ERDC/EL TR-08-16

US Army Corps
of Engineers®
Engineer Research and
Development Center

Dredging Operations Environmental Research Program

Dredged Material Analysis Tools

Performance of Acute and Chronic Sediment Toxicity Methods

Jeffery Steevens, Alan Kennedy, Daniel Farrar,
Cory McNemar, Mark R. Reiss, Roy K. Kropp, Jon Doi,
and Todd Bridges

July 2008 Revised

Environmental Laboratory

Dredging Operations Environmental
Research Program

ERDC/EL TR-08-16
July 2008 Revised

Dredged Material Analysis Tools

Performance of Acute and Chronic Sediment Toxicity Methods

Jeffery Steevens, Alan Kennedy, Daniel Farrar, and Todd Bridges

Environmental Laboratory
U.S. Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Cory McNemar

Specpro, Inc.
3532 Manor Drive
Vicksburg, MS 39180

Mark R. Reiss

U.S. Environmental Protection Agency, Region 2
290 Broadway
New York, NY 10007

Roy K. Kropp

Marine Sciences Laboratory
Battelle Pacific Northwest Division
Sequim, WA 98382

Jon Doi

Aqua Survey, Inc.
469 Point Breeze Road
Flemington, NJ 08822

Final report

Prepared for U.S. Army Corps of Engineers
Washington, DC 20314-1000

Abstract: This report and research were supported by the U.S. Army Corps of Engineers New York District and U.S. Environmental Protection Agency Region 2. The work was conducted to provide insight into the potential advantages and disadvantages of using chronic sediment toxicity tests with relevant benthic macroinvertebrates as part of dredged material evaluations, as described in the Inland and Ocean Testing Manuals (USEPA/USACE 1991, 1998). Nine sediments collected from the New York Harbor (NYH) were used to assess test methods in a preliminary evaluation at one test facility and an interlaboratory evaluation at three test facilities. The two acute test methods (10-day *Ampelisca abdita* and *Americamysis bahia*) currently used in evaluations of NYH material were compared to available chronic protocols to gauge relative performance of the toxicity tests. Acute tests are typically short-term (e.g., 10-day) lethality assessments conducted over a small portion of the test organism's life cycle, while chronic tests are longer-term and assess sublethal measurement endpoints (e.g., growth and reproduction) in addition to lethality. The available chronic test methods used in this study were the 28-day test using the estuarine amphipod, *Leptocheirus plumulosus*, and 20-day and 28-day tests using the marine polychaete *Neanthes arenaceodentata*. Use of chronic tests is recommended or required by dredged material evaluation guidance and regulations, respectively. The sublethal endpoints measured in chronic tests may be more sensitive measures of toxicity and more predictive of longer-term population effects. Of the tests compared, the currently used acute (10-day) *Ampelisca abdita* test and the available chronic (28-day) *L. plumulosus* test were the most responsive (i.e., sensitive) to the tested NYH sediments. Response is defined as the amount an endpoint (e.g., survival) was reduced for test organisms in site sediments relative to that same endpoint in the control sediment. Of these two test methods, neither clearly demonstrated better capability to identify contaminated sediments (i.e., "hits"). The *A. abdita* test was more consistent in performance and exhibited greater statistical power but demonstrated lesser response to the sediments and lower correlation with sediment chemistry. The sublethal endpoints used in the *L. plumulosus* test were more responsive to the sediments and more closely related to sediment contamination but had lower statistical power than lethality endpoints. An acute (10-day) test using *L. plumulosus* was also conducted in one laboratory and similar responsiveness was found relative to the acute *A. abdita* test. The remaining toxicity tests, including the currently applied acute *A. bahia* test and the 28-day *N. arenaceodentata test* were not responsive to the tested sediments in this evaluation and thus did not suggest toxicity in any of the tested sediments. Specific conclusions and recommendations on the application of these test methods are offered at the end of this document.

Contents

Figures and Tables

Figures

Tables

Preface

This report was prepared through a research effort funded, in part, by the U.S. Army Engineer District, New York, and the U.S. Environmental Protection Agency (USEPA). It describes the results of research focusing on the use of chronic toxicity test methods for assessing sediments proposed for dredging under the Clean Water Act and the Marine Protection, Research, and Sanctuaries Act.

The research effort was conducted and the report prepared by Dr. Jeffery A. Steevens, Dr. Todd S. Bridges, Alan J. Kennedy, and Daniel Farrar of the Environmental Laboratory (EL), U.S. Army Engineer Research and Development Center (ERDC); Cory McNemar of Specpro, Inc.; Mark R. Reiss of USEPA Region 2; Dr. Roy K. Kropp of Battelle Pacific Northwest Division; and Dr. Jon Doi of Aqua Survey, Inc. Technical advice on this effort was provided by Monte P. Greges and Oksana S. Yaremko, New York District, and Doug Pabst, Buddy Lobue, and M. R. Reiss, USEPA Region 2. Analytical chemistry was completed by Drs. Douglas B. Taggart, Richard Karn, and Anthony J. Bednar of the EL Environmental Chemistry Branch. Two contract laboratories, Aqua Survey, Inc. (Dr. Doi), and Battelle's Marine Sciences Laboratory (Dr. Kropp and Betsy Barrows), participated in the interlaboratory analysis. The authors acknowledge support provided by Thomas L. Wyche (New York District) with collection of the sediments.

This report was prepared under the U.S. Army Corps of Engineers' Dredging Operations and Environmental Research Program, for which Dr. Todd S. Bridges serves as Program Manager. Dr. Mike Passmore was Deputy Director, EL, and Dr. Beth Fleming was Director, EL.

COL Richard B. Jenkins was Commander and Executive Director of ERDC. Dr. James R. Houston was Director.

1 Introduction

Background

Dredged material management is regulated by Section 404 of the Clean Water Act, Section 10 of the Rivers and Harbors Act, and Section 103 of the Marine Protection, Research and Sanctuaries Act (MPRSA). According to Section 404 of the Clean Water Act, a discharge into an aquatic system must offer the least-adverse management alternative, satisfy legal standards, not result in significant environmental degradation, and apply all practical measures to reduce impacts. Section 103 of MPRSA states that "the transportation of dredged material for the purpose of dumping it into ocean waters [is allowable] where...the dumping will not unreasonably degrade or endanger human health, welfare, or amenities, or the marine environment, ecological system, or economic potentialities." Managers must characterize dredged material contamination and subsequently estimate risk of disposing of the sediments in open waterways. Similar multi-tiered approaches are recommended by the U.S. Environmental Protection Agency (USEPA) and U.S. Army Corps of Engineers (USACE) in both the *Inland Testing Manual*, or ITM (USEPA/USACE 1998), and the *Ocean Testing Manual*, or OTM (USEPA/USACE 1991). The OTM, which is most relevant to the testing discussed in this report, describes a screening-level tier, a sediment collection and chemical analysis tier to obtain further information and additional tiers (i.e., Tiers III and IV) if a management decision still cannot be made.

An important component of dredged material assessment is the application of sediment toxicity tests, Tiers III and IV (USEPA/USACE 1991, 1998), which use relevant benthic test organisms and apply measurement endpoints to gauge the extent of contamination in the material. These tests are a useful tool to assess potential risk to the disposal location and can help circumvent knowledge gaps for complex questions such as chemical mixtures/synergisms, unknown contaminants, and chemical bioavailability, where sediment-screening techniques may fall short. Several organizations (e.g., USEPA, ASTM International, Environmental Canada) have published protocols (e.g., ASTM 2002) that specifically evaluate toxicity in either freshwater (USEPA 2000; ASTM 2000a) or marine sediments (e.g., USEPA 1994; ASTM 1998; USEPA 2000; ASTM 2000b; ASTM 2000c; USEPA 2001; ASTM 2002; ASTM

2003). Such toxicity tests, however, present their own unique limitations, including consistency of results and interpretation of uncertainty for field extrapolations. Vorhees et al. (2002) provide a detailed analysis of uncertainty related to improving dredged material management decisions.

Toxicity assays can be conducted for an *acute* or *chronic* duration. Acute assays, a Tier III activity, typically assess test organism mortality over short-term exposure durations relative to the organism's life cycle. Chronic assays, a Tier IV activity, are carried out over a larger portion of the organism's life cycle (e.g., at least one-tenth; Simini et al. 2000) and usually measure sublethal effects on activities such as growth and reproduction. Currently, the dredged material management program in New York Harbor (NYH) routinely applies only acute (10-day) sediment toxicity tests using the marine mysid shrimp, *Americamysis bahia* (formerly *Mysidopsis bahia*), and the marine amphipod, *Ampelisca abdita*.

Application of chronic tests is addressed in the OTM (USEPA/USACE 1991) as a latter activity that is conducted if "Tier III test results lead to equivocal interpretation." In the context of CWA and MPRSA, some researchers support use of chronic tests (e.g., Munns et al. 2002). Wenning et al. (2005) state that acute tests may not be as predictive as chronic tests for gauging long-term population-level effects in a sediment quality guideline (SQG) context. For example, the longer duration of exposure used in chronic tests is more representative of an environmental exposure received by benthic organisms. Longer exposure allows concentrations of contaminants that accumulate at a relatively slower rate (e.g., high molecular weight organics) to reach steady state in test organism tissue[1]. Sublethal endpoints also may be more relatable to field-population responses given that growth and reproduction contribute to population dynamics, especially when contaminant exposures are at lower, subacute concentrations where the effects more subtle than lethality (as measured by acute tests) may occur. More importantly, Federal regulations require consideration of sublethal effects in assessments of dredged material targeted for ocean disposal. According to 40 CFR 227.27(b), "The limiting permissible concentration of the... solid phases of a material means concentration which will not cause unreasonable acute or chronic toxicity or other sublethal adverse effects based on test results...

[1] Standardized bioaccumulation tests are conducted 28-days for this reason.

using appropriate sensitive benthic organisms." Several chronic test methods have been developed and published (e.g., USEPA/USACE 2001; ASTM 2000b; ASTM 2003) that would satisfy this regulatory requirement.

Overall, dredged material assessments of marine sediments rely more extensively on acute rather than chronic sediment toxicity tests. Some management programs, such as the Dredged Material Management Program (DMMP), formerly the Puget Sound Dredged Disposal Analysis (PSDDA) Program, have administered standard use a chronic test (i.e., 20-day *Neanthes arenaceodentata*) with a sublethal endpoint (Johns et al. 1990). The DMMP is an interagency management program for the Puget Sound area overseen by the Washington Department of Natural Resources, USACE (Seattle District), and USEPA (Region 10). The performance of chronic sediment toxicity tests, however, has not been as widely verified as the acute lethality tests currently in use, which may explain the discrepancy in frequency of implementation. Perhaps more importantly, performing chronic toxicity tests is more time-consuming and labor-intensive than performing acute tests; the cost is almost twofold greater than acute tests, providing a significant reason for resistance.

Objective

This study was conducted to address the four following questions, which must be resolved before chronic protocols can be recommended for use as a routine component of a dredged material management process:

1. How does the performance of chronic tests compare to the currently utilized acute sediment toxicity tests?
2. Can different laboratories successfully perform chronic sediment toxicity tests with similar results?
3. Can the higher cost of conducting chronic toxicity tests be justified for dredged material management programs?
4. What is the utility of chronic tests, at their current stage of development, in dredged material management?

(The scope of the fourth question is too broad to be answered using the limited data in this report alone, and therefore warrants dedicated emphasis in future discussion at a national level with USEPA and USACE.)

Approach

To address the research questions, sediments containing varying levels of contamination were collected from NYH. The sediments, collected from locations in or near Federal project channels, were used to compare acute toxicity tests currently conducted in NYH dredged material evaluations with available chronic toxicity protocols. Discrete sediments, in addition to sediment dilutions, were initially tested in a preliminary evaluation at one laboratory and then in an interlaboratory evaluation at three laboratories. The inter- and intralaboratory variability, consistency, and responsiveness to contamination were assessed for each test method. Performance comparisons of the marine/estuarine sediment toxicity test methods were previously conducted between single (Mearns et al. 1985; Bay et al. 2003) or multiple (Schlekat et al. 1995) acute protocols, single chronic protocols (USEPA 2001; Johns et al. 1990) and for both acute and chronic protocols using the same test organism (Anderson et al. 1998; McGee et al. 2004; Farrar et al. 2005a). However, no such evaluation has been conducted using sediments and test protocols specific to dredged material evaluations in NYH.

2 Methods

Study site

Nine sampling stations were selected in NYH (Figure 1) based on varying historic contamination levels (i.e., low, moderate, and high). Thirty-five to 40 gallons of sediment were collected (Table 1) at each station from 27-29 September 2004 using a 2.5 gallon Van Veen Grab sampler (Figure 2), according to USEPA (2001) guidance. Sediments from each site were collected into seven or eight new HDPE 5 gallon buckets. At the end of each sampling day, sediments were stored in refrigerated cabinets (4 ± 2 °C) at the Caven Point Field Station (Jersey City, NJ). A reference sand designated for use by the EPA Region 2 dredging program located south of NYH was acquired from Aqua Survey Inc. (Flemington, NJ), and a control sediment was collected from Sequim Bay, WA (Battelle, Sequim, WA) for comparison to NYH site sediments. Sediments were shipped overnight to the Waterways Experiment Station (ERDC, Vicksburg, MS), by a refrigerated truck (4 ± 2 °C). Sediments from each site were thoroughly homogenized in a 55 gallon high-density polyethylene (HDPE) container using a large motorized mixer, model ND-1A (Lightnin, Rochester, New York), redistributed into buckets and stored in a walk-in cooler at 4 ± 2 °C.

Sediment chemistry

Each of the homogenized sediments was sampled for chemical analysis and submitted to the Environmental Chemistry Branch of the ERDC Environmental Laboratory (Vicksburg, MS) for analysis of pesticides, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), total metals, simultaneously extracted metals (SEM), and acid volatile sulfides (AVS). Samples were submitted to Severn Trent Laboratories (Knoxville, TN) for dioxins, furans, alkyl tins, and particle size analysis. All chemical analyses followed USEPA 846 methodology. Organic compounds were extracted by method 3545 using accelerated solvent extraction. Cleanup of organic extracts was accomplished for PAHs using a modification of method 3630 (silica gel), pesticides using a modification of method 3630 (florisil), and PCBs using a modification of method 3665 (sulfuric acid). PAH analysis was performed according to method 8270 by gas chromatograph using selective ion monitoring mode.

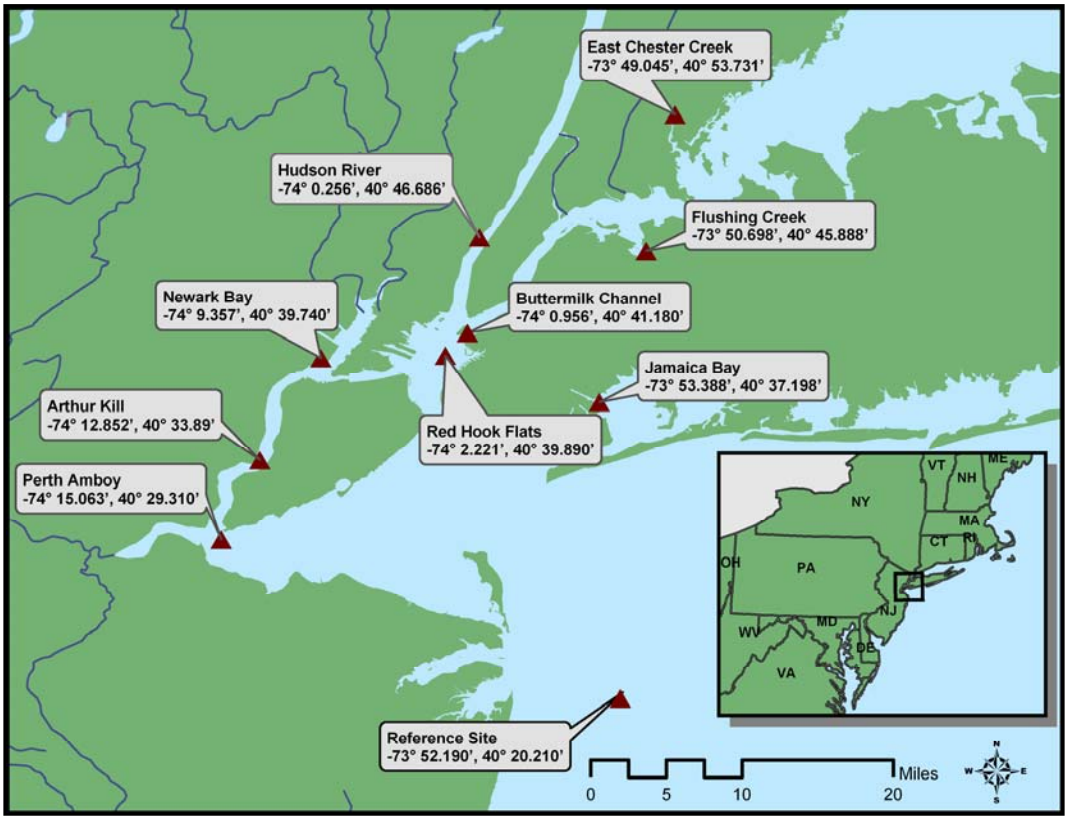

Figure 1. Sampling stations in New York Harbor.

Table 1. Sample station descriptions, date of collection and biological observations.

Sediment	Location	Depth (ft)	Collection Date	Benthic Organisms Observed
Arthur Kill	40° 33.89' N 74° 12.852' W	36	9/27/2003	None
Buttermilk Channel	40° 41.180' N 74° 0.956' W	20	9/26/2003	Crabs, Mussels, Oligochaetes
East Chester Creek	40° 53.731' N 73° 49.045' W	20 - 23	9/28/2003	Clams, Hermit crabs, Oligochaetes
Flushing Creek	40° 45.888' N 73° 50.698' W	18	9/28/2003	None
Hudson River	40° 46.686' N 74° 0.256' W	31	9/26/2003	None
Jamaica Bay	40° 37.198' N 73° 53.388' W	39	9/27/2003	Amphipods, Hardshell clams, Hermit crabs, Horseshoe crabs
Newark Bay	40° 39.740' N 74° 9.357' W	49	9/27/2003	Oligochaetes
Perth Amboy	40° 29.310' N 74° 15.063' W	26	9/27/2003	None
Red Hook Flats	40° 39.890' N 74° 2.221' W	35	9/26/2003	None

Figure 2. Van Veen grab used in collection of the sediments. Picture was taken at the Arthur Kill sampling location.

Pesticides and PCBs were analyzed using an Agilent 5890 gas chromatograph with electron capture detection. Metals analyses were performed following methods 6010B and 6020 using a PerkinElmer Optima 3000DV Inductively Coupled Plasma Atomic Emission Spectrometer and an Elan 6000 Inductively Coupled Plasma Mass Spectrometer, respectively. Total mercury analysis was performed using a P.S. Analytical Atomic Fluorescence Detector following method 7471A. Methylmercury was determined using ethylation followed by carbon-trapping and analysis by gas chromatography with fluorescence detection. Dioxins and furans were measured using method 8290 by high-resolution gas chromatograph/mass spectrophotometer. Grain size distribution of sediments was determined using ASTM method D422.

Sediment toxicity tests

The acute protocols included in this comparison are 10-d tests currently used in evaluations of NYH sediments with the marine amphipod, *Ampelisca abdita*, and marine mysid shrimp, *Americamysis bahia* (Figure 3). The methods followed were from USEPA (1994) for *A. abdita* tests and from Aqua Survey (2001) and ASTM (2002) for *A. bahia* tests. Field-collected *A. abdita* were purchased from Aquatic Research Organisms (Hampton, NH) for the preliminary evaluation and from Aqua

Survey, Inc. (Flemington, NJ) for the interlaboratory comparison. Laboratory-cultured *A. bahia* were obtained from Applied Biosystems (Fort Collins, CO). Briefly, acute tests were 10-day static non-renewal, with each treatment consisting of five replicate 1 L beakers containing 175 mL sediment, 775 mL overlying water, and 20 organisms (Table 2). The *A. bahia* received a daily feeding ration of brine shrimp (*Artemia*) while the *A. abdita* test involved no feeding. An acute (10-d) *Leptocheirus plumulosus* test (USEPA 1994) using organisms from ERDC in-house cultures was also conducted for comparative purposes (Appendix A1) but was not included in the interlaboratory comparison. The endpoint assessed for acute protocols was percent survival. More detail about the acute protocols is provided in Appendix A1.

