N62661.AR.002408
NS NEWPORT
5090.3a

LETTER AND U S NAVY RESPONSE TO AGENCIES COMMENTS REGARDING DRAFT PHASE 2 REMEDIAL INVESTIGATION AND BASELINE ECOLOGICAL RISK ASSESSMENT FOR SITE 17 BUILDING 32 GOULD ISLAND NS NEWPORT RI
06/28/2011
TETRA TECH NUS

TETRA TECH

C-NAVY-06-11-4446

June 28, 2011

Project Number 112G00949

Ms. Kymberlee Keckler, Remedial Project Manager
U.S. EPA Region I
5 Post Office Square, Suite 100
Boston, Massachusetts 02114-3912

Mr. Gary Jablonski, Remedial Project Manager
Rhode Island Department of Environmental Management
235 Promenade Street
Providence RI, 02908-5767

Reference: CLEAN Contract No. N62467-04-D-0055
Contract Task Order No. 458

Subject: Response to Comments on Draft Phase 2 RI and BERA
IR Site 17, Building 32 Area, Gould Island
Naval Station Newport, Newport RI

Dear Ms. Keckler, Mr. Jablonski:

On behalf of Ms. Maritza Montegross, U.S. Navy NAVFAC Mid-Atlantic, attached, I am providing to you the response to comments on the Draft Phase 2 RI and Baseline Ecological Risk Assessment for the site referenced above. This response summary is provided based on the resolutions reached by following discussions held with yourselves and the rest of the technical project team copied on this transmittal. Such discussions were held April 15, May 9, May 16 and June 13, 2011.

Based on these responses, the draft final document will be prepared and submitted. If you have any questions, please do not hesitate to contact me at 978-474-8434.

Very truly yours,

Stephen S. Parker, LSP
Project Manager

Enclosures

c: A. Bernhardt TtNUS (w/encl.)
D. Barclift, NAVFAC (w/encl.)
K. Finkelstein, NOAA (w/encl.)
A. Gavaskar, NAVFAC (w/encl.)
G. Glenn, TtNUS (w/o encl.)
B. Hoskins, USEPA (w/encl.)
W. Johnson, NAVFAC (w/encl.)
K. Munney, USF&W (w/encl.)
P. Steinberg, Mabbett Associates (w/encl.)
D. Moore, NAVSTA (w/encl.)
Site File (c/o G. Wagner TtNUS)(w/encl.)
File 112G00949-8.0 (w/encl.), 3.1 (w/o encl.)

Tetra Tech
250 Andover Street, Suite 200, Wilmington, MA 01887-1048
Tel 978.474.8400 Fax 978.474.8499 www.tetratech.com

ATTACHMENT A

NAVY RESPONSES TO U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA) COMMENTS (DATED FEBRUARY 1, 2011)
PHASE 2 REMEDIAL INVESTIGATION, AND BASELINE ECOLOGICAL RISK ASSESSMENT – DECEMBER 2010,
SITE 17, GOULD ISLAND
NAVSTA NEWPORT, NEWPORT RHODE ISLAND

Navy responses to the EPA comments on the Phase 2 Remedial Investigation and Baseline Ecological Risk Assessment – December 2010, Site 17, Gould Island, Naval Station Newport, Newport Rhode Island (February 2010) are presented below. Comments & responses were discussed 4/15/11, 5/9/11, 5/16/11, and 6/13/11. The comments are presented first (in italics) followed by Navy's responses based on these discussions.

General Comments

Comment 1*: While the receptors, exposure assumptions, NOAELs and LOAELS, for the food chain modeling are appropriate and adhere to the December 2009 Sampling and Analysis Plan, it is unclear why uptake and risk were only calculated for exposure to sediment. Surface water and biota data were not evaluated in most cases, even though the data were available, and uptake parameters were selected for these pathways. Please incorporate these data to the calculations.*

Response: Risks to mammals and birds from chemicals in surface water were not evaluated in the BERA because the surface water is saline and would not be used as a drinking water source for mammals and birds. The water ingestion rates in Table 6-3 are not applicable and will be deleted from this table and any appendix tables where they appear. Risks to aquatic organisms from chemicals in surface water were evaluated in the December 2006 RI report. Because no additional surface water data were collected as part of the BERA sampling, it was not necessary to re-evaluate risks to those receptors.

As discussed on April 15, 2011, *tissue samples will be included in the food chain model regardless of whether their concentrations exceed PALs. Chemicals with site tissue concentrations greater than reference tissue concentrations will be carried through the conservative and refined food chain models, but a comparison to background concentrations will then be presented in Section 6.4.3.*

Comment 2*: The interpretation of the toxicity tests and derivation of NOEC and LOEC separately for each endpoint is not appropriate. EPA considers a sample toxic when there were effects for any of the endpoints. Identification of NOEC and LOEC should be based on this assumption. While various levels of interpretation have been provided in the BERA, priority should be given to NOEC and LOEC based on all toxicity test endpoints (survival, growth, and reproduction) together.*

Response: The Navy did give priority to the NOECs and LOECs developed using all endpoints. In fact, the NOECs and LOECs used to evaluate the results in Figures 6-1 through 6-3 were developed using all of the endpoints.

On April 15, 2011, the approach used was discussed and it was agreed that for discussion purposes, the Navy would revise Table 6-27 to include any sample on Table 6-13 that had lower survival, growth, or reproduction versus the laboratory control as a "toxic" sample. The attached Draft Table 6-31 shows this revision. As can be seen from the table, that even using this very conservative approach, very few NOECs and LOECs actually changed because most of the samples that were added as "toxic" samples had low chemical concentrations.

The topic was discussed again on 5/16/11, and again on 6/13/11. The central issue appeared to be how to determine which samples are toxic vs. non-toxic within this data set. Tables showing replicate survival results were generated and are provided attached (Table A and Table B). It was noted that sample 509 was listed as toxic (at 63% survival) whereas sample 519 was listed as non-toxic (also at 63% survival). The laboratory's statistical tests were evaluated and it was concluded that this was an area of uncertainty originating with the variability between replicate chambers of the test organisms. Alternate statistics were not determined to be a good resolution. Because the uncertainty could not be resolved, it was agreed on 6/13 to republish Table 6-27 as was originally done in the Baseline ERA Risk Characterization section of the BERA report. However, a revised Table 6-27 (it will be Table 6-31 in the BERA report) will be included in the uncertainty section and will categorize any sample on Table 6-13 that had lower survival, growth, or reproduction versus the laboratory control as a "toxic" sample. In addition, it would categorize a sample as "toxic" if the sample had less than 70% survival. It was recognized during discussions that the PRG values need to be established in risk management steps of the FS and the PRG may be based on the NOECS and LOECS from either the original Table 6-27, which will remain in the report or the revised Table 6-31.

Specific Comments

***Comment 1, p. E-2 ¶2**: In the last sentence remove "conducted."*

Response: Comment noted, the word "conducted" will be removed from the last sentence of the above mentioned paragraph.

***Comment 2, p. E-4 ¶1:** In the first sentence following the bullet please change "insoluble in water" to "relatively insoluble in water "or "only slightly soluble in water." Make the same correction on pages 1-20, 5-5, and 5-11.*

Response: Comment noted, the changes requested above will be made to the document in the appropriate locations.

***Comment 3, p. E4 ¶2**: In the second sentence after the bullet please change "are not likely to breakdown or degrade" to "are likely to breakdown or degrade only very slowly."*

Response: The requested revision will be made.

***Comment 4, p. 1-10 §1.4.4**: The last sentence refers to trace amounts of PCBs in groundwater, but the concentrations cited are 500 times the MCL and therefore are not considered trace . Please edit the text here and on pages 1-14, 5-2, and 7-3 to correct this.*

Response: Comment noted, the words "trace" and "low" will be removed from describing the concentrations and the concentrations that were detected will be stated without interpretation.

***Comment 5, p. 1-23 §1.6:** In the last sentence please delete "in" following CSM.*

Response: Comment noted, change requested above will be made.

***Comment 6, p. 2-2 §2.1.1 ¶4:** The third sentence refers to 2-foot intervals, but the surface soil sample was collected over a 1-foot interval. Please correct.*

Response: This paragraph will be changed to state the following "The four soil borings located at the former rigging platform (SB-410 through SB-413) were also advanced using HSA drilling techniques. In accordance with the SAP, samples that were collected for laboratory analysis were collected from depths as follows: a) 0-1 foot, b) the 2 foot interval immediately above the saturated zone, and c) a 2 foot interval determined by the field geologist base on visual evidence of contamination, staining or sheen, or olfactory evidence of contamination. From each interval that was retrieved from the boring, two sample aliquots were collected; one sample aliquot was used for field screening of VOCs (jar headspace analysis), and the second aliquot was set aside for possible laboratory analysis (aliquots collected to represent surface soil, 0-1 foot, were collected only from the top foot of the 0-2 foot split barrel sampler). Samples collected from these borings and selected for laboratory analysis were analyzed for PCBs, PAHs, and TAL metals."

Comment 7, p. 2-5 §2.2.1.1: *Please discuss how to handle deeper sediment profile samples, especially the use of core baskets, or core catchers. There may be a better way to handle core loss in future sampling events (such as pre-remediation sampling for depth profiling contaminants). The use of a basket can be helpful for soft sediments, or for coarser sediments prone to falling out of a core upon retrieval, however they have some significant disadvantages. First, they tend to striate and smear the core as it enters the tube, creating a distortion of the core profile. Second, they can create resistance that pushes very soft sediment aside and prevents entry into the core tube, such that the top of the sample does not truly represent the top of the sediment profile. Often enough, at a certain depth, a core sampler will meet significant resistance, regardless of whether a core basket is used. Often this occurs in a sand layer that compacts at the bottom end of the tube. At this point, the end of the core tube is effectively plugged against core loss, and the core can be retrieved intact without the use of a core basket and the concurrent problems they present. The cores taken at Gould Island are likely adequate, but there is some uncertainty in the depth profiling that could possibly have been avoided. Please add this to the uncertainty discussion.*

Response: The field team made multiple attempts to retrieve full recovery at each location, but as stated in the report, was unsuccessful at all but two locations. At multiple locations attempts were made to collect sediment using the core tube with no basket. All attempts to collected sediment without the core basket inserted into the core tube either destroyed the core tube because the stainless steel tip was part of the core basket, or there was simply no recovery. This section will be revised to include a description of these failed attempts as well as a discussion regarding the potential use of other methods of sediment coring in areas with different sediment characteristics for future reference at this site or others.

Comment 8, p. 2-6 §2.2.1.2: *The second paragraph states that three locations where samples were planned were not sampled. Why weren't substitute samples collected?*

Response: SD-429 and SD-438 were not collected because of weather restrictions during the first round of sediment sampling. These were the last samples to be collected during the round 1 sediment sampling and weather created an unsafe sampling condition. These two samples were collected and sent to the laboratory for analysis of PCB homologues, PAHs, Metals and TOC, during the round 2 sediment sampling. An explanation of this will be included to this section.

Sediment sample SD-448 was not collected because there was no sediment present at this location and moving this station immediately offshore was not possible due to large pieces of concrete preventing boat traffic in the area and a sheetpile wall preventing access from shore.

The Final SAP called for the collection of 30 surface sediment samples as part of the ecological investigation. Analytical results from the round 1 sediment samples were supposed to be used to choose these 30 locations. During a conference call on 2/25/2010, where the EPA, RIDEM, NOAA and the Navy were represented, locations for the ecological investigation sediment sampling were chosen. During this conference call the RIDEM and NOAA each requested one additional sample station. The Navy agreed to move forward with these requests and added ecological sediment sample locations SD-508 and SD-509. Therefore, although it was not a primary objective, substitute samples were added during the ecological sediment sampling event.

Comment 9, p. 4-1 §4.0: *The percent moisture discrepancy between the SAP Project Action Levels (PALs) in dry weight and the lab-reported concentrations in wet weight were resolved using 85% moisture for tissue. Please clarify whether this was a sample-specific or generic tissue-type value.*

Response: The percent moisture value for tissue was not sample specific. The 85% moisture for tissue was based on a value of 82% moisture for bi-valves in EPA (1993), which was "rounded" to 85% because the value of 82% had a standard deviation of 4.5. The percent moisture value for crabs in EPA (1993) is lower (74%) so the 85% value is considered to be conservative. Note that the 85% moisture value was not used to modify the actual tissue concentrations. The 85% moisture value used to modify the ingestion rate of the raccoon, which was then used to calculate the PALs in Table 6-9 and was used for the food chain model calculations. That is because the source of the ingestion rate for the raccoon (Nagy et al., 1999), predicted ingestion rates on a dry-weight basis. The 85% moisture value was not used to convert the food ingestion rate for the herring gull, because the food ingestion rate for the gull was already on a wet-weight basis. However, it was used to convert the food ingestion rate to a dry weight basis to evaluate incidental ingestion of sediment, because the sediment was reported on a dry weight basis. This is a conservative estimate because the percent moisture of sediment is generally less than 40%.

Comment 10, p. 4-2 §4.0: *In the first full sentence, please change "statistical calculates" to "statistical calculations."*

Response: Comment noted, change requested above will be made.

Comment 11, p. 4-10 §4.1.4: *In the last sentence please change "to not pose" to "do not pose."*

Response: Comment noted, change requested above will be made.

Comment 12, p. 4-46 §4.3.7: *The first sentence needs to be corrected. PCB tissue samples from the Stillwater Basin biota and the Northwest Shoreline biota had greater PCB concentrations than biota from the Northeast Shoreline.*

Response: The sentence will be changed to state that "Tissue samples collected from the Northeast shoreline contained PCB Homologues concentrations at levels slightly below levels detected in samples collected from the Stillwater Area, and at levels that are greater than levels detected in samples collected from the Northwest Shoreline." Basis for this conclusion is presented as follows:

NE average – 46.4 µg/kg
NW average – 6.9 µg/kg
STILL average – 60.7 µg/kg

Comment 13, p. 5-7 §5.5 ¶2: *Please correct the last sentence because mercury is present in potentially significant concentrations in coal. This site operated a coal-fired power plant,*

and therefore the mercury present at the site could reasonably be attributed to its presence in the coal.

Response: The use and burning of coal at and prior to the time period that it was done at this site was a common occurrence across the industrialized northeast US. While the Navy does not disagree that mercury could have been produced as such, the deposition from other coal burning operations all over the area cannot be separated from that produced at this location. Therefore, it would be identified as an anthropogenic contaminant. The statement will be revised as follows: "Mercury was also detected exceeding the ERM during Phase 1, but was not suspected to be related to site activities."

Comment 14, p. 6-16 §6.4.1: *The first sentence is unclear and does not really explain the need for toxicity tests. Please consider rephrasing to something like: "Several of the sediment risk screening levels used to select COPCs are conservative and exceedances of them do not necessarily mean that adverse impacts to the benthic community are occurring."*

Response: The sentence will be modified as suggested

Comment 15, pp. 6-17 to 6-19 §6.4.1.1: *The decision rules described require that a sample must be statistically different from two of the three reference sample in order to be considered toxic. In this instance, most samples that were toxic to one reference sample were also toxic to two or more, so this rule did not drive the process significantly. EPA recommends that the data be examined differently, such as ranking the samples in descending order of toxicity in order to make sense of the contamination and toxicity data. This decision rule was not used in other Navy sites, where toxicity to any reference sample was sufficient to consider the sample impacted.*

Response: The objective of the rule that requires that a sample to be statistically different from two of the three reference sample in order to be considered toxic was not intended to drive any process. It was intended to determine which samples were considered toxic versus reference samples. Also, there were several samples that only had lower growth or reproduction compared to one of the reference samples. A similar approach was used at the Area A Wetland at New London, except only two reference samples were available. At that site, a comparison to the combined reference set provided the "third" sample. That being said, the Navy did examine the data differently as discussed in the Navy's response to EPA general comment No. 2. Note that NOECs and LOECs were developed for the chemicals considered the primary risk drivers at the site (PAHs, PCBs, and ERM-Q, which includes several metals).

Comment 16, p. 6-21 §6.4.1.3 ¶3: *Regarding the last sentence, the analysis is reasonable but because of the extraordinarily high metals concentrations found in a few locations, contaminant migration is a concern so these isolated areas should not be disregarded for remediation.*

Response: Any potential remedial actions taken for the site would be discussed and evaluated in the feasibility study, not this risk assessment step. The statement does not disregard metals for this analysis in the FS. That being said, the need to remediate the site would have to consider that fact that the areas with elevated concentrations of metals are small and could not be duplicated when re-sampled. Therefore, even if the sediment were to be dispersed and migrate to other areas, it is unlikely that these areas would pose a significant risk to sediment invertebrates because the concentrations would be lower in these areas.

Comment 17, p. 6-22 §6.4.1.3: *Please review the tabulated NOEC and LOEC values for total PCBs as they are not consistent with the text in the preceding paragraph.*

Response: The LOEC for total PCBs is 2,720. The table will be corrected to show that value.

Comment 18, p. 6-27 §6.5.2 ¶3: *a) Please supplement the second sentence to balance the statement by adding: "… conversely, chemicals that are biodegradable under field conditions may become non-biodegradable when removed from the field."*

b) Please delete the fifth sentence because not all chemical concentrations in the sediment can be considered low; some are quite high at many stations. The discussion should also note that uncertainty exists regarding consistency between the chemical concentrations in the samples sent for toxicity testing and the samples analyzed by the laboratory.

Response: a) The suggested text will be added with the exception that the word "biodegradable" will be replaced with "bioavailable."

b) The fifth sentence will be changed to read as follows: "This likely is partially due to the low concentrations of chemicals in some of the sediment samples considered to be toxic." Also, the following sentence will be added to the beginning of the fourth paragraph of section 6.5.2: "Although the sediment samples were well homogenized before they were sub-divided, there may be differences between the chemical concentrations in the samples sent for toxicity testing and the samples analyzed by the laboratory."

Comment 19, p. 6-28 §6.5.2: *Please revise the last sentence regarding the use of "small" for all areas of contamination. The area of contamination in the Stillwater Basin is not small.*

Response: The last sentence was meant to reference the small areas for metals. Therefore, the word "contaminant" will be changed to "metals."

Comment 20, p. 7-4 §7.1.5 ¶3: *The second sentence concludes that neither the coal pile nor the sanitary sewer was a source of the PAHs detected. Please explain the basis for that determination.*

Response: The paragraph will be revised to state: "Subsurface soil data did not show presence of PAHs or fuels in the area of the leach field, Building 33, or the coal pile upgradient of the roadway where the PAHs were found in the phase 1 RI. Therefore…"

Comment 21, p. 7-11 §7.1.8.4: *In the penultimate sentence change "to" to "do."*

Response: Comment noted, change requested above will be made.

Comment 22, p. 7-16 §7.4: *Please explain the basis for the third sentence and incorporate the discussion to the RI.*

Response: All buildings in the vicinity of, and including, Building 32 were demolished from 2000 through 2002. This negates the risk of vapor intrusion into indoor air spaces. A discussion of this will be included into this paragraph.

Comment 23, p 7-18 §7.5 ¶4: *It is not apparent why the RI would suggest that indicator PAHs be used to focus remediation since there is no apparent cost benefit for this approach. Please clarify the rationale for using indicator PAHs.*

Response: The use of an indicator PAH or limited COC approach provides for a single COC to be used to direct remedial action efforts, instead of organizing an area based on a large number of related COCs such as similar PAHs. It is agreed that if the risk is a result of a group of PAHs, such as HMW PAHs, it may not matter because the PRG is set for the group as a whole.

ATTACHMENT B

Responses to Comments from RIDEM
Draft Phase II Remedial Investigation, Baseline Ecological Risk Assessment
Site 17, Building 32, Gould Island, NETC
Comments Dated 2/14/11

1. Page E-1, Executive Summary; Paragraph 3, 4th sentence.

"The concrete roadways were removed at that time, and polychlorinated biphenyl (PCB) releases were discovered near transformer buildings around Building 32."

PCBs were found at other structures, such as the Riggers Storage Building and in the roadway. Further, Transformer Buildings 61 & 62 are located at the end of the connecting pier close to Building 35. Please modify the above as follows and modify any other similar statements found elsewhere in the report: "The concrete roadways were removed at that time, and polychlorinated biphenyl (PCB) releases were discovered near transformer buildings, <u>other structures, and in the roadway</u>."

Response: The requested change will be made.

2. Page E-2, Executive Summary; Paragraph 2.

"The affected groundwater is limited to areas associated with the drainage systems outside the southwest corner of Building 32."

MCLs were also exceeded at the northeast corner of Building 32. Please remove this statement and any other similar statements from the report and modify the above to state that exceedances of MCLs were observed outside of the southwest and northeast corners of Building 32.

Response: The requested change will be made.

3. Page E-3, Executive Summary; Paragraph 1.

This paragraph states that the concentration of PCBs in the Stillwater Area is higher in the 6-24 inch interval and the contaminant distribution also points to the deposition of clean material over the contaminated sediments. A review of the data in Figures 4-6 through 4-10 reveals that this is not the case. It is also possible that the observed distribution of contaminants may have resulted in part due to groundwater upwelling. Please remove these statements from both this and any other section of the report and language discussing the actual observed depositional rate.

Response: The data support the statement made in the cited passage (first bullet for Phase 2 data, second sentence). The summary figures 4-6 through 4-10 may not show this clearly because they combine data across several time periods. However, the data does support this conclusion. In particular the data from 2010 in the Stillwater area where only surface sediments were collected show lower concentrations than the data collected in 2009 where surface and subsurface intervals overlap to some extent.

There is no information to suggest groundwater upwelling is bringing PCBs to the subsurface or surface sediment in the Stillwater basin. No changes are recommended based on this comment, although revisions of figures described elsewhere in this response summary will make this conclusion more evident.

4. Page E-4, Executive Summary; Paragraph 1.

"Pesticides, PCBs and metals found in soils are stable and insoluble in water. However, through erosion and migration of contaminated soils, these compounds could be transported from soils to marine sediments, and could also migrate with sediment within the marine environment, particularly in disturbed or high energy areas at the shoreline."

Pesticides, PCBs, and metals have limited solubility and can be found in the groundwater. Please remove this statement and modify the above and any other similar statements in the report as follows: "Pesticides, PCBs and metals found in soil are stable and <u>have limited solubility</u> in water. Through erosion and migration of contaminated soils, <u>as well as migration via groundwater</u>, these compounds could be transported from soils to the marine environment and could also migrate with sediments within the marine environment particularly in the disturbed or high energy areas of the shoreline."

Response: The Navy concurs with regards to the use of the term "limited solubility". There is no data to suggest pesticides, PCBs or metals in sediment have been brought to the sediment via groundwater flow.

5. Page E-5, Executive Summary; Paragraph 3.

The report notes that the site exceeds MCLs; however, groundwater is currently not used at the site. The report should note that the site is in a GA aquifer and in the past groundwater was used as a potable water supply. Please add the following to this section and any other similar section of the report: "The site is located in a GA aquifer which has a remedial objective of MCLs. Though not in use today, historically groundwater was used on the site."

Response: The report will be revised to cite the presence of the state GA groundwater designation, and the past use of groundwater at the site as a potable water source. The remedial objectives are not identified at this point and should not be speculated.

6. Page E-5, Executive Summary, Recommendations.

The report proposes a series of recommendations for the FS which are primarily based upon the results of the Phase II investigation. Please be advised that the data from the SASE, Phase I/II RI, and any other investigations and removal actions conducted at the site must be compiled and included in the evaluation of areas of concern for the site. This evaluation should include figures which delineate exceedances of criteria or regulations.

Response: The conclusions section of the executive summary includes the information pertinent to the recommendations for the FS. Other changes to the document as described elsewhere in this response will support these conclusions. A data compilation will be provided in Appendix D and Figures will be revised as described elsewhere in this response summary.

7. Page 1-14, Section 1.5.2, Site History; Paragraph 5.

The report states that PCBs were found in the shoreline near Building 54 but not extended out beyond the shoreline. Please be advised that PCBs up to 8800 ppb were found in sediment samples collected at this location. Elevated levels were also found on the shoreline. A series of removal actions were conducted in this area including the removal of contaminated soil, sediments and shoreline material. Please modify the report to state that PCBs were found in sediment and shoreline samples collected at the site and a removal action was conducted in this area for both soil and sediment.

Response: This information will be added to this section and onto page 1-10, within the last paragraph that describes the PCB removal actions in this area. The information available for post excavation data will be provided.

8. Page 2-2, Section 2.2, Sediment Investigation; Whole Section.

Sediment samples were collected during the three different sampling rounds employing three different collection methods (divers, ponar, and tubes). Please include a discussion of the methods employed for each sampling event, and the procedures used to mark the locations of the samples in the field, including the plus or minus accuracy bars. As an illustration, the first sampling event employed divers. Was an underwater GPS unit employed? Was the location estimated from a surface craft using GPS? Were measurements made from fixed locations? etc. Please include this information in the response to comments.

Response: All sediment samples that were collected during the Phase 2 RI were located using a DGPS Leica Model 421B that is accurate to 1-3 meters. The unit was either mounted on the boat for samples accessed using a boat or hand-held for samples that were accessed via shore. Sediment samples that were collected during the Phase 1 RI were located using a Trimble Pro XR DGPS and post corrected to achieve sub-meter accuracy. Further details regarding Phase 1 RI sediment sampling can be found in section 2.5 and Appendix E-2 of the Phase 1 RI (TtNUS, 2006).

Section 2.2 of the report will be revised to better explain the techniques used to locate sediment samples at Gould Island during the Phase 2 RI, as described above.

9. Page 2-3, Section 2.1.2, Soil Boring Investigation, Field Measurements and Observations; Last Paragraph, Table 2-2.

The report notes that a number of the soil borings were designed to investigate potential petroleum sources of contamination in the area. The report does not present the information that was used to position the borings in the field. As an illustration, SB 407 is located in the general vicinity where underground storage tanks had been removed by the Navy, and free product was encountered on top of the water table after the removal action. The figure in the report does not depict the location of the underground storage tanks, and the report does not indicate how the location of the tanks was determined in the field (GPS coordinates from removal actions, measurements taken from field notes during removal action, engineering plans, etc). Please modify the figure to depict the location of these areas of concern, discuss in the report how the borings were located and include in an

appendix the original sources of information which were used to locate the borings (i.e., a copy of the figure from the tank removal action depicting the location of the tank).

Response: The location of the UST was speculated to be at the southwest corner of Building 33. However, the records available do not show the specific location. It is our understanding that the removals conducted in this area were completed as appropriate. If RIDEM has information on the location of the tank or photos of the excavation showing contamination, or evidence that the work was done incompletely, the Navy could use that information to direct appropriate actions.

10. Page 2-3, Section 2.1.2, Soil Boring Investigation, Field Measurements and Observations; Last Paragraph, Table 2-2.

Soil borings were collected in the area of the drainage system in order to ascertain if there was a source. At Boring 402, the interval which exhibited the highest degree of field contaminant (333 or 129 ppm interval) was not tested. Instead, the interval with a headspace reading of 30 ppm was tested for laboratory analysis. Similarly, at boring 404, the interval with the highest degree of contaminant (800 ppm headspace reading) was not tested for lab analysis. Please explain why the intervals with the highest PID readings were not tested for lab analysis.

Response: Worksheets #11, 14 and 17 of the Final SAP required that soil samples be collected at the 0-6 inch, 12-18 inch, 24-30 inch and 42-48 inch intervals. Although the SAP does call for samples to be collected at 2 foot intervals for the entire length of the boring for classification (which jar headspace analysis is a part of) it did not require additional samples be collected for laboratory analysis at locations where elevated field screening readings occurred. However, the presence of these headspace concentrations measured is important in the interpretation of the data, and the text of Section 4.1.1.1 will be revised to note that soil headspace screening results in the 4-6 foot interval at Borings SB402 and SB404 indicate that the soils could contain TPH above the concentrations of TPH measured in the samples collected in the upper intervals.

11. Page 2-4, Section 2.2, Sediment Investigation; Figure 2-2.

Please include a box around sample location SD-406 in the Stillwater Area labeled SD-501 indicating a 2010 BERA sediment sample location. Also, please depict the terminus of the discharge pipes based upon information from both engineering plans and field observations.

Response: Figure 2-2 will be updated to correct sample location SD406 and include sample location SD501. The Phase 1 RI report presents a discussion on the terminus of each of the discharge pipes found (section 2.5.1). During phase 1, one sample station was positioned at each pipe terminus found by divers and designated with an “F” (Figure 2-2). This will be explained in the cited section.

12. Page 2-4, Section 2.1.2, Soil Boring Investigation, Field Measurements and Observations; Paragraph 2.

The report notes that Soil Borings 410-415 were installed during the removal action associated with the Riggers Storage Building. Please include on the figure the outline of the area which was subject to the removal action. In addition, please include pertinent

information in the report such as the depth for the removal action and the results of the confirmatory samples.

Response: The information presented in Appendix A of the Phase 1 RI report will be used to present the extents of the PCB removal actions, these extents will be added to Figures 1-3 and/or 2-1, as appropriate.

13. Page 4-1, Section 4.0, Nature and Extent of Contamination; whole section.

This section of the report deals with the nature and extent of contamination based upon information obtained from the Phase II Investigations. The report also includes tables and figures with the results of the Phase II Investigation. While this is acceptable, the Phase II Investigation, as noted in Section 1.0 of the report, was to provide additional information to access potential source areas and address extent issues. As such, the results of the Phase I and previous studies must be included in the report, tables, and figures. Therefore, please modify Section 4.0 to include the results of the previous investigations and please produce tables and figures with this information. The figures should delineate sample locations which exceed criteria along with the concentrations observed and depth of samples.

An approach similar to the 4-1 through 4-5 series in the Phase I RI should be employed (soil figures should also include any pertinent observations such as staining or product, sediment figures with analytical results should also include toxicity test results, ERQ and designation of samples which exceed ERLs, ERMs and are considered toxic). Tables should include colored codes exceedances, (as an illustration exceedance of ERL and ERM on the same table). The tables and figures must be submitted as part of the response to comments. Also, please include the results from the Phase I, II and other investigations into Section 5, 6 and 7 of the report.

Response: While it is understood that having all the data compiled together would be a convenient summary of information, it is accepted practice to provide separate reports for separate investigatory efforts. This is especially practical when the datasets are as large as the one for this site. In this case, one report (the Phase 1 RI) identified and targeted gaps for a later report, which is the Phase 2 RI and BERA.

However, in order to address the request, Appendix D, which presents the analytical data sets for each media, will be expanded to provide all the data available for the site that is representative of the existing condition (i.e. samples of soil excavated will not be included). Also, please note that the sediment chemical data from Phase 1 is already incorporated into the distribution maps of the representative chemical constituents, although it was realized that EPA sediment data from 2003 was inadvertently left off these maps, and they will be revised accordingly.

The use of constituent – specific tag maps as were done for the Phase 1 RI report was carefully evaluated. It is our recommendation that this Phase 2 RI and BERA which targets certain media and contaminants and seeks to resolve ecological risk to those contaminants would not benefit from large - scale tag maps. However, sediment contaminant distribution maps were expanded and revised significantly, and proposed revised example figures for Section 4 are attached to this response summary. The use of these, along with the revised

Appendix D will provide all the information in one document as requested. If the review parties need to see similar information for additional contaminants not yet mapped, additional maps for those can likely be generated quickly.

Regarding the text, the previous data are properly referenced and described already, and adding further discussion of the old data into text discussions will result in lengthy unusable text.

14. Page 4-5, Section 4.1.1, Southwest Corner of Building 32; Paragraph 3.

The report notes that data from the Phase I & II RI was used to evaluate whether a release is present near the former locations of underground storage tanks at Building 33 (Power Plant) and other locations. Releases at these locations were discovered during the removal of the USTs, as well as during other removal actions conducted at the site. As these samples and observations (observations of product or contaminated soil) would have been taken when the releases were discovered, they must be included in the report and used as part of the overall evaluation. Please include figures and tables with this information (the figures should also include the Phase 2 sample locations as well as the historic locations), and include an appendix with the appropriate sections from the original reports.

Response: Please refer to the response to Comment #9, above. The purpose of this sampling effort was to determine if a possible ugradient source was present. The data provided by the borings indicates that there is not such a continuing source. A summary of the tank closure documents is not needed for the conclusions of this Phase 2 RI.

15. Page 4-5, Section 4.1.1, Southwest Corner of Building 32; Tables 4-11/4-12.

These tables contain the analytical data from the soil samples collected at the site. Exceedances of regulatory criteria are shaded. A review of the tables reveals that all exceedances are not shaded, (i.e., TPH in TP-09, TP-10a, etc). Please review these tables and shade the areas which exceed criteria. Also, please produce a figure(s) which depicts the sample locations from the Phase I /II and other investigation/remedial actions which exceed criteria. The concentrations observed in the exceedance should be included in the figure. Please submit this figure as part of the response to comments.

Response: The data has been checked and the tables will be revised. In the instance of the TPH, the PAL was missing from the criteria column, so the highlights were not generated. This error will be corrected.

16. Page 4-5, Section 4.1.1.1, Building 32 Coal Pile Area.

The report notes that due to the lower concentration of contaminants detected in the shallow soil samples, the street drain and not the coal pile is the source of TPH observed in the area. A review of the data reveals that elevated levels of TPH (up to 3500 ppm) were found in the 0-2 foot interval in test pits dug adjacent to the street on the west side of Building 32. TPH contaminated surface soil was also observed in this area during the previous removal actions. As such, it appears that there is a surface source of TPH which required remediation in this area. In regards to the storm drain system, it is acknowledged that elevated levels of TPH, which also require remediation, exist in this area. Both of the above

should be noted in the report. Finally, based upon the information presented, it does not appear that samples were collected from the coal pile. If this is the case, the report should refrain from making statements concerning the potential sources at the coal pile and simply state that samples were not collected from this area. Please modify the report to note the above.

Response: The data collected indicate a surface release from the building is the source of the PAHs and TPH that was captured by the storm drain, also referred to in the Phase 1 report as the trench drain on the west side of Building 32. Coal was found in some soil samples (SB404 2-4 feet and SB405 1-2 feet) and as such, it is believed that the samples were taken on the downgradient edge of the coal pile. This will be clarified, and it will be noted that samples were not taken across the entire coal pile for the characterization of the entire feature.

