RETURN TO UAO AR INDEX

The New Bedford Harbor Superfund site

Long Term Monitoring Program (1993-2009)

William Nelson and Barbara Bergen (Atlantic Ecology Division, Office of Research and Development, National Health and Environmental Effects Laboratory, U.S. EPA, Narragansett, RI, USA)

Abstract New Bedford Harbor (NBH), located in southeastern Massachusetts, was designated as a marine Superfund site in 1983 due to sediment contamination by polychlorinated biphenyls (PCBs). Based on human health and environmental concerns, the decision was made to dredge the PCB-contaminated sediments from the harbor. The Environmental Protection Agency's (EPA) Region I office (Boston, MA) and the U.S. Army Corps of Engineers (New England Division) required a long-term monitoring program to assess the effects and effectiveness of remedial activities at this site over the duration of the project. Staff at the Atlantic Ecology Division (AED), of the EPA's National Health and Environmental Effects Research Laboratory (NHEERL), had extensive experience in near-coastal monitoring and designed the long-term monitoring program for this site.

The goal of the New Bedford Harbor Long-Term Monitoring Program (NBH-LTM) is to measure spatial and temporal chemical and biological changes in sediment, water and biota to assess the effects and effectiveness of the remedial activities. A systematic, probabilistic sampling design was used to select sediment sampling stations. This unbiased design allowed the three segments of the harbor to be compared spatially and temporally to quantify changes

SDMS DocID 507955

resulting from dredging the contaminated sediments. Sediment is collected at each station and chemical (e.g., PCBs, metals), physical (e.g., grain size), and biological (e.g., benthic community) measurements conducted on all samples. To date, sediment samples have been collected five times: 1) 1993 - baseline collection, 2) 1995 - following the “Hot Spot” sediment removal from the upper harbor, 3) 1999 - before dredge field testing studies, 4) 2004 - prior to full-scale upper harbor remediation, and 5) 2009 - after several years of full-scale remediation. Also, PCB tissue concentration in deployed mussels was measured twice each year to assess water column PCB bioaccumulation.

This paper describes the overall NBH-LTM approach and the results from the five rounds of sample collections. There is a decreasing spatial gradient in PCB concentrations from the northern portion of the site (upper harbor, the primary source of PCBs) to the southern boundary (outer harbor, Buzzards Bay). Concurrently, there is an increasing spatial gradient in biological condition (e.g., benthic community diversity) from north to south. The contaminant and biological gradients have been maintained since the 1993 baseline collection; however, since the onset of full-scale remediation, PCB concentrations have decreased throughout the site and one of the benthic indices has shown significant improvement in the southern areas of the site.

Introduction

The introduction of anthropogenic contaminants into the environment has resulted in numerous aquatic areas throughout the United States where sediment is contaminated to potentially harmful levels (U.S. EPA, 2004). These sites are in both freshwater and marine locations, and pose various degrees of risk to both human health and the environment. The highest risk sites are identified on the Environmental Protection Agency's (EPA) National Priorities List (NPL) for cleanup under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and additional sites under the Resource Conservation and Recovery Act (RCRA), (U.S. EPA, 2004). Due to the costs and potential risks associated with contaminated site remediation, it is crucial that both the environmental effects and effectiveness of remedial activities be rigorously documented.

The need to implement effective monitoring programs at contaminated sediment sites has been emphasized many times (NRC, 1997; Goldberg and Betine, 2000; U.S EPA, 2005). Most recently, the National Research Council (NRC, 2007) conducted a review of contaminated sites and indicated that: 1) pre-remediation monitoring is essential to develop baseline information, 2) monitoring should be supported with rigorous statistical analyses, 3) although measuring fish-tissue contaminant concentrations is important, site-related impacts are possibly better assessed by examining benthic community condition, and 4) because of the extended time-frame associated with contaminated sediment remediation (especially “mega-sites,” where remedial costs are > $50 million), consistency and flexibility in monitoring are both critical to make useful assessments. Even before these national calls for robust monitoring programs, it was decided

that because of the potential cost of remediation at the New Bedford Harbor (NBH) Superfund Site, a long-term monitoring (LTM) program would be essential to document the benefits gained for the costs incurred; therefore, the NBH-LTM program was designed and first implemented in 1993 (Nelson et al., 1996a). The program incorporates all the recommendations called for in the 2007 NRC report and is described in this paper.

The NBH Superfund Site, located in Bristol County, Massachusetts (MA), extends from the shallow northern reaches of the Acushnet River estuary, south through the commercial harbor of New Bedford, MA, and into 17,000 acres of Buzzards Bay (Fig.1). Industrial and urban development surrounding the harbor has resulted in sediments becoming contaminated with high concentrations of many pollutants, notably polychlorinated biphenyls (PCBs) and heavy metals (e.g., cadmium, chromium, copper, and lead) with contaminant gradients decreasing from north to south (Nelson et al, 1996a). In 1979, fishing, shellfish harvesting, and lobstering were prohibited in NBH due to elevated levels of PCBs in seafood. The site was proposed for the EPA's NPL in 1982, and finalized on the NPL in September 1983. EPA's site-specific investigations began in 1984, including pilot studies on various remedial approaches (Nelson and Hansen, 1991).

Remedial activities in NBH have varied in size (i.e., number of yds^3 dredged) and duration (Fig. 2). In the area north of the Coggeshall St. Bridge (termed the upper harbor - UH, Fig. 1) several remedial activities have occurred in the subtidal area. The "Hot Spot" remediation removed sediments with PCB concentrations greater than 4,000 ppm and was completed in the fall of 1995 (Bergen et al., 2005). Limited dredging occurred between 2000-04, and full scale operations to remove the remaining contaminated sediments from the UH began in 2004. The

area between the Coggeshall St. Bridge and the Hurricane Barrier (termed the lower harbor - LH, Fig. 1) has several smaller areas identified for remediation. Several locations from the Hurricane Barrier to the outer closure line (termed the outer harbor - OH, Fig. 1) are also being investigated currently for possible remediation. Additional site history and information is available online at the New Bedford Harbor web site (http://www.epa.gov/region01/nbh).

Prior to the 1994 dredging of the "Hot Spot" in the upper harbor, it became evident that it would be necessary to answer the basic question, "How can the effectiveness of the remedial actions taken in NBH best be documented?" Based on that question, a comprehensive LTM program was designed in 1993 to assess the overall effectiveness of the remedial activities by documenting the environmental benefits gained for the money expended. There is no all-encompassing monitoring blueprint because Superfund sites can be quite different, both with respect to types of contaminants (e.g., PCBs, metals, dioxins) and physical characteristics (e.g., freshwater streams, estuaries, etc.). However, certain elements are inherent in any well-designed monitoring program. These include: 1) identifying the specific goals of the monitoring effort, 2) selecting the most appropriate variables to measure so that managers and/or scientists can determine whether those goals are met, 3) designing a statistically rigorous program to provide quantitative assessments, 4) establishing quality assurance guidelines that are flexible enough to incorporate technological improvements (e.g., analytical advances, new metrics, etc.), and 5) ensuring that the results can be understood by and transferred to a broad spectrum of stakeholders (e.g., managers, scientists, public).

