this journal).

This synthesis also comprises my personal assessment of the status and outlook for sediment quality assessment following this Symposium. I do not expect (and would be disappointed and surprised by) total agreement with what I have to say. However, I hope that disagreements will be provided as testable hypotheses, whose proof or disproof will continue to enrich and improve the field of sediment quality assessment. To this end I provide my own hypotheses on major subject areas. Some of these hypotheses will appear simplistic to those with experience and expertise in sediment quality assessment. However, participants at the Symposium encompassed a wide range of knowledge and experience; it is not unreasonable to expect the same from readers. My challenge to readers is to support or refute my hypotheses with good science, and to report the results at the second

International Symposium on this subject, which will be held in late 1996, Pallanza, Italy.

2. Initial premises

The organizers of this Symposium had two major premises: (1) that sediment quality has a major impact on ecosystem health, and (2) that sediments are emerging as one of the most common problems across the world deserving remedial actions. The first premise received general agreement from all participants. However, the second premise was and is arguable on two counts: first, are contaminated sediments more than a 'hot-spot' problem?; second, is the 'no-action' alternative a valid remedial action for sediments? The 'hot-spot' issue arises from the differentiation between contamination (a chemical or other parameter out of place) and pollution (contamination which causes adverse biological effects). The 'no-action' alternative recognizes the reality that one can never return to pristine (or pre-disturbance) conditions after pollution. Both issues can only be resolved by establishment of reasonable determination and, ultimately, some type of prediction of cause-and-effect. Prediction is, of course, preferable to an after-the-fact determination, but such a goal will not easily be attained.

3. Sediments

Sediments comprise a complex matrix, which varies both areally and with depth in the sediments. This matrix comprises the physico-chemical three-dimensional context within which organisms live (or not) and within which contaminants may be bound and thus not bioavailable, or the reverse (which can change due to a variety of factors, on both a micro- and a macro-scale). It is rare to find only one contaminant in sediments, which generally comprise complex mixtures. The biotic component of sediments comprises complex organism interactions, including competition and predation, habitat effects (e.g., grain-size, depth, non-anthropogenic water quality parameters such as salinity), and contaminant effects. The latter include bioaccumulation and toxicity.

Physical factors affecting sediments include grain size, porosity, and permeability (the ability of water to move through the sediments; in sediments with low permeability, interstitial water will have a very different composition from the overlying water). Sediments are a dynamically evolving medium comprising solid matter and water.

Contaminants affecting sediments can be divided into two general categories, which was done with separate platform presentations on the first day of the Symposium: metals (including other inorganic contaminants), and organic contaminants. Other major categories are: large-scale and biological factors.

It was apparent during the Symposium (and is hopefully reflected in this synthesis), that the appropriate use of different tools provides the most useful information. The tools will improve, but meanwhile we must use what we have intelligently, being wary of generalizations and accepting the fact that a complex environment merits non-simplistic approaches.

Given the complexity of sediment-contaminant interactions, the following hypothesis is proposed:

> Hypothesis 1: There is no single 'perfect' method for characterizing sediment pollution status; there are only 'tools in the toolbox'.

3.1. *Metals*

Calmano (AEHMS, 1994) provided an overview of this subject from a chemist's perspective, which was supplemented by five following platform presentations. Metals in sediments are affected by various factors including redox (e.g., Eh, which generally differentiates oxic from anoxic sediments), pH, sulfides (including acid volatile sulfides [AVS]; an important point made during the Symposium was that all sulfides are not the same), carbonates, total organic carbon (TOC), speciation, iron and manganese oxides. The importance of the latter was emphasized for at least freshwater sediments by several speakers and has been noted elsewhere (e.g., Allen, 1993). Calmano (AEHMS, 1994) also referred to acid producing capacity (APC) and acid neutralizing capacity (ANC) as possible important (but not well studied) factors determining metals bioavailability.

Other factors include the importance of microbial processes on mercury in sediments (e.g., methyl mercury can be biomagnified up the food chain to human beings). It was also noted that metals other than mercury show much greater bioavailability in fresh as opposed to salt waters.

Sequential extraction was noted by several speakers as a surrogate for direct measures of bioavailability. However, it was also noted that this is not an exact predictive tool. Neither, it appears, is AVS for other than specific circumstances involving particular metals (SAB, 1995).

3.2. *Organics*

The session on organic contaminants in sediments suffered from the sudden withdrawal of the keynote speaker. Thus there was no broad overview of the subject. However, three platform presentations and some information from the Poster session provided insights.

The type and concentration of TOC has a major role in the partitioning of organic contaminants (particularly non-polar compounds) to and within sediments. Fractionation was noted as a tool whose parameters need to be better known. For example, differences result from using lake as opposed to distilled water for fractionation (Kukkonen – AEHMS, 1994).

3.3. *Large-scale factors*

Large-scale factors affecting sediments and discussed at the Symposium were lake depth and residence time affecting resuspension, remobilization, burial (including reburial) (Broberg, AEHMS, 1994). In shallow lakes morphometry and wind can have a major influence. Other factors also of importance in some areas and locations include: ice scour, the salt wedge in such estuaries, waves, and river freshet.

3.4. *Biological factors*

Feeding (selective and non-selective) has major effects on bioavailability and hence, toxicity. Sediments consist of micro-environments, within which life-strategies are extremely important. For instance, tube builders, sediment feeders and organisms which move through and feed within the sediments will not be exposed to the same contaminant concentrations. This is a major consideration not only for sediment assessments involving community structure, btu also for tests designed to assess bioavailability.

Organisms have a major effect on contaminant bioavailability through their actions. In the case of animals, bioturbation can affect persistence and availability either by increasing or decreasing each. For instance, lugworms can dilute effects by mixing 'clean' with contaminated sediment (Kure – AEHMS, 1994). So far the direction of such affect(s) is not predictable. Plants in, for example, marshes, have affect(s) through root activity. Microbial activity also has an affect (e.g., degradation). There was a suggestion that seasonal changes in bioavailability could be due to non-anthropogenic contamination and pollution (Borchert *et al.* – AEHMS, 1994), but this remains a hypothesis to be tested.

> Hypothesis 2: Non-anthropogenic contamination and pollution may be important in certain areas and situations, and need to be differentiated from anthropogenic sources and influences.

4. Tools

A wide variety of 'tools in the tool box' were mentioned and discussed. These can be broadly divided into two categories: laboratory and field. Such tools include bioassays and benthic community structure, which are discussed separately.

4.1. *Laboratory*

Several papers dealt with the issue of spiked sediments, in particular metals uptake by fish and fluoranthene toxicity to *Chironomus riparius*. The former represents research into single contaminant cause-and-effect questions which are valuable, but whose results must be interpreted in the context that sediments consist of mixtures, not individual contaminants. The latter involved more than one study and presentation, including both radio-labelled and conventional fluoranthene additions (which each have advantages and disadvantages). Spiking sediments remains more art than established technique; in this context, Stewart & Thompson (AEHMS, 1994) found similarities between three spiking techniques (standard, coat, and dry), but not a fourth (wet spiking).

The use of artificial sediments was detailed in various papers throughout the Symposium. The point of such sediments is to provide standardized, comparable data (e.g., for toxicity), and in particular to facilitate prediction. Improvements have been made in using artificial sediments, in particular in spiked sediment toxicity tests (Stewart & Thompson – AEHMS, 1994).

> Hypothesis 3: At some point in the future, artificial rather than 'natural' sediments will form the basis for comparisons to determine sediment quality.

Laboratory tests can involve single species or multiple species (e.g., microcosms). They can also involve cause-and-effect investigations triggered by the finding of an effect either in the laboratory (e.g., sediment toxicity) or in the field (e.g., impaired benthic community structure). Toxicity identification evaluation (TIE), reviewed by Ankley & Schubauer-Berigan (AEHMS, 1994) appears to be a most promising tool for sediments. For instance, sediment TIE have been successfully used to determine whether (or not) ammonia is causing observed toxicity. Methodology development is proceeding rapidly and TIE is expected to be a major tool in future for determining cause-and-effect relationships for all contaminants. A major research needs is a standard 'best' method of obtaining pore water.

4.2. *Field*

Suspended sediments are useful for monitoring movements and inputs of contaminants. Two different techniques were described: *in situ* centrifugation, in this case for organic contaminants in river water (Sekela *et al.* – AEHMS, 1994) and sediment traps (three different presentations: metals in the Baltic, organic contaminants in river water, and Cesium-137 in shallow lakes).

> Hypothesis 4: Knowledge of and understanding of suspended sediments if essential for assessing and predicting sediment pollution.

There was substantial discussion regarding the usefulness of complex bioassay systems such as mesocosms (e.g., from whole lakes to field environmental enclosures) for prediction and assessment. No conclusion was reached.

5. Bioassays

5.1. *General issues*

Bioassays (exposure of an organism or biotic system to a stimulus such as a contaminant) include both bioaccumulation tests (which measure a phenomenon) and toxicity tests (which measure an effect). A key issue in sediment toxicity tests is differentiating internal dose and external concentration. External concentration represents what is measured in, for instance, the external medium (e.g., pore water and/or whole sediment) during a test. This is what the organism (or biotic system) is exposed to. Internal dose represents what actually enters through biological membranes. In the case of some contaminants such as the PAH, which are metabolized, measurements of PAH in tissues provides no useful information on toxicity to the test organism. This is emphasized by the fact that bioaccumulation tests for PAH use organisms which do not metabolize this group of organic contaminants and which are thus not affected by them (e.g., bivalves). However, the consumer(s) of such organisms may be affected. In the case of other contaminants, synoptic measurements of dose and concentrations can provide extremely useful information on cause-and-effect. This was exampled by Besser *et al.* (AEHMS, 1994) who showed uptake of copper by *C. tentans* prior to the onset of observable effects.

A major research area involves differentiating whether sediment toxicity is due in any part to either or both of hydrogen sulfide and ammonia in sediments, as opposed to metals and organic contaminants. It is noteworthy that both of these 'conventional' contaminants primarily have acute effects (i.e., affect survival).

> Hypothesis 5: 'Conventional' contaminants such as hydrogen sulfide and/or ammonia may be responsible for some sediment toxicity, but cannot explain all sediment toxicity, particularly where observed effects are not acute.

5.2. *End-points*

Major end-points in sediment toxicity tests are: death, growth, fecundity (i.e., reproductive impairment). Other end-points whose significance is less certain include: genotoxicity, cytotoxicity, tetatogenicity and mutagenicity.

> Hypothesis 6: Measures of genetic, cytotoxic, tetatogenic, and mutagenic effects of sediments are subject to non-anthropogenic influences (c.f. Hypothesis 2) and are, as yet, too poorly understood to be used for assessment or remediation purposes.

5.3. *Sampling and subsequent sediment treatment*

Sediment sampling is an area of uncertainty, yet key to the assessment process. Depending on the questions being asked, data quality is entirely dependent upon the effectiveness, representativeness and quality of sampling. Methods for sampling are not a research issue but rather an issue for consensus on practice among workers. Storage time issues are discussed below. In general, sediment treatment during testing also needs consensus on practice among workers.

5.4. *Storage time(s)*

The effect of sediment storage is an important variable affecting bioavailability which must be evaluated. It was noted that different effects will occur on different organisms over different times. The effects of storage time on contaminant bioavailability appear to be most pronounced for moderately contaminated and toxic sediments.

> Hypothesis 7: The effects of sediment storage time cannot be predicted and should be standardized, including sampling.
>
> Subhypothesis 7a: Reference (i.e., non-toxic, non-contaminated) sediments can be stored for relatively long periods of time, whose exact extent remains to be determined.
>
> Subhypothesis 7b: Highly toxic, highly contaminated sediments can be stored for relatively long periods of time, whose exact extent remains to be determined.
>
> Subhypothesis 7c: Moderately toxic, moderately contaminated sediments (i.e., >70% of sediments) cannot be stored for any significant period of time, and predictions cannot be made, hence for these sediments two 'rules' apply: test as soon as possible, at least within a preset time period.

I noted that the oft-quoted 2-week maximum holding time was a 'guesstimate' I made during the early days of sediment toxicity testing, and never intended as a final measure (Chapman, 1995a). I expressed my concern that this measure has now been 'cast in stone', and raised concern that other measures we now use and will use in future be more rigorously assessed before being 'generalized'.

5.5. *Exposure routes*

There are four general, major exposure routes used more or less routinely for sediment bioassays: whole sediment, elutriate, interstitial (= pore) water, and chemical extracts. An issue which received some discussion was the use of settled as opposed to resuspended sediments in tests depending on the final use of the information. For instance, testing with resuspended sediments may be more useful for assessing dredged disposal operations than settled sediment tests.

A relatively new route of exposure discussed by Andreasson & Dave (AEHMS, 1994) was the use of higher organism fluids following their exposure to contaminated sediments. These researchers used *Daphnia* to test the toxicity of bile and blood plasma from trout exposed to metals in sediments. However, this technique is limited to contaminants which bioconcentrate (i.e., Cu, Pb and Hg in their study; PAH and other organics also bioconcentrate and are found in bile). It is anticipated that studies with this, still experimental, route of exposure, will increase (e.g., Harris *et al.*, 1994).

Although theories such as EqP assume that pore water is the major route of uptake for sediment contaminants, sediment ingestion is an additional route of uptake for some organisms (cf. Chandler *et al.*, 1994).

> Hypothesis 8: Because feeding is a major route of uptake for some organisms, pore water analyses cannot explain all uptake and effects related to contaminated sediments.

Burton (AEHMS, 1994) pointed out the importance of phototoxicity in assessing possible effects of PAH in sediments. This cannot be done by standard laboratory toxicity tests and is a relatively new research area for sediments related to possible increased light penetration of waters with thinning of the ozone layer.

5.6. *Types of tests*

Tests can be generally differentiated into three generations of whole organism tests (after Chapman *et al.*, 1992a), 'kit' tests, and biomarkers. First generation tests are those in which water column tests are adapted for sediments. Second generation tests are those which are specifically designed for sediments and which measure primarily survival. Third generation tests are second generation tests which also measure chronic responses, in particular fecundity and/or growth. Very few of the tests described during the Symposium fit in this final category; clearly, there is a need for such tests, in particular those which more closely approximate the full life-cycle of the test organism.

The suitability of these different generations of tests for assessing sediment toxicity was not discussed, however, it was generally agreed that, in terms of toxicity: pore water > whole sediment > elutriates. Harkey *et al.* (1994) have suggested that the aqueous exposure route is not appropriate for either benthic organisms or water-column species. If these authors are correct, then first generation sediment toxicity tests are not appropriate for sediment assessment or monitoring.

> Hypothesis 9: Whole organisms sediment bioassay tests give the most useful information when conducted with whole sediments using benthic organisms (i.e., second and third generations tests as defined by Chapman *et al.*, 1992a).

Bioavailability was also noted to change depending on whether testing was conducted under static, water renewal, or flow-through conditions (Burton – AEHMS, 1994). However, static testing remains the simplest, least costly, and most generally used technique.

'Kit' tests are defined here as those which are delivered with pre-packaged organisms and techniques. Examples are the Microtox test (which comes in varieties such as the elutriate and whole sediment test), the Comet test, Algaltoxkit, Rotoxkit, and Thamnotoxkit. Where such 'kits' involve whole organisms, these are a form of microbiotest, defined as whole organism tests which are rapid, sensitive, use small volumes, and are less expensive to conduct than larger volume tests.

Biomarkers are defined here as tests which involve biotic levels of organization below the whole organism (e.g., biochemical changes). Munkittrick *et al.* (AEHMS, 1994) presented preliminary but promising results relating fish liver detoxification enzyme changes to reproductive abnormalities in the same fish and to whole sediment acute toxicity tests with *Daphnia* and *Hyalella*. Hansen's (AEHMS, 1994) presentation illustrated the complexity embodied in biomarkers, where different mechanism of action provide different responses but whose utility we do not yet understand.

5.7. *Test organisms*

Burton (AEHMS, 1994) provided detailed comments regarding selection of test organisms. Three major discussion points merit repetition here: sensitivity *versus* discrimination – some tests are too sensitive, and both inter- and intra-laboratory variability must be established; elimination of redundancy – in terms of similar information given by different tests in a test battery; surrogates *versus* native species – the latter give the most 'realistic' information.

Specific recommendations for organisms to use in a test battery were provided by den Besten *et al.* and Thain *et al.* (AEHMS, 1994). Both recommend a crustacean; in addition, the former recommend using a fish and considering an alga, in contrast, the latter recommend a polychaete as a second test species.

Assessing moderately polluted dredged material in Holland was addressed by both Lourens *et al.* and Stonkhurst & van den Hurk (AEHMS, 1994). Their toxicity testing attempts to cover similar taxonomic groups as done in North America, accepting the U.S. EPA precedent. Tests used in both Holland and North America are the oyster (*Crassostrea gigas*) fertilized egg to prodissoconch I larva test, the Microtox test, and tests with Sheepshead minnows (*Cyprinodon variegatus*). The former two tests, together with the amphipod *Bathyporeia sarsi* provided the highest level of discriminatory power in these difficult to characterize sediments.

5.8. *Test standardization*

Several papers discussed standardization. For instance, an international ring test using *C. riparius* exposed to pesticide spiked into the overlying water found the test method to be robust (Heimbach & Hammer – AEHMS, 1994). Flemming (AEHMS, 1994) described a European collaboration testing several species against poorly soluble substances (e.g., Lindane, PCBs), which found *C. riparius* (freshwater) and *Corophium volutator* (salt water) 10-d survival and growth to be robust tests.

6. Benthos

Benthic communities were a component of many of the papers, but were only the focus of one: Diaz and Rosenberg (AEHMS, 1994). Various points were made during this presentation and subsequent discussion which bear repeating.

Benthos is usually collected by grab samples which do not differentiate layers in sediment. This can lead to apparent problems in interpretation when, for instance, the benthos is living on a thin, non-toxic sediment layer overlying toxic sediments. If benthic community structure analyses is done without knowing this, it would indicate less impairment than toxicity tests or chemical analyses, which could result in an incorrect assessment of sediment quality.

Sediment contamination can result in major energetic changes affecting the food chain, in particular the large bioturbators and predators, resulting in a shortened food chain consisting of smaller, faster growing species. The effects of benthos on contaminated sediments relate to biogenic alteration of sediment and pore water. These effects are reduced when contamination results in larger animals being replaced by smaller, simply because larger animals move more sediment than smaller animals.

Sediment contamination affects both the structure and function of benthic communities. However, so too do conventional loadings (e.g., resulting in increased ammonia and sulfides, low dissolved oxygen concentrations), physical disturbance (e.g., habitat, water quality parameters such as salinity), and biotic interactions (e.g., competition, predation). Using the benthos alone (e.g., without considering, for instance, chemistry and toxicity), it is almost impossible to distinguish and determine the status of moderately contaminated and toxic sediments, which comprise the majority (>70%) of all sediments.

7. Strategies

This section summarizes attempts to use two major strategies: holistic assessment to determine site- and situation-specific conditions; and, generic methods in the form of sediment quality criteria for achieving the same purpose.

7.1. *Holistic assessment*

Holistic assessment is defined here as approaches which use more than one 'tool in the toolbox' in an integrated approach appropriate for a specific site or situation. The 'typical' generic approach, in contrast, involves determining numerical criteria (or guidelines, standards or similar term) for pass/fail determinations.

Barret (AEHMS, 1994) described a study which progressed from laboratory experiments (fate and effects = testable predictions) to an outdoor microcosm (test predictions in a more complex system = new predictions) to field experiments (test new predictions, make final determinations). This progression, a tiered approach, allowed for elaboration of more physical and chemical features of the contaminant in question, Prochloraz, than is normally possible. For instance, phototoxicity was found to be an issue (also an issue for PAH as noted above). The final field experiments allowed for assessing both planned and unplanned (i.e., more realistic) events.

A commendable attempt to classify estuarine sediments (arguably the most difficult sediments to work on, as noted below) by Flemming *et al.* (AEHMS, 1994) involves parallel but independent assessment by bulk chemical analyses and toxicity testing. The chemistry component, which follows the approach of Long *et al.* (1995) provides cost-effective *screening* (not definitive) information to assess and prioritize (i.e., for further study or possibly action). Toxicity data will serve to clarify the chemical data. Ultimately, an integrative approach will be done involving the Triad of chemistry, toxicology and benthic community structure, at a limited subset of sites prioritized by the initial chemistry and toxicity data.

The above estuarine classification scheme is inferential, as are most sediment classification schemes. In other words, direct measurements of ecosystem health are not attempted but, rather, inferences are drawn based on indirect information (i.e., chemical contamination). A direct assessment classification scheme for freshwater sediments was described by Goyvaerts for Dutch waters and by Reynoldson & Zarull (AEHMS, 1994) for the Great Lakes. In Holland 'acceptable' benthic communities are determined and normalized to habitat (in this case, the focus is on chironomids and oligochaetes). These are then compared to other areas to determine and prioritize polluted sites (as defined by the biota, not by the chemistry). In the Great Lakes direct measures of benthic community structure are supported (but not driven) by chemical analyses and toxicity testing.

The above two direct assessment schemes follow work and techniques pioneered in U.K. rivers (Wright *et al.*, 1984), and show great promise in my opinion. However, they have not yet been applied outside freshwater, require expert knowledge of the benthos, and have intensive up-front costs. The latter consideration has deterred some groups from attempting this in their own jurisdictions, but ignores the cost implicit in making a wrong decision based on simplistic (but initially relatively inexpensive) information (see Hypothesis 1 and Final Comments).

Aanes (AEHMS, 1994) described a lake assessment which began using 'standard' first generation laboratory test organisms (*Daphnia pulex, Selenastrum capricornutum*), then progressed to include the amphipod *Gammarus lacustris.* This latter species used to live in the lake but no longer does due to sediment contamination. This approach is laudatory because it involves ultimately testing what you want to protect, and doing so by 'benchmarking' (EPA/ACOE, 1995) the non-standard test organism against the two 'standard' tests. Standard or benchmark tests may not be ecologically relevant to the ecoregion or location being assessed, which was why Aanes (AEHMS, 1994) conducted additional testing with *G. lacustris*.

7.2. *Generic sediment quality criteria (SQC)*

There were relatively few papers on this subject, specifically on using chemical numbers preferentially over toxicity and/or community structure studies. However, Canadian sediment quality guidelines were described (Smith *et al.* – AEHMS, 1994). In addition, there was discussion (as previously noted) regarding the failure of AVS to provide a basis for metals SQC, and regarding the limitation of EqP theory in providing a basis for non-polar organic contaminant SQC. For organisms for which sediment ingestion is an important or the major route of exposure to sediment contamination (i.e., the trophic route), EqP theory does not appear to hold (cf. Chandler *et al.*, 1994; Odin *et al.*, 1994). Also, as noted by van Beelen & van Vlaardingen (1994), SQC derived solely from tests with aerobic organisms do not protect important microbial processes in anaerobic sediments. Finally, the compatibility (or lack thereof) of sediment and soil criteria was discussed but without resolution.

It is important to note that most SQC provide only one value, which can lead the unwary (or the uncaring) to assume a high level of certainty when, in fact, there is only a high level of uncertainty. It would be useful if SQC or other screening values include both a 'no known effect' level and a 'known effect by any measure' level.

8. Challenges

Sediments are a complex matrix which, in too many cases, we have attempted to define simplistically. Rather, sediments require integrated environmental management (as defined by Cairns *et al.*, 1994). It appears clear, and was generally endorsed by all participants, that simplistic approaches and solutions are not likely to work. For example, attempts to develop generic environmental quality criteria using broad-scale normalizing factors such as AVS and EqP have failed as an overall answer, but do provide tools for screening purposes where they are applicable (which is not in all situations or sediments). In this section, various challenges (i.e., future research areas) are outlined. In general, it was felt by Symposium participants that more *in situ* work would be useful, where feasible and appropriate.

8.1. *Metals bioavailability*

With the realization that AVS is not the single major factor controlling metals availability in sediments, has come a need to better understand other factors which may play a major role under certain conditions. In this regard, iron and manganese oxides may be of great importance. However, the entire subject of metals bioavailability is a major area for research.

8.2. *TOC*

TOC is recognized as a major factor controlling contaminant (both metals and organics) bioavailability in sediments. Two major questions arose from this Symposium and require answering: (1) is all TOC the same?; (2) are we measuring it correctly? (cf. Malley *et al.* – AEHMS, 1994).

8.3. *Estuarine sediments*

Estuarine sediments (i.e., where there is tidal influence and/or salinity is intermediate between fresh (0 ppt) and marine ($\geq$ 25 ppt)) are a major site for anthropogenic inputs since many industries and cities are located where rivers discharge into the sea. However, there is a dearth of sediment toxicity tests for this region, which is the most complex of all sedimentary environments. For instance, changes in salinity greatly influence bioavailability of heavy metals except mercury, and that of various organic contaminants. There is a critical and pressing need for second and third generation estuarine sediment toxicity tests.

8.4. *Generic assessment*

It is questionable whether predicting biological effects will ever be possible, but attempts to do so continue. Attempts such as sediment quality criteria (SQC) are to date only useful, as previously discussed, for screening purposes and/or as one of a set of 'tools in the tool box' whose utility is determined on a site- and situation-specific basis. If accurate, independent prediction is ever possible, it is likely that it will not be generally useful, and that clear differentiation must be made of when (and when not) direct testing is necessary. It is, however, clear that accurate prediction will not be possible without resolving the relative importance of sediment ingestion as opposed to interstitial water as routes of uptake of contaminants in sediments.

8.5. *Toxicity tests*

A battery of tests is preferred since a single test cannot provide all necessary information on the differential responses of the varied and complex biotic environment. However, for cost reasons one or two tests are often used instead. Given this reality, numerous questions (of which these are only a few) remain to be answered: (1) Can a single toxicity test provide an effective screen? (2) Which test, if any, is best for this in what situation(s)? (3) Potential effects on microbial communities are not well characterized (the Microtox test is of arguable use in this regard), should they be and, if so, how? (4) Potential effects on plant communities are not well characterized (the freshwater *Selenastrum* test is of arguable use in this regard), should they be and, if so, how? (5) What is the role of biomarkers? [I believe their present role is for screening or for supplemental investigations, not for definitive assessment since we are not yet sure what such measures are telling us, although this will hopefully change in future.] (6) What is the role of subcellular tests (e.g., genotoxicity)? [I believe there is even more uncertainty regarding what these tests are telling us that is the case of biomarkers and that, accordingly, these tests are presently only appropriate for research, see Hypothesis 6.] (7) Given that TOC affects bioavailability, should we feed organisms in bioassay tests, thus reducing bioavailability? (8) How do we ensure that future test development balances ecological relevance with our ability to conduct the tests (the most sensitive species probably cannot be tested in the laboratory), without undue reliance on test developers' particular preferences and biases? (Note that we can, and sometimes do, increase the sensitivity of those species we can use in the laboratory by conducting tests under environmental conditions near their tolerance, which pre-stress the organisms, increasing their sensitivity to any additional stresses.) (9) How do we balance sensitivity and discriminatory power such that tests are neither too sensitive (false positives) nor the reverse (false negatives)? (10) When and how should we standardize sediment: water exposure ratios? (Whole sediment tests described in this Symposium ranged from 1:1 and 1:4 water to sediment, to a 2 cm sediment layer in 1 L of water.)

8.6. *Standard tests versus elegant experiments*

Both standard tests and elegant experiments were discussed, but rarely in the same study. Chapman *et al.* (1994) have suggested that we are less capable of conducting elegant experiments than in the past, because of the regulatory focus on standardized tests which can be widely used. Whether this is either true or of any significance, remains to be determined.

8.7. *Data interpretation*

This issue poses some of the greatest challenges, in particular: How to distinguish and anthropogenic toxics effect from (1) 'noise' due to, for instance, natural variability in the ecosystem; (2) other confounding factors (e.g. physical and chemical sediment characteristics and non-anthropogenic toxics)? And, if one can actually determine such an effect, how is its ecological significance to be evaluated?

8.8. *Validation and extrapolation*

'Validation' is often recommended for sediment bioassays (e.g., Crane *et al.* – AEHMS, 1994). This is generally mentioned in the context of ensuring that benthic responses are reflected in laboratory responses. Although validation is appropriate in theory, it is not appropriate in practice. The reason for integrated assessment using a variety of different tools is that no single tool provides 'the answer' with regard tos sediment pollution (e.g., see Chapman *et al.*, 1995). Similarly, and as detailed by Chapman (1995b), no single tool serves to validate any other. However, extrapolation from the laboratory to the field is clearly easiest, and often most successful, if one is testing what one wants to protect.

> Hypothesis 10: Sediment bioassays cannot be validated solely by comparison with the benthos; 'validation' is not always possible, and when possible in some cases will result from relatively simple comparisons, in other cases will only result from complex integrative assessments.

8.9. *Regulation*

Two regulatory problems were discussed not in presentations but during discussion. These remain problems: (1) how to reconcile the fact that regulations are based on single number exceedances with the reality that toxicity tests do not give a single number but rather a range of values (as demonstrated when repeating the same test or conducting a ring test)? (2) how to convince regulators not to use estimates such as NOECs which are concentration-dependent, when point measures such as EC50s are more appropriate (cf. Chapman *et al.*, 1996).

8.10. *Remediation*

There were not a lot of papers on remediation, though this clearly is an important topic. Liming of a lake polluted by metals was found to be effective in reducing bioavailability and restoring a viable aquatic ecosystem (Baudo *et al.* – AEHMS, 1994). In this example the metals remained; is this an acceptable long-term solution, simply changing conditions (another method could be burial of contaminated sediments with clean material – either by natural means or by directed capping), so that the sediments are not polluted although they remain contaminated?

This was not judged acceptable in the case of PCBs in a lake where it was estimated that if there were no action, they would adversely affect the environment for some 100 years (Gullbring and Hammar – AEHMS, 1994). The lake was dredged.

Remediation of contaminated sediments is still in its infancy. Hopefully major progress will be reported at the next Symposium in 1996.

8.11. *Cooperation and harmonization*

A whole session of the Symposium was dedicated to 'International Cooperation and Harmonization', which are clearly required since sediments (like most environmental problems in the world today) do not respect geopolitical boundaries. Interestingly enough, unofficial contacts between individuals seem to result in much more progress than official contacts. However, the present focus of both official and unofficial efforts appears to be reductionist (e.g., standardization of methods). The future focus of such efforts must be broader, for example, joint assessment and remediation of international problem sediments (e.g., as is presently occurring in the Great Lakes: Munawar *et al.* – AEHMS, 1994) and of sediments contaminated by airborne chemicals from non-adjoining countries. Similar comments apply to different political regions (e.g., states, provinces) within countries as illustrated by van de Guchte (AEHMS, 1994) for the Netherlands. Data bases clearly need to be both compatible and amenable to exchange, but standardization of such basic issues as sediment collection depths for different purposes needs to occur before this is useful. For instance, in Scandinavia Dave (AEHMS, 1994) described the situation as paralysis due to lack of consensus on such basic assessment elements as how to sample, store and test sediments. Standardization takes the form of both ring tests and routine use, and needs to focus on balancing four factors: reliability, reproducibility, sensitivity and discriminatory capacity. However, standardization cannot exclude flexibility necessary for dealing with complex and special situations.

9. Final comments

One of the two initial premises of this Symposium, and the one that was arguable, was that sediments are one of the most common world-wide problems for remediation. To determine whether remediation is necessary ultimately requires a risk assessment. One of the basic conclusions arising from this Symposium is that sediments are a special case for risk assessment and that we need to develop a framework specific to sediments. Work in this regard is presently underway, through a Workshop conducted in April 1995 (Ansilomar, California; sponsored by the Society of Environmental Toxicology and Chemistry [SETAC]). Of particular importance in a sediment risk assessment is the time-frame. Since sediments comprise both a sink and a source for contaminants, any risk assessment must not only consider the immediate but also consider unlikely events which could occur over long time periods and which might render today's relatively innocuous sediments a 'time bomb'. Moreover, such a risk assessment must incorporate and make provision for communicating uncertainty inherent in the process, particularly for the majority of sediments which are neither 'clean' nor highly contaminated and toxic.

For the present, I propose two major conclusion arising from the Symposium:

Conclusion 1: We cannot [yet] generalize regarding sediment quality assessments.

Conclusion 2: The significance of sediment contamination and bioavailability depends on the reference comparison.

The former conclusion is self-explanatory. The latter has both scientific and non-scientific implications. Specifically, we can now test and assess sediments chemically and biologically using a wide variety of methods (and in future will have many more methods

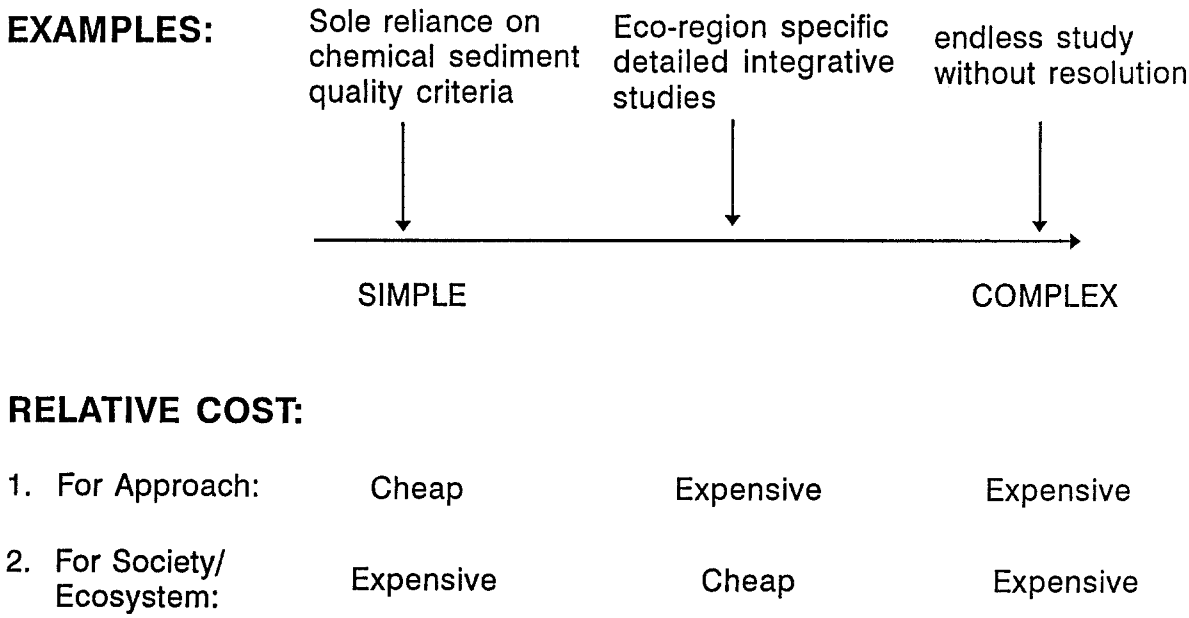

Fig. 1. Relative cost of different approaches to sediment quality assessment.

available). However, comparisons vary by investigator, jurisdiction, and geographic area from statistical analyses *versus* controls (most rigorous and of questionable environmental significance), to burden of evidence comparisons using best professional judgement (not amenable to quantification and not favoured by rigorous statisticians or regulators). The reference comparison depends on determining what is: normal, acceptable, and desirable (and, in future, may involve artificial sediments, see Hypothesis 3).

In this regard there are three questions (aside from the obvious, overriding 'So what?' question which must be borne in mind by scientists, managers, regulators and other stakeholders related to sediment quality assessment: (1) Why are we doing this? (2) What are we trying to protect? (3) Does it matter? The answer to the second question involves protection of human health and [or?] the ecosystem; the answer to the third question involves both ethics and social values. The first question has (K. Barret, pers. comm.) three answers: to assess cleanliness, to monitor remediation, and for prediction/risk assessment. These three answers must provide the basis for any framework developed for sediment quality assessment, from which the various 'tools in the toolbox' can be deployed. These tools, it was agreed by all participants, fall into the following categories (after Chapman *et al.*, 1992b): chemistry, toxicity, benthos, bioaccumulation, others (e.g., bottomfish histopathology). No framework was developed during discussions at this First Symposium. Hopefully the Second Symposium (1996) will include various proposals for such a framework and some mechanism for arriving at a consensus position among participants.

In deriving such a framework it is important to differentiate between the tools available for sediment quality assessment and the monies available for such work. Tiered testing is one solution to effective use of available funds (i.e., tiers increase in cost and complexity commensurate with the degree of complexity of the problem and the level of effort required to reduce uncertainty). However, it is also important to remember that relatively inexpensive methods may not be the least expensive ultimately for either society or the environment (cf. Fig. 1).

Acknowledgements

I thank the organizers of the Symposium, in particular Goran Dave, for inviting me and asking me to provide a daily synthesis of the presentations, and for funding this. Without such encouragement and assistance, this paper would not have been written. I also thank Diane Malley for providing her notes on the final discussions at the Symposium.

References

AEHMS, 1994. Final Program and Abstract Book, the Aquatic Ecosystem Health and Management Society, First International Symposium on Sediment Quality Assessment: Rationale, Challenges and Strategies. August 22–25, 1994. Goteborg, Sweden, 75 pp.

Allen, H. E., 1993. The significance of trace metal speciation for water, sediment and soil quality criteria and standards. Sci Tot Environ Suppl 1993: 23–45.

Cairns, J., Jr., T. V. Crawford & H. Salwasser (eds), 1994. *Implementing Integrated Environmental Management.* Virginia Tech, Blacksburg, Virginia. 137 pp.

Chandler, G. T., B. C. Coull & J. C. Davis, 1994. Sediment- and aqueous-phase fenvalerate effects on meiobenthos: implications for sediment quality criteria development. Mar Environ Res 37: 313–327.

Chapman, P. M., 1995a. How should numerical criteria be used? Human Ecol Risk Assess 1: 1–4.

Chapman, P. M., 1995b. Do sediment toxicity tests require field validation? Environ Toxicol Chem. 14 (in press).

Chapman, P. M., R. C. Swartz, B. Roddie, H. Phelps, P. van den Hurk & R. Butler, 1992a. An international comparison of sediment toxicity tests in the North Sea. Mar Ecol Prog Ser 91: 253–264.

Chapman, P. M., E. A. Power & G. A. Burton, Jr., 1992b. Integrative assessments in aquatic ecosystems. In: G. A. Burton, Jr. (ed.), *Contaminated Sediment Toxicity Assessment.* pp. 313–340. Lewis Publishers, Chelsea, Michigan.

Chapman, P. M., M. D. Paine, T. Moran & T. Kierstead, 1994. Refinery water (intake and effluent) quality: update of 1970s with 1990s toxicity testing. Enviro Toxicol Chem 13: 897–909.

Chapman, P. M., M. D. Paine, A. D. Arthur & L. A. Taylor, 1995. A triad study of sediment quality associated with a major, relatively untreated marine sewage discharge. Mar Pollut Bull. (in press).

Chapman, P. M., R. Caldwell & P. Chapman, 1996. A warning: NOECs are inappropriate for regulatory use. Environ Toxicol Chem. (in press).

EPA/ACOE, 1995. *Evaluation of Dredged Material Proposed for Discharge in Waters of the U.S. – Testing Manual.* U.S. Environmental Protection Agency EPA – 823-B-94-002. Washington, D.C.

Harkey, G. A., P. F. Landrum & S. J. Klaine, 1994. Comparison of whole-sediment, elutriate and pore water exposures for use in assessing sediment-associated organic contaminants in bioassays. Environ Toxicol Chem 13: 1315–1329.

Harris, G. E., T. L. Metcalfe, C. D. Metcalfe & S. Y. Huestis, 1994. Embryotoxicity of extracts from Lake Ontario rainbow trout (*Oncorhynchus mykiss*) to Japanese medake (*Oryzias latipes*). Environ Toxicol Chem 13: 1393–1403.

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

Odin, M., A. Feurtet-Mazel, F. Ribeyre & A. Boudou, 1994. Actions and interactions of temperature, pH and photoperiod on mercury bioaccumulation by nymphs of the burrowing mayfly *Hexagenia rigida*, from the sediment contamination source. Environ Toxicol Chem 13: 1291–1302.

SAB, 1995. An SAB Report: Review of the Agency's approach for developing sediment criteria for five metals. U.S. Environmental Protection Agency, Science Advisory Board, Washington, D.C. EPA-SAB-EPEC-95-0XX.

van Beelen, P. & P. van Vlaardingen, 1994. Toxic effects of pollutants on the mineralization of 4-chlorophenol and benzoate in methanogenic river sediment. Environ Toxicol Chem 13: 1051–1060.

Wright, J. F., D. Moss, P. D. Armitage & M. T. Furse, 1984. A preliminary classification of running water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. Freshw Biol 14: 221–256.