17. Page 4-6, Section 4.1.1.2, Building 42, Former Riggers Storage Building.

Note the Riggers Storage Building is Building 52, please revise section heading.

The soil borings collected in this area were installed in the former footprint of Building 41, the former Torpedo Building. Soil borings were not located at the former location of Building 52. If the samples were not collected either through or on either side of the removal actions conducted at Building 52, the report should note this and state that additional samples will be collected in this area during predesign studies.

Response: It is acknowledged that the building identifiers for buildings 41 and 52 are in error in several portions of the report, and will be corrected. The excavation area around building 52 was deemed completed in 2002, a summary of the post excavation data will be provided in Section 1 and referenced in this section.

18. Page 4-10, Section 4.2, Sediments; Whole Section.

Please include large fold out figures similar to those in the Phase I RI in which all contaminants which exceed ERLs and ERMs are color coded (Phase I & II), and also include the toxicity test results, tissue sample results, and EQM. It is recommended that separate figures be created for each area, Stillwater Basin, Northeast Area, Northwest Area, and Reference Area (this will provide the needed space to accommodate the above information).

Response: The sediment contaminant distribution maps are revised to show evidence of ERL and ERM exceedances by date, and by depth. Refer also to the response to comment No. 13. Revised example maps are provided attached to this response summary. The revised Figure 6-3 was discussed on April 15, and it was agreed that this figure provided the information desired.

19. Page 4-11, Section 4.2, Sediments; Paragraph 3, Figures 2-2 and 4-3.

Sediment samples varied from approximately 0-4 feet across the site. As deep samples were not taken at all locations, please include a figure depicting sample locations and depths collected.

Response: The information is provided on Figure 2-3, and will be added to Figure 2-2 to show spatial orientation of deep sediment samples

20. Page 4-11, Section 4.2.1, Step 1 (2009) Sample Results; Table 4-4.

The report notes that the maximum values from the 2009 and 2010 reference samples are used in the comparison study. A review of the data reveals that the 2009 data employed the UCL in lieu of the max. Please revise the table to use the maximum observed concentration. Also, please include a table with the analytical results from the 2009 and 2010 data, and include an evaluation of the reference stations in the text (that is, note any difference between 2009 and 2010 data for the same reference locations, note any difference between sampling stations, etc).

Response: Tables 4-1 through 4-3 and 4-5 present the site data compared to the reference data. The maximum reference sample concentration is used to compare to the site data. 95% UCL data are provided for informational purposes because the 95% UCL values are used in Section 6.

Table 4-4 presents a comparison of the 2009 (six samples) and 2010 (three samples) reference data to all nine reference data combined. These comparisons will be clarified.

21. Page 4-12, Section 4.2.1.1, Stillwater Basin PAHs; Paragraph 4.

The report notes that the potential source of PAHs observed in the Stillwater Basin would be the fuel transfer from barges on the north side of Building 32. The report should also note that the source of PAHs may be associated with the carrier oil used in the PCB transformers. In regards to fuel loading operations, there were piers located on the east and west side of Building 32 which could also have serviced fuel ships. Unless it is known that the fuel transfer pipe is located on the north side of Building 32, the report should state that fuel off loading may have occurred at the north, west, and east side of Building 32.

Response: The source of PAHs as fuel in the Stillwater area is speculative. It is acknowledged that fueling operations could have occurred in any area and the report will be revised to reflect this. Carrier oil from PCB transformers could be a source as well. Finally, PAHs are suspected as components in parts cleaning solutions historically used at the overhaul shop, and as such, transfer of these materials at the rigging platform could also be a source. All this information will be added to the cited discussion.

22. Page 4-15, Section 4.2.1.2, Northeast Shoreline; Paragraph 4.

The report notes that elevated PAHs were found in a storm/parking lot drain found on the west side of Building 32. A parking lot is not depicted on the figure in this report. Please indicate the location of the parking lot and include the condition maps or other sources of information which indicated that the area in question was a parking lot. If this information is not available, please eliminate references to parking lots.

Be advised that Section 3.5 of the Phase I RI report notes that the system was a trench drain designed to collect road runoff and groundwater from the southwest corner of Building 32. As such, in this and any other section of the report which references this drain, please

refer to this drain as an outfall for the trench drain which collected groundwater and road runoff.

Response: The term "parking lot" is erroneous and will be struck. The area in question is the paved roadway between the southwest corner of Building 32 and the former coal pile. The clarifications requested will be made.

23. Page 4-16, Section 4.2.1.2, Northeast Shoreline; Paragraph 4, Last Sentence.

The report notes that the difference observed between the Phase I and Phase II sample results is attributable to the fact that Phase I targeted source areas while Phase II was designed to determine the outer extent of contamination. In regards to samples which were designed to be collocated, the difference may be attributable to the difficulty in obtaining a sample from the same location during the different sampling events. Therefore, please add the following after the last sentence of this paragraph: "The difference observed in samples which were designed to be collocated may be attributed to the difficulty in obtaining marine samples from the same location." (Note: This comment also applies to the Stillwater and Northwest Areas).

Response: The requested clarification will be made, in relation to the heterogeneity of marine sediment in an active shoreline as well as the precision of survey instrumentation used during different events.

24. Page 4-29, Section 4.2.2.5, Reference Locations (Potters Cove and Cranston Cove); Whole Section.

The report notes that the maximum detected PAHs occurred at JPC-02 (Potters Cove) and that all of the maximums for the PAHs also occurred at this location. Similarly, 16 of the maximum detected metals were also observed at JPC-02. Considering the observed distribution of contaminants and the fact that the maximum values were observed at one background station, the report must evaluate the suitability of this station and whether it should be retained as a reference station. Please include this evaluation in the response to comments.

Response: The suitability of the sample station was evaluated as part of the first step sampling effort. It was agreed by all parties at that time that the station was adequate and suitable. Differences in concentrations measured at this station between the 2009 and 2010 events are simply another example of the heterogeneity of marine sediment.

25. Page 5-1, Section 5.1, Overview of Contaminants at Site 17; Whole Section.

This section of the report deals with the results of past investigations to discuss the nature and extent of contamination at the site. Please add sections discussing the findings of the previous investigations, remedial actions, and include corresponding figures and tables.

Response: Please refer to the response to Comment 13. The information in this section is largely derived from the 2006 Phase 1 RI report, augmented using new data as described in the Phase 2 report. If there are specific items that need further evaluation, these issues can be expanded.

26. Page 5-2, Section 5.1.1, Summary of Phase I Contaminants at Site 17; Paragraph 1, Last Sentence.

This section of the report states that NAPL was not encountered in any of the monitoring wells or test pits at the site. The Phase 1 report notes that sheens and suspended product was found in the test pits. Further, product was also observed in the excavation associated with Transformer Vault 54, and the crushed stone with the perforated pit was installed for monitoring and potential future remediation. Therefore, please remove this statement from the report.

Response: The statement will be revised to be more accurate and complete, referencing the sheens, and product identified in the previous efforts.

27. Page 4-38, Section 4.3, Biota; Paragraph 1, First Sentence, Tables 4-6 - 4-10.

This report compares the tissue concentrations to PALs. Please note in the report the basis for the PALs, and explain any difference that may be observed with the PALs and values used elsewhere in the report. As an illustration, in Section 6.42 the CTL for cadmium is listed as 0.15 ppm, while the PAL is listed as 1.79 ppm (the PALs appear to be applicable to food chain while the CTLs are applicable to impacts to shellfish). Finally, as CTLs are also PALs, the report should clearly state that CTLs and PALs are used in the COCs.

Response: Tissue Project Action Levels (PALs) are presented in the SAP, and will be described in this section as well. The term "PAL" is often confused. The PAL is only a quantitative level to which the data must be reported to for the planned use of the data, and is selected to give the laboratory a reporting limit. The PAL is not a "remediation action level", meaning remediation will be undertaken if the concentration is exceeded.

28. Page 4-38, Section 4.3, Biota; Paragraph 1, Last Sentence.

"PALs were converted from those presented in the SAP assuming a 85% moisture of the body tissue."

Based upon studies by Ricciardi & Bourget, 1998, an assumed moisture content of 5% for blue mussel (*Mytilus edulis*) and 14% for green crabs (*Carcinus maenas*) should be employed. Please incorporate these values into the report.

Response: Please see the Navy's response to EPA Comment No. 9 for the basis of the percent moisture values and how they were used. The article referenced in the comments was reviewed. The Navy does not believe that the values cited in the comment are correct. It appears that the cited values are actually ash-free dry weight (AFDW) to weight-weight (WW) conversion factors. The WW values used to develop the conversion factors included the shells for mollusks, which is not representative of how the site samples were analyzed. For some species, including blue mussels (but not green crabs), a shell-free dry weight (SFDW) to WW conversion factors was calculated. However, the SFDW was obtained by drying specimens to constant weight in an oven at temperatures of 60°C to 110°C. It is not sure whether this is a standard method for determining percent moisture values. That being said, Dr. W. Edward Johnson from NOAA's National Status and Trends Program was contacted via e-mail about typical percent

moisture values in mussels. He responded in an e-mail on March 16, 2011 that "As a general rule mussels are about 80% moisture. A good review of the subject is found here: http://www.int-res.com/articles/meps/163/m163p245.pdf", which happened to be the same article cited above by RIDEM. In addition, the following article indicates that mussels are about 88% water: http://ccma.nos.noaa.gov/publications/nst_natldistrib.pdf.

Therefore, the assumption of 85% moisture is appropriate.

29. Page 4-45, Section 4.3.7, Stillwater Basin; Whole Section.

A review of the data indicates that the lipid concentration in the Stillwater Basin is 0.7-1%, which is lower than the lipid concentration (3-5%) found in other areas, such as the reference locations. This is indicative of potential stress on the organisms due to the presence of contaminants. This should be noted in this section of the report and factored into a weight-of-evidence approach for overall risk.

Response: At the request of RIDEM, the Navy conducted an evaluation of lipid concentrations compared to PCB concentrations in clam tissue. This species and contaminant combination was selected due to the best abundance of data across each area. A summary chart is presented as Figure 1, attached to this response. While this chart could be interpreted to indicate a possible correlative effect of PCB concentrations and reduced lipid concentrations, it would be an improper leap of judgment to conclude that the PCBs are reducing the lipid content. This evaluation does not consider several other physical factors that may affect lipid content, including date of collection, and abundance of local food sources. In addition, the Navy could find no documentation in the literature to indicate that PCBs, or other chemicals, have the potential to decrease the percentage of lipids in clams. While this is an interesting observation, it does not change the conclusions that there is risk to the ecological system from contaminants in the sediment. If the Navy was to consider this a quantitative line of evidence, additional work would first need to be done to determine if other factors are affecting the lipid content and whether chemicals cause the percentage of lipids in clams to be reduced.

30. Page 5-2, Section 5.1.1, Summary of Phase I Contaminants at Site 17; Paragraph 2, 3rd, 4th and 5th Sentences.

"The PCB removals were conducted to address transformer buildings and former releases from those buildings; however, once the cleanup goals for that project were met, the efforts were terminated. The concentrations of PCBs detected in the soil were below the cleanup goals for the previous removal actions, and those actions did not pursue PCB contamination to the marine sediment. While PCBs were not detected in groundwater collected from monitoring wells, previous samples have noted low concentrations (less then 1.0 ppm) in pooled water in excavations conducted at PCB removal areas."

Please be advised that PCBs were found at other structures besides the transformer buildings. Further, it was anticipated that additional investigations or actions would be required at certain locations. As an illustration, at Building 56, TPH contaminated soil was observed and samples were collected from this area, as it was anticipated that additional

action would be carried out at a future date. At Building 54, PCB sheens were observed in the northern side of the excavation. After the removal action was completed, a PVC pipe was installed in this area to allow for additional monitoring and/or removal. In regards to the sediments, a limited removal action of sediments was conducted and the cleanup objective was 1 ppm. Finally, 430,000 ppm of PCBs was found in an oil/water sample collected from groundwater during one for the removal actions. Therefore, please modify the above as follows:

"The PCB removals were conducted to address transformer buildings and other structures or locations where PCB releases were observed. The removal actions were terminated once the cleanup objective (10 ppm for soil/1 ppm for sediment) was obtained. Additional action was deemed necessary at certain sites such as Building 56, where TPH contamination was found and samples were collected in anticipation of future action. At Building 54, PCB sheen was observed flowing in at the northern side of the removal action. A PVC pipe was installed in order to facilitate additional investigation and/or action. While PCBs were not detected in groundwater collected from monitoring wells, previous samples have detected levels up to 430,000 ppm in groundwater samples containing free product collected during the removal action."

Response: The information requested will be checked and additional detail will be presented in Section 1.4.4. A summary similar to the above will be added to the paragraph, though it may be modified slightly for clarity. It is important to note that the highly contaminated groundwater was removed, and the post excavation samples met the cleanup objectives. The observations of the PCB sheen will be clarified.

31. Page 5-5, Section 5.3, Fate and Transport of Selected PAHs; Paragraph 1.

The report lists the potential source of PAHs found in the sediments. The focus appears to be on PAHs associated with fuel products from sumps or fueling operations via direct discharge or erosion/overland flow. The report should also note that PAHs may also be associated with the release of PCBs. Further, as TPH/PCB contaminated groundwater has been found at the site, groundwater transport is also a possible mechanism; this should be noted in the report.

Response: This section of the report provides the fate and transport processes available to the contaminants found. The possible sources of PAHs are noted for clarity, and the requested information will be provided. However the data available do not suggest that groundwater transport of PCBs and TPH is a significant pathway in the areas of interest, and the revisions will note this.

32. Page 5-5, Section 5.4, Fate and Transport of PCBs; Paragraph 3.

Elevated levels of PCBs were found in groundwater samples containing free product, which represent a potential migration pathway. This should be noted in the report

Response: Please reference the response to comment 30 above. The past presence of PCBs in free product will be referenced as described in Section 1.

33. Page 5-5, Section 5.4, Fate and Transport of PCBs; Paragraph 6, 2nd sentence.

Please remove this sentence from the document. Please refer to comment #3 mentioned above.

Response: Please refer to the response to comment #3.

34. Page 5-6, Section 5.5, Fate and Transport of Selected Metals; Paragraph 6, Last sentence.

The report notes that the metals concentration in the Stillwater Area increases with depth and those metals are becoming encapsulated by sediments. At other locations on the site, the concentration of metals decreases with depth; therefore, the report should refrain from making general statements and simply note that the concentration of metals may increase or decrease with depth depending upon location. Further, the report must state that the majority of the deeper cores were collected in the Stillwater Area and there is limited data for the other areas.

Response: The paragraph will be revised as follows:

> "In the Stillwater area, where surface and subsurface sediment samples were collected at most locations, metals were found at higher concentrations at depth than at the surface, indicating past releases and sediment deposition covering those releases. However, such data was not available to support such a finding at the Northeast shoreline where a less depositional environment is present. Therefore, concentrations of metals at the site may increase or decrease with depth depending on location and the physical forces acting on the sediment."

35. Page 5-9, Section 5.5.4, Lead.

The concentration of lead observed at sampling station 304 F (and its equivalent) during the Phase I and Phase II investigations was 21,200 and 21 ppm respectively. Exceedances of the ERL for lead occur at several locations (SD517, SD304, SD438, and SD531) that are within 60-170 ft from the "hotspot". Exceedances of ERL and ERMs were also observed at other locations at the site. Further, toxicity was evident at certain locations. Therefore, lead should be retained as a COPC for the FS.

Response: The last sentences will be revised so as to not indicate that lead is being dismissed.

36. Page 6-1, Section 6.1, Introduction; Paragraph 2.

The report notes that COPCs were identified in the Phase I report. Please attach a table with the list of COPCs. The report should also include a table with the final list of COPCs and justification statements for any contaminants which were deleted or added to the COPCs from the Phase I report. In addition, please add the following to this section of the report: "Based upon the results of the BERA, the list of COPCs could be expanded or reduced."

Response: The list of COPCs from the Phase I report will be added to the text in the form of a bulleted list instead of a table. Adding a table would require renumbering all of the tables in Section 6 and the list of COPCs from the Phase I report is relatively

small. The list will be added to Section 6.2. The suggested sentence will be added to Section 6.1.

37. Page 6-7, Section 6.3.1.1, Sediment; Paragraph 2.

Please note that the Buchman (2006) tabulated sediment screening level reference has a newer version (2008). Please review and incorporate into the report as necessary.

Response: Buchman (2008) was used as the source of some screening tables. The citation will be corrected on p. 6-7.

38. Page 6-17, Section 6.4.1.1, Summary of Sediment Toxicity Testing; Paragraph 3, Appendix G.5.

Appendix G.5 contains the results of the toxicity tests. The appendix notes that an evaluation of normality will be conducted and either a parametric or a non-parametric test will be employed. However, based upon the information presented, it does not appear that the critical values for all tests were incorporated into the tables and the tables themselves require some clarification in the footnotes. As an illustration, for the Sharpio Wilkes Test for normality, the test stat is provided (assumed to be the site specific results); however, the critical value is not. Also, the p value is noted but not explained. In regards to the final statistical test which was employed, the report must clearly state whether the test was either a parametric or a non-parametric test, the critical values for the test, the critical value obtained from the analysis of the site samples and any limitations associated with the test (minimum number of samples, etc). Please add the appropriate information to the text and appendix.

Response: The toxicity reports were prepared by EnviroSystems and are presented as received; therefore, additional details are not available. The laboratory was contacted to provide a more detailed response, which will be included in the revised document.

39. Page 6-17, Section 6.4.1.1, Summary of Sediment Toxicity Testing; Paragraph 3, Appendix G.5.

In the ANOVA, it is not clear whether the site sample locations which were less toxic than the background were incorporated into the overall data set of site samples when assigning toxicity for an individual site sample station. Please note whether this was done. It is recommended by the State that these sample locations not be employed as part of the overall site sample data set.

Response: Each site sample was compared separately to each reference sample. The entire site sample set was not compared to the entire reference sample set. The first sentence in the fifth paragraph in Section 6.4.1.1 will e changed as follows: "The survival, growth, and reproduction rates of the amphipods ***in each sample*** were statistically compared to the survival, growth, and reproduction rates in ~~the~~ ***each*** reference sample (analysis of variance [ANOVA] p value less than 0.05) to determine whether the survival, growth, or reproduction rates of the amphipods in ***any of*** the site samples were lower than the survival, growth, or reproduction rates in the reference samples."

40. Page 6-17, Section 6.4.1.1, Summary of Sediment Toxicity Testing; Paragraph 5.

The report notes that an ANOVA was used to compare the site sample group to the reference group in order to assign toxicity to a particular sample location. There are a number of concerns associated with the toxicity results, such as the observed variance of the toxicity test in the triplicate analysis, combining results from less contaminated areas with more contaminated areas, etc. An approach which would resolve many of these issues would be to state that an affect of 25% on the endpoint would indicate positive toxicity (as an illustration, in terms of survival, samples with survival of 75% or lower would be considered toxic).

Response: Please see the Navy's response to RIDEM Comment No. 39 regarding the fact that each site sample was compared separately to each reference sample. The entire site sample set was not compared to the entire reference sample set. Please clarify RIDEM's concerns with the observed variance of the toxicity test in the triplicate analysis. The Navy does not agree that assuming an effect of 25% on the endpoint would indicate positive toxicity because the common standard is a statistically significant difference from reference samples. However, the Navy indicated in their response to EPA Comment No. 15 that they will attempt to examine the data differently, such as ranking the samples in descending order of toxicity in order to make sense of the contamination and toxicity data.

41. Page 6-20, Section 6.4.1.2, Development of Dose-response Plots; Paragraph 2, Table 6-16.

This section of the report deals with the development of the ERM-Q. These quotients are based upon ERMs, which translates into areas with a high risk. Values in-between ERLs and ERMs are considered intermediate risk. As such, ERL-Qs should be developed using the same protocols used for ERMs and incorporated into the overall risk assessment.

Response: The possibility of development of ERL-Qs was discussed on April 15, 2011. It was recognized that a comparison to the ER-L values were presented in Table 6-15 in the report, but the ERL-Qs were not actually calculated. The table will be revised to include hazard quotients based on the ER-L, but a mean ERL-Q, as was calculated for the ER-M will not be calculated. Also, no decisions (including the development of NOECs or LOECS) will be based on the hazard quotients based on the ERL because this metric is not used by the authors that developed the ERL and ERM (see attached articles). The reason is that according to Long et al., (1998, attached) ERLs "are intended to represent chemical concentrations toward the low end of the effects ranges, that is, below which adverse biological effects were rarely observed." The purpose of the ERM-Q, however, is to try and predict the toxicity of samples based on the chemical concentrations. That is the reason that they were used in the BERA.

42. Page 6-21, Section 6.4.1.3, Development of NOECs and LOECs; Paragraph 1.

The report states that contaminants which exceeded ERLs in one or more sample stations selected for toxicity were evaluated in the NOEC and LOEC process. A review of the information provided in the tables indicates that a number of contaminants which met these criteria were excluded. As an illustration, individual PAHs which exceeded the criteria and

had toxicity tests were not included in the evaluation. These and any other contaminants must be included in the evaluation. Please modify the report accordingly.

Response: It is common to evaluate risks from total PAH groupings (i.e., HMW PAHs, LMW PAHs, and total PAHs) as opposed to individual PAHs. The same is true for PCBs. As stated in page 7-18, ecological risk for PAHs are based on the total PAH value; therefore, the individual PAHs were not included in the evaluation. NOEC and LOECs for individual PAHs are presented in the new Table 6-31. Note that the Navy is not suggesting or recommending the PRGs should be developed for individual PAHs, though, because the distribution in concentrations for the individual PAHs follows the same pattern for the total PAH groupings. This will be clarified in the end of the first paragraph of section 6.4.1.3 and in the uncertainty analysis section.

43. Page 6-21, Section 6.4.1.3, Development of NOECs and LOECs; Paragraph 4.

The report notes that NOECs and LOECs were not developed for metals where toxicity tests were conducted due to the limited number of sampling stations. These sampling or distribution limitations should not be used to state that values will not be generated for these contaminants. Therefore, please propose values for these contaminants.

Response: Section 6.4.1.3 does not state that NOECs and LOECs were not developed for metals where toxicity tests were conducted due to the limited number of sampling stations. The text states that "Although NOECs and LOECs were developed for several of the metals, there is considerable uncertainty in the LOECs so they will not be used to evaluate the data. The primary reason is that elevated levels of metals were not found across the site; elevated levels were only found in a few locations. Although several of these locations were re-sampled for toxicity testing, the elevated concentrations were not found. This indicates that the elevated concentrations are not wide-spread in those areas, and as such are not likely to present a significant concern for ecological receptors." The reason for the uncertainty is discussed in the next paragraph. To clarify, the words "(as discussed in the next paragraph)" will be added before the comma in the first sentence of the referenced paragraph.

44. Page 6-22, Section 6.4.1.3, Development of NOECs and LOECs; Whole Page.

The report notes that there were a number of problems associated with the development of dose response curves (in a number of cases, curves could not be developed and/or the results were inconsistent) and the associated of LOELs and NOELs. Based upon the proposed values and the elimination of certain contaminant types and/or contaminants, it is apparent that additional evaluation is needed. Perhaps a risk ranking, weight of evidence, or similar system could be employed. Please evaluate this option and implement the suggestion made in the other comments and submit new values as part of the response to comments.

Response: As indicated in the response to EPA general comment No. 2, the Navy revised Table 6-27 to include additional samples in the "toxic" grouping. In addition, a weight of evidence approach for evaluating ecological risk was discussed on April 15 and then again by the Navy internally after the call. However, because the sampling plan was not designed to support a WOE approach, and because

data lines were not developed to support it, it is not recommended to make this change at this time. The group discussed it on June 13, and it was clarified that four lines of evidence were evaluated in the BERA and of those, two are used to determine if there is overall risk to the ecosystem in an area, or for the site as a whole (tissue testing, and food chain modeling). Neither indicated presence of unacceptable risk. The other two lines of evidence, (chemistry and toxicity tests) were evaluated in the BERA and can be used to develop point-specific risk and can be used to calculate cleanup goals. The WOE approach was discussed during the development of the SAP, but it was determined at that time that other test endpoints are not available that would provide reliable, site specific data that can be used to derive cleanup goals for the site.

45. Page 6-23, Section 6.4.2, Biota; Whole Section, Table 6-6.

It appears that certain contaminants, such as the majority of PAHs detected in tissue samples were not retained due to their low frequency of detection. In the Stillwater Area, the frequency of detection is 1 out of 1. Further, the majority of PAHs were detected in this sample. The limited sampling that was conducted in that area shouldn't be considered as grounds for rejection. Therefore, please retain PAHs for samples collected in the Stillwater Basin.

Response: Certain contaminants, including PAHs, were eliminated because of low frequency of detection. The Navy agrees that PAHs were detected in the only 2005 tissue collected from the Stillwater Basin, but risks from PAHs were evaluated in the Phase I RI. The project team agreed that the Phase 2 tissue samples did not need to be analyzed for PAHs. Therefore, because no new data are available, there was not a need to re-evaluate risks from PAHs in tissue, especially when PAHs were only detected in 1 of 21 samples, and PAHs are not retained for further evaluation.

46. Page 6-23, Section 6.4.2, Biota; Whole Section, and Table 6-6.

PCBs were detected in high frequencies and were retained as COCs for biota samples. This section of the report however does not include a discussion of PCBs. Please modify the report to include a discussion of PCBs. In regards to Table 6-6, it appears that the CTL employed for PCBs was based upon guidance from the Oregon Department of Environmental Quality. The value employed (930 ppb), which applies to both fin fish and shellfish, was obtained by multiplying the ambient water quality criteria by a bioconcentration factor. The discussion in the text should state why this value was employed in lieu of values that were obtained from literature studies similar to that performed for the other COCs. Please include this discussion in the response to comments.

Response: As presented in Section 6.3.4.2, CTLs from ORDEQ (2007) are conservative values meant to protect 95 percent of all aquatic organisms. Because none of the PCB concentrations exceeded its CTL in Section 6.3.4.2, they were not discussed further in Section 6.4.2. A literature search was only performed if CTLs were not available in ORDEQ (2007) or for chemicals that exceeded the conservative value. Additional discussion between the risk assessors will be required to reach agreement on the approach proposed by RIDEM.

47. Page 6-24, Section 6.4.2, Biota; Paragraph 1, Last Sentence.

This sentence notes that CTLs were not available for the listed organics and inorganics. It is not clear why this subset of contaminants is listed. That is, does this represent all of the contaminants found above CTL and/or reference concentrations? Please explain. The report should include a table with a list of all contaminants in tissue samples which exceed reference concentrations (including contaminants not tested for in reference) by location, Northeast, Northwest, Stillwater Area, etc. The range, max and average concentration should be included. The table should also include the species specific tissue residual value and if one is not available the surrogate value (the max and average value should be included). In the case of surrogates, in order to gauge applicability, the report should compare the sensitivity of the surrogate to the site-specific organism (that is comparing the values for contaminants in common between the site-specific species and the surrogate to see if they are similar). Contaminants with tissue concentrations above, species specific, similar sensitive surrogate species, and reference (for contaminants with no TRV), should be considered indicative of impact and carried forth in the evaluation. All of the above should be discussed in the text. Please include this information in the response to comments.

Response: Section 6.3.4.2 presents the initial screening of tissue concentrations to the CTLs. The second paragraph of this section states: "As the first part of this screening, the maximum chemical concentrations in the site samples were initially compared to the maximum chemical concentrations in the reference samples to determine whether the chemical were bioaccumulating in the site tissue samples. If the maximum chemical concentration in the site sample was less than the maximum chemical concentration in the reference samples, it was determined that the chemical was not site related in the organism and that chemical was not further evaluated in the ERA for the tissue samples. If the chemical concentration in the biota sample was greater than maximum chemical concentration in the reference samples, the chemical was initially selected as a COPC for further evaluation. Also, chemicals that were detected infrequently in one or two samples were eliminated from further evaluation." Therefore, the subset of contaminants listed in Section 6.4.2 is the group of contaminants that were retained as COPCs from the initial screening in Section 6.3.4.2.

Tables 6-5 and 6-6 include a list of chemicals in the tissue samples compared to the reference concentrations (including chemicals not tested for in reference samples) by location. The range and maximum concentrations are presented in those tables, although the average concentration is not. The Navy does not believe that it is necessary to include the average concentration in the table because the table is meant to present a comparison of the maximum detected concentrations in the reference samples to the maximum detected concentrations in the site samples. Similar information is presented in Tables 4-23 through 4-31.

The Navy does not believe that it is necessary to include the species-specific tissue residue in the Section 6 tables, because the available data are discussed in the text (Section 4) of the report. Toxicity data are not available to conduct the comparison that RIDEM is requesting. That was the concern the Navy raised when RIDEM requested this evaluation be conducted during the development of the SAP

48. Page 6-24, Section 6.4.2, Biota; Paragraph 2.

The report notes that tissue residual values were not found for a number of contaminants using the two cited references. Please indicate whether the EPA-cited reference is ECOTOX (integrates AQUIRE, PHYTOTOX, TERRETOX). If not, perhaps this reference can be of use. http://www.epa.gov/med/Prods_Pubs/ecotox.htm

Response: The EPA-cited reference is for Jarvinen, A.W., and G.T. Ankley. 1999. Linkage of effects to tissue residues: Development of a comprehensive database for aquatic organisms exposed to inorganic and organic chemicals. SETAC Press, pp. 1-358. http://www.epa.gov/med/Prods_Pubs/tox_residue.htm this reference will be added to the reference list. The website above will be reviewed to see if additional data is available for mussels/crabs. ECOTOX does not contain any tissue residue values.

49. Page 6-24, Section 6.4.2, Biota; Paragraph 4.

The report should include a discussion of 2-methyl phenol which was detected in a number of samples. The discussion should note whether this is a metabolite of PAHs, note it observed site concentrations with respect to background and the applicability of CTLs. The report should consider retaining this compound for the overall risk evaluation along with the associated uncertainty.

Response: 2-Methylphenol is as synonym for o-Cresol. Although there are many anthropogenic and natural sources of cresols to the environment, there is not information to indicate that they are a breakdown product of PAHs. Furthermore, according to the ATSDR Toxicological Profile for Cresols, "Few data are available describing the food chain bioaccumulation of cresols. The available experimental data (Freitag et al. 1985) are consistent with estimated values obtained from regression equations which suggest that it will not bioconcentrate to any significant extent (Thomas 1982). Information concerning the potential for biomagnification has not been described, although the log Kow values are small and biomagnification is expected to be insignificant. Therefore, no data needs exist at this time." Therefore, the source of the 2-methylphenol in the tissue samples is not known, but could be naturally occurring. A discussion of this will be added to this section of the report.

50. Page 6-24, Section 6.4.2, Biota; Paragraph 4.

The report notes that benzaldehyde in tissues samples is not a contaminant of concern as it is not found in sediment samples. Please note whether this compound is a metabolite of PAHs. If it is, this should be noted in the report as a possible source of the observed concentrations and benzaldehyde should be retained as an indicator COC for PAHs.

Response: No information was found to indicate that benzaldehyde is a metabolite of PAHs.

51. Page 6-24, Section 6.4.2, Biota; Paragraph 4.

Benzaldehyde was not analyzed for in the reference samples but it was in the site samples. Please explain why this PAH and the others were not analyzed in the reference samples.

Response: SVOCs were only analyzed for in the 2005 tissue samples. Reference seamples were not collected as part of that first phase study. SVOCs were not included in the Phase 2 RI, as per the approved SAP.

52. Page 6-24, Section 6.4.2, Biota; Paragraph 4.

It is noted that the highest concentration of benzaldehyde was detected in the South Area. Please note in the report the range that was observed in the Northwest, Northeast and Stillwater Areas.

Response: Table 6-5 presents the maximum detected concentrations in the Northwest, Northeast and Stillwater Areas.

53. Page 6-25, Section 6.4.2, Biota; Bullet 2, Last Sentence.

Considering the limited TLV value for chromium (one surrogate species study) and the concentration observed at the site with respect to both the NOEL/ LOEL and background, the report should remove the last sentence and note that there are impacts with respect to chromium and carry this forth through the evaluation process.

Response: The Navy does not believe that chromium should be carried forth through any further evaluation for reasons presented in the referenced paragraph.

54. Page 7-1, Section 7.0, Summary and Conclusions; Whole Section.

As noted in the comments above, there are a number of concerns with respect to the nature and extent of contamination, associated risk, areas of concern, and COCs. As such, changes are expected with respect to the conclusions and recommendations in this section. RIDEM requests that the associated changes in this section be submitted as part of the response to comments.

Response: The revisions of the findings based on the responses to comment stated above will be incorporated into the revised document.

55. Page 7-8, Section 7.1.7, Human Health Risk Assessment; Paragraph 5.

The report states that the human health risk assessment from the Phase I RI was not revised based upon the findings of the Phase II RI. The report notes that samples collected during the Phase II RI, depending upon location, were either above, below or equivalent to the results of the Phase I RI. The report should state whether the observed difference would change the conclusion of the human health risk assessment of the Phase I RI. It is recommended that a table be created for soil, groundwater and sediment which compares the maximum, range and average concentrations observed in the Phase II RI compared to the results of the Phase I RI.

Response: It is agreed that a table to support the claim of the data falling within previously collected data is appropriate. Such a table will be developed and provided in this section. It should be noted that no groundwater or shoreline sediment were collected in Phase 2, so no changes to those data sets are described. Also note that there is no human exposure anticipated for subtidal sediment. The table will

compare shellfish and soil for the basis of the statement, and clarify that other media were not collected.

56. Page 7-8, Section 7.1.7, Human Health Risk Assessment; Paragraph 5.

The report discusses the human health risk assessment results for the site. The report should note that exceedances of the RIDEM residential criteria were observed in soils (note: criteria also includes TPH), exceedances of MCLs were also observed in groundwater at the site. Further, RIDEM Regulations, including the Remediation Regulations, are ARARs for the Site (this should also be noted in the report). Therefore, these exceedances must be carried forth to the FS as they must be evaluated as part of the remedy for the site.