The primary goal of the NBH-LTM program is to assess the effectiveness of the remediation by quantifying spatial and temporal biological and chemical changes in different

environmental compartments. Because the primary focus of remedial activities in NBH is the sediment (i.e., main repository for PCBs and other contaminants), concentrations of PCBs (measured as 18 congeners) and nine metals were quantified. Sediment toxicity tests were also conducted to ascertain short-term acute effects, and the benthic invertebrate community was characterized to assess the longer-term chronic effects of contaminants on biota. Several other variables such as sediment grain size, concentrations of total organic carbon (TOC) and acid volatile sulfide (AVS) were measured because they can affect chemical bioavailability and are important when interpreting the biological effects.

A secondary goal of this LTM program is to show compliance with applicable or relevant and appropriate requirements (ARARs), a requirement at all Superfund sites (USEPA, 2005). In this case, two of the most important ARARs for measuring remedial success are compliance with water quality standards and Food and Drug Administration (FDA) standards for PCB levels in seafood. PCB tissue concentrations in selected marine organisms from NBH are being monitored separately by the State of Massachusetts to assess the risks of consuming fish and shellfish by humans and wildlife; however, directly linking some biota, such as fish, to specific remedial actions can be difficult because of their mobility and exposure to offsite conditions (NRC, 2007).

An alternative approach is to use deployed organisms to quantify changes in PCB water column concentrations. Filter-feeding bivalves (e.g., mussels, oysters) "sample" the water column almost constantly, thus integrating the effects of tides and weather. Because of this, blue mussels have been used extensively for many years to quantify chemical pollution (Farrington, 1983; Tanabe et al., 1987). Caged mussels have been used in several previous monitoring and

research efforts in NBH, and an extensive data set has been compiled on PCB accumulation in blue mussels, *Mytilus edulis* (Bergen et. al., 1993; Nelson et al., 1995). These studies demonstrated a good relationship between mussel PCB tissue concentrations and temporally integrated water column concentrations. Therefore, PCB bioaccumulation in blue mussels was selected for the NBH-LTM program as a way to monitor changes in PCB water column concentrations over time.

With this framework in mind, this paper describes the LTM program established for the NBH Superfund site to document the effectiveness of remedial activities. First implemented in 1993 (baseline), the program is currently on-going with five sampling events completed to date, and it incorporates all the elements discussed above and meets the criteria suggested by other groups, such as the NRC. It uses the general approaches of other monitoring programs (e.g. EPA's Environmental Monitoring and Assessment Program, EMAP, Paul, et. al., 1999), such as a probabilistic sampling design and the "triad" approach (Chapman, 1990) which combines chemistry, toxicity, and benthic community analysis. Finally, as with many monitoring programs, questions arise as to why certain endpoints were selected, how sampling stations are located, etc. For the NBH-LTM, initially designed and implemented in 1993, a balance was struck between the goals of the program, budgetary realities, and state-of-the-science available at the time. This overview paper describes the overall monitoring program and approach and presents selected data collected during the five sampling periods spanning the years from 1993 to 2009. More comprehensive and detailed analyses delving deeper into the specific data will be presented in subsequent papers.

Methods

Station Locations

To meet the goals of the NBH-LTM program, a probabilistic sampling design was used and stations were selected by applying a systematic hexagonal grid, then sampling as close as possible to the center of each hexagon in 1993 (Fig. 1). Each subsequent LTM collection used these original locations. Because different remedial activities are planned within the upper harbor (UH), lower harbor (LH) and outer harbor (OH) areas, spatial and temporal changes within each of these three segments, as well as the entire Superfund site, needed to be quantified separately. To accomplish this, the hexagon size in the sampling grid was adjusted to include approximately 30 stations per segment (Fig. 1); however, due to the shoreline configuration, and non-sampleable areas within each segment, the actual number of stations was 27 in the UH, 29 in the LH, and 23 in the OH.

Sediment Sampling

Sediment collection procedures were identical at each of the sampling locations (U.S. EPA, 1995). Sediment was collected using a Young-modified van Veen grab sampler (440 cm^2 in surface area). At each site, numerous grabs were taken to obtain an adequate amount of sediment for chemical and toxicological analyses. The NBH-LTM program is designed to quantify changes over an extended period of time, especially changes resulting from remedial activities; therefore, only the top 2 cm of these grabs were used in the composite for chemical analysis because these are most reflective of current sediment concentrations.

From each individual grab, a sub-sample was removed for acid volatile sulfide (AVS) analysis. The remainder of the top 2 cm from each grab at a given site was composited and homogenized before sub-samples were taken for chemistry, sediment toxicity, and grain size analyses. Prior to laboratory analysis, PCB and TOC sediment samples were kept cold at 4°C, while the metals and AVS samples were frozen. Sediment for toxicity testing was taken from the same composite, press-sieved through a 2-mm mesh stainless-steel screen, thoroughly homogenized, and stored at 4°C until testing. Finally, approximately 100 ml of the composited sediment were placed in a polyethylene bag for grain size analysis.

Three additional grabs were collected and processed separately at each station for benthic community analysis. These grabs were a minimum of 7 cm deep to ensure sampling of the benthic community. Each grab was processed according to the procedures described in Reifsteck *et al*. (1993). A small core was taken from each grab for sediment grain size characterization, independent of the grain size analysis for the chemistry composite described previously. The remaining sample was sieved through a 0.5 mm screen using a backwash technique to minimize damage to soft-bodied animals. Samples were preserved in a buffered 10% formalin and seawater solution with rose bengal added as a vital stain. Two of the three samples were processed for taxonomic identification, the third was archived.

A full suite of variables was measured at each station for the first four sampling events (Table 1); however, due to funding constraints, and an evaluation of results from the previous four sampling events, the last round of monitoring in 2009 did not include metals analysis or sediment toxicity (see Discussion).