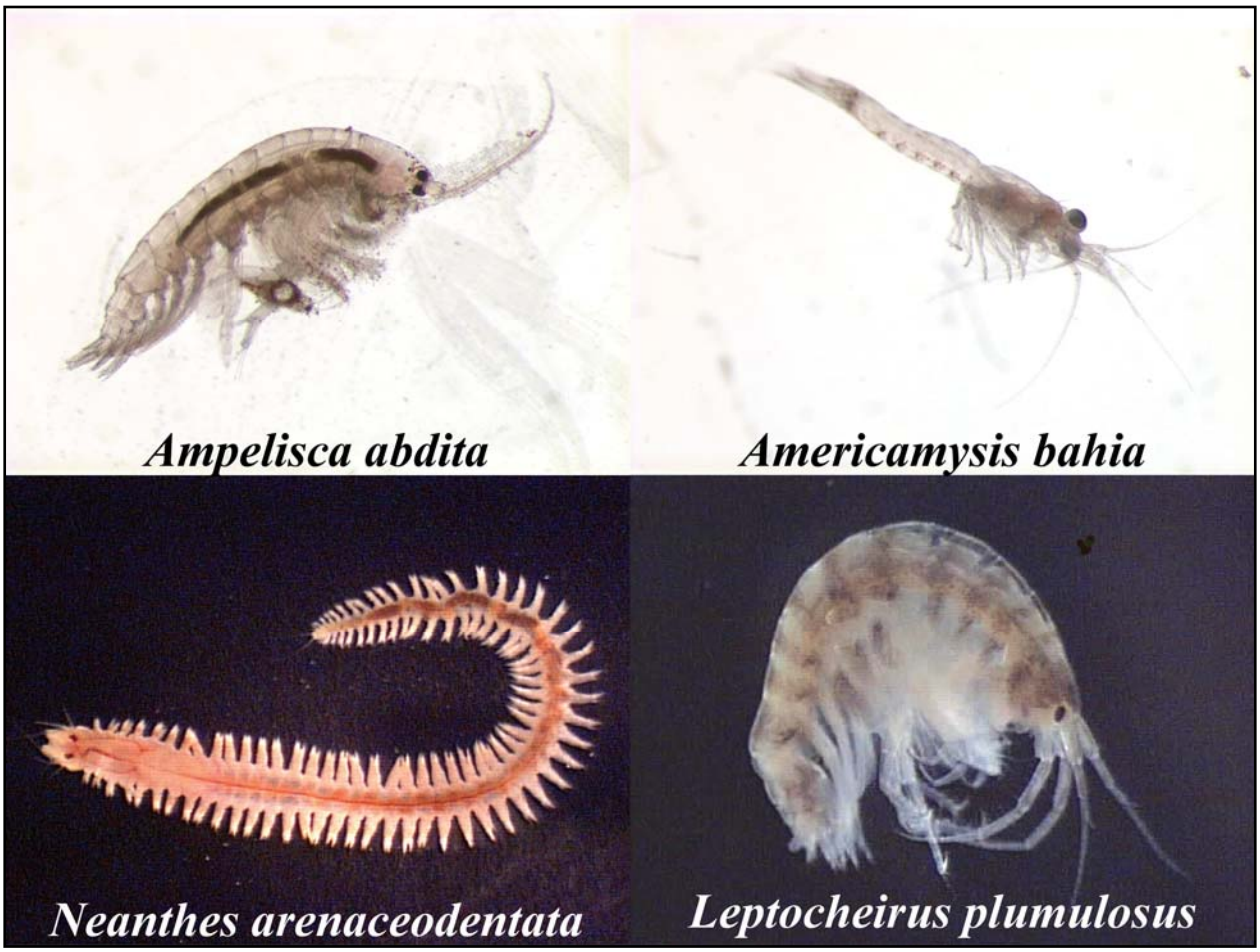

Figure 3. Benthic invertebrates used in sediment toxicity tests. New York Harbor evaluations currently apply acute tests (10-day) using *Ampelisca abdita* and *Americamysis bahia*. Available chronic protocols in this study include 20- or 28-day tests using *Leptocheirus plumulosus* and *Neanthes arenaceodentata.*

Table 2. Specifics for sediment toxicity protocols used in the interlaboratory comparison. More detailed information can be found in Appendix A.

Protocol Type	Test Organism	Duration (days)	Test Type	Test Vessels	Replicates	Organisms/ Replicate	Feeding Regime	Endpoints
Acute	*Americamysis*	10	Static Non-renewal	1 L	5	20	Daily	Survival
	Ampelisca	10	Static Non-renewal	1 L	5	20	None	Survival
Chronic	*Leptocheirus*	28	Static Renewal (3 times/week)	1 L	5	20	3 times/ week	Survival Growth Reproduction
	Neanthes (ERDC)	28	Static Renewal (1 time/week)	300 mL	10	1	2 times/ week	Survival Growth
	Neanthes (PSDDA)	20	Static Renewal (1 time/ 3 days)	1 L	5	5	1 time/ 2 days	Survival Growth
	Neanthes (PSDDA) modified	20	Static Renewal (1 time/ 3 days)	1 L	5	5	2 times/ week	Survival Growth

Chronic protocols used the estuarine amphipod, *L. plumulosus* (ERDC in-house cultures) and the marine polychaete, *Neanthes arenaceodentata* (Don Reish, California State University–Long Beach), shown in Figure 3. The 28-d *L. plumulosus* test treatments included five replicate 1 L beakers containing 175 mL sediment, 775 mL overlying water, and 20 amphipods, and was conducted according to standard guidance (USEPA 2001). The 28-d *N. arenaceodentata* test was conducted with modifications from published ASTM (2000b) guidance, using treatments of 10 replicate 300 mL tall-form beakers containing 75 mL sediment, 75 mL overlying water, and one worm (Bridges et al. 1997). In addition, a 20-d *N. arenaceodentata* test method used in the DMMP according to Johns et al. (1990) was evaluated, involving five replicate 1 L beakers per treatment, each containing 175 mL of sediment, 775 mL overlying water, and five worms. Finally, a modification of the 20-d PSDDA *N. arenaceodentata* test was conducted, reducing the 20-d feeding ration (8 mg Tetramarine® / worm / 2 days) to that of the 28-d *N. arenaceodentata* test (4 mg Tetramarine® and 1 mg alfalfa / worm / week). The four tests were static, renewal, involving regular water exchanges and feeding rations (Table 2). The endpoints assessed were survival and growth. Reproduction, standardized as number of neonates per surviving adult, as recommended by Gray et al (1998), was also included as an endpoint in the *L. plumulosus* study. More detail about the chronic protocols is provided in Appendix A2.

Criteria that must be met for test method acceptability included specified overlying water quality ranges (see Appendices A1, A2), ≥ 90 percent control survival for acute tests, 28-d *N. arenaceodentata* test and 20-d *N. arenaceodentata* tests, and ≥ 80 percent survival for 28-d *L. plumulosus* test. In addition, *L. plumulosus* neonates must be observed in each control/reference replicate. Test temperature and photoperiod were regulated using water baths equipped with timer-controlled lights and water recirculating REMCOR heating/cooling units (REMCOR Products Company, Glendale Heights, IL). Exposure chambers received trickle-flow aeration. The quality of the overlying water in the sediment toxicity tests was measured using a model ABMTC handheld refractometer (Aquafauna Bio-Marine, Hawthorne, California) for salinity, a model 315i meter (WTW; Weilheim, Germany) for pH, and a model Oxi 330 meter (WTW; Weilheim, Germany) for D.O. Porewater ammonia samples were taken by centrifuging 45 mL of sediment at 4000 rpm for 15 minutes using VWR Brand 50 mL centrifuge tubes (Cat. No. 21008-177). The porewater, or supernatant, was then decanted and analyzed for total ammonia concentrations using an 720A ion-selective electrode (ISE) meter (Thermo Orion Electron Corp., Beverly, MA) equipped with an 95-12 ammonia-sensitive electrode (Thermo Orion Electron Corp., Beverly, MA). Total porewater ammonia in the bulk NYH sediments ranged from 20-80 mg/L (Table 3), and concentrations in some sediments exceeded recommended thresholds for *A. abdita* (30 mg/L), *L. plumulosus* (60 mg/L; USEPA 1994), *A. bahia* (0.6 mg unionized NH_3/L, Battelle 2000) and *N. arenaceodentata* (20 mg/L; Dillon et al. 1993). Since ammonia is not typically a contaminant of concern in dredged material evaluations, sediments were purged before addition of organisms according to USEPA (1994) to ≤ 20 mg/L, a more conservative level recommended in Ferretti et al. (2000). Given that the objective of this study was to evaluate the protocols using the NYH sediments rather than the converse, all sediments were purged equally for simplicity to ensure that tests were initiated on the same days at each facility.

Table 3. Total ammonia concentrations (mg/L) in bulk sediments reported for the preliminary (a) and interlaboratory (b) evaluations.

Sediment	Total ammonia (mg/L)	Unionized Ammonia (mg/l) T = 20, pH = 8	Unionized Ammonia (mg/l) T = 25, pH = 8
(a)			
Control	2.5	0.10	0.13
Arthur	77.6	2.96	4.18
Buttermilk	60.6	2.23	3.26
Chester	36.3	1.39	1.95
Flushing	57.1	2.18	3.07
Hudson	47.3	1.18	2.55
Jamaica	62.4	2.38	3.36
Newark	32.6	1.25	1.76
Perth	19.0	0.73	1.02
Red Hook	35.1	1.34	1.89
(b)			
Control	24.2 ± 5.7	0.92 ± 0.22	1.30 ± 0.31
Hudson	41.4 ± 6.8	1.58 ± 0.26	2.23 ± 0.37
Chester	31.4 ± 6.4	1.20 ± 0.24	1.69 ± 0.34
Red Hook	34.8 ± 5.6	1.33 ± 0.21	1.87 ± 0.30
100% Newark	34.4 ± 6.7	1.31 ± 0.26	1.85 ± 0.36
75% Newark	31.3 ± 4.4	1.20 ± 0.17	1.69 ± 0.24
50% Newark	30.1 ± 6.2	1.15 ± 0.24	1.62 ± 0.33
25% Newark	26.7 ± 4.8	1.02 ± 0.18	1.44 ± 0.26

Estimations for the unionized fraction are provided for the indicated temperatures and pH. Ammonia concentrations for the interlaboratory evaluation are the means (± one standard deviation) of the measurements reported by the three participating facilities. The reference sand was not analyzed due to insufficient porewater volume. The total ammonia criterion at test initiation for *Ampelisca abdita* and *Leptocheirus plumulosus* is 30 and 60 mg/L, respectively. Test sediments were purged until total ammonia concentrations were less than 20 mg/L.

Study design

All sediments were homogenized with an impeller mixer within a bucket before use in toxicity tests. This study involved preliminary and inter-laboratory evaluations to assess the performance of the toxicity test methods. The sediments tested in both the preliminary and inter-laboratory evaluations were submitted on separate occasions for chemical analysis since the sediments were stored for 6 months between the two evaluations.

The preliminary evaluation was conducted only at ERDC from December 2003 to February 2004 using all nine field-collected sediments (Table 1). The purpose of this evaluation was to gain baseline information to select sediments for the interlaboratory evaluation. The testing protocols used were the 10-d acute *A. bahia* and *A. abdita* tests and the 28-d chronic *L. plumulosus* and *N. arenaceodentata* tests. A 10-d *L. plumulosus* test was conducted for comparison purposes. To identify the most appropriate sediment for use in a contamination gradient (i.e., dilution series) an additional 10-d *L. plumulosus* test was used to evaluate Hudson and Newark sediments; gradients were created on a dry weight basis by diluting the sediments with the control (i.e., Sequim Bay) sediment. The reference sediment was not used in the dilutions due to its unrepresentative coarse grain size, which would not allow for homogenous mixing.

The interlaboratory comparison, conducted from August to September 2004, involved three facilities experienced in sediment toxicity testing: Battelle's Marine Sciences Laboratory (Sequim, WA), Aqua Survey, Inc. (Flemington, NJ), and ERDC-EL (Vicksburg, MS). To provide anonymity, the laboratories are referred to as Lab A, B, and C (in no particular order) for the remainder of this report. Four distinct sediments (Hudson, Chester, Red Hook, and Newark) and a Newark gradient were selected based on preliminary results. Newark and Hudson served as highly contaminated sediments while Chester was moderately contaminated and Red Hook was low in contamination. The Newark sediment was diluted with the control sediment on a dry weight basis (75 percent, 50 percent, 25 percent). A 10 gallon aliquot of each sediment, including the three diluted Newark sediments, control, and reference sediment, was homogenized at ERDC and placed into 2 gallon buckets and shipped overnight to participating facilities in coolers with ice (4 ± 2°C). The protocols used were the 10-d *A. abdita* test, the 28-d *L. plumulosus,* and 28-d *N. arenaceodentata* tests. The 20-d DMMP *N. arenaceodentata* protocol and a modification of the DMMP protocol that used the 28-d *N. arenaceodentata* feeding ration were also used in the interlaboratory comparison. The 10-d *A. bahia* test was excluded because no organism mortality was observed in any of the sediments in the preliminary evaluation. A 10-d *L. plumulosus* test was conducted at one facility for comparative purposes (initiated on the same day as the 28-d *L. plumulosus* test). Each test for a particular protocol was initiated on the same day at the three facilities, and organisms tested were obtained from the same source.

Statistical analysis

All toxicity tests were conducted in a completely randomized and blind fashion. Data normality (Kolmogorov-Smirnov test), homogeneity (Levene's Test), and treatment differences compared with the control sediment (one-way ANOVA, Dunnett's Method) were determined at the $\alpha = 0.05$ level using SigmaStat statistical analysis software (SPSS, Chicago, IL). Survival data were arcsine-square root transformed while sublethal endpoints were square root transformed when necessary to satisfy normality assumptions; when normality could not be achieved, the Kruskal-Wallis test on ranks was applied. In addition to statistical significance compared with the control, a toxicological decision criterion (TDC) was assessed, which is defined as a reduction in survival that was at least 10 percent (20 percent for amphipod lethality) lower than that of the control (USEPA/USACE 1998). Thus, both statistical significance and ecological significance in these test methods were required to suggest toxicity in any particular sediment (i.e., a "hit"). For sublethal endpoints, only a statistically detectable decrease in growth or reproduction relative to control was required. The lethal median concentration producing 50 percent mortality (LC50) in sediment dilutions was determined by the Spearman-Karber method (Gulley 1996).

The response magnitude (RM) was determined to estimate endpoint response/sensitivity to sediments. The RM was essentially the grand mean of inter-treatment reductions compared with the control. It is important to point out that the RM is a measure of endpoint response to the sediment but not necessarily to contamination. The minimum significant difference (MSD), calculated as in Chapman et al. (1995), represented the smallest test endpoint reduction relative to the control in a sediment treatment where statistical significance was possible. Calculations of the RM[1] and MSD are provided below:

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

[1] Treatment endpoint values greater than control values were made equal to the control value (e.g., if survival for a site sediment was 97 percent and control survival was 95 percent, the site sediment was brought to 95 percent).

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

where k = number of treatments; x_c = control mean; x_i = treatment mean; d = critical value for Dunnett's procedure; and s_w = square root of within mean square.

The RM and MSD values were expressed as a percentage of the control mean and were modifications from a previous study (Gray et al. 1998). These are relative parameters and thus were only used to compare protocols evaluating the same sediments during the same evaluations. Low MSDs and large RMs (in contaminated sediments) are most desirable. Kendall's concordance test (W), determined as in Zar (1984), was used to assess interlaboratory agreement between ranks of endpoints to examine the consistency and repeatability of tests. Finally, Spearman correlations of ranks (SPSS, Chicago, IL) were used to assess strength of associations between test endpoint magnitude (i.e., response) and sediment contamination levels.

3 Results

Sediment chemistry and characterization

The objective of this study was not to determine the specific contaminants driving toxicity, although knowledge of dominant chemical classes in the sediments is useful provided that the associated impacts may vary between test organisms and endpoints. PCBs, PAHs, metals, dioxins, furans, and alkyl tins were detected at varying magnitudes in all sediments. Sediment quality guidelines, or SQGs, were used as a simple screen to assess the potential for sediment toxicity to benthos. In this study, SQG values were used to determine that a sediment is not likely to cause effects to benthos or identify a potential class of contaminants that may be responsible for toxicity. However, SQGs used in this context have several limitations. The SQG values do not address more than one chemical or its interactions, they do not address or quantify exposure, and are not site specific. They also have a high rate of false positives and false negatives, as documented by O'Connor et al. (1998) and Long et al. (1998).

Generally, SQGs in the NOAA SQuiRT Tables (1999), such as effects range median (ERM), were exceeded for PAHs, PCBs, ΣDDT, and metals in some sediments. Tributyl tins (TBT) did not exceed available PSDDA (73 μg/kg) screening values, and 2,3,7,8-TCDD did not exceed the available AET screening value of 0.0036 μg/kg (NOAA 1999). Total organic carbon (TOC) in the sediments ranged from 1.7 – 5.2 percent. Notable decreases in volatile LMW PAH concentrations were observed for sediments used in the interlaboratory comparison (Table 4) compared with the same sediments tested in the preliminary evaluation (Table 5). The Flushing, Chester, and Newark sediments had the highest PAH concentrations; Hudson, Flushing, Newark, Arthur, and Buttermilk had the highest PCB concentrations; Arthur and Newark had the highest concentrations of dioxins and furans; and Arthur had the highest pesticides (primarily ΣDDT). Comprehensive chemical analyses for both evaluations are provided in Appendix B. Although there were individual exceedances in metal SQGs, metals were not estimated to be a major contributor to toxicity based on equilibrium partitioning guidelines for metals SEM-AVS (i.e., SEM < AVS; USEPA 2002). Similarly, the ΣPAH model (Swartz 1999) did not indicate high probability of PAH-related toxicity in these sediments (using preliminary evaluation sediment chemistry). The model

(Table 6) summarized the probability of no toxic effect, in addition to more than 13 percent and 24 percent amphipod mortality. The probability of no toxicity to amphipods ranged from 60 – 100 percent, with indication of low levels of mortality for only the Chester and Newark sediments. Overall, the probability of PAH toxicity was less than 0.40 for all tested sediments. The site sediments were predominantly fines (> 90 percent), with the exception of Jamaica Bay (73 percent fines), Chester (42 percent fines) and Red Hook (46 percent fines). The reference sediment, required for comparative use in NYH dredged material evaluations, was only 2 percent fines (i.e., mostly sand). There was > 70 percent similarity between grain size for all sediments except for Chester, Red Hook, and the reference. Full particle size distributions for the test sediments are provided in Figure 4.

Table 4. Summary of major classes of organic compounds detected in sediments used for the preliminary evaluation.

Sediment	%TOC, µg/kg	∑ LPAH, µg/kg	∑ HPAHs, µg/kg	∑ PCBs, µg/kg	ΣDDT, µg/kg	Dioxins Furans, ng/kg
ER-M	NA	3160	9600	180	46	NA
Arthur	4.5	1186	7519	204	555	3807
Buttermilk	3.6	2250	9442	195	15	1763
Chester	2.7	2258	13374	122	21	1380
Flushing	5.2	2223	13820	271	40	2160
Hudson	3.4	1711	6391	675	79	2470
Jamaica	3.8	266	1423	51	0	881
Newark	3.2	4975	11006	238	78	3718
Perth	3.5	540	2817	112	31	1646
Red Hook	1.7	954	4958	94	31	478.5

The effects range medium (ER-M) is given to provide a frame of reference for potential biological effects. Underlined values are above the ER-M.
TOC = total organic carbon
LPAH/HPAH = low/high molecular weight polycyclic aromatic hydrocarbon
PCB = polychlorinated biphenyl
DDT = dichlorodiphenyltrichloroethane

Table 5. Summary of major classes of organic compounds detected in sediments used for the interlaboratory comparison.

Sediment	%TOC	∑ LPAH, µg/kg	∑ HPAHs, µg/kg	∑ PCBs, µg/kg	∑ DDT, µg/kg	Dioxins Furans, ng/kg
ER-M	NA	3160	9600	180	46	NA
Hudson	2.9	1260	4659	720	147	3668
Chester	2.0	2455	11243	162	53	411
Red Hook	1.5	1017	5034	101	31	519
100% Newark	3.3	2253	6218	251	120	6679
75% Newark	2.6	1722	5582	208	95	754
50% Newark	2.6	797	2762	151	49	418
25% Newark	2.5	394	1537	82	32	270

The table provides quick reference to relative sediment contamination and illustrates the dilution of contaminants for the Newark gradient. The effects range medium (ER-M) is given to provide a frame of reference for potential biological effects. Underlined values are above the ER-M.
TOC = total organic carbon
LPAH/HPAH = low/high molecular weight polycyclic aromatic hydrocarbon
PCB = polychlorinated biphenyl
DDT = dichlorodiphenyltrichloroethane

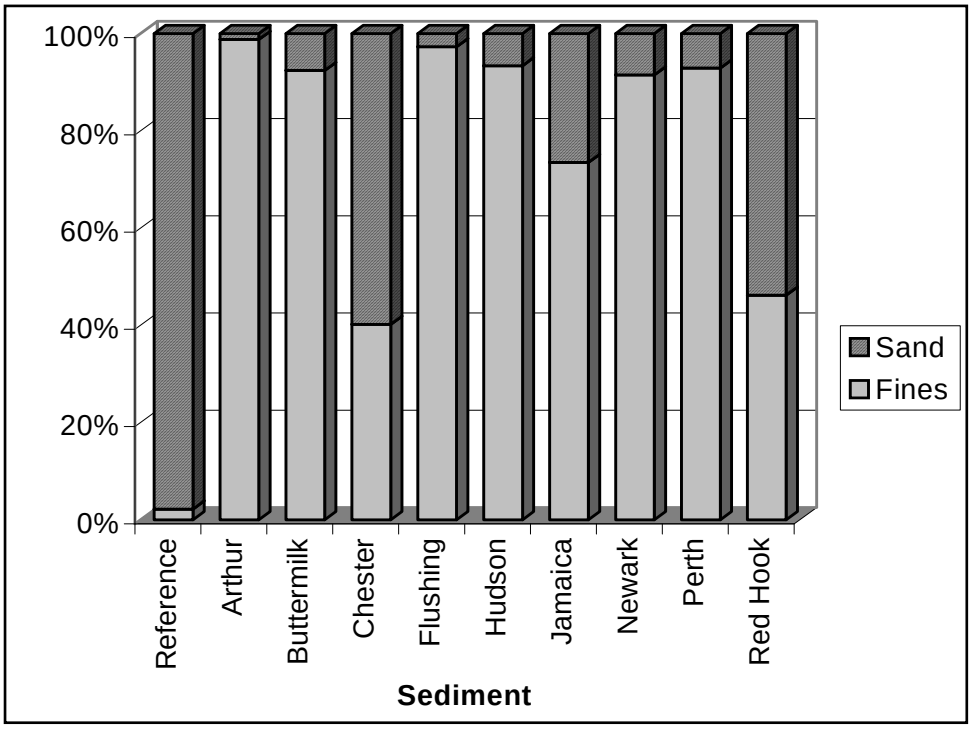

Figure 4. Comparison of particle size of sediments from New York Harbor. All sediments consisted of 0 percent gravel.