Response: The role of TPH as an ARAR under CERCLA is not clear, so for this report, it will be stated that direct exposure criteria established by RIDEM Remediation regulations are considered ARARs, and as such, COCs that exceed ARARs will be identified in Section 2 of the FS report for this site. The fact that MCLs were exceeded in groundwater will be included in Section 7.1.5 and 7.1.7.

57. Page 7-14, Section 7.4, Soil; Whole Section.

The report notes that there are exceedances of TPH, organics and inorganics in sumps and vaults which are easily addressed via removal actions. The report also notes other areas where exceedances were observed. Additional areas of concern were also found in the Phase I RI and other studies, as noted in the comments above (areas which exceed RIDEM criteria and/or risk assessment values). The report should also note that the other areas with elevated levels could also be addressed via removal actions or other remedial measures. Please revise the report accordingly.

Response: Please note that areas that are to be remediated under CERCLA will be identified by comparison of data to selected PRGs in the FS. The RI report does not make this distinction. However, further discussion prior to submission of the FS may be necessary if the responses to the comments and the revised RI do not identify the areas that RIDEM believes are of concern and subject to CERCLA action.

58. Page 7-16, Section 7.4, Groundwater; Paragraph 1.

This section deals with groundwater use at the site. Please add the following after sentence 2:
"Groundwater has been used as potable water at the site in the past. The site lies within a state classified GA aquifer and the cleanup objective under State regulations is MCLs."

Response: The past use of groundwater as potable supply will be added. The exceedance of MCLs is noted in this paragraph. The cleanup objectives are not the subject of this RI.

59. Page 7-16, Section 7.4, Sediment & Shellfish; Whole Sections.

As noted in the previous comments, there are concerns with respect to the risk evaluations which were performed, contaminants of concern, and potential areas of concern, which may necessitate changes in these two sections.

Response: Revisions to other sections of the report will be carried through the conclusions as appropriate.

60. Page 7-17, Section 7.5, Recommendations.

Please add the following to the potential human health risks:

- *Residential groundwater consumption (see comment #5)*
- *TPH in soil/sediment (see comment #16)*

Please revise the potential for ecological risk from "total PAHs in sediment" to "total <u>and individual PAHs</u> in sediment". (see comment #39)

Please revise *"ERMQ from arsenic, cadmium, chromium, nickel, zinc, PAHs, and PCBs in sediment"* to *"<u>ERML and</u> ERMQ from arsenic, cadmium, chromium, nickel, zinc, PAHs, PCBs, <u>and other individual PAHs</u> in sediment."*

Response:

a) Revisions to the groundwater risk assessment, and reinterpretation of the findings are not the subject of this document.
b) This report will not speculate on risks to human health from TPH in soil or sediment. TPH is addressed as an ARAR at the FS steps.
c) Ecological risk is assessed as total PAHs in sediment, not for PAHs individually. See the response to RIDEM comment No. 42.

61. Page 7-18, Section 7.5, Recommendations; 4th paragraph.

The PAHs in sediment should be addressed individually as well as a group (Total PAHs). See comment #39 mentioned above.

Response: Ecological risk is assessed as total PAHs in sediment, not for PAHs individually. See the response to RIDEM comment No. 42.

62. Page 7-18, Section 7.5, Recommendations.

As indicated in the previous comments, there are additional contaminants of concern and therefore additional areas of potential concern, such as evaluating individual PAHs as well as total PAHs. Please revise the text in paragraphs 2-6 to reflect the modifications to the bullets as noted in comment #57.

Response: The section can be revised to cite that other actions may be appropriate to meet ARARs (i.e. MCLs and RIDEM DECs) and those actions will be defined as appropriate in the FS. However the RI is intended to identify the risks and COCs, and should not go into any detail about the actions.

ATTACHMENT C

RESPONSES TO COMMENTS FROM NOAA
PHASE 2 REMEDIAL INVESTIGATION, AND BASELINE ECOLOGICAL RISK ASSESSMENT – SITE 17, GOULD ISLAND
NAVSTA NEWPORT, NEWPORT RHODE ISLAND

COMMENTS DATED FEBRUARY 1, 2011

1. *Tables 6-12 and 6-13 shows survival often well below 80%, yet not significantly different from the Reference. This deserves further discussion. Of particular note is Station SD509 showing 63% survival and labeled as toxic when compared to the lab control yet station SD519 showing the same percentage of survival is not.*

Response: A review of Table 6 (page 10 of 24) in the toxicity test report in Appendix G.5 shows that the coefficient of variation (CV) for the three reference samples [SD524 (35%), SD525 (29%), and SD526 (22%)] is greater than the CV for the laboratory control sample (15%). The CV for SD509 (30%) is lower than the CV for SD519 (35%). Because the higher CV in sample SD519 compared to sample SD509 caused survival in SD519 not to be different than the reference stations and lab control, even though survival is sample SD509 was different than the lab control. Ad discussion of this will be added to the uncertainty analysis section of the report.

2. *The concentrations of COCs sometimes do not match the locations where most/least toxicity is found; either based on individual sample concentration or ERM-Q. See Tables 6-12, 13, 14 and 6-16. Note stations SD520 and SD566.*

Response: The Navy agrees that the there was little correlation between toxicity and chemical concentrations in sediment samples at several locations. That is why the NOEC/LOEC approach was selected. It is not clear why Stations SD520 and SD566, which are both considered toxic samples are being singled out in the comment. This was discussed on January 31, 2011, and again on April 15, 2011, and June 13, 2011. The central issue appeared to be how to determine which samples are toxic vs. non-toxic within this data set. Tables showing replicate survival results were generated and are provided attached (Tables A and B). It was noted that sample 509 was listed as toxic (at 63% survival) whereas sample 519 was listed as non-toxic (also at 63% survival). The laboratory's statistical tests were evaluated and it was concluded that this was an area of uncertainty originating with the variability between replicate chambers of the test organisms. Alternate statistics were not determined to be a good resolution. Because the uncertainty could not be resolved, it was agreed on 6/13 to republish Table 6-27 as was originally done in the Baseline ERA Risk Characterization section of the BERA report. However, a revised version of Table 6-27 (it will be Table 6-31 in the Draft Final BERA report) will be included in the uncertainty section and will categorize any sample on Table 6-13 that had lower survival, growth, or reproduction versus the laboratory control as a "toxic" sample. In addition, it would categorize a sample as "toxic" if the sample had less than 70% survival or greater as a cut-off for non-toxic samples. It was recognized during discussions that the PRG values need to be established in risk

management steps of the FS and the PRG may be based on the NOECS and LOECS from either the existing Table 6-27 or the new Table 6-31 (attached to this response summary).

3. *The duplicate sample of SD509 (SD DUP01-043010) is shown on the new Figure 6-3 (1/28/11) along with the original SD 509, however both samples are labeled the same. I would prefer to see the duplicate sample labeled correctly on the map (hence, the upper left one is the duplicate)*

Response: Revised maps will be tagged with the sample ID, instead of the location ID, and be presented for each round of data collected. Revised example maps are provided attached to this response summary. This should resolve the problem.

4. *I understand how Table 6-29 can show no LOECs at all for survival related to any chemical despite showing such toxicity at three locations (509, 520, 566); this given your Footnote 3 in the Summary of NOECs and LOECs but I would like further discussion. Could one not argue that if a station is toxic then the chemicals found at those toxic concentrations are also toxic? What has been done elsewhere? Can we use the ERM-Q against toxicity to see if there is an agreeable cutoff where above a certain value most samples are toxic? That I have done elsewhere.*

Response: Although it is not clear which Footnote 3 is being referred to, the primary reason that no LOECs were developed for the survival endpoint is that only three samples were considered toxic for that endpoint, and no maximum chemical concentrations were detected in those samples (see Tables 6-17 and 6-18). This is discussed briefly in the third paragraph on page 6-21. If the suggested argument in the comment was used, than you would potentially be identifying locations with greater chemical concentrations as toxic, even though the toxicity test indicated that they were not toxic. This is the same approach that was agreed to and used at Area A Wetland, and New London. The ERM-Q was evaluated as shown on Table 6-29.

5*. How we will rank the toxicity test endpoints - Survival, growth, and reproduction?*

Response: This may need to be a project team decision. Note that the last table in Table 6-29 (middle of page 3 of 3) presents the NOECs and LOECS developed by considering a sample to be toxic if **<u>any</u>** of the endpoints were impacted.

6. *Given #2 above, the entire NOEC and LOEC analysis is put into some doubt although the data presentation in Tables 6-17 - 6-29 is fine.*

Response: Although there is some uncertainty in the evaluations, the NOECs and LOECs developed considering a sample to be toxic if **<u>any</u>** of the endpoints were impacted for PAHs, PCBs, and the ERM-Q seem reasonable. Based on the relatively low concentrations of metals in the samples selected for toxicity testing, it is not surprising that LOECs were not developed for metals (except for nickel, which is not considered an accurate LOEC for reasons discussed in the text).

***7**. Please note the difficulty of using the Tables and Figures. Most Tables are shown with Stillwater Basin Samples at top but not all of them, making comparisons by the reader difficult (e.g. using Tables 6-14 with 6-16). In addition, the Figures were hard to use when trying to find specific chemical concentrations by locations. One must first use Figures 4-1 to 4-29 to find a location of interest (likely a red dot) then go to Figure 2-2 to find out the Station number/code and then move into the tables to learn the exact chemical concentration. Its torture! Rather, the concentrations of key chemicals should be placed in boxes around the station locations on Figure 2-2. Lastly, the station location with the highest chemical concentrations from the most recent sampling- SD501 - is not found on Figure 2-2. Most of us usually look for the potentially worst location, and it isn't recorded in the Figure!*

Response: Revised maps will be tagged with the sample ID, instead of the location ID, and be presented for each round of data collected. This should resolve the problem. Inclusion of key sample data in boxes tagged to each location was considered, and is complicated by the quantity of data available. The Navy would prefer to provide individual maps for individual contaminants if they are necessary.

ATTACHMENT D

RESPONSES TO COMMENTS FROM USF&W
PHASE 2 REMEDIAL INVESTIGATION, AND BASELINE ECOLOGICAL RISK ASSESSMENT – SITE 17, GOULD ISLAND
NAVSTA NEWPORT, NEWPORT RHODE ISLAND

COMMENTS DATED FEBRUARY 14, 2011

1. We are in agreement with and support the comments provided by EPA and NOAA with regards to contaminant and toxicity issues in the offshore areas of the site. The major agency issues were discussed during our 1-31-11 conference call, with all vested parties. We are particularly interested in follow-up statistical analysis of the toxicity tests, which EPA will be conducting. Toxicity tests that show <70% survival, even though not showing significant difference from reference areas, are problematic and need to be further reviewed/discussed. As mentioned by EPA, sample locations showing potentially toxic response in any of the toxicity test endpoints (survival, growth, or reproduction) should be considered impacted for corresponding NOEC/LOEC determinations. Several iterations of NOEC/LOEC determinations have been conducted but all endpoints combined per area, may be the most pertinent. This may significantly alter existing NOEC/LOEC determinations per area.

It would make more sense to present conservative and average terrestrial food chain scenario tables in succession, for review. We are interested review The Phase 1 RI screening level terrestrial risk assessment that did not trigger the identification of any COPCs. Please send us a CD when possible. if available. NOAA has also made some suggestions about data presentation that would facilitate data review.

Response: The conservative and average terrestrial food chain scenario tables will be presented in succession to facilitate review. The Phase 1 RI was provided to USF&W February 11, 2011.

Some section-specific comments:

2. Section 2.3, pg. 2-9: *Clarify that both spider and green crabs were analyzed, as per Table 2-6.*

Response: While green crabs were targeted for analysis, none were found at many of the target stations. However, spider crabs were abundant, and were collected in lieu of green crabs where insufficient mass was acquired. This will be clarified. The data tables reflect the species collected.

3. Section 3.1.3, pg 3-4: *The indicator species great blue heron and raccoon, should be mentioned as present on island or their use explained if they are not present.*

Response: Both the great blue heron and raccoon are believed to be present but unconfirmed at Gould Island. This will be clarified. The surrogate species used in the food

chain model were the green heron and raccoon. Please see the Navy's response to USF&WS Comment No. 6 below.

4. Section 4.2: *It may be helpful to differentiate 2005 samples from 2009/2010 samples on the figures. As NOAA noted, it would also be helpful to identify elevated locations by ID to aid in cross-checking data or provide a summary figure with pertinent elevated COCs per select location in highlighted boxes.*

Response: Revised maps will be tagged with the sample ID, instead of the location ID, and be presented for each round of data collected. This should resolve the problem.

5. Section 4.3: *It would be beneficial to denote which species are represented per location on Figures 4-26-29. It is unfortunate that PAHs were not included in the tissue analysis.*

Response: The species of biota sampled at each station can be added to the tissue results figures. PAHs were not included in the tissue analysis as per the work plan.

6. Section 6.2.4.3: *Earlier in the document great blue heron was mentioned as an indicator species.*

Response: The referenced sentence in Section 3.1.2 will be changed, as follows: "The ~~great blue heron *(Ardea herodias)* and the~~ herring gull *(Larus argentatus)* ***is the*** ~~are~~ avian species that ***has*** ~~have~~ been identified as ***a*** target receptor~~s~~ of concern evaluated in the risk assessments; however, ~~neither of these birds~~ ***it*** is ***not*** identified on the federal or state endangered or threatened species list for Rhode Island (RIDEM, 1999).

7. *It is unclear if this species was only used in the screening ERA and why it or black-crowned night heron were not carried through to this phase.*

Response: Please see the Navy's response to F&WS Comment No. 6 regarding the blue heron. The black-crowned night heron was not evaluated previously and was not carried through this BERA because the herring gull was used as the receptors species for piscivorous birds, because it was used in the SERA.

8. Section 6.3.2: *Table 6-9 should reflect sediment ingestion, in addition to food ingestion, for total exposure potential, as shown in Table 6-3.*

Response: It is not possible to account for sediment ingestion in the development of PALs for tissue because there could be different combinations of the tissue and sediment concentrations that could yield an HQ greater than 1.0. Sediment ingestion is included in the food chain models. A review of the average food chain models in Appendix G.4 shows that the dose from sediment ingestion is very low as compared to the NOAEL and LOAEL TRVs. Therefore, using the tissue PALs to eliminate chemicals from being included in the food chain model would not have changed the overall risk conclusions for piscivorous wildlife.

9. Section 6.3.4.2, pg 6-14: *Table 6-7 footnote PALs: It is unclear how PALs for crabs and clams/mussels were determined. Please clarify. Crabs exceeded PALs for several metals according to text on 4-46 and therefore should be included in food chain modeling for specific COCs. Please clarify. Additionally, PALs for crabs and clams/mussels do not compare equally to upper food chain PALs and therefore should not be used as exclusionary measures in food chain modeling.*

Response: The PALs were developed as described on Table 9. Although the Navy agrees that several chemicals in tissue exceeded PALs, as presented in the SAP, food chain modeling was only to be conducted for chemicals that were determined to be bioaccumulating in the tissue. This was to be done by comparing the analytical data to the maximum chemical concentration in the reference samples to determine whether the chemical concentrations are elevated above reference concentrations. If the chemical concentration in the biota sample is less than maximum chemical concentration in the reference samples, it was to be determined that the chemical is not site related in the organism. This comparison was presented on Tables 6-5 and 6-6 of the BERA. Most chemicals were detected at concentrations that were lower than background concentrations and there therefore not included in the food chain model. Only chemicals that exceeded reference concentrations and PALs were included in the food chain model. The last sentence in the comment is unclear and may require further discussion.

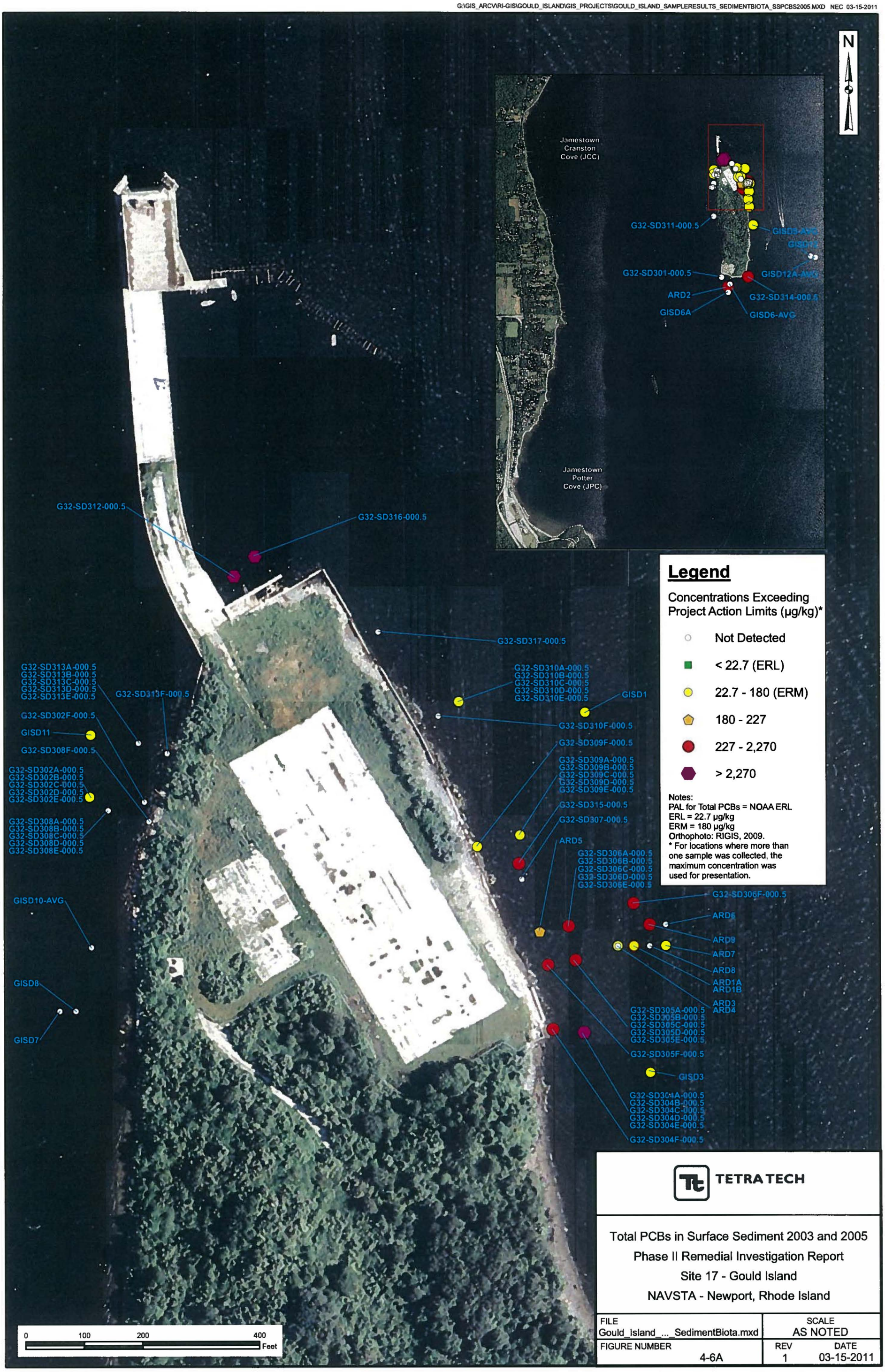
10. Section 6.3.4.3: *It is unclear why crabs are not included in the food chain uptake modeling, especially if they have different uptake potential than clams/mussels, for specific COC (i.e. Cu, Fe, Hg, etc., pg-4-45), as noted above.*

Response: No chemicals detected in the crab samples were included in the food chain modeling because all detected concentrations were less than reference concentrations and/or PALs (see last bullet on page 6-14).

11. *We will continue to examine the toxicity test data and food chain modeling and may have additional questions pertaining to those topics.*

Response: Comment noted.

G:\GIS_ARCV\RI-GIS\GOULD_ISLAND\GIS_PROJECTS\GOULD_ISLAND_SAMPLERESULTS_SEDIMENTBIOTA_SSPCBS2005.MXD NEC 03-15-2011



TETRA TECH

Total PCBs in Surface Sediment 2003 and 2005

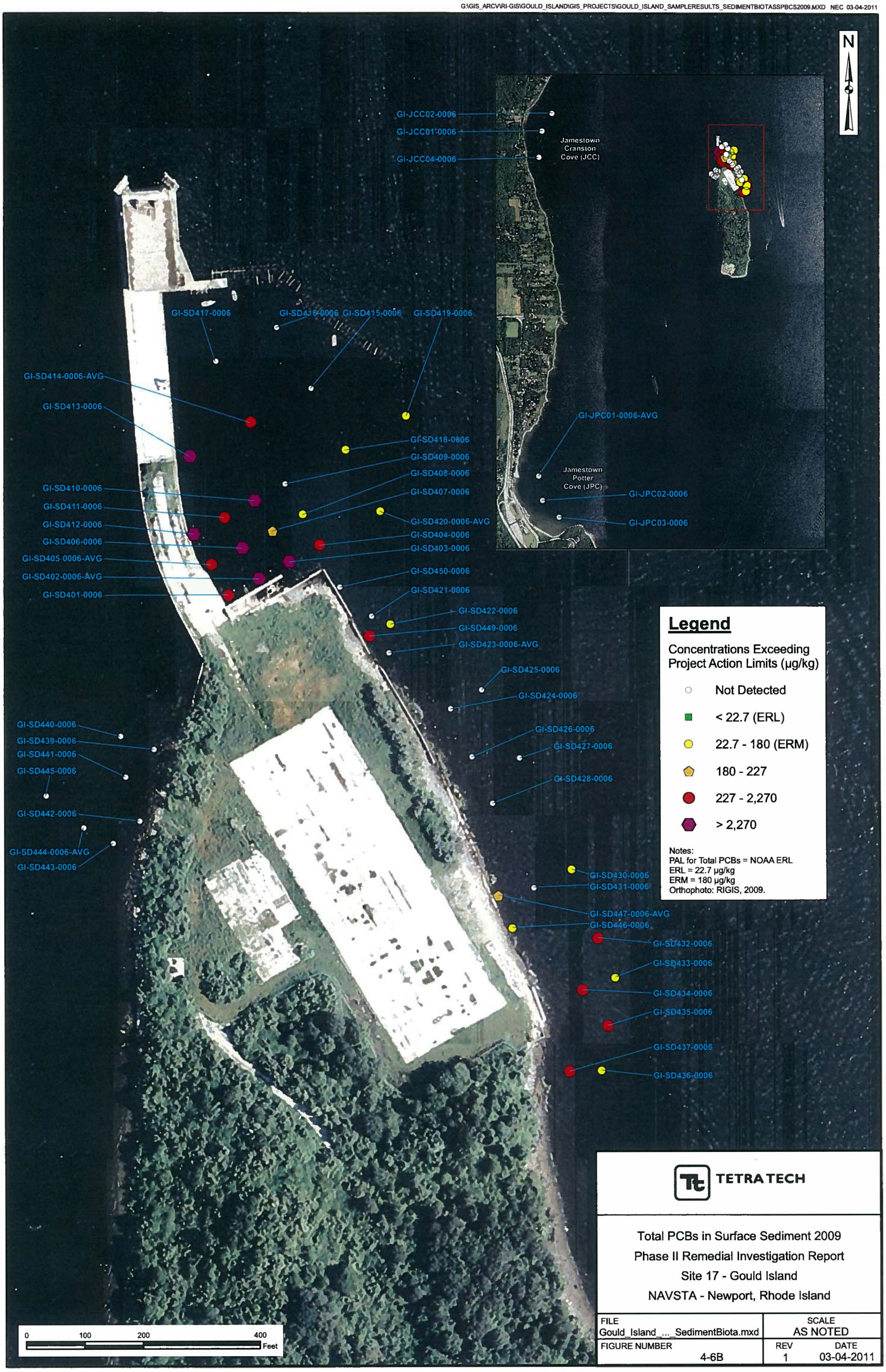
Phase II Remedial Investigation Report

Site 17 - Gould Island

NAVSTA - Newport, Rhode Island

FILE	SCALE	
Gould_Island_..._SedimentBiota.mxd	AS NOTED	
FIGURE NUMBER	REV	DATE
4-6A	1	03-15-2011

G:\GIS_ARCV\RI-GIS\GOULD_ISLAND\GIS_PROJECTS\GOULD_ISLAND_SAMPLERESULTS_SEDIMENTBIOTASSPBCS2009.MXD NEC 03-04-2011

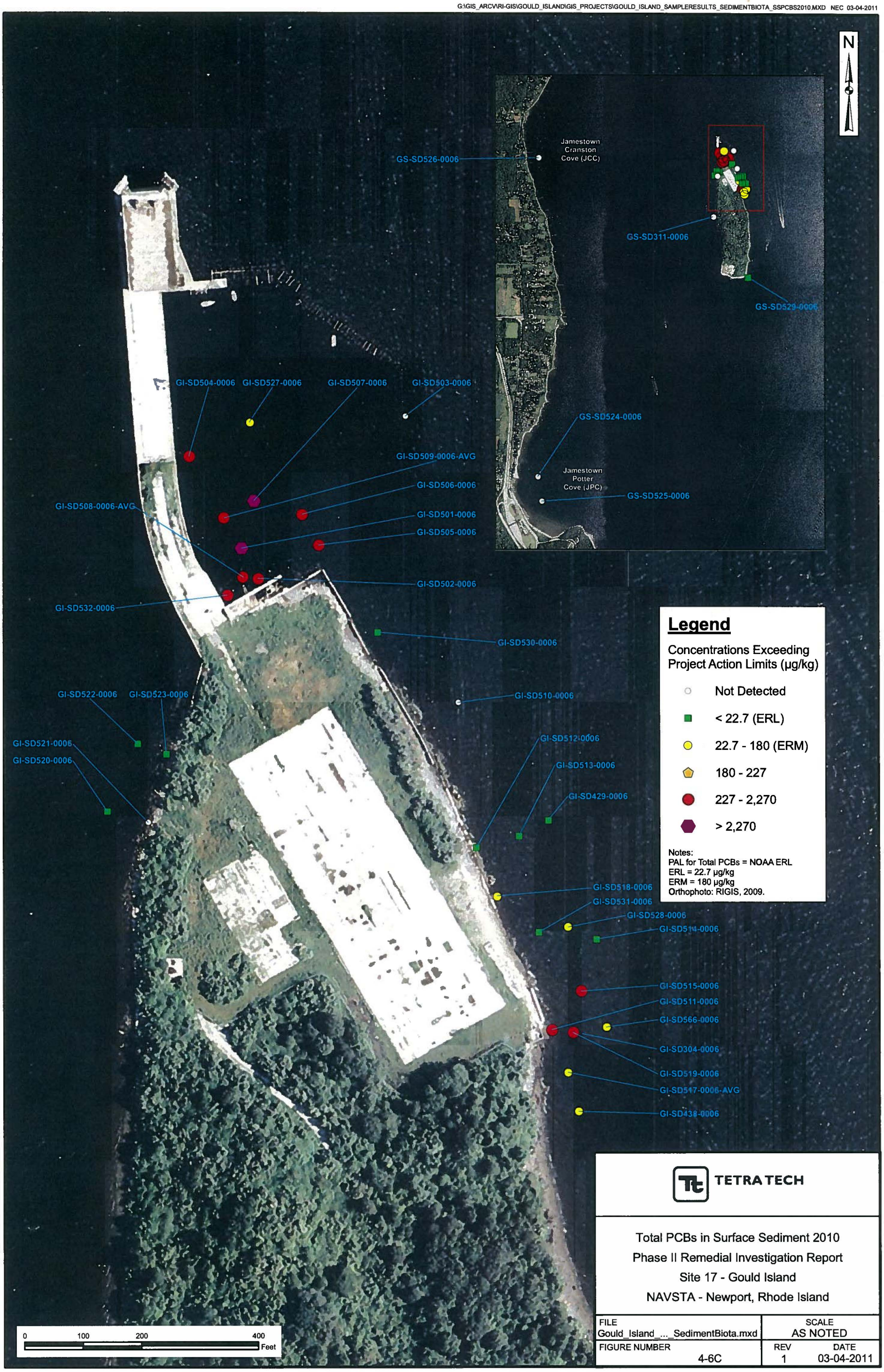


TETRA TECH

Total PCBs in Surface Sediment 2009
Phase II Remedial Investigation Report
Site 17 - Gould Island
NAVSTA - Newport, Rhode Island

FILE	SCALE
Gould_Island_..._SedimentBiota.mxd	AS NOTED
FIGURE NUMBER 4-6B	REV 1 / DATE 03-04-2011

G:\GIS_ARCV\RI-GIS\GOULD_ISLAND\GIS_PROJECTS\GOULD_ISLAND_SAMPLERESULTS_SEDIMENTBIOTA_SSPCBS2010.MXD NEC 03-04-2011

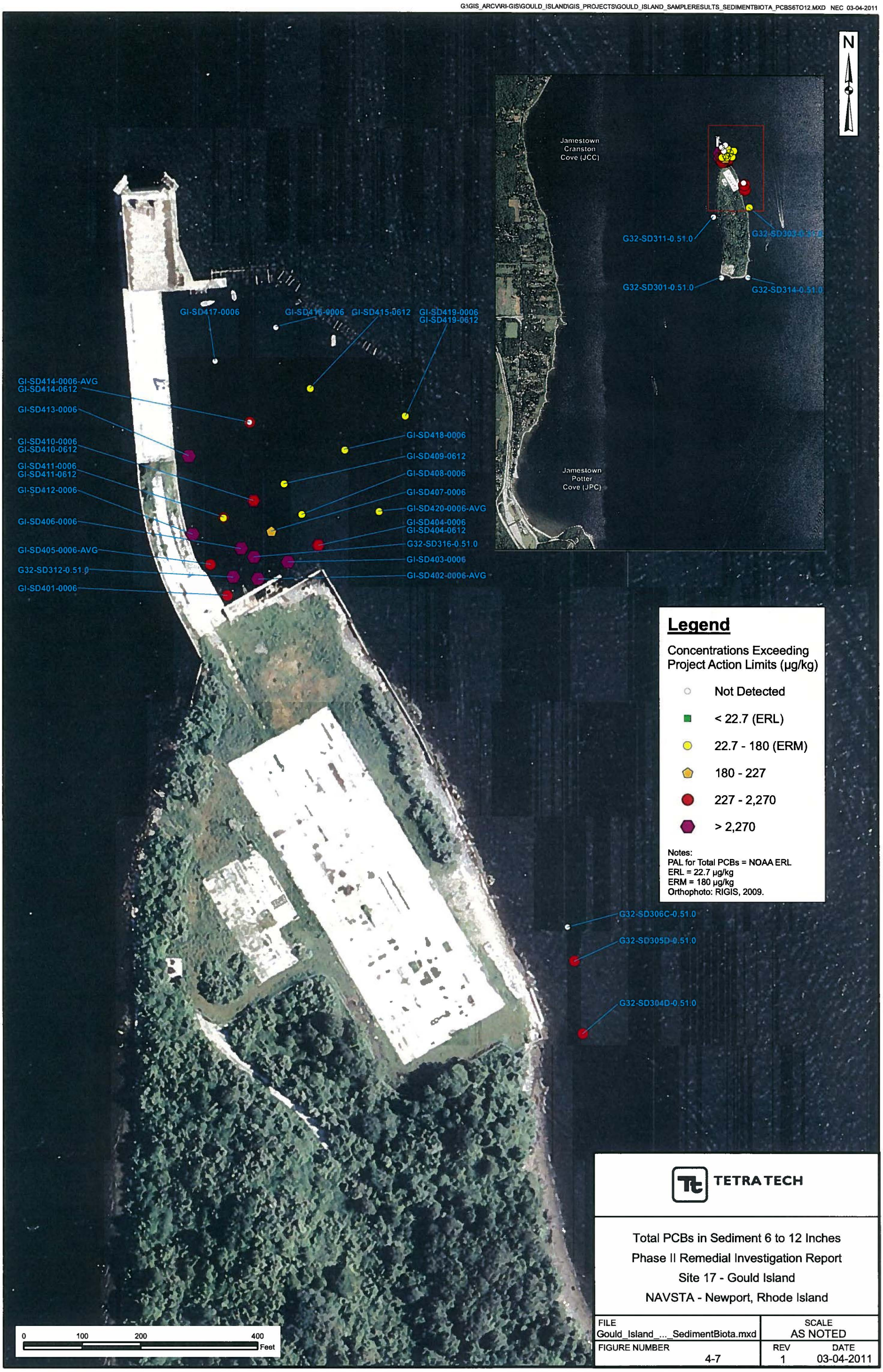


TETRA TECH

Total PCBs in Surface Sediment 2010
Phase II Remedial Investigation Report
Site 17 - Gould Island
NAVSTA - Newport, Rhode Island

FILE	SCALE
Gould_Island_..._SedimentBiota.mxd	AS NOTED
FIGURE NUMBER 4-6C	REV 1 / DATE 03-04-2011

G:\GIS_ARCV\RI-GIS\GOULD_ISLAND\GIS_PROJECTS\GOULD_ISLAND_SAMPLERESULTS_SEDIMENTBIOTA_PCBS6TO12.MXD NEC 03-04-2011



TETRA TECH

Total PCBs in Sediment 6 to 12 Inches
Phase II Remedial Investigation Report
Site 17 - Gould Island
NAVSTA - Newport, Rhode Island

FILE	SCALE
Gould_Island_..._SedimentBiota.mxd	AS NOTED
FIGURE NUMBER 4-7	REV 1 / DATE 03-04-2011

TABLE A FOR RESPONSE TO COMMENTS
Results of Leptocheirus plumulosus 28 Day Toxicity Test
Gould Island Site 17

Endpoint	GI-SD507-0006 (Lab ID 010)					GI-SD519-0006 (Lab ID 022)				
	A	B	C	D	E	A	B	C	D	E
Number of Organisms Surviving per Replicate	10	14	21	11	12	15	11	19	8	10
Proportion Survived	0.5	0.7	1	0.55	0.6	0.75	0.55	0.95	0.4	0.5
Mean Dry Weight (mg)	0.51	1.2	1.9	0.64	0.91	1.8	1.3	1.8	2.2	1.8
Mean Dry Biomass (mg)	0.26	0.86	1.9	0.35	0.55	1.4	0.73	1.7	0.87	0.89
Overall Reproduction	2.6	5.9	8.0	7.8	9.3	3.9	4.3	5.6	1.5	5.1
Offspring/Surviving Female	3.7	10.3	16.7	8.6	12.4	19.7	11.8	26.5	6	25.5