Sediment Chemistry

A rigorous quality assurance project plan (QAPP) was developed to ensure data comparability from one sampling event to another and to be flexible enough to allow for innovation in analytical techniques. The QAPP is performance based, delineating the guidelines that must be met by participating laboratories (e.g., analysis of certified standard reference sediment, matrix spike recoveries, field duplicate agreement, blanks, etc.) but does not specify particular analytical methods. In the interest of brevity, the methods reported here are general, with each year of the program having some slight modifications to analytical techniques. However, the same quality assurance standards have been met every year and specific details of the quality assurance standards promulgated for the NBH-LTM program can be found elsewhere (U.S. ACOE, 1999). Other information and technical documents related to the site, including details of all analytical techniques used in each sampling year, can be found on the internet at http://www.epa.gov/region01/nbh.

PCB analysis While there are multiple ways to quantify PCBs (e.g., Aroclor® mixtures, homologues, individual congeners), the sediments in the NBH-LTM program were analyzed for 18 of the 209 PCB congeners. These are the same 18 congeners that are used in the National Oceanic and Atmospheric Administration's (NOAA) National Status and Trends Program to assess marine environmental quality (Calder, 1986).

For PCB analysis, approximately one gram of sediment was solvent extracted with acetone and methylene chloride. Samples were cleaned by open column chromatography with activated copper. Extracts were further cleaned by treatment with concentrated sulfuric acid,

then solvent-exchanged to hexane, and analyzed by gas chromatography with an electron capture detector for 18 individual PCB congeners listed in Table 1. Total PCB concentrations were calculated as the sum of these 18 congeners. As part of this analysis, a second gram of sediment was weighed, dried and weighed again to determine moisture content.

Metals analysis Although sediment PCB concentrations are the primary contaminant of concern in NBH, metals are present in high concentrations in some areas. Because one goal of the remediation is to remove these co-located contaminants, sediments were also analyzed for metals. Nine metals were selected for analysis including: copper (Cu), cadmium (Cd), lead (Pb), zinc (Zn), arsenic (As), selenium (Se), mercury (Hg), chromium(Cr), and nickel (Ni).

Sediments were analyzed for metal concentrations by ultra sonicating a mixture of approximately five grams of sediment and 2M nitric acid. The resultant extract was brought to volume and analyzed on an inductively coupled plasma spectrophotometer (ICP) or an atomic absorption spectrophotometer (AA) as necessary. Total bioavailable metals used for comparison with biological responses is defined as the molar sum of all metals measured, using a 2 M nitric acid extraction. In this study, a sum of the six divalent metals (Ni, Pb, Cd, Cu, Zn and Hg) was calculated to estimate bioavailable metals.

Acid Volatile Sulfide (AVS) AVS was quantified as the amount of sulfide released upon extraction of the sediment with a cold 1M hydrochloric acid (HCl) solution. The sulfide liberated in this process was trapped and measured with a sulfide-specific electrode (Boothman, et al., 1992).

Total organic carbon (TOC) The analysis of TOC consisted of first removing the sediment

inorganic carbon from carbonates and bicarbonates by acid treatment. Organic compounds were combusted and the resultant carbon dioxide measured by direct non-dispersive infrared detection.

Sediment Toxicity

The test species used in this project is the euryhaline benthic amphipod, *Ampelisca abdita*, and it has been used previously in NBH (Otis and Averett, 1988). Following EPA procedures (U.S. EPA, 1994), sediments were added to exposure chambers one day before the amphipods. The exposure chambers were 1-L canning jars with an inverted glass dish as a cover. Two hundred ml of control or NBH sediment were placed on the bottom of each of five replicates per sediment sample and covered with approximately 600 ml of filtered Narragansett Bay seawater. Gentle aeration of the overlying water and lighting were continuous during the test period. Twenty sub-adult amphipods (passing through a 1.0 mm, but retained on a 0.71 mm screen) were added randomly to each replicate chamber. Amphipods were exposed to test sediments for 10 days under static conditions. Salinity of the overlying water ranged from 30 to 35 parts per thousand (ppt) and temperature was maintained at 20 $\pm$ 1°C. Additional performance control sediments were used to assess the survival of amphipods under a set of standardized conditions.

Sediment grain size distribution

The sediments were analyzed in a two-phase process and categorized as gravel (> 2 mm), sands (> 63 μm and < 2 mm) and silt/clay (< 63 μm). First, the silt/clay fraction was separated from the rest of the sediment using wet-sieving and pipette analytical techniques. Shell

fragments and organic debris larger than 1 mm were removed prior to this analysis. Next, sediments were classified by their size distribution to determine the relative proportions of gravel, sand, and silt/clay.

Benthic Community Analysis

In the laboratory, macrobenthos were identified to lowest practical taxonomic level (generally species) and counted. A complete description of the methods can be found in the Environmental Monitoring and Assessment Program (EMAP) Laboratory Methods Manual (U.S. EPA, 1995). All analyses and sample collections were performed in accordance with a strict Quality Assurance/ Quality Control program (Valente and Strobel, 1993).

As an initial step prior to quantitative analysis of the benthic data, a rigorous evaluation of the five collections was conducted to ensure as much taxonomic comparability among the collections as possible. For example, a species identified as "X" in 1993 could have undergone a taxonomic name change and be called "Y" in 2009. Further, an animal identified to species one year, may only be identified to genus another year due to it's physical condition after sieving (e.g., small worms) or the proficiency of the taxonomist. It was necessary also to select a specific level of taxa across all five collections to provide a consistent evaluation. For example, if only animals identified to species were used, it would omit anything identified only to genus or family. This would leave out "difficult to identify" animals, such as marine oligochaetes, which may be present in large numbers in impacted areas like NBH, and still be meaningful in an overall benthic assessment. Therefore, it was decided to include all identified taxa when

evaluating benthic condition (approximately 90% were identified to the species/genus level).

Four metrics were used to assess the status of the benthic community. Taxonomic richness was determined as the total number of unique taxa per station, cumulative for both grabs per station. Dominance (i.e., evenness) was quantified by determining the number of taxa that constituted 75% of the total number of organisms at each station. This was done by totaling the number of individuals in each station replicate, calculating what percent each taxon comprised of that total, then ordering the cumulative percentages from highest to lowest until 75% was reached. The mean was then calculated for each station, harbor segment and collection year. Two other indices, Shannon's H′ (Washington, 1984), a widely used diversity index, and the EMAP benthic index for the Virginian Biogeographic Province (Paul, et. al., 2001), where NBH is located, were also calculated for each station. These four metrics were selected for use because they each evaluate different aspects of the benthic community; however, collectively, they provide a comprehensive suite to assess the overall condition of the benthos throughout the entire Superfund site.