Sediment toxicity tests

Water quality parameters were within the acceptable ranges for the test organisms, as specified by the standard protocols (Appendix C). There was one deviation in temperature (17.9 °C), however, from the specified range in one chronic *L. plumulosus* test that was quickly rectified. This deviation did not appear to impact organism condition during the experiment. Total porewater ammonia concentrations (following purging) at test initiation and termination were below thresholds cited in the guidance (Appendix C6). In general, lethal and sublethal endpoint responses in the reference sand were low relative to the control in acute and chronic tests using *L. plumulosus* test, while both the reference and control sediments were similar in the *N. arenaceodentata* tests. Given that there was poor test organism performance in the reference sand for which mortality approximated or exceeded test sediments (Tables 6, 8, and 9), statistical comparisons were related to the control sediment.

Preliminary evaluation

Toxicity test results from the preliminary evaluation are presented in Table 7. Mean survival in the control was ≥ 90 percent, except for in the *A. abdita* test (87 percent). The *A. bahia* test had high overall survival in all sediments (>75 percent) and no significant differences among test sediments. Significant reductions in the survival endpoint relative to that of the control were observed in the 10-d *A. abdita* and *L. plumulosus* tests, associated with 6 and 7 (of 9) sediments, respectively. Some *A. abdita*, however, were observed swimming in water above the sediments in some treatments on test days 9 and 10, suggesting avoidance. Significant reductions in endpoints were observed in the 28-d *L. plumulosus* test for five site sediments, with four and three for survival and reproduction, respectively. Statistical differences were not found for *L. plumulosus* growth. Using Kendall's test, the ranks of the magnitudes of the three 28-d *L. plumulosus* endpoints ($W = 0.78$, $p < 0.05$), 10-d *L. plumulosus* ($W = 0.67$, $p < 0.05$) and *A. abdita* tests ($W = 0.62$, $p < 0.05$) were in significant agreement with one another,. The 28-d *N. arenaceodentata* test resulted in no significant reductions in survival or growth for any NYH sediment. Rather, biomass in four sediments was significantly higher than for the control.

Table 6. Toxicity testing results for the preliminary evaluation.

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

Asterisks represent a significant reduction compared to the control and underlines represent an ecologically significant endpoint response (i.e., > 20% reduction) relative to the control. Number signs represent a significantly increased response relative to the control.

Gradients of the two sediments for which the protocols detected the largest significant endpoint reductions (i.e., Hudson, Newark) were tested using the 10-d *L. plumulosus* protocol (Table 8). Survival in the 100 percent Hudson (11 ± 4 vs. 12 ±12 percent) and 100 percent Newark (27 ± 14 vs. 20 ± 11 percent) sediments corresponded well with previous testing (Table 7). A dose-dependant relationship was observed, with 10-d LC50 values of 32 (22-46) percent and 48 (43-54) percent for Hudson and Newark, respectively, based on the dry weight content of each sediment.

Table 7. Survivorship results for 10-day *Leptocheirus plumulosus* sediment dilution toxicity tests with Hudson (a) and Newark (b) sediments.

Nominal Treatment	Mean Survival (%)
(a)	
Control (0% Hudson)	95 ± 6
31% Hudson	48 ± 13*
43% Hudson	40 ± 12*
59% Hudson	37 ± 17*
82% Hudson	4 ± 4*
100% Hudson	12 ±12*
10-d LC50: 32 (22-46)% Hudson	
(b)	
Control (0% Newark)	95 ± 6
24% Newark	63 ± 10*
35% Newark	65 ± 12*
50% Newark	45 ± 14*
77% Newark	24 ± 12*
100% Newark	20 ± 11*
10-d LC50: 48 (43-54)% Newark	
Indicated dilution percentages are by sediment dry weight, executed by addition of Sequim Bay sediment. Asterisks represent a significant reduction compared to the control and underlines represent an ecologically significant endpoint response (i.e., > 20% reduction) relative to the control. LC50 values are also provided.	

Interlaboratory evaluation

The sediments used in the interlaboratory comparison were stored (4 °C) for 11 months from the time of collection (6 months following the preliminary evaluation). Overall, there were notable reductions (primarily LMW PAHs) in the measured concentrations of organic compounds (Tables 4 and 5) and subsequent increases in overall test organism survivorship in the sediments, compared with the same sediments used in the preliminary evaluation. There were two tests that failed to meet control endpoint acceptability criteria. The control and reference survival in the

10-d *A. abdita* test conducted by one lab was < 90 percent (Table 9) and the 28-d *L. plumulosus* test conducted by a different lab failed to meet the control survival (< 80 percent) and reproduction acceptability criteria (i.e., neonates in all control replicates; Table 10). Failure was unlikely the result of measured confounding factors (e.g., temperature, ammonia, low D.O.), as there were no recorded deviations in specified ranges (Appendix C). The failed 10-d *A. abdita* was conducted in reconstituted seawater while the failed 28-d *L. plumulosus* test was conducted in natural seawater. Acceptable 10-d *A. abdita* tests were completed by two of the three laboratories. Ranks of survival from the failed test agreed closely with the other two ($W = 0.86$, $p < 0.03$), but the results must be interpreted cautiously due to control acceptability failure; the focus will therefore be on the two tests that met criteria. Statistically significant reductions in survival which were beyond the test decision criterion (i.e., > 20 percent reduction relative to control) and were recorded for three (of seven) site sediments observed in both of the acceptable *A. abdita* tests (Table 9). In both cases, survival was significantly reduced in the 100 percent Newark and Chester sediments. There were statistically significant reductions in survival in one of the two *A. abdita* tests for the Hudson and Red Hook sediments. A dose-dependent survival reduction was not observed in the diluted Newark sediments. Survival was reduced in the 25 percent Newark sediment (78 ± 16 percent) in one lab; this particular sediment was unlikely "toxic" given: (1) survival in sediments with higher Newark content was not significantly reduced relative to the control in the same test and (2) the reduction was not beyond the decision criterion of 20 percent. In the 10-d *L. plumulosus* test conducted by Laboratory A, significant reductions in survival were observed for two of the site sediments (i.e., undiluted Hudson and Newark) and the reference. Although not significant, a clear dose-dependent relationship in survival was observed in the Newark dilution. Acute testing suggested toxicity in the highly contaminated sediments (i.e., Hudson, Newark) but not in other sediments.

Table 8. Survivorship results for the 10-day *Ampelisca abdita* tests and one 10-d *Leptocheirus plumulosus* test in the interlaboratory evaluation.

Treatment	10-d *Ampelisca*			10-d *Leptocheirus*
	Lab A[1]	Lab B	Lab C	
Control	84 ± 16	97 ± 4	95 ± 6	92 ± 8
Reference	NA	79 ± 8*	94 ± 5	52 ± 17*
Hudson	56 ± 12	74 ± 7*	85 ± 5	32 ± 18*
Chester	49 ± 21	72 ± 12*	32 ± 16*	76 ± 15
Red Hook	58 ± 18	82 ± 6	65 ± 24*	77 ± 16
100% Newark	56 ±17	75 ± 12*	71 ± 11*	52 ± 23*
75% Newark	66 ± 11	88 ± 7	85 ± 8	65 ± 26
50% Newark	61 ± 21	90 ± 8	86 ± 4	77 ± 27
25% Newark	58 ± 12	78 ± 16*	82 ± 21	92 ± 18

Asterisks represent a significant reduction compared to the control and underlines represent an ecologically significant endpoint response (i.e., > 20% reduction) relative to the control.
[1] This test failed acceptability criteria because control survival was less than 90%. Survival was significantly reduced in all test sediments except 75% and 50% Newark, compared to the control significant reduced.

Table 9. Endpoint results for the 28-day *Leptocheirus plumulosus* tests in the interlaboratory evaluation.

Treatment	Lab A			Lab B			Lab C[1]		
	Survival	Neonates/ Survivor	Biomass	Survival	Neonates/ Survivor	Biomass	Survival	Neonates/ Survivor	Biomass
Control	94 ± 4	2.5 ± 0.9	1.0 ± 0.2	92 ± 6	3.5 ± 1.2	2.0 ± 0.2	66 ± 11	0.1 ± 0.1	0.5 ± 0.1
Reference	85 ± 9	0.4 ± 0.1*	0.6 ± 0.1*	57 ± 16*	0.1 ± 0.1*	0.4 ± 0.2*	79 ± 4	0.0 ± 0.0	0.2 ± 0.0
Hudson	86 ± 9	0.1 ± 0.2*	0.5 ± 0.2*	84 ± 5	0.4 ± 0.4*	0.9 ± 0.2*	13 ± 8	0.0 ± 0.0	0.4 ± 0.1
Chester	84 ± 17	0.7 ± 0.4*	0.7 ± 0.3	87 ± 9	2.4 ± 1.8	1.4 ± 0.6	84 ± 9	0.0 ± 0.0	0.6 ± 0.1
Red Hook	98 ± 4	1.3 ± 0.7	1.0 ± 0.2	92 ± 3	3.6 ± 2.6	1.8 ± 0.6	87 ± 8	0.0 ± 0.0	0.8 ±0.2
100% Newark	81 ± 14	0.0 ± 0.1*	0.6 ± 0.1*	79 ± 9	0.8 ± 0.9*	1.2 ± 0.4*	55 ±15	0.0 ± 0.0	0.4 ± 0.1
75% Newark	89 ± 16	0.2 ± 0.2*	1.0 ± 0.3	81 ± 12	1.1 ± 0.6	1.3 ± 0.6	68 ± 12	0.0 ± 0.0	0.4 ± 0.1
50% Newark	98 ± 3	0.8 ± 0.6*	1.0 ± 0.1	85 ± 8	3.2 ± 2.6	1.1 ± 0.5*	73 ± 8	0.0 ± 0.1	0.6 ± 0.2
25% Newark	97 ± 4	2.3 ± 1.1	1.1 ± 0.1	94 ± 11	3.8 ± 1.5	1.3 ± 0.4	82 ±8	0.0 ± 0.1	0.6 ± 0.1

Asterisks represent a significant reduction compared to the control and underlines represent an ecologically significant endpoint response (i.e., > 20% reduction) relative to the control.
[1] This test failed acceptability criteria due to low survival and reproduction in the control.

Two of the three 28-d *L. plumulosus* tests satisfied control acceptability criteria. Again, results from the failed tests are provided, but only the successful tests will be discussed in detail. Neither successful test produced significant reductions in 28-d survival, unlike in the preliminary study (Table 10). The test conducted by Lab A exhibited significant reductions in reproduction (neonates/survivor) and biomass for five and two sediments, respectively; overall, one or more endpoints were significantly reduced in five sediments. In the Lab B test, there were also significant reductions for reproduction (two sediments) and biomass (three sediments); overall, reduction in at least one endpoint was found for three sediments. One Lab B control replicate was excluded as an outlier for the reproduction endpoint because it was assumed a second brood was produced based on the number of neonates (i.e., 333), which was beyond two standard deviations from the mean of the other replicates. Biomass obtained in Lab B was at least twofold greater than that obtained by Labs A and C. Overall, there were notable dose-dependent relationships in the Newark gradient observed for the survival and reproduction endpoints for both tests. There was reasonable consensus that the highly contaminated sediments (i.e., Hudson, Newark) were toxic based on the acute amphipod tests and the chronic *L. plumulosus* tests. Both chronic *L. plumulosus* tests determined significant reductions in one or more of the diluted Newark concentrations.

There were no statistically significant reductions in endpoints for any of the chronic *N. arenaceodentata* test methods; biomass, however, was generally higher in the site sediments than in the control, with significant increases in some sediments observed (Tables 11, 12, and 13). Additionally, overall polychaete biomass was four times greater in the 20-d PSDDA protocol (i.e., 8 mg Tetramarine® per worm per week) compared with the two protocols using the ERDC feeding ration (i.e., 4 mg Tetramarine® and 1 mg alfalfa per worm per week). These tests all satisfied acceptability criteria.

Table 10. Survivorship results for the 28-day *Neanthes arenaceodentata* tests in the interlaboratory evaluation. Number signs represent significant increases compared to the control.

Treatment	Lab A		Lab B		Lab C	
	Survival	Biomass	Survival	Biomass[1]	Survival	Biomass
Control	100 ± 0	3.3 ± 0.8	100 ± 0	6.0 ± 1.3	100 ± 0	2.9 ± 1.1
Reference	100 ± 0	2.2 ± 1.4	100 ± 0	6.2 ± 1.9	100 ± 0	2.6 ± 0.6
Hudson	100 ± 0	2.5 ± 1.2	100 ± 0	5.8 ± 1.0	100 ± 0	2.5 ± 0.7
Chester	100 ± 0	3.7 ± 3.0	100 ± 0	7.0 ± 1.9	100 ± 0	3.4 ± 0.8
Red Hook	90 ± 32	4.3 ± 1.2	100 ± 0	6.8 ± 1.9	100 ± 0	3.5 ± 1.1
Newark	90 ± 32	2.6 ± 0.9	100 ± 0	6.3 ± 1.4	100 ± 0	3.4 ± 0.5
75% Newark	100 ± 0	3.8 ± 2.2	100 ± 0	6.0 ± 1.8	100 ± 0	4.0 ± 1.1#
50% Newark	100 ± 0	3.7 ± 0.8	100 ± 0	6.8 ± 1.5	100 ± 0	4.5 ± 0.6#
25% Newark	100 ± 0	4.1 ± 1.0	90 ± 32	7.1 ± 1.3	100 ± 0	4.5 ± 0.7#

[1] Worms were almost twofold larger than in other laboratories.

Table 11. Survivorship results for the 20-day *Neanthes arenaceodentata* tests in the interlaboratory evaluation. Number signs represent significant increases compared to the control.

Treatment	Lab A		Lab B		Lab C	
	Survival	Biomass	Survival	Biomass	Survival	Biomass
Control	80 ± 20	11.6 ± 7.3	100 ± 0	16.0 ± 2.7	100 ± 0	7.9 ± 2.4
Reference	80 ± 20	7.4 ± 3.3	100 ± 0	18.4 ± 4.5	92 ± 18	12.8 ± 1.6
Hudson	68 ± 18	5.3 ± 2.9	96 ± 9	14.7 ± 2.6	100 ± 0	8.7 ± 2.8
Chester	80 ± 14	11.8 ± 8.8	96 ± 9	15.8 ± 0.8	96 ± 9	11.1 ± 3.2
Red Hook	68 ± 11	13.2 ± 8.6	100 ± 0	16.9 ± 3.1	88 ± 18	13.5 ± 3.7#
Newark	68 ± 27	7.9 ± 3.7	92 ± 11	12.7 ± 3.0	92 ± 18	10.5 ± 2.8
75% Newark	76 ± 17	6.9 ± 1.4	100 ± 0	16.4 ± 1.7	100 ± 0	8.1 ± 3.4
50% Newark	48 ± 23	9.9 ± 2.6	96 ± 9	16.2 ± 2.8	96 ± 9	9.8 ± 3.7
25% Newark	64 ± 17	7.5 ± 1.5	100 ± 0	15.2 ± 1.6	92 ± 11	11.7 ± 1.7

Table 12. Survivorship results for the modified 20-day *Neanthes arenaceodentata* tests in the interlaboratory evaluation. There were no significant differences observed.

Treatment	Lab A		Lab B		Lab C	
	Survival	Biomass	Survival	Biomass	Survival	Biomass
Control	92 ± 11	1.5 ± 0.1	96 ± 9	3.7 ± 0.8	96 ± 9	1.9 ± 0.3
Reference	80 ± 14	1.7 ± 0.4	84 ± 9	4.2 ± 1.1	88 ± 11	2.2 ± 0.5
Hudson	76 ± 17	2.4 ± 0.4	100 ± 0	4.1 ± 0.4	84 ± 17	1.8 ± 0.7
Chester	72 ± 18	2.5 ± 0.7	100 ± 0	4.4 ± 0.6	68 ± 11	2.7 ± 0.9
Red Hook	84 ± 26	2.0 ± 1.0	96 ± 9	3.8 ± 1.1	84 ± 17	2.1 ± 0.3
Newark	88 ± 18	1.6 ± 0.3	92 ± 11	4.4 ± 1.0	96 ± 9	1.6 ± 0.9
75% Newark	92 ± 11	1.7 ± 0.3	96 ± 9	4.5 ± 0.5	88 ± 11	2.3 ± 0.9
50% Newark	84 ± 17	1.9 ± 0.6	92 ± 11	4.5 ± 0.4	96 ± 9	2.2 ± 0.6
25% Newark	96 ± 9	1.6 ± 0.3	100 ± 0	3.7 ± 1.0	92 ± 11	2.4 ± 0.6

Comparison of test performance

The response magnitude (RM) was plotted against the minimum significant difference (MSD) for each test endpoint to assess performance (Figures 5 and 6). The upper left corner of the graphics represents best performance (i.e., high statistical power, high response) while the lower right represents worst performance (i.e., low power, low response). Plots of endpoints with standard deviations above the line (slope = 1; MSD = RM) were most capable of achieving significantly reduced responses (i.e., RM > MSD). This comparison is independent of achieving the test decision criterion (i.e., >20 percent or >10 percent). Performance was best for endpoints that had larger portions of the bars (± 1 SD) above the line. In both evaluations, higher statistical power (i.e., low MSD) was observed for survival endpoints and lower power (i.e., high MSD) for sublethal endpoints. *Ampelisca abdita* (and 10-d *L. plumulosus*) survival had relatively low variability and the mean RM was generally greater than the mean MSD (i.e., plots above the line), indicating good performance. The 28-d *L. plumulosus* sublethal endpoints, however, had higher RM values, which at least partially compensated for lower power (i.e., high MSD) in terms of ability to detect statistical significance (a surrogate for sediments identified as “toxic,” or “hits”). *Leptocheirus plumulosus* reproduction (open diamonds) consistently had the highest RM values of all endpoints. Interestingly, in the interlaboratory evaluations, reproduction outperformed (i.e., mean RM > mean MSD) biomass in one 28-d *L. plumulosus* test, while biomass outperformed reproduction in the other test. Reproduction in the preliminary evaluation did outperform biomass. The remaining test organism endpoints, including 28-d *L. plumulosus* survival, demonstrated low responsiveness (RM) to the test sediments.

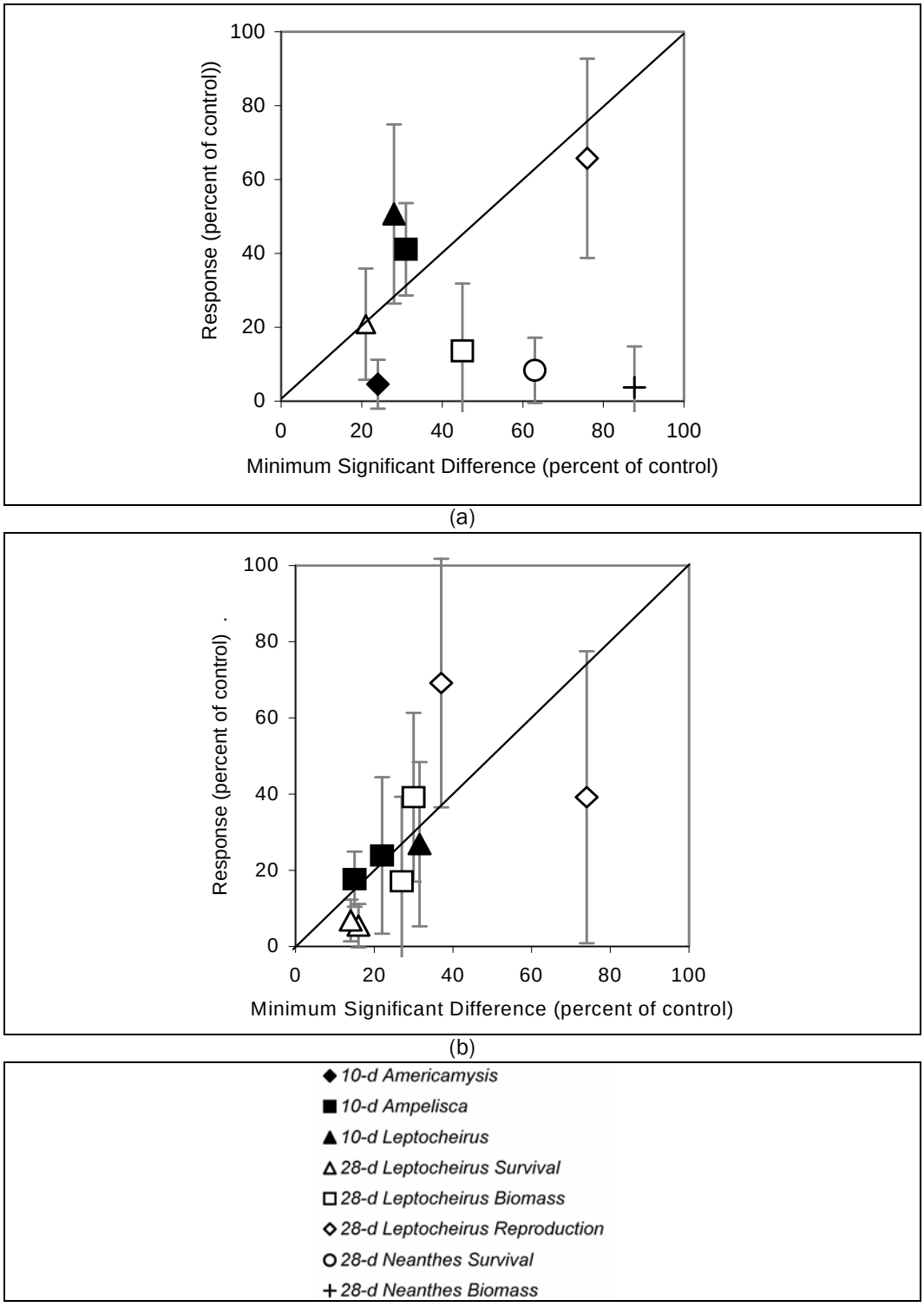

Figure 5. (MSD) for the preliminary (a) and interlaboratory (b) evaluations. The slope of the diagonal line is 1 (indicating RM = MSD). Performance was considered best for endpoints above the line. Acute endpoints = filled symbols, chronic endpoints = hollow symbols. The bars show one standard deviation from the mean.