TABLE B - RESPONSE TO COMMENTS
Leptocheirus plumulosus 28 Day Survival
Gould Island Site 17

Sample ID	Laboratory ID	Number of Organisms per Replicate					Percent Survival	Survival Statistically Different Compared to:		Sample Considered Toxic Based on Any Endpoint
		A	B	C	D	E		Lab Control	Reference Samples	
Laboratory Control										
NA	Lab control	14	21	19	17	15	86	--	NO	NO
Reference Samples										
GI-SD524-0006	017	19	19	17	19	6	80	NO	--	NO
GI-SD525-0006	016	20	20	15	21	9	85	NO	--	NO
GI-SD526-0006	015	20	18	15	19	11	83	NO	--	NO
Site Samples										
GI-SD501-0006	009	16	20	16	17	15	84	NO	NO	**YES**
GI-SD502-0006	018	18	18	19	20	17	92	NO	NO	NO
GI-SD503-0006	007	13	10	17	16	15	71	NO	NO	**YES**
GI-SD504-0006	011	21	15	21	18	20	95	NO	NO	NO
GI-SD505-0006	023	19	15	19	17	15	85	NO	NO	**YES**
GI-SD506-0006	021	21	16	10	20	8	75	NO	NO	NO
GI-SD507-0006	010	10	14	21	11	12	68	NO	NO	NO
GI-SD508-0006	019	10	18	18	15	13	74	NO	NO	**YES**
GI-SD509-0006	001	18	15	11	9	10	63	**YES**	NO	**YES**
GI-SD-DUP01-043010[1]	012	6	17	16	8	7	54	**YES**	**YES**	**YES**
GI-SD510-0006	008	11	16	14	19	17	77	NO	NO	**YES**
GI-SD511-0006	002	16	11	17	16	17	77	NO	NO	NO
GI-SD512-0006	028	14	15	15	14	16	74	NO	NO	**YES**
GI-SD513-0006	029	20	22	19	16	18	95	NO	NO	NO
GI-SD514-0006	005	17	18	17	12	7	71	NO	NO	NO
GI-SD515-0006	006	20	14	19	10	12	75	NO	NO	NO
GI-SD517-0006	030	22	11	18	13	18	82	NO	NO	NO
GI-SD518-0006	014	16	12	16	20	17	81	NO	NO	**YES**
GI-SD519-0006	022	15	11	19	8	10	63	NO	NO	NO
GI-SD520-0006	031	11	8	16	10	5	50	**YES**	**YES**	**YES**
GI-SD521-0006	027	11	15	15	16	16	73	NO	NO	**YES**
GI-SD522-0006	033	18	17	17	20	15	87	NO	NO	NO
GI-SD523-0006	032	16	20	17	15	17	85	NO	NO	**YES**
GI-SD527-0006	003	17	17	15	20	12	81	NO	NO	NO
GI-SD528-0006	025	18	18	17	8	20	81	NO	NO	**YES**
GI-SD529-0006	020	20	18	15	15	21	89	NO	NO	NO
GI-SD530-0006	026	17	18	18	18	13	84	NO	NO	NO
GI-SD531-0006	024	20	19	13	8	19	79	NO	NO	**YES**
GI-SD532-0006	013	24	20	17	21	11	93	NO	NO	NO
GI-SD566-0006	004	8	13	14	10	8	53	**YES**	**YES**	**YES**

Table presents total number of L.plumulosus alive after 28 day toxicity test.
Shading indicates sample is considered toxic.
1 Duplicate of GI-SD509-0006
NA - Not applicable

TABLE 6-31
UNCERTAINTY ANALYSIS - SUMMARY OF NOECS AND LOECS FOR LEPTOCHEIRUS PLUMULOSUS BASED ON ALL ENDPOINTS
PHASE 2 REMEDIAL INVESTIGATION
SITE 17, GOULD ISLAND
NAVSTA NEWPORT, NEWPORT RI

Sample Area	Sample Number	Total Organic Carbon (%)	Concentration (mg/kg)						Concentration (ug/kg)							
			Antimony	Copper	Lead	Mercury	Nickel	Silver	2-Methyl naphthalene	Acenaphthene	Acenaphthylene	Anthracene	Benzo(a) anthracene	Benzo(a) pyrene	Benzo(b) fluoranthene	Benzo(g,h,i) perylene
	Toxic Samples[1]															
GI_SD_STILLWATER_2010	GI-SD501-0006	1.17	0.79 UJ	28.7 J	35.2 J	0.107	7.7 J	0.145 J	520 U	130 J	210 J	4400	5200	3800	4100	1600
GI_SD_STILLWATER_2010	GI-SD503-0006	0.8	0.7 UJ	9.6 J	13.9 J	0.059	6.7 J	0.23 UJ	93 U	42 J	24 J	300	1000	760	770	330
GI_SD_STILLWATER_2010	GI-SD505-0006	1.38	0.208 J	25 J	20.4	0.079	6.7	0.271 U	110 U	38 J	54 J	270	580	600	590	320
GI_SD_STILLWATER_2010	GI-SD507-0006	0.724	0.69 UJ	12.5 J	12.4 J	0.045	6.5 J	0.143 J	46 UJ	62 J	51 J	230 J	750 J	740 J	770 J	330 J
GI_SD_STILLWATER_2010	GI-SD508-0006-AVG	0.288	0.65 UJ	14.8 J	10.6 J	0.0195 J	8.9 J	0.0435 J	21.5 U	18 J	12 J	94.5 J	185 J	118 J	145 J	61.5
GI_SD_STILLWATER_2010	GI-SD509-0006	0.57	0.68 UJ	11.8 J	10.6 J	0.0285	5.85 J	0.106 J	17.8 J	44 J	17 J	136 J	300 J	285 J	260 J	155 J
GI_SD_STILLWATER_2010	GI-SD509-0006-D	0.59	0.68 UJ	11.8 J	10.6 J	0.0285	5.85 J	0.106 J	17.8 J	44 J	17 J	136 J	300 J	285 J	260 J	155 J
GI_SD_NE_2010	GI-SD510-0006	0.474	0.63 UJ	17.7 J	10.7 J	0.012 J	10.8 J	0.017 U	8.4 UJ	11 J	12 J	76 J	96 J	130 J	120 J	71 J
GI_SD_NE_2010	GI-SD512-0006	0.233	1.77 UJ	30.6 J	8.4 J	0.006 J	33.4	0.2 UJ	42 J	190 J	32 UJ	360 J	420 J	380 J	280 J	190 J
GI_SD_NE_2010	GI-SD518-0006	0.97	0.57 UJ	17.9 J	14.5 J	0.023	9.6 J	0.19 U	37 U	70	13 J	160	560	540	430	330
GI_SD_NE_2010	GI-SD519-0006	0.524	0.665 UJ	10.9 J	17.9	0.02 J	6	0.222 UJ	1.8 J	6.8 J	2.3 J	25	68	68	60	37
GI_SD_NW_2010	GI-SD520-0006	0.429	0.67 UJ	6.5 J	9.3 J	0.015 J	6.7	0.435 J	8.9 J	21 J	12 J	70 J	150 J	150 J	110 J	92 J
GI_SD_NW_2010	GI-SD521-0006	0.296	0.58 UJ	22.9 J	13.2 J	0.019 J	9.4	0.19 UJ	31 J	290 J	12 J	780 J	1600 J	1400 J	990 J	680 J
GI_SD_NW_2010	GI-SD523-0006	0.366	0.6 UJ	37.7 J	35.9 J	0.005 J	17.2	0.277 UJ	4.8 J	30	2 J	65	150	150	110	89
GI_SD_NE_2010	GI-SD528-0006	0.44	0.67 UJ	16.4 J	96.8 J	0.018 J	7.3	0.254 UJ	3.1 J	13	2.2 J	22	55	53 J	46	38 J
GI_SD_NE_2010	GI-SD531-0006	0.646	0.75 UJ	102 J	199 J	0.229	8.2	0.25 UJ	3.5 J	17	6 J	32	69	77 J	93	78 J
GI_SD_NE_2010	GI-SD566-0006	1.03	0.71 UJ	16 J	17.7 J	0.03	8.8 J	0.24 UJ	13 J	91 J	5.7 J	190 J	340 J	330 J	250 J	180 J
	Non-Toxic Samples[1]															
GI_SD_STILLWATER_2010	GI-SD502-0006	0.946	0.81 J	21.2 J	11.3 J	0.02 J	10 J	0.27 U	320 U	200 J	110 J	2300	2200	1500	1700	620
GI_SD_STILLWATER_2010	GI-SD504-0006	1.13	0.72 UJ	17.8 J	21 J	0.083	7 J	0.223 J	96 U	110	29 J	240	650	590	490	340
GI_SD_STILLWATER_2010	GI-SD506-0006	0.72	0.692 UJ	11.1 J	11.4	0.035	6.5	0.231 UJ	46 U	19 J	35 J	97	290	260	280	150
GI_SD_NE_2010	GI-SD511-0006	0.59	0.71 UJ	10.6 J	20.7 J	0.029	7.6 J	0.24 UJ	9.3 U	4.2 J	1.9 J	11	25	28	27	21
GI_SD_NE_2010	GI-SD513-0006	0.67	2.05 UJ	6.8 J	7.3 J	0.039 U	6.2	0.524	1.4 J	9 U	12	9 U	28	31 J	22	25 J
GI_SD_NE_2010	GI-SD514-0006	1.37	0.89 UJ	13.9 J	23 J	0.087	6.3 J	0.3 U	117 UJ	220 J	47 J	500 J	910 J	890 J	700 J	490 J
GI_SD_NE_2010	GI-SD515-0006	1.47	0.253 J	18.7 J	18.3 J	0.239	8.2 J	0.27 U	26 J	150 J	14 J	260 J	510 J	530 J	440 J	290 J
GI_SD_NE_2010	GI-SD517-0006-AVG	0.825	1.46 UJ	13.8 J	41.6 J	0.031 J	8.5	0.245 UJ	1.95 J	2.75 J	3.75 J	11 UJ	37 J	37 J	38.5 J	24 J
GI_SD_NW_2010	GI-SD522-0006	1.19	2.6 UJ	40.6 J	22.2 J	0.039	21.6	0.492 J	2.9 J	7.5 J	5.2 J	25	130	81 J	76	43 J
GI_SD_REF_2010	GI-SD524-0006	0.59	0.68 UJ	6.2 J	9.2 J	0.028	7 J	0.097 J		9.1 U	4.6 J	4.1 J	22	24	19	20
GI_SD_REF_2010	GI-SD525-0006	1.16	0.83 UJ	15 J	20.6 J	0.083	10.2 J	0.19 J		3.9 J	13 J	15 J	61 J	73 J	58 J	48 J
GI_SD_REF_2010	GI-SD526-0006	0.849	0.68 UJ	9.9 J	10.6 J	0.039	11.7 J	0.144 J		8.9 U	3.6 J	5 J	15	18	15	14
GI_SD_STILLWATER_2010	GI-SD527-0006	1.14	0.71 UJ	14.9 J	15 J	0.054	7.7 J	0.23 UJ	18 U	13 J	17 J	57	160	160	160	98
GI_SD_S_2010	GI-SD529-0006	0.83	0.718 UJ	8.1 J	12.7	0.027 J	8.4	0.252 UJ	13	57	2.4 J	98	180	180	140	110
GI_SD_NE_2010	GI-SD530-0006	0.5	0.69 UJ	186 J	20.9 J	0.031	32.8	0.23 UJ	3.2 J	7.9 J	6.9 J	23	68	67 J	59	40 J
GI_SD_STILLWATER_2010	GI-SD532-0006	0.442	0.71 UJ	10.4 J	7.1 J	0.023 J	8.4 J	0.24 U	290 U	230 J	87 J	2400	2900	2100	2600	970
	NOEC[2]		0.81	186	41.6	0.239	32.8	0.524	26	230	110	2400	2900	2100	2600	970
	[Max.] in toxic sample		0.208	102	199	0.229	33.4	0.435	42	290	210	4400	5200	3800	4100	1600
	LOEC[3]		NA	NA	96.8	NA	33.4	NA	31	290	210	4400	5200	3800	4100	1600
	Geometric mean of NOEC and LOEC		NA	NA	63.5	NA	33.1	NA	28	258	152	3250	3883	2825	3265	1246

Footnotes:
1 - Uncertainty analysis risk characterization is presented in Section 6.5.4
2 - Greatest concentration in a non-toxic sample.
3 - Lowest concentration in a toxic sample that is greater than the maximum concentration in a non-toxic sample.

ERM-Q - Effects range medium quotient
NOEC - No observed effects level
LOEC - Lowest observed effects level

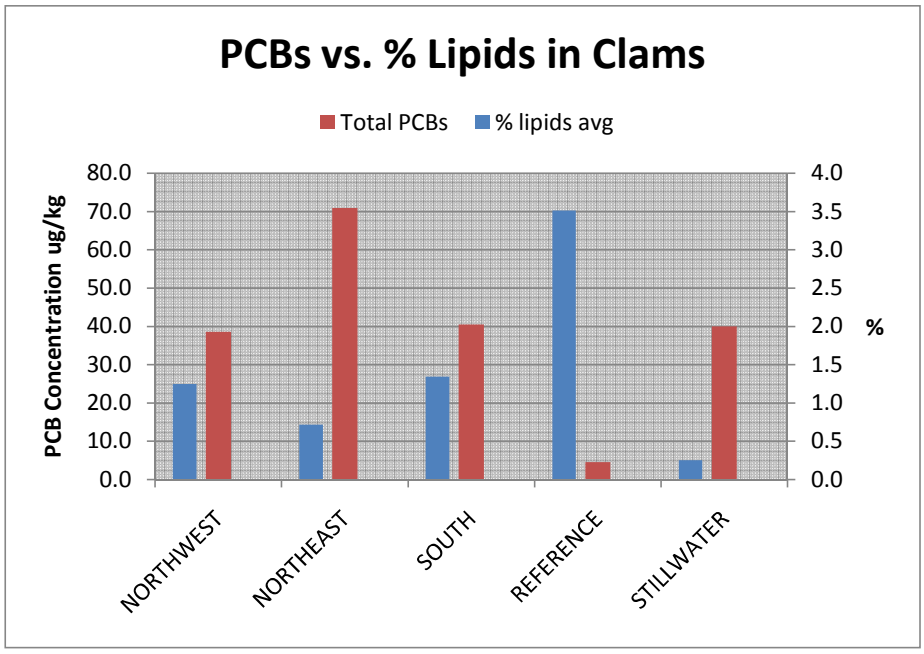
TABLE 6-31
UNCERTAINTY ANALYSIS - SUMMARY OF NOECS AND LOECS FOR LEPTOCHEIRUS PLUMULOSUS BASED ON ALL ENDPOINTS
PHASE 2 REMEDIAL INVESTIGATION
SITE 17, GOULD ISLAND
NAVSTA NEWPORT, NEWPORT RI

Sample Area	Sample Number	Concentration (ug/kg)													Average ERM-Q
		Benzo(k) fluoranthene	Chrysene	Dibenzo(a,h) anthracene	Fluoranthene	Fluorene	Indeno(1,2,3-cd)pyrene	Naphthalene	Phenanthrene	Pyrene	HMW PAHs	LMW PAHs	Total PAHs	Total PCB Homologs	
	Toxic Samples[1]														
GI_SD_STILLWATER_2010	GI-SD501-0006	3900	6900	580	12000	210 J	1600	520 U	1600	8900	48600	6550 J	55100 J	2720 J	2.29
GI_SD_STILLWATER_2010	GI-SD503-0006	580	1100	130	2100	38 J	330	93 U	540	1800	8900	944 J	9840 J	5.71 U	0.17
GI_SD_STILLWATER_2010	GI-SD505-0006	520	840	110	830	49 J	280	110 U	390	840	5510	801 J	6310 J	950	0.66
GI_SD_STILLWATER_2010	GI-SD507-0006	590 J	1100 J	120 J	1500 J	72 J	330 J	67 J	610 J	1200 J	7430 J	1090 J	8520 J	2590	1.59
GI_SD_STILLWATER_2010	GI-SD508-0006-AVG	125 J	270 J	20.5 J	655	21.5 J	56.5	21.5 U	180 J	515	2150 J	326 J	2480 J	895 J	0.58
GI_SD_STILLWATER_2010	GI-SD509-0006	225 J	425 J	55 J	625 J	44 J	155 J	36.5 UJ	345 J	520 J	3000 J	593 J	3600 J	460 J	0.34
GI_SD_STILLWATER_2010	GI-SD509-0006-D	225 J	425 J	55 J	625 J	44 J	155 J	36.5 UJ	345 J	520 J	3000 J	593 J	3600 J	460 J	0.34
GI_SD_NE_2010	GI-SD510-0006	93 J	140 J	26 J	150 J	23 J	73 J	12 J	110 J	120 J	1020 J	244 J	1260 J	2.14 UJ	0.07
GI_SD_NE_2010	GI-SD512-0006	260 J	370 J	70 J	1000 J	170 J	190 J	130 J	990 J	750 J	3910 J	1880 J	5790 J	3.2 J	0.21
GI_SD_NE_2010	GI-SD518-0006	360	560	110	1000	57	310	37 U	540	900	5100	840 J	5940 J	164 J	0.22
GI_SD_NE_2010	GI-SD519-0006	49	74	13	120	6.3 J	34	2.7 J	67	110	633	112 J	745 J	286	0.21
GI_SD_NW_2010	GI-SD520-0006	110 J	160 J	29 J	360 J	34 J	82 J	19 J	260 J	280 J	1520 J	425 J	1950 J	6.7 J	0.06
GI_SD_NW_2010	GI-SD521-0006	1100 J	1500 J	270 J	3300 J	290 J	690 J	78 UJ	2400 J	2600 J	14100 J	3800 J	17900 J	1.42 U	0.32
GI_SD_NW_2010	GI-SD523-0006	110	150	30	320	28	84	7.9 UJ	230	260 J	1450 J	360 J	1810 J	3.92 J	0.12
GI_SD_NE_2010	GI-SD528-0006	40	54	12	120	9.3	33	8.8 U	74	100 J	551 J	124 J	675 J	56.1 J	0.12
GI_SD_NE_2010	GI-SD531-0006	64	82	25	140	13	66	10 U	95	130 J	824 J	166 J	990 J	19.1 J	0.19
GI_SD_NE_2010	GI-SD566-0006	230 J	340 J	74 J	770 J	66 J	190 J	33 J	550 J	590 J	3290 J	949 J	4240 J	129 J	0.18
	Non-Toxic Samples[1]														
GI_SD_STILLWATER_2010	GI-SD502-0006	1400	3800	250 J	8800	440	690	320 U	4700	6000	27000 J	7750 J	34700 J	862 J	1.06
GI_SD_STILLWATER_2010	GI-SD504-0006	530	730	110	1200	87 J	320	96 U	770	1000	5960	1240 J	7200 J	394 J	0.38
GI_SD_STILLWATER_2010	GI-SD506-0006	240	480	49	940	26 J	140	7 J	310	740	3570	494 J	4060 J	547	0.39
GI_SD_NE_2010	GI-SD511-0006	24	31	7.1 J	49	4.7 J	16	9.3 U	34	43	271 J	55.8 J	327 J	276 J	0.20
GI_SD_NE_2010	GI-SD513-0006	24	31	6.8 J	48	9 U	19	9 U	25	48 J	283 J	38.4 J	321 J	4.3 J	0.04
GI_SD_NE_2010	GI-SD514-0006	630 J	940 J	170 J	2000 J	200 J	470 J	117 UJ	1500 J	1500 J	8700 J	2470 J	11200 J	16 J	0.22
GI_SD_NE_2010	GI-SD515-0006	340 J	500 J	100 J	1100 J	100 J	300 J	54 J	790 J	860 J	4970 J	1390 J	6360 J	290 J	0.31
GI_SD_NE_2010	GI-SD517-0006-AVG	32 J	42 J	7.7 J	56.5 J	3.75 J	22 J	10 UJ	25 J	45.5 J	342 J	37.2 J	380 J	150 J	0.15
GI_SD_NW_2010	GI-SD522-0006	89	160	15	490	6.9 J	42	11 U	69	360 J	1490 J	116 J	1600 J	6.4 J	0.13
GI_SD_REF_2010	GI-SD524-0006	18	27	4.1 J	40	2.3 J	15	9.1 U	20	40	229 J	31 J	260 J	0.543 U	0.04
GI_SD_REF_2010	GI-SD525-0006	52 J	71 J	15 J	130 J	6.7 J	45 J	11 UJ	58 J	120 J	673 J	96.6 J	770 J	0.669 U	0.07
GI_SD_REF_2010	GI-SD526-0006	13	17	3.2 J	36	1.8 J	12	8.9 U	17 U	30	173 J	10.4 J	184 J	0.538 U	0.05
GI_SD_STILLWATER_2010	GI-SD527-0006	140	240	33	260	16 J	91	18 U	130	230	1570	233 J	1800 J	25.3	0.08
GI_SD_S_2010	GI-SD529-0006	150	180	36	400	47	98	30	320	310	1780	567 J	2350 J	16	0.08
GI_SD_NE_2010	GI-SD530-0006	55	100	15	140	7.9 J	37	9.2 U	74	130 J	711 J	123 J	834 J	8.1 J	0.19
GI_SD_STILLWATER_2010	GI-SD532-0006	2000	4900	360	8900	330	980	290 U	3600	6300	32000	6650 J	38700 J	1240	1.27
	NOEC[2]	2000	4900	360	8900	440	980	54	4700	6300	32000	7750	38700	1240	1.27
	[Max.] in toxic sample	3900	6900	580	12000	290	1600	130	2400	8900	48600	6550	55100	2720	2.29
	LOEC[3]	3900	6900	580	12000	NA	1600	67	NA	8900	48600	NA	55100	2590	1.59
	Geometric mean of NOEC and LOEC	2793	5815	457	10334	NA	1252	60	NA	7488	39436	NA	46178	1792	1.42

Footnotes:
1 - Uncertainty analysis risk characterization is presented in S
2 - Greatest concentration in a non-toxic sample.
3 - Lowest concentration in a toxic sample that is greater than

ERM-Q - Effects range medium quotient
NOEC - No observed effects level
LOEC - Lowest observed effects level

FIGURE 1
RESPONSE TO COMMENTS, DRAFT PHASE 2 RI BERA
SITE 17, GOULD ISLAND



	NW	NE	S	REF	SB
Avg L =	77	82	79	59	36
# samples	3	9	4	3	1

each sample consists of >3 individuals.
data includes 1/2 ND values
PCB data reflects total aroclor and total homologues

SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 17, No. 4, pp. 714–727, 1998
Printed in the USA
0730-7268/98 $6.00 + .00

PREDICTING TOXICITY IN MARINE SEDIMENTS WITH NUMERICAL SEDIMENT QUALITY GUIDELINES

EDWARD R. LONG,*† L. JAY FIELD,‡ and DONALD D. MACDONALD§
†Coastal Monitoring and Bioeffects Assessment Division, ‡Hazardous Materials Response and Assessment Division, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, Washington 98115, USA
§MacDonald Environmental Sciences, Ltd., 1733 Idaho Place, Nanaimo, British Columbia V9X 1C6, Canada

(*Received* 5 *February* 1997; *Accepted* 29 *July* 1997)

Abstract—Matching synoptically collected chemical and laboratory bioassay data (n = 1,068) were compiled from analyses of surficial sediment samples collected during 1990 to 1993 to evaluate the predictive ability of sediment quality guidelines (SQGs), specifically, effects range—low (ERL), effects range—median (ERM), threshold effects level (TEL), and probable effects level (PEL) values. Data were acquired from surveys of sediment quality performed in estuaries along the Atlantic, Pacific, and Gulf of Mexico coasts. Samples were classified as either nontoxic ($p > 0.05$ relative to controls), marginally toxic ($p < 0.05$ only), or highly toxic ($p < 0.05$ and response greater than minimum significant difference relative to controls). This analysis indicated that, when not exceeded, the ERLs and TELs were highly predictive of nontoxicity. The percentages of samples that were highly toxic generally increased with increasing numbers of guidelines (particularly the ERMs and PELs) that were exceeded. Also, the incidence of toxicity increased with increases in concentrations of mixtures of chemicals normalized to (divided by) the SQGs. The ERMs and PELs indicated high predictive ability in samples in which many substances exceeded these concentrations. Suggestions are provided on the uses of these estimates of the predictive ability of sediment guidelines.

Keywords—Sediment quality guidelines Predictive ability Laboratory toxicity tests

INTRODUCTION

Using similar empirical approaches, sediment quality guidelines (SQGs) were prepared for salt water [1–3] and freshwater [4,5] as informal (nonregulatory) benchmarks to aid in the interpretation of sediment chemistry data. For marine sediments, effects range—low (ERL) and effects range—median (ERM) concentrations for 9 trace metals, 3 chlorinated organics, and 13 polynuclear aromatic hydrocarbons (PAHs) were identified [1]. Threshold effects level (TEL) and probable effects level (PEL) concentrations for 9 trace metals, 8 chlorinated organics, 1 phthalate, and 13 PAHs were published [2]. These guidelines were not based upon experiments in which causality was determined. Rather, both sets of marine guidelines were based upon empirical analyses of data compiled from numerous field and laboratory studies performed in many estuaries and bays of North America. These studies included chemistry data and a variety of different types of biological data for numerous taxa derived from either bioassays of field-collected samples, laboratory toxicity tests of clean sediments spiked with specific toxicants, benthic community analyses, or equilibrium-partitioning models.

The objectives of the ERL and TEL values and of the ERM and PEL values were comparable. The ERLs and TELs were intended to represent chemical concentrations toward the low end of the effects ranges, that is, below which adverse biological effects were rarely observed. The ERMs and PELs were intended to represent concentrations toward the middle of the effects ranges and above which effects were more frequently observed. As estimates of reliability, the incidence of adverse effects within concentration ranges defined by these SQGs were determined using data with which they were derived [1,2]. Generally, adverse effects occurred in less than 10% of studies in which concentrations were below the respective ERL or TEL values and were observed in more than 75% or 50% of studies in which concentrations exceeded the ERMs or PELs, respectively.

Since they were published, the guidelines [1,2] have been used as interpretive tools in many sediment assessments throughout North America and elsewhere. Generally, the ERLs and TELs have been used to identify relatively uncontaminated samples that pose a limited risk of toxicity. The ERMs and PELs have been used to identify those samples and areas in which chemical concentrations were sufficiently elevated to warrant further evaluation. Because these guidelines were based upon analyses of large databases, mostly composed of field-collected data in which mixtures of toxicants were encountered, it was assumed [1,2] that the guidelines would provide relatively accurate tools for classifying newly collected samples as potentially toxic or nontoxic. Thus far, however, the accuracy of the two sets of guidelines in predicting nontoxic and toxic conditions correctly has not been evaluated. Therefore, because of the widespread use of these guidelines, we concluded there was a need for analyses of their predictive ability with data independent of those with which the SQGs were derived.

The objectives of this paper are to quantify the frequency with which ERL/ERM and TEL/PEL guidelines correctly classify samples as either nontoxic or toxic; to quantify the incidence of toxicity among samples in which different numbers of SQGs were exceeded; to determine the incidence of toxicity over ranges in concentrations of chemical mixtures normalized

* To whom correspondence may be addressed (edward_long@hazmat.noaa.gov).

Table 1. Sources of data and the toxicity tests performed in each study

			Bioassays performed						
Survey area	Year sampled	No. samples	Amphipod survival	Clam embryo survival	Clam embryo development	Microbial bioluminescence	Urchin egg fertilization	Urchin embryo development	Abalone embryo development
Hudson–Raritan estuary	1991	38	X	X	X	X			
Newark Bay	1993	20	X						
Long Island Sound	1991	63	X	X	X	X			
Boston Harbor	1993	30	X			X	X		
Tampa Bay phase 1	1992	16	X			X	X		
Tampa Bay phase 2	1993	45	X				X		
San Diego Bay	1993	121	X					X	
San Pedro Bay	1992	45	X						X
Charleston Harbor	1993	79	X			X	X		
EMAP—Estuaries[a]	1990–1992	611	X						
Total		1,068							

[a] EMAP = Environmental Monitoring and Assessment Program; data from mysid tests not included.

to the SQGs; and to compare the relative predictive ability of the two sets of guidelines. The design followed that of a previous study in freshwater [5] in which type I and type II errors were determined for ERL/ERM and TEL/PEL values. Type I errors (false positives) are those in which toxicity was expected (based upon high chemical concentrations), but was not observed. Type II errors (false negatives) are those in which no toxicity was expected (low chemical concentrations), but was actually observed.

METHODS

Matching, synoptically collected, sediment chemistry and bioassay data for 1,068 samples were compiled from studies performed by the National Oceanic and Atmospheric Administration (NOAA) and U.S. Environmental Protection Agency (U.S. EPA) (Table 1). Regional sediment quality assessments were conducted as a part of NOAA's National Status and Trends Program (NS&TP) and included those performed in (all in the USA) the Hudson–Raritan estuary in New York and New Jersey [6], Newark Bay in New Jersey [6], the bays adjoining Long Island Sound in New York and Connecticut [7], Boston Harbor in Massachusetts [8], Tampa Bay in Florida [9], San Diego Bay [10] and San Pedro Bay [11] in southern California, and Charleston Harbor in South Carolina (unpublished). The U.S. EPA data were generated in Environmental Monitoring and Assessment Program (EMAP) studies of the Virginian and Louisianian estuarine provinces [12–14].

All of these data were generated during surveys performed to quantify the spatial extent, patterns, and severity of adverse biological effects attributable to toxic substances. Samples from the upper 2 to 3 cm of the sediments were collected with grab samplers throughout each survey area to characterize surficial sediment contamination and toxicity. Data for these analyses were selected because they were generated with similar protocols, included matching chemistry and toxicity results, indicated a range in toxicity responses, and represented conditions from all three coastlines.

Sample collection and handling methods, toxicity testing methods, chemical analytical protocols, and raw data are included in the respective technical reports. All analytical laboratories followed the performance-based protocols of the NS&TP and EMAP—Estuaries to ensure comparability among results [15]. All toxicity testing laboratories followed standardized protocols: American Society for Testing and Materials [16,17] for the amphipod survival tests, U.S. National Biological Service [18] for the urchin tests, and U.S. EPA [19] and Schiewe et al. [20] for the Microtox™ tests (AZUR Environmental, Carlsbad, CA, USA). All bioassay data were expressed as percent of negative, laboratory controls (not reference samples) to account for variability among studies and laboratories in organism viability.

We considered several different approaches to the classification of samples as either nontoxic or toxic. In an interlaboratory comparison of performance, results of amphipod survival tests were classified as either nontoxic (mean survival 96–96.5%), marginally toxic (mean survival 76.5–83%), clearly toxic (mean survival $< 76\%$), highly toxic (mean survival $< 20\%$) [21]. Swartz et al. [22] classified results of amphipod survival tests as either not toxic ($<13\%$ mortality), uncertain (13–24% mortality), or toxic ($>24\%$ mortality). Statistical tests were recommended [16] to determine if differences in results of tests of field-collected samples and controls are statistically significant. An alternative approach [23], based upon results of power analyses of amphipod survival tests, recommended the use of minimum significant differences (MSDs) from controls as criteria for classifying samples as toxic.

We chose to use a combination of these approaches to classify samples. Following standardized procedures [16], samples in which test results were not statistically different from negative controls (i.e., $p > 0.05$) were classified as nontoxic and samples in which results were significantly different from controls were classified as toxic. However, to further distinguish differences in degrees of toxicity, sample classifications followed the recommendations of Thursby et al [23]. Samples in which test results were significant relative to controls, but were less than MSDs were labeled as marginally toxic and those in which results were both significant and greater than MSDs were labeled as highly toxic. The highly toxic label does not imply that toxicity was severe; rather, it was used to identify those results for which statistical certainty was greatest. The MSD values calculated and published for *Ampelisca abdita* [23] were used for all amphipod test results. The MSD values for Microtox tests [24], *Arbacia punctulata* fertilization tests [25], and and all other tests were determined empirically with power analyses of the frequency distributions of data from each test.

Marginally toxic samples are subject to type I error (declaring a sample as toxic when it is not) when variance among

laboratory replicates is very small [23]. However, these samples could not be classified as nontoxic because they were significantly different from controls. Therefore, we chose to classify them separately as neither nontoxic nor highly toxic. Because of the uncertainty associated with marginally toxic results, this evaluation focuses mainly upon the nontoxic and highly toxic categories.

Following the completion of an electronic database, several analyses were performed to determine the predictive ability of the guidelines. In these analyses, the guidelines for nickel were excluded because of the low degree of reliability determined for these values [1,2]. Also, the sums of low- and high-molecular-weight PAHs and total PAHs were excluded to avoid redundancy with the data for individual compounds. In summations of total polychlorinated biphenyls (PCBs), total dichlorodiphenyltrichloroethanes (DDTs), and total PAHs, concentrations of individual compounds were treated as zeroes when they were below method detection limits (MDLs). The MDLs achieved differed slightly among laboratories; therefore, the use of zeroes minimized inconsistencies in data treatments. In any case the use of either one half of the MDL or zeroes had no effect upon classification of samples relative to the SQGs.

Three data analyses were performed. First, the predictive abilities of individual SQGs were determined. Second, the incidence of toxicity was determined among samples in which none of the substances equaled or exceeded the ERL concentrations; in which one or increasing numbers of substances exceeded ERL concentrations, but none exceeded any ERM; and in which one or increasing numbers of substances exceeded ERM concentrations. The same approach was used to evaluate the predictive ability of the TEL/PELs. We scored samples as exceeding SQGs when a chemical concentration either equaled the value or exceeded it by any amount.