Mussel PCB Bioaccumulation

Detailed methods for collecting and deploying blue mussels are found in Nelson and Gleason (1996b). Briefly, mussels (5 -7 cm in length) were collected from uncontaminated locations in either MA or RI and placed in polyethylene mesh bags. Four independent replicate bags with 25 mussels each were deployed 1 meter above the bottom at three sites: the Coggeshall St. Bridge (NBH-2), at the boundary between the UH and LH; the Hurricane Barrier

(NBH-4), at the boundary between the LH and OH; and at West Island (NBH-5), a reference site in Buzzards Bay (Fig 1). After 28 days, the mussels were retrieved and frozen. Prior to analysis, mussels were thawed, shucked and homogenized. Two grams of the homogenate were extracted with acetonitrile and pentane, solvent-exchanged to hexane and analyzed by gas chromatography for the same 18 PCB congeners quantified in the sediments (Bergen et al., 1993). Since 1993, mussel deployments have occurred twice a year at each station, except monthly during the Hot Spot remediation, for a total of 41 deployments.

Statistical Analysis

Spatial Interpolation Spatial interpolation of the PCB sediment data was conducted by kriging using the Geostatistical Analyst Extension for ESRI's ArcGIS mapping and analysis software (Johnston et al., 2001). The ordinary kriging method was optimized for each harbor segment using the 1993 NBH-LTM PCB data. The following parameters were adjusted to optimize the method for each segment: model type, lag size, number of lags, and search neighborhood type and geometry. The "optimized" model was selected on the basis of minimizing prediction error diagnostics including: mean, root-mean-square, average standard error, mean standardized, and root-mean-square standardized. Once a model was selected for a harbor segment, it was then applied to the 1995, 1999, 2004, and 2009 data for that segment. The resulting surface was saved to ESRI GRID format, then clipped to the outline of the appropriate segment. The value table (contains concentration and the number of cells with each concentration) was exported to dbase format (.dbf) and used to generate contour maps (Fig 3).

Spatial and Temporal Analysis Statistical differences ($p < 0.05$) among the harbor segments within a collection year and temporal changes from the 1993 baseline sampling were tested using a repeated measures analysis of variance, with an unstructured variance-covariance matrix for collection years, to test for segment (i.e., UH, LH, OH) by year interaction. When the segment-by-year interaction was significant, individual harbor segments were tested for year differences and individual years were tested for segment differences. Because of the wide range in PCB concentrations from the UH to the OH, these values were transformed ($\log_{10}$) prior to analysis; all other variables were not.

Results

Chemical Endpoints

PCBs The spatial PCB distributions for five ranges (0-1, 1-10, 10-50, 50-100 and >100ppm), and the overall surface weighed average concentration (SWAC) for each segment/year combination, are presented in Figure 3. The largest source of PCBs to NBH was in the northern-most portion of the UH, and this is reflected in a distinct spatial gradient in PCB concentration from the UH to the LH to the OH (Fig. 3). For example, in 1993, the UH SWAC PCB sediment concentration was 109 ppm, with significantly decreasing concentrations in the LH (7.4 ppm) and OH (0.8 ppm). This spatial pattern of significantly decreasing PCB concentrations from the UH to OH is consistent for each of the five LTM collections.

All PCB concentration temporal comparisons are relative to the 1993 baseline collection.

In the UH, there was a significant increase in PCB concentration in 1995, which decreased back to the baseline concentration in 1999, followed by a statistically significant ($p < 0.05$) decrease in the 2004 and 2009 collections (Fig. 3). The LH concentration decreased significantly in 2004 and 2009 relative to the baseline concentration, and the OH concentrations in 1995, 1999, 2004 and 2009 were significantly less than the baseline value.

Metals The overall spatial trend for the metals is similar to that for PCBs in that there is generally a decreasing gradient from the UH (i.e., highest mean concentrations) to the OH (lowest mean concentrations) (Table 2). For all eight metals, the concentrations are lowest in the OH when compared to the UH and LH. With the exception of Cd and Ni in 1995 and Cd in 1999, all of the OH metal concentration are significantly lower than the LH concentrations. The UH concentrations are also significantly higher than the LH concentrations for Cr, Pb, Hg (except 1993), Ni, and Zn. For As, Cu (except 1995) and Se, the concentrations are not significantly different between the UH and LH.

Temporally, all of the metals with the exception of Cd, show significant increases in the UH and LH in 1995 compared to 1993 (Table 2). In the 1999 collection, the patterns begin to diverge. Cadmium and Pb are back to baseline concentrations in both the UH and LH, while As, Hg, and Ni are still significantly higher in both areas and Se is significantly lower in both. Copper remains higher in the UH, then returns to baseline in the LH, while Zn is significantly higher than baseline in the LH. By 2004, Cd and Cr are similar to baseline concentrations in both segments, while As, Cu, Ni, Hg, and Se are significantly higher than baseline in both areas. Lead and Zn are only higher in the LH. In the 1995 OH samples, all of the metals with the exception

of Cd, Cu, and Hg were statistically increased. In 1999, they were all back to 1993 baseline concentrations. In 2004, As, Cr, Pb, Ni, and Zn again demonstrated significant increases over baseline.

Sediment Toxicity

The sediment toxicity data show a consistent spatial response for each of the four sampling years, with greatest mean survival in the OH, followed by the LH and the UH (Table 3). In all comparisons, these differences were statistically significant ($p < 0.05$), except in 1993 where survival in the UH and LH were not different.

The temporal comparisons to the 1993 baseline data indicate that in both the UH and LH, survival was lower in 1995 and 1999; however, 2004 was not significantly different from 1993. In the OH, survival was significantly lower for 1995, 1999, and 2004 collections compared to the 1993 baseline collection.

Benthic Community

The simplest benthic health measure presented here is richness, or the number of taxa present at each station (Table 4). There is a consistent spatial gradient with significantly lower taxa per station in the UH, except for 1995 when the UH and LH were the same. The OH exhibited significantly higher numbers of taxa than either the UH or LH in each of the five collections. Temporally, there was a significant reduction in UH richness in 1999 and 2004

compared to the 1993 baseline value; however, the mean value was higher again in 2009 and not different from 1993. The LH exhibited a similar pattern with 2004 being significantly reduced compared to 1993, but also rising in 2009. The OH taxa numbers were significantly lower in both 1999 and 2004, but were again comparable to 1993 levels by 2009.

A second method used to investigate the status of the benthic community was the number of dominant species in each segment of the harbor. As an example, the results for the 1993 baseline sampling are shown in Figure 4, although a similar pattern was observed for the other four years also. A clear spatial gradient was observed in 1993 with the UH having three dominant species, the LH five, and the OH twenty-six. Further, the dominant species in the UH was *Streblospio benedicti*, an indicator of impacted benthic condition. The dominant species in the LH was the opportunistic clam, *Mulinia lateralis*. The OH displayed a much more diverse benthic community with smaller numbers of larger animals, indicative of a healthier condition. Temporally, the same spatial pattern for each of the three harbor segments was observed for the other four LTM collections.

