A Comparison of Acute and Chronic Toxicity Methods for Marine Sediments

Alan J. Kennedy, Jeffery A. Steevens, Guilherme R. Lotufo, John D. Farrar,
Mark R. Reiss, Roy K. Kropp, Jon Doi, Todd S. Bridges

► To cite this version:

Alan J. Kennedy, Jeffery A. Steevens, Guilherme R. Lotufo, John D. Farrar, Mark R. Reiss, et al..
A Comparison of Acute and Chronic Toxicity Methods for Marine Sediments. *Marine Environmental
Research*, 2009, 68 (3), pp.118. 10.1016/j.marenvres.2009.04.010 . hal-00563079

HAL Id: hal-00563079

<https://hal.science/hal-00563079>

Submitted on 4 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

A Comparison of Acute and Chronic Toxicity Methods for Marine Sediments

Alan J. Kennedy, Jeffery A. Steevens, Guilherme R. Lotufo, John D. Farrar,
Mark R. Reiss, Roy K. Kropp, Jon Doi, Todd S. Bridges

PII: S0141-1136(09)00044-0
DOI: [10.1016/j.marenvres.2009.04.010](https://doi.org/10.1016/j.marenvres.2009.04.010)
Reference: MERE 3337

To appear in: *Marine Environmental Research*

Received Date: 24 October 2008
Revised Date: 16 April 2009
Accepted Date: 25 April 2009

Please cite this article as: Kennedy, A.J., Steevens, J.A., Lotufo, G.R., Farrar, J.D., Reiss, M.R., Kropp, R.K., Doi, J., Bridges, T.S., A Comparison of Acute and Chronic Toxicity Methods for Marine Sediments, *Marine Environmental Research* (2009), doi: [10.1016/j.marenvres.2009.04.010](https://doi.org/10.1016/j.marenvres.2009.04.010)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Comparison of Acute and Chronic Toxicity Methods for Marine Sediments

Alan J. Kennedy^{a,*}, Jeffery A. Steevens^a, Guilherme R. Lotufo^a, John D. Farrar^a, Mark R. Reiss^b,
Roy K. Kropp^c, Jon Doi^d, Todd S. Bridges^a

^aU.S. Army Engineer Research and Development Center, Environmental Laboratory, 3909 Halls
Ferry Road, Vicksburg, Mississippi 38180

^bU.S. Environmental Protection Agency, Region 2, 290 Broadway, New York, New York, 10007

^cMarine Sciences Laboratory, Battelle Pacific Northwest Division, Sequim, WA, 98382

^dAqua Survey Inc., 469 Point Breeze Rd., Flemington, NJ, 08822

* To whom correspondence may be addressed (Alan.J.Kennedy@usace.army.mil)

Alan J. Kennedy

U.S. Army Engineer Research and Development Center

Environmental Laboratory, CEERD-EP-R

3909 Halls Ferry Rd

Vicksburg, MS 39180-6199

Phone: 601-634-3344

Fax: 601-634-2263

Email: Alan.J.Kennedy@usace.army.mil

Abstract. Sediment toxicity tests are valuable tools for assessing the potential effects of contaminated sediments in dredged material evaluations because they inherently address complexity (e.g., unknown contaminants, mixtures, bioavailability). Although there is a need to understand the chronic and sublethal impacts of contaminants, it is common to conduct only short-term lethality tests in evaluations of marine sediments. Chronic toxicity methods for marine sediments have been developed but the efficacy of these methods is less documented. In this evaluation of marine sediments collected from the New York / New Jersey (NY/NJ) Harbor, three 10-day acute toxicity test methods (*Ampelisca abdita*, *Leptocheirus plumulosus*, *Americamysis bahia*) and three chronic and sublethal test methods (28-d *L. plumulosus*, 20- and 28-d *Neanthes arenaceodentata*) were applied by three testing laboratories. Although the *N. arenaceodentata* and *A. bahia* tests did not indicate significant toxicity for the sediments tested in this study, these methods have been reported useful in evaluating other sediments. The 10-d *A. abdita*, 10-d *L. plumulosus* and 28-d *L. plumulosus* tests were comparable between laboratories, indicating 29 – 43%, 29%, and 43 – 71% of the tested sediments as potentially toxic. The 28-d *L. plumulosus* method was the only chronic toxicity test that responded to the test sediments in this study. The 28-d *L. plumulosus* endpoint magnitudes were related to sediment chemistry and the sublethal endpoints were reduced as much or more than acute lethality endpoints. However, intra-treatment sublethal endpoint variability was greater, compromising detection of statistical significance. In this study, the chronic *L. plumulosus* test method was less consistent among laboratories relative to acute test methods, identifying potential for toxicity in a similar number (or slightly more) NY/ NJ Harbor sediments.

Keywords – Sediment toxicity, bioassay, chronic, sublethal, amphipod, polychaete

1. Introduction

More than 55 million cubic yards of sediment are dredged from harbors and channels annually in the United States and placed in the ocean. Of this material, it is estimated approximately 2.5 million cubic yards require special management because of chemical contamination (U.S. ACE, 2008). Sediment can be placed in the ocean if it is determined that the material does not "...unreasonably degrade or endanger: human health, welfare, or amenities, marine environment, ecological systems, or economic potentialities" (Marine Protection Research and Sanctuaries Act of 1972). A risk-based chemical and biological evaluation must be conducted to determine suitability for ocean placement of dredged sediments (U.S. EPA/ U.S. ACE, 1991). This assessment includes an evaluation of direct toxicity of sediments to benthic aquatic invertebrates.

Whole-sediment toxicity tests are commonly used in contaminated site assessments and dredged material evaluations. Several organizations (U.S. Environmental Protection Agency (U.S. EPA), American Society for Testing Materials (ASTM), Environment Canada (EC) and the International Standards Organization (ISO)) have published test methods to specifically evaluate the toxicity potential of freshwater (e.g., U.S. EPA, 2000; ASTM, 2005) or marine (e.g., U.S. EPA, 1994; ASTM, 2000; ASTM, 2003) sediments. Toxicity tests can be conducted for acute or chronic exposure durations. Acute tests typically assess mortality over short-term exposures, although some freshwater (U.S. EPA, 2000; ASTM, 2005) and marine and estuarine (Moreira et al., 2005) sediment tests also assess sublethal endpoints. Chronic tests represent a larger portion of the organism's life cycle and directly address sublethal endpoints (e.g., growth, reproduction) in addition to survival.

It is common for evaluations of marine sediments to employ only acute benthic toxicity test methods. Exceptions include use of a 20-day (-d) chronic polychaete test method (Johns et al., 1990) in the Dredged Material Management Program (DMMP) and consideration of several chronic methods in southern California (Bay et al., 2007). Given that acute tests for marine sediments may not be predictive of population-level effects, there is support for the use of chronic test methods (Munns et al., 2002; Wenning et al., 2005). These longer exposure tests that directly measure sublethal endpoints may provide information amenable for population modeling (McGee and Spencer, 2001; Ingersoll et al., 2005). Additionally, Finkelstein and Kern (2005) presented evidence to suggest that acute tests may result in more false negatives (i.e., failure to identify toxic sediments) compared to chronic methods when interpreted relative to sediment chemistry data expressed as effects range medium quotients. Chronic, or sublethal, sediment toxicity methods for estuarine and marine sediments have been published in guidance documents (e.g., ASTM, 2000; Johns et al., 1990; U.S. EPA/ACE, 2001) and in peer-reviewed journals (e.g., McGee and Spencer, 2001; Heuvel-Greve et al., 2007). Although chronic test endpoint sensitivity for freshwater sediment method has been studied widely (Ingersoll et al., 2001), the effectiveness of chronic test methods for marine sediments is less known.

Other comparisons of marine and estuarine sediment toxicity test performance have been conducted between single (Mearns et al., 1986; Bay et al., 2003) and multiple (Schlekat et al., 1995; Casado-Martinez et al., 2006) acute test methods, single chronic test methods (Johns et al., 1990; U.S. EPA/ACE, 2001) and acute and chronic test methods that use the same (McGee et al., 2004; Kuhn et al., 2002; Castro et al., 2006; Gale et al., 2006) or different (Anderson et al., 1998; Pinza et al., 2002; Greenstein et al., 2008) test organisms. The overall goal of this study was to conduct an interlaboratory comparison of a suite of acute and chronic test methods relevant to

evaluate New York / New Jersey (NY/NJ) Harbor (USA) sediments that are applicable to a dredged material management program that involves annual dredging of more than five million cubic yards of sediment for navigation and construction projects. The U.S. EPA requires 10-d acute *Ampelisca abdita* and *Americamysis bahia* bioassays as one line of evidence for the North Atlantic region to determine the suitability of sediments for ocean placement. We conducted tests and compared results between these currently applied acute sediment toxicity tests to an alternative acute test method (10-d *Leptocheirus plumulosus*) and available chronic toxicity test methods (28-d *L. plumulosus* and 20-d, 28-d *N. arenaceodentata*) to determine relative utility to this region-specific dredged material management program. Specific objectives were to investigate (1) the relative test method performance, (2) the consistency of results among three laboratories, and (3) whether the chronic tests offered an enhanced level of sensitivity to contaminants in the tested sediments.

2. Materials and Methods

2.1. Study design

This study involved preliminary and interlaboratory evaluations of sediments collected from the New York / New Jersey (NY / NJ) Harbor (Figure 1). Sediment aliquots were analyzed for contaminants on separate occasions because of a six-month storage period that exceeded the recommended eight-week holding time for sediments. Traditional holding time recommendations were not met because of the need for a preliminary assessment of the sediment toxicity prior to the interlaboratory comparison. However, holding time requirements were not directly applicable since the objective of this study was to compare the relative responses of different test methods to sediments rather than to assess the potential for site toxicity. The preliminary evaluation (December 2003 to February 2004) was conducted at only Lab A to

evaluate sediments from all nine sample sites. The toxicity tests conducted were the acute (10-d) *A. abdita*, *A. bahia*, and *L. plumulosus* and the chronic (28-d) *L. plumulosus* and 28-d *N. arenaceodentata* test methods. Additionally, control sediment (full description provided in Section 2.2) was used to dilute the Newark (100, 77, 50, 35, 24% dry weight) and Hudson (100, 82, 59, 43, 31% dry w/w) sediments using the 10-d *L. plumulosus* test to assess toxicity.

The interlaboratory evaluation (August – September 2004) involved three labs experienced in sediment toxicity testing (see author affiliations). Four distinct sediments of varying contamination (Hudson, Chester, Red Hook, and Newark) and a Newark dilution series (75%, 50%, 25% dry w/w) were selected based on preliminary results from the nine collected sediments. Methods used were the acute *A. abdita* test, chronic *L. plumulosus*, and three different *N. arenaceodentata* test methods summarized in Table 1. The 10-d *A. bahia* test was excluded because of low overall response in the preliminary evaluation. An acute *L. plumulosus* test was initiated at one laboratory (Lab A) on the same day as the chronic *L. plumulosus* tests. Each test method was initiated on the same day at the three facilities using organisms from the same sources.

[Table 1 position]

2.2. Test Sediments

Nine sites (Figure 1) in the NY/NJ Harbor were selected based on varying historic levels of contamination. Sediment was collected (27 – 29 September 2003) using a 10-L Van Veen grab and stored in 20-L HDPE buckets according to guidance (U.S. EPA, 2001). Buckets were shipped overnight (4 ± 2 °C) to the U.S. Army Engineer Research and Development Center (ERDC, Vicksburg, MS, USA). A reference sediment located south of New York Harbor (40° 20.210' N, 73° 52.190' W), designated for testing by the U.S. EPA Region 2 dredging program,

was acquired by Aqua Survey, Inc (Flemington, NJ, USA). A control sediment was collected from Sequim Bay, WA (Battelle, Sequim, WA, USA, 48° 06.52' N, 123° 01.410' W). All sediments were thoroughly homogenized to consistent texture with an impeller mixer before use. [Figure 1 position]

2.3. Sediment chemistry

Chemical analyses followed U.S. EPA SW-846 methodology (unless otherwise specified). The homogenized sediments were analyzed for pesticides, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), total metals, simultaneously extracted metals (SEM), and acid volatile sulfides (AVS) by the ERDC Environmental Chemistry Branch. Samples were submitted to Severn Trent Laboratories (Knoxville, TN) to analyze dioxins, furans, alkyl tins (GC-MS; method 8270), and sediment grain size (ASTM D422). To determine ammonia concentrations, sediment porewater was extracted by 15-minute centrifugation of sediment at 2,700 g and analyzed using a 720A ion selective electrode meter, equipped with a 95-12 ammonia-sensitive electrode (Thermo Orion Electron Corp.).

2.4. Sediment toxicity tests

Three acute toxicity tests and three chronic toxicity test methods were conducted (Table 1). The *A. abdita* and *L. plumulosus* tests were conducted according to standard guidance (U.S. EPA, 1994) while the *A. bahia* sediment test followed Appendix E of the Inland Testing Manual (U.S. EPA/ACE, 1998), since no U.S. EPA or ASTM sediment testing methods were available. The chronic *L. plumulosus* survival, growth, and reproduction test followed U.S. EPA / U.S. ACE (2001) while the three chronic *N. arenaceodentata* followed various methods (ASTM, 2000; Johns et al., 1990; Bridges et al., 1997). *Ampelisca abdita* was field-collected from the Blackwater River at Sea Brook, NH, USA (Aquatic Research Organisms, Hampton, NH) for the

preliminary comparison or from Atlantic Highlands, NJ, USA (Aqua Survey, Inc.) for the interlaboratory comparison. The remaining three organisms were laboratory-reared (*A. bahia*; obtained from Applied Biosystems (Fort Collins, CO, USA); *L. plumulosus* obtained from ERDC in-house cultures; *N. arenaceodentata* were obtained from Don Reish (California State University-Long Beach, CA, USA)). Since the concern of this study was endpoint response related to persistent contaminants rather than ammonia, sediments were purged for 48 h according to U.S. EPA (U.S. EPA, 1994) prior to test organism addition to a more conservative level of 20 mg/L for interstitial water (Ferretti et al., 2000). Ammonia concentrations were determined in separate test chambers not used in tests. Criteria governing test acceptability included a method-specific minimum control survival value (Table 1) and water quality within prescribed limits (U.S. EPA, 1994; Johns et al., 1990; U.S. EPA/ACE, 2001). The overlying water quality (temperature, pH, salinity, dissolved oxygen) was measured daily.

2.5. Statistical analysis

Tests were conducted in a randomized and blind fashion. Data normality (Kolmogorov-Smirnov test), homogeneity (Levene's Test), and treatment differences compared to the control sediment (one way ANOVA and Dunnett's Method, two way ANOVA and Tukey's test) were determined at the $\alpha = 0.05$ level using SigmaStat software (SPSS, Chicago, IL). Survival data were arcsine-square-root transformed and sublethal endpoints were square-root transformed when necessary for normality. When normality could not be achieved, the Kruskal-Wallis one-way ANOVA on ranks was applied. Since the study objective was to compare relative test method performance, not to characterize toxicity at a site, the control sediment was used in statistical comparisons rather than the site reference sediment (comprised only of 2% fines), which did not have representative characteristics of the predominantly fine test sediments. In

addition to statistical significance compared to the control, a toxicological decision criterion (TDC), defined as a reduction in survival that was at least 10% (20% for amphipod lethality) less than that of the control (U.S. EPA/ACE, 1998), was applied. For survival, both statistical significance and the TDC were required to classify sediments as potentially toxic, while only a statistically detectable decrease relative to the control was needed for sublethal endpoints. The lethal median concentration producing 50% mortality (LC50) in sediment dilutions was determined by the Spearman-Kärber method using Toxstat[®] software (Gulley 1996, University of Wyoming).

To assess test endpoint performance, four metrics were calculated: (1) the response magnitude (RM), (2) the minimum significant difference (MSD), (3) Kendall's concordance test, and (4) Spearman correlations. The RM estimated the relative amount of response between test endpoints due to sediment exposure by taking the grand mean of inter-treatment reductions compared to the control (modified from Gray et al., 1998).

$$RM = \frac{\frac{1}{k} \sum_{i=1}^k |x_i - x_c|}{x_c} * 100 \quad (1)$$

Where, k = number of treatments, x_i = treatment mean, x_c = control mean

The MSD measured the smallest endpoint reduction needed in a sediment treatment for a statistically significant difference relative to the control (Chapman et al., 1995).

$$MSD = d * s_w \sqrt{\frac{1}{n_o} + \frac{1}{n_c}} \quad (2)$$

d = critical value for Dunnett's procedure, s_w = square root of within mean square, n_o = treatment replicates, n_c = control replicates

Calculations of the RM and MSD were expressed as a percentage of the control mean to facilitate comparison between lethal and sublethal endpoints. Treatment endpoint values greater than control values were adjusted to the control value. Kendall's concordance test (W) was used

to assess interlaboratory agreement between ranks of endpoints to examine the consistency of tests (U.S. EPA/ACE, 2001; Mearns et al., 1986; Bay et al., 2003; Schlegel et al., 1995; Zar, 1984).

3. Results

3.1. Sediment chemistry and characterization

A summary of physical and chemical characteristics is provided in Table 2. Most sediments consisted of greater than 90% fines, with some exceptions. Reductions in PAH concentrations during storage were observed in the sediments for the interlaboratory evaluation. A more detailed summary of analytical chemistry data and correlation analysis is reported in Steevens et al. (2008).

[Table 2 position]

3.2. Sediment toxicity tests

Water quality parameters were within test method specified ranges, with the exception of one overnight temperature deviation (17.9 °C) in one chronic *L. plumulosus* test in the interlaboratory evaluation (Lab A) that was quickly rectified and did not appear to impact organism condition. Total porewater ammonia concentrations (following purging) at test initiation and termination were below recommended thresholds (< 20 mg/L) in all tests. In general, test endpoints (e.g. survival, reproduction) in the reference sediment (comprised of 98% sand) were reduced relative to the control sediment in the acute and chronic *L. plumulosus* tests but were variable in the *A. abdita* tests and similar in the *N. arenaceodentata* tests relative to the control (Table 3, Figure 2).

3.3. Preliminary evaluation

Toxicity test results are presented in Table 3. Mean survival in the control met acceptability criteria in all tests, with the exception of the *A. abdita* test (87%). In the *A. bahia* test, mean survival was relatively high ($> 75\%$), with no statistically significant differences among sediment treatments. The 28-d *N. arenaceodentata* test also resulted in no significant survival (range: 60 – 100%) or growth (Table 3) reduction effects for any sediment. Significant reductions in test endpoints were observed for the acute (*A. abdita*, *L. plumulosus*) and chronic (*L. plumulosus*) test methods using amphipods.

In a separate 10-d *L. plumulosus* test (data not presented), a concentration-dependent relationship was observed for dilutions series of the Hudson and Newark sediments, with 10-d LC50 values of 32 (22 – 46, 95% CIs) % and 48 (43 – 54, 95% CIs) %, respectively. This testing was used to select a sediment for dilution in the interlaboratory comparison. Also, survival in the undiluted Hudson ($12 \pm 12\%$) and undiluted Newark ($20 \pm 11\%$) treatments were similar to those obtained from previous testing (compare to Table 3).

[Table 3 position]

3.4. Interlaboratory evaluation

The test sediments were stored for 11 months (at 4 °C) after collection, 6 months following the preliminary evaluation, prior to use in the interlaboratory evaluation. Reductions in PAH concentrations (Table 2) and increases in test organism survival were observed (Table 3 and Figure 2). Interestingly, slight but consistent increases in PCBs, dioxins and DDTs were observed in the sediments following storage. This did not influence the use of the sediments for relative comparison between test sensitivity, which was the objective of the study. Two tests were excluded because of failures to meet control acceptability criteria; 10-d *A. abdita* control

survival was less than 90% (Lab A) and 28-d *L. plumulosus* failed to meet the control survival (< 80%) and reproduction acceptability criteria (Lab C).

For the two acceptable *A. abdita* tests, statistically significant survival reductions greater than the TDC were recorded (Figure 2). Two-way ANOVA determined no significant differences between laboratories after allowing for variability that was due to sediment effects, although individual significant differences in survival were determined between laboratories for the Reference and Chester sediments.

In contrast to the preliminary evaluation, no statistically significant reductions in survival were observed among sediments for the two acceptable 28-d *L. plumulosus* tests (Figure 2). However, there were statistically significant differences in survival between the two laboratories for the Reference, Red Hook and 50% Newark sediments. Significant reductions in either biomass or reproduction were observed for five and three (out of seven) sediments. After factoring out variability between the two laboratories by using two-way ANOVA, significant reductions in the biomass and reproduction endpoints were observed in both tests for the Reference, Hudson and 100% Newark sediments; reproduction was significantly reduced in 75% Newark sediment.

As in the preliminary study, survival was high in all chronic *N. arenaceodentata* tests, with no statistically significant endpoint reductions. Significant increases in biomass occurred in some sediments (Figure 3). Overall biomass was statistically significantly greater (four-fold) in the 20-d DMMP protocol compared to the other two methods that used lower feeding rations. Biomass was generally determined to be significantly greater in the Lab B tests by two-way ANOVA. Mean variability was comparable for the 28-d (coefficient of variation (CV) = $30 \pm$

16%) and the 20-d ($CV = 30 \pm 18\%$) test methods and slightly lower for the modified 20-d test method ($CV = 24 \pm 12\%$).

The RM, a measure of sensitivity, and MSD, a measure of statistical power, values are graphically summarized for each test endpoint in Figure 4. Test endpoints generally showed greater response in the preliminary evaluation relative to the interlaboratory evaluation. Among test methods, the lowest responses were for interlaboratory 10-d *A. bahia*, 28-d *L. plumulosus* survival (Figure 4), and *N. arenaceodentata* endpoints. The highest response was observed in 28-d *L. plumulosus* reproduction. However, higher statistical power (i.e., smaller MSD) was observed for survival endpoints (acute and chronic) than for sublethal endpoints. Among sensitive sediment toxicity tests, the response to test sediments was similar for individual test methods with good overall agreement between ranks of endpoint magnitudes. Kendall's concordance test (W) determined high coefficients of agreement for *A. abdita* survival ($W = 0.85$, $p < 0.25$), *L. plumulosus* survival ($W = 0.80$, $p < 0.25$), *L. plumulosus* biomass ($W = 0.77$, $p < 0.25$), and *L. plumulosus* reproduction ($W = 0.95$, $p < 0.10$). Lower significance was because of test failures (n value reduced from three to two tests).

[Figure 2 position]

[Figure 3 position]

[Figure 4 position]

4. Discussion

The overall objective of this study was to compare the performance of acute whole sediment toxicity tests commonly used in evaluations of NY/NJ harbor sediment to alternative chronic methods discussed in regulations but not commonly applied in practice. Varying endpoint response to the sediments evaluated was observed among test methods, ranging from

low response (acute *Americamysis bahia*, chronic *Neanthes arenaceodentata*) to moderate response (acute *Leptocheirus plumulosus* and *Ampelisca abdita* survival) to high response (chronic *L. plumulosus* reproduction) (Figure 4). Relative differences in endpoint response has been previously reported among test methods, testing facilities, evaluations, and sediments with differing predominant contaminants (Bay et al., 2007; Mearns et al., 1986; Bay et al., 2003; Kuhn et al., 2002). Generally, no one species or test method is universally most sensitive across sediments and conditions, supporting guidance recommending the use of multiple species to assess sediment toxicity.

Endpoint response to sediments is not necessarily indicative of contaminant effects. Several factors may complicate the interpretation of sediment tests, including uncertainty (Vorhees et al., 2002) and confounding factors (e.g., grain size, ammonia, sulfides, indigenous animals) (Postma et al., 2002). Grain size, for instance, is a well-known confounding factor in sediment tests (U.S. EPA, 1994; U.S.EPA/ACE, 2001) that can cause amphipod mortality by physical or energetic stress when outside species-specific tolerance ranges (Emery et al., 1997). With the exception of the reference sediment (98% coarse grains), the grain-size distributions of sediments tested in this study were similar (Table 2) and unlikely to produce confounding mortality. The different grain size of the reference sediment was not representative of the predominantly fine-grained test sediments and may have caused the reduced *L. plumulosus* endpoints observed in the reference sediment (Table 3, Figure 2). Porewater ammonia is another stressor in NY / NJ Harbor sediment evaluations that can confound tests if not purged as described in guidance. In this study porewater ammonia was purged to a more conservative value of 20 mg/L to remove it as a confounding factor to focus more on persistent contaminants of concern.

While the acute *Americamysis bahia* test method provided reasonable statistical power (Figure 4), the survival endpoint of this test was not as sensitive to the tested NY/NJ Harbor sediments as the other acute methods examined. This sediment test method is applied to evaluate NY/NJ Harbor sediments to provide a taxon other than amphipods, and it has responded to contaminated sediments in other studies (e.g., Ho et al., 2000). The sensitivity and exposure of *A. bahia* has been best demonstrated in the water column (U.S. EPA, 2002) and the Inland Testing Manual (U.S.EPA/ACE, 1998) provides the only national guidance for sediment evaluations using this species. This test method supplies an epibenthic exposure to sediments, since *A. bahia* dwells predominantly in the water-sediment interface, and includes a daily feeding ration (unlike the other acute test methods). The above factors should be considered when the intention is to be protective of infauna that ingest sediments containing hydrophobic contaminants. This method may be useful for test sediments with physical characteristics that are not suitable to infaunal life (e.g., clays).

The chronic *N. arenaceodentata* test endpoints were not responsive to the evaluated NY / NJ Harbor sediments. However, others (Anderson et al., 1998; Bridges et al., 1997; Green et al., 1999) have reported reductions in *N. arenaceodentata* endpoints, more often expressed for growth than survival, in exposures to contaminated sediments. The 20-d method is documented in a routine testing program (DMMP). In previous evaluations of test sediments from other regions, 28-d *N. arenaceodentata* growth responded similarly to 10-d *L. plumulosus* survival (Farrar, unpublished data). The 20-d *N. arenaceodentata* test method was reported to outperform the chronic (28-d) *L. plumulosus* test method (Pinza et al., 2002) and has responded to more (Pinza et al., 2002) and fewer (Anderson et al., 1998) sediments relative to the 10-d acute amphipod *R. abronius* method. The sensitivity of the *N. arenaceodentata* tests may vary

remarkably among major chemical classes and may be less responsive to PAHs because of the species' efficient metabolism (Christensen et al., 2002). Although the *N. arenaceodentata* endpoints were not responsive in this study, differences were observed among methods. The 20-d method involved a larger feeding ration and produced larger worms than the 28-d and modified 20-d tests (Figure 3). Bridges et al. (1997) demonstrated that relatively large amounts of supplemental food can reduce the toxicity of test sediments. Also, the 20-d methods use five worms per test chamber (compared to one per chamber in the 28-d method), which may inherently produce greater growth variability because of dominant worm interactions within replicates (Anderson et al., 1998; Bridges et al., 1997).

Only the test methods employing amphipods resulted in statistically reduced endpoint responses in exposure to the tested NY / NJ Harbor sediments, and the 28-d *L. plumulosus* method was the only chronic test in this sediment-specific evaluation to result in statistically significant reductions in test endpoints (Table 3, Figure 2). Mean percent survival was generally higher in the interlaboratory evaluation, possibly resulting from contaminant loss (e.g., PAHs, Table 2) after extended storage (Norton et al., 1999). Amphipod test methods suggested toxicity in at least one endpoint relative to the control for the most highly contaminated sediments (i.e., Hudson, 100% Newark). The presented data set conveys no clear advantage between the acute amphipod test methods in terms of statistical significance (Figure 4). However, logistically *L. plumulosus* offered the following advantages; (1) stronger correlations between survival and sediment chemistry (Steevens et al., 2008), (2) existing laboratory cultures making them available throughout the year; and (3) burrowing activity into sediments without tube-building (U.S. EPA, 1994; Bay et al., 2003; Ho et al., 2000) that provides a clear vector for sediment exposure, and more expedient processing at test termination.

Test endpoint performance was comparable among laboratories for the responsive test methods (i.e., acute *A. abdita*, chronic *L. plumulosus*) based upon good agreement (Kendall's W) between ranks of endpoint magnitudes. Overall, "testing laboratory" was not a significant factor on the 10-d *A. abdita* data but did contribute to significant interlaboratory differences in the 28-d *L. plumulosus* test method (also indicated in Figure 4). Thus, the chronic test method may have been more susceptible to differences among laboratory conditions.

The sublethal reproduction endpoint in the chronic *L. plumulosus* test method offered similar or enhanced response to the tested sediments relative to lethality (Figure 2, Figure 4). However, this magnitude of response is not meaningful in the context of management decisions that apply sediment test data based on presence or absence of statistically significant endpoint reductions relative to the reference (U.S. EPA, 1994; U.S. EPA/ACE, 1998). In this regard, the associated statistical power (i.e., MSD) of the *L. plumulosus* sublethal endpoints in this study was lower than survival endpoints, diminishing the ability of statistical tests to detect differences. The acute *A. abdita* and *L. plumulosus* survival endpoints performed more consistently than the sublethal *L. plumulosus* endpoints (Figure 4) because of high intra-treatment sublethal endpoint variability. Some studies have found enhanced sensitivity in chronic sediment test methods relative to acute methods assessing estuarine and marine sediments using amphipods (Castro et al., 2006), a copepod (Bejarano et al., 2004) and a polychaete (Moreira et al., 2005; Rice et al., 1995). However, the results from this study and from previous studies (McGee et al., 2004; Pinza et al., 2002; Greenstein et al., 2008) indicate that the specific chronic sediment toxicity methods we evaluated performed less consistently relative to acute methods.

In addition to confounding factors relevant to all tests (described above), multiple factors specific to chronic sediment toxicity test methods for marine sediments may contribute to

differences in performance (McGee et al., 2004; Kuhn et al., 2002). The response of different endpoints (e.g., survival vs. growth) may reflect differing modes of action (Kuhn et al., 2002), inherent sublethal endpoint variability (McGee et al., 2004; Pinza et al., 2002), or different procedures applied in the test methods (Table 1). Although intra-treatment variation may be reduced through increased replication or expression of neonates produced as a function of number of sexually mature females, the increased labor of such approaches may not be cost effective (Gray et al., 1998). Also, the addition of supplemental food to the sediment matrix in chronic toxicity test methods for marine or estuarine sediments may reduce chemical bioavailability (McGee et al., 2004) relative to organisms in acute tests that receive no supplemental food (Lotufo et al., 2001). Such differences do not exist between acute and chronic test methods for freshwater sediments, as both employ feeding rations and water exchanges (U.S. EPA, 2000; McGee and Spencer, 2001; Bridges et al., 1997).

5. Conclusion

The results of this study indicated performance differences for several marine sediment toxicity test methods. While the 10-d *A. abdita* method performed comparably to the 10-d *L. plumulosus*, the other acute method employed in these evaluations (10-d *A. bahia*) did not respond to the sediments in this study. Of the chronic test methods investigated for use, only the 28-d *L. plumulosus* method responded to the tested sediments. Advantages and disadvantages of the chronic *L. plumulosus* method were highlighted for NY / NJ Harbor dredged material evaluations. The potential benefits of the 28-d *L. plumulosus* test include theoretical improvements in level of protection through direct measure of sublethal effects, large sublethal endpoint response, strong relationships with sediment contamination, use of endpoints needed for population modeling, and fewer false negatives (Finkelstein and Kern, 2005). The

drawbacks of the 28-d *L. plumulosus* test may include, difficulties related to sublethal endpoint variability, longer turn-around time to obtain results, difficulty meeting acceptability criteria (more can go wrong in longer tests), and increased labor intensiveness leading to increased cost. Chronic estuarine and marine sediment toxicity test methods generally cost 1.7 (28-d *N. arenaceodentata*) to 2.3 (28-d *L. plumulosus*) times more than acute tests (based on per treatment costs reported by Bay et al., (2007)). The ability of the 28-d *L. plumulosus* chronic toxicity test method to indicate greater sensitivity to contaminated sediments relative to the 10-d test method was not consistent (McGee et al., 2004; this study). Overall, evidence exists for lower response (DeWitt et al., 1996; Fuchsman et al., 1998; Green et al., 1999; McGee et al., 2004) and enhanced response (McGee et al., 1993; Sferra et al., 1999) of *L. plumulosus* sublethal endpoints. Although the results of this study provided good indication of which test methods (i.e., acute and chronic amphipods tests) have the strongest application to NY / NJ sediments and associated contaminants of concern, more rigorous testing is needed before attempting to generalize to other regions. Test method selection should be contingent on management goals (e.g., desired level of protection, required use of indigenous taxa, desire for estimation of population level effects) and organism suitability to site-specific sediment composition

Acknowledgements. We thank U.S. Army Corps of Engineers New York District and the U.S. Environmental Protection Agency for sponsoring this research. We thank Monte Greges, Oksana Yaremko (U.S. ACE), Walter Berry (U.S. EPA), and Ken Finkelstein (NOAA) and three anonymous reviewers for enhancing technical clarity. Mark Graves (U.S. ACE) provided GIS mapping assistance. Analytical chemistry was completed by the analytical chemistry branch (Doug Taggart, Richard Karn, Anthony Bednar). We acknowledge support for the sediment

collection effort by Mr. Tom Wyche (CENAN). Deb Oestreicher and Jen Chappell provided technical editing. Permission has been granted by the Chief of Engineers to publish this material. The work presented in this paper has been internally reviewed by U.S. EPA, its publication however does not signify that the contents reflect the views of the Agency.

REFERENCES

- Anderson, B.S., Hunt, J.W., Phillips, B.M., Tudor, S., Fairey, R., Newman, J., Puckett, H.M., Stephenson, M., Long, E.R., Tjeerdema, R.S., 1998. Comparison of marine sediment toxicity test protocols for the amphipod *Rhepoxynius abronius* and the polychaete worm *Nereis (Neanthes) arenaceodentata*. *Environ Toxicol Chem* 17, 859–866.
- American Society for Testing and Materials (ASTM), 2000. Standard guide for conducting sediment toxicity tests with polychaetous annelids. Method E1611-00, West Conshohocken, PA.
- American Society for Testing and Materials (ASTM), 2003. Standard test method for measuring the toxicity of sediment-associated contaminants with estuarine and marine invertebrates. Method E1367-03, West Conshohocken, PA.
- American Society for Testing and Materials (ASTM), 2005. Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. Method E1706-05, West Conshohocken, PA.
- Bay, S.M., Jirik, A., Asato, S., 2003. Interlaboratory variability of amphipod sediment toxicity tests in a Cooperative Regional Monitoring Program. *Environ Monit Assess* 81:257-268.
- Bay, S., Greenstein, D., Young, D., 2007. Evaluation of methods for measuring sediment toxicity in California Bays and Estuaries. Southern California Coastal Water Research Project. Technical Report 503.
- ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/503_toxicity_indicator_methods.pdf

- Bejarano, A.C., Maruya, K.A., Chandler, G.T., 2004. Toxicity assessment of sediments associated with various land-uses in coastal South Carolina, USA, using a meiobenthic copepod bioassay. *Mar Pollut Bull* 49:23-32. Ref ID:8112.
- Bridges, T.S., Farrar, J.D., Duke, B.M., 1997. The influence of food ration on sediment toxicity in *Neanthes arenaceodentata*. *Environ Toxicol Chem* 16, 1659-1665.
- Casado-Martinez, M.C., Beiras, R., Belzunce, M.J., Gonzalez-Castromil, M.A., Marin-Guirao, L., Postma, J.F., Riba, I., DelValls, T.A., 2006. Interlaboratory assessment of marine bioassays to evaluate the environmental quality of coastal sediments in Spain. IV. Whole sediment toxicity test using crustacean amphipods. *Ciencias Marinas* 32,149-157.
- Castro, H., Ramalheira, F., Quintino, V., Rodrigues, A.M., 2006. Amphipod acute and chronic sediment toxicity assessment in estuarine environmental monitoring: An example from Ria de Aveiro, NW Portugal. *Mar Pollut Bull* 53,91-99.
- Chapman, G.A., Denton, D.L., Lazorchak, J.M., 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast estuarine and marine organisms. EPA-600-R-95-136. U.S. Environmental Protection Agency, Cincinnati, OH.
- Christensen, M., Anderso, O., Banta, G.T., 2002. Metabolism of pyrene by the polychaetes *Nereis diversicolor* and *Arenicola marina*. *Aquat Toxicol* 58, 15-25.
- DeWitt, T.H., Swartz, R.C., Hansen, D.J., McGovern, D., Berry, W.J., 1996. Bioavailability and chronic toxicity of cadmium in sediment to the estuarine amphipod *Leptocheirus plumulosus*. *Environ Toxicol Chem* 15,2095-2101.
- Emery, V.L., Moore, D.W., Gray, B.R., Duke, B.M., Gibson, A.B., Wright, R.B., Farrar, J.D., 1997. Development of a chronic sublethal sediment bioassay using the estuarine amphipod *Leptocheirus plumulosus* (Shoemaker). *Environ Toxicol Chem* 16, 1912-1920.

- 493
494 Ferretti, J.A., Calesso, D.F., Hermon, T.R., 2000. Evaluation of methods to remove ammonia
495 interference in marine sediment toxicity tests. *Environ Toxicol Chem* 19, 1935-1941.
496
497 Finkelstein, K., Kern, J., 2005. Improvement in Correlation Between Chemistry and Toxicity
498 using the 28-day Sediment Toxicity Test. Abstract A1-09 in: Olfenbittel RF, White PJ (eds),
499 Remediation of Contaminated Sediments – 2005: Finding Achievable Risk Reduction Solutions.
500 Proceedings of the Third International Conference on Remediation of Contaminated Sediments
501 (New Orleans, Louisiana; Jan 24-27, 2005). ISBN 1-57477-150-7, published by Battelle Press,
502 Columbus, OH, www.battelle.org/bookstore.
503
504 Fuchsman, P.C., Barber, T.R., Sheehan, P.J., 1998. Sediment toxicity evaluation for
505 hexachlorobenzene- spiked sediment tests with *Leptocheirus plumulosus*, *Hyaella azteca*, and
506 *Chironomus tentans*. *Arch Environ Contam Toxicol* 35, 573-579.
507
508 Gale, S.A., King, C.K., Hyne, R.V., 2006. Chronic sublethal sediment toxicity testing using the
509 estuarine amphipod, *Melita plumulosa* (Zeidler): Evaluation using metal-spiked and field-
510 contaminated sediments. *Environ Toxicol Chem* 25, 1887-1898.
511
512 Gray, B.R., Emery, V.L., Brandon, D.L., Wright, R.B., Duke, B.M., Farrar, J.D., Moore, D.W.,
513 1998. Selection of optimal measures of growth and reproduction for the sublethal *Leptocheirus*
514 *plumulosus* sediment bioassay. *Environ Toxicol Chem* 17, 2288-2297.
515
516 Green, A.S., Moore, D.W., Farrar, J.D., 1999. Chronic toxicity of 2,4,6-trinitrotoluene to a
517 marine polychaete and an estuarine amphipod. *Environ Toxicol Chem* 18, 1783-1790.
518
519 Greenstein, D., Bay, S., Anderson, B., Chandler, G.T., Farrar, J.D., Keppler, C., Phillips, B.,
520 Ringwood, A., Young, D. 2008 Comparison of methods for evaluating acute and chronic
521 toxicity in marine sediments. *Environ Toxicol Chem*, 27: 933 – 944.
522
523 Heuvel-Greve, M., Postina, J., Jol, J., Kooman, H., Dubbeldam, M., Schipper, C., Kater, B.,

2007. A chronic bioassay with the estuarine amphipod *Corophium volutator*: Test method description and confounding factors. *Chemosphere* 66,1301–1309.
- Ho, K.T., Kuhn, A., Pelletier, M., Mcgee, F., Burgess, R.M., Serbst, J., 2000. Sediment toxicity assessment: Comparison of standard and new testing designs. *Arch Environm Contamin Toxicol* 39, 462 – 468.
- Ingersoll, C.G., MacDonald, D.D., Wang, N., Crane, J.L., Field, L.J., Haverland, P.S., Kemble, N.E., Lindskoog, R.A., Severn, C., Smorong, D.E., 2001. Predictions of sediment toxicity using consensus-based freshwater sediment quality guidelines. *Arch Environ Contam Toxicol* 41,8–21.
- Ingersoll, C.G., Wang, N., Hayward, J.M.R., Jones, J.R., Jones, S.B., Ireland, D.S., 2005. A field assessment of long-term laboratory sediment toxicity tests with the amphipod *Hyaella azteca*. *Environ Toxicol Chem* 24, 2853–2870.
- Johns, D.M., Ginn, T.C., Reish, D.J., 1990. Protocol for juvenile *Neanthes arenaceodentata* sediment bioassay. EPA 68-D8-0085. U.S. Environmental Protection Agency, Seattle, WA.
- Kuhn, A., Munns, W.R., Serbst, J., Edwards, P., Cantwell, M.G., Gleason, T., Pelletier, M.C., Berry, W., 2002. Evaluating the ecological significance of laboratory response data to predict population-level effects for the estuarine amphipod *Ampelisca abdita*. *Environ Toxicol Chem* 21:865-874.
- Lotufo, G.R., Landrum, P.F., Gedeon, M.L., 2001. Toxicity and bioaccumulation of DDT in freshwater amphipods in exposures to spiked sediments. *Environ Toxicol Chem* 20: 810-825.
- McGee, B.L., Fisher, D.J., Yonkos, L.T., Ziegler, G.P., Turley, S., 1999. Assessment of Sediment Contamination, Acute Toxicity, and Population Viability of the Estuarine Amphipod *Leptocheirus plumulosus* in Baltimore Harbor, Maryland, USA. *Environ Toxicol Chem* 18, 2151-2160.

- McGee, B.L., Schlekat, C.E., Reinhartz, E., 1993. Assessing sublethal levels of sediment contamination using the estuarine amphipod *Leptocheirus plumulosus*. Environ Toxicol Chem 12, 577-587.
- McGee, B.L., Spencer, M., 2001. A field-based population model for the sediment toxicity test organism *Leptocheirus plumulosus*: II. Model application. Mar Environ Res 51:347-363.
- McGee, B.L., Fisher, D.J., Wright, D.A., Yonkos, L.T., Ziegler, G.P., Turley, S.D., Farrar, J.D., Moore, D.W., Bridges, T.S., 2004. A field test and comparison of acute and chronic sediment toxicity with the marine amphipod *Leptocheirus plumulosus* in Chesapeake Bay, USA. Environ Toxicol Chem 23,1751-1761.
- Mearns, A.J., Swartz, R.C., Cummins, J.M., Dinnel, P.A., Plesha, P., Chapman, P.M., 1986. Inter-Laboratory Comparison of a Sediment Toxicity Test Using the Marine Amphipod, *Rhepoxynius abronius*. Mar Environ Res 19:13-37.
- Moreira, S.M., Moreira-Santos, M., Guilhermino, L., Ribeiro, R., 2005. Short-term sublethal *in situ* toxicity assay with *Hediste diversicolor* (Polychaeta) for estuarine sediments based on postexposure feeding. Environ Toxicol Chem 24, 2010-2018.
- Munns, W.R., Berry, W.J., Dewitt, T.H., 2002. Toxicity testing, risk assessment, and options for dredged material management. Mar Pollut Bull 44, 294-302.
- Norton, B.L., Lewis, M.A., Mayer, F.L., 1999. Storage duration and temperature and the acute toxicities of estuarine sediments to *Mysidopsis bahia* and *Leptocheirus plumulosus*. Bull Environ Contam Toxicol 63,157-166.
- Pinza, M.R., Ward, J.A., Kohn, N.P., 2002. Results of interspecies toxicity comparison testing associated with contaminated sediment management, Sediment Management Annual

Review Meeting, Seattle, WA.

www.nws.usace.army.mil/publicmenu/DOCUMENTS/dmmo/SMARM_2002_minutes.pdf

Postma, J.F., de Valk, S., Dubbeldam, M., Maas, J.L., Tonkes, M., Schipper, C.A., Kater, B.J., 2002. Confounding factors in bioassays with freshwater and marine organisms. *Ecotoxicol Environ Safety* 53, 226-237.

Rice, C.A., Plesha, P.D., Casillas, E., Misitano, D.A., Meador, J.P., 1995. Growth and survival of three marine invertebrate species in sediments from the Hudson-Raritan estuary, New York. *Environ Toxicol Chem* 14, 1931-1940.

Schlekat, C.E., Scott, K.J., Swartz, R.C., Albrecht, B., Antrim, L., Doe, K., Douglas, S., Ferretti, J.A., Hansen, D.J., Moore, D.W., Mueller, C., Tang, A., 1995. A Interlaboratory comparison of a 10-day sediment toxicity test method using *Ampelisca abdita*, *Eohaustorius estuarius* and *Leptocheirus plumulosus*. *Environ Toxicol Chem* 14, 2163-2174.

Sferra, J.C., Fuchsman, P.C., Wenning, R.J., Barber, T.R., 1999. A Site-Specific Evaluation of Mercury Toxicity in Sediment. *Arch of Environ Contam Toxicol* 37,488-495.

Steevens, J., Kennedy, A., Farrar, D., McNemar, C., Reiss, M. R., Kropp, R. K., Doi, J., and Bridges, T., 2008. "Dredged Material Analysis Tools; Performance of Acute and Chronic Sediment Toxicity Methods," ERDC/EL TR-08-16, U.S. Army Engineer Research and Development Center, Vicksburg, MS.

U.S. Army Corps of Engineers (U.S. ACE). 2008. Overview of Corps National Dredging Program and Regulations. From Dredged Material Management Training. Sacramento, CA. April 14. <http://el.erdc.usace.army.mil/dots/training/dmams.cfm>

U.S. Environmental Protection Agency, (U.S. EPA), 1994. Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods. U.S. Environmental Protection Agency, EPA/600/R-94/025, Washington, D.C.

- 616
617 U.S. Environmental Protection Agency, (U.S. EPA). 2000. Methods for measuring the toxicity
618 and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. U.S.
619 Environmental Protection Agency, EPA/600/R-99/064, Washington, D.C.
620
621 U.S. Environmental Protection Agency, (U.S. EPA). 2001. Methods for collection, storage, and
622 manipulation of sediments for chemical and toxicological analyses: Technical Manual. U.S.
623 Environmental Protection Agency, EPA-823-B-01-002, Washington, D.C., 20460.
624
625 U.S. Environmental Protection Agency, (U.S. EPA) 2002. Methods for Measuring the Acute
626 Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms.
627 EPA/812/R/02/012, Office of Water, Washington, D.C., 20460.
628
629 U.S. Environmental Protection Agency, (U.S. EPA)/Army Corps of Engineers, (ACE). 1991.
630 Evaluation of Dredged Material Proposed for Ocean Disposal (Ocean Testing Manual). EPA
631 503/8-91/001, U.S. Environmental Protection Agency, Washington DC.
632
633 U.S. Environmental Protection Agency, (U.S. EPA)/Army Corps of Engineers, (ACE). 1998.
634 Evaluation of material proposed for discharge to waters of the U.S. - Testing Manual (Inland
635 Testing Manual). EPA/823/B-98/004, U.S. Environmental Protection Agency, Washington DC.
636
637 U.S. Environmental Protection Agency, Army Corps of Engineers, (U.S. EPA/ U.S. ACE), 2001.
638 Methods for Assessing the Chronic Toxicity of Marine and Estuarine Sediment-Associated
639 Contaminants with the Amphipod *Leptocheirus plumulosus*. U.S. Environmental Protection
640 Agency, EPA/600/R-01/020, Washington, D.C.
641
642 Vorhees, D.J., Kane Driscoll, S.B., von Stackelberg, Bridges, T.S., 2002. An evaluation of
643 sources of uncertainty in a dredged material assessment. Hum Ecol Risk Assess 8, 369-389.
644
645 Wenning, R.J., Batley, G.E., Ingersoll, C.G., Moore, D.W., 2005. Use of sediment quality

- 646 guidelines & related tools for the assessment of contaminated sediments. Society of
647 Environmental Toxicology and Chemistry Press. Pensacola, FL, USA. 816 pp.
648
649 Zar, J.H. 1984. *Biostatistical Analysis*. 2nd Edition. Prentice-Hall, Inc., Englewood Cliffs, NJ.
650 718 pp.

ACCEPTED MANUSCRIPT

Figure legends

Figure 1. Sampling stations in the New York Harbor.

Figure 2. Test results for the interlaboratory comparison. Survival results for the 10-d *Ampelisca abdita*, 10-d *Leptocheirus plumulosus* and 28-d *L. plumulosus* test methods are shown in panel (a). Biomass for the 28-d *L. plumulosus* test method is shown in panel (b). Reproduction for the 28-d *L. plumulosus* for the test method is shown in panel (c). Histogram columns indicate mean values and bars represent one standard deviation from the mean. Asterisks represent a statistically significant reduction compared to the control and an endpoint response greater than the toxicological decision criterion (i.e., > 20% reduction relative to the control).

Figure 3. *Neanthes arenaceodentata* toxicity test results for the 28-d, 20-d DMMP and 20-d DMMP with modified feeding ration test methods. Histogram columns indicate mean values and bars represent one standard deviation from the mean. Number signs represent significant increases compared to the control.

Figure 4. Plot of the Response Magnitude (RM) relative to the control versus the Minimum Significant Difference (MSD) needed for statistical comparisons for the toxicity test endpoints evaluated in this study. Data from the preliminary data (open points) and interlaboratory (closed points) are indicated. Dotted vertical lines around points indicate the range in endpoint data while the diagonal dashed line indicates a one-to-one relationship.

674

677

ACCEPTED

Table 1. Summary of whole sediment toxicity test methods.

Test Method	Temperature (°C)	Water Renewal	Replicates (Chamber Size)	Organisms per Replicate	Feeding Regime	Endpoint(s)	Control Acceptability Criteria
10-day <i>Americamysis bahia</i>	20 ± 2	None	n = 5 (1 L)	20	<i>Artemia nauplii</i> daily	Survival	≥ 90% survival
10-day <i>Ampelisca abdita</i>	20 ± 2	None	n = 5 (1 L)	20	None	Survival	≥ 90% survival
10-day <i>Leptocheirus plumulosus</i>	25 ± 2	None	n = 5 (1 L)	20	None	Survival	≥ 90% survival
28-day <i>Leptocheirus plumulosus</i>	25 ± 2	Three times per week	n = 5 (1 L)	20	Three times a week (M, W, F): Days 0-14: 20 mg Tetramin [®] /beaker Days 15-28: 40 mg Tetramin [®] /beaker	Survival Growth Reproduction	Survival ≥ 80%, measurable growth Neonates in all replicates
28-day <i>Neanthes arenaceodentata</i>	20 ± 2	Once a week	n = 10 (0.3 L)	1	Twice a week: Tuesdays: 2 mg Tetramarin [®] /beaker Fridays: 2 mg Tetramarin [®] 1 mg alfalfa/beaker	Survival Growth	≥ 80% Survival, measurable growth
20-day <i>Neanthes arenaceodentata</i>	20 ± 2	Once every three days	n = 5 (1 L)	5	Every other day: 40 mg Tetramarin [®] /beaker	Survival Growth	≥ 80% Survival, measurable growth

20-day <i>Neanthes</i> <i>arenaceodentata</i> (modified)	20 ± 2	Once every three days	n = 5 (1 L)	5	<p>Twice a week:</p> <p>Tuesdays: 10 mg Tetramarin[®]/beaker</p> <p>Fridays: 10 mg Tetramarin[®], 5 mg alfalfa/beaker</p>	Survival Growth	≥ 80% Survival, measurable growth
---	--------	--------------------------	----------------	---	--	--------------------	---

Table 2. Summary of physical characteristics and major classes of organic compounds detected in sediments ($\mu\text{g/kg}$) used in the preliminary evaluation (PE) and interlaboratory evaluation (IE). TOC = total organic carbon, PAH = polycyclic aromatic hydrocarbon,

Sediment	Particle Size (%)	TOC (%)	Bulk		Σ PAH $\dagger$ ($\mu\text{g/kg}$)	Σ PCBs ($\mu\text{g/kg}$)	Σ DDT ($\mu\text{g/kg}$)	2,3,7,8 TCDD ($\mu\text{g/kg}$)
			Porewater	Ammonia (mg/L) $\ddagger$				

PCB = polychlorinated biphenyl, DDT = dichlorodiphenyltrichloroethane, TCDD = tetrachlorodibenzodioxin.

	Fines	Sand	PE	IE	PE	IE	PE	IE	PE	IE	PE	IE
Control	71	29	0.6		2.5		< 600		< 30		< 3	
Reference	2	98	<0.04	n.a.	n.a.	n.a.	1.6	n.a.	<0.39	n.a.	< 1	n.a.
Arthur	99	1	4.5	n.a.	77.6	n.a.	7,811	n.a.	204	n.a.	555	n.a.
Buttermilk	92	8	3.6	n.a.	60.6	n.a.	10,405	n.a.	195	n.a.	15	n.a.
Chester	40	60	2.7	2.0	36.3	31.4	14,802	13,093	122	162	21	53
Flushing	97	3	5.2	n.a.	57.1	n.a.	14,173	n.a.	271	n.a.	40	n.a.
Hudson	93	7	3.4	2.9	47.3	41.4	7,382	5,090	675	720	79	147
Jamaica Bay	74	27	3.8	n.a.	62.4	n.a.	1,530	n.a.	51	n.a.	nd	n.a.
100% Newark	92	9	3.2	3.3	32.6	34.4	15,206	8,000	238	251	78	120
75% Newark	n.a.	n.a.	n.a.	2.6	n.a.	31.3	n.a.	6,869	n.a.	208	n.a.	95
50% Newark	n.a.	n.a.	n.a.	2.6	n.a.	30.1	n.a.	3,354	n.a.	151	n.a.	49
25% Newark	n.a.	n.a.	n.a.	2.5	n.a.	26.7	n.a.	1,825	n.a.	82	n.a.	32
Perth	93	8	3.5	n.a.	19.0	n.a.	2,984	n.a.	112	n.a.	31	n.a.
Red Hook	46	54	1.7	1.5	35.1	34.8	5,166	94	94	101	31	31

n.a. = not available

† 13 PAHs from (Swartz, 1999)

‡ All sediments purged to 20 mg/L total ammonia prior to the addition of test organisms.

Table 3. Toxicity test results from the preliminary evaluation. Mean endpoint responses ($\pm$ one standard deviation from the mean) are presented.

Sediment	10-d Acute Tests Mean Survival			28-d <i>Leptocheirus plumulosus</i>			28-d <i>Neanthes arenaceodentata</i>	
	<i>Americamysis bahia</i> (%)	<i>Ampelisca abdita</i> (%)	<i>Leptocheirus plumulosus</i> (%)	Mean Survival (%)	Biomass (mg)	Neonates/ Survivor	Mean Survival (%)	Biomass (mg)
Control	90 $\pm$ 9	87 $\pm$ 10 [†]	91 $\pm$ 2	96 $\pm$ 6	0.9 $\pm$ 0.5	2.4 $\pm$ 1.5	80 $\pm$ 42	1.2 $\pm$ 0.5
Reference	86 $\pm$ 7	42 $\pm$ 6*	72 $\pm$ 15	60 $\pm$ 15*	0.4 $\pm$ 0.0	0.2 $\pm$ 0.2*	80 $\pm$ 42	1.1 $\pm$ 0.5
Arthur	93 $\pm$ 8	58 $\pm$ 15	29 $\pm$ 12*	71 $\pm$ 17*	0.9 $\pm$ 0.1	0.6 $\pm$ 0.5	70 $\pm$ 48	0.8 $\pm$ 0.4
Buttermilk	78 $\pm$ 30	45 $\pm$ 18*	53 $\pm$ 12*	90 $\pm$ 10	1.2 $\pm$ 0.1	1.7 $\pm$ 1.9	80 $\pm$ 42	2.6 $\pm$ 0.9#
Chester	95 $\pm$ 5	55 $\pm$ 23*	40 $\pm$ 25*	65 $\pm$ 15*	0.5 $\pm$ 0.1	0.2 $\pm$ 0.2*	60 $\pm$ 52	2.7 $\pm$ 0.9#
Flushing	88 $\pm$ 3	42 $\pm$ 19*	37 $\pm$ 27*	79 $\pm$ 13	0.8 $\pm$ 0.2	0.1 $\pm$ 0.2*	90 $\pm$ 32	2.2 $\pm$ 0.8#
Hudson	90 $\pm$ 8	48 $\pm$ 10*	11 $\pm$ 4*	46 $\pm$ 13*	0.6 $\pm$ 0.2	1.9 $\pm$ 2.6	70 $\pm$ 48	2.1 $\pm$ 0.7
Jamaica Bay	94 $\pm$ 8	59 $\pm$ 28	79 $\pm$ 18	89 $\pm$ 14	1.3 $\pm$ 0.4	0.8 $\pm$ 0.4	70 $\pm$ 48	1.5 $\pm$ 0.9
Newark	82 $\pm$ 10	36 $\pm$ 11*	27 $\pm$ 14*	71 $\pm$ 11*	0.6 $\pm$ 0.3	0.2 $\pm$ 0.4*	70 $\pm$ 48	1.7 $\pm$ 0.5
Perth	75 $\pm$ 26	46 $\pm$ 23*	72 $\pm$ 11	84 $\pm$ 12	1.0 $\pm$ 0.3	1.1 $\pm$ 1.0	100 $\pm$ 0	1.5 $\pm$ 0.4
Red Hook	97 $\pm$ 3	72 $\pm$ 10	56 $\pm$ 15*	89 $\pm$ 10	1.0 $\pm$ 0.2	0.8 $\pm$ 0.8	100 $\pm$ 0	2.5 $\pm$ 0.7#

[†] Control survival did not meet acceptability criteria (≥ 90)

* Statistically significant reduction compared to the control and a response greater than the toxicological decision criterion (i.e., $> 20\%$ reduction relative to the control).

Statistically significant increase in response relative to the control.