In the third analysis, the incidence of toxicity over ranges in mean SQG quotients [5,25] was determined. The concentrations of individual chemicals were divided by their respective ERMs or PELs and the means of these concentration-to-SQG quotients were determined. The means of these quotients were determined to account for differences among studies in the numbers of chemicals for which analyses were performed. Predictive ability was calculated with samples classified as either nontoxic or highly toxic, excluding the marginally toxic results.

Similar to the criteria used to determine guideline reliability [1], we considered the guidelines to be predictive if the incidence of toxicity was less than 25% when all concentrations were less than the ERLs or TELs and greater than 75% when at least one concentration exceeded an ERM or PEL. Therefore, our target level for both false negatives and false positives was $\leq$25%.

Data are reported for the results of amphipod survival tests alone and for any one of the battery of two to four tests performed. In the latter analyses, samples were classified as marginally or highly toxic if one or more of the bioassays met the criteria for these classifications.

RESULTS

The database

Data were compiled from 1,068 samples analyzed during EMAP and NOAA studies conducted during 1990 to 1993. Roughly one third of the data were obtained from the NOAA surveys of selected urbanized bays with sample sizes ranging from 20 to 121 (Table 1). The EMAP data from the Atlantic and Gulf coasts constituted the remaining two thirds of the database (n = 611).

Amphipod survival was determined for all samples; one to three additional tests were performed on all samples except those collected in the EMAP and Newark Bay studies (n = 437). The data from bioassays performed with mysids by the EMAP were not used because these tests failed to indicate toxicity. Amphipod survival was determined with *A. abdita* in Atlantic and Gulf coast surveys and with *Rhepoxynius abronius* in California surveys. Other tests included bivalve (*Mulinia lateralis*) embryo survival and development with exposures to elutriates; microbial bioluminescence (Microtox) in exposures to organic solvent extracts; and pore-water tests of echinoderm (*A. punctulata*) fertilization in Gulf and Atlantic coast areas, echinoderm (purple urchin, *Strongylocentrotus purpuratus*) embryo development in San Diego Bay, and embryological development of red abalone (*Haliotis rufescens*) embryos in San Pedro Bay. Insufficient numbers of samples were tested in any of these nonamphipod tests to warrant analyses alone; therefore, the results of these tests were combined.

The chemical data from each survey indicated that samples contained mixtures of contaminants, including trace metals, PAHs, and chlorinated hydrocarbons. The numbers of samples analyzed for each chemical ranged from 399 to 1,060 (Table 2). Analyte concentrations exceeded the MDL in a majority of the samples. The concentrations of most trace metals ranged over two to three orders of magnitude, and those of most organic compounds ranged over four to six orders of magnitude. Concentrations of the PAHs were most often less than the MDL.

None of the samples exceeded the ERM value for arsenic and <1.0% exceeded the ERMs for cadmium and chromium (Table 2). Relatively small proportions of the samples had chemical concentrations that exceeded ERM values, indicating that the data were not skewed toward waste sites with unusually high concentrations. Undoubtedly, some samples contained chemicals that were not quantified or for which there were no SQGs.

Among the different tests performed, 15 to 91% of the samples were at least marginally toxic (Table 3). Bioassay results showed a wide range of response, often from 0 to >100% of mean control responses. In the amphipod tests 36 to 52% of the samples were toxic whereas in the tests of pore water 56 to 91% of samples were toxic.

The frequency distributions of the data from most of the tests were similar, that is, responses in most samples were >80% of control responses (Table 3). Many of the EMAP samples were marginally toxic in amphipod tests. The data from embryological tests with the purple urchin (*S. purpuratus*) and red abalone (*H. rufescens*) indicated similar frequency distributions, both suggesting higher sensitivities to the samples than found in the amphipods. Empirically derived MSDs for each bioassay were very similar, ranging from 80 to 87%.

Incidence of toxicity

Concentrations greater than individual SQGs. Table 4 summarizes the percentages of samples that were not toxic, were marginally toxic, and were highly toxic in the amphipod tests alone and in any of the two to four tests performed when the concentrations of substances equaled or exceeded individual ERMs. For 18 of the 27 ERMs evaluated, highly toxic

Table 2. Ranges in chemical concentrations, numbers of samples in which concentrations were less than or greater than method detection limits (MDLs), and percentages of samples in which effects range—median (ERM) values were exceeded

Chemical[a]	Units	No. samples	% > ERM[b]	No. > MDL	Range in detected concentrations		Range in concn. below detection limits[c]		No. < MDL
					Lowest	Highest	Lowest	Highest	
Arsenic	ppm	920	0.0	913	0.1	41	1.2	1.7	7
Cadmium	ppm	987	0.2	987	0.03	19.8	0.01	0.05	0
Chromium	ppm	1,058	0.5	1,045	1	1,220	1.2	18	13
Copper	ppm	1,057	2.4	1,031	0.7	1,770	0.2	1	26
Lead	ppm	1,052	3.4	1,038	1.4	510	0.3	1.3	14
Mercury	ppm	994	12.7	994	0.01	15	0.001	0.01	0
Nickel	ppm	1,042	2.1	1,006	0.3	136	0.2	1.7	36
Silver	ppm	919	4.4	866	0.01	10.1	0.01	0.7	53
Zinc	ppm	1,060	5.3	1,060	1	1,880	NA	NA	0
2-Methylnaphthalene	ppb	921	1.0	591	0.40	15,557	0.20	10	330
Dibenz[*a*,*h*]anthracene	ppb	399	11.8	363	0.40	4,534	0.70	10	36
Acenaphthene	ppb	977	3.3	394	0.10	56,338	0.10	80	583
Acenaphthylene	ppb	807	2.8	254	0.40	12,915	0.02	100	553
Anthracene	ppb	997	4.8	521	0.20	89,366	0.03	90	476
Benz[*a*]anthracene	ppb	996	7.2	652	0.30	59,298	0.02	130	344
Benzo[*a*]pyrene	ppb	980	10.0	631	0.20	54,862	0.02	170	349
Chrysene	ppb	997	5.5	688	0.20	60,331	0.10	130	309
Fluoranthene	ppb	1,000	4.2	755	0.30	108,236	0.20	110	245
Fluorene	ppb	945	3.2	530	0.10	54,209	0.10	90	415
Naphthalene	ppb	900	0.9	456	0.70	17,414	0.40	70	444
Phenanthrene	ppb	1,054	5.1	779	0.40	194,343	0.40	90	275
Pyrene	ppb	1,029	8.1	819	0.40	143,132	0.10	120	210
Total LMW PAHs	ppb	956	5.0	956	0.2	552,124	NA	NA	0
Total HMW PAHs	ppb	925	8.2	925	2	461,675	NA	NA	0
Total PAHs	ppb	1,003	1.1	1,003	0.2	1,013,799	NA	NA	0
p,p′-DDE	ppb	789	12.0	741	0.004	2,900	0.03	0.3	48
p,p′-DDD	ppb	742	No ERM	666	0.004	784	0.1	1	76
p,p′-DDT	ppb	656	No ERM	543	0.004	3,517	0.02	1	113
Total DDTs	ppb	813	13.2	813	0.01	4,631	NA	NA	0
Total PCBs	ppb	830	23.4	830	0.1	16,675	NA	NA	0
Dieldrin	ppb	615	No ERM	490	0.002	21.2	0.03	0.5	125
Lindane	ppb	533	No ERM	306	0.01	157	0.05	1	227

[a] LMW = low-molecular-weight, PAH = polynuclear aromatic hydrocarbon, HMW = high-molecular-weight, DDE = dichlorodiphenyldichloroethylene, DDT = dichlorodiphenyltrichloroethane, PCB = polychlorinated biphenyl.
[b] Percent of samples with detectable concentrations.
[c] NA = not applicable for summed concentrations.

results occurred in amphipod tests in 40 to 65% of the samples. The percentages of samples that were highly toxic in amphipod tests ranged from 40% for the ERM value for total PCB to 100% for the cadmium and chromium ERMs. The target percent of false positives (≤25% not toxic) was observed for 13 of the ERMs. The ERMs for six substances correctly classified ≥75% of samples as highly toxic in amphipod tests. Marginally toxic samples contributed relatively little (0–20%) to over-

Table 3. Frequency distribution of toxicity responses (expressed as percent of the total number of samples tested within categories of toxicologic responses), incidence of toxicity, and minimum significant differences (MSDs) for each test

Test medium/species[a]	Endpoint	Duration	*n*	% Control response <20%	20–39.99%	40–59.99%	60–80%	>80%	% Samples toxic[b]	MSD value
Solid phase										
Ampelisca abdita—NOAA	Survival	10 d	289	6.6	4.8	4.5	11.8	72.3	36.3	80
A. abdita—EMAP	Survival	10 d	611	1.4	1.0	2.3	12.1	83.1	38.3	80
Rhepoxynius abronius	Survival	10 d	166	6.0	8.4	6.0	18.7	60.8	51.8	80
Solvent extract										
Photobacterium phosphoreum	Bioluminescence	15 min	224	17.4	12.1	9.8	12.9	47.8	44.6	80
Elutriate										
Mulinea lateralis	Survival	48 h	100	1.0	8.0	12.0	11.0	68.0	29.0	80
M. lateralis	Normal development	48 h	100	7.0	3.0	0.0	1.0	89.0	15.0	80
Porewater										
Arbacia punctulata	Fertilization	1 h	168	24.4	5.9	5.4	5.4	58.9	56.0	87
Strongylocentrotus purpuratus	Normal development	1 h	52	86.5	0.0	3.8	1.9	7.7	90.4	85
Haliotis rufescens	Normal development	48 h	45	71.1	4.4	4.4	6.7	13.3	91.1	85

[a] NOAA = National Oceanic and Atmospheric Administration, EMAP = Environmental Monitoring and Assessment Program.
[b] Marginally + highly toxic ($p < 0.05$, *t* tests).

Table 4. Incidence of toxicity in either amphipod tests alone or any of the two to four tests performed among samples in which individual effects range—median (ERM) values were exceeded

	Amphipod tests (n = 1,068)					Any test performed[b] (n = 437)				
Chemical[a]	No.	% Not toxic	% Marginally toxic	% Highly toxic	% Total toxic	No.	% Not toxic	% Marginally toxic	% Highly toxic	% Total toxic
Metals										
Cadmium	2	0	0	100	100	0	NA	NA	NA	NA
Chromium	5	0	0	100	100	2	0	0	100	100
Copper	25	48	0	52	52	22	18	0	82	82
Lead	35	17	6	77	83	20	5	0	95	95
Mercury	126	34	12	54	66	81	10	6	84	90
Nickel	21	24	14	62	76	5	0	0	100	100
Silver	38	34	18	47	65	22	0	14	86	100
Zinc	56	34	5	61	66	32	13	0	88	88
PAHs										
2-Methylnaphthalene	6	0	17	83	100	4	0	0	100	100
Dibenz[*a*,*h*]anthracene	43	28	2	70	72	31	19	0	81	81
Acenaphthene	13	23	15	62	77	7	0	0	100	100
Acenaphthylene	7	0	14	86	100	6	0	0	100	100
Anthracene	25	24	20	56	76	19	11	0	89	89
Benz[*a*]anthracene	47	23	13	64	77	30	10	0	90	90
Benzo[*a*]pyrene	63	37	8	56	64	46	17	0	83	83
Chrysene	38	32	13	55	68	26	8	0	92	92
Fluoranthene	32	28	13	59	72	21	5	0	95	95
Fluorene	17	29	12	59	71	10	10	0	90	90
Naphthalene	4	25	0	75	75	4	0	25	75	100
Phenanthrene	40	25	15	60	75	25	4	0	96	96
Pyrene	66	33	9	58	67	46	11	0	89	89
Sum LMW PAHs	48	21	13	67	80	31	3	6	90	96
Sum HMW PAHs	76	39	9	51	60	56	16	0	84	84
Sum total PAHs	11	9	18	73	91	6	0	0	100	100
Chlorinated hydrocarbons										
p,p′-DDE	89	45	7	48	55	70	9	6	86	92
Total DDTs	107	38	11	50	61	82	5	9	87	96
Total PCBs	194	49	11	40	51	162	17	6	78	84

[a] PAH = polynuclear aromatic hydrocarbon, LMW = low-molecular weight, HMW = high-molecular-weight, DDE = dichlorodiphenyldichloroethylene, DDT = dichlorodiphenyltrichloroethane, PCB = polychlorinated biphenyl.
[b] Excludes Environmental Monitoring and Assessment Program and Newark Bay samples; NA = not applicable.

all predictive ability. However, based upon sums of the marginally toxic and highly toxic responses, the number of ERMs that correctly predicted toxicity in ≥75% of samples increased from 6 to 13.

Relative to results of the amphipod tests, predictive ability increased considerably when the results were considered for all of the tests performed; ≥75% for all substances that exceeded the ERM concentrations (Table 4). The target percent of false positives (≤25%) was observed for all ERMs and was ≤10% for 18 substances. As with the amphipod data, the marginally toxic results in all tests performed contributed relatively little to overall predictive ability; that is, the samples often were either nontoxic or highly toxic.

Predictive ability observed with the individual PELs was slightly lower than that of equivalent ERMs (Table 5). The percentages of samples exceeding PELs that were highly toxic in amphipod tests ranged from 15% (lindane) to 73% (dieldrin). For 25 of the 31 PELs, highly toxic conditions in amphipod tests occurred in 40 to 65% of the samples. Predictive ability of ≥75% was observed for none of the PELs with only highly toxic responses and with three PELs (cadmium, acenaphthylene, and dieldrin) with marginally plus highly toxic responses combined. The target percent of false positives (≤25%) was observed for the same three PELs. When the results of any of the tests performed were considered, the percent of false positives for the PELs was ≤25% for all except one substance (p,p′-dichlorodiphenyldichloroethylene [p,p′-DDE]) and was ≤10.0% for 15 PELs. For most substances, marginally toxic results contributed 5 to 10% to overall predictive ability in both the amphipod tests alone and in all tests considered. Predictive ability of ≥75% (with highly toxic responses) was observed in any of the tests performed for all PELs except that for p,p′-DDE.

Concentrations above and below all ERL or TEL concentrations. Among the 329 samples in which none of the chemical concentrations exceeded any ERL values, 68% were not toxic, 21% were marginally toxic, and 11% were highly toxic in the amphipod tests (Table 6). Among samples in which multiple bioassays were performed, 46% were not toxic in all tests and 41% were highly toxic in at least one test when all chemical concentrations were less than the ERLs.

Of the samples tested with amphipods, 448 were found in which one or more of the 24 concentrations were greater than or equal to the ERL, but none of the concentrations were greater than or equal to the ERM values; 63% were nontoxic, 20% were marginally toxic, and 18% were highly toxic. A total of 64% of 173 samples was highly toxic in any test performed when one or more ERLs was exceeded and no ERMs were exceeded. The percent of false positives for one or more ERLs exceeded was 63% for amphipod tests alone and 20% for all tests performed.

Generally, the incidence of toxicity increased with the number of chemicals greater than or equal to the ERL concentrations; however, this pattern was variable and inconsistent (Ta-

Table 5. Incidence of toxicity in either amphipod tests alone or any of the two to four tests performed among samples in which individual probable effects levels (PELs) were exceeded

	Amphipod tests (n = 1,068)					Any test performed[b] (n = 437)				
Chemical[a]	No.	% Not toxic	% Marginally toxic	% Highly toxic	% Total toxic	No.	% Not toxic	% Marginally toxic	% Highly toxic	% Total toxic
Metals										
Cadmium	21	19	10	71	81	6	0	0	100	100
Chromium	41	34	7	59	66	24	8	0	92	92
Copper	179	41	11	48	59	146	13	6	81	87
Lead	122	37	11	52	63	85	8	6	86	92
Mercury	127	35	12	54	66	82	11	6	83	89
Nickel	74	34	12	54	66	37	5	5	89	94
Silver	109	41	10	49	59	82	12	11	77	88
Zinc	126	38	10	52	62	87	14	2	84	86
PAHs										
2-Methylnaphthalene	47	28	13	60	73	22	5	9	86	95
Dibenz[*a*,*h*]anthracene	80	36	3	61	64	65	15	2	83	85
Acenaphthene	84	38	8	54	62	56	5	7	88	95
Acenaphthylene	47	23	9	68	77	40	3	8	90	98
Anthracene	131	44	7	49	56	100	11	5	84	89
Benz[*a*]anthracene	116	39	9	52	61	93	12	4	84	88
Benzo[*a*]pyrene	126	41	9	50	59	100	12	3	85	88
Chrysene	116	43	9	47	56	93	12	4	84	88
Fluoranthene	103	42	10	49	59	80	13	5	83	88
Fluorene	74	30	12	58	70	51	6	6	88	94
Naphthalene	38	26	11	63	74	23	0	13	87	100
Phenanthrene	106	40	11	49	60	77	8	5	87	92
Pyrene	117	40	9	51	60	94	11	4	85	89
Sum LMW PAHs	117	36	9	55	64	79	8	5	87	92
Sum HMW PAHs	114	42	7	51	58	90	12	3	84	87
Sum total PAHs	56	32	11	57	68	38	11	0	89	89
Chlorinated hydrocarbons										
p,p′-DDE	3	67	0	33	33	3	33	0	67	67
p,p′-DDD	144	35	11	54	65	115	8	7	85	92
p,p′-DDT	97	33	11	56	67	68	6	7	87	94
Total DDTs	101	36	12	52	64	78	5	9	86	95
Total PCBs	191	50	10	39	49	159	17	6	77	83
Dieldrin	41	20	7	73	80	25	4	0	96	96
Lindane	54	81	4	15	19	50	14	0	86	86

[a] PAH = polynuclear aromatic hydrocarbon, LMW = low-molecular-weight, HMW = high-molecular-weight, DDE = dichlorodiphenyldichloroethylene, DDD = dichlorodiphenyldichloroethane, DDT = dichlorodiphenyltrichloroethane, PCB = polychlorinated biphenyl.
[b] Excludes Environmental Monitoring and Assessment Program and Newark Bay samples.

ble 6). Because of the relatively small numbers of samples in which many ERLs were exceeded, the incidence of toxicity also was calculated for several combined ERL categories. In the amphipod tests (n = 777), the incidence of highly toxic responses was 9% with only 1 ERL exceeded, 13% with 1 to 4 ERLs exceeded, 22% with 5 to 9 ERLs exceeded, and peaked at 67% with 15 to 19 ERLs exceeded.

The proportion of samples that was highly toxic in any test performed was 67% when only one ERL was exceeded (Table 6). The incidence of highly toxic samples increased quickly with the number of ERLs that were exceeded, reaching ≥89% when 10 to 14 concentrations were greater than or equal to the ERLs. With several exceptions (notably one sample in which 22 ERLs were exceeded), generally the proportions of samples that were marginally toxic decreased with increases in the number of concentrations greater than or equal to the ERLs.

Among the 233 samples in which all concentrations were less than the TELs; 65% were not toxic, 26% were marginally toxic, and 9% were highly toxic in amphipod tests (Table 7). A total of 62% of samples (n = 26) were not toxic in all tests performed when all concentrations were less than the TELs. The incidence of toxicity did not increase consistently in either amphipod tests alone or in any tests with increases in the number of TELs exceeded. Sample sizes in which multiple bioassays were performed were relatively small and, partly as a consequence, results were highly variable.

Concentrations above and below all ERM and PEL concentrations. Among the 1,068 samples included in this analysis, 777 and 683 had chemical concentrations less than all ERMs and less than all PELs, respectively (Tables 8 and 9). In amphipod tests, 15 and 13%, respectively, of these samples were highly toxic (false negatives). The incidence of highly toxic responses when one or more concentrations was greater than or equal to the ERM or greater than or equal to the PEL was 39 and 35%, respectively, in amphipod tests and 78 and 77%, respectively, in any test performed. With both the marginally and highly toxic responses combined, the incidence of toxicity in samples with concentrations greater than or equal to one or more ERMs or PELs increased slightly to 52 and 48%, respectively, in the amphipod tests and 86.2 and 86.1%, respectively, in any test.

In both the amphipod tests and any tests performed, the incidence of highly toxic responses generally increased and the incidence of marginally toxic responses markedly decreased with increases in the numbers of ERMs or PELs that were exceeded (Tables 8 and 9). The incidence of highly toxic responses in amphipod tests increased from 23% with only 1

Table 6. Incidence of toxicity in either amphipod tests alone or in any test performed among samples with concentrations of 0 to 24 substances greater than or equal to the effects range—low (ERL) values, but all less than the effects range—median (ERM) values

	Amphipod survival only (n = 777)				Any test performed[a] (n = 212)			
No. ERL values exceeded	No. samples	% Not toxic	% Marginally toxic	% Highly toxic	No. samples	% Not toxic	% Marginally toxic	% Highly toxic
0	329	68	21	11	39	46	13	41
1	143	68	23	9	15	13	20	67
2	66	71	15	14	13	46	8	46
3	37	62	22	16	12	42	17	42
4	43	63	16	21	21	33	14	52
5	30	60	17	23	13	38	8	54
6	33	64	12	24	24	21	13	67
7	20	55	35	10	15	27	20	53
8	15	53	27	20	12	8	33	58
9	8	50	13	38	6	0	33	67
10	9	89	0	11	6	0	0	100
11	12	42	25	33	7	0	14	86
12	9	78	11	11	8	0	13	88
13	2	0	0	100	2	0	0	100
14	4	25	50	25	3	0	33	67
15	2	50	50	0	2	0	50	50
17	1	0	0	100	1	0	0	100
18	2	0	0	100	2	0	0	100
19	1	0	0	100	1	0	0	100
20	4	50	0	50	3	0	0	100
21	4	25	25	50	4	0	0	100
22	1	0	100	0	1	0	100	0
23	1	0	0	100	1	0	0	100
24	1	0	0	100	1	0	0	100
1 or more	448	62.7	19.6	17.6	173	20.2	15.6	64.2
1 to 4	289	67.1	20.1	12.8	61	32.8	14.8	52.5
5 to 9	106	58.5	19.8	21.7	70	21.4	18.6	60.0
10 to 14	36	58.3	16.7	25.0	26	0.0	11.5	88.5
15 to 19	6	16.7	16.7	66.7	6	0.0	16.7	77.8
20 to 24	11	27.3	18.2	54.5	10	0.0	10.0	90.0

[a] Excludes Environmental Monitoring and Assessment Program and Newark Bay data.

ERM exceeded to 32% with 1 to 5 ERMs exceeded, to 52% with 6 to 10 ERMs exceeded, and peaked at 85% with ≥11 ERMs exceeded (Table 8). The lowest percent false positives (10%) occurred among samples with 11 to 20 ERMs exceeded. In samples in which multiple bioassays were performed, incidence of highly toxic responses increased from 70% with only 1 ERM exceeded, to 89% with 6 to 10 ERMs exceeded, and peaked at 100% with ≥11 ERMs exceeded. Results were variable among samples with greater than or equal to eight ERMs exceeded because of the small sample sizes.

The predictive ability of the PELs was somewhat lower than that of the ERMs, but, nevertheless, indicated a similar pattern of increasing incidence of highly toxic responses with increasing numbers of PELs exceeded (Table 9). In the amphipod tests, the incidence of highly toxic responses was 14% with 1 PEL exceeded, 24% with 1 to 5 PELs exceeded, 40% with 6 to 10 PELs exceeded, 50% with 11 to 20 PELs exceeded, and 88% with ≥21 PELs exceeded. The lowest percent false positives (17%) occurred among samples with ≥21 PELs exceeded. The proportion of samples showing highly toxic results was much higher when all bioassays were considered, averaging 80% with 6 to 10 PELs exceeded and peaking at 100% with ≥21 PELs exceeded. Percent false positives in any of the tests performed was <25% when one or more PEL was exceeded.

Over ranges in mean SQG quotients. In the preceeding analyses, the methods did not account for the degree to which the chemical concentrations exceeded the different SQGs. That is, samples in which chemical concentrations exceeded SQGs by very different amounts were scored the same. Given similar sediment characteristics and toxicant bioavailability, the probability of toxicity could increase with increasing concentrations. Therefore, to account for both the actual concentrations of individual substances and the combinations of chemicals occurring as mixtures, the predictive abilities of the mean SQG quotients were determined.

The relationships between the incidence of highly toxic responses in the amphipod tests and mean SQG quotients are illustrated in Figures 1 and 2. To clarify these relationships, the chemical concentrations are shown as medians of 39 SQG quotient intervals, each consisting of at least 25 samples. These relationships were considerably more variable when marginally toxic responses were included; therefore, the plots are shown only for highly toxic responses. The incidence of highly toxic responses was most variable and ranged from 0 to 40% among samples with the lowest mean ERM quotients (0.001–0.02) and PEL quotients (0.006–0.05). A gradual, albeit variable, pattern of increasing incidence of toxicity beginning at mean ERM and PEL quotients of 0.04 and 0.07, respectively, was evident. Among samples with mean ERM or PEL quotients ≥1.0 or ≥1.6, respectively, 60 to 80% were highly toxic in the amphipod tests. Percent false positives decreased to <25% with mean ERM or PEL quotients >1.2 or >2.3, respectively.

Some of the samples with the lowest mean ERM and PEL quotients were highly toxic, as indicated in the left tails of the distributions (Figs. 1 and 2). These samples shared very few of the same characteristics. They were scattered among many

Table 7. Incidence of toxicity in either amphipod tests alone or in any test performed among samples with concentrations of 0 to 27 substances greater than or equal to the threshold effects level (TEL) values, but all less than the probable effects level (PEL) values

	Amphipod survival only (n = 683)				Any test performed[a] (n = 142)			
No. TEL values exceeded	No. samples	% Not toxic	% Marginally toxic	% Highly toxic	No. samples	% Not toxic	% Marginally toxic	% Highly toxic
0	233	65	26	9	26	62	15	23
1	102	74	15	12	9	22	11	67
2	67	67	24	9	5	40	20	40
3	62	69	21	10	10	40	30	30
4	46	65	11	24	5	0	0	100
5	28	61	25	14	7	29	14	57
6	15	53	33	13	4	50	0	50
7	10	70	20	10	5	20	20	60
8	15	73	7	20	6	17	0	83
9	5	60	20	20	3	0	0	100
10	12	67	8	25	6	33	17	50
11	11	27	27	45	6	17	17	67
12	15	67	0	33	11	27	0	73
13	10	90	0	10	5	40	0	60
14	7	71	29	0	4	50	25	25
15	3	33	33	33	2	50	50	0
16	4	75	25	0	1	0	100	0
17	2	50	50	0	1	0	0	100
18	4	75	0	25	2	0	50	50
19	8	63	25	13	3	33	0	67
20	5	40	20	40	4	0	25	75
21	4	0	75	25	4	0	25	75
22	3	33	33	33	3	33	33	33
23	9	33	44	22	7	0	43	57
24	1	0	0	100	1	0	0	100
25	1	0	0	100	1	0	0	100
27	1	0	100	0	1	0	100	0
1 or more	450	65.1	19.1	15.8	116	23.3	17.2	59.5
1 to 5	305	68.9	18.4	12.8	36	27.8	16.7	55.6
6 to 9	45	64.4	20.0	15.6	18	22.2	5.6	72.2
10 to 14	55	63.6	10.9	25.5	32	31.3	9.4	59.4
15 to 19	21	61.9	23.8	14.3	9	22.2	33.3	44.4
20 to 27	24	25.0	41.7	33.3	21	4.8	33.3	61.9

[a] Excludes Environmental Monitoring and Assessment Program and Newark Bay data.

Table 8. Incidence of toxicity in either amphipod tests alone or in any of two to four tests performed among samples with concentrations of 0 to 20 substances greater than or equal to the effects range—median (ERM) concentrations

	Amphipod survival only (n = 1,068)				Any test performed[a] (n = 437)			
No. ERM values exceeded	No. samples	% Not toxic	% Marginally toxic	% Highly toxic	No. samples	% Not toxic	% Marginally toxic	% Highly toxic
0	777	65	20	15	212	25	15	60
1	95	59	18	23	69	17	13	70
2	66	52	12	36	62	11	11	77
3	34	56	21	24	30	17	7	77
4	19	32	5	63	10	20	0	80
5	11	45	0	55	9	11	0	89
6	11	55	9	36	10	20	0	80
7	10	40	0	60	8	13	0	88
8	4	25	25	50	4	0	0	100
9	11	9	9	82	7	0	0	100
10	10	50	20	30	6	17	0	83
11	4	0	25	75	1	0	0	100
12	6	33	0	67	3	0	0	100
13	4	0	0	100	3	0	0	100
14	3	0	0	100	1	0	0	100
15	1	0	0	100	1	0	0	100
17	1	0	0	100	0	0	0	0
20	1	0	0	100	1	0	0	100
1 or more	291	47.8	13.4	38.8	225	13.8	8.0	78.2
1 to 5	225	53.3	14.7	32.0	180	15.0	10.0	75.0
6 to 10	46	37.0	10.9	52.2	35	11.4	0.0	88.6
11 to 20	20	10.0	5.0	85.0	10	0.0	0.0	100.0

[a] Excludes Environmental Monitoring and Assessment Program and Newark Bay data.

Table 9. Incidence of toxicity in either amphipod tests alone or in any of two to four tests performed among samples with concentrations of 0 to 20 substances greater than or equal to the probable effects level (PEL) concentrations

	Amphipod survival only (n = 1,148)				Any test performed[a] (n = 517)			
No. PEL values exceeded	No. samples	% Not toxic	% Marginally toxic	% Highly toxic	No. samples	% Not toxic	% Marginally toxic	% Highly toxic
0	683	65	22	13	142	30	17	53
1	106	73	13	14	79	15	10	75
2	49	45	18	37	42	19	14	67
3	36	53	14	33	25	12	16	72
4	17	47	29	24	11	18	36	45
5	16	56	13	31	13	15	0	85
6	10	30	20	50	7	0	0	100
7	4	25	25	50	3	0	33	67
8	18	56	6	39	16	31	0	69
9	11	55	0	45	7	0	0	100
10	9	67	11	22	7	29	0	71
11	19	53	16	32	17	12	0	88
12	8	38	13	50	7	14	0	86
13	13	54	8	38	11	18	0	82
14	8	38	13	50	8	0	13	88
15	11	36	9	55	10	20	0	80
16	8	25	13	63	7	0	14	86
17	7	43	0	57	5	0	0	100
18	10	40	0	60	8	0	13	88
19	5	0	20	80	2	0	0	100
20	3	0	33	67	3	0	0	100
21	2	0	0	100	1	0	0	100
22	5	0	0	100	1	0	0	100
23	5	20	0	80	1	0	0	100
24	4	25	0	75	3	0	0	100
26	1	0	0	100	1	0	0	100
1 or more	385	51.7	13.0	35.3	295	13.9	8.8	77.3
1 to 5	224	60.3	15.6	24.1	170	15.9	12.9	71.2
6 to 10	52	50.0	9.6	40.4	40	17.5	2.5	80.0
11 to 20	92	39.1	10.9	50.0	78	9.0	3.8	87.2
21 to 26	17	11.8	0.0	88.2	7	0.0	0.0	100.0

[a] Excludes Environmental Monitoring and Assessment Program and Newark Bay data.

of the different NOAA and EMAP study areas. These samples often, but not always, had relatively low organic carbon content (<1.0%) and percent fine-grained materials (<50%) and detectable concentrations of butyl tins, chlorinated pesticides, alkyl-substituted PAHs, ammonia, or other substances not accounted for with the SQGs.

DISCUSSION AND CONCLUSIONS

Sediment quality guidelines [1,2] were based upon empirical analyses of data compiled from many different studies. The SQGs were intended to provide informal (nonregulatory), effects-based benchmarks to aid in the interpretation of sed-

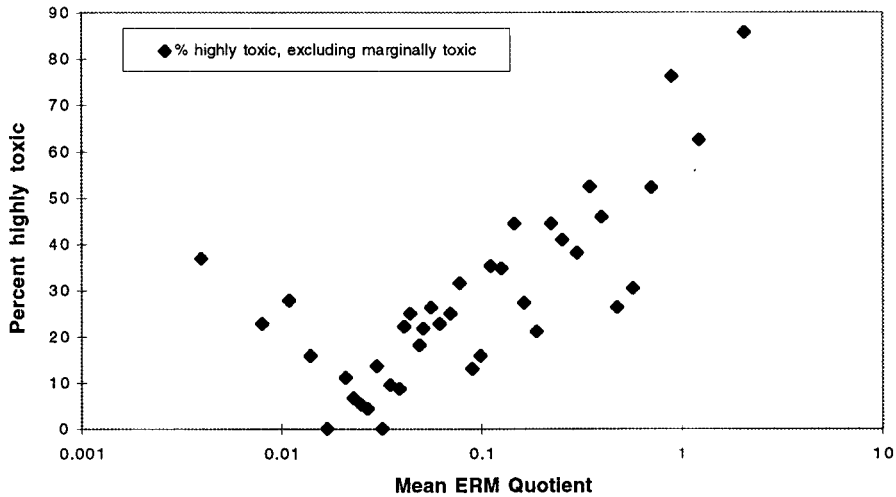


Fig. 1. The relationship between the incidence of toxicity in amphipod survival tests and mean effects range—median (ERM) quotients (plotted as the medians of 39 quotient intervals, each consisting of 25 samples).

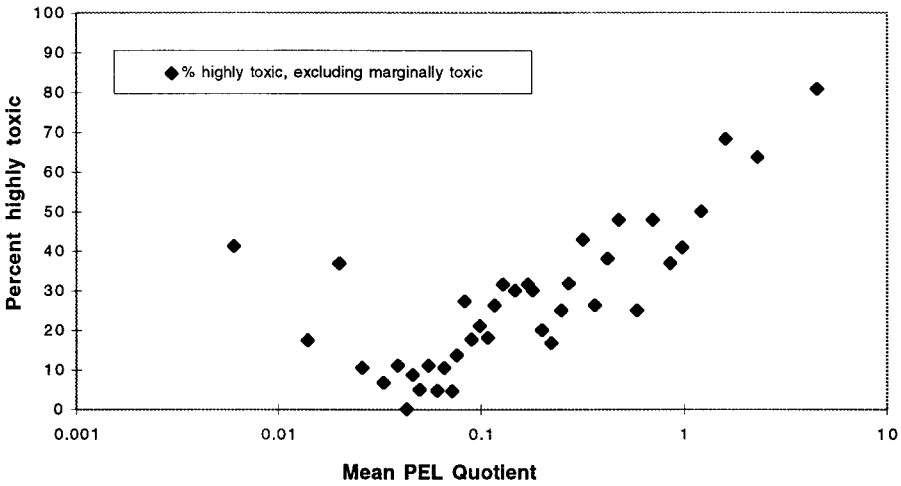


Fig. 2. The relationship between the incidence of toxicity in amphipod survival tests and mean probable effects level (PEL) quotients (plotted as the medians of 39 quotient intervals, each consisting of 25 samples).

iment chemistry data. The ERL and TEL values were intended to represent chemical concentrations below which the probability of toxicity and other effects was minimal. In contrast, the ERM and PEL values were intended to represent mid-range concentrations above which adverse effects were more likely, although not always expected. Intermediate frequencies of effects were expected at chemical concentrations between the ERLs and ERMs and between the TELs and PELs. In this analysis of independent data sets, we attempted to determine if the incidence of toxicity in selected, acute laboratory bioassays would follow the same pattern as observed with multiple measures of effects in the databases used to derive the guidelines.

The majority of the data compiled to develop the guidelines was generated in field studies in which different chemical mixtures were encountered. In these field studies causality could not be determined. The intent of this study was to also use data from surveys of numerous saltwater areas to determine the frequency with which the guidelines correctly predicted nontoxic and toxic conditions.

Unlike SQGs based upon the apparent effects threshold approach [26], the ERL/ERMs and TEL/PELs were not intended to represent concentrations above which adverse effects were always observed. Because the ERLs and TELs were intended to represent conservative concentrations below which toxicity was not frequently expected, we estimated the frequency of false negatives as the incidence of toxicity among samples in which all concentrations were lower than these values. Earlier [1,2], as a measure of reliability, we reported that the frequency of false negatives among the data sets used to derive the guidelines was ≤25% for most chemicals and ≤10% for many chemicals. Specifically, at concentrations below the individual ERL and TEL values for nine trace metals, the incidence of effects ranged from 1.9 to 9.4% and from 2.7 to 9.0%, respectively. For organic compounds, the incidence of effects was more variable, ranging from 5.0 to 27.3% for 19 ERLs and from 0.0 to 47.6% for 25 TELs when concentrations were below these levels.

The same criterion (≤25% false positives) previously used for estimates of reliability was used as the target for estimates of predictive ability in this analysis. Based upon the highly toxic responses, the ERLs and TELs indicated 11 and 9% false negatives (toxicity observed when not expected), respectively, in the tests of amphipod survival, thus bettering the target of ≤25%. The incidence of false negatives also was relatively low (41 and 23% for the ERLs and TELs, respectively) in any one of the two to four tests performed. Based again upon the highly toxic responses, the incidences of false negatives in amphipod tests were, as expected, slightly higher (15 and 13%, respectively) for the ERMs and PELs than for the ERLs and TELs. Therefore, the probabilities of highly toxic responses in amphipod survival tests are relatively low (≤16%) among samples in which all chemical concentrations are lower than both sets of SQGs. However, the incidences of false negatives among any of the tests performed were 60 and 53% (highly toxic responses) for the ERMs and PELs, respectively. These data suggest that there remains a moderate probability of toxicity among samples with all chemical concentrations less than the ERMs or less than the PELs if a battery of relatively sensitive, sublethal bioassays is considered.

In the amphipod tests, the incidences of highly toxic responses and total toxic responses were 18 to 20% and 16 to 19%, respectively, when one or more chemicals exceeded the ERLs and/or TELs. These results agreed well with the original intent of the ERLs and TELs as indicators of the lower end of the possible effects range. These results also agreed very well with the estimates of reliability (calculated with the database used to derive the SQGs) for most ERLs and TELs (30–50% effects) [1,2]. However, when predictive ability was estimated with data from more sensitive sublethal tests, toxicity was observed much more frequently than in the amphipod tests alone.

The ERMs and PELs were derived as mid-range points within the distributions of effects data for each chemical. The ERMs were calculated as the medians (50th percentiles) of chemical concentrations associated with measures of adverse effects. The derivation of the PELs incorporated both the no-effects data along with effects data into the calculations of mid-range concentrations. Neither set of guidelines was intended as a toxicity threshold above which effects were always expected. The incidence of highly significant toxicity in the amphipod survival tests among samples that exceeded individual ERMs and PELs generally agreed with the intent of these values (i.e., as mid-range values). That is, 40 to 65% of

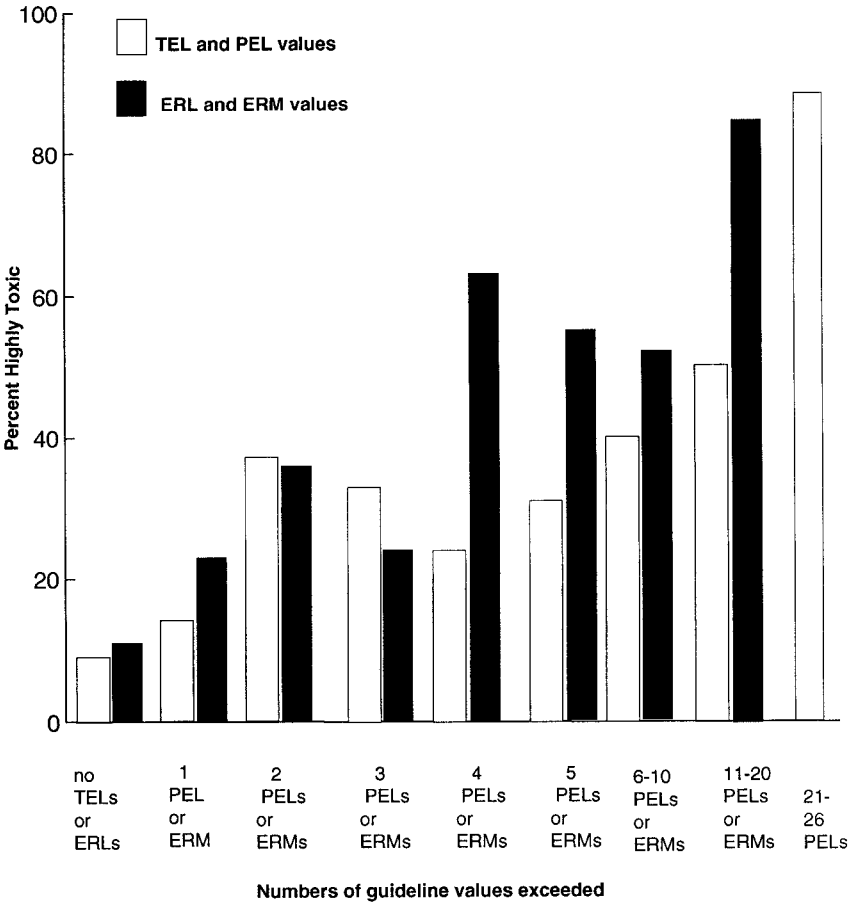


Fig. 3. Summary of the predictive ability of threshold effects level/probable effects level (TEL/PEL) values and effects range—low/effects range—median (ERL/ERM) values in amphipod survival tests (as percent highly toxic among the total numbers of samples).

the samples were highly toxic in amphipod tests at concentrations above most of these individual values. Also, the incidence of total toxicity (marginally + highly toxic) was 52 and 48% when the concentrations of one or more chemicals exceeded ERMs and PELs, respectively. When results from any one of a battery of bioassays were considered, the percentages of samples that were highly toxic increased remarkably to ≥85% for 19 of the ERMs and for 19 of the PELs and to 77 to 78% when one or more ERMs and/or PELs were exceeded.

In all analyses performed on the predictive ability of the SQGs, the percentages of samples demonstrating toxicity were lowest when either no chemicals or the least number of chemicals exceeded the lower range guidelines and increased with increases in the numbers of mid-range guidelines that were exceeded (Fig. 3). Results were variable at intermediate concentrations, but, nevertheless, the data indicated an overall pattern of increasing incidence of toxicity with increasing numbers of ERMs and PELs exceeded. Percent false positives in amphipod tests (no toxicity observed when toxicity was expected) dropped to <25% among samples in which 11 to 20 ERMs (n = 20) and 21 to 26 PELs (n = 17) were exceeded.

Because the two sets of SQGs were derived with slightly different procedures, one objective of this evaluation was to compare their predictive ability. The results indicated that the two sets of SQGs were very similar in predicting toxicity (Fig. 3). The percentages of false negatives for the ERLs and TELs were 11 and 9%, respectively, in the amphipod tests. The incidences of highly toxic responses in amphipod tests were slightly higher for the PELs than for the ERMs among samples in which two or three chemicals exceeded the guideline concentrations. Otherwise, the incidence of toxicity often was higher when chemical concentrations exceeded the ERMs as compared to when the concentrations exceeded the PELs.

Based upon these data, users of the SQGs can identify the probability that their samples would be toxic by comparing the chemical concentrations in their samples to the appropriate SQGs and then to the incidence of toxicity shown in this paper. For example, highly toxic responses would be expected in amphipod survival tests in only approximately 9 to 11% of the samples when all chemical concentrations are below the TELs or ERLs (Fig. 3). Among samples in which only one ERL or TEL value is exceeded and no other chemicals exceeded any other ERL/ERMs or TEL/PELs, toxicity in amphipod tests would be expected in only 9 and 12% of the samples, respectively.

The probability of toxicity in amphipod survival tests is not very high (23 and 14%, respectively) among samples in which only one ERM or only one PEL value is exceeded (Fig. 3). However, the probabilities of toxicity increase with the number of ERMs and PELs exceeded. Based upon the results of this evaluation (n = 1,068), users can expect toxicity in a large majority of samples, that is, in >85% of the samples in amphipod tests (n = 20, n = 17) and in 100% of samples in any one of a battery of sensitive bioassays (n = 9 or 6) when 11 or more ERMs or 21 or more PELs are exceeded. Therefore,

Table 10. Incidence of toxicity in amphipod tests only within three ranges in mean sediment quality quideline quotients

	No. samples	% Not toxic	% Marginally toxic	% Highly toxic
Mean effects range—median quotients				
<0.1	653	67.8	20.5	11.6
0.11 to 1.0	364	51.6	16.5	31.9
>1.0	51	23.5	5.9	70.6
Mean probable effects level quotients				
<0.1	481	67.6	22.0	10.4
0.11 to 1.0	474	58.6	17.1	24.3
>1.0	113	35.4	8.8	55.8

the probability of incorrectly classifying samples as toxic would be 15 and 0%, respectively, in these highly contaminated samples.

The data from the analyses of the mean SQG quotients suggest that the probability of observing toxicity was a function of not only the number of guidelines exceeded but the degree to which they were exceeded. Therefore, the probabilities of highly toxic responses would be relatively low (<12% in amphipod tests) among samples with mean SQG quotients <0.1 (Table 10). The probabilities of toxicity increase to 32 and 24%, respectively, with mean ERM and PEL quotients of 0.11 to 1.0 and increase again to 71 and 56%, respectively, with quotients >1.0.

Despite the selection of high-quality data sets from NS&TP and EMAP—Estuaries studies, the analyses of predictive ability had a number of limitations or potential sources of error. Different results may have been obtained if other data had been used in this evaluation of predictive ability.

The core bioassay upon which these analyses focused was the amphipod survival test. This bioassay has become the most widely applied sediment toxicity test in North America and provides important information for many research, monitoring, and management programs. Amphipod survival tests have been used in both the derivation and field validation of various guidelines [22,26]. However, because different taxa have different sensitivities to toxicants, the use of a battery of toxicity tests is widely accepted and highly recommended in sediment quality assessments [27]. Furthermore, the use of multiple tests increases the number of surrogates of sediment-dwelling taxa. Considerable gains in predictive ability were attained by the addition of data from other tests to those from the amphipod tests. Because only one, two, or three (not, say, 10) tests accompanied the amphipod bioassays, we attribute the gains in predictive ability not to the number of tests performed, but, rather, to the greater sensitivity of the tests to the chemicals in the sediments.

Tests of invertebrate gametes and embryos exposed to pore waters and bioluminescent bacteria exposed to solvent extracts have been used widely in U.S. estuaries [24] and generally are more sensitive than are test with amphipods to the same samples. The large differences in sensitivity between the amphipod survival tests and the other tests performed is reflected in the data that were analyzed. The probabilities of observing toxicity in the more sensitive sublethal tests would be much higher than in the amphipod tests. Users are advised to consider the data from both categories of bioassays when using the guidelines, especially because highly sensitive tests such as those performed with urchin gametes [25] and bioluminescent bacteria [28] have shown strong associations with chemical concentrations.

Sediment quality guidelines were not available for many substances that were measured in the samples. Some substances may have occurred at concentrations above toxicologic thresholds. Other substances that were not measured probably occurred in many or all samples. Also, some samples may have had high concentrations of ammonia and hydrogen sulfide that covaried with anthropogenic substances and contributed to toxicity. Together, the effects of these substances may have contributed to the false negatives observed. However, our nationwide experience indicates that toxicants often covary with each other to a large degree [7,25] and the quantified substances for which SQGs were available should have served as reasonable surrogates for the covariates. Furthermore, our experience in assessments of surficial sediments suggests that ammonia and sulfides occur in either pore water or overlying water in test chambers at toxicologically significant concentrations in <10% of the samples. Nevertheless, the contribution of all potentially toxic substances in the samples could not be accounted for.

Although standardized and widely accepted methods and protocols were used, some interlaboratory and interstudy differences in methods may have occurred. Some variability in results may have been attributable to merging data from different studies and geographic areas. For example, data were compiled from tests performed with two species of amphipods to increase the sample size and to include data from all three coastlines. Differences in sensitivity between these two amphipod species may have contributed to variability in the results. Also, variability may have been increased by merging data from different species of urchins and molluscs along with data from the Microtox tests into one category.

Most of the samples were not collected within hazardous waste sites and most were not highly contaminated (zero to five SQGs exceeded). The relatively small numbers of highly contaminated samples appeared to contribute to variability in results. Additional data from highly contaminated sites would be useful in further clarification of predictive ability.

Despite these potential limitations of this study, the predictive ability estimated with these data often matched their previously reported reliability. Also, the results of this analysis agreed relatively well with the the estimates of reliability reported [5] for freshwater sediment effects concentrations. The results of this analysis [5] determined type I (false positive) and type II (false negative) errors for freshwater ERL/ERM and TEL/PEL values based upon data from individual samples from numerous studies. For most substances, the errors ranged from 5 to 30%. The paired sets of values, however, differed somewhat in absolute concentrations and error rates.

The toxicity/chemistry relationships observed in this study may not apply in all situations, especially in sediments in which contaminants are found in forms such as copper slag [29] or coal pitch in organically enriched mud [30]. The guidelines are most useful when applied to fine-grained, sedimentary deposits such as those sampled during the NOAA and EMAP—Estuaries studies.

In conclusion, the results of these analyses indicate the following: the probabilities of highly toxic responses occurring in amphipod survival tests among samples in which all chemical concentrations are less than ERLs and/or TELs are 9 to 11%; the probabilities of highly toxic responses occurring in amphipod survival tests among samples in which mean SQG

quotients are <0.1 are 10 to 12%; the probabilities of highly toxic responses occurring when one or more ERLs or TELs are exceeded and no ERMs or PELs are exceeded are 16 to 18% in amphipod tests alone and 60 to 64% in any one of a battery of sensitive tests performed; the probabilities of either marginally or highly toxic responses occurring are 48 to 52% in amphipod tests and 86% in any one of a battery of sensitive tests performed when concentrations exceed one or more ERMs or PELs; consistent with their original intent, the ERMs and PELs are considerably better at predicting toxicity than are the ERLs and TELs. Furthermore, the probabilities of toxicity occurring generally increase with increasing numbers of chemicals that exceed the ERM and PEL concentrations; the probabilities of toxicity occurring generally increase with increasing mean SQG quotients; and the incidence of false negatives is slightly lower for the TELs than for the ERLs, but the incidence of false positives is generally higher for the PELs than for the ERMs; however, there is good overall agreement in the predictive ability of the TEL/PELs and the ERL/ERMs.

Based upon these analyses of predictive ability and previous analyses of reliability, it appears that the SQGs provide reasonably accurate estimates of chemical concentrations that are either nontoxic or toxic in laboratory bioassays. However, we urge that all SQGs should be used with caution, because, as observed in this analysis, they are not perfect predictors of toxicity. Especially among samples with intermediate chemical concentrations, the SQGs are most useful when accompanied by data from in situ biological analyses, other toxicologic assays, other interpretive tools such as metals : aluminum ratios, and other guidelines derived either from empirical approaches and/or cause–effects studies.

Acknowledgement—Funding for much of the NOAA data collection was provided by the Coastal Ocean Program and the National Status and Trends Program, both of NOAA. Helpful comments on an initial version of the mansucript were provided by Chris Ingersoll, Jeff Hyland, R. Scott Carr, Rick Swartz, and Tom O'Connor. The EMAP data were provided by the U.S. EPA; database development was provided by the Office of Research and Development, U.S. EPA, and the Hazardous Materials Response and Assessment Division, NOAA. Funding for data analyses was provided by the Hazardous Materials Response and Assessment Division, and Coastal Monitoring and Bioeffects Assessment Division, NOAA. Corinne Severn and Carolyn Hong provided valuable assistance in data management.

REFERENCES

1. Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manage* 19: 81–97.
2. MacDonald DD, Carr RS, Calder FD, Long ER. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5:253–278.
3. Long ER, MacDonald DD. 1992. National Status and Trends Program approach. Sediment classification methods compendium. 823-R-92-006. U.S. Environmental Protection Agency, Washington, DC.
4. Smith SL, MacDonald DD, Keenleyside KA, Ingersoll CG, Field LJ. 1996. A preliminary evaluation of sediment quality assessment values for freshwater sediments. *J Great Lakes Res* 22:624–638.
5. Ingersoll CG, et al. 1996. Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironomus riparius. J Great Lakes Res* 22:602–623.
6. Long ER, Wolfe DA, Scott KJ, Thursby GB, Stern EA, Peven C, Schwartz T. 1995. Magnitude and extent of sediment toxicity in the Hudson–Raritan estuary. NOAA Technical Memorandum NOS ORCA 88. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
7. Wolfe DA, Bricker SB, Long ER, Scott KJ, Thursby GB. 1994. Biological effects of toxic contaminants in sediments from Long Island Sound and environs. NOAA Technical Memorandum NOS ORCA 80. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
8. Long ER, Sloane GM, Carr RS, Scott KJ, Thursby GB, Wade TL. 1994. Sediment toxicity in Boston Harbor: Magnitude, extent, and relationships with chemical toxicants. NOAA Technical Memorandum NOS ORCA 96. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
9. Long ER, et al. 1994. Magnitude and extent of sediment toxicity in Tampa Bay, Florida. NOAA Technical Memorandum NOS ORCA 78. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
10. Fairey R, et al. 1996. Chemistry, toxicity and benthic community conditions in sediments of the San Diego Bay region. Final Report. California State Water Resources Control Board, Sacramento, CA, USA.
11. Sapudar RA, et al. 1994. Sediment chemistry and toxicity in the vicinity of the Los Angeles and Long Beach harbors. Project Report. California State Water Resources Control Board, Sacramento, CA, USA.
12. Schimmel SC, Melzian BD, Campbell DE, Strobel CJ, Benyi SJ, Rosen JS, Buffum HW. 1994. Statistical summary. EMAP—Estuaries. Virginian Province—1991. EPA 620/R-94/005. U.S. Environmental Protection Agency, Washington, DC.
13. Strobel CJ, Buffum HW, Benyi SJ, Petrocelli EA, Reifsteck DR, Keith DJ. 1995. Statistical summary. EMAP—Estuaries. Virginian Province—1990 to 1993. EPA 620/R-94/-26. U.S. Environmental Protection Agency, Narragansett, RI.
14. Summers JK, Macauley JM. 1993. Statistical summary: EMAP—Estuaries Louisianian Province—1991. EPA Report 600/R-93-001. U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.
15. Lauenstein GG, Cantillo AY. 1993. Sampling and analytical methods of the National Status and Trends Program, National Benthic Surveillance and Mussel Watch projects. 1984–1992. Vol 1. Overview and summary of methods. NOAA Technical Memorandum NOS ORCA 71. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
16. American Society for Testing and Materials. 1990. Standard guide for conducting solid phase, 10-day, static sediment toxicity tests with marine and estuarine infaunal amphipods. E 1367-90. In *Annual Book of ASTM Standards.* Philadelphia, PA.
17. American Society for Testing and Materials. 1993. Standard guide for conducting solid phase, 10-day, static sediment toxicity tests with marine and estuarine infaunal amphipods. E 1367-92. In *Annual Book of ASTM Standards.* Philadelphia, PA.
18. Carr RS, Chapman DC. 1992. Comparison of solid-phase and pore water approaches for assessing the quality of marine and estuarine sediments. *Chem Ecol* 7:19–30.
19. U.S. Environmental Protection Agency. 1990. Recommended protocols for conducting laboratory bioassays on Puget Sound sediments. Revised. Puget Sound Estuary Program. Seattle, WA.
20. Schiewe MH, Hawk EG, Actor DI, Krahn MM. 1985. Use of a bacterial bioluminescence assay to assess toxicity of contaminated marine sediments. *Can J Fish Aquat Sci* 42:1244–1248.
21. Mearns AJ, Swartz RC, Cummins JM, Dinnel PA, Plesha P, Chapman PM. 1986. Inter-laboratory comparison of a sediment toxicity test using the marine amphipod, *Rhepoxynius abronius. Mar Environ Res* 19:13–37.
22. Swartz RC, Schults DW, Ozretich RJ, Lamberson JO, Cole FA, DeWitt TH, Redmond MS, Ferraro SP. 1995. SigmaPAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments. *Environ Toxicol Chem* 14: 1977–1987.
23. Thursby GB, Heltshe J, Scott KJ. 1997. Revised approach to toxicity test acceptability criteria using a statistical performance assessment. *Environ Toxicol Chem* 16:1322–1329.
24. Long ER, Robertson A, Wolfe DA, Hameedi J, Sloane GM. 1996. Estimates of the spatial extent of sediment toxicity in major US estuaries. *Environ Sci Technol* 30:3585–3592.
25. Carr RS, Long ER, Windom HL, Chapman DC, Thursby G, Sloane GM, Wolfe DA. 1996. Sediment quality assessment studies of Tampa Bay, Florida. *Environ Toxicol Chem* 15:1218–1231.
26. Barrick R, Becker S, Brown L, Beller H, Pastorok R. 1988. Sed-

iment quality values refinement: 1988 update and evaluation of Puget Sound AET. Vol 1. PTI Environmental Services, Bellevue, WA, USA
27. Ingersoll CG, et al. 1997. Workgroup summary report on uncertainty evaluation of measurment endpoints used in sediment ecological risk assessment. In Ingersoll CG, Dillon T, Biddinger GR, eds, *Ecological Risk Assessment of Contaminated Sediment*. SETAC, Pensacola, FL, USA.
28. Johnson BJ, Long ER. 1998. Rapid toxicity assessment of sediments from large estuarine ecosystems: A new tandem in vitro testing approach. *Environ Toxicol Chem* (in press).
29. Lee GF, Jones RA. 1991. Overview: Aquatic life risk assessment in copper-contaminated sediments at National City marine terminal. Woodward-Clyde Consultants, San Diego, CA, USA.
30. Paine MD, Chapman PM, Allard PJ, Murdoch MH, Minifie D. 1996. Limited bioavailability of sediment PAH near an aluminum smelter: Contamination does not equal effects. *Environ Toxicol Chem* 15:2003–2018.

SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 19, No. 10, pp. 2598–2601, 2000
Printed in the USA
0730-7268/00 $9.00 + .00

CLASSIFYING PROBABILITIES OF ACUTE TOXICITY IN MARINE SEDIMENTS WITH EMPIRICALLY DERIVED SEDIMENT QUALITY GUIDELINES

EDWARD R. LONG,*† DONALD D. MACDONALD,‡ CORINNE G. SEVERN,§ and CAROLYN B. HONG§
†National Oceanic and Atmospheric Administration, National Centers for Coastal Ocean Science, 7600 Sand Point Way NE, Seattle, Washington 98115, USA
‡MacDonald Environmental Sciences Ltd., 2376 Yellow Point Road, R.R. 3, Nanaimo, British Columbia, Canada V9X 1W5
§EVS Environment Consultants, 200 West Mercer Street, Suite 403 Seattle, Washington 98119, USA

(*Received* 15 *June* 1999; *Accepted* 18 *April* 2000)

Abstract—Matching, marine sediment chemistry, and toxicity data (n = 1,513), compiled from three studies conducted in the United States, were analyzed to determine both the frequency of acute toxicity to amphipods and average percentage survival in laboratory bioassays within ranges in toxicant concentrations. We determined that the probability of observing acute toxicity was relatively low (<10%) and that average control-adjusted survival equaled or exceeded 92% in samples in which sediment quality guidelines were not exceeded. Both the incidence of toxicity increased and average survival decreased as chemical concentrations increased relative to the guidelines. In sediments with highest contaminant concentrations, 73 to 83% of the samples were highly toxic, and average control-adjusted amphipod survival was 37 to 46%. Results of this study confirm that the relationships between sediment chemical concentrations and toxicity reported in a previous study were robust. Further, they indicate that numerical guidelines for saltwater sediments can be used to estimate the probability of observing toxic effects in acute amphipod tests.

Keywords—Sediment quality guidelines Sediment toxicity

INTRODUCTION

Empirically derived sediment quality guidelines (SQGs) were developed on the basis of the associations observed between measures of adverse biological effects and the concentrations of potentially toxic substances in marine and estuarine sediments. Two sets of SQGs developed for saltwater include the effects range—low (ERL) and effects range—median (ERM) values for 25 chemicals [1] and the threshold effects level (TEL) and probable effects level (PEL) values for 31 chemicals [2]. The relative abilities of individual SQGs to identify chemical concentrations that were either rarely or frequently associated with adverse effects were previously reported [1,2], using the data with which the guidelines were derived. Subsequently, the predictive abilities of the SQGs were quantified [3] using an independent data set (i.e., not used in the guidelines derivation). Data used to calculate the predictive abilities of the SQGs were compiled in a large database from studies conducted in estuaries of the Atlantic, Gulf of Mexico, and Pacific coasts. All these studies provided matching data on chemical contamination and acute toxicity using comparable methods (n = 1,068).

Recommended applications of the SQGs were based on observations of their predictive abilities [4]. In that study, four independent indices of chemical contamination were identified that defined four categorical ranges in the incidence of acute toxicity in amphipod survival tests. Acute toxicity measured in laboratory bioassays occurred rarely (10–12% of samples) in sediments in which none of the guidelines was exceeded. The percentages of samples in which acute toxicity was observed increased with the numbers of guidelines exceeded. In addition, they increased with the concentrations of mixtures of substances normalized to the guidelines (expressed as mean SQG quotients). This classification system provided users of the guidelines with a basis for estimating the probabilities that sediment samples would be acutely toxic in laboratory tests using the data on chemical concentrations alone.

The primary objective of this paper was to further evaluate the predictive abilities of the SQGs using an expanded database and therefore to determine the robustness of the chemistry/toxicity relationships observed in the previous study. Another objective was to summarize the probabilities of observing acute toxicity to amphipods on the basis of evaluations of the sediment chemistry data, using the guidelines. Finally, to determine the applicability of the guidelines in sediments with somewhat atypical geochemical properties, we analyzed data available from Biscayne Bay, Florida, USA, and Pearl Harbor, Hawaii, USA.

METHODS

Data from solid-phase tests of sediments were selected for these evaluations because of their relatively low uncertainty and many other favorable attributes [5]. The most data available nationwide are from tests performed with the amphipod *Ampelisca abdita.* A database of matching sediment chemistry and toxicity data was compiled from three sources. The majority of the data (consisting of information from 1,068 samples collected in estuaries of the Atlantic, eastern United States; Gulf of Mexico, southern United States; and Pacific coast, western United States) was compiled for a previous analysis of the predictive ability of sediment guidelines [3]. These data are hereafter referred to as the national data set.

Additional data (n = 226) were obtained from a study conducted during 1995–1996 in Biscayne Bay by the U.S. National Oceanic and Atmospheric Administration [6]. Another data set (n = 218) was obtained from a study of Pearl Harbor

* To whom correspondence may be addressed (ed.long@noaa.gov).

Table 1. Percentage incidence of highly toxic samples and average percentage amphipod survival in marine sediment samples classified according to numerical sediment quality guidelines[a]

Chemical characteristics relative to sediment guidelines	% Highly toxic[b] samples in amphipod survival tests				Average control-adjusted amphipod survival			
	National database[c] (n = 1,068)	Biscayne Bay, FL (n = 226)	Pearl Harbor, HI (n = 219)	Combined summary (n = 1,513)	National database[c] (n = 1,068)	Biscayne Bay, FL (n = 226)	Pearl Harbor, HI (n = 219)	Combined summary (n = 1,513)
Category 1								
Mean ERM quotients <0.1	11	3	4	9	93	95	94	93
Mean PEL quotients <0.1	10	3	6	8	93	95	92	93
No ERLs exceeded	11	5	14	9	92	94	93	92
No TELs exceeded	9	6	0	8	92	93	102	92
Category 2								
Mean ERM quotients 0.11–0.5	30	15	5	21	81	85	97	86
Mean PEL quotients 0.11–1.5	25	23	12	21	84	81	93	86
1–5 ERMs exceeded	32	46	24	32	79	66	86	79
1–5 PELs exceeded	24	37	2	18	83	70	99	88
Category 3								
Mean ERM quotients 0.51–1.5	46	73	61	49	74	48	68	70
Mean PEL quotients 1.51–2.3	50	67	33	49	66	46	76	68
6–10 ERMs exceeded	52	67	71	57	63	47	66	59
6–20 PELs exceeded	47	59	48	48	71	57	73	70
Category 4								
Mean ERM quotients >1.5	75	75	100[d]	76	43	31	44	41
Mean PEL quotients >2.3	77	75	33[e]	73	47	31	56	46
>10 ERMs exceeded	85	75	33[e]	80	41	31	56	41
>20 PELs exceeded	88	75	50[d]	83	38	10	45	37

[a] ERM = effects range—median; PEL = probable effects level; ERL = effects range—low; TEL = threshold effects level.
[b] Mean survival significantly different from controls and <80% of controls.
[c] [3].
[d] 100% (two of two) either marginally toxic ($p < 0.05$, mean survival ≥80% of control) or highly toxic.
[e] 100% (three of three) either marginally toxic ($p < 0.05$, mean survival ≥80% of control) or highly toxic.

conducted in 1997 by the U.S. Navy (unpublished data). These two data sets were selected for evaluation primarily because they were developed with the same methods used in the previous analysis [4] for sample collections, chemical analyses, and amphipod toxicity tests. They were selected also because of the somewhat unusual geochemical properties in both bays. Most of the samples from Biscayne Bay were very sandy and composed primarily of carbonate shell debris. Other samples from adjoining canals were fine-grained silts high in organic carbon content. The Pearl Harbor samples consisted primarily of fine-grained materials accumulated from erosion of iron-rich soils and lava.

Amphipod survival data are reported as percentage of nontoxic controls (average control-adjusted survival). Test results were considered to be nontoxic when mean survival in a sample was not significantly different ($p > 0.05$) from those performed with negative controls. As previously described [3], results were considered as marginally toxic when mean survival was significantly lower than in negative controls ($p < 0.05$) but exceeded 80% of the controls. Highly toxic samples were those in which survival was significantly lower than in controls and <80% of mean survival in controls. The statistical significance of average amphipod survival of less than 80% in classification of sediments as toxic was determined in power analyses [7]. Sediments in which average control-adjusted amphipod survival is below 80% often are defined as toxic in dredging studies [8]. Previously described methods [3] were used to calculate mean SQG quotients.

Statistical methods previously described [3,4] to evaluate the predictive abilities of the SQGs were used in the data analyses. In the previous study [4], we determined the chemical concentrations relative to the SQGs that resulted in four categories of toxicity (i.e., indices of ~10%, ~25%, ~50%, and ~75%). Samples were classified as such with four independent indices: numbers of ERLs/ERMs exceeded, numbers of TELs/PELs exceeded, mean ERM quotients, and mean PEL quotients. Following the previously used procedures [4], we determined the percentages of samples that were highly toxic to amphipods with the new data for each of the four indices of chemical contamination within each of the four classification categories. In addition, the average percentage survival of the amphipods in the toxicity tests was determined for all samples included in each category. Data from the national data set (n = 1,068) were combined with those from Biscayne Bay and Pearl Harbor to provide new estimates (n = 1,513) of the probabilities of observing acute toxicity.

RESULTS

Results of the analyses conducted with the national database indicated that only 9 to 11% of samples were highly toxic in samples classified as Category 1, that is, in which none of the ERLs or TELs were exceeded or mean SQG quotients were <0.1 (Table 1). Average control-adjusted survival in these samples ranged from 92 to 93%, well above the critical threshold of 80% survival.

In the national database, the incidence of highly toxic responses increased sequentially from 9 to 11% in Category 1 to 24 to 32% in Category 2, to 46 to 52% in Category 3, and to 75 to 88% in Category 4 as chemical concentrations increased (Table 1). Average amphipod survival decreased se-

quentially in these samples from 92 to 93%, to 79 to 84%, to 63 to 74%, and to 38 to 47%. It is noteworthy that average survival approximated the critical threshold of 80% in Category 2 samples.

The data from the Biscayne Bay study followed a pattern similar to that of the national database (Table 1). However, the percentages of toxic samples in Category 1 (3–6%) were slightly lower than in the national data base (9–11%). Average survival was less than 80% in samples with 1 to 5 ERMs or PELs exceeded. In addition, as compared to the national database, the percentages of samples in Biscayne Bay that were highly toxic were higher and the average amphipod survival was lower in samples with chemical characteristics equivalent to those in the national database of Category 3.

In the Pearl Harbor data set, both the incidence of toxicity and the average control-adjusted survival were more variable within categories than they were in the national or Biscayne Bay data sets (Table 1). Within Category 1, the incidence of toxicity and average survival varied from 0 to 14% and from 92 to 102%, respectively. Similar variability was apparent in the Category 2, 3, and 4 samples. However, the incidence of toxicity generally increased between categories, and average survival decreased to 66 to 76% in Category 3 and to 44 to 56% in Category 4. Only two or three samples were classified in Category 4, and this reduced our ability to interpret these data. Nevertheless, amphipod survival was at least significantly reduced relative to controls in all the Category 4 samples from Pearl Harbor.

With the data from the three studies combined (n = 1,513), the patterns of both increasing incidence of highly toxic responses and decreasing average survival relative to increasing chemical concentrations were similar to those observed in the national database (Table 1). In samples not expected to be toxic (Category 1) 8 to 9% of samples were highly toxic, and average survival was ≥92%. In the Category 2 samples, however, average incidence of toxicity was somewhat lower and average survival slightly higher than in the national database. Therefore, in the three different data sets, average amphipod survival was either slightly above or below the control-adjusted threshold of 80% in sediments with chemical characteristics equal to those of Category 2. In Category 3 sediments, approx. one-half of the samples were highly toxic, and average survival was ≤70%. In samples with the highest chemical concentrations, 73 to 83% were highly toxic, and average survival was ≤46%.

DISCUSSION AND CONCLUSION

In summarizing his observations of the many attempts by scientists to establish causal links between anthropogenic stresses and effects on biological resources, Sindermann [9] suggested the principal criteria for establishing such relationships were (1) to establish the statistical relationship and demonstrate clear differences between exposed and unexposed populations and (2) to demonstrate that the relationships are consistent over space and time; (3) establish the precision with which the relationship is specific to either species, location, or time; (4) establish that exposure precedes the response, not vice versa; (5) determine that the relationship is plausible and consistent with previous hypotheses; (6) quantify the statistical significance or probability of the relationship; and (7) describe the predictive performance drawn from the observed association. However, he cautioned that satisfaction of all these criteria did not necessarily constitute scientific proof of causality but rather provided a weight of evidence with which to infer a causal relationship.

In this paper, the relationships between the concentrations of chemical substances in sediments and the incidence of toxicity in laboratory tests were quantified using numerical guidelines. The relationships satisfied many of Sindermann's criteria. The data indicated clear differences in toxicity between exposed and unexposed populations, that is, in samples with no SQGs exceeded versus those with many (e.g., >10 ERMs) exceeded. Generally, the incidence of toxicity and average survival changed very little (<5%) by addition of the new data to the national database, demonstrating that the relationships were robust and consistent over space and were not location-specific. Increasing toxicity (response) was related to and tracked with increasing chemical concentrations (exposure), and this relationship constituted a plausible exposure–response association consistent with what had been previously reported. The data assembled in Table 1 summarize the likelihood (probability) of acute toxicity occurring within ranges in chemical concentrations and thus provide a set of predictive tools for assessing the relative quality of sediments.

The data from this study indicate that there is a very low likelihood of acute toxicity in amphipod survival tests and that average survival likely will exceed 80% in sediments with all chemical concentrations below the ERL or TEL values or with mean ERM or mean PEL quotients <0.1. Therefore, the probabilities of misclassifying such samples as acceptable or likely nontoxic or background when they are actually toxic are very small (<10%). In contrast, there is a much higher probability (≥48%) that samples would be toxic in which six or more ERM or PEL values are exceeded or in which mean ERM quotients exceed 0.5 or mean PEL quotients exceed 1.5. Samples with chemical characteristics equivalent to those of Category 2 sediments may or may not be acutely toxic, depending on their specific geochemical characteristics. Therefore, these are sediments with the least certainty of accurately predicting toxicity.

The data indicated that a small degree of variability in toxicological responses that is likely attributable to regional differences in the geochemistry of sediments and the relative bioavailability of sediment-associated toxicants can lead to differences in the predictive abilities of sediment guidelines. Contaminants in the coarse carbonate sands of Biscayne Bay would be expected to be much more bioavailable than those found in the silty clays of other estuaries. The relatively high incidence of toxicity and relatively low percentage survival in Category 2 and 3 samples from Biscayne Bay seemed to substantiate this hypothesis. Users of SQGs should be aware of this regional variability when classifying samples as acceptable and unacceptable with the guidelines. However, despite these slight regional differences in numerical results, the data showed the same pattern in all areas; that is, the probability of acute toxicity generally increased with increasing chemical contamination of the sediments as calibrated to the SQGs. Therefore, we conclude that the sediment guidelines can be used to reliably estimate the probability of acute toxicity in laboratory bioassays.

Acknowledgement—Data from Pearl Harbor were kindly provided by Peter Nakamura (Pacific Division, U.S. Naval Engineering Command, PACNAVFACENGCOM, Pearl Harbor, HI, USA), with assistance from William Lester and John Clayton (Ogden Environmental, San Diego, CA, USA).

REFERENCES

1. Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manag* 19: 81–97.
2. MacDonald DD, Carr RS, Calder FD, Long ER, Ingersoll CG. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5:253–278.
3. Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environ Toxicol Chem* 17:714–727.
4. Long ER, MacDonald DD. 1998. Recommended uses of empirically derived, sediment quality guidelines for marine and estuarine ecosystem. *J Hum Ecol Risk Assess* 4:1019–1039.
5. Ingersoll CG, et al. 1997. Workgroup summary report on uncertainty evaluation of measurement endpoints used in sediment ecological risk assessments. In Ingersoll CG, Dillon T, Biddinger GR, eds, *Ecological Risk Assessment of Contaminated Sediments.* SETAC Special Publication. Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA, pp 271–296.
6. Long ER, et al. 1999. Magnitude and extent of chemical contamination and toxicity in sediments of Biscayne Bay and vicinity. NOAA/NOS CCMA 141. Technical Memorandum. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
7. Thursby GB, Heltshe J, Scott KJ. 1997. Revised approach to toxicity test acceptability criteria using a statistical performance assessment. *Environ Toxicol Chem* 16:1322–1329.
8. U.S. Environmental Protection Agency, Army Corps of Engineers. 1991. Evaluation of dredged material proposed for ocean disposal. EPA 503/8-91-001. U.S. Testing Manual. Washington, DC.
9. Sindermann CJ. 1997. The search for cause and effect relationships in marine pollution studies. *Mar Pollut Bull* 34:218–221.

SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 20, No. 10, pp. 2276–2286, 2001
Printed in the USA
0730-7268/01 $9.00 + .00

AN EVALUATION OF METHODS FOR CALCULATING MEAN SEDIMENT QUALITY GUIDELINE QUOTIENTS AS INDICATORS OF CONTAMINATION AND ACUTE TOXICITY TO AMPHIPODS BY CHEMICAL MIXTURES

RUSSELL FAIREY,*† EDWARD R. LONG,‡ CASSANDRA A. ROBERTS,† BRIAN S. ANDERSON,§ BRYN M. PHILLIPS,§ JOHN W. HUNT,§ HOWARD R. PUCKETT,‖ and CRAIG J. WILSON#

†Moss Landing Marine Laboratories, San Jose State University Foundation, 7544 Sandholdt Road, Moss Landing, California 95039, USA
‡NOAA, NCCOS/Center for Coastal Monitoring and Assessment, 7600 Sand Point Way NE, Seattle, Washington 98115, USA
§University of California-Davis, Marine Pollution Studies Laboratory, 34500 Highway 1, Monterey, California 93940, USA
‖California Department of Fish and Game, Marine Pollution Studies Laboratory, 34500 Highway 1, Monterey, California 93940, USA
#California State Water Resources Control Board, Bays and Estuaries Unit, PO Box 944213, Sacramento, California 92444-2130, USA

(*Received* 5 *July* 2000; *Accepted* 18 *February* 2001)

Abstract—Mean sediment quality guideline quotients (mean SQGQs) were developed to represent the presence of chemical mixtures in sediments and are derived by normalizing a suite of chemicals to their respective numerical sediment quality guidelines (SQGs). Mean SQGQs incorporate the number of SQGs exceeded and the degree to which they are exceeded and are used for comparison with observed biological effects in the laboratory or field. The current research makes it clear, however, that the number and type of SQGs used in the derivation of these mean quotients can influence the ability of mean SQGQ values to correctly predict acute toxicity to marine amphipods in laboratory toxicity tests. To determine the optimal predictive ability of mean SQGQs, a total of 18 different chemical combinations were developed and compared. The ability of each set of mean SQGQs to correctly predict the presence and absence of acute toxicity to amphipods was determined using three independent databases (n = 605, 2753, 226). Calculated mean SQGQ values for all chemical combinations ranged from 0.002 to 100. The mean SQGQ that was most predictive of acute toxicity to amphipods is calculated as SQGQ1 = ((Σ ([cadmium]/4.21)([copper]/270)([lead]/112.18)([silver]/1.77)([zinc]/410)([total chlordane]/6)([dieldrin]/8)([total PAH_{OC}]/1,800)([total PCB]/400))/9). Both the incidence and magnitude of acute toxicity to amphipods increased with increasing SQGQ1 values. To provide better comparability between regions and national surveys, SQGQ1 is recommended to serve as the standard method for combination of chemicals and respective SQGs when calculating mean SQGQs.

Keywords—Sediment quality guideline quotients Amphipod Toxicity Chemical mixtures

INTRODUCTION

Although the use of empirically derived sediment quality guidelines (SQGs) in sediment monitoring and assessment has been the subject of debate, recent reports suggest SQGs continue to be widely used to predict when chemical concentrations are likely to be associated with a measurable biological response [1–5]. Use of SQGs has been encouraged by recent research that indicates reasonable predictive ability of SQGs [6], in combination with limited application or regional specificity of chemical guidelines derived through other methods [7–11]. The increasing use of SQGs results from a practical need for protective management tools when identifying areas where anthropogenic chemicals may present a risk to benthic biota. Unfortunately, the inappropriate application of SQGs has resulted in criticism of their use in sediment management. There is a need to continually reexamine the appropriate use of SQGs as management tools and to refine uses of SQGs to better predict toxicity and/or biological community impairment.

Empirically derived SQGs have been generated from large sets of synoptically collected chemical and biological data [6]. Sediment quality guidelines were developed primarily from field-collected sediments using statistical approaches that associate chemical concentration and biological response. They were established to demonstrate the individual chemical concentrations at which biological effects were expected to be present or absent [12]. Chemical mixtures can also be represented using SQGs and are generally calculated by normalizing each chemical found in the sediment to its respective SQG value and then averaging the resulting normalized values for a given suite of chemicals. These normalized chemical summaries (hereafter referred to as mean sediment quality guideline quotients, or mean SQGQs) represent complex chemical mixtures within each unique sediment sample as a quantitative numeric value that incorporates both the magnitude and number of SQGs exceeded. Mean SQGQs can be used to predict the probability of biological effects in the laboratory or field as previously demonstrated [12–14]. In this article, we report and compare a variety of methods for calculating mean SQGQs with the objective of proposing a standardized method for SQGQ calculation that improves the ability of SQGQs to predict whole sediment acute effects on amphipods.

METHODS

This research evaluated the type and number of analytes in the SQGQ calculation to find chemical combinations that best predicted biological effects, as indicated by marine amphipod mortality in sediment toxicity tests. Mean SQGQS were calculated using effects range-median (ERM) [15,16] and probable effects level (PEL) [17] SQG values for normalizing sediment chemical concentrations. The ERM and PEL sediment quality guidelines have been published for 10 individual trace metals, three individual pesticides, 13 individual polycyclic aromatic hydrocarbons (PAHs), three groupings of individual PAHs,

* To whom correspondence may be addressed (fairey@mlml.calstate.edu).

Table 1. Mean sediment quality guideline quotients (SQGQ) chemical combinations using effects range median (ERM) sediment quality guideline (SQG) values

Mean SQGQs using ERM guidelines[a]					
ERMQ1	ERMQ2	ERMQ3	ERMQ4	ERMQ5	ERMQ6
Antimony	Arsenic	Cadmium	Antimony	Arsenic	Arsenic
Arsenic	Cadmium	Chromium	Arsenic	Cadmium	Cadmium
Cadmium	Chromium	Copper	Cadmium	Chromium	Chromium
Chromium	Copper	Mercury	Chromium	Copper	Copper
Copper	Lead	Zinc	Copper	Lead	Lead
Lead	Mercury	Total DDT	Lead	Mercury	Mercury
Mercury	Silver	Dieldrin	Silver	Silver	Silver
Silver	Zinc	Total PCBs	Zinc	Zinc	Zinc
Zinc	2-Methylnaphthalene	Low mol wt PAHs	Total DDT	Total chlordane	Total chlordane
Total DDT (OC)	Dibenz[*a*,*h*]anthracene	High mol wt PAHs	Total chlordane	Dieldrin	Dieldrin
Total chlordane	Acenaphthene		Dieldrin	Total PCBs	Total PCBs
Dieldrin	Acenaphthylene		Total PCBs	Low mol wt PAHs	Low mol wt PAHs
Endrin	Anthracene		Low mol wt PAHs	High mol wt PAHs	Total PAHs
Total PCBs	Benz[*a*]anthracene		Total PAHs		
Low mol wt PAHs	Benzo[*a*]pyrene				
High mol wt PAHs	Chrysene				
	Fluorene				
	Fluoranthene				
	Naphthalene				
	Phenanthrene				
	Pyrene				
	p′p′-DDE				
	Total DDT				
	Total PCBs				

[a] ERMQ chemical combinations used effects range median SQGs [15,16] as critical values for mean quotient calculation except ERMQ1, which used an organic carbon normalized SQG [33] value for total DDT. OC = organic carbon; PCBs = polychlorinated biphenyls; PAHs = polycyclic aromatic hydrocarbons; mol wt = molecular weight.

two summed organochlorine pesticides, and the sum of 18 polychlorinated biphenyl congeners (PCBs) [15–17]. Guideline values for these 32 chemicals were selected because of published studies demonstrating the reliability and predictive ability of the individual guideline values [6,16]. Two different chemical combinations that were used in previous studies were compared initially and evaluated for their ability to correctly classify sediments as toxic or nontoxic. Based on the results of this comparison, 10 additional chemical combinations were sequentially tested using various combinations of the same chemicals and their respective SQGs. Each successive combination was an attempt by the investigators to improve mean SQGQ predictive ability based on the results of the previous combination. Six of the 12 calculation methods were based on various combinations of chemicals that were divided by their respective ERM guidelines (Table 1). Six PELQ calculation methods were based on various combinations of chemicals that were divided by their respective PEL guidelines (Table 2).

In an attempt to evaluate mean SQGQs that incorporated additional types of SQGs, an assessment of the predictive ability of individual SQGs was performed. The assessment included individual SQGs developed using correlative approaches, consensus approaches, and theoretical approaches. Individual SQGs were selected for inclusion in mean SQGQ quotient calculation methods when the values best predicted acute toxicity to amphipods for that chemical or chemical class and were representative of chemical concentrations most commonly found in field samples from the database (italicized entries in Table 3). A total of 15 individual SQGs were selected for use in additional chemical combinations. These SQGs were incorporated into six SQGQ calculation methods that were based on combinations of chemicals divided by a mixture of respective ERM, PEL, or other selected individual SQGs (Table 4).

Each mean quotient calculation method included chemicals and their respective SQGs for each of four major chemical groups (metals, pesticides, PAHs, PCBs). A total of 18 chemical combinations were evaluated.

All mean quotient calculations attempted to use the same summation methods for chemical classes (total chlordane, total DDT, total PCBs, low molecular weight PAHs, high molecular weight PAHs, total PAHs). This step was necessary for valid comparison of mean quotient calculations but presented limitations with large multistudy data sets because of differences in analyte lists among studies. The methods for summation of chemical classes used in this study are given in the Appendix [15,18–20]. Individual chemical quotients are calculated by dividing the measured concentration or class summation concentration in the sediment sample by the respective published sediment quality guideline value for each chemical for which individual SQGs were derived. A value greater than one indicates the chemical concentration in that sample exceeded its respective SQG. The mean SQGQ is obtained by calculating the mean of the resulting individual quotients for a given combination of chemicals. A generalized example of the calculation is

$$\text{mean sediment quality guideline quotient} = \left(\sum_n ([\text{arsenic/SQG}][\text{chromium/SQG}][\text{copper/SQG}] \ldots \times [\text{total chlordane/SQG}][\text{total PCB/SQG}])\right) \Big/ n$$

where n = total number of analytes.

Samples that were found to have chemical concentrations less than the method detection limit were adjusted, for purposes

Table 2. Mean sediment quality guideline quotient (SQGQ) chemical combinations using probable effects level (PEL) SQG values

Mean SQGQs using PEL guidelines[a]					
PELQ1	PELQ2	PELQ3	PELQ4	PELQ5	PELQ6
Arsenic	Cadmium	Cadmium	Arsenic	Arsenic	Arsenic
Cadmium	Chromium	Chromium	Cadmium	Cadmium	Cadmium
Chromium	Copper	Copper	Chromium	Chromium	Chromium
Copper	Lead	Mercury	Copper	Copper	Copper
Lead	Mercury	Zinc	Lead	Lead	Lead
Mercury	Nickel	Total DDT	Silver	Mercury	Mercury
Silver	Silver	Dieldrin	Zinc	Silver	Silver
Zinc	Zinc	Total PCBs	Total DDT	Zinc	Zinc
Total DDT (OC)	2-Methylnaphthalene	Low mol wt PAHs	Total chlordane	Total chlordane	Total chlordane
Total chlordane	Dibenz[*a*,*h*]anthracene	High mol wt PAHs	Dieldrin	Dieldrin	Dieldrin
Dieldrin	Acenaphthene		Total PCBs	Lindane	Lindane
Lindane	Acenaphthylene		Low mol wt PAHs	Total PCBs	Total PCBs
Total PCBs	Anthracene		Total PAHs	Low mol wt PAHs	Low mol wt PAHs
Low mol wt PAHs	Benz[*a*]anthracene			High mol wt PAHs	Total PAHs
High mol wt PAHs	Benzo[*a*]pyrene				
	Chrysene				
	Fluorene				
	Fluoranthene				
	Naphthalene				
	Phenanthrene				
	Pyrene				
	p′p′-DDE				
	Total DDT				
	Total PCBs				

[a] PELQ chemical combinations used probable effects level (PEL) SQGs [17] as critical values for mean quotient calculations except for PELQ1, which used an organic carbon normalized SQG [33] value for total DDT. OC = organic carbon; PCBs = polychlorinated biphenyls; PAHs = polycyclic aromatic hydrocarbons; mol wt = molecular weight.

of the summation, to a value of one half of the given method detection limits. When one or more required chemicals were not analyzed for a particular sample, that sample was removed from the base data set and not included in evaluations of mean quotient calculation methods requiring that chemical. This step was necessary for valid comparison of the summation methods and the data sets to which they could be applied.

This study used chemical and biological data from three readily available sources as the basis for comparisons of predictive ability. The sources were the State of California's Bay Protection and Toxic Cleanup Program database (BPTCP), the National Oceanic and Atmospheric Administration's (NOAA's) national sediment toxicity (SEDTOX) database, and NOAA's Biscayne Bay survey in Florida. The BPTCP database included 605 coastal sediment samples for which synoptic chemical analyses and toxicity tests (*Eohaustorius estuarius* and *Rhepoxynius abronius*) were performed [21–27]. These data were used for initial comparisons between calculation methods developed by Fairey et al. [21] and Long et al. [6] and 10 additional chemical mixture combinations using ERM and PEL guidelines for mean quotient calculation. These data also were used to evaluate the predictive ability of a variety of individual SQGs with the objective of evaluating the predictive ability of six additional chemical mixture combinations that combined SQGs from different studies. The SEDTOX database included 2,753 sediment samples from studies throughout the coastal United States for which synoptic chemical analyses and whole sediment acute toxicity tests (*Ampelisca abdita, R. abronius*) were paired [28]. The SEDTOX data was used to confirm predictive accuracy of mean quotient calculation methods using BPTCP data and to further evaluate mean quotient calculation methods that included SQGs from mixed sources. The Biscayne Bay database [29] included 226 samples analyzed for sediment chemistry and toxicity (*A. abdita)* and was used as a third independent database for testing predictive accuracy.

Chemical data were combined in dBase IV database files and manipulated using dBase command programs. Mean SQGQ values derived from the various unique chemical combinations in each sample then were compared with synoptically measured toxicity test responses to assess how well each chemical combination predicted acute toxicity. Samples were classified as toxic for the BPTCP study when sample survival was significantly different from controls and less than a critical value determined using the minimum significant difference 90th percentile that was generated from BPTCP data (*Eohaustorius estuarius* survival in sample was less than 75% of survival in negative controls, *R. abronius* survival in sample was less than 77% of survival in negative controls) [21–27,30]. Samples were classified as toxic for the remaining studies when samples were significantly different from controls and less than 80% of negative controls [31].

MacDonald et al. [32] compared mean quotient values at five critical levels to assess associations with measured biological effects (<0.1, <0.5, >0.5, >1.0, and >1.5). Long and MacDonald [12] used similar levels but also included an additional level (>2.3) to better evaluate greater mean SQGQ values. Seven critical levels (Tables 5 & 6) were used in the current study, with one additional critical level (<0.2) included to better evaluate mean SQGQ values in the lower ranges, where the majority of samples fell.

In these analyses, we compared both the incidence (percentages) of acute toxicity and average amphipod survival above and below mean SQGQ critical values. The criteria for identification of optimal predictive ability were that the incidence of toxicity increased steadily and average survival decreased steadily as mean SQGQs increased; incidence of toxicity was low ($<5\%$) and average survival high ($>80\%$) when

Table 3. Sample counts and the incidence of acute toxicity to amphipods when effects range medians (ERMs), probable effects levels (PELs), or other individual sediment quality guidelines (SQGs) were exceeded using the Bay Protection and Toxic Cleanup Program (BPTCP) data (n = 605)

Chemical name	ERM[b]	No. samples > ERM	% Toxic	PEL[c]	No. samples > PEL	% Toxic	Other SQGs	No. samples > SQG	% Toxic
Antimony	*25 μg/g*	3	67	NA	NA	NA	200 μg/g [11]	0	NA
Arsenic	*70 μg/g*	2	100	41.6 μg/g	6	67	700 μg/g [11]	1	100
Cadmium	9.6 μg/g	3	100	*4.21 μg/g*	11	100	NA	NA	NA
Chromium	*370 μg/g*	7	71	160.4 μg/g	87	36	270 μg/g [11]	22	50
Copper	*270 μg/g*	69	70	108.2 μg/g	206	56	1,300 μg/g [11]	3	100
Lead	218 μg/g	15	93	*112.18 μg/g*	64	75	660 μg/g [11]	3	100
Mercury	0.7 μg/g	110	55	0.696 μg/g	110	55	*2.1 μg/g* [11]	19	68
Nickel	51.6 μg/g	76	41	42.8 μg/g	108	44	NA	NA	NA
Silver	3.7 μg/g	6	100	*1.77μg/g*	44	75	6.1 μg/g [11]	1	100
Zinc	*410 μg/g*	76	78	271 μg/g	141	65	1,600 μg/g [11]	5	100
Total chlordane	*6 ng/g* [15]	185	54	4.79 ng/g	210	52	NA	NA	NA
Total DDT	46.1 ng/g	214	49	51.7 ng/g	203	50	100 μg/g OC [33]	9	33
Dieldrin	*8 ng/g*	30	90	4.3 ng/g	53	79	20 μg/g OC [9]	0	NA
Endrin	45 ng/g	0	NA	NA	NA	NA	*0.76 μg/g OC* [9]	2	100
Lindane	NA	NA	NA	0.99 ng/g	9	67	*0.37 μg/g OC* [9]	2	100
Total PCB	180 ng/g	164	53	188.79 ng/g	157	54	*400 ng/g* [35]	80	65
Acenaphthene	500 ng/g	3	33	88.9 ng/g	42	53	230 μg/g OC [9]	0	NA
Acenaphthylene	640 ng/g	0	NA	127.89 ng/g	31	58	NA	NA	NA
Anthracene	1,100 ng/g	27	41	245 ng/g	107	51	NA	NA	NA
Fluorene	540 ng/g	10	50	144.35 ng/g	48	48	NA	NA	NA
2-Methylnaphthalene	670 ng/g	3	67	201.28 ng/g	19	63	NA	NA	NA
Naphthalene	2,100 ng/g	1	100	390.64 ng/g	9	44	NA	NA	NA
Phenanthrene	1,500 ng/g	27	55	543.53 ng/g	86	57	240 μg/g OC [9]	8	50
Low mol wt PAHs	3,160 ng/g	25	44	1,442 ng/g	72	51	24,000 ng/g [11]	1	0
Benz[*a*]anthracene	1,600 ng/g	32	43	692.53 ng/g	100	52	NA	NA	NA
Benzo[*a*]pyrene	1,600 ng/g	59	53	763.22 ng/g	115	54	NA	NA	NA
Chrysene	2,800 ng/g	31	45	845.98 ng/g	103	52	NA	NA	NA
Dibenz[*a*,*h*]anthracene	260 ng/g	73	62	134.61 ng/g	115	54	NA	NA	NA
Fluoranthene	5,100 ng/g	18	50	1,493.54 ng/g	84	48	300 μg/g OC [9]	13	38
Pyrene	2,600 ng/g	44	48	1,397.6 ng/g	111	54	NA	NA	NA
High mol wt PAHs	9,600 ng/g	71	51	6,676.14 ng/g	100	51	69,000 ng/g [11]	1	100
Total PAHs	44,792 ng/g	5	60	16,770.54 ng/g	45	47	*1,800 μg/g OC* [20]	10	60

[a] The number of samples exceeding individual SQGs and the percentage of those samples determined to be toxic were used in selecting which SQGs were reasonably predictive of toxicity to amphipods and were representative of concentrations commonly found in BPTCP samples. Selected SQGs (italics) were used in mean calculation methods that incorporated individual SQGs from a variety of sources. PCB = polychlorinated biphenyl; PAH = polycyclic aromatic hydrocarbon; OC = organic carbon; NA = not applicable; mol wt = molecular weight.
[b] [16].
[c] [17].

Table 4. Mean sediment quality guideline quotient (SQGQ) chemical combinations using effects range median (ERM) probable effects level (PEL) and other individual SQGs values

Mean SQGQs using combinations of individual SQGs[a]					
SQGQ1	SQGQ2	SQGQ3	SQGQ4	SQGQ5	SQGQ6
Cadmium	Cadmium	Cadmium	Cadmium	Arsenic	Antimony
Copper	Chromium	Chromium	Chromium	Cadmium	Arsenic
Lead	Copper	Copper	Copper	Copper	Cadmium
Silver	Lead	Lead	Lead	Lead	Chromium
Zinc	Silver	Silver	Silver	Silver	Copper
Total chlordane	Zinc	Zinc	Zinc	Zinc	Lead
Dieldrin	Total chlordane	Lindane	Total chlordane	Total chlordane	Mercury
Total PCBs	Dieldrin	Dieldrin	Dieldrin	Dieldrin	Silver
Total PAHs	Endrin	Endrin	Total PCBs	Lindane	Zinc
	Total PCBs	Total PCBs	Total PAHs	Total PCBs	Total chlordane
	Total PAHs	Total PAHs		Total PAHs	Dieldrin
					Endrin
					Lindane
					Total PCBs
					Total PAHs

[a] SQGQ chemical combinations used ERM, PEL, and other individual SQGs as critical values for mean quotient calculation. PCB = polychlorinated biphenyl; PAHs = polycyclic aromatic hydrocarbons.

Table 5. Comparison of mean sediment quality guideline quotient chemical combinations using the Bay Protection and Toxic Cleanup Program (BPTCP) data (n = 605[a]); ERM = effects range median; PEL = probable effects level

SQGQ range[c]	ERMQ1 (n = 528)[b]	ERMQ2 (n = 395)	ERMQ3 (n = 569)	ERMQ4 (n = 533)	ERMQ5 (n = 533)	ERMQ6 (n = 531)
>2.3	n = 6, 100%	n = 9, 33%	n = 24, 50%	n = 19, 58%	n = 9, 100%	n = 8, 100%
>1.5	n = 17, 88%	n = 13, 46%	n = 46, 61%	n = 34, 65%	n = 23, 83%	n = 20, 85%
>1.0	n = 50, 78%	n = 31, 42%	n = 92, 62%	n = 52, 63%	n = 63, 78%	n = 55, 78%
>0.5	n = 122, 62%	n = 94, 50%	n = 205, 51%	n = 161, 53%	n = 147, 57%	n = 136, 60%
<0.5	n = 406, 37%	n = 301, 37%	n = 364, 38%	n = 372, 38%	n = 386, 37%	n = 395, 36%
<0.2	n = 227, 36%	n = 175, 36%	n = 149, 38%	n = 182, 37%	n = 202, 36%	n = 212, 35%
<0.1	n = 65, 31%	n = 65, 39%	n = 38, 29%	n = 38, 29%	n = 53, 23%	n = 53, 23%
SQGQ range[c]	PELQ1 (n = 528)	PELQ2 (n = 394)	PELQ3 (n = 569)	PELQ4 (n = 528)	PELQ5 (n = 533)	PELQ6 (n = 531)
>2.3	n = 14, 100%	n = 20, 35%	n = 39, 59%	n = 29, 66%	n = 16, 94%	n = 13, 100%
>1.5	n = 50, 76%	n = 45, 44%	n = 85, 64%	n = 61, 66%	n = 55, 76%	n = 48, 75%
>1.0	n = 91, 70%	n = 36, 50%	n = 142, 58%	n = 113, 63%	n = 95, 69%	n = 87, 74%
>0.5	n = 197, 53%	n = 127, 48%	n = 274, 51%	n = 241, 49%	n = 202, 53%	n = 199, 54%
<0.5	n = 331, 36%	n = 267, 36%	n = 295, 35%	n = 287, 37%	n = 331, 35%	n = 332, 35%
<0.2	n = 88, 29%	n = 118, 37%	n = 66, 35%	n = 63, 33%	n = 81, 30%	n = 82, 29%
<0.1	n = 29, 21%	n = 33, 24%	n = 7, 14%	n = 2, 0%	n = 5, 0%	n = 5, 0%
SQGQ range[c]	SQGQ1 (n = 561)	SQGQ2 (n = 561)	SQGQ3 (n = 563)	SQGQ4 (n = 561)	SQGQ5 (n = 528)	SQGQ6 (n = 528)
>2.3	n = 16, 94%	n = 14, 93%	n = 3, 100%	n = 16, 94%	n = 11, 100%	n = 5, 100%
>1.5	n = 42, 76%	n = 32, 81%	n = 7, 100%	n = 39, 82%	n = 20, 82%	n = 14, 100%
>1.0	n = 81, 70%	n = 58, 72%	n = 18, 72%	n = 71, 69%	n = 39, 77%	n = 30, 77%
>0.5	n = 160, 60%	n = 149, 62%	n = 98, 73%	n = 163, 59%	n = 123, 64%	n = 108, 71%
<0.5	n = 400, 36%	n = 411, 36%	n = 465, 36%	n = 398, 36%	n = 405, 36%	n = 420, 35%
<0.2	n = 187, 30%	n = 132, 28%	n = 178, 31%	n = 126, 28%	n = 207, 32%	n = 203, 36%
<0.1	n = 50, 32%	n = 3, 66%	n = 10, 30%	n = 17, 18%	n = 57, 30%	n = 27, 30%

[a] In the BPTCP database, 250 sediment samples were acutely toxic to amphipods (41%) while 355 were not toxic.
[b] The count reflects the number of BPTCP samples for which the particular chemical combination could be calculated.
[c] For each SQGQ range, the number of samples with mean quotient values exceeding the respective critical value are given as well as the percentage of those samples exceeding the critical value that were acutely toxic to amphipods.

mean SQGQs were low (<0.1); incidence of toxicity was high (>90%) and average survival was low (<50%) when mean SQGQs were high (>2.3); and when predictive ability was similar across SQGQ ranges for different calculation methods, sample counts were compared to find methods that correctly predicted toxicity for the largest number of samples in each data set.

RESULTS

As shown by the sample count for the BPTCP database (Table 5), the number of samples that lend themselves to a particular calculation method differs dramatically among methods (n = 394–569 out of 605). This variability is due to periodic expansions of the analyte list for the BPTCP over the program's seven years. In general, the lower the number of analytes included in the calculation method, the more likely the derivation could be completed for the majority of samples. Results demonstrate an increased number of samples being available for SQGQ calculation and biological comparison when SQG chemicals were limited.

Using the BPTCP database, toxicity to marine amphipods (*E. estuarius* and *R. abronius*) in 605 sediment samples was compared with calculated mean quotient values. Based on BPTCP comparisons with controls and the minimum significant difference, 250 samples were classified as acutely toxic (41%) and 355 were classified as nontoxic (59%). Calculated mean quotient values from sediment samples for all derivation methods ranged from 0.031 (lowest multiple chemical indicator) to 10.9 (greatest multiple chemical indicator).

In the BPTCP data, there was a wide range in the patterns in toxicity among the different calculation methods (Table 5). Generally, SQGQ values above the greater critical values were associated with greater proportions of toxic samples while those below the lower critical values were associated with reduced proportions of toxic samples. In two cases (ERMQ2, PELQ2), the incidence of toxicity did not increase appreciably with increasing chemical concentration. In four cases (ERMQ3, ERMQ4, PELQ3, PELQ4), the increase in toxicity was minimal with increasing chemical concentration. One hybrid combination of several types of sediment quality guidelines for mean quotient calculation (SQGQ6) resulted in the greatest association with the incidence of toxicity to amphipods (100% for values >2.3 and 100 % for values >1.5). In general, predictive ability improved when mean quotient calculations utilized SQGs developed using a variety of comparative statistical approaches (e.g., ERM, PEL, consensus, equilibrium partitioning). Associations with the incidence of toxicity were lowest for all methods when mean quotient values dropped below 0.1.

Using SEDTOX data, 2,753 samples were examined for associations between toxicity to marine amphipods (*E. estuarius*, *A. abdita*, and *R. abronius)* and all 18 mean quotient calculation methods. Of these samples, 484 were classified as toxic (18%) and 2,269 were classified as nontoxic (82%). The lower incidence of toxicity reflects the primary program goal of identifying general spatial status and temporal trends along the coastal margins of the United States rather than investigation of contaminated sites as with BPTCP. Calculated mean

Table 6. Comparison of mean sediment quality guideline quotient (SQGQ) calculation methods using SEDTOX database, Biscayne Bay (FL, USA) and combined data; effects range median = (ERM); probable effects level = (PEL)

SQGQ range[c]	SQGQ1 (n = 1,435)	SQGQ2 (n = 977)	SQGQ3 (n = 1,134)	SQGQ4 (n = 1,367)	SQGQ5 (n = 1,385)	SQGQ6 (n = 946)	PELQ1 (n = 1,286)	ERMQ2 (n = 1,668)	ERMQ6 (n = 1,399)
Comparison of SQGQ calculation methods to toxicity using SEDTOX data (n = 2,753[a])[b]									
>2.3	n = 19, 89%	n = 17, 53%	n = 12, 0%	n = 19, 79%	n = 15, 73%	n = 9, 11%	n = 28, 71%	n = 13, 54%	n = 19, 79%
>1.5	n = 33, 88%	n = 29, 66%	n = 24, 30%	n = 32, 81%	n = 30, 77%	n = 25, 60%	n = 58, 69%	n = 25, 64%	n = 31, 84%
>1.0	n = 70, 78%	n = 52, 70%	n = 49, 51%	n = 58, 76%	n = 58, 72%	n = 40, 60%	n = 116, 56%	n = 46, 72%	n = 61, 77%
>0.5	n = 216, 48%	n = 184, 42%	n = 183, 40%	n = 205, 45%	n = 176, 46%	n = 136, 48%	n = 241, 40%	n = 127, 52%	n = 168, 47%
<0.5	n = 1,218, 8%	n = 792, 6%	n = 949, 9%	n = 1,161, 6%	n = 1,206, 8%	n = 809, 7%	n = 1,045, 5%	n = 1,541, 8%	n = 1,231, 7%
<0.2	n = 1,022, 5%	n = 636, 3%	n = 766, 5%	n = 955, 4%	n = 1,007, 4%	n = 666, 3%	n = 822, 3%	n = 1,378, 6%	n = 1,049, 4%
<0.1	n = 768, 4%	n = 459, 2%	n = 515, 3%	n = 656, 3%	n = 746, 3%	n = 487, 2%	n = 537, 3%	n = 1,196, 5%	n = 806, 4%
Comparison of SQGQ calculation methods to toxicity using Biscayne Bay data (n = 226[d])									
SQGQ range	SQGQ1 (n = 226)	SQGQ2 (n = 226)	SQGQ3 (n = 226)	SQGQ4 (n = 226)	SQGQ5 (n = 226)	SQGQ6 (n = 219)	PELQ1 (n = 226)	ERMQ2 (n = 226)	ERMQ6 (n = 226)
>2.3	n = 6, 100%	n = 4, 100%	n = 0	n = 6, 100%	n = 4, 100%	n = 3, 100%	n = 5, 100%	n = 0	n = 4, 100%
>1.5	n = 9, 100%	n = 7, 100%	n = 3, 100%	n = 7, 100%	n = 7, 100%	n = 6, 100%	n = 12, 83%	n = 3, 67%	n = 9, 89%
>1.0	n = 18, 83%	n = 15, 87%	n = 6, 100%	n = 16, 81%	n = 14, 86%	n = 9, 100%	n = 20, 75%	n = 6, 67%	n = 14, 86%
>0.5	n = 24, 71%	n = 23, 74%	n = 18, 78%	n = 23, 74%	n = 23, 74%	n = 19, 79%	n = 26, 65%	n = 17, 77%	n = 24, 71%
<0.5	n = 202, 4%	n = 203, 4%	n = 208, 5%	n = 203, 4%	n = 203, 4%	n = 200, 5%	n = 200, 4%	n = 208, 6%	n = 202, 4%
<0.2	n = 187, 4%	n = 188, 4%	n = 193, 4%	n = 187, 4%	n = 188, 4%	n = 188, 4%	n = 180, 4%	n = 198, 5%	n = 189, 4%
<0.1	n = 156, 4%	n = 163, 4%	n = 167, 4%	n = 158, 4%	n = 162, 4%	n = 166, 4%	n = 131, 5%	n = 180, 4%	n = 162, 4%
Comparison of SQGQ calculation methods to toxicity using minimized common data from 3,584 samples (n = 1,692[e])									
SQGQ range[c]	SQGQ1 (n = 1,692)	SQGQ2 (n = 1,692)	SQGQ3 (n = 1,692)	SQGQ4 (n = 1,692)	SQGQ5 (n = 1,692)	SQGQ6 (n = 1,692)	PELQ1 (n = 1,692)	ERMQ2 (n = 1,692)	ERMQ6 (n = 1,692)
>2.3	n = 32, 100%	n = 32, 75%	n = 15, 20%	n = 34, 94%	n = 26, 92%	n = 17, 53%	n = 45, 87%	n = 18, 56%	n = 28, 89%
>1.5	n = 62, 87%	n = 58, 76%	n = 33, 48%	n = 59, 88%	n = 50, 84%	n = 45, 78%	n = 112, 76%	n = 37, 65%	n = 55, 85%
>1.0	n = 127, 80%	n = 107, 76%	n = 65, 63%	n = 115, 78%	n = 98, 78%	n = 79, 71%	n = 214, 65%	n = 75, 65%	n = 118, 81%
>0.5	n = 315, 56%	n = 316, 54%	n = 263, 56%	n = 321, 56%	n = 282, 57%	n = 263, 60%	n = 416, 50%	n = 210, 57%	n = 290, 57%
<0.5	n = 1,377, 13%	n = 1,376, 13%	n = 1,429, 10%	n = 1,371, 13%	n = 1,410, 12%	n = 1,429, 11%	n = 1,276, 13%	n = 1,481, 17%	n = 1,402, 15%
<0.2	n = 1,039, 8%	n = 946, 7%	n = 977, 8%	n = 960, 6%	n = 1,073, 9%	n = 1,057, 9%	n = 859, 5%	n = 1,245, 14%	n = 1,091, 10%
<0.1	n = 724, 4%	n = 619, 3%	n = 604, 3%	n = 655, 3%	n = 740, 4%	n = 680, 4%	n = 540, 3%	n = 1,009, 11%	n = 766, 4%

[a] In the SEDTOX database, 484 sediment samples were acutely toxic to amphipods (18%) while 2,269 were not toxic.
[b] The count (n values in the column heads) reflects the number of samples in each database for which the particular chemical combination could be calculated.
[c] For each SQGQ range, the number of samples with mean quotient values exceeding the respective critical value are given as well as the percentage of those samples exceeding the critical value that were acutely toxic to amphipods.
[d] In the Biscayne Bay database, 25 sediment samples were acutely toxic to amphipods (11%) while 201 were not toxic.
[e] In the minimized common database, 371 sediment samples were acutely toxic to amphipods (22%) while 1,321 were not toxic.

quotient values for all 18 methods ranged from 0.01 to 100. The unique combination most predictive of toxicity to amphipods was SQGQ1. This combination of mixed sediment quality guidelines for mean quotient calculation resulted in the greatest association with the incidence of toxicity to amphipods (89% for values >2.3 and 88% for values >1.5; Table 6). Other derivation methods yielded reduced associations (30%–81%) with observed toxicity for values greater than 1.5 (Table 6). The incidence of toxicity invariably was less than 10% in samples with lesser mean quotient values (<0.5).

As with the BPTCP database, there was a significant difference in the total number of SEDTOX samples for which mean quotient values could be calculated using different chemical combinations. Totals ranged between 946 and 1,668 and represented 35 to 60% of the available data. The method that used the greatest number of chemicals (24 for ERMQ2) in the mean quotient calculations yielded the largest number of calculated SQGQ values (Table 6) but also demonstrated some of the lowest incidences of associated toxicity. Including the large number of individual PAHS in the ERMQ2 calculation was conducive to use of the data set but did not improve associations with acute amphipod toxicity. The SQGQ1, which incorporated only nine chemicals (no individual PAHs) and most accurately predicted toxicity, could be calculated for 1,435 samples, representing slightly more than one half of the available data.

Most mean quotient calculation methods provided good predictive ability in the Biscayne Bay data (Table 6). Both SQGQ6 and SQGQ1 predicted equally well based on the number and percentage of samples that were toxic (100%) when mean quotient values were greater than 1.0 and 1.5, respectively. The incidence of toxicity was greater than 70% for SQGQ6 and SQGQ1 when values were greater than 0.5, while values less than 0.5 were toxic less than 5% of the time.

To demonstrate general relationships between acute toxicity to amphipods and SQGQ values, data from all three sources were combined into a single data set. Reduction of the combined data set of 3,584 samples to only those samples that could be used in all 18 quotient calculation methods resulted in a common data set of 1,692 samples. Calculated mean quotient values for all 18 methods ranged from 0.002 to 100. The SQGQ1 resulted in the greatest association with the incidence of toxicity to amphipods (100% for values >2.3 and 87% for values >1.5). A low incidence of toxicity (4%) was observed when SQGQ1 values were less than 0.1 (Table 6). Table 7 identifies the percentages of samples that were acutely toxic for 12 ranges of SQGQ1 values. Also expressed is the magnitude of the toxic response (as mean % survival) over the same SQGQ1 ranges. The SQGQ1 values were highly correlated with the incidence ($r^2 = 0.901$, $p < 0.001$) and magnitude ($r^2 = 0.913$, $p < 0.001$) of acute toxicity (Fig. 1).

Table 7. Acute amphipod toxicity associated with SQGQ1 ranges for common samples from combined NOAA/BPTCP/Biscayne Bay database (n = 1,692)[a]

SQGQ1 range	Range average	No. of samples	% of Samples toxic	Average % survival	Average no. ERMs exceeded
0–0.1	0.04	724	4	95	0.03
0.1–0.25	0.16	389	19	83	0.3
0.25–0.5	0.35	262	33	76	1.6
0.5–0.75	0.62	113	32	76	3.3
0.75–1.0	0.87	77	52	70	5.4
1.0–1.25	1.1	44	68	53	6.3
1.25–1.5	1.36	21	81	56	6.7
1.5–2	1.62	21	67	45	7.6
2–2.5	2.2	11	91	36	9.5
2.5–3.0	2.72	10	100	35	9.4
3.0–3.5	3.23	10	100	20	10.2
>3.5	4.63	10	100	13	12.2

[a] NOAA = National Oceanic and Atmospheric Administration; BPTCP = Bay Protection and Toxic Cleanup Program; SQGQ = sediment quality guideline quotient; ERM = effects range medians.

DISCUSSION

Mean SQGQs have been used previously to compare toxicity test response and observed benthic community response to concentrations of a mixture of 16 chemicals in coastal sediments from California, USA [21]. Similarly, Long et al. [6] used mean SQGQs to compare observed toxicity test response with a mixture of 24 chemicals in test sediments from a national database (n = 1,068). Both these studies found a pattern of increasing incidence of toxicity in sediments with increasing mean SQGQ values and identified critical levels above which biological effects could be predicted. Each of these studies used effects range median values to derive mean SQGQs, but chemicals used in the quotient calculation methods were not identical. Long et al. [6] used 13 individual SQGs to represent the individual polycyclic aromatic hydrocarbon compounds in the average, while Fairey et al. [21] represented the PAHs by using two individual SQGs, one each for low and high molecular weight PAH classes. Another difference is that Long et al. [6] represented the organochlorine pesticides using ERMs for *p′,p′*-DDE and total DDT, while Fairey et al. [21] represented organochlorine pesticides using ERMs for total chlordane, dieldrin, endrin, and an organic carbon normalized value for DDT [33]. These seemingly minor differences had the potential of yielding dramatically different results. For example, there is poor correlation (Pearson's correlation, $r^2 = 0.119$) between mean SQGQ values calculated by these two mean SQGQ calculation methods in the present paper (ERMQ1 [21] and ERMQ2 [6]; Table 2) using chemical data from California's Bay Protection and Toxic Cleanup Program (BPTCP; n = 605). Investigation of outliers that grouped far from predicted values indicated that chemical mixtures were dominated by DDT or PAHs in those particular California samples. As demonstrated in Table 3, the predictive abilities of ERM SQGs for DDT and most individual PAHs were low using BPTCP data. Inclusion of these individual SQGs in the mean SQGQ calculation method reduced the overall predictive ability of the mean ERMQ2 values. Using ERMQ1, 88% of the most contaminated samples (ERMQ1 > 1.5) were observed to be toxic to amphipods (*R. abronius* and *E. estuarius*), while the incidence of associated toxicity drops to 46% for ERMQ2 values >1.5 (Table 5). Also of note is that the mean quotient calculation method requiring the greatest number of chemicals (ERMQ2) resulted in the fewest samples available for mean quotient calculation. Missing chemical analytes over multiple years of the program limited the utility of the ERMQ2 values as indicators of chemical mixture concentration in this data set. These results demonstrate how selection of individual SGQs to be included in the mean SQGQ, in this case particularly with DDT and PAHs, can influence the outcome of any subsequent analysis.

In the data analysis reported here, the ability of mean SQGQs to accurately predict the probability of observing an associated biological effect is dependent on the combination

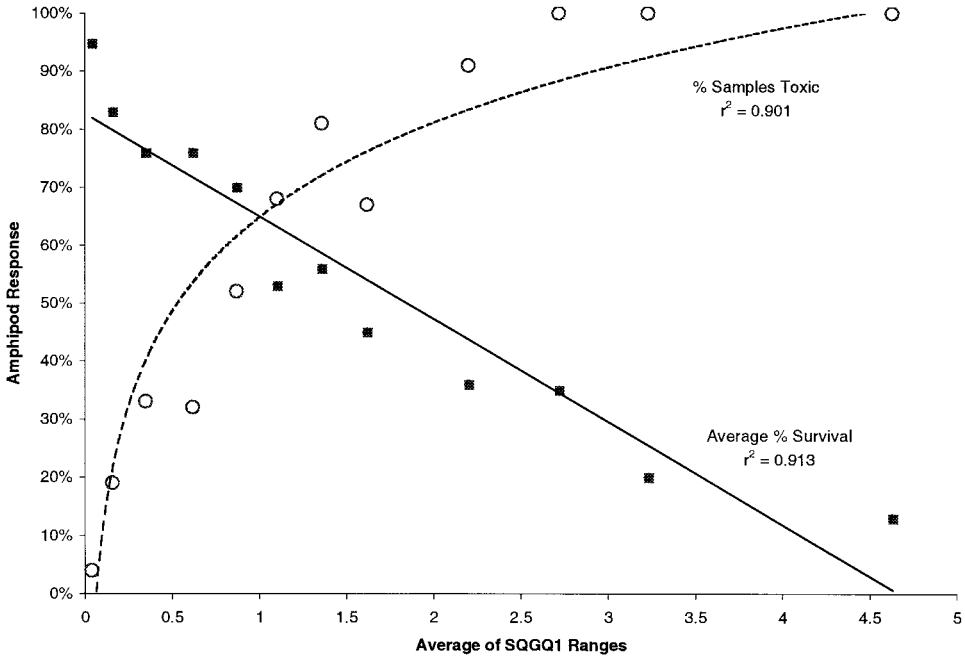


Fig. 1. Amphipod response over the range of average sediment quality guideline quotients (SQGQ1) values. Values were calculated from 1,692 samples for which all 18 chemical combinations could be calculated from the combined National Oceanic and Atmospheric Administration/Bay Protection and Toxic Cleanup Program (NOAA/BPTCP), Biscayne Bay databases. Logarithmic fit equation for percent of samples toxic versus SQGQ is $y = 0.2352 \ln(x) + 0.6487$ ($r^2 = 0.901$, $p < 0.001$). Linear fit equation for average % survival versus SQGQ is $y = -0.1774(x) + 0.8277$ ($r^2 = 0.913$, $p < 0.001$).

of chemicals and SQGs used in the calculation. This relationship becomes clear from the wide range of predictive success observed among derivation methods presented here. It is also clear that use of different data to assess mean SQGQ calculation influences the estimates of predictive ability for sediment toxicity. This demonstrates that calculation methods for mean SQGQs must be refined and standardized using a method for which predictive ability and interstudy comparability are not compromised. This point is critical to our future use of mean SQGQs and will directly impact their utility as assessment tools.

To refine calculation of mean SQGQs, a major assumption must be made that chemical analytes used in the mean SQGQ calculation are indeed representative of, or surrogates of, the toxicologically significant chemical mixture in the samples regardless of which chemicals were quantified in the analyses. This is a simplistic approach because of the seemingly infinite number of chemical combinations present in field-collected sediments. In addition, each research or monitoring program that generates sediment chemistry data has its own objectives and resultant analyte lists, based on economic and regional environmental considerations. Use of chemicals that occur most commonly across many programs will maximize data use and allow development of mean SQGQs more applicable to a wide range of environmental conditions and objectives. The choice of toxicologically representative chemicals is limited, however, by current analytical methodologies and our knowledge of those chemicals in the literature for which sediment quality guidelines have been established. This operational limitation mandates the assumption that those few published SQG chemical analytes are representative of anthropogenic contamination in the sample or that they serve well as covarying surrogates. It is constructive, though, because, as demonstrated, limiting the number of chemicals in the calculation allows more data to be included in correlative analyses. The observation that only 4% of samples were toxic when toxicity was not predicted (i.e., mean SQGQ1 < 0.1; Table 6) suggests that chemicals for which analyses were not performed, or for which there are no individual SQGs, infrequently occur at toxicologically significant concentrations in otherwise uncontaminated samples.

An implication of the mean SQGQ approach is that toxicological mechanism(s) of the representative chemicals are additive. Independent experimental evidence has demonstrated that acute toxicity to amphipods by some substances, such as PAHs, are similar and additive when known combinations of chemicals are spiked into clean sediment [34]. The research presented here does not provide direct experimental evidence of additivity. The results are, however, consistent with the hypothesis because sediments are toxic more frequently when chemical concentrations simultaneously exceed increasing numbers of individual SQGs. Table 7 demonstrates that, as increasing numbers of effects range median SQGs are exceeded, the incidence and magnitude of toxicity correspondingly increases. An additional observation in the current research is that associations between toxicity and mean SQGQs improve by removing chemicals with less predictive SQGs (Table 3) from the quotient calculations. This suggests that effectively representing toxicological modes of action may be more important to SQGQ predictive accuracy than simply including additional toxic chemicals to the representative chemical matrix.

Consideration of the issues discussed here and the results presented here leads to the recommendation that a standard method for combination of chemicals and respective SQGs should be adopted for the calculation of mean SQGQs to allow comparability between regions and surveys. The current research has demonstrated that, for the methods and data tested, SQGQ1 best meets the criterion for predictive accuracy and data utility. The SQGQ1 combination achieves this by incor-

Table 8. Chemicals and sediment quality guidelines used in chemical combination SQGQ1[a]

SQGQ1 chemicals	SQG concentration	Guideline type
Cadmium	4.21 μg/g dry wt	PEL [17]
Copper	270 μg/g dry wt	ERM [16]
Silver	1.77 μg/g dry wt	PEL [17]
Lead	112.18 μg/g dry wt	PEL [17]
Zinc	410 μg/g dry wt	ERM [16]
Total chlordane	6 ng/g dry wt	ERM [15]
Dieldrin	8 ng/g dry wt	ERM [16]
Total PAHs	1,800 μg/g OC	Consensus [20]
Total PCBs	400 ng/g dry wt	Consensus [35]

[a] SQGQ = sediment quality guideline quotients; PEL = probable effect level; ERM = effect range median; PAH = polycyclic aromatic hydrocarbon; PCB = polychlorinated biphenyl; OC = organic carbon.

porating a number of theoretical and empirical sediment quality guidelines (Table 8) from a variety of reliable sources [16,17,20,35]. As evaluated with the criteria for identification of optimal predictive ability, the incidence of toxicity increased steadily and average survival decreased steadily as mean SQGQ1 values increased. The incidence of toxicity was less than 5% and average survival was over 90% when mean SQGQ1 values were less than 0.1. The incidence of toxicity was greater than 90% and average survival was less than 50% when mean SQGQ1 values were greater than 2.3. The SQGQ1 calculation method incorporates SQGs for nine of the most commonly measured chemicals and best utilizes available data in the tested data sets. Until a more predictive method has been demonstrated, SQGQ1 is recommended to serve as the standard for calculation of mean SQGQs when comparing with acute toxicity to amphipods. SQGQ1 is calculated as

$$\begin{aligned}\text{SQGQ1} &= \Big(\Big(\sum ([\text{cadmium}]/4.21)([\text{copper}]/270)([\text{lead}]/112.18) \\ &\quad \times ([\text{silver}]/1.77)([\text{zinc}]/410)([\text{total chlordane}]/6) \\ &\quad \times ([\text{dieldrin}]/8)([\text{total PAH}_{\text{OC}}]/1800)([\text{total PCB}]/400)\Big) \\ &\quad \div 9\Big)\end{aligned}$$

An additional recommendation concerns sediments where unique chemical conditions prevail or regional anthropogenic activities warrant special consideration. In these situations, investigators might adjust the mean quotient derivation method to better fit local environmental or management concerns. A unique chemical example would be in a large agricultural watershed. Where endrin has been found to be a primary chemical of concern, managers may wish to incorporate the U.S Environmental Protection Agency's EqP-based sediment quality guideline for endrin into their mean quotient derivation method. This flexibility will ensure a major component in the watershed's chemical signature is not ignored when investigating the relationship between chemical mixtures and biological effects. With this type of specific objective, the derivation method that best demonstrates associations between chemicals of interest and biological effects will be the most effective tool for utilizing probabilities of toxicity and focusing local causality studies. It is important, however, that investigators report differences in the mean SQGQ calculation methods. We recommend use of the SQGQ1 method reported here for providing national and regional perspective and, if needed, include a local method for addressing any unique local chemical mixtures or specific management objectives. The ability to exercise flexibility with derivation methods should encourage and enhance the effective use of mean SQGQ values.

As with other empirical approaches to evaluating sediment contamination using sediment quality guidelines, limitations in the use of the mean SQGQs must be acknowledged. The SQGQ1 approach is meant to serve as a central tendency indicator (i.e., as means of multiple individual quotients) of contamination for a complex sediment matrix. The disadvantage of central tendency indicators is that they minimize the potential for impact from any one component. Therefore, when performing a sediment toxicologic evaluation, it is prudent to consider chemical exposure on an individual chemical basis in addition to the chemical matrix basis described here. Consideration of individual SQGs may help in situations when exposure to single chemicals is poorly represented by overall contamination within a sample. The point to emphasize is that mean SQGQs should not be used as the sole indicator of sediment contamination. They should be used as additional tools in our efforts to better understand the relationships between chemical exposure and biological response. Use of mean SQGQ1 should fit well in the conceptual framework of the sediment quality triad that emphasizes a weight-of-evidence approach to sediment quality assessment [36].

The current research has focused on manipulation of multiple chemical constituents to investigate their relationships with a single biological response. Critical values selected for classifying toxic biological responses are a major factor in the results presented here, and it should be noted that the critical values used as biological benchmarks for comparison over selected SQGQ ranges are themselves currently subject to investigation. New methods for determining toxicity are being examined [30,37,38] while others are reviewing current methods with more extensive data sets than have previously been assembled [28]. Results may influence the selection of critical values, such as *t* test and minimum significant difference determinates used in this study. As methods for determination of toxicity are revised, the need to revisit relationships between toxicity and chemistry also may follow.

The current research has focused on acute toxicity of sediment to marine amphipods as the sole measure of biological response. There is a need to further this research and derive optimal chemical mixtures for correlating with other test species and response parameters. Long et al. [6] reported that the incidence of toxic samples increased markedly when data from sublethal tests were added to those from the acute amphipod bioassays. Evaluation of mean SQGQs using large databases that include bioassays with reproductive, developmental, and chronic endpoints are needed to further validate use of mean SQGQs. Subsequent research should continue to expand investigation of impacts of multiple contaminants to measures in other environments and to measures of individual and community response in the field. Work with freshwater species is being reported [32], and relationships similar to those reported here are emerging between toxicity to several freshwater test species and mean SQGQs. Current efforts by Hyland et al. [14], Lowe and Thompson [39], and Fairey et al. [21] include investigations of marine benthic community response and have demonstrated increased probabilities of community impacts as mean SQGQ values increase. These efforts and the current research reported here should serve to strengthen empirical

evidence supporting the use of mean SQGQs as interpretive tools.

Acknowledgement—Funding support for the current research was from California's State Water Resources Control Board Bay Protection and Toxic Cleanup Program. Data from the SEDTOX database and Biscayne Bay were graciously supplied by NOAA through Jay Field (Office of Response and Restoration).

REFERENCES

1. Lee GF, Jones-Lee A. 1996. Can chemically based sediment quality criteria be used as reliable screening tools for water quality impacts? Learned discourses: Timely scientific opinions. *SETAC News* 16:14–15.
2. Long ER. 1996. Numerical sediment quality guidelines can be used as reliable interpretive tools. Learned discourses: Timely scientific opinions. *SETAC News* 16:18–19.
3. Sampson JR, Pastorak RA, Ginn TC. 1996. ER-L and ER-M values should not be used to assess contaminated sediments. Learned discourses: Timely scientific opinions. *SETAC News* 16: 29–31.
4. O'Conner TP. 1999. Sediment quality guidelines do not guide. Learned discourses: Timely scientific opinions. *SETAC News* 19: 28–29.
5. Long ER, MacDonald DD. 1999. Sediment guidelines do guide those willing to see—A reply to Dr. Tom O'Conner. Learned discourses: Timely scientific opinions. *SETAC News* 19:21–22.
6. Long ER, Field LJ, MacDonald DD. 1998. Predicting toxicity in marine sediments with numerical sediment quality guidelines. *Environ Toxicol Chem* 17:714–727.
7. Long ER, MacDonald DD, Cubbage JC, Ingersoll CG. 1998. Predicting the toxicity of sediment-associated trace metals with simultaneously extracted trace metal: Acid volatile sulfide concentrations and dry weight-normalized concentrations: A critical comparison. *Environ Toxicol Chem* 17:972–974.
8. O'Conner TP, Daskalakis KD, Hyland JL, Paul JF, Summers JK. 1998. Comparisons of sediment toxicity with predictions based on chemical guidelines. *Environ Toxicol Chem* 17:468–471.
9. U.S. Environmental Protection Agency. 1993. Technical basis for establishing sediment quality criteria for nonionic contaminants for the protection of benthic organisms by using equilibrium partitioning. EPA 822-R-93-011. Office of Science and Technology, Washington, DC.
10. Di Toro DM, Mahoney JD, Hansen DJ, Scott KJ, Carlson AR, Ankley AR. 1992. Acid-volatile sulfide predicts the acute toxicity of cadmium and nickel in sediments. *Environ Sci Technol* 26: 96–101.
11. PTI Environmental Services. 1991. Pollutants of concern in Puget Sound. EPA 910/9-91-003. U.S. Environmental Protection Agency, Seattle, WA.
12. Long ER, MacDonald DD. 1998. Recommended use of empirically-derived sediment quality guidelines for marine and estuarine ecosystems. *Human Ecol Risk Assess* 4:1019–1039.
13. Carr RS, Long ER, Windom HL, Chapman DC, Thursby G, Sloane GM, Wolfe DA. 1996. Sediment quality assessment studies of Tampa Bay, FL, USA. *Environ Toxicol Chem* 15:1218–1231.
14. Hyland JL, Van Dolah RF, Snoots TR. 1999. Predicting stress in benthic communities of southeastern U.S. estuaries in relation to chemical contamination of sediments. *Environ Toxicol Chem* 18: 2557–2564.
15. Long ER, Morgan LG. 1990. The potential for biological effects of sediment-sorbed contaminants tested in the National Status and Trends Program. NOAA Technical Memorandum NOS OMA 52. National Oceanic and Atmospheric Administration, Seattle, WA, USA, p 175.
16. Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manag* 19: 81–97.
17. MacDonald DD, Carr RS, Calder FD, Long ER, Ingersoll CG. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5:253–278.
18. U.S. Environmental Protection Agency. 1995. *Guidance for Assessing Chemical Contaminant Data for Use in Fish Advisories*, Vol 1—Fish Sampling and Analysis, 2nd ed. EPA 823-R-95-007. Office of Water, Washington, DC.
19. National Oceanic and Atmospheric Administration. 1989. Standard analytical procedures of the NOAA Analytical Facility, 2nd ed. NOAA Technical Memorandum NMFS F/NWC-92, 1985-86. National Status and Trends Program, U.S. Department of Commerce, Rockville, MD.
20. Swartz RC. 1999. Consensus sediment quality guidelines for polycyclic aromatic hydrocarbon mixtures. *Environ Toxicol Chem* 18:780–787.
21. Fairey R, et al. 1996. Chemistry, toxicology, and benthic community conditions in sediments of the San Diego Bay Region. CA State Water Resources Control Board Final Report. Sacramento, CA, USA.
22. Anderson BS, et al. 1997. Chemistry, toxicity and benthic community conditions in sediments of the southern California bays and estuaries. California State Water Resources Control Board, Sacramento, CA, USA.
23. Downing J, et al. 1998. Chemical and biological measures of sediment quality in the central coast region. California State Water Resources Control Board, Sacramento, CA, USA.
24. Jacobi M, et al. 1998. Chemical and biological measures of sediment quality and tissue bioaccumulation in the north coast region. California State Water Resources Control Board, Sacramento, CA, USA.
25. Phillips BM, et al. 1998. Sediment chemistry, toxicity, and benthic community conditions in selected water bodies of the Santa Ana region. California State Water Resources Control Board, Sacramento, CA, USA.
26. Anderson BS, et al. 1998. Sediment chemistry, toxicity, and benthic community conditions in selected water bodies of the Los Angeles region. California State Water Resources Control Board, Sacramento, CA, USA.
27. Hunt JW, et al. 1998. Sediment quality and biological effects in San Francisco Bay. California State Water Resources Control Board, Sacramento, CA, USA.
28. Field LJ, MacDonald DD, Norton SB, Severn CG, Ingersoll CG. 1999. Evaluating sediment chemistry and toxicity data using logistic regression modeling. *Environ Toxicol Chem* 18:1311–1322.
29. Long ER, et al. 1999. Magnitude and extent of chemical contamination and toxicity in the sediments of Biscayne Bay and vicinity. NOAA Technical Memorandum NOS ORCA 141. National Oceanic and Atmospheric Administration, Silver Spring, MD, USA.
30. Phillips BM, Hunt JW, Anderson BS. 2001. Statistical significance of sediment toxicity test results: Threshold values derived by the detectable significance approach. *Environ Toxicol Chem* 20:371–373.
31. Thursby GB, Schlekat CE. 1993. Statistical analysis of 10-day solid phase toxicity data for amphipods. *Abstracts*, 14th Annual Meeting, Society of Environmental Toxicology and Chemistry, Houston, TX, USA, p 99.
32. MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluations of consensus-based sediment quality guidelines for freshwater ecosystems. *Arch Environ Contam Toxicol* 39:20–31.
33. Swartz RC, Cole FA, Lamberson JO, Ferraro SP, Schults DW, DeBen WA, Lee II H, Ozretich RJ. 1994. Sediment toxicity, contamination and amphipod abundance at a DDT- and dieldrin-contaminated site in San Francisco Bay. *Environ Toxicol Chem* 13:949–962.
34. Swartz RC, Kemp PF, Schults DW, Lamberson JO. 1988. Effects of mixtures of sediment contaminants on the marine amphipod, *Rhepoxynius abronius*. *Environ Toxicol Chem* 7:1013–1020.
35. MacDonald DD, DiPinto LM, Fields J, Ingersoll CG, Long ER, Swartz RC. 2000. Development and evaluation of consensus-based sediment effect concentrations for polychlorinated biphenyls. *Environ Toxicol Chem* 19:1403–1413.
36. Chapman PM, et al. 1997. General guidelines for using the sediment quality triad. *Mar Pollut Bull* 34:368–372.
37. Smith R. 2001. The use of random-model tolerance intervals in environmental monitoring and regulation. *J Agric Biol Environ Stat* (in press).
38. Thursby GB, Heltshe J, Scott KJ. 1997. Revised approach to toxicity test acceptability criteria using a statistical performance assessment. *Environ Toxicol Chem* 16:1322–1329.
39. Lowe S, Thompson B. 1999. Results of the Benthic Pilot Study, 1994–1997. II: Identifying benthic responses to contamination in San Francisco Bay. San Francisco Estuary Regional Monitoring Program for Trace Substances. Technical Report 39. San Francisco Estuary Institute, Richmond, CA.

APPENDIX

Summations used in sediment quality guideline quotient (SQGQ) chemical combinations; references are given for source of summation technique

Total DDT [15,18] = $\sum$	([*o*′,*p*′-DDD][*p*′,*p*′-DDD][*o*′,*p*′-DDE][*p*′,*p*′-DDE][*o*′,*p*′-DDT][*p*′,*p*′-DDT])
Total chlordane [18] = $\sum$	([*cis*-chlordane][*trans*-chlordane][*cis*-nonachlor][*trans*-nonachlor][oxychlordane])
Total PCB [19][a] = $\sum$	([PCB8][PCB18][PCB28][PCB44][PCB52][PCB66][PCB101][PCB105][PCB118][PCB128][PCB138] [PCB153][PCB170][PCB180][PCB187][PCB195][PCB206][PCB209])2)
Low molecular weight PAHs [20] = $\sum$	([acenapthene][acenaphthylene][anthracene][fluorene][naphthalene][phenanthrene])
High molecular weight PAHs [20] = $\sum$	([benz[*a*]anthracene][benzo[*a*]pyrene][benzo[*b*]fluoranthene]‹benzo[*k*]fluoranthene][chyrsene] [fluoranthene][pyrene])
Total PAHs [20] = $\sum$	([low molecular weight PAHs][high molecular weight PAHs])

[a] This summation is based on work of O'Conner [19] for the National Oceanic and Atmospheric Administration's National Status and Trends Program while developing comparability between individual PCB congener summations and historic Aroclor equivalents. PCB = polychlorinated biphenyl; PAHs = polycyclic aromatic hydrocarbons; DDE = (1,1-dichloro-2,2-bis[p-chlorophenyl]ethylene).