Shannon's H′ index is a measure of benthic diversity (higher values indicate greater diversity) that incorporates both abundance and evenness (Washington, 1984). When applied to the NBH benthic data, a consistent spatial pattern is observed relative to the three segments of the harbor (Table 4). For each of the five LTM collections, the UH was always significantly lower than the LH, which was always significantly lower than the OH. Temporally, in the UH, the collections in 1995 and 2004 were significantly higher than 1993. In the LH, each of the collections were not different from 1993 except for 2009, which was significantly higher. Relative to the OH, the mean values in 1995, 1999, and 2004 were significantly lower than 1993,

but increased and were not different again by 2009.

The final metric applied to the NBH benthic data was the EMAP benthic index (EMAP-BI) for the Virginian Biogeographical Province (Paul, et. al, 2001). This biodiversity index was developed to assess estuarine benthic condition from Cape Cod, MA to the mouth of Chesapeake Bay, VA. It incorporates a salinity-normalized Gleason's D biodiversity metric as well as the abundances of spionid polychaetes and salinity-normalized tubificid oligochaetes. Positive EMAP-BI values indicate a good condition while negative values are indicative of impaired communities. Consistent with the other measures used here, there was a similar spatial pattern for the EMAP-BI; the UH exhibited the worst condition, as evidenced by the large negative values observed each collection year (i.e., degraded condition), the LH was significantly improved relative to the UH, with values near zero, and the OH was always significantly highest with positive values, indicative of a good benthic community (Table 4, Fig. 5). Temporally, the EMAP-BI did not change significantly from 1993 in the UH for any of the four subsequent collections. The LH showed a similar pattern until 2009, when the index switched to positive and was also significantly higher than 1993. The same pattern occurred in the OH, with each year similar to the 1993 baseline until 2009, when a significantly higher value was obtained.

Mussel Bioaccumulation

The means of total PCBs (as the sum of 18 congeners) in the blue mussel tissue for the 41 deployments at the three stations are shown in Figure 6a. There is a significant spatial gradient among stations, with an approximate four fold decrease in overall mean concentration between

stations NBH-2 (19 ppm) and NBH-4 (5.1 ppm) and an order of magnitude decrease between station NBH-4 (5.1 ppm) and NBH-5 (0.5 ppm). These PCB differences among stations are maintained over time (Fig. 6b, note the difference in scales among stations). Each station exhibits seasonal variability due to the mussel reproductive cycle where lipid-rich gametes (along with lipophilic organic contaminants such as PCBs) increase during the year, then decrease dramatically during spawning.

Discussion

In the five NBH-LTM collections to date, a large amount of data were collected and analyzed. While data collection can be relatively straight forward, a greater challenge involves interpreting the results in the context of the two programmatic goals to assess remedial effects and effectiveness. In this paper, remedial effects are evaluated based on spatial and temporal changes in the individual measurements, relative to remedial activities. Remedial effectiveness on the other hand, goes beyond merely documenting changes, and involves looking at those changes in the context of site-specific or general criteria (e.g., sediment clean-up levels), and includes a more integrative assessment of all the data collectively (e.g., how do changes in chemical concentrations relate to biological metrics?). In addition, it must be remembered that the remediation of PCB contaminated sediment in NBH is only about 30% complete as of the 2009 LTM collection. Therefore, this paper is not a final evaluation of the overall NBH remediation; rather, it is an interim appraisal of the NBH-LTM program's ability to document effects to date, as well as the effectiveness of those changes.

Overall, the results of NBH-LTM to date present a fairly consistent narrative relative to remedial effects. At the spatial scale of the entire Superfund site, the PCB data demonstrate a clear decreasing gradient in sediment contaminant concentrations from the UH to the LH to the OH during each of the five collection events (Fig. 3), and the biological data (e.g., benthic community) show an increasing gradient along the same transect (Fig 5). These data can also be inspected at a finer spatial scale to document effects within a harbor segment relative to remedial activities. For example, there is a significant increase in the 1995 PCB concentration in the UH, which can be related to the "Hot Spot" dredging from 1994-95 (Fig. 2), a mass removal operation to dredge sediments with PCB concentrations > 4,000 ppm. Figure 7 shows the PCB concentrations at each UH station as well as their location. From the graph, it can be seen that there are increases in sediment PCB concentrations at some stations during the 1995 collection (dashed red lines); however, this occurred only in those stations closest to the Hot Spot dredging area (Fig 7b). These data show that there was a localized effect, mostly likely due to resuspension while dredging PCB-contaminated sediments with concentrations ranging from 4,000 – 100,000 ppm. A similar pattern was observed in 2009, where several of the individual UH stations exhibited slight increases in PCBs in 2009 (Fig 7a), presumably due to resuspended PCBs during the full-scale dredging activities in the UH. Once again, these increases are localized to the immediate vicinity of the dredge area within the UH and had less of an impact than the Hot Spot dredging (i.e., 2009 UH mean PCB concentration was still significantly lower than the 1993 baseline value).

The overall strategy used during the Hot Spot and subsequent dredging was that localized, short-term increases in sediment PCB concentrations were acceptable, as long as there was no

significant PCB transport to the LH and OH due to the dredging. The rationale for this approach was that the entire UH, including the Hot Spot area, was to be dredged again during future remedial activities to obtain final PCB clean-up concentrations. An extensive operational water quality monitoring program was implemented to ensure this strategy worked (Bergen, *et. al.* 2005) and operational water column monitoring has demonstrated that PCBs are not being transported to cleaner areas of the harbor. The LTM data show that the UH PCB sediment concentrations were back to baseline in 1999 (Fig 3). In addition, the LH and OH PCB concentrations from 1995 show no increase from 1993, further demonstrating the localized nature of the dredging effects and lack of far-field PCB transport.

From a temporal perspective, the major observation from Figure 3 is that there is a significant decrease ($p < 0.05$) in average PCB concentration within each segment in 2004 and 2009 compared to 1993. The remedial timeline (Fig. 2) indicates that up to the 1999 LTM collection, only the "Hot Spot" remediation had occurred; however, by the 2004 LTM collection, approximately 38,600 yd^3 of PCB contaminated sediment had been removed as a result of the Hot Spot and other limited dredging activities. The subsequent effect of these remedial activities is reflected in significantly lower PCB concentrations in each segment relative to the 1993 baseline data (Fig. 3), indicating that removing the PCB source in the UH has corresponded with beneficial effects observed throughout NBH. From the 2004 collection to the 2009 collection, another 159,000 yd^3 of contaminated sediment were removed from the UH and the trend of significantly lower mean PCB concentrations, compared to the baseline, has continued in each segment of the harbor. Looking collectively at the spatial and temporal changes in the LTM PCB sediment data, in the context of the goals to limit transport while removing significant amounts

of contaminated sediment that serve as the primary PCB source for the entire Superfund site, it is clear that the dredging to date can be deemed to be effective, but unfinished.