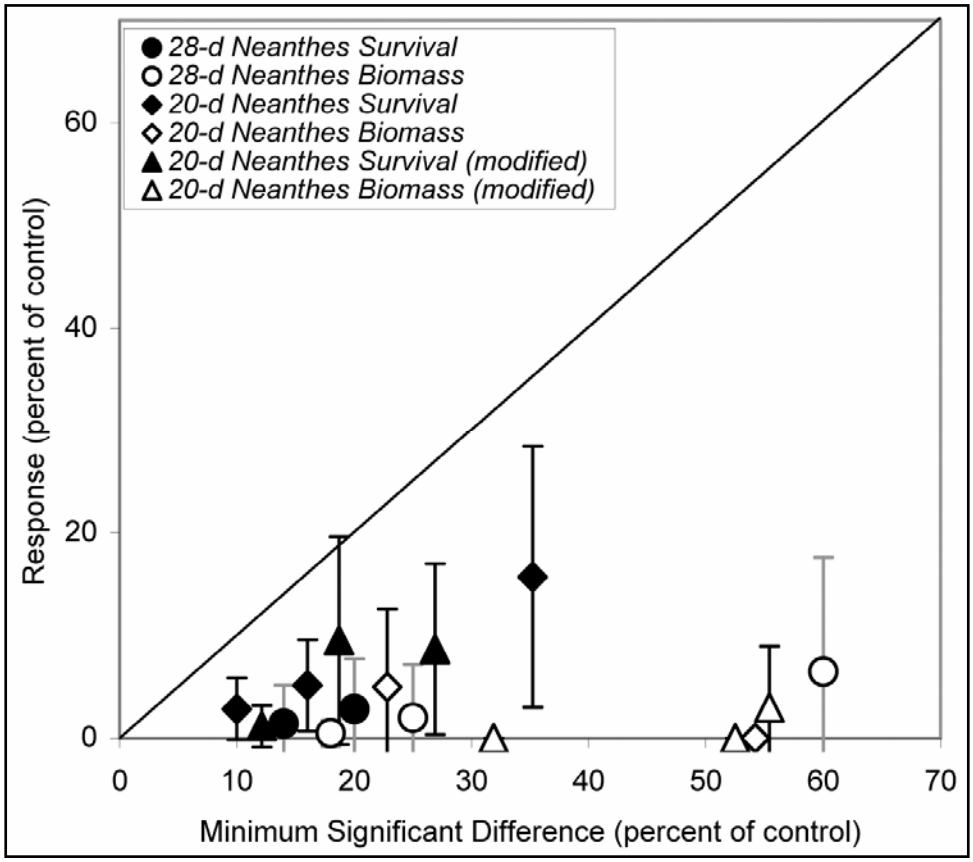

Figure 6. Plots of Neanthes arenaceodentata endpoint response magnitude (RM) and minimum significant difference (MSD) in the interlaboratory evaluation. The slope of the diagonal line is 1 (indicating RM = MSD). Performance was considered best for endpoints above the line. Survival = filled symbols; biomass = hollow symbols. The bars show one standard deviation from the mean.

There was good overall agreement in ranks of endpoint magnitude within the different test methods conducted in the different laboratories. Kendall's concordance test determined high coefficients of agreement between of endpoint ranks 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 ($W = 0.95$, $p < 0.10$). These relationships, however, were not significant ($\alpha < 0.05$) due to failures in one of the three *A. abdita* and *L. plumulosus* tests, which resulted in a smaller "n" value. The amphipod test methods in the interlaboratory evaluation generally distributed the sediments into low (Red Hook, 25 percent Newark, 50 percent Newark), medium (Chester, 75 percent Newark) and high (Hudson, 100 percent Newark) endpoint effect categories by ranks (Table 14). Due to lack of endpoint responsiveness to the sediments, there was no concordance between labs in *N. arenaceodentata* survival for the 28-d ($W = 0.20$, $p < 0.75$), 20-d

Table 13. Ranks of endpoint magnitudes for the preliminary (a) and interlaboratory (b) evaluations.

Sediment	10-d *Leptocheirus* Rank	10-d *Ampelisca* Rank	28-d *Leptocheirus* Rank[1]	Overall Mean Rank
(a)				
Control	1	1	2.5 ± 2.6	1.5 ± 0.9
Jamaica	2	3	3.3 ± 2.3	2.8 ± 0.7
Red Hook	5	2	4.2 ± 1.2	3.7 ± 1.5
Perth	3.5*	7*	4.2 ± 0.8	4.9 ± 1.9
Buttermilk	6*	8*	2.3 ± 0.6	5.4 ± 2.9
Arthur	9*	4	6.7 ± 1.0*	6.6 ± 2.5
Chester	7*	5*	9.3 ± 0.6*	7.1 ± 2.2
Reference	3.5	9.5*	10.0 ± 1.0*	7.7 ± 3.6
Hudson	11*	6*	7.2 ± 4.6*	8.1 ± 2.6
Flushing	8*	9.5*	8.0 ± 2.6*	8.5 ± 0.9
Newark	10*	11*	8.3 ± 0.8*	9.8 ± 1.3

Sediment	10-d *Leptocheirus* Rank	10-d *Ampelisca* Rank			28-d *Leptocheirus* Rank			Overall Mean Rank
	Lab A	B	C	Mean	A	B	Mean	
(b)								
Control	1	1	1	1.0 ± 0.0	2.8	2.2	2.5 ± 0.5	1.6 ± 0.9
50% Newark	3.5	2	3	2.5 ± 0.7	3.0*	5.3*	4.2 ± 1.6	3.4 ± 1.2
25% Newark	2	6	6	6.0 ± 0.0	2.0	2.2	2.1 ± 0.1	3.6 ± 2.2
Red Hook	3.5	4	8*	6.0 ± 2.8	2.7	2.2	2.4 ± 0.4	4.1 ± 2.3
75% Newark	6	3	4.5	3.8 ± 1.1	5.2	5.8	5.5 ± 0.5	4.9 ± 1.2
Reference	7.5*	5	2	3.5 ± 2.1	6.7*	9.0*	7.8 ± 1.6	6.0 ± 2.7
Chester	5	9*	9*	9.0 ± 0.0	6.3*	4.0	5.2 ± 1.6	6.7 ± 2.3
Hudson	9*	8*	4.5	6.3 ± 2.5	7.7*	7.3*	7.5 ± 0.2	7.3 ± 1.7
100% Newark	7.5*	7*	7*	7.0 ± 0.0	8.7*	7.0*	7.8 ± 1.2	7.4 ± 0.7

Only the test methods that exhibited statistically significant responses were included. The number "1" designates the highest endpoint magnitude (i.e., lowest toxicity) while higher numbers designate lower endpoint magnitude (i.e., highest toxicity). One standard deviation from the arithmetic mean is indicated. Sediments are listed from lowest to highest overall mean ranks. Asterisks indicate endpoint reductions of statistical significance.

[1] Represents the average of survival, reproduction and biomass ranks.

(W = 0.48, $p < 0.25$), or 20-d modified (W = 0.27, $p < 0.75$) tests. Surprisingly, there was significant agreement between ranks of 28-d *N. arenaceodentata* biomass between labs (W = 0.78, $p < 0.03$) despite the lack of an adverse response. It is important to recall this correspondence was related to increases, not reductions, in growth

compared to the control. Such agreement was not observed for 20-d *N. arenaceodentata* biomass ($W = 0.50$, $p < 0.25$) and modified 20-d *N. arenaceodentata* biomass ($W = 0.58$, $p < 0.1$).

For simplicity, Spearman correlations (r) were used to relate endpoint ranks (e.g., survival, biomass) with ranks of each major chemical class analyzed in the sediments. This comparison does not establish a causal link between contamination and test endpoints but is commonly applied in this fashion and is useful for comparing these methods and whether endpoints corresponded with measured contaminants and particle size issues. There were significant correlations ($p < 0.10$) between PAHs, PCBs, dioxins, and pesticides, which co-varied with one another (Tables 15 and 16). PAH concentrations, however, were not highly correlated with the other discussed chemical classes ($0.36 < r < 0.64$, $0.10 < p < 0.39$), except for some metals. In both evaluations, acute *L. plumulosus* survival was closely and inversely related to all chemical classes. *Ampelisca abdita* survival displayed a weak inverse relationship with PAHs. Significant inverse correlations were observed for 28-d *L. plumulosus* survival (ΣDDT, PCBs), biomass (HMW PAHs, PCBs) and reproduction (HMW PAHs). There were no strong relationships between *A. bahia* survival and organic chemical classes, although there were weak correlations ($-0.6 < r < -0.7$, $p < 0.1$) with some metals (Sb, As, Be, Se) and a strong correlation with Be ($r = -0.72$, $p < 0.02$). Occurrences of significant relationships for 28-d and 20-d PSDDA *N. arenaceodentata* endpoints to chemical classes were inconsistently observed in the interlaboratory evaluation. The highest Spearman coefficients were observed for the 10-d *L. plumulosus* tests and the two interlaboratory 28-d *L. plumulosus* tests for PCBs, ΣDDT and dioxins ($r > 0.8$). Percent fines (i.e., silts and clays) were positively correlated with PCBs ($r = 0.63$, $p < 0.10$), ΣDDT ($r = 0.70$, $p < 0.05$), dioxin ($r = 0.75$, $p < 0.05$), and most metals ($r > 0.8$, $p < 0.05$). Percent sand was sporadically positively correlated with *N. arenaceodentata* survival in two of the three modified 20-d tests ($r = 0.73$, $p < 0.10$; $r = 0.78$, $p < 0.05$). *Neanthes arenaceodentata* biomass was positively correlated with sand in one of the 28-d tests (0.69, $p < 0.10$) and negatively correlated with sand in one of the modified 20-d tests ($r = -0.80$, $p < 0.05$). No other significant relationships were observed.

Table 14. Spearman correlation coefficients between ranks of endpoint magnitude and ranks of selected chemical classes for preliminary evaluation data.

Endpoint	LMW PAHs	HMW PAHs	PCBs	Σ DDT	Dioxin
10-d *Leptocheirus* Survival	-0.63#	-0.57#	-0.92*	-0.81*	-0.82*
10d *Ampelisca* Survival	-0.65*	-0.60#	-0.63#	-0.21	-0.55
10d *Americamysis* Survival	-0.18	-0.10	-0.35	-0.10	-0.45
28d *Leptocheirus* Survival	-0.40	-0.37	-0.62#	-0.65*	-0.52
28d *Leptocheirus* Biomass	-0.64*	-0.63#	-0.64*	-0.55	-0.41
28d *Leptocheirus* Reproduction	-0.44	-0.64#	-0.19	-0.21	-0.13
28d *Neanthes* Survival	-0.42	-0.28	-0.21	-0.07	-0.29
28d *Neanthes* Biomass	0.54	0.51	0.03	-0.40	-0.40
Asterisks indicate significance at the 0.05 significance level; number signs indicate significance at the 0.10 level.					

Table 15. Spearman correlation coefficients between ranks of endpoint magnitude and ranks of selected chemical classes for interlaboratory evaluation data.

Endpoint	Facility	LMW PAHs	HMW PAHs	PCBs	Σ DDT	Dioxin
10-d *Leptocheirus* Survival	A	-0.67 #	-0.51	-0.99 *	-0.94 *	-0.88 *
10d *Ampelisca* Survival	A	-0.55	-0.47	-0.33	-0.33	-0.17
	B	-0.57	-0.46	-0.39	-0.43	-0.18
	C	-0.49	-0.61 #	0.18	0.25	0.05
28d *Leptocheirus* Survival	A	-0.81 *	-0.69 #	-0.70 #	-0.76 *	-0.54
	B	-0.57	-0.46	-0.86 *	-0.82 *	-0.85 *
	C	-0.18	-0.04	-0.82 *	-0.89 *	-0.69 #
28d *Leptocheirus* Biomass	A	-0.67 #	-0.52	-0.89 *	-0.82 *	-0.75 #
	B	0.14	0.36	-0.58	-0.67 #	-0.36
	C	-0.42	-0.23	-0.85 *	-0.93 *	-0.70 #
28d *Leptocheirus* Reproduction	A	-0.71 #	-0.57	-0.96 *	-0.93 *	-0.88 *
	B	-0.64 #	-0.46	-1.00 *	-0.96 *	-0.85 *
	C	0.0	0.0	0.0	0.0	0.0
28d *Neanthes* Survival	A	-0.16	-0.32	0.0	0.16	-0.48
	B	0.61	0.61	0.61	0.41	0.62
	C	0.0	0.0	0.0	0.0	0.0
28d *Neanthes* Biomass	A	-0.45	-0.23	-0.85 *	-0.88 *	-0.56
	B	-0.31	-0.18	-0.85 *	-0.76 *	-0.86 *
	C	-0.69 #	-0.58	-0.76 *	-0.67 #	0.68 #
20d *Neanthes* Survival	A	0.85 *	0.85 *	0.41	0.33	0.30
	B	-0.50	-0.39	-0.62	-0.62	-0.47
	C	0.28	0.07	0.62	0.68 #	0.22
20d *Neanthes* Biomass	A	0.07	0.29	-0.54	-0.68 #	-0.34
	B	-0.21	-0.04	-0.50	-0.64 #	-0.34
	C	-0.25	-0.07	-0.71 #	-0.75 *	-0.47
Modified 20d *Neanthes* Survival	A	-0.45	-0.40	-0.34	-0.22	-0.10
	B	0.0	-0.09	-0.08	0.02	-0.35
	C	-0.38	-0.38	-0.07	0.02	0.08
Modified 20d *Neanthes* Biomass	A	0.34	0.30	0.20	0.07	-0.05
	B	0.40	0.36	0.40	0.36	0.24
	C	-0.04	0.0	-0.54	-0.43	-0.81 *
Asterisks indicate significance at the 0.05 significance level; number signs indicate significance at the 0.10 level.						

A rough graphical method to gauge test performance focusing on sediment quality guidelines was adopted from a figure in Wenning et al. (2005), who indicated SQGs are most useful for initial screening-level assessments when actual sediment chemistry values are below SQGs, suggesting no effect and thus no further action. By plotting binary sediment toxicity from tests (i.e., toxic or not toxic) against sediment contamination, a distribution of test correspondence with chemistry was elucidated. This representation is applicable because management decisions are often based on whether sediments pass or fail bioassay tests, regardless of the degree to which survival is affected. For this purpose, organic sediment contamination was simplified into mean sediment quality guideline quotients (SQG –Qs) of PAHs, PCBs, and DDTs (i.e., Mean of [PAH] / $PAH_{ER\text{-}M}$ + [PCB] / $PAH_{ER\text{-}M}$ + [DDT] / $DDT_{ER\text{-}M}$). The contaminants selected for inclusion in calculating the SQG-Qs were the same in the comparison of the acute and chronic tests and based on those that had available ERMs. Although this approach is an oversimplification of actual sediment toxicology and is criticized due to sediment-specific bioavailability and partitioning issues, it has some utility for comparing the potential false positives and negatives in the tests conducted in this study. The 28-d *L. plumulosus* tests conducted had a narrower zone of uncertainty and no apparent false negatives at higher sediment contamination levels, a concern for conservative management decisions, compared with the 10-d *A. abdita* tests (Figure 7).

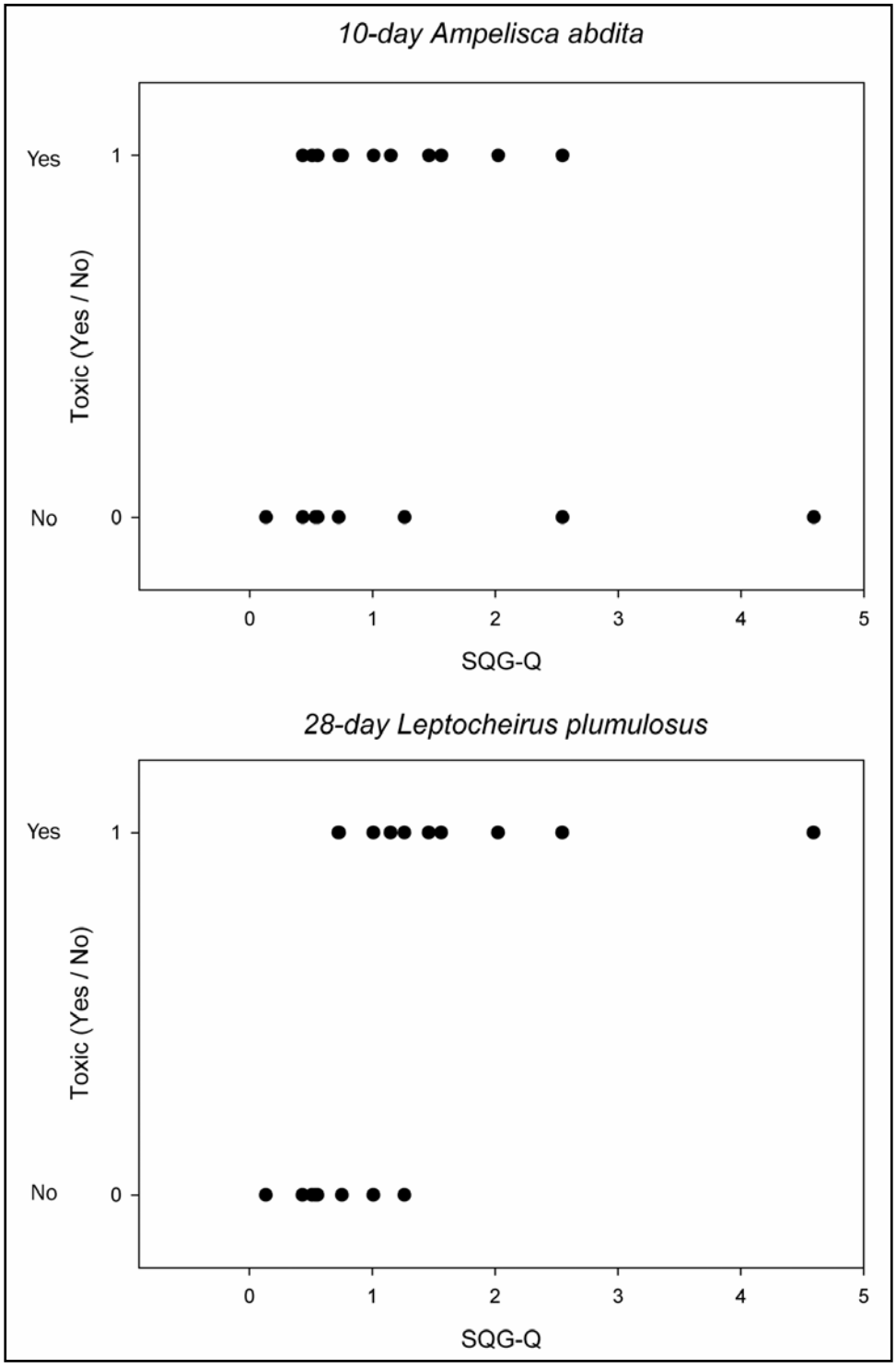

Figure 7. Plots of sediments suggested toxic or non-toxic by responsive test methods. Data are aggregated from both evaluations are compared to mean sediment quality guideline quotients (SQG -Q = [PAH]/ERM + [PCB]/ERM + [DDT]/ERM).

4 Discussion

Based on our data and previous study (McGee et al. 2004; Farrar et al. 2005a; Farrar et al. 2005b), the relative sensitivity of acute and chronic amphipod protocols tested in the same sediments can vary between sediments, evaluations, duration, and endpoints. That is, a particular test method may be less responsive than another method in one region but may be the most responsive test in another region. For instance, Bay et al. (2005) reported that the 10-d method using *Ampelisca abdita* was less responsive than other amphipods (i.e., *Eohaustorius estaurius* and *Rhepoxynius abronius*) in sediments they tested, while *A. abdita* is widely known as a responsive test organism in other evaluations. Farrar et al. (2005b) reported that the relative responsiveness of the 10-d and 28-d *L. plumulosus* test methods and the 28-d *N. arenaceodenta* protocol depended on whether the major chemical class was PAHs, PCBs, or metals. Ciarelli et al. (1998) found more significant endpoint reductions in chronic than in acute tests using *Corophium* spp. to assess dredged material, stating that application of chronic tests is more useful at levels resulting in low to moderate toxicity.

In the current study, the acute *Americamysis bahia* test was unresponsive to all sediments tested while there were no significant endpoint reductions observed in any of the *Neanthes arenaceodentata* tests. The *N. arenaceodentata* test endpoints have responded with significant endpoint reductions compared with sediments in other studies (e.g., Bridges et al. 1997; Farrar et al. 2005a; Farrar et al. 2005b) and have proven useful in the DMMP (PSDDA 2002). It is unknown at this time whether *N. arenaceodentata* may be less sensitive to organic sediment contaminants (e.g., Anderson et al. 1998) or if the higher organic content of the test sediments may have masked reductions in growth (Bridges et al. 1997). Farrar et al. (2005b) reported that while the 28-d *N. arenaceodentata* protocol was less responsive to PAHs than *L. plumulosus*, it was equally responsive to a PCB-contaminated sediment and equally or more responsive to a metals-contaminated sediment. The Anderson study, however, attributed the lower number of significant 20-d *N. arenaceodentata* endpoint reductions, relative to a 10-d amphipod, *R. abronius*, test to lower statistical power inherent in the design (five replicates, five worms / replicate). In this evaluation, however, only the

tests that utilized amphipods (i.e., *Ampelisca abdita*, *Leptocheirus plumulosus*) provided significant endpoint reductions relative to the control, and the 28-d *L. plumulosus* test was the only responsive (i.e., sensitive) chronic test method. The 28-d *L. plumulosus* and 10-d *A. abdita* protocols had similar abilities to identify statistically significant endpoint reductions (e.g., toxic sediments) in this evaluation (Tables 6, 8, and 9). By applying Kendall's concordance test to endpoint ranks in a similar application as other comparison studies (Mearns et al. 1985; Schlekat et al. 1995; USEPA 2001; Bay et al. 2003), it was determined that there was good agreement between responsive endpoints (i.e., *A. abdita* survival, *L. plumulosus* endpoints) in different laboratories, suggesting all tests were valid and repeatable.

Table 16. Probability of PAH toxicity to amphipods as predicted by the Sum PAH model (Swartz et al. 1995).

Sediment	Mortality > 13%	Mortality > 24%	Uncertain Toxicity	Not Toxic
Arthur	0.12	0.00	0.12	0.88
Buttermilk	0.24	0.00	0.24	0.76
Chester	0.40	0.17	0.23	0.60
Flushing	0.25	0.00	0.25	0.75
Hudson	0.15	0.00	0.15	0.85
Jamaica	0.00	0.00	0.00	1.00
Newark	0.32	0.08	0.24	0.68
Perth	0.00	0.00	0.00	1.00
Red Hook	0.28	0.03	0.25	0.72
Output suggests a relatively low likelihood of PAH toxicity (0.0 to 0.17).				

The acute and chronic test methods that used amphipods identified toxicity for some tested sediments. All amphipod test methods suggested toxicity via significant reductions in at least one endpoint relative to the control for the most highly contaminated sediments tested (i.e., Hudson, 100 percent Newark) with slight distinctions. Among tests that satisfied acceptability criteria, the acute *A. abdita* tests unanimously exhibited survival reductions for the Newark and Chester sediments while the chronic *L. plumulosus* tests unanimously exhibited endpoint reductions for the Newark and Hudson sediments. Chester and Newark had the highest PAH contamination, and among the tested sediments were assigned the greatest probability of inducing toxicity by the Sum PAH model (Swartz et al. 1995). Although PAHs in these sediments exceeded

ER-Ms (NOAA SquiRT Tables 1999) and consensus standards (Swartz 1999), the Sum PAH Model (Table 6), which accounts for bioavailability, predicted a low probability (≤ 0.17) of substantial PAH toxicity (> 24 percent mortality) in any of the sediments (Table 6). This result may be interpreted to mean that any observed toxicity is more likely attributable to other contaminants. Fay et al. (2000) suggested low amphipod, in this case *A. abdita*, sensitivity to certain high molecular weight PAHs. The Hudson sediment, which also induced significant endpoint reductions, was predominantly contaminated with PCBs, dioxins, and furans. Barber et al (1998) provided evidence that benthic macroinvertebrates (e.g., amphipods) are not sensitive to TCDD due to lack of an Ah receptor (for this reason dioxins were not TEQ adjusted). Therefore, observed toxicity may be more of a response to PCBs (or pesticides in the case of the Arthur sediment), which is supported by the consensus-based median standard of 340 μg/kg developed by MacDonald et al. 2000). An alternative hypothesis for the response of *A. abdita* in Chester sediment to PAH contamination may be related to the more coarse grain size of that sediment (Figure 4), a confounder in toxicity tests that is discussed below. The chronic *L. plumulosus* test did suggest toxicity and responded in a dose-dependant fashion in the Newark dilution, while the acute *A. abdita* tests did not. There was a statistically significant reduction relative to the control (but less than the decision criterion of > 20 percent) in *A. abdita* survival in one of the two tests observed in the 25 percent Newark treatment This was not intuitive given lack of significant effect observed in the 75 percent and 50 percent Newark treatments in that same test. The acute *L. plumulosus* test did exhibit a dose-dependent, but not statistically different, response to the Newark dilution. *Ampelisca abdita* and *L. plumulosus* endpoints may respond differently to different chemical stressors. Not enough is understood about such interactions between the test organisms and the complex sediment matrix to directly conclude that one organism or test method is more sensitive than another to contaminated sediments. These observations highlight the need to test with multiple species and endpoints, even among very sensitive amphipods, to aid in accounting for uncertainty between organism and endpoint response to different stressors (Cairns 1986).

Performance parameters indicated different advantages of the assessed endpoints. The lethality endpoints (both acute and chronic) in this study exhibited lower MSD values than sublethal endpoints, and thus smaller differences relative to the control are needed to detect statistical

differences; the detection of such differences, however, is contingent on a large enough response of the endpoint to the sediments. Sublethal endpoints, notably *L. plumulosus* reproduction, were more responsive to the same sediments and corresponded more closely with contamination levels (Tables 13 and 14). More variation, however, was associated with sublethal responses, requiring larger differences in the sample populations to establish statistically significant reductions in sublethal measurements. The relative responsiveness and MSD values of the sublethal endpoints varied between the different tests (Figure 5), further illustrating the range in performance.

While none of the *N. arenaceodentata* endpoints were reduced significantly by exposure to these sediments, differences did exist between protocols. The 20-d PSDDA protocol produced much larger polychaete worms (mean = 11.6 ± 3.4 mg / worm) than the 28-d and modified 20-d protocols that allocated less supplemental food to test organisms (4.5 ± 1.5 and 2.7 ± 1.1 mg / worm, respectively). This larger feeding ration could explain the higher porewater ammonia levels at termination in the 20-d PSDDA test than in the other two *N. arenaceodentata* tests (Appendix C6.). Additionally, the 20-d protocol (modified or not) may inherently produce greater growth variability (and less power) due to dominant worm interactions within replicates containing five test organisms (Bridges et al. 1997; Anderson et al. 1998), although more power for the survivorship endpoint is available (five versus one worm). There is concern that the greater amount of supplemental food allocated to polychaetes in the 20-d protocol may reduce endpoint responsiveness to the sediments, and thus reduce measurements of toxicity; this concept is discussed by several researchers (Bridges et al. 1997, McGee et al. 2004). Lotufo et al. (2001) suggested that amphipods with higher lipid content may be more tolerant, based on lipid normalized lethal residue values inducing 50 percent morality to DDT. Thus, feeding organisms during tests maintains or increases their fat reserves, potentially making them less sensitive to contamination. This logic may also apply to comparisons of survival in the 10-d (not fed) and 28-d (fed) tests with *L. plumulosus*, where significant decreases were observed using the 10-d method but not the 28-day method (Tables 7, 9, and 10) when the amphipods were exposed to the same sediments.

Multiple factors (Kuhn et al. 2002; McGee et al. 2004) may contribute to lowering the number of statistically significant reductions in chronic test

endpoints (i.e., survival, growth, and reproduction). First, sublethal endpoints used in chronic tests are inherently more variable than is survival. Attempts have been made to standardize reproduction assessments as a means of reducing variation in endpoint measurements, but the costs of some of those approaches are likely to outweigh the potential benefit of their use (Gray et al. 1998). In addition, non-treatment-induced variation can be decreased by increasing replication within each treatment/sediment; some clients have requested 10 replicates in chronic *L. plumulosus* tests to increase the power of sublethal measurements. Second, the addition of supplemental food (Bridges et al. 1997; McGee et al. 2004) to the sediment matrix in chronic toxicity tests may reduce chemical exposure by three mechanisms:

- Chemicals may be less available due to the increased percentage of organic matter (i.e., TOC) in the sediments,
- Some test species may not burrow into or ingest as much contaminated sediment because food is readily available on the surface, and
- The nutritional condition of test organisms (e.g., higher lipid content) in chronic tests may be greater than the vigor of organisms in acute tests that receive no supplemental food over a 10-day period (Lotufo et al. 2001).[1]

Third, renewing the overlying water may effectively remove soluble contaminants from sediments. *Leptocheirus plumulosus* survival in evaluations of contaminated sediments, for instance, can be greater in 28-d tests than 10-d tests (as observed in Table 7), a phenomenon that may be explained by decreases in contaminant availability or concentration due to supplemental feeding and/or water exchanges unique to the chronic test. Fourth, differences between acute survival and chronic sublethal endpoint response may be related to differences in mode of action rather than relative sensitivity (Kuhn et al. 2002). That is, a particular chemical concentration may increase amphipod mortality but no influence on physiological factors governing the growth of surviving individuals.

Interpretation of sediment toxicity assay results is further complicated by uncertainty (Vorhees et al. 2002) and confounding factors (Postma et al.

[1] Incidentally, such food-ration effects are of less concern in freshwater sediment toxicity test methods (i.e., *Chironomus tentans, Hyallela azteca*) because both acute and chronic guidance involves feeding (USEPA 2000).

2002). In general, it is important to recognize that endpoint response to sediments is not necessarily indicative of contamination; although coarse grain size (Figure 4) was unlikely to significantly influence comparisons of sediments in the current study based on Spearman correlations between percent sand and endpoint magnitude, it is a well known confounding factor (USEPA 1994; USEPA 2001). The exception to this factor was probably the reference sand, which differed greatly from the test sediments in terms of particle size. This knowledge may create implications for use of the reference sand in NYH evaluations. A more relevant confounder to this study was ammonia, because concentrations in bulk NYH sediments are often elevated. To alleviate this concern, sediments were purged well below relevant threshold values (USEPA 1994), as recommended by Ferretti et al. (2000). Ammonia concentrations were monitored at test initiation and termination (Appendix C6), and they were unlikely to confound test results in this study due to purging measures and lack of correlation with endpoint magnitude. They should, however, be monitored carefully in NYH evaluations.

Advantages and disadvantages have been identified for the chronic *L. plumulosus* test. These considerations likely apply to chronic *N. arenaceodentata* protocols and chronic amphipod tests conducted in other studies (Ciarelli et al. 1998; Kuhn et al. 2002) that lack standard methods. The benefits of the *L. plumulosus* test, based on the current study, include marginally more sediments identified with statistically significant endpoint reductions (i.e., potential for toxicity), larger sublethal endpoint response to sediment contamination than lethality endpoints, stronger relationships with sediment contamination than acute survival (Tables 15 and 16), use of endpoints that satisfy regulatory requirements, and fewer false negatives (Finkelstein and Kern 2005; Figure 7). An additional advantage of using *L. plumulosus* over *A. abdita* to consider is that the former organism is laboratory-cultured while the latter must be obtained from the field, a condition that may lead to some level of stress in laboratory assays and greater inter-test variability. The drawbacks of the 28-d *L. plumulosus* test include the following: (1) laboratories may experience greater difficulty meeting acceptability criteria, (2) difficulties may arise related to teasing out statistical differences from inherently more variable sublethal endpoint measurements and increased labor intensiveness, and (3) consequent cost, as compared with acute tests. In a survey of four test facilities that conduct whole-sediment toxicity tests, the cost of performing an assessment of one

sediment (i.e., five replicates) using 10-d *L. plumulosus* and *A. abdita* tests ranged from $630 – $1,700, while the 28-d *L. plumulosus* test ranged from $1,800 – $2,500 and the 28-d *N. areanceodenta* test ranged from $1,080 – $2,000.

5 Recommendations

Sediment toxicity tests that use relevant benthic test organisms are a crucial component of dredged material evaluations. These tests apply measurement endpoints to gauge the extent of toxicity in materials where complex or unknown chemical interactions and bioavailability issues cannot be explained by chemistry or models alone. Test method selection should be contingent on management goals (e.g., desired level of protection, indigenous taxa) and knowledge of organism suitability to site-specific sediment composition (e.g., grain size, ammonia levels). Based on the data set presented in this study and relevant literature, we offer the following recommendations for whole sediment toxicity evaluations:

1. Regulations implementing the MPRSA require evaluating the potential for dredged material to cause chronic toxicity. The use of chronic toxicity tests provides a direct approach for evaluating this potential. However, the chronic tests evaluated in this study were not consistently more sensitive to contaminated sediments, a finding supported by other comparisons of acute and chronic sediment tests. In some cases, in fact, sediments can produce a toxic response in an acute test but no response in a chronic test. The chronic tests require a longer period to obtain results, are more likely to fail to meet performance standards for test acceptability and to require more maintenance, and they are approximately 1.5 – 3 times as expensive as acute tests. At this time, we recommend the continued use of acute tests. The results of this study demonstrate that acute tests are predictive of chronic toxicity and meet the bioassay requirements for the assessment that "no significant undesirable effects will occur due either to chronic toxicity..." (CFR 227.6(c)(3)). The need for chronic tests should be determined on a project-by-project basis. In some cases, more information is desired or the need for using a chronic exposure is identified (e.g., when toxicity is expected to be driven by highly hydrophobic contaminants that may not reach a sufficient portion of steady state in a shorter exposure period).

2. The 10-d *Americamysis bahia* test method should be discontinued and replaced. The *A. bahia* method uses an epibenthic organism with a loose association with and limited exposure to the sediment, and it was

not responsive to the sediments tested in this study. It should be used only to assess the effects of sediment on water column organisms that do not feed on benthic material or organisms.

3. The *L. plumulosus* 10-d protocol should be selected for assessment of sediments proposed for open-water disposal. In some cases, as described in the first recommendation above, the 28-day *L. plumulosus* protocol can be used as determined by project-specific needs for decision-making. Protocols using *L. plumulosus* are desirable because they
   a. are sensitive to NYH sediments as demonstrated in this study
   b. experience direct and continuous exposure to the sediment
   c. are easy to culture in the laboratory (*A. abdita* is field collected)
   d. have a proven track record using both natural and reconstituted seawaters.

4. The current study, by itself, does not provide evidence to support the use of the 20-d or 28-d *Neanthes arenaceodentata* test methods to evaluate New York Harbor sediments because no significant decreases were observed for any of the test sediments collected from NYH. On a national level, however, these methods have been found to be useful in other applications (e.g., DMMP, SCWRRP).

5. The current reference sediment is unsuitable for dredged material evaluations. That sediment has a very different grain size distribution and organic carbon content than much NYH dredged material. These differences create a strong potential for confounding factors to complicate the interpretation of toxicity test results. A more suitable reference sediment should be identified for evaluating sediments from NYH.

References

American Society of Testing Materials (ASTM). 1998. Standard Guide for Conducting 10-day Static Sediment Toxicity Tests with Marine and Estuarine Amphipods. Method E1367-99. West Conshohocken, PA: ASTM International.

American Society of Testing Materials (ASTM). 2000a. Standard Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates. Method E1706-00. West Conshohocken, PA: ASTM International.

American Society of Testing Materials (ASTM). 2000b. Standard Guide for Conducting Sediment Toxicity Tests with Polychaetous Annelids. Method E1611-00. West Conshohocken, PA: ASTM International.

American Society of Testing Materials (ASTM). 2000c. Standard Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates. Method E1706-00. West Conshohocken, PA: ASTM International.

American Society of Testing Materials (ASTM). 2002. Standard Guide for Designing Biological Tests with Sediments. Method E1525-02. West Conshohocken, PA: ASTM International.

American Society of Testing 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: ASTM International.

Anderson, B. S., J. W. Hunt, B. M. Phillips, S. Tudor, R. Fairey, J. Newman, H. M. Puckett, M. Stephenson, E. R. Long, and R. S. Tjeerdema. 1998. Comparison of marine sediment toxicity test protocols for the amphipod *Rhepoxynius abronius* and the polychaete worm *Nereis (Neanthes) arenaceodentata*. *Environmental Toxicology and Chemistry* 17: 859–866.

Aqua Survey, Inc. 2001. Standard Operating Procedure: Static Non-renewal Test for Determining Acute Toxicity of Sediments to Mysid Shrimp. OP No. SOP/SED/1560. Flemington, NJ: Aqua Survey, Inc.

Barber, T. R., D. J. Chappie, D. J. Duda, P. C. Fuchsman, and B. L. Finley. 1998. Using a spiked sediment test to establish a no-effect concentration for dioxin exposure to the amphipod Ampelisca abdita. *Environmental Toxicology and Chemistry* 17: 420–424.

Battelle. 2000. Sediment Bioassay Testing using *Mysidopsis bahia*. SOP-MSL-T-024-02. Sequim, WA: Battelle Pacific Northwest National Laboratories, Marine Science Laboratory.

Bay, S. M., A. Jirik, and S. Asato. 2003. Interlaboratory variability of amphipod sediment toxicity tests in a Cooperative Regional Monitoring Program. *Environmental Monitoring Assessment* 81:257-268.

Bay, S., T. Gries, B. Anderson, B. Phillips, L. Field, D. Moore, and D. Greenstein. 2005. Comparison of marine amphipod test species responsiveness to contaminated field sediments. Poster presentation at the national meeting of the Society of Environmental Toxicology and Chemistry, Baltimore, MD, 14–17 November 2005.

Bridges, T. S., J. D. Farrar, and B. M. Duke. 1997. The influence of food ration on sediment toxicity in *Neanthes arenaceodentata*. *Environmental Toxicology and Chemistry* 16: 1659–1665.

Cairns, J., Jr. 1986. The myth of the most sensitive species. *BioScience* 36: 670–672.

Chapman, G. A., D. L. Denton, and J. M. Lazorchak. 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. Cincinnati, OH: USEPA.

Ciarelli, S., W. A. P. M. Vonck, N. M. van Straalen, and J. Stronkhorst. 1998. Ecotoxicity assessment of contaminated dredged material with the marine amphipod *Corophium volutator*. *Archives of Environmental Contamination and Toxicology* 34: 350–356.

Dillon, T. M., D. W. Moore, and A. B. Gibson. 1993. Development of a chronic sublethal bioassay for evaluating contaminated sediment with the marine polychaete worm *Nereis* (*Neanthes*) *arenaceodentata*. *Environmental Toxicology and Chemistry* 12: 589–605.

Farrar, J. D., A. J. Kennedy, G. R. Lotufo, and D. J. Greenstien. 2005a. Assessment of California coastal sediments using acute and chronic sediment bioassays. Draft technical report for the Dredging Operations and Environmental Research Program (in press). Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Farrar, J. D., A. J. Kennedy, G. R. Lotufo, and C. M. McNemar. 2005b. Comparative evaluation of the efficacy of acute and chronic sublethal sediment tests for assessing sediment quality. Poster presentation at the national meeting of the Society of Environmental Toxicology and Chemistry, Baltimore, MD, 14–17 November 2005.

Fay, A. A., B. J. Brownawell, A. A. Elskus, and A. E. McElroy. 2000. Critical body residues in the marine amphipod *Ampelisca abdita*: Sediment exposures with nonionic organic contaminants. *Environmental Toxicology and Chemistry* 19: 1028–1035.

Ferretti, J. A., D. F. Calesso, and T. R. Hermon. 2000. Evaluation of methods to remove ammonia interference in marine sediment toxicity tests. *Environmental Toxicology and Chemistry* 19(8): 1935–1941.

Finkelstein, K., and J. Kern. 2005. Improvement in correlation between chemistry and toxicity using the 28-day sediment toxicity test. Abstract A1-09 in Remediation of contaminated sediments – 2005: Finding achievable risk reduction solutions, R. F. Olfenbuttel and P. J. White, ed. *Proceedings of the Third International Conference on Remediation of Contaminated Sediments,* New Orleans, LA; January 24–27, 2005. ISBN 1-57477-150-7. Columbus, OH: Battelle Press.

Gray, B. R., V. L. Emery, D. L. Brandon, R. B. Wright, B. M. Duke, J. D. Farrar, and D. W. Moore. 1998. Selection of optimal measures of growth and reproduction for the sublethal *Leptocheirus plumulosus* sediment bioassay. *Environmental Toxicology and Chemistry* 17: 2288–2297.

Gulley, D. D. 1996. TOXSTAT® Version 3.3. Laramie, WY: University of Wyoming, Department of Zoology and Physiology.

Johns, D. M., T. C. Ginn, and D. J. Reish. 1990. Protocol for juvenile *N. areanceodentata* sediment bioassay. EPA 68-D8-0085. Seattle, WA: USEPA.

Kuhn, A., W. R. Munns, J. Serbst, P. Edwards, M. G. Cantwell, T. Gleason, M. C. Pelletier, and W. Berry. 2002. Evaluating the ecological significance of laboratory response data to predict population-level effects for the estuarine amphipod *Ampelisca abdita*. *Environmental Toxicology and Chemistry* 21: 865–874.

Long, E. R., and D. D. MacDonald. 1998. Recommended uses of empirically derived, sediment quality guidelines for marine and estuarine ecosystems. *Human and Ecological Risk Assessment* 4:1019–1039.

Lotufo, G. R., P. F. Landrum, and M. L. Gedeon. 2001. Toxicity and bioaccumulation of DDT in freshwater amphipods in exposures to spiked sediments. *Environmental Toxicology Chemistry* 20: 810-825.

MacDonald, D. D., L. M. Dipinto, J. Field, C. G. Ingersoll, E. R. Long, and R. C. Swartz. 2000. Development and evaluation of consensus-based sediment effect concentrations for polychlorinated biphenyls. *Environmental Toxicology and Chemistry* 19: 1403–1413.

McGee, B. L., D. J. Fisher, D. A. Wright, L. T. Yonkos, G. P. Ziegler, S. D. Turley, J. D. Farrar, D. W. Moore, and T. S. Bridges. 2004. A field test and comparison of acute and chronic sediment toxicity with the marine amphipod *Leptocheirus plumulosus* in Chesapeake Bay, USA. *Environmental Toxicology and Chemistry* 23: 1751–1761.

Mearns, A. J., R. C. Swartz, J. M. Cummins, P. A. Dinnel, P. Plesha, and P. M. Chapman. 1986. Inter-laboratory comparison of a sediment toxicity test using the marine amphipod, *Rhepoxynius abronius*. *Marine Environmental Research* 19:13-37.

Munns, W. R., W. J. Berry, and T. H. Dewitt. 2002. Toxicity testing, risk assessment, and options for dredged material management. *Marine Pollution Bulletin* 44: 294–302.

National Oceanic and Atmospheric Administration (NOAA). 1999. *Screening quick reference tables*. Silver Spring, MD: National Oceanic and Atmospheric Administration Coastal Protection and Restoration Division, Office of Response and Restoration.

O'Connor, T. P., K. D. Daskalaskis, J. L. Hyland, J. F. Paul, and J. K. Summers. 1998. Comparisons of sediment toxicity with predictions based on chemical guidelines. *Environmental Toxicology and Chemistry* 17:468–471.

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

Puget Sound Dredged Disposal Analysis (PSSDA). Issue Paper. 2002. *Implementation of the Neanthes 20-day sediment bioassay*. Olympia, WA: Washington State Department of Ecology.

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

Simini, M., R. T. Checkai, and M. E. Maly. 2000. *Tri-service remedial project manager's handbook for ecological risk assessment.* SFIM-AEC-ER-CR-200015. Aberdeen Proving Ground, MD: U.S. Army Environmental Center.

Swartz, R. C., D. W. Schultz, R. J. Ozretich, and J. O. Lamberson. 1995. ΣPAH model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments. *Environmental Toxicology and Chemistry* 14: 1977–1987.

Swartz, R. C. 1999. Consensus sediment quality guidelines for polycyclic aromatic hydrocarbon mixtures. *Environmental Toxicology and Chemistry* 18: 780–787.

U.S. Environmental Protection Agency (USEPA) / U.S. Army Corps of Engineers (USACE). 1991. *Evaluation of dredged material proposed for ocean disposal.* EPA 503/8-91/001. Washington, DC: USEPA.

USEPA/USACE. 1994. *Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods*. EPA/600/R-94/025. Washington, DC: USEPA.

USEPA/USACE. 1998. *Evaluation of material proposed for discharge to waters of the U.S. — Testing manual (inland testing manual)*. EPA/823/B-98/004. Washington, DC: USEPA.

USEPA. 2000. *Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates*. EPA/600/R-99/064. Washington, DC: USEPA.

USEPA/USACE. 2001. *Methods for assessing the chronic toxicity of marine and estuarine sediment-associated contaminants with the amphipod* Leptocheirus plumulosus. EPA/600/R-01/020. Washington, DC: USEPA.

USEPA. 2001. *Methods for collection, storage, and manipulation of sediments for chemical and toxicological analyses: Technical manual.* EPA-823-B-01-002. Washington, DC: USEPA.

USEPA. 2002. *Equilibrium partitioning sediment guidelines for the protection of benthic organisms: Metal mixtures (cadmium, copper, lead, silver, and zinc)*. Draft Document. EPA/822/R-02/045. Washington, DC: USEPA..

Vorhees, D. J., S. B. Kane Driscoll, K. A. von Stackelberg, and T. S. Bridges. 2002. An evaluation of sources of uncertainty in a dredged material assessment. *Human and Ecological Risk Assessment* 8: 369–389.

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

Zar, J. H. 1984. *Biostatistical Analysis*. 2nd Edition. Englewood Cliffs, NJ: Prentice-Hall.

Appendix A: Conditions for Conducting Sediment Toxicity Tests

Table A1. Conditions for conducting acute toxicity tests.

Test method	10-d *Ampelisca abdita*	10-d *Americamysis bahia*	10-d *Leptocheirus plumulosus*
Test type	Static non-renewal	Static non-renewal	Static non-renewal
Test duration	10 days	10 days	10 days
Temperature	20.0 ± 1.0°C	20.0 ± 1.0°C	25.0 ± 1.0°C
Salinity	28 ± 2 ppt	28 ± 2 ppt	20 ± 2 ppt
pH	7.8 ± 0.5	7.8 ± 0.5	7.8 ± 0.5
Light quality	Ambient Laboratory	Ambient Laboratory	Ambient Laboratory
Light intensity	500 – 1000 lux	500 – 1000 lux	500 – 1000 lux
Photoperiod	24:0 hr (light:dark)	16:8 hr (light:dark)	24:0 hr (light:dark)
Test chamber size	1L	1L	1L
Sediment volume (depth)	175 mL (2 cm)	175 mL (2 cm)	175 mL (2 cm)
Overlying water volume	Fill to 950 mL	Fill to 950 mL	Fill to 950 mL
Sediment settling time	Overnight	Overnight	Overnight
Water renewal	None	None	None
Age of test organisms	Immature (3 – 5 mm)	Juvenile (3 – 5 d)	Neonates (0.5 – 0.75 mm)
Organisms/chamber	20	20	20
Replicates/treatment	5	5	5
Organisms/treatment	100	100	100
Feeding regime	None	~150 *Artemia* daily	None
Test chamber cleaning	None	None	None
Test solution aeration	>40% O_2 saturation	>40% O_2 saturation	>40% O_2 saturation
Dilution water	28 ppt	28 ppt	20 ppt
Dilution series	None	None	None
Endpoint(s)	Survival	Survival	Survival

Table A2. Conditions for conducting chronic toxicity tests.

Test method	28-d *L. plumulosus*	28-d *N. arenaceodentata*	20-d *N. arenaceodentata*
Test type	Static non-renewal	Static non-renewal	Static non-renewal
Test duration	28 days	28 days	20 days
Temperature	25.0 ± 1.0°C	20.0 ± 1.0°C	20.0 ± 1.0°C
Salinity	20 ± 2 ppt	30 ± 2 ppt	28 ± 2 ppt
pH	7.8 ± 0.5	7.8 ± 0.5	7.8 ± 0.5
Light quality	Ambient Laboratory	Ambient Laboratory	Ambient Laboratory
Light intensity	500 – 1000 lux	500 – 1000 lux	500 – 1000 lux
Photoperiod	16:8 hr (light:dark)	16:8 hr (light:dark)	16:8 hr (light:dark)
Test chamber size	1L	300 mL	1L
Sediment volume (depth)	175 mL (2 cm)	75 mL (2 cm)	175 mL (2 cm)
Overlying water volume	Fill to 950 mL	Fill to 150 mL	Fill to 950 mL
Sediment settling time	Overnight	Overnight	Overnight
Water renewal	50% (M, W, F)	50% once per week	33% once every 3 days
Age of test organisms	Neonates (0.25-0.6 mm)	≤7 day old juveniles	2-3 week old juveniles
Organisms/chamber	20	1	5
Replicates/treatment	5	10	5
Organisms/treatment	100	10	25
Feeding regime	Three times weekly: Days 0-14: 20 mg Tetramin®/beaker Days 15-28: 40 mg Tetramin®/beaker	Tuesdays: 2 mg Tetramarin®/beaker Fridays: 2 mg of Tetramarin® and 1 mg of alfalfa per beaker	40 mg Tetramarin®/ beaker every other day[1]
Test chamber cleaning	None	None	None
Test solution aeration	Trickle flow (>40%)	Trickle flow (>40%)	Trickle flow (>40%)
Dilution water	20 ppt	28 ppt	20 ppt
Dilution series	None	None	None
Endpoint(s)	Survival, Growth, Reproduction	Survival, Growth	Survival, Growth

[1] The modified 20-d test used the same feeding regime as the 28-d *N. arenaceodentata*.

Appendix B: Sediment Chemistry and Composition

Table B1. Polycyclic aromatic hydrocarbons (PAH) concentrations (µg/kg) in the Preliminary Evaluation. Values surpassing the effects range medium (ER-M) and probable effects level (PEL) are indicated by underlining and asterisks, respectively.

Chemical	Reference	Arthur	Buttermilk	Chester	Flushing	Hudson	Jamaica	Newark	Perth	Red Hook	ER-M	PEL
NAPHTH	1.6	213	347	75.4	213	334	37.3	229	112	169	2100	391
ACENAY	<3.9	17.9	45.3	11.1	67.9	30.7	<8.9	18.9	11.1	25.3	640	128
ACENAP	<3.9	54.6	127*	66.2	130*	118*	9.8	869	21.3	48	500	88.9
FLUORE	<3.9	92.3	172	148	174	198	24	683	41.7	57	5100	1494
PHENAN	<3.9	570*	1020*	991*	996*	636*	135	2630	235	413	1500	544
ANTRAC	<3.9	238	539*	966*	642*	394*	59.6	545*	119	242	1100	245
FLANTHE	<3.9	1480	2120*	4040*	2490*	1500*	335	3670*	543	871	5100	1494
PYRENE	<3.9	1722*	2200*	3840	3340	1490*	308	3110	610	1110	2600	1398
CHRYSE	<3.9	954*	1180*	1520*	1740*	797*	161	1010*	364	614	2800	846
BAANTHR	<3.9	697*	1070*	1310*	1580*	716*	150	959*	292	649	1600	693
BBFLANT	<3.9	970	952	1060	1470	603	172	832	333	510	NA	NA
BKFLANT	<3.9	803	760	775	1330	565	138	650	302	458	NA	NA
BAPYRE	<3.9	893*	1160*	829*	1870	720	159	775*	373	746	1600	763
I123PYR	<3.9	690	648	527	1140	415	115	469	271	363	NA	NA
DBAHANT	<3.9	133	149*	111	260	96	23.1	101	57	87.1	260	135
B-GHI-PY	<3.9	627	557	438	1030	358	102	436	253	319	NA	NA
2MeNAPH	<3.9	140	268*	66.8	229*	217*	40.9	135	76.5	133	670	201

Table B2. Polychlorinated biphenyls (PCB) concentrations (μg/kg) in the preliminary evaluation. Values surpassing the effects range medium (ER-M) and probable effects level (PEL) are indicated by underlining and asterisks, respectively.

Chemical	Reference	Arthur	Buttermilk	Chester	Flushing	Hudson	Jamaica	Newark	Perth	Red Hook
14DClB	<0.39	81.6	38.6	40.5	74.0	91.0	24.2	69.1	42.1	32.9
PCB 8	<0.39	6.7	9.8	4.2	11.9	31.3	3.5	8.2	4.8	5.4
PCB 15	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PCB 18	<0.39	7.7	12.1	7.1	12.6	63.0	2.6	12.0	5.1	6.9
PCB 28	<0.39	21.0	35.9	11.5	39.9	113.0	8.5	28.0	14.4	16.2
PCB 44	<0.39	16.6	11.5	10.5	18.1	60.7	3.0	18.6	8.1	7.0
PCB 49	<0.39	15.0	16.4	9.2	15.4	55.6	3.3	18.7	8.2	8.3
PCB 52	<0.39	23.8	20.6	14.8	30.6	73.5	<0.52	31.2	11.9	9.9
PCB 87	<0.39	9.2	5.9	6.1	11.5	25.6	1.4	10.3	3.9	2.5
PCB 101	<0.39	13.3	9.8	9.0	17.9	36.4	2.9	16.1	7.8	5.3
PCB 105	<0.39	6.8	5.0	5.0	8.6	18.3	1.8	8.0	3.8	2.0
PCB 118	<0.39	12.7	10.9	8.4	15.6	32.6	3.5	14.5	7.9	5.4
PCB 128	<0.39	3.2	2.9	2.4	4.3	8.5	1.1	3.5	2.1	1.5
PCB 138	<0.39	17.1	12.3	9.7	23.4	38.7	5.4	17.4	8.8	5.5
PCB 153	<0.39	NR	NR	NR	NR	NR	NR	NR	NR	NR
PCB 170	<0.39	4.3	2.4	1.9	5.4	8.3	1.1	4.2	2.0	1.3
PCB 180	<0.39	10.3	7.0	4.1	12.1	16.6	2.4	9.3	4.3	2.7
PCB 183	<0.39	2.6	1.8	1.8	3.8	6.4	1.0	3.0	1.4	0.8
PCB 187	<0.39	NR	NR	NR	NR	NR	NR	NR	NR	NR
PCB 195	<0.39	1.2	0.6	0.9	1.9	2.5	0.7	1.4	0.7	0.4
PCB 206	<0.39	1.8	2.0	1.3	3.7	5.0	<0.52	3.2	1.0	0.7
PCB 209	<0.39	<0.57	1.6	0.6	2.4	2.5	0.4	<0.46	<0.49	0.7
PCB 66	<0.39	23.7	22.3	11.8	27.2	64.8	6.4	24.3	14.2	10.0
PCB 184	<0.39	<0.57	<0.56	<0.36	<0.59	<0.44	<0.52	<0.46	<0.49	<0.51
PCB 81	<0.39	<0.57	<0.56	<0.36	<0.59	<0.44	<0.52	<0.46	<0.49	<0.51
PCB77 C	<0.39	3.1	3.7	1.5	3.9	9.4	1.6	5.1	2.2	1.7
PCB 123	<0.39	<0.57	<0.56	<0.36	<0.59	<0.44	<0.52	<0.46	<0.49	<0.51
PCB114	<0.39	<0.57	<0.56	<0.36	<0.59	<0.44	<0.52	<0.46	<0.49	<0.51
PCB 126	<0.39	<0.57	<0.56	<0.36	<0.59	<0.44	<0.52	<0.46	<0.49	<0.51
PCB 167 C	<0.39	4.0	0.6	0.6	1.2	1.8	<0.52	0.8	<0.49	0.28J
PCB 156	<0.39	NR	NR	NR	NR	NR	NR	NR	NR	NR
PCB 157	<0.39	NR	NR	NR	NR	NR	NR	NR	NR	NR
PCB 169	<0.39	<0.57	<0.56	<0.36	<0.59	<0.44	<0.52	<0.46	<0.49	<0.51
PCB 189	<0.39	<0.57	<0.56	<0.36	0.50J	0.5	<0.52	0.46J	<0.49	<0.51
Σ PCB[1]	0.0	285.7	233.7	162.9	345.4	766	74.8	306.9	154.7	127.1

[1] The ER-M for total PCBs = 180 μg/kg.

Table B3. Pesticides concentrations (µg/kg) in the preliminary evaluation. Values surpassing the effects range medium (ER-M) and probable effects level (PEL) are indicated by underlining and asterisks, respectively.

	Reference	Arthur	Buttermilk	Chester	Flushing	Hudson	Jamaica	Newark	Perth	Red Hook	ER-M	PEL
ALDRIN	<0.50	<1.23	<1.16	<0.75	<1.23	<0.93	<1.09	<0.95	<1.05	<1.05	NA	NA
A-BHC	<0.50	2.25*	2.36*	1.5*	<1.23	2.52*	<1.09	1.81*	1.12*	1.12*	NA	0.99
B-BHC	<0.50	<1.23	<1.16	<0.75	<1.23	<0.93	<1.09	<0.95	<1.05	<1.05	NA	0.99
G-BHC	<0.50	<1.23	<1.16	<0.75	<1.23	<0.93	<1.09	<0.95	<1.05	<1.05	NA	0.99
D-BHC	<0.50	<1.23	<1.16	<0.75	<1.23	<0.93	<1.09	<0.95	<1.05	<1.05	NA	0.99
PPDDD	<1.00	204	8.16	6.75	22.4	38.6	<2.18	22.1	23.2	23.2	27	374
PPDDE	<1.00	73.1	7.21	13.8	17.1	40.7	<2.18	48.4	7.48	7.48	27	374
PPDDT	<1.00	203	<2.31	<1.51	<2.46	<1.86	<2.18	<1.91	<2.10	<2.10	27	4.77
OP-DDE	<1.00	30.6	<2.31	<1.51	<2.46	<1.86	<2.18	7.75	<2.10	<2.10	27	374
OP-DDD	<1.00	44.1	<2.31	<1.51	<2.46	<1.86	<2.18	<1.91	<2.10	<2.10	270	7.81
OP-DDT	<1.00	<2.46	<2.31	<1.51	<2.46	<1.86	<2.18	<1.91	<2.10	<2.10	270	4.77
Total DDT	0	554.8	15.37	20.55	39.5	79.3	0	78.25	30.68	30.68	NA	NA
HPTCL	<0.50	1.18	<1.16	1.57	1.67	<0.93	0.81	<0.95	<1.05	<1.05	NA	NA
DIELDRIN	<1.00	<2.46	<2.31	<1.51	<2.46	<1.86	<2.18	<1.91	<2.10	<2.10	NA	4.3
ENDOI	<0.50	<1.23	<1.16	<0.75	<1.23	<0.93	<1.09	<0.95	<1.05	<1.05	NA	NA
ENDOII	<1.00	<2.46	<2.31	<1.51	<2.46	<1.86	<2.18	0.67	<2.10	<2.10	NA	NA
ENDOSU	<1.00	<2.46	<2.31	<1.51	<2.46	<1.86	<2.18	<1.91	<2.10	<2.10	NA	NA
ENDRIN	<1.00	3.29	1.73	2.27	3.1	<0.93	<2.18	2.13	3.81	3.81	NA	NA
ENDALD	<1.00	<2.46	<2.31	5.57	<2.46	<1.86	<2.18	<1.91	<2.10	<2.10	NA	NA
HPTCLE	<0.50	<1.23	<1.16	<0.75	<1.23	<0.93	<1.09	<0.95	<1.05	<1.05	NA	NA
METOXYCL	<5.00	<12.3	<11.6	<7.50	<12.3	<9.30	<10.9	<9.50	<10.5	<10.5	NA	NA
TOXAPHEN	<10.0	<24.6	<23.1	<15.1	<24.6	<18.6	<21.8	<19.1	<21.0	<21.0	NA	NA
A-CHL	<0.50	10.29	1.31	2.19	10.6	2.94	0.7	3.61	3.75	3.75	NA	NA

Table B4. Dioxins and furans concentrations (pg/g) in the preliminary evaluation.

Dioxin	Reference	Arthur	Buttermilk	Chester	Flushing	Hudson	Jamaica	Newark	Perth	Red Hook
2,3,7,8-TCDD	ND	8.50	2.00	0.58	2.50	18.00	1.20	28.00	2.10	0.86
TOTAL TCDD	ND	55.00	34.00	11.00	31.00	34.00	20.00	69.00	22.00	8.80
1,2,3,7,8-PeCDD	ND	3.60	1.40	0.88	2.10	2.40	0.89	1.90	1.20	0.41
TOTAL PeCDD	ND	42.00	19.00	13.00	23.00	22.00	9.60	29.00	14.00	4.70
1,2,3,4,7,9-HxCDD	ND	3.50	1.60	1.40	2.40	2.70	1.10	2.20	1.40	0.52
1,2,3,6,7,8-HxCDD	ND	11.00	7.00	5.00	8.60	12.00	3.90	9.70	6.90	2.10
1,2,3,7,8,9-HxCDD	0.26	9.50	5.70	3.90	7.20	7.40	3.10	6.80	4.70	1.80
TOTAL HxCDD	0.87	120.00	77.00	53.00	79.00	110.00	44.00	120.00	65.00	21.00
1,2,3,4,6,7,8-HpCDD	1.20	180.00	110.00	89.00	140.00	170.00	58.00	230.00	75.00	29.00
TOTAL HpCDD	2.20	410.00	260.00	210.00	290.00	390.00	140.00	920.00	170.00	65.00
OCDD	4.30	2300.00	990.00	800.00	1200.00	1200.00	520.00	1900.00	1100.00	260.00
2,3,7,8-TCDF	ND	15.00	8.90	6.70	9.20	11.00	6.40	9.20	7.10	2.90
TOTAL TCDF	0.30	200.00	120.00	82.00	140.00	220.00	44.00	160.00	79.00	39.00
1,2,3,7,8,-PeCDF	ND	7.80	3.30	2.80	4.00	5.60	3.50	4.30	3.40	1.00
2,3,4,7,8-PeCDF	ND	9.50	4.90	3.10	5.40	6.40	2.30	6.30	4.00	1.30
TOTAL PeCDF	ND	170.00	77.00	70.00	140.00	150.00	33.00	130.00	54.00	24.00
1,2,3,4,7,8-HxCDF	ND	25.00	6.40	5.70	9.30	17.00	3.50	25.00	7.40	2.00
1,2,3,6,7,8-HxCDF	ND	15.00	4.70	3.20	9.50	12.00	1.70	4.90	3.90	1.60
2,3,4,6,7,8-HxCDF	ND	6.90	2.40	2.40	3.30	3.40	1.30	3.90	2.40	0.73
1,2,3,7,8,9-HxCDF	ND	1.20	ND	ND	ND	ND	ND	ND	ND	0.23
TOTAL HxCDF	ND	140.00	56.00	50.00	98.00	110.00	23.00	110.00	43.00	17.00
1,2,3,4,6,7,8-HpCDF	0.53	98.00	41.00	28.00	50.00	94.00	18.00	85.00	30.00	10.00
1,2,3,4,7,8,9-HpCDF	0.33	10.00	3.00	1.70	3.20	4.20	ND	5.90	3.40	1.20
TOTAL HpCDF	0.86	190.00	78.00	57.00	97.00	160.00	30.00	170.00	57.00	22.00
OCDF	1.10	180.00	52.00	34.00	62.00	74.00	17.00	110.00	42.00	17.00

Table B5. Metals concentrations (mg/kg) in the preliminary evaluation. Values surpassing the effects range medium (ER-M) and probable effects level (PEL) are indicated by underlining and asterisks, respectively.

Metal	Reference	Arthur	Buttermilk	Chester	Flushing	Hudson	Jamaica	Newark	Perth	Red Hook	ER-M	PEL
Hg	0.0075	3.6	1	0.41	1.9	2	0.54	2.5	1.8	0.53	0.71	0.696
Me-Hg	<0.05	9.48	0.86	4.34	0.37	3.51	0.54	0.79	6.41	9.31	NA	NA
Sb	<0.100	1.05	0.59	0.53	0.58	1.75	0.34	0.75	0.84	0.23	NA	NA
As	4.02	20.5	11.5	6.52	11.6	13.9	7.04	12.4	16.3	6.21	70	41.6
Be	0.108	1.01	0.828	0.528	0.989	0.888	0.599	0.929	1.02	0.421	NA	NA
Cd	<0.020	1.9	0.706	1.59	3.01	2.71	0.699	1.97	0.943	0.544	9.6	4.21
Cr	6.1	118	77	59.8	122	141	58.2	116	94.5	38.2	370	160
Cu	2.7	269*	111*	107	251*	178*	67.6	177*	160*	52.7	270	108
Pb	7.3	182*	118*	119*	196*	140*	74.6	134*	135*	72.2	218	112
Ni	2.29	36.3	31.5	30.3	46*	34.8	20.9	34.8	31.9	19.1	51.6	42.8
Se	<0.200	2.21	1.14	0.827	1.43	0.967	1.01	0.929	1.61	0.669	NA	NA
Ag	<0.100	3.67*	3.7	1.67	4.44	6.1	1.61	2.65*	3.19*	1.62	3.7	1.77
Tl	<0.200	0.27	0.21	0.219	0.38	0.209	0.2	0.24	0.248	<0.200	NA	NA
Zn	11.7	289*	179	240	337*	233	125	252	231	94.3	410	271

Table B6. Alkyltins concentrations (μg/kg) in the preliminary evaluation.

Alkyltin	Reference	Arthur	Buttermilk	Chester	Flushing	Hudson	Jamaica	Newark	Perth	Red Hook	PSDDA[1]
Tetrabutyltin	2.0	5.7	5.3	2.9	5.2	4.0	4.8	4.0	4.9	2.7	NA
Tributyltin	1.8	16	4.7	2.5	6.2	3.5	4.3	52	4.3	2.4	NA
Dibutyltin	1.5	11	4.0	2.2	5.1	3.0	3.7	9.0	3.7	2.1	NA
Monobutyltin	1.2	3.3	3.1	1.7	3.0	2.3	2.8	2.4	2.8	1.6	NA
Total	6.5	36	17.1	9.3	19.5	12.8	15.6	67.4	15.7	8.8	73
[1] PSDDA (1996)											

Table B7. Polycyclic aromatic hydrocarbons (PAH) concentrations (μg/kg) for sediments used in interlaboratory evaluation. Values surpassing the effects range medium (ER-M) and probable effects level (PEL) are indicated by underlining and asterisks, respectively.

Chemical	Hudson	Chester	Red Hook	100% Newark	75% Newark	50% Newark	25% Newark	ER-M	PEL
NAPHTH	290	59.8	156	148	150	77.4	37.4	2100	391
ACENAY	20.2	7.36	21.7	13.5	11	5.68	<11	640	128
ACENAP	78.3	41.9	42	483*	411*	234*	110*	500	88.9
FLUORE	131	201	58.1	331	232	107	52.8	5100	1494
PHENAN	497	1460*	458	929*	644*	253	133	1500	544
ANTRAC	243	685*	281*	348*	274*	120	60.5	1100	245
FLANTHE	1030	3430*	989	1880*	1760*	913	491	5100	1494
PYRENE	950	3350	1100	1700*	1390	734	438	2600	1398
CHRYSE	555	1530*	686	636	595	250	141	2800	846
BAANTHR	448	933*	642	554	500	224	125	1600	693
BBFLANT	634	1020	713	730	688	331	174	NA	NA
BKFLANT	213	375	255	247	214	105	62.7	NA	NA
BAPYRE	476	605	649	471	435	205	105	1600	763
I123PYR	353	381	363	306	298	143	75.9	NA	NA
DBAHANT	74.9	82.4	81.7	68.8	59.8	29.4	14.3	260	135
B-GHI-PY	346	446	333	317	282	138	79.2	NA	NA
2MeNAPH	162	57.1	111	85.4	74.4	42.8	25.3	670	201

Table B8. Polychlorinated biphenyls (PCB) concentrations (μg/kg) for sediments used in interlaboratory evaluation.

Chemical	Hudson	Chester	Red Hook	100% Newark	75% Newark	50% Newark	25% Newark
14DCIB	106.00	26.10	18.90	31.20	32.40	26.60	12.90
PCB 8	21.40	2.19	3.48	5.52	4.36	3.14	<0.84
PCB 15	NA	NA	NA	NA	NA	NA	NA
PCB 18	78.60	9.27	8.20	17.50	13.40	8.82	4.73
PCB 28	80.60	8.37	12.10	22.20	15.90	11.20	5.94
PCB 44	58.80	8.58	5.45	16.50	13.40	10.20	5.69
PCB 49	59.70	7.64	6.81	18.60	14.70	10.30	7.20
PCB 52	74.90	12.40	8.99	24.30	20.10	14.20	7.76
PCB 87	16.50	4.26	1.58	<0.66	<0.71	<0.79	<0.84
PCB 101	21.50	10.20	5.34	17.60	13.50	10.70	6.43
PCB 105	16.30	4.23	1.89	6.11	5.29	3.80	2.24
PCB 118	48.80	10.60	5.71	20.00	14.40	9.52	3.72
PCB 128	3.63	1.95	1.03	2.64	2.15	1.71	1.04
PCB 138	50.70	11.90	5.75	22.80	17.10	12.30	6.28
PCB 153	NR	NR	NR	NR	NR	NR	NR
PCB 170	10.00	2.41	1.14	4.57	3.43	2.77	1.22
PCB 180	9.86	4.28	2.27	7.46	6.49	4.85	3.64
PCB 183	2.99	20.30	1.11	3.54	2.86	1.93	0.98
PCB 187	NR	NR	NR	NR	NR	NR	NR
PCB 195	2.57	0.55	0.35	1.00	0.88	0.58	<0.84
PCB 206	5.49	1.75	0.87	3.02	2.11	1.44	0.75
PCB 209	4.84	1.63	0.96	2.42	2.12	1.73	0.94
PCB 66	41.20	12.70	9.06	22.50	21.20	15.00	9.49
PCB 184	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	<0.84
PCB 81	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	<0.84
PCB77 C	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	<0.84
PCB 123	0.95	<0.51	<0.47	0.41	0.29	<0.79	<0.84
PCB114	1.76	<0.51	<0.47	0.61	0.59	<0.79	<0.84
PCB 126	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	<0.84
PCB 167 C	2.97	0.51	0.35	<0.66	0.87	0.59	0.41
PCB 156	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	<0.84
PCB 157	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	0.42
PCB 169	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	<0.84
PCB 189	<0.66	<0.51	<0.47	<0.66	<0.71	<0.79	<0.84
Σ PCB[1]	720.1	161.8	101.3	250.5	207.5	151.4	81.8

[1] The ER-M for total PCBs = 180 μg/kg.

Table B9. Pesticides concentrations (μg/kg) for sediments used in interlaboratory evaluation. Values surpassing the effects range medium (ER-M) and probable effects level (PEL) are indicated by underlining and asterisks, respectively.

Chemical	Hudson	Chester	Red Hook	100% Newark	75% Newark	50% Newark	25% Newark	ER-M	PEL
ALDRIN	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	NA
A-BHC	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	0.99
B-BHC	<1.43	<1.15	3.65	<1.30	<1.45	<1.30	<1.40	NA	0.99
G-BHC	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	0.99
D-BHC	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	0.99
PPDDD	37	6.82	3.72	15.8	18	7.41	5.49	27	374
PPDDE	34.8	13	5.19	37.3	33.5	17.4	10.2	27	374
PPDDT	45.6*	18.2*	9.39*	23.4*	19.3*	10.1*	6.0*	27	4.77
OP-DDE	30.1	15.3	8.84	27.3	24.5	13.9	8.31	27	374
OP-DDD	<1.43	<1.15	3.72	15.8*	<1.45	<1.30	1.69	270	7.81
OP-DDT	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	270	4.77
Total DDT	147.5*	53.3*	30.86	119.6*	95.3*	48.81	31.65	46.1	51.7
HPTCL	<1.43	<1.15	<1.18	<1.30	<1.45	1.16	0.77	NA	NA
DIELDRIN	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	4.3
ENDOI	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	NA
ENDOII	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	NA
ENDOSU	7.73	<1.15	<1.18	6.43	5.1	<1.30	<1.40	NA	NA
ENDRIN	<1.43	2.26	1.24	2	1.59	0.96	<1.40	NA	NA
ENDALD	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	NA
HPTCLE	<1.43	<1.15	<1.18	9.33	9.05	5.22	3.1	NA	NA
METOXYCL	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	NA
A-CHL	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	NA
TRANSNON	<1.43	<1.15	<1.18	<1.30	<1.45	<1.30	<1.40	NA	NA

Table B10. Dioxins and furans concentrations (pg/g) for sediments used in interlaboratory evaluation.

Chemical	Hudson	Chester	Red Hook	100% Newark	75% Newark	50% Newark	25% Newark	ER-M	PEL
2,3,7,8-TCDD	42	ND	1	7	4	2	1	NA	NA
TOTAL TCDD	65	2	9	16	10	3	2	NA	NA
1,2,3,7,8-PeCDD	3	0	1	1	ND	ND	ND	NA	NA
TOTAL PeCDD	33	3	5	13	6	1	1	NA	NA
1,2,3,4,7,9-HxCDD	3	1	1	1	0	ND	ND	NA	NA
1,2,3,6,7,8-HxCDD	16	2	2	7	2	1	1	NA	NA
1,2,3,7,8,9-HxCDD	9	2	2	7	1	ND	ND	NA	NA
TOTAL HxCDD	150	17	21	180	25	16	8	NA	NA
1,2,3,4,6,7,8-HpCDD	260	25	30	350	47	29	17	NA	NA
TOTAL HpCDD	580	75	74	3300	220	95	100	NA	NA
OCDD	1700	220	280	2700	330	220	120	NA	NA
2,3,7,8-TCDF	40	3	6	4	5	2	2	NA	NA
TOTAL TCDF	330	29	47	77	47	18	9	NA	NA
1,2,3,7,8,-PeCDF	7	1	1	2	1	0	0	NA	NA
2,3,4,7,8-PeCDF	8	1	1	3	2	ND	1	NA	NA
TOTAL PeCDF	270	21	27	67	35	16	7	NA	NA
1,2,3,4,7,8-HxCDF	32	3	3	13	5	2	1	NA	NA
1,2,3,6,7,8-HxCDF	10	1	1	2	1	ND	0	NA	NA
2,3,4,6,7,8-HxCDF	5	1	1	1	1	ND	ND	NA	NA
1,2,3,7,8,9-HxCDF	ND	ND	ND	ND	ND	ND	ND	NA	NA
TOTAL HxCDF	190	18	18	76	24	11	5	NA	NA
1,2,3,4,6,7,8-HpCDF	120	8	11	47	15	10	4	NA	NA
1,2,3,4,7,8,9-HpCDF	5	1	1	3	1	ND	ND	NA	NA
TOTAL HpCDF	220	18	28	160	34	18	10	NA	NA
OCDF	130	8	10	90	23	20	8	NA	NA

Table B11. Metals concentrations (mg/kg) for sediments used in interlaboratory evaluation. Values surpassing the effects range medium (ER-M) and probable effects level (PEL) are indicated by underlining and asterisks, respectively.

Chemical	Hudson	Chester	Red Hook	100% Newark	75% Newark	50% Newark	25% Newark	ER-M	PEL
Hg	2.46	0.532	1.14	2.97	2.28	1.56	0.907	0.71	0.696
Me-Hg	0.0431	0.054	0.0602	0.0432	0.0393	0.0363	0.0335	NA	NA
Sb	0.81	0.5	0.35	0.9	0.8	0.63	0.51	NA	NA
As	13.5	6.09	7.08	11.8	10.4	9.53	8.05	70	41.6
Be	0.957	0.502	0.499	1.09	0.949	0.726	0.569	NA	NA
Cd	2.72	1.53	0.586	2.05	1.79	1.62	1.28	9.6	4.21
Cr	135	58	41.6	1.04	87.4	76.6	59	370	160
Cu	151*	88.6	43.1	144*	113*	85.4	60.6	270	108
Pb	130*	130*	61.6	122*	94.8	72.3	41.5	218	112
Ni	37.2	30.7	22.5	36.9	37.2	36.2	34.9	51.6	42.8
SE	0.296	0.333	<0.200	0.655	0.418	0.419	0.44	NA	NA
Ag	0.564	1.59	1.67	2.3*	1.84*	1.34	0.793	3.7	1.77
Tl	0.214	<0.200	<0.200	0.254	0.245	0.236	0.234	NA	NA
Zn	238	189	116	278*	223	178	135	410	271

Appendix C: Overlying Water Quality Summary

Table C1. Mean overlying water quality data for the 10-d *Ampelisca abdita* tests conducted in the interlaboratory comparison. Ranges are provided in parentheses.

Treatment	Lab	Temp. (°C)	Salinity (ppt)	pH (SU)	D.O. (mg/L)
Control	A	19.6 (19.1 - 20.0)	29 (28 - 29)	8.06 (7.98 - 8.14)	8.4 (7.8 - 8.7)
	B	20.8 (20.7 - 20.9)	27 (27 - 28)	8.04 (7.80 - 8.20)	7.1 (6.9 - 7.2)
	C	19.8 (19.2 - 20.4)	29 (28 - 29)	8.02 (7.87 - 8.26)	7.3 (7.1 - 7.4)
Reference	B	20.8 (20.7 - 20.9)	27 (27 - 28)	7.91 (7.90 - 8.00)	7.2 (6.9 - 7.5)
	C	19.8 (19.2 - 20.3)	29 (28 - 29)	8.05 (7.93 - 8.22)	7.3 (7.0 -7.5)
Hudson	A	19.6 (19.1 - 20.0)	28 (28 - 28)	8.12 (8.01 - 8.27)	8.6 (8.5 - 8.8)
	B	20.8 (20.7 - 20.9)	27 (27 - 27)	7.97 (7.80 - 8.10)	7.1 (6.8 - 7.3)
	C	19.8 (19.2 - 20.4)	28 (28 - 28)	8.04 (7.80 - 8.30)	7.2 (6.5 - 7.4)
Chester	A	19.6 (19.1- 20.0)	29 (28 - 29)	8.26 (7.96 - 8.53)	8.4 (7.2 - 8.8)
	B	20.8 (20.7 - 20.9)	27 (27 - 27)	8.28 (7.80 - 8.80)	7.0 (6.7 - 7.3)
	C	19.8 (19.2 - 20.4)	28 (28 - 29)	8.17 (7.86 - 8.55)	7.2 (6.6 - 7.4)
Red Hook	A	19.6 (19.1 - 20.1)	29 (28 - 29)	8.19 (8.00 - 8.38)	8.7 (8.6 - 8.8)
	B	20.8 (20.7 - 20.9)	27 (27 - 27)	8.08 (7.80 - 8.40)	7.0 (6.7 - 7.3)
	C	19.8 (19.2 - 20.4)	28 (28 - 29)	8.17 (7 92 - 8.63)	7.3 (7.2 - 7.5)
100% Newark	A	19.6 (19.0 - 20.2)	28 (27 - 29)	8.14 (7.97 - 8.29)	8.5 (8.4 - 8.8)
	B	20.8 (20.7 - 20.9)	27 (27 - 27)	7.99 (7.70 - 8.20)	7.0 (6.6 - 7.3)
	C	19.8 (19.3 - 20.4)	28 (28 - 28)	8.15 (7.81 - 8.52)	7.3 (7.0 - 7.5)
75% Newark	A	19.6 (19.0 - 20.1)	28 (27 - 28)	8.16 (8.02 - 8.33)	8.4 (7.8 - 8.6)
	B	20.8 (20.7 - 20.9)	27 (27 - 27)	8.07 (7.70 - 8.40)	7.0 (6.8 - 7.3)
	C	19.8 (19.3 - 20.3)	28 (28 - 29)	8.14 (7.85 - 8.50)	7.3 (7.2 - 7.4)
50% Newark	A	19.6 (19.1- 20.1)	31 (28 - 33)	8.17 (8.06 - 8.27)	8.6 (8.6 - 8.8)
	B	20.8 (20.7 - 20.9)	27 (27 - 28)	8.08 (7.80 - 8.40)	7.0 (6.7 - 7.2)
	C	19.9 (19.2 - 20.4)	28 (28 - 29)	8.14 (7.87 - 8.44)	7.3 (7.1 - 7.7)
25% Newark	A	19.6 (19.1- 20.1)	28 (28 - 28)	8.16 (8.05 - 8.28)	8.7 (8.4 - 8.8)
	B	20.8 (20.7 - 20.9)	27 (27 - 28)	8.10 (7.80 - 8.40)	7.0 (6.8 - 7.2)
	C	19.8 (19.2 - 20.3)	28 (28 - 29)	8.12 (7.86 - 8.35)	7.3 (7.1 - 7.5)

Table C2. Mean overlying water quality data for the 28-d *Leptocheirus plumulosus* tests conducted in the interlaboratory comparison. Ranges are provided in parentheses.

Treatment	Lab	Temp. (°C)	Salinity (ppt)	pH (SU)	D.O. (mg/L)
Control	A	22.8 (17.9[1] - 24.1)	20 (20 - 20)	8.02 (7.89 - 8.21)	7.1 (6.2 - 8.2)
	B	25.3 (23.8 - 26.4)	20 (18 - 20)	7.89 (7.70 - 8.00)	6.5 (5.9 - 7.2)
	C	24.9 (24.4 - 25.7)	21 (20 - 21)	7.77 (7.50 - 8.01)	6.5 (8.5 - 7.0)
Reference	A	22.8 (17.9† - 24.1)	20 (19 - 20)	8.08 (7.84 - 8.27)	7.2 (6.4 - 8.1)
	B	25.3 (23.9 - 26.4)	19 (18 - 20)	7.95 (7.70 - 8.10)	6.7 (6.3 - 7.1)
	C	25.1 (24.4 - 25.7)	21 (20 - 21)	7.81 (7.30 - 8.02)	6.6 (6.0 - 7.7)
Hudson	A	22.8 (17.9† - 24.1)	20 (19 - 21)	8.10 (7.84 - 8.31)	7.1 (6.2 - 8.2)
	B	25.2 (23.9 - 26.2)	19 (18 - 20)	7.98 (7.70 - 8.30)	6.5 (6.0 - 7.0)
	C	25.1 (24.5 - 25.7)	21 (20 - 21)	7.88 (7.41 - 8.36)	6.6 (6.0 - 7.1)
Chester	A	22.7 (17.9† - 24.2)	20 (19 - 20)	8.22 (7.27 - 8.60)	7.1 (6.3 - 8.0)
	B	25.3 (23.7 - 26.3)	19 (18 - 20)	8.15 (7.80 - 8.80)	6.6 (6.1 - 7.2)
	C	25.1 (24.6 - 25.7)	21 (20 - 22)	8.12 (7.74 - 8.95)	6.4 (5.2 - 6.9)
Red Hook	A	22.7 (17.9† - 24.2)	20 (20 - 21)	8.24 (7.93 - 8.68)	7.0 (6.2 - 8.2)
	B	25.2 (23.9 - 26.2)	19 (18 - 20)	8.12 (7.90 - 8.60)	6.5 (5.8 - 7.1)
	C	25.2 (24.6 - 25.7)	21 (21 - 21)	8.10 (7.72 - 8.85)	6.4 (5.5 - 7.0)
100% Newark	A	22.7 (17.9† - 24.1)	20 (19 - 21)	8.20 (7.94 - 8.51)	7.0 (6.2 - 8.6)
	B	25.2 (24.0 - 26.2)	19 (18 - 20)	8.90 (7.80 - 8.50)	6.5 (6.1 - 7.1)
	C	25.2 (24.6 - 25.7)	21 (20 - 21)	8.03 (7.58 - 8.56)	6.5 (5.4 - 7.5)
75% Newark	A	22.6 (17.9† - 24.3)	20 (20 - 21)	8.09 (7.12 - 8.47)	7.1 (6.1 - 8.4)
	B	25.3 (24.0 - 26.3)	19 (18 - 20)	8.10 (7.85 - 8.50)	6.5 (6.0 - 7.2)
	C	25.1 (24.6 - 25.7)	21 (21 - 22)	8.05 (7.70 - 8.57)	6.5 (5.8 - 7.8)
50% Newark	A	22.7 (17.9† - 24.2)	20 (19 - 20)	8.16 (7.85 - 8.44)	6.9 (6.3 - 7.8)
	B	25.2 (24.0 - 26.3)	19 (19 - 20)	8.07 (7.90 - 8.50)	6.5 (5.9 - 7.0)
	C	25.1 (24.2 - 25.8)	21 (21 - 22)	8.07 (7.77 - 8.66)	6.7 (6.2 - 7.9)
25% Newark	A	22.7 (17.9† - 24.2)	20 (20 - 20)	8.11 (7.81 - 8.38)	7.0 (6.3 - 8.0)
	B	25.3 (24.0 - 26.3)	19 (18 - 20)	8.04 (7.80 - 8.50)	6.5 (6.0 - 7.2)
	C	25.1 (24.0 - 25.7)	21 (20 - 21)	8.03 (7.60 - 8.56)	6.6 (6.0 - 7.9)

[1] Result of an equipment malfunction and was quickly rectified.

Table C3. Mean overlying water quality data for the 28-d *Neanthes arenaceodentata* tests conducted in the interlaboratory comparison. Ranges are provided in parentheses.

Treatment	Lab	Temp. (°C)	Salinity (ppt)	pH (SU)	D.O. (mg/L)
Control	A	19.2 (18.8 - 21.0)	32 (30 - 34)	8.13 (7.82 - 8.38)	7.6 (6.3 - 9.1)
	B	20.2 (19.5 - 21.2)	32 (29 - 33)	7.81 (7.50 - 8.10)	7.1 (6.6 - 7.4)
	C	19.7 (18.8 - 20.4)	30 (29 - 32)	7.93 (7.80 - 8.10)	7.3 (7.0 - 7.6)
Reference	A	19.2 (18.8 - 20.7)	33 (30 - 38)	8.16 (7.82 - 8.24)	7.9 (6.8 - 8.7)
	B	20.2 (19.5 - 21.1)	32 (29 - 34)	7.88 (7.70 - 8.00)	7.2 (7.0 - 7.5)
	C	19.7 (18.7 - 20.7)	30 (29 - 30)	7.93 (7.89 - 8.00)	7.3 (7.0 - 7.6)
Hudson	A	19.2 (18.7 - 20.9)	32 (30 - 35)	8.29 (8.10 - 8.46)	7.8 (6.8 - 8.6)
	B	20.1 (19.5 - 21.0)	31 (29 - 33)	7.88 (7.70 - 8.00)	7.1 (6.4 - 7.4)
	C	19.7 (18.8 - 20.3)	29 (28 - 31)	8.11 (8.00 - 8.20)	7.3 (7.1 - 7.6)
Chester	A	19.3 (18.8 - 20.9)	32 (30 - 34)	8.38 (8.13 - 8.65)	7.7 (6.7 - 8.5)
	B	20.2 (19.5 - 21.0)	31 (29 - 34)	8.14 (7.80 - 8.60)	7.1 (6.7 - 7.4)
	C	19.6 (18.8 - 20.4)	29 (28 - 30)	8.24 (8.10 - 8.40)	7.3 (7.1 - 7.6)
Red Hook	A	19.2 (18.8 - 20.7)	31 (30 - 33)	8.36 (7.95 - 8.79)	7.4 (5.0 - 8.6)
	B	20.2 (19.5 - 20.9)	32 (30 - 34)	8.19 (7.80 - 8.40)	7.2 (6.9 - 7.4)
	C	19.7 (18.8 - 20.4)	30 (28 - 32)	8.26 (8.00 - 8.50)	7.3 (7.1 - 7.5)
100% Newark	A	19.2 (18.7 - 20.8)	31 (30 - 35)	8.35 (8.00 - 8.60)	7.5 (5.9 - 8.5)
	B	20.1 (19.5 - 20.9)	31 (29 - 33)	8.12 (7.70 - 8.40)	7.1 (6.7 - 7.8)
	C	19.7 (18.8-20.5)	29 (28 - 31)	8.23 (8.10 - 8.40)	7.3 (6.9 - 7.6)
75% Newark	A	19.2 (18.7 - 20.8)	31 (30 - 34)	8.36 (8.10 - 8.59)	7.4 6.0 - 8.6)
	B	20.2 (19.5 - 20.8)	31 (29 - 33)	8.12 (7.80 - 8.40)	7.2 (7.0 - 7.8)
	C	19.7 (18.8 - 20.4)	30 (29 - 31)	8.25 (8.10 - 8.40)	7.3 (7.0 - 7.6)
50% Newark	A	19.2 (18.7 - 20.6)	31 (30 - 32)	8.36 (8.08 - 8.57)	7.5 (6.8 - 8.5)
	B	20.2 (19.5 - 21.0)	31 (30 - 33)	8.13 (7.80 - 8.40)	7.1 (6.7 - 7.4)
	C	19.7 (18.8 - 20.4)	30 (29 - 31)	8.23 (8.00 - 8.40)	7.1 (5.9 - 7.5)
25% Newark	A	19.2 (18.8 - 20.6)	31 (30 - 34)	8.36 (8.06 - 8.61)	7.6 (6.8 - 9.0)
	B	20.2 (19.5 - 21.2)	32 (29 - 33)	8.13 (7.80 - 8.4)	7.1 (6.9 - 7.3)
	C	19.7 (18.8 - 20.3)	30 (29 - 31)	8.22 (8.00 - 8.4)	7.2 (6.6 - 7.5)

Table C4. Mean overlying water quality data for the 20-d *Neanthes arenaceodentata* tests conducted in the interlaboratory comparison. Ranges are provided in parentheses.

Treatment	Lab	Temp. (°C)	Salinity (ppt)	pH (SU)	D.O. (mg/L)
Control	A	19.5 (18.9 - 20.2)	30 (28 - 30)	8.03 (7.85 - 8.20)	7.1 (5.0 - 9.1)
	B	20.1 (19.5 - 20.6)	28 (28 - 29)	7.71 (7.50 - 7.90)	6.9 (6.2 - 7.4)
	C	19.7 (19.4 - 20.4)	27 (26 - 28)	7.77 (7.34 - 8.14)	7.2 (5.2 - 8.5)
Reference	A	19.5 (18.8 - 20.3)	30 (28 - 30)	7.60 (7.89 - 8.23)	7.1 (4.3 - 9.2)
	B	20.1 (19.6 - 20.6)	28 (28 - 29)	7.72 (7.60 - 7.90)	7.0 (6.3 - 7.2)
	C	19.6 (19.2 - 20.4)	27 (26 - 28)	7.77 (7.56 - 7.96)	7.3 (6.2 - 8.7)
Hudson	A	19.5 (18.8 - 20.2)	30 (28 - 31)	8.08 (7.86 - 8.19)	7.1 (5.0 - 9.0)
	B	20.1 (19.6 - 20.6)	28 (28 - 29)	7.72 (7.50 - 7.90)	6.8 (6.3 - 7.1)
	C	19.9 (19.6 - 20.5)	27 (26 - 28)	7.86 (7.56 - 8.19)	7.3 (6.0 - 8.0)
Chester	A	19.5 (18.8 - 20.2)	30 (28 - 31)	8.21 (7.99 - 8.39)	7.1 (5.2 - 9.1)
	B	20.0 (19.5 - 20.5)	28 (28 - 29)	7.85 (7.60 - 8.10)	6.8 (5.9 - 7.1)
	C	19.7 (19.0 - 20.4)	27 (26 - 29)	7.96 (7.25 - 8.34)	7.3 (5.9 - 8.2)
Red Hook	A	19.5 (18.8 - 20.2)	30 (28 - 30)	8.11 (7.96 - 8.27)	7.0 (4.9 - 9.0)
	B	20.1 (19.5 - 20.6)	28 (28 - 29)	7.81 (7.60 - 8.20)	6.8 (6.0 - 7.1)
	C	19.7 (19.3 - 20.4)	27 (26 - 28)	7.88 (7.62 - 8.28)	7.1 (5.4 - 8.2)
100% Newark	A	19.5 (18.8 - 20.2)	30 (29 - 30)	7.65 (8.01 - 8.25)	7.1 (5.1 - 9.1)
	B	20.1 (19.5 - 20.6)	28 (28 - 29)	7.75 (7.60 - 8.00)	6.8 (6.2 - 7.0)
	C	19.6 (19.3 - 20.4)	27 (26 - 28)	7.88 (7.66 - 8.19)	7.2 (5.7 - 8.2)
75% Newark	A	19.5 (18.8 - 20.2)	30 (29 - 31)	8.09 (7.70 - 8.42)	7.0 (5.0 - 9.1)
	B	20.1 (19.6 - 20.6)	28 (28 - 29)	7.78 (7.50 - 8.00)	6.7 (5.8 - 7.1)
	C	19.8 (19.5 - 20.4)	27 (26 - 28)	7.98 (7.75 - 8.33)	7.2 (5.6 - 8.0)
50% Newark	A	19.5 (18.8 - 20.2)	30 (28 - 31)	8.18 (7.91 - 8.49)	7.1 (4.5 - 9.1)
	B	20.0 (19.5 - 20.6)	28 (28 - 29)	7.78 (7.60 - 8.00)	6.8 (6.1 - 7.2)
	C	19.9 (19.6 - 20.6)	27 (26 - 28)	7.98 (7.81 - 32)	7.3 (5.7 - 7.7)
25% Newark	A	19.5 (18.8 - 20.2)	30 (28 - 31)	8.13 (7.93 - 8.38)	7.1 (5.1 - 9.0)
	B	20.1 (19.6 - 20.6)	28 (28 - 29)	7.78 (7.50 - 7.90)	6.8 (6.2 - 7.0)
	C	19.7 (19.3 - 20.3)	27 (26 - 28)	7.93 (7.69 - 8.35)	7.2 (5.7 - 7.6)

Table C5. Mean overlying water quality data for the modified 20-d *Neanthes arenaceodentata* tests conducted in the interlaboratory comparison. Ranges are provided in parentheses.

Treatment	Lab	Temp. (°C)	Salinity (ppt)	pH (SU)	D.O. (mg/L)
Control	A	20.0 (19.6-20.2)	29 (28-31)	8.14 (7.98-8.36)	7.8 (5.4-8.8)
	B	19.9 (19.6-20.5)	28 (28-29)	7.81 (7.60-8.00)	7.3 (6.9-7.5)
	C	19.7 (19.4-20.2)	30 (26-29)	7.85 (6.82-8.14)	7.3 (5.7-8.4)
Reference	A	20.0 (19.6-20.2)	29 (28-31)	8.19 (8.07-8.41)	7.9 (5.1-9.0)
	B	19.9 (19.6-20.5)	28 (28-28)	7.87 (7.60-7.90)	7.3 (6.9-7.6)
	C	19.7 (19.1-20.1)	27 (26-28)	7.93 (7.46-8.17)	7.4 (6.5-8.3)
Hudson	A	20.0 (19.5-20.2)	29 (28-31)	8.15 (7.91-8.48)	7.8 (5.2-8.7)
	B	19.9 (19.6-20.5)	28 (28-28)	7.79 (7.70-8.00)	7.2 (6.7-7.4)
	C	19.8 (19.3-20.2)	27 (26-28)	7.95 (7.68-8.17)	7.3 (6.2-8.3)
Chester	A	20.0 (19.5-20.2)	29 (28-30)	8.24 (8.13-8.64)	7.6 (5.2-8.7)
	B	19.9 (19.6-20.5)	28 (28-29)	8.01 (7.50-8.40)	7.0 (3.4-7.5)
	C	19.6 (19.1-20.0)	27 (26-28)	8.06 (7.83-8.41)	7.3 (6.4-8.3)
Red Hook	A	20.0 (19.5-20.2)	28 (27-30)	8.18 (7.95-8.37)	7.8 (5.4-8.8)
	B	19.9 (19.6-20.5)	28 (28-29)	7.97 (7.80-8.30)	7.2 (6.7-7.5)
	C	19.6 (19.0-20.1)	30 (26-28)	8.07 (7.85-8.34)	7.4 (6.7-8.3)
100% Newark	A	20.0 (19.5-20.2)	30 (28-30)	8.20 (8.03-8.48)	7.7 (5.7-8.7)
	B	19.9 (19.6-20.5)	28 (28-29)	7.94 (7.70-8.30)	7.2 (6.8-7.5)
	C	19.6 (19.1-19.9)	27 (26-28)	8.03 (7.83-8.20)	7.3 (6.7-8.3)
75% Newark	A	20.0 (19.5-20.2)	29 (28-30)	8.20 (8.14-8.35)	7.8 (5.5-8.8)
	B	19.9 (19.6-20.5)	28 (28-29)	8.00 (7.80-8.20)	7.2 (6.8-7.4)
	C	19.7 (19.1-20.2)	27 (26-28)	8.08 (7.84-8.25)	7.3 (6.7-8.2)
50% Newark	A	20.0 (19.5-20.2)	29 (28-30)	8.25 (8.10-8.40)	7.8 (5.7-8.8)
	B	19.9 (19.6-20.5)	28 (28-29)	8.01 (7.80-8.30)	7.2 (6.3-7.5)
	C	19.8 (19.2-20.5)	27 (26-28)	8.07 (7.84-8.39)	7.4 (6.7-8.3)
25% Newark	A	20.0 (19.5-20.3)	29 (28-31)	8.23 (8.10-8.40)	7.8 (5.5-8.8)
	B	19.9 (19.6-20.5)	28 (28-29)	8.01 (7.80-8.30)	7.3 (7.0-7.5)
	C	19.7 (19.1-20.2)	28 (26-28)	8.04 (7.85-8.31)	7.3 (6.6-8.0)

Table C6. Mean total porewater ammonia concentrations (mg/L) for each test sediment and protocol between labs. Porewater ammonia levels are provided at the initiation and termination of tests.

Sediment		10-d *Ampelisca*	28-d *Leptocheirus*	28-d *Neanthes*	20-d *Neanthes*	20-d Mod *Neanthes*
Control	Start	14.5 ± 3.8	9.8 ± 3.5	12.4 ± 1.4	11.2 ± 2.5	11.2 ± 2.3
	End	7.1 ± 2.7	1.0 ± 0.8	0.7 ± 0.1	4.9 ± 3.4	0.7 ± 0.7
Hudson	Start	23.5 ± 5.2	20.7 ± 3.8	26.7 ± 4.8	20.2 ± 7.2	20.1 ± 3.4
	End	3.5 ± 3.4	1.9 ± 1.8	0.4 ± 0.1	1.1 ± 0.8	1.9 ± 2.5
Chester	Start	17.9 ± 2.7	16.7 ± 2.6	21.0 ± 2.7	10.0 ± 8.9	15.7 ± 4.9
	End	2.4 ± 2.0	1.5 ± 0.4	0.7 ± 0.2	1.7 ± 1.4	0.5 ± 0.1
Red Hook	Start	18.7 ± 3.8	17.0 ± 5.2	20.7 ± 2.9	17.7 ± 6.2	17.0 ± 4.6
	End	4.2 ± 1.8	1.5 ± 0.5	1.1 ± 0.4	3.3 ± 3.1	0.9 ± 0.1
100% Newark	Start	18.3 ± 4.9	18.3 ± 1.8	24.2 ± 3.5	18.3 ± 5.8	21.1 ± 0.7
	End	2.3 ± 1.9	1.4 ± 0.5	0.7 ± 0.2	1.1 ± 0.1	0.7 ± 0.3
75% Newark	Start	17.2 ± 1.8	17.2 ± 4.9	19.0 ± 1.3	13.7 ± 6.5	16.3 ± 3.2
	End	1.8 ± 1.4	1.6 ± 0.8	0.4 ± 0.1	1.7 ± 2.0	0.3 ± 0.2
50% Newark	Start	13.9 ± 3.1	14.3 ± 3.3	17.6 ± 0.8	13.4 ± 3.6	12.6 ± 3.7
	End	1.4 ± 1.7	1.2 ± 0.5	0.5 ± 0.0	0.7 ± 0.6	0.2 ± 0.2
25% Newark	Start	13.7 ± 3.0	12.3 ± 3.1	14.1 ± 1.2	14.4 ± 0.7	10.2 ± 2.8
	End	1.5 ± 1.6	1.5 ± 0.5	0.5 ± 0.0	0.7 ± 0.4	0.2 ± 0.0

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
July 2008 Revised	Final report	

4. TITLE AND SUBTITLE

Dredged Material Analysis Tools; Performance of Acute and Chronic Sediment Toxicity Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jeffery Steevens, Alan Kennedy, Daniel Farrar, Cory McNemar, Mark R. Reiss, Roy K. Kropp, Jon Doi, and Todd Bridges

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Environmental Laboratory, U.S. Army Engineer Research and Development Center, 3909 Halls Ferry Road, Vicksburg, MS 39180-6199; Specpro, Inc., 3532 Manor Drive, Vicksburg, MS 39180; U.S. Environmental Protection Agency, Region 2, 290 Broadway, New York, NY 10007; Marine Sciences Laboratory, Battelle Pacific Northwest Division, Sequim, WA 98382; Aqua Survey, Inc., 469 Point Breeze Road, Flemington, NJ 08822

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDC/EL TR-08-16

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Corps of Engineers
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report and research were supported by the U.S. Army Corps of Engineers New York District and U.S. Environmental Protection Agency Region 2. The work was conducted to provide insight into the potential advantages and disadvantages of using chronic sediment toxicity tests with relevant benthic macroinvertebrates as part of dredged material evaluations, as described in the Inland and Ocean Testing Manuals (USEPA/USACE 1991, 1998). Nine sediments collected from the New York Harbor (NYH) were used to assess test methods in a preliminary evaluation at one test facility and an interlaboratory evaluation at three test facilities. The two acute test methods (10-day *Ampelisca abdita* and *Americamysis bahia*) currently used in evaluations of NYH material were compared to available chronic protocols to gauge relative performance of the toxicity tests. Acute tests are typically short-term (e.g., 10-day) lethality assessments conducted over a small portion of the test organism's life cycle, while chronic tests are longer-term and assess sublethal measurement endpoints (e.g., growth and reproduction) in addition to lethality. The available chronic test methods used in this study were the 28-day test using the estuarine amphipod, *Leptocheirus plumulosus*, and 20-day and 28-day tests using the marine polychaete *Neanthes arenaceodentata*. Use of chronic tests is recommended or required by dredged material evaluation guidance and regulations, respectively. The sublethal endpoints measured in chronic tests may be more sensitive measures of toxicity and more predictive of

(*Continued*)

15. SUBJECT TERMS

Sediment toxicity	Sublethal	Amphipod
Chronic	Bioassay	Polychaete

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED		72	

14. ABSTRACT (Concluded)

longer-term population effects. Of the tests compared, the currently used acute (10-day) *Ampelisca abdita* test and the available chronic (28-day) *L. plumulosus* test were the most responsive (i.e., sensitive) to the tested NYH sediments. Response is defined as the amount an endpoint (e.g., survival) was reduced for test organisms in site sediments relative to that same endpoint in the control sediment. Of these two test methods, neither clearly demonstrated better capability to identify contaminated sediments (i.e., "hits"). The *A. abdita* test was more consistent in performance and exhibited greater statistical power but demonstrated lesser response to the sediments and lower correlation with sediment chemistry. The sublethal endpoints used in the *L. plumulosus* test were more responsive to the sediments and more closely related to sediment contamination but had lower statistical power than lethality endpoints. An acute (10-day) test using *L. plumulosus* was also conducted in one laboratory and similar responsiveness was found relative to the acute *A. abdita* test. The remaining toxicity tests, including the currently applied acute *A. bahia* test and the 28-day *N. arenaceodentata test* were not responsive to the tested sediments in this evaluation and thus did not suggest toxicity in any of the tested sediments. Specific conclusions and recommendations on the application of these test methods are offered at the end of this document.