Ecotoxicological testing of sediments and dredged material: an overlooked opportunity?

Susanne Heise, Marc Babut, Carmen Casado, Ute Feiler, Benoit J D Ferrari, Laura Marziali

HAL Id: hal-03252994
https://hal.inrae.fr/hal-03252994

Submitted on 8 Jun 2021

ADVANCES IN SEDIMENT SCIENCE AND MANAGEMENT

Ecotoxicological testing of sediments and dredged material: an overlooked opportunity?

Susanne Heise[1] · Marc Babut[2] · Carmen Casado[3] · Ute Feiler[4] · Benoit J. D. Ferrari[3] · Laura Marziali[5]

Abstract

Purpose Basing decisions for the management of contaminated sediments on ecotoxicological data is still often met with skepticism by European stakeholders. These concerns are discussed as they pertain to bioassays to show how ecotoxicological data may provide added value for the sustainable management of sediment in aquatic systems.

Materials and methods Five "concerns" are selected that are often raised by stakeholders. The ecotoxicological practice is discussed in light of the knowledge gained in recent decades and compared with chemical sediment analysis and chemical data.

Results and discussion Common assumptions such as a higher uncertainty of biotest results for sediments compared to chemical analyses are not supported by interlaboratory comparisons. Some confusion also arises, because the meaning of biotest data is often misunderstood, questioning their significance in light of a limited number of organisms and altered test conditions in the lab. Because biotest results describe a sediment property, they should not be directly equated with an impact upon the biological community. To identify a hazard, however, the possibility of false-negative results due to the presence of contaminants that are not analyzed but are toxic is lower.

Conclusions The cost of increased investment in ecotoxicological tests is, in our view, small compared with that of making false-negative assessments of sediment/dredged material that can ultimately have long-term environmental costs. As such, we conclude that ecotoxicological testing is an opportunity for sediment management decision-making that warrants more attention and confidence in Europe.

Keywords Sediment · Ecotoxicological testing · Bioassays · Stakeholders

1 Introduction

Ecotoxicological testing as an approach to assess sediment quality was first described in the late 1970s (e.g., Anderson

and Prater 1977; Swartz et al. 1979) and gained interest in the following years. Giesy et al. (1988) have reasoned that chemical analyses of sediments should be complemented by toxicity tests, because the large number of potentially toxic substances would make the assessment of potential biotic impacts time consuming and costly, and neither bioavailability nor interactions among substances could be accounted for. Consequently, toxicity tests of elutriates, porewater, and/or sediments were introduced into management guidelines. Among the first examples, the US Army Corps of Engineers and the US Environmental Protection Agency developed a decision-making framework for the management of dredged material in 1985 that included biotesting of dredged material in addition to chemical evaluation if aquatic disposal was the preferred option (Lee and Peddicord 1988). Dutka et al. (1988) have stated the importance of applying biotests in combination ("biotest battery") in assessing sediment quality, owing to the differing sensitivity of organisms. Several years later, Chapman (1990) described the sediment quality triad

✉ Susanne Heise
  Susanne.Heise@haw-hamburg.de

1   Hamburg University of Applied Sciences, Ulmenliet 20,
    Hamburg D-21033, Germany

2   RiverLy Research Unit, INRAE, 5, avenue de la Doua,
    F-69625 Villeurbanne, France

3   Swiss Centre for Applied Ecotoxicology (Ecotox Centre),
    CH-1015 Lausanne, Switzerland

4   Federal Institute of Hydrology, Am Mainzer Tor 1,
    D-56068 Koblenz, Germany

5   CNR-IRSA Water Research Institute, Italian National Research
    Council, Via del Mulino 19, I-20861 Brugherio, MB, Italy

approach, in which a combination of chemical and ecotoxicological data, in addition to information on the biological community, was proposed as a powerful tool to determine pollution-induced degradation.

In Europe, compared with the USA, ecotoxicological testing started later but gained attention in light of the large amounts of sediments that must be dredged and managed in ports such as Rotterdam, Antwerp, and Hamburg. The Port of Hamburg had ecotoxicological tests performed on samples of dredged material for the first time in the mid-1990s, and around the same time, the Federal Institute of Hydrology in Germany began gathering biotest data on sediments along the Rhine and Elbe rivers. Since then, ecotoxicological investigation of sediment and dredged material has become much more common, but inclusion in national regulatory frameworks of European states remains slow. In an overview in 2003, den Besten et al. (2003) analyzed the regulatory implementation of bioassays in sediment and dredged material assessment in Europe. They showed a large variation among European countries in decision-making frameworks, which ranged from solely chemical-based assessment to biotests playing an important role in decision support systems. During a workshop in 2018, co-organized by SedNet and the NSR-Interreg project Sullied Sediments, data on how regulations in Europe have changed over time, as well as information regarding the roles of biological effect-based data, were gathered. Although a comparison between 2003 and 2018

showed some manifestation of bioassays in decision-making frameworks, the opposite trend has also been observed (Fig. 1). The Netherlands, which relied on bioassays for their dredged material assessment in 2003, has now excluded them from their regulations. Italy, in contrast, significantly increased the extent of biotesting and integrated biotesting into the assessment of marine dredged materials before management. Other countries, such as France and Italy, have included bioassays in their national guidelines, but have not yet integrated them into national regulations.

Even in countries with long histories of bioassay use, such as France, The Netherlands, and Germany, and decades after the first inclusion of ecotoxicological testing in decision-making frameworks, concerns remain among decision-makers and other stakeholders. The following concerns may explain why ecotoxicological testing and data are often met with skepticism:

– "Biotest results are much less precise and accurate than chemical data."
– "The low number of test organisms cannot represent the ecosystem sensitivity."
– "Because of altered environmental conditions, laboratory testing does not reflect natural conditions and thus cannot be related to bioavailability in situ."
– "Agreement on how to assess biotest data is lacking."
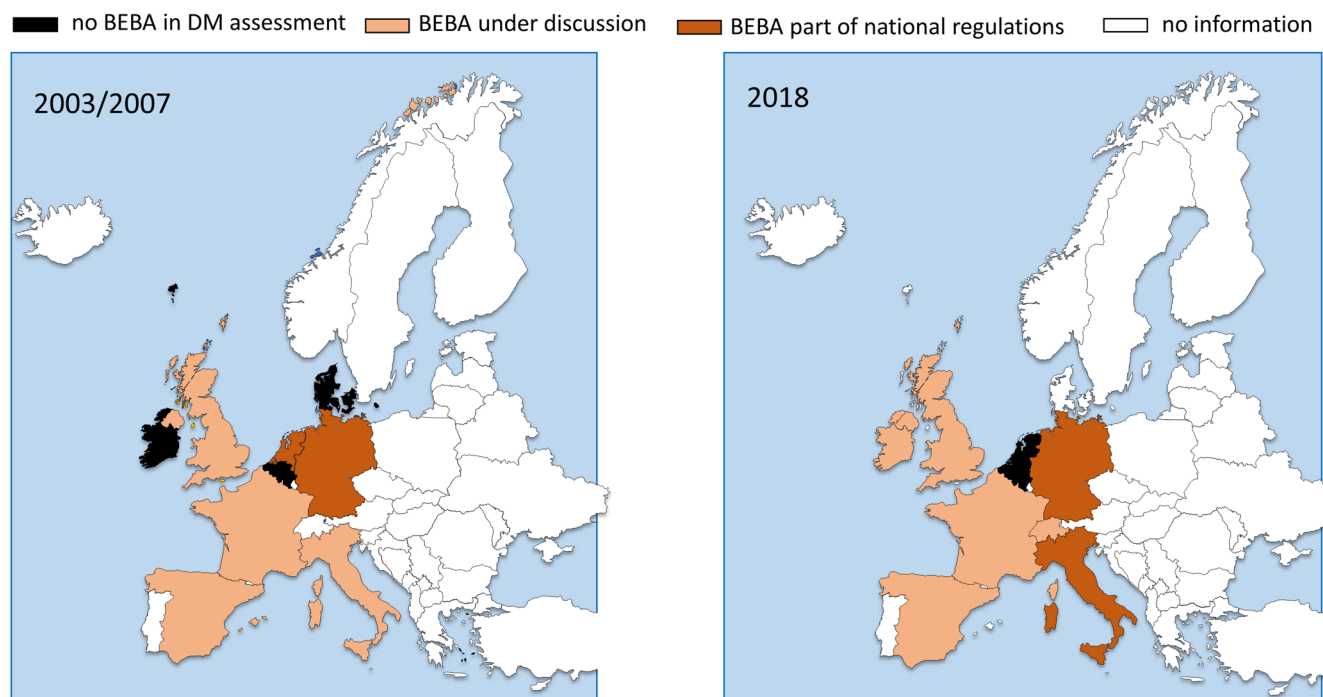

Fig. 1 Status of the inclusion of biological effect-based assessments (BEBA) into national regulatory frameworks for dredged material (DM) in European states in 2003/2007 compared to 2018 (based on den Besten et al. 2003; den Besten 2007 and the outcome of the SedNet & Sullied Sediments Workshop 2018)

– "Biotesting significantly increases the costs of sediment management."

The authors of this article observe, with reservations of their own, that misconceptions and assumptions, which may be mistaken for facts, can lead to decisions that are harmful to the environment and could be avoided. In this article, we thus scrutinize each of those concerns in light of what we have learned over the past several decades, and we hope to start a discussion on how stakeholders and scientists can work together to improve decision-making in the management of sediments and dredged material.

2 Responses to stakeholders' concerns

2.1 "Biotest results are much less precise and accurate than chemical data"

If chemical analysis were a precise and accurate tool, different laboratories should detect very similar concentrations of compounds in the same sediment sample, and the same laboratory should detect the same concentration when measuring a sample several times. Interlaboratory proficiency-testing exercises allow participating laboratories to test their regular in-house analytical methods. For regulatory bodies, they serve as an external quality control evaluation of monitoring data.

Table 1 summarizes the outcomes of the interlaboratory exercises in the analysis of contaminants in sediments, organized by the International Atomic Energy Agency (IAEA) in 1998, 1999, and 2001. The interlaboratory coefficient of variation (CV) for organic compounds exceeded 100% for some analytes in all chemical substance groups tested. According to de Mora et al. (2007), these outcomes were in line with those of other proficiency tests, e.g., from Quality Assurance of Information in Marine Environmental Monitoring (QUASIMEME), and no improvement in analytical performance was detected between 1996 and 2006. However, for

interlaboratory comparisons, laboratories perform analyses according to their in-house protocols. Thus, methods and potentially sediment pretreatment may differ. Consequently, interlaboratory differences tend to be larger than intralaboratory differences. In an exercise conducted by the National Institute of Standards and Technology (NIST) on analysis of PAHs, PCBs, chlorinated pesticides, and PBDE congeners, the CVs for three replicates, indicating the precision of in-house analyses, were below 10% for most laboratories but sometimes exceeded 50%, depending on the substance (Schantz et al. 2006).

For trace metals, the reproducibility and accuracy of analyses are better when outliers are removed. Table 2 shows results from the IAEA-405 intercomparison exercise for arsenic and heavy metals that are routinely analyzed in most sediment and dredged material samples. With the exception of cadmium, all CVs were below 20% and thus were considered acceptable (Wyse et al. 2004). The data also showed a wide range of results when outliers were not removed, thus demonstrating the uncertainty that can accompany chemical data, particularly if quality control procedures are omitted.

With regard to sediment bioassays, Dillon (1994) has stressed the necessity for intra- and interlaboratory comparison before these bioassays are included in regulatory decision support systems. Standardization procedures, and in this context round robin tests, are usually a prerequisite before bioassays are relied upon. Interlaboratory comparisons that precede standardization of bioassays are usually conducted on spiked rather than natural sediment or water samples. For this discussion, however, the focus is on the reproducibility of the results for natural (environmental) samples from the following selected biotests that are being or have been used for sediment and dredged material assessment in Europe, such as the whole sediment assays with *Corophium volutator* (amphipod), *Echinocardium cordatum* (sea urchin), and *Caenorhabditis elegans* (nematode); solid-phase tests with *Aliivibrio fischeri* (luminescent bacteria); sediment contact tests with *Myriophyllum aquaticum* (aquatic plant); and elutriate tests

Table 1 Reproducibility of chemical analyses of sediment contaminants in IAEA interlaboratory assessments (coefficient of variation, CV range in % for different compounds), compiled from Villeneuve et al. (2000, 2002), de Mora et al. (2007), and Wyse et al. (2004)

Analytes	CV range for compounds/ congeners measured	IAEA code	Number of reporting laboratories
Petroleum hydrocarbons	20–100%	383 (1998)	22
	10–140%	408 (1999)	24
	6–140%	417 (2001)	55
Organochlorine pesticides	47–150%	383 (1998)	48
	41–120%	408 (1999)	43
	45–140%	417 (2001)	78
PCBs	13–150%	408 (1999)	25
	1–160%	417 (2001)	53
Trace metals	2–117%	405 (1998)	86

Table 2 Extracted results of the intercomparison exercise IAEA-405 for commonly regulated trace metals in sediments (Wyse et al. 2004)

Analyte	Accepted results (outlying means excluded)				All reported results	
	No. of labs	Reference value (mg kg^{-1})	Range of lab means (mg kg^{-1})	CV (%)	No. of labs	Range of lab means (mg kg^{-1})
As	47	23.6	16.8–28.1	10.5	62	4.272–275
Cd	63	0.73	0.124–1.26	28.1	89	0.056–11.69
Cr	63	84	50.8–119.2	17.8	93	0.292–141.0
Cu	80	47.7	35.0–60.0	10.6	101	0.030–229.7
Hg	60	0.81	0.494–1.193	17.1	75	0.013–8.200
Ni	61	32.5	19.95–47.10	16.4	76	0.039–686
Pb	74	74.8	50.5–100	12.6	102	0.063–125.019
Zn	87	279	184.9–378	12.3	106	0.326–831.2

with an embryo-larvae development bioassay with *Crassostrea gigas* (oyster), *Paracentrotus lividus* (sea urchin), and *Daphnia magna* (water flea) (Table 3).

Inter- and intralaboratory CVs differ strongly between bioassays. The variability would be expected to be highest for tests with larger organisms, and/or if test organisms are sampled from the field and not cultured in a laboratory. The latter is often the case for marine test species. Test organisms, collected in the field, are genetically more diverse, usually have longer life cycles, and fewer organisms are used per test replicate. However, the CVs for tests with marine organisms are most often within the range reported for ISO sediment toxicity tests performed with freshwater laboratory-cultured organisms (e.g., Feiler et al. 2014).

Because of the heterogeneity of sediments, lower precision might have been assumed for direct contact tests. Available data (Table 3) does not support this general assumption. Moreover, most results are well within the commonly accepted criterion of a CV of less than 30 to 40% (Environment Canada 1990; Moore et al. 2000).

On the basis of these interlaboratory comparisons, the assumption that ecotoxicological results in general are less reliable than chemical data can thus not be confirmed. Despite the more recent use of biotesting compared to chemical analytics, and although biological organisms naturally have variable phenotypes, CVs in chemical and ecotoxicological results for sediment quality assessment are in the same range, and sediment contact tests are not necessarily less reproducible than elutriate tests.

However, sampling methodology (pooling of subsamples, homogenization, and sample volume) has been shown to have a large effect on the reproducibility of solid-phase toxicity, and it cannot be recorrected after a sample is brought to the laboratory (Ferrari et al. 1999). Similarly, sampling has a key influence on the variability of the results of chemical characterization, often to a greater extent than analytical variability (Schiavone et al. 2011). Moreover, sediment storage and pretreatment significantly affect test results (e.g., De Lange et al. 2008).

Another major source of interlaboratory variability in ecotoxicological testing, as suspected by Stronkhorst et al. (2004), is the degree of experience of laboratory technicians with bioassays. Effort should be made to provide specific training in performing ecotoxicological tests if the results are used for regulatory purposes. This aspect is particularly important when the evaluation of test endpoints has some degree of subjectivity, such as the development of sea urchin larvae (Casado-Martínez et al. 2006b). One possibility for improving the performance of ecotoxicological testing of laboratories is the initiation of frequent interlaboratory comparisons for bioassays.

Although here we compare the precision of numerical endpoint results with those from analytical techniques through CVs, the use of CVs alone for assessing the results of toxicity tests has been challenged. Whereas extremely toxic or nontoxic samples may result in very low CVs (Burton Jr. et al. 1996), good agreement in the classification of samples according to toxicity and no toxicity may also be achieved with high CVs (Thursby et al. 1997; Casado-Martínez et al. 2006a, b). As Norberg-King et al. (2006) indicated: "it is important to keep in mind that the purpose of a toxicity test is not to find statistical differences; rather, it is to decide, with an acceptable degree of uncertainty, whether a sample is toxic."

In conclusion The statement that biotest results are generally less precise and accurate than chemical data cannot be confirmed. Nevertheless, more intra- and interlaboratory comparisons would help to harmonize procedures (sampling, pretreatment, and standard operation procedures) and to train technicians.

2.2 "The low number of test organisms cannot represent the ecosystem sensitivity"

This statement refers to the application of biotests to assess in situ sediment quality and to protect the environment against stress from contaminants. The sensitivity and stress levels of

Table 3 Examples of reproducibility (coefficient of variation, CV in %) of sediment toxicity tests in interlaboratory assessments

Test matrix	Species	Outcome	Reference
Solid-phase/whole sediment	10-day amphipod survival (*Corophium volutator*)	$CV_{interlab} = 15\%$ (4 laboratories from which one excluded, 3 moderately contaminated sediments) $CV_{intralab} \leq 11\%$	Stronkhorst et al. (2004)
	10-day amphipod survival (*Ampelisca brevicornis, Corophium volutator, Corophium multisetosum,* and *Microdeutopus gryllotalpa*)	$CV_{interlab} = 33\%$ (5 laboratories after excluding results from sediment storage longer than 4 weeks, 6 low to highly contaminated sediments)	Casado-Martínez et al. (2006a)
	14-day sea urchin survival (*Echinocardium cordatum*)	$CV_{interlab} = 2$ and 9% (4 laboratories from which one excluded, results for 2 moderately contaminated sediments causing low toxicity)	Stronkhorst et al. (2004)
	72-h nematode growth and reproduction (*Caenorhabditis elegans*)	$CV_{interlab}$ (growth) < 15% $CV_{interlab}$ (reproduction) < 40%	Höss et al. (2012)
	30-min bacterial bioluminescence inhibition (*Aliivibrio fischeri*)	[1] $CV_{interlab} = 12\%$ [2] $CV_{interlab} = 48$ and 58% (average for 2 consecutive interlab comparisons and 2 test protocols)	[1] Stronkhorst et al. (2004) [2] Casado-Martínez et al. (2006b)
	10-day macrophyte growth (*Myriophyllum aquaticum*)	$CV_{interlab}$ (growth rate) < 25%, $CV_{interlab}$ (growth inhibition) < 40% (average for 4 sediments and 21 laboratories, 86% laboratories met validity criterion)	Feiler et al. (2014)
Elutriate/pore water	48-h oyster embryo-larvae development (*Crassostrea gigas*)	$CV_{interlab} = 100\%$ (average for 3 laboratories, 3 moderately contaminated sediments, elutriates produced by testing lab)	Stronkhorst et al. (2004)
	Paracentrotus lividus (sea urchin), embryo-larvae development	$CV_{interlab} = 58\%$ (average for 4 laboratories and 6 from low to highly contaminated sediments, elutriates produced at testing lab)	Casado-Martinez et al. (2006b)
	48-h *Daphnia magna* (water flea)	$CV_{interlab} = 33\%$, $CV_{intralab} = 16\%$ (effluents)	Grothe and Kimerle (1985)

an ecosystem can be best assessed by studying the benthic community. However, changes in diversity can also be due to noncontaminant stressors, such as temperature or light; therefore, the "triad approach" combines benthic community data with toxicity data (and chemical data) (Chapman et al. 1997). Sometimes hypothetical "most sensitive test organisms" reflecting the sensitivity of the biological community have been desired to allow for cost efficient and fast determination of the chemical stress in situ.

This statement misinterprets the importance of ecotoxicological testing, and the search for the most sensitive organism will not be successful anyway, as Cairns (1986) has explained. Species differ in sensitivity toward chemicals with different modes of action; the same species may be very sensitive to substance A yet tolerant to substance B. Consequently, the search for a species "representative" of an ecosystem's status is necessarily flawed. What we can expect from a biotest is information, such as the presence or availability of (undefined) substances that have the potential to disturb and affect organisms in the field. If basic biological traits are inhibited, such as photosynthesis, reproduction, or energy metabolism, the probability of implications for the ecosystem rises.

For the selection of a test species for sediment toxicity test development, practical reasons will prevail (e.g., availability and handleability). The utility of such tests can be greatly improved if the proposal for a species is accompanied by appropriate information regarding its sensitivity to contamination, its ecological importance, and its exposure pathways (Dillon 1994). As an indicator of potential risk to the biological community, a given biotest must be sensitive to chemical stress. An excessive tolerance would increase the likelihood of false-negative responses. Accordingly, field validation is needed, during which reactions of biotests are compared with measurable changes in the biological community, so that regulatory agencies can assess the relevance of bioassay results. Although there has been some debate regarding the need for field validation for sediment toxicity testing (Chapman 1995), a workshop to evaluate the uncertainty of measurement endpoints used in sediment ecological risk assessment highlighted the inadequate field validation of sediment toxicity tests in 1996 (Ingersoll et al. 1997). To overcome this bottleneck, several initiatives in the USA have demonstrated the ecological relevance of amphipod sediment toxicity testing. Long et al. (2001) have studied the relationship between acute sediment toxicity tests with marine and estuarine amphipods and benthic community structure metrics (abundance and diversity) in more than 1400 samples from studies conducted in the USA. Although the authors found considerable variability among the datasets, they concluded that ecologically relevant

losses in the abundance and diversity of the benthic infauna frequently corresponded to decreased amphipod survival in laboratory tests. In > 90% of the samples classified as toxic, at least one measure of benthic diversity or abundance was < 50% of the average reference value. No amphipods were found in 39% of samples classified as toxic, although amphipods were also absent from 28% of the nontoxic samples. However, the abundance of crustaceans (notably amphipods) decreased in the infauna as amphipod survival decreased in the laboratory tests in many of the studied areas. A break point in the data indicated that, in general, amphipod abundance in the field was lowest when survival in the laboratory tests decreased below 50% that of controls.

A field validation study was also completed at a PAH-contaminated Superfund site USA, involving a 10-day toxicity test with the marine amphipods *Leptocheirus plumosus* and *Rhypoxynius abronius* (Ferraro and Cole 2002). Both toxicity tests were validated as indicators of changes in several macrofaunal community metrics that had low but sufficient statistical power to discriminate ecologically important effects: the percentage loss of the indices increased relative to values determined for nontoxic reference areas, as the average survival in laboratory toxicity tests decreased. Losses of benthic resources reached 50% when the survival dropped to 0%.

According to Borgmann et al. (2005), the freshwater amphipod *Hyalella azteca* is frequently one of the most sensitive organisms in sediment toxicity tests, according to the results of risk assessment of chemical substance registration. A close correlation between toxicity to *H. azteca* in laboratory toxicity tests and an abnormally low abundance of sensitive benthic species, such as amphipods, mayflies, sphaeriid clams, and tanytarsid midges, in the field has been shown to predict effects on sensitive species in situ.

The nematode *Caenorhabditis elegans* has become another frequently used organism in freshwater sediment toxicity tests. Haegerbaeumer et al. (2018) have compared the sensitivity of 27 wild nematode species extracted from freshwater sediments with that of *C. elegans* toward metals and PAHs. Although *C. elegans* is more tolerant to chemical stress than the average freshwater nematode species, the sensitivity of the extracted animals varied over a wide range, and the *C. elegans* responses were well within that range, except for benzo[a]pyrene.

In conclusion Single-test species should not be expected to represent the sensitivity of ecosystems but should be regarded as indicators of available and harmful substances that may affect biological communities. From this perspective, more information must be compiled and provided on the sensitivity of test species toward relevant substances in comparison to that of biological communities, to provide information on the possibility of false-negative outcomes in batteries of biotests.

2.3 "Because of altered environmental conditions, laboratory testing does not reflect natural conditions and thus cannot be related to bioavailability in situ"

As indicated above, a direct extension of laboratory results to situations in situ is certainly not possible, as also described by Ferrari et al. (2019). The above statement reveals a misunderstanding of the purpose of performing biotests. These tests are intended to show whether there is a hazard for the aquatic or benthic community.

Therefore, experimental conditions may change as long as they remain environmentally relevant, and different scenarios may be tested.

In Europe, discussions of biotests in a regulatory context apply primarily to dredged material assessment. Deciding on management options for dredged material requires deciding on treatments, during which the material undergoes several physico-chemical changes, as do sediment samples in preparation for ecotoxicological testing. Bioassays usually require oxic conditions and more water than was present in the original sediment. Resuspension in a greater water volume and oxidation of samples will also occur during relocation of dredged material, and thus preparation of samples for biotesting simulates realistic conditions. Test conditions such as pH, temperature, or salinity, however, depend on the requirements of the given test organisms and must be kept within a certain range, even if it does not reflect the *environmental* situation. Ecotoxicological tests must be understood not to predict with high certainty what will happen in the environment but to characterize environmental samples on the basis of their properties under fixed conditions. The toxicity measured in the laboratory reflects the capability of the sediment to do harm under certain conditions and thus indicates toxicity potential. The information on ecotoxicity becomes meaningful in an environmental context, considering the exposure. Management decisions, e.g., to dredge or to relocate sediment, should be performed on the basis of its toxic potential to ensure that the material's properties do not adversely affect the environment.

The same applies to chemical data for sediments. Bioassays, elution, or leaching tests are performed in standardized conditions that do not necessarily represent the in situ bioavailability of contaminants. Moreover, as reported above, sampling strategy, storage, and pretreatment may also alter contaminant bioavailability, thus affecting the results of chemical analysis. For example, De Lange et al. (2008) have reported the analysis of acid volatile sulfide (AVS) and simultaneously extracted metals (SEM) stored under different conditions: AVS increased significantly during cool storage, whereas SEM was not affected. The authors found different AVS values according to the sediment layer (i.e., 0–2 cm vs. 2–5 cm). In addition, the choice of digestion procedure may significanlty affect the results of trace element analysis (e.g.,

Mossop and Davidson 2003). Therefore, chemical analysis, like ecotoxicological tests, is performed to characterize environmental samples under fixed conditions, which may not reflect the in situ status.

In conclusion Both biotesting and chemical analysis characterize sediments under standardized conditions that do not necessarily represent in situ conditions. Despite this limitation, both can reveal potential hazards to aquatic communities. However, ecotoxicological tests are more powerful in detecting the effects of pollutant mixtures and of chemicals that are not assayed.

2.4 "Agreement on how to assess biotest data is lacking"

Interpretation of individual test results within a biotest battery is performed differently depending on the laboratory and the guidelines. For whole organism tests, the results are often expressed as percentage inhibition of a certain endpoint such as mortality, photosynthesis, growth, or reproduction, compared with that of an unaffected control. Thresholds that differentiate "toxicity classes," indicating, e.g., low, moderate, or high toxicity, often appear to be set arbitrarily and not to account for the characteristics of the test systems.

Different toxicity endpoints of different organisms have different response ranges, sensitivity, and precision, thus calling into question the use of strict threshold values in ecotoxicology (Ahlf and Heise 2005; Höss et al. 2010).

The issue becomes even more complicated in interpreting the results of biotest batteries, because the bioassays usually yield differing responses. Interpretation of multi-test results providing information on sediment or dredged material quality range from always considering the most sensitive organisms in a test battery (e.g., in Germany, according to GÜBAK-WSV (2009)) to integrating biotest data by more complex classification techniques, such as the Hasse diagram technique, fuzzy logic expert systems (Hollert et al. 2002), or toxicity profiling (Hamers et al. 2010). These integrating assessment approaches, although more complicated and less transparent, have the potential to improve decision-making on the basis of sound science and have found acceptance, e.g., in the Italian regulation for disposal of dredged marine sediments at sea in other than National Relevance Sites (SedNet and Sullied Sediments 2018).

In quantifying single and multiple responses in bioassays to assess their relevance in providing information on environmental toxicity of sediments, chemical analyses face a similar problem. Sediment quality guidelines (SQGs) are intended to relate the chemical concentrations in sediments to hazards. They have been developed to protect the biological community, to predict effects on benthic organisms, or both. Most

have been derived through empirical or theoretical/ mechanistic approaches.

Many (controversial) discussions have debated the design, implementation, and limitations of SQGs, thus resulting in a large variation in guidelines. DelValls et al. (2004) have reviewed SQGs from different European countries and have shown that they differ by two orders of magnitude for some substances (e.g., As, Cu, and seven PCBs). Most of the limitations listed and discussed at the Pellston Workshop on "use of sediment quality guidelines and related tools for the assessment of contaminated sediments" in 2002 have not been addressed to date for existing SQGs; e.g., they deliver no or limited information on the ecologically important aspects of chronic toxicity to sediment-dwelling organisms and cause-effect relationships, in addition to the questionable transferability of SQGs, derived from one endpoint in the laboratory, to, e.g., effects on organisms in the environment (Wenning et al. 2005). Moreover, existing SQGs cover tens of substances at best, and therefore substances of emerging concern cannot be reliably assessed with this tool. The same applies to chemical guidelines developed by countries to manage dredged material (action levels). These action levels vary substantially among countries and cover only a very limited number of substances (see Röper and Netzband 2011).

In conclusion There is indeed no agreement yet on how to assess biotest results, although several approaches that account for test-specific characteristics have been reported. Contrary to the common perception, however, sediment quality guidelines and action levels also substantially vary among countries and, even if effect-based, have limited ability to predict adverse effects or protect benthic communities. Complementary application of chemical analyses and ecotoxicological testing still appears to be the best way to decrease the probability of false-negative results from sediment or dredged material analyses. Sediment toxicity testing with carefully selected organisms to target contaminants with a special mode of action could become a cost-effective monitoring technique.

2.5 "Biotesting significantly increases the costs of sediment management"

A brief study of testing costs was performed during drafting of a guidance document dedicated to the hazard assessment of sediments in French waterways (Stamm and Babut 2019). Two types of biotests were considered: miniaturized tests intended for a screening tier and classic biotests intended for an in-depth assessment if the screening tier did not lead to making a decision. The associated costs are shown in Table 4.

Well-known, commonly used tests, such as ostracod (ISO 14371) or Microtox™, have unit costs similar to the costs of "simple" analyses, such as those of trace elements, PCBs

(except dioxin-like congeners) or PAHs, which are mostly automated in chemical laboratories. Other tests appear to be more expensive for several reasons. A longer test duration (e.g., *Gammarus*), which entails a higher workload, leads accordingly to a more expensive test. The cost cited by potential contractors is also influenced by the potential demand (i.e., the number of tests the contractor expects to perform), which in turn is associated with the investment needed and the number of laboratories accredited for those tests.

Thus, currently, according to these tariffs, the cost of the screening tier would amount to approximately 2000 € per sample when biotests are included or 1000 € when they are not, in which case, the list of chemicals analyses would be limited to trace elements, PCBs, except dioxin-like congeners, and PAHs. If the concentrations of more chemicals of emerging concern are required in a screening tier, only the cost of chemical analyses would increase, because the biotest battery would have the same level of response (i.e., taking into account that bioassays assess the potential effects of all chemicals in the samples).

Costs can be further cut substantially with the use of smaller test organisms, such as bacteria and algae, thus enabling the test procedures to be miniaturized (Rojíčková et al. 1998; Heise and Ahlf 2005; Wadhia and Thompson 2007; Paixão et al. 2008).

More broadly, including any additional lines of evidence in the assessment framework would increase the overall expense leading to decision. This trend would be true for not only biotests but also for chemicals of emerging concern beyond the current lists of priority substances.

In conclusion In our view, the cost issue should be discussed in relation to the needs—what information is required to reach a decision—and the cost of mismanagement, that is, of making a wrong decision. With analytical and bioassay data complementing each other, the risk of false-negative results which would guide the decision in the wrong—and costly— direction would be reduced. Objecting that biotests are expensive does not make much sense: it is a simplistic argument with no rational grounds.

Table 4 Expected unit costs (excl. VAT) for a range of biotests. *na* not available

Organism	Standard	Cost range (€)
C. elegans	NF ISO 10872	850–1000
H. incongruens	NF ISO 14371	250–525
A. fischeri	Microtox	90–240
C. riparius	XP T90-339-1	≈ 500
Gammarus spp.	na	1700

3 Opportunities in biotesting environmental samples

3.1 Environmentally safer decisions

The following examples demonstrate the risk of overlooking the adverse effects that chemicals that are not routinely measured or unknown might have on the environment if ecotoxicological testing of sediments is not performed.

- de Baat et al. (2019) have performed sediment toxicity tests with the midge *Chironomus riparius* and chemical analyses in sediments from 12 areas in the Netherlands, where the major pollution sources were urban, agricultural, or from wastewater treatment plant (WWTP). They measured traditional contaminants, such as metals and PAHs, as well as emerging contaminants, which usually are not included in monitoring programs (WWTP markers such as bisphenol A and nonylphenol; pesticides such as prosulfocarb and triallate). Although the overlying water did not show toxicity to the sensitive invertebrate *Daphnia magna*, sediments from all sites had lethal and/or sublethal effects on midge larvae (*Chironomus riparius*). Chemical analyses showed that metals and PAHs were predominantly present in the pore waters of sediments at urban locations. Sediments from the areas of the WWTP and agricultural use, however, were dominated by WWTP markers and pesticides. The effects of these sediments were more pronounced than those from urban area material. This impact would not have been detected if only the "usual" list of contaminants had been measured and no ecotoxicological testing had been performed.

- Feiler et al. (2013) applied a test battery consisting of five sediment contact tests—a plant test (*Myriophyllum aquaticum*, ISO 16191), nematode test (*Caenorhabditis elegans*, ISO 10872), oligochaete test (*Lumbriculus variegatus*, OECD 225), fish embryo test (*Danio rerio*, on the basis of DIN 38415–6), and bacteria test (*Arthrobacter globiformis*, ISO/DIS 10871)—to 21 native freshwater sediments characterized by a broad variety of geochemical properties and anthropogenic contamination. On the basis of the toxicity pattern derived from the test battery, the sediments were assessed according to a classification system for sediment toxicity. For sediments with high toxicity potential, the test-derived classification agreed well with the application of consensus-based sediment quality guidelines, whereas in sediments with low to medium toxic potential, SQGs often underestimated the toxicity detected by the sediment contact tests.

These examples also demonstrate the importance of performing tests in direct contact with sediments, although

also elutriate or pore water tests are meaningful, given the possibility of sediment resuspension into the water phase.

- Claus et al. (2009) have investigated the observation that phytotoxicity of elutriates and pore water from Elbe river sediments in Germany showed an increasing trend between 1992 and 2007, although the overall contamination decreased substantially since the fall of the Iron Curtain in 1989. Through an effect-directed analytical approach, the authors were able to characterize the substance that caused high toxicity in algae growth inhibition tests with *Desmodesmus subspicatus* as a low volatile, thermostable, nonpolar, and highly lipophilic compound, but were unable to structurally identify the contaminant.

3.2 Less overprotective measures

Another opportunity for ecotoxicological testing is the identification of cases in which low bioavailability of contaminants decreases the need for costly management actions.

- Marziali et al. (2017a) have reported the results of test batteries applied on sediments collected in a North Italian reservoir where arsenic was the contaminant of concern, showing values up to 20 times higher than the chemical threshold (33 mg kg^{-1} d.w., according to the Probable Effect Concentration by MacDonald et al. 2000), which were significantly higher than the concentrations in the downstream river. To recover water storage capacity and maintain proper functioning of the dam, desiltation was necessary. Even if they were derived mainly from natural weathering, such values of arsenic would normally result in the mechanical removal of tons of sediments from the reservoir, with high costs and low potential for reuse. However, the application of test batteries on whole sediments and elutriates showed no/slight toxicity to test organisms, thus demonstrating low bioavailability of the toxicant, even after sediment remobilization and mixing. According to these results, sediment flushing in the downstream river stretch was considered feasible, according to a proper operational plan. Similar results were obtained in other North Italian reservoirs where SQGs were exceeded because of the accumulation of geogenic trace metals, atmospheric deposition, or local anthropogenic enrichment (Marziali et al. 2017b).

There are few examples of cases of management issues in which low toxicities despite high contamination have led to less strict management decisions, because the regulatory context for SQG is clear, and confidence in chemical data is high. While the potential of information on nonbioavailable substances from combined chemical and ecotoxicological data

should not be underestimated in decision-making frameworks (Ahlf et al. 2002, 2008), analysis would require intensified biotesting with whole sediment bioassays to simulate the long-term exposure of organisms in the downstream river stretch, and chronic or subchronic tests, such as the whole sediment test with *Heterocypris incongruens* (Marziali et al. 2017a).

4 Conclusion and outlook: ecotoxicological testing of sediments—an overlooked opportunity

In the ongoing debate with stakeholders on the inclusion of bioassays in sediment management decision frameworks, tests are considered not meaningful but unnecessary and consequently cost inefficient. However, as shown above, their reliability is largely similar to that of chemical analysis of sediment. The examples in Section 3 demonstrate their added value in determining potential risks that would be overlooked if solely chemical data were relied upon. Whether the costs justify the outcome may be a matter of perspective, but if the aim is to perform management activities in a way that respects and protects the environment in a sustainable way, ecotoxicological testing provides the opportunities to do so.

Moreover, this article shows that many challenges still remain, some unique to bioassays and some shared with chemical analyses. The following challenges should be addressed to facilitate future biotesting of sediments for regulatory purposes:

- Standardization of more biotests with a strong focus on sediment contact tests should be promoted.
- The full paradigm for sediment toxicity test development by Dillon (1994) should be followed, and this work should be supported by regulatory agencies.
- More intra- and interlaboratory comparisons should be performed to train laboratory assistants and to reduce the variability of biotest results.
- Research into the automation of biotests should be increased to decrease costs and improve sample turnover.
- To facilitate the interpretation of ecotoxicity data, test-specific effect-based thresholds should be developed.

Expertise in ecotoxicology should be strengthened in industry and regulatory agencies to maximize the opportunities that ecotoxicological testing offers.

Compliance with ethical standards

References

Ahlf W, Heise S (2005) Sediment toxicity assessment: rationale for effect classes. J Soils Sediments 5:16–20

Ahlf W, Hollert H, Neumann-Hensel H, Ricking M (2002) A guidance for the assessment and evaluation of sediment quality a German approach based on ecotoxicological and chemical measurements. J Soils Sediments 2:37–42

Ahlf W, Feiler U, Heininger P, Heise S (2008) Monitoring sediment quality using toxicity tests as primary tools for any risk assessment. In: Quevauviller P, Borchers U, Thompson KC, Simonart T (eds) The water framework directive—ecological and chemical status monitoring. Water Quality Measurements Series. John Wiley and Sons, pp 255–269

Anderson BL, Prater MA (1977) A 96-hour sediment bioassay of Duluth and Superior Harbor basins (Minnesota) using *Hexagenia limbata*, *Asellus communis*, *Daphnia magna*, and *Pimephales promelas* as test organisms. Bull Environ Contam Toxicol 18:159–169

Borgmann U, Grapentine L, Norwood WP et al (2005) Sediment toxicity testing with the freshwater amphipod *Hyalella azteca*: relevance and application. Letter to the Editor. Chemosphere 61:1740–1743

Burton GA Jr, Norberg-King TJ, Benoit DA, Ankley GT, Winger PV, Kubitz J, Lazorchak JM, Smith ME, Call DJ, Ingersoll CG, Greer E, Dwyer FJ, Day KE, Kennedy P, Stinson M (1996) Interlaboratory study of precision: *Hyalella azteca* and *Chironomus tentans* freshwater sediment toxicity assays. Environ Toxicol Chem 15:1335–1343

Cairns J (1986) The myth of the most sensitive species. BioScience 36:670–672

Casado-Martínez MC, Beiras R, Lloret J et al (2006a) Inter-laboratory assessment of marine bioassays to evaluate environmental quality of coastal sediments in Spain: III. Bioassay using larval stages of the marine sea urchin Paracentrotus lividus (Lamark, 1816). Ciencias Marinas 32:139–144

Casado-Martínez MC, Campisi T, Díaz A et al (2006b) Inter-laboratory assessment of marine bioassays to evaluate environmental quality of coastal sediments in Spain: II. The bioluminescence inhibition test for rapid sediment toxicity assessment. Ciencias Marinas 32:129–138

Chapman PM (1990) The sediment quality triad approach to determining pollution-induced degradation. Sci Total Environ 97-98:815–825

Chapman PM (1995) Do sediment toxicity tests require field validation? Environ Toxicol Chem 14:1451–1453

Chapman PM, Anderson B, Carr S, Engle V, Green R, Hameedi J, Harmon M, Haverland P, Hyland J, Ingersoll C, Long E, Rodgers J Jr, Salazar M, Sibley PK, Smith PJ, Swartz RC, Thompson B, Windom H (1997) General guidelines for using the sediment quality triad. Marine Pollut Bull 34:368–372

Claus E, Neumann-Hensel H, Heininger P et al (2009) Effektorientierte Untersuchungen zur Algentoxizität fraktionierter Porenwässer und Eluate aus Elbesedimenten. UWSF 21:267–271

de Baat ML, Wieringa N, Droge STJ, van Hall BG, van der Meer F, Kraak MHS (2019) Smarter sediment screening: effect-based quality assessment, chemical profiling, and risk identification. Environ Sci Technol 53:14479–14488

De Lange HJ, Van Griethuysen C, Koelmans AA (2008) Sampling method, storage and pretreatment of sediment affect AVS concentrations with consequences for bioassay responses. Environ Pollut 151:243–251

de Mora S, Villeneuve J-P, Wyse E (2007) An overview of laboratory performance studies conducted by the IAEA for marine pollution studies of metals and organic contaminants. Accred Qual Assur 12:587–592

DelValls TA, Andres A, Belzunce MJ et al (2004) Chemical and ecotoxicological guidelines for managing disposal of dredged material. Trends Anal Chem 23:819–828

den Besten P (2007) Risk assessment approaches in European countries. In: Heise S (ed.) Sediment risk management and communication. Sustainable management of sediment resources. Elsevier, Amsterdam, pp 153–205

den Besten PJ, de Deckere E, Babut MP, Power B, DelValls TA, Zago C, Oen AMP, Heise S (2003) Biological effects-based sediment quality in ecological risk assessment for European waters. J Soils Sediments 3:144–162

Dillon TM (1994) A paradigm for developing sediment toxicity bioassays for the regulatory evaluation of dredged material. Dredging Operations Technical Support Program US Army Corps of Engineers, Waterways Experiment Station. Vicksburg, MS. 26

Dutka BJ, Jones K, Kwan KK, Bailey H, McInnis R (1988) Use of microbial and toxicant screening tests for priority site selection of degraded areas in water bodies. Water Res 22:503–510

Environment Canada (1990) Guidance document on control of toxicity test precision using reference toxicants. Environment Canada, Ottawa

Feiler U, Höss S, Ahlf W, Gilberg D, Hammers-Wirtz M, Hollert H, Meller M, Neumann-Hensel H, Ottermanns R, Seiler TB, Spira D, Heininger P (2013) Sediment contact tests as a tool for the assessment of sediment quality in German waters. Environ Toxicol Chem 32:144–155

Feiler U, Ratte M, Arts G, Bazin C, Brauer F, Casado C, Dören L, Eklund B, Gilberg D, Grote M, Gonsior G, Hafner C, Kopf W, Lemnitzer B, Liedtke A, Matthias U, Okos E, Pandard P, Scheerbaum D, Schmitt-Jansen M, Stewart K, Teodorovic I, Wenzel A, Pluta HJ (2014) Inter-laboratory trial of a standardized sediment contact test with the aquatic plant *Myriophyllum aquaticum* (ISO 16191). Environ Toxicol Chem 33:662–670

Ferrari B, Radetski CM, Veber A et al (1999) Ecotoxicological assessment of solid wastes: a combined liquid and solid-phase testing approach using a battery of bioassays and biomarkers. Environ Toxicol Chem 18:1195–1202

Ferrari BJD, Vignati DAL, Roulier JL, Coquery M, Szalinska E, Bobrowski A, Czaplicka A, Dominik J (2019) Chromium bioavailability in aquatic systems impacted by tannery wastewaters. Part 2:

new insights from laboratory and in situ testing with *Chironomus riparius* Meigen (Diptera, Chironomidae). Sci Total Environ 653:1–9

Ferraro SP, Cole FA (2002) A field validation of two sediment-amphipod toxicity tests. Environ Toxicol Chem 21:1423–1437

Giesy JP, Graney RL, Newsted JL, Rosiuand CJ, Benda A, Kreis RG Jr, Horvath FJ (1988) Comparison of three sediment bioassay methods using Detroit river sediments. Environ Toxicol Chem 7:483–498

Grothe DR, Kimerle RA (1985) Inter- and intralaboratory variability in *Daphnia magna* effluent toxicity test results. Environ Toxicol Chem 4:189–192. https://doi.org/10.1002/etc.5620040208

GÜBAK-WSV (2009) Joint transitional arrangements for the handling of dredged material in German costal waterways by federal authorities and the Federal States (Bundesländer) Bremen, Hamburg, Mecklenburg-Western Pomeriana, Lower Saxonia and Schleswig-Holstein. https://www.bafg.de/Baggergut/DE/04_Richtlinien/guebag_en.pdf?__blob=publicationFile

Haegerbaeumer A, Höss S, Heininger P, Traunspurger W (2018) Is *Caenorhabditis elegans* representative of freshwater nematode species in toxicity testing? Environ Sci Pollut Res 25:2879–2888

Hamers T, Leonards PE, Legler J et al (2010) Toxicity profiling: an integrated effect-based tool for site-specific sediment quality assessment. Integr Environ Assess Manag 6:761–773

Heise S, Ahlf W (2005) A new microbial contact assay for marine sediments. J Soils Sediments 5:9–15

Hollert H, Heise S, Pudenz S, Brüggemann R, Ahlf W, Braunbeck T (2002) Application of a sediment quality triad and different statistical approaches (Hasse diagrams and fuzzy logic) for the comparative evaluation of small streams. Ecotoxicology 11:311–321

Höss S, Ahlf W, Bergtold M, Bluebaum-Gronau E, Brinke M, Donnevert G, Menzel R, Möhlenkamp C, Ratte H-T, Traunspurger W, von Danwitz B, Pluta H-J (2012) Interlaboratory comparison of a standardized toxicity test using the nematode Caenorhabditis elegans (ISO 10872). Environ Toxicol Chem 31:1525–1535. https://doi.org/10.1002/etc.1843

Höss S, Ahlf W, Fahnenstich C, Gilberg D, Hollert H, Melbye K, Meller M, Hammers-Wirtz M, Heininger P, Neumann-Hensel H, Ottermanns R, Ratte HT, Seiler TB, Spira D, Weber J, Feiler U (2010) Variability of sediment-contact tests in freshwater sediments with low-level anthropogenic contamination—determination of toxicity thresholds. Environ Pollut 158:2999–3010

Ingersoll CG, Ankley GT, Baudo R et al. (1997) Workgroup summary report on uncertainty evaluation of measurement endpoints used in sediment ecological risk assessment. In: Ingersoll CG, Dillon T, Biddinger GR (eds) Pensacola FL, 1997. Ecological risk assessment of contaminated sediments. SETAC Pellston Workshop on Sediment Ecological Risk Assessment. SETAC Pr., pp. 297–352

Lee CR, Peddicord RK (1988) Decision-making framework for management of dredged material disposal. In: Salomons W, Förstner U (eds) Environmental management of solid waste—dredged material and mine tailings. Springer, Heidelberg, pp 324–371

Long ER, Hong CB, 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

MacDonald DD, Ingersoll CG, Berger TA (2000) Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39:20–31

Marziali L, Guzzella L, Tartari G et al (2017a) Protocollo per la valutazione della tossicità dei sedimenti degli invasi in Regione Lombardia ai fini della fluitazione. Biologia Ambientale 31:219–225

Marziali L, Rosignoli F, Drago A, Pascariello S, Valsecchi L, Rossaro B, Guzzella L (2017b) Toxicity risk assessment of mercury, DDT and arsenic legacy pollution in sediments: a triad approach under low concentration conditions. Sci Total Environ 593-594:809–821

Moore DRJ, Warren-Hicks W, Parkhurst BR, Teed RS, Baird RB, Berger R, Denton DL, Pletl JJ (2000) Intra- and intertreatment variability in reference toxicant tests: implications for whole effluent toxicity testing programs. Environ Toxicol Chem 19:105–112

Mossop KF, Davidson CM (2003) Comparison of original and modified BCR sequential extraction procedures for the fractionation of copper, iron, lead, manganese and zinc in soils and sediments. Anal Chim Acta 478:111–118

Norberg-King T, Sibley P, Burton GA et al (2006) Interlaboratory evaluation of *Hyalella azteca* and *Chironomus tentans* short-term and long-term sediment toxicity tests. Environ Toxicol Chem 25:2662–2674

Paixão S, Silva L, Fernandes A et al (2008) Performance of a miniaturized algal bioassay in phytotoxicity screening. Ecotoxicology 17:165–171

Rojíčková R, Dvořáková D, Maršálek B (1998) The use of miniaturized algal bioassays in comparison to the standard flask assay. Environ Toxicol Water Qual 13:235–241

Röper H, Netzband A (2011) Assessment criteria for dredged material with special focus on the North Sea Region. Hamburg Port Authority. https://sednet.org/download/Dredged_Material_Criteria_North_Sea_0611.pdf. Accessed 21.12.2019

Schantz MM, Kucklick JR, Parris RM et al. (2006) NIST Intercomparison program for organic contaminants in the marine environment: description and results of 2005 Organic Intercomparison Exercises

Schiavone S, Ghestem JP, Coquery M (2011) Incertitudes liées à l'échantillonnage: exemple d'estimation pour la mesure des micropolluants dans les sédiments. Cemagref. 88 p

SedNet, Sullied Sediments (2018) Report on the workshop on sediment classification and management decisions, *in situ* and *ex situ*. Compiled by Susanne Heise, HAW-Hamburg. Hamburg, Germany, 21 p

Stamm E, Babut M (2019) Guide opérationnel pour l'évaluation du danger des sédiments contaminés: Une démarche graduée pour estimer le danger. Irstea/Agence Française de la Biodiversité. 54 p

Stronkhorst J, Ciarelli S, Schipper CA, Postma JF, Dubbeldam M, Vangheluwe M, Brils JM, Hooftman R (2004) Inter-laboratory comparison of five marine bioassays for evaluating the toxicity of dredged material. Aquat Ecosyst Health Manag 7:147–159

Swartz RC, DeBen WA, Cole FA (1979) A bioassay for the toxicity of sediment to marine macrobenthos. Water Pollut Control Fed 51:944–950

Thursby GB, Heltshe J, Scott KJ (1997) Revised approach to toxicity test acceptability criteria using a statistical performance assessment. Environ Toxicol Chem 16:1322–1329

Villeneuve J-P, de Mora SJ, Cattini C, Carvalho FP (2000) Determination of organochlorinated compounds and petroleum hydrocarbons in sediment sample IAEA-408. Results from a world-wide intercalibration exercise. J Environ Monit 2:524–528

Villeneuve J, de Mora S, Cattini C (2002) Global intercomparison for the determination of chlorinated pesticides, PCBs and petroleum hydrocarbons in sediment sample IAEA-417. Environ Technol 23:1203–1217

Wadhia K, Thompson KC (2007) Low-cost ecotoxicity testing of environmental samples using microbiotests for potential implementation of the Water Framework Directive. Trends Anal Chem 26:300–307

Wenning R, Batley G, Ingersoll C et al. (2005) Use of sediment quality guidelines & related tools for the assessment of contaminated sediments. Society of Environmental Toxicology and Chemistry (SETAC), Pensacola (FL)

Wyse EJ, Coquery M, Azemard S, de Mora SJ (2004) Characterisation of trace elements and methylmercury in an estuarine sediment reference material, IAEA-405. J Environ Monit 6:48–57