The spatial PCB gradient observed in the sediment data also is reflected in the deployed blue mussel PCB tissue data (Fig 6a), and has been observed for indigenous ribbed mussel populations in NBH as well (Bergen et al., 2001). Previous studies in NBH have demonstrated that blue mussels closely follow water column PCB concentrations (Bergen et al, 1993) and that mussel PCB bioaccumulation was more closely linked to storm events, than any dredging activity (Bergen et al., 2005). Temporally, the entire data set shows that between 1993 and 2009, no net change in mussel PCB bioaccumulation has occurred, indicating that overall water column PCB concentration has not been altered dramatically along this gradient over time either (Fig 6b), which may seem inconsistent with the significant changes in PCB sediment concentrations over the same time period. While reductions in PCB sediment concentrations in the UH indicate that the overall mass of PCBs removed has been significant, they have occurred over a relatively small area of the northern UH (Fig 7a). Most of the UH is yet to be remediated, therefore, the overall UH surficial area where sediment PCBs are available to partition into the water column is still large and water column concentrations have not changed enough to impact mussel PCB bioaccumulation. It is reasonable to expect that once remediation is completed throughout the entire UH (i.e., PCB sediment concentration of 10 ppm), water column PCB concentrations will also decrease, leading to decreased mussel PCB tissue concentrations. Continued mussel deployments will track these changes to determine if this is the case.

It is very important to not only examine the LTM results in terms of spatial and temporal changes, but also in the context of how effective these changes are in improving the overall

environmental quality in NBH. In order to make this assessment, the data need to be evaluated in terms of both site-specific criteria, as well as a more integrated approach such as the sediment quality triad (Chapman, 1990), which examines chemistry, sediment toxicity, and benthic community data in a comprehensive manner. For the NBH-LTM data set, a comprehensive table was generated to compare the site-specific PCB sediment criteria with other measured variables for which "threshold" values (discussed below) were assigned (Table 5). For example, the PCB clean-up level established in the NBH Record of Decision (http://www.epa.gov/region1/superfund/sites/newbedford/38206.pdf) is 10 ppm in the UH, 50 ppm in the LH, and the "presumptive" clean-up concentration for the OH is 10 ppm (currently, the OH is undergoing human health and ecological risk assessments to establish a final value). The mean PCB sediment concentration for each segment-collection year combination was compared statistically to its corresponding criterion value using a single sample t-test. If the mean was significantly worse than the criterion, the box in Table 5 was colored red, not significantly different from it, yellow, and significantly better, green. The bolded value in the colored boxes indicates the percent of stations within that segment-collection year combination that violated the criterion. Because a systematic grid was used in the sampling design, this number corresponds directly to the actual area within a segment that violates the criterion (within a few percent, some hexagons partially overlap land so it is not exactly 100%). This same approach (i.e., statistical test, box colors) was applied to several other variables that do not have NBH site-specific criteria, but were assigned "threshold values" based on the literature. The values in the non-colored boxes are the means of additional calculated values for which it is difficult to assign a "good" or "bad" value; however, they are included in Table 5 to help put the

NBH data in context of other commonly used metrics of sediment quality.

For the contaminant data, PCBs and metals are separated to help identify relationships to corresponding biological effects (Table 5). It is apparent that the PCB criterion of 10 ppm in the UH is violated for each of the five collections, although the percent of area exceeding the criteria is decreasing during 2004 and 2009, consistent with the significant decreases in mean PCB concentrations due to remedial dredging activities (Fig. 3). With respect to the LH and OH, each of the segment-collection year combination boxes is green, and the values are all zero, indicating that there are no stations violating the existing criteria (although this may change in the OH when the final PCB clean-up concentration is set).

With respect to metals, there are no NBH site-specific remedial criteria for sediment metals concentrations; however, it is possible to estimate their effect by quantifying the amount of biologically available metals, evaluated as the difference between simultaneously extracted metals (SEM) and total acid volatile sulfides (AVS) in the sediments (DiToro, 1990). A "threshold value" of SEM-AVS < 0 is considered "good," because AVS is greater than SEM and theoretically metals are not available to cause toxicity. Table 5 shows that at the spatial scale of a segment, bioavailable metals should not be a major toxicity problem in NBH for any segment-collection combination (green boxes), although the bolded numbers in those boxes indicate that there are some areas within each segment that have sediment metals concentrations that are potentially toxic. These data also indicate that as PCB remediation has continued in the UH, co-located bioavailable metals concentrations in the sediment have also been reduced (e.g., the UH area with SEM-AVS>0 has decreased from 37% in 1993 to 11% in 2009). In fact, these data were used to assess whether to collect metals data for the LTM 2009 collection as a cost-saving

measure, which is discussed below.

To further put the chemistry data in perspective, the mean effects range median quotient (ERM-Q) was calculated for both PCBs (sediment PCB concentration divided by effects range median, ERM) and metals (sediment As, Cd, Cr, Cu, Pb, Hg, Ni, Zn concentrations divided by their respective ERM value, then summed at each station) and the values shown in Table 5. While this is not a site-specific criterion value, it has been used to identify potential sediment toxicity (Long, *et. al.*, 1998). Values < 1 are generally associated with minimal toxicity, while values above 1.6 are associated with significant toxicity. The metals ERM-Q values in the LH and OH are below 1, indicating minimal effects, while those values in the UH are borderline, where some toxicity may be expected. In contrast, the PCB ERM-Q's are orders of magnitude higher than the metals, well above any borderline values and would be expected to be very toxic using this screening metric.

The threshold value used here for the sediment toxicity data is subjective because there is no absolute survival that is "good." However, based on the experience of many familiar with this test, a value of 80% survival or above is considered generally acceptable as "good" survival. Therefore, for the purposes of this paper, 80% survival or greater was considered good (i.e., green), significantly lower was "bad" (i.e., red), and not different from, but less than 80% (e.g., 78%) was considered borderline and shaded yellow. Table 5 clearly shows that both the UH and LH are considered bad for each of the years that sediment toxicity was evaluated. In the OH, only the baseline year of 1993 would be considered good, while the other years were borderline.

The results of the amphipod sediment toxicity test are not always consistent with the chemistry and benthic community data for the four collections. For example, looking at Table 5,

the OH PCB and metals concentrations were consistently in the “good” category (green), as was the EMAP-BI, yet the sediment toxicity fluctuated between “good” to “borderline” (yellow) during the same time period. In addition, when looking at the percent survival within each segment (Table 3), sediment toxicity was much more variable than the benthic metrics (Table 4). In addition, it is clear from the data that there is a lot of variability in mean survival among sampling years, with ranges of 53% in the UH (2 - 55%), 36% in the LH (32 - 68%), and 16% in the OH (75 - 91%). Further, it is difficult to explain the cause of this variability in the context of remedial activities, especially considering that the collection year with the lowest absolute survival values for each segment (i.e., 1999) had no remedial activities. The one consistent overall pattern is that the harbor segments with the greatest variability and lowest survival correlate directly with the existing PCB and metal contamination gradient from the UH to the OH.

A number of metrics were used to quantify the third component of the sediment quality triad (i.e., benthic condition). As with sediment toxicity, selecting a definitive “good/bad” criterion number is difficult and subjective. For this paper, we selected the EMAP-BI metric to assign a threshold value to for two reasons: 1) it was used extensively throughout the Virginian Biogeographical Province (where NBH is located) to assess benthic condition, and 2) negative EMAP-BI values are indicative of an impaired benthic community, therefore, it is possible to assign a good/bad numerical assessment value. In Table 5, values significantly greater than zero are considered good, (i.e., green), significantly less are considered bad (i.e., red), and not different from zero are borderline (i.e., yellow). The EMAP-BI demonstrates that the benthic community in the UH is severely impacted over the entire time period, while the OH has been

consistently in the “good” category. The most interesting response for this metric is in the LH, where the index has transitioned from bad in 1993 and 1995 to borderline in 1999 and beyond. Further, the spatial extent of the borderline condition has decreased to 45% in 2009, which was also a significant improvement from the 1993 baseline year (Fig. 5). For the purpose of putting these numbers into context, two other indices are presented also, Shannon’s H’ and the number of taxa per station. It is more difficult to assign a specific criterion or threshold value for these indices (e.g., are 28 species good and 25 bad?); however, they both exhibit the same general pattern of improving benthic condition from the UH to OH as the EMAP-BI.

By examining all three components of the sediment quality triad in Table 5 collectively, it is reasonable to conclude that PCBs are the primary contaminant of concern in NBH, not metals, and that this is conclusion is clearest in the UH. Further, since 2004, when full-scale remediation was initiated, the area above the PCB criterion value of 10 ppm in the UH has been reduced (Table 5) and the mean concentration in this segment has been significantly reduced (Fig. 3). Metals may be a small contributor to sediment toxicity and impaired benthic condition in the UH, but are dramatically overshadowed by the PCB effect. These results are consistent with previous toxicity identification evaluation (TIE) work conducted in NBH by Ho, *et. al.* (1997), where PCBs were also observed to be the major contaminant contributing to toxicity. Because metals tend to be co-located with PCBs in the UH, the area with SEM-AVS values greater than zero is also being reduced as the PCBs are remediated. While the remediation appears to be having a positive effect relative to the PCB criterion, the corresponding sediment toxicity and benthic community data demonstrate that this has not translated into significantly improved biological response to date. This is not surprising given that only about 28% of the UH remediation is

complete; however, as remediation continues and PCB levels decrease, the current LTM program will document future changes.

The picture in the LH is more interesting in that as PCB sediment concentrations decrease, as occurred in 2004 and 2009, the benthic community has improved and is significantly better in 2009 compared to the 1993 baseline (Fig. 5). The reason for this is probably due to the reduced transport from the primary source of contaminants in the UH. During the Hot Spot remediation in 1995 (Bergen, *et. al*, 2005), it was determined that there is a net PCB transport from the UH to the LH under ambient conditions due to tidal overturn and weather events. It follows that as continued remediation reduces the overall UH PCB sediment concentrations, less will be transported to the LH, as documented by the significantly reduced PCB levels in 2004 and 2009, and the benthic community should continue to improve there.

The OH is currently undergoing a remedial investigation to determine a final PCB concentration criterion. The "presumptive" clean-up concentration of 10 ppm used in this paper may change or isolated areas may need to be remediated based on the on-going human health and ecological risk assessments. The LTM data indicate that as a whole, the OH has very low PCB and metals concentrations and the benthic community is healthy.

As is common at most, if not all Superfund sites, the costs associated with monitoring must be balanced with other remedial needs to maximize clean-up; therefore, flexibility in monitoring programs is key to accomplishing this end. This adaptive management approach (NRC 2003; Linkov *et al*, 2008) was used in the NBH-LTM, specifically for the 2009 collection. Under the adaptive management paradigm, the focus is on reaching a specified objective in a specified period of time, and not on establishing a set approach that is unchangeable (Gustavson

et al, 2008). When the decision was made in the spring of 2009 to conduct another round of monitoring, all of the data from the previous years were carefully evaluated to determine if some of the information was redundant, or otherwise less informative to the overall purpose of the LTM (i.e., assessing effects and effectiveness of the remediation). It was obvious that the metals data were not that informative relative to assessing effects and effectiveness. In fact, when a correlation analysis was conducted with all the SEM-AVS and sediment toxicity data for all the years, no significant relation was observed at all with sediment toxicity ($r = -0.04$). In addition, SEM-AVS was not correlated with the two benthic indices either ($r = 0.17$ and $r = 0.19$ for the EMAP-BI and Shannon's H' index, respectively). Based on these facts, it was determined to not include metals in the 2009 LTM collection.

Further, the sediment toxicity was highly variable during the course of the project. While the overall pattern was similar to the benthic condition data, the response was more variable and less correlated to contaminant levels. For example, there was no correlation between sediment toxicity and metals ($r = -0.04$) and less of a correlation with PCBs ($r = -0.42$) than the EMAP-BI and PCBs ($r = -0.86$). Because sediments used for toxicity testing are homogenized, which has been shown to effect interstitial water PCB geochemistry (Burgess and McKinney, 1997), this may have had an effect on the increased variability compared to the benthic community evaluations, which are more reflective of *in-situ* longer-term responses to environmental conditions. Based on the fact that the LTM is conducted at approximately five year intervals, is concerned with longer-term changes, and that the remediation is expected to proceed for an extended period of time at current spending levels, it was decided that quantifying the benthic community was more consistent with the goals of the LTM program, and as a result sediment

toxicity testing was also dropped from the 2009 collection. This is not to say that either metals or sediment toxicity will not be incorporated again at some point in the future, or that they cannot provide useful information at another site. For example, it may be desirable to demonstrate a significant decrease in sediment toxicity in the UH once remediation is complete, which would require reinstating the sediment toxicity test at some point. However, based on the evaluation of the previous four LTM collections, both the metals and sediment toxicity measurement were not monitored during the 2009 LTM collection.

Conclusions

In summary, there is a significant decreasing sediment PCB gradient from the UH to the LH to the OH. This gradient has been maintained temporally from baseline sampling in 1993 to the latest sampling in 2009; however, remedial activities in the UH between 1999 and 2009 have resulted in significantly reduced PCB concentrations in all three segments of the harbor in 2009.

The higher contaminant concentrations along this gradient are associated with negative benthic community impacts, increased sediment toxicity, and greater PCB bioaccumulation in deployed blue mussels. While these biological effects are maintained over time in the benthic community indices and mussel bioacculation, a significant increase in the EMAP-BI was observed in 2009 in the LH and OH, indicating the possible beginning of longer-term improvements.

As designed, the NBH-LTM program has been able to quantify both spatial and temporal effects relative to the remedial activities at the site. The spatial coverage is sufficient to detect fine-scale localized effects (i.e., individual stations) due to the dredging activities (e.g., 1995 UH

dredging), and also adequate to determine larger scale segment-wide changes (e.g., PCB and biological indices). The temporal coverage links chemical and biological variables at appropriate time scales (i.e., long-term benthic community changes do not relate with yesterday's dredging). To date, less than 30% of the scheduled remediation has occurred, and this is reflected in slow changes in the longer-term indicators, like the benthic community and bioaccumulation in blue mussels; however, in areas where PCB concentrations are significantly lower (e.g., LH and OH), the benthic community is beginning to improve (e.g., EMAP - BI).

Finally, this monitoring program has been flexible enough to maintain a consistent level of sampling over a long time frame, but also adaptive enough to accommodate inevitable changes during that same period. For example, the duration of activities at Superfund sites, especially expensive ones like NBH, can be very protracted due to litigation, politics, and variable funding. Further, the personnel involved in these projects typically change over time, including remedial project managers (e.g., there have been four at this site since 1993) as well as the contractors that perform much of the sampling. Finally, new techniques that may be more informative and less expensive (e.g., analytical chemistry) need to be incorporated, while at the same time be comparable to older data. We believe the NBH-LTM program is consistent with these requirements, has been, and will continue to be, an effective method to assess remedial effects and effectiveness, and that other contaminated sites can benefit from the monitoring approach used in NBH.

Acknowledgements Since the inception of the NBH-LTM, dozens of people from various Federal, state, local, and private entities have contributed to this project. In particular, the

authors thank NBH remedial project managers Gayle Garman and David Dickerson, Joseph Mackay of the U.S. Army Corps of Engineers, New England Division, Saro Jayaraman of AED for mussel PCB analysis, James Helsche, Computer Sciences Corporation, for statistical analyses, and Chuck Audette, Raytheon, for benthic community analysis. This manuscript greatly benefited from technical discussions with Anne Kuhn, Glenn Thursby and Daniel Campbell of AED and the thorough reviews of Mohamed Abdelrhman, John Kiddon, and Walter Galloway. Although the research described in this article has been partially funded by the U.S. Environmental Protection Agency it has not been subjected to Agency review. Therefore, it does not necessarily reflect the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This is contribution number AED-11-034 of the Atlantic Ecology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, U. S. Environmental Protection Agency.

References

Bergen, B.J., Nelson, W.G., & Pruell, R.J. (1993). The bioaccumulation of PCB congeners by blue mussels (*Mytilus edulis*) deployed in New Bedford Harbor, MA. *Environmental Toxicology and Chemistry*, 12, 1671-1681.

Bergen, B.J., Nelson, W.G., Quinn, J.G., & Jayaraman, S. (2001). Relationships among total lipids, lipid classes and PCB concentrations in two indigenous populations of ribbed mussels (*Geukensia demissa*) over an annual cycle. *Environmental Toxicology and Chemistry*, 20(3), 575-581.

Bergen, B.J., Nelson, W.G., Mackay, J., Dickerson, D., & Jayaraman, S. (2005). Environmental monitoring of remedial dredging at the New Bedford Harbor, MA, Superfund site. *Environmental Monitoring and Assessment*, 111, 257-275.

Boothman, W.S. & Helmstetter, A. (1992). *Determination of acid-volatile sulfide and simultaneously-extracted metals in sediments using sulfide-specific electrode detection*. Standard Operating Procedures, U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, R.I.

Burgess, R.M. & McKinney. R.A. (1997). Effects of sediment homogenization on interstitial water PCB chemistry. *Archives of Environmental Contamination and Toxicology*, 33, 125-129.

Calder, J.A, (1986). Marine environmental quality: NOAA's National Status and Trends Program. In: *Oceans '86 Conference Record*. Marine Technical Society IEEE Cat. No. 86CH2363-0. Washington, DC. and Inst. Electrical Electronics Eng., Piscataway, New Jersey, 1351-1354.

Chapman, P.M. (1990). The sediment quality triad approach to determining pollution-induced degradation. *Science of the Total Environment,* 97/98, 815 - 825.

Di Toro, D.M., Mahony, R.A., Hansen, D.J., Scott, K.J., Hicks, M.B., Mayr, S.M., et al. (1990). Toxicity of cadmium in sediments: the role of acid volatile sulfide. *Environmental Toxicology and Chemistry,* 9, 1487-1502.

Farrington, J.W. (1983). Bivalves as sentinels of coastal chemical pollution: The mussel watch. *Oceanus*, 26, 18-29.

Goldberg, E.D., & Bertine, K.K. (2000). Beyond the Mussel Watch – new directions for monitoring marine pollution. *Science of the Total Environment,* 247, 165-174.

Gustavson, K.E., Burton, G.A., Francingues Jr., N.R., Reible, D.D., Vorhees, D.J. & Wolfe J.R . (2008). Evaluating the effectiveness of contaminated sediment dredging. *Environmental Science and Technology*, 42, 5042-5047.

Ho, K.T, McKinney, R.A., Kuhn, A., Pelletier, M.C., & Burgess, R.M. (1997). Identification of acute toxicants in New Bedford Harbor sediments. *Environmental Toxicology and Chemistry*, 16(3), 551-558.

Johnston, K., Ver Hoef, J., Krivoruchko, K., & Lucas, N. (2001). Using ArcGIS Geostatistical Analyst. ESRI. Redlands, C.A., 48 pp.

Linkov, I., Satterstrom, F.K., Kiker, G., Batchelor C., Bridges, T., & Ferguson, E. (2008). From comparative risk assessment to multi-criteria decision analysis and adaptive management: recent developments and applications. *Environment International*, 32, 1072-1093.

Long, E.R., Field, L,J., & MacDonald, D.D. (1998). Predicting toxicity in marine sediments with numerical sediment quality guidelines. Environmental Toxicology and Chemistry 17, 714 - 727.

National Research Council (NRC). (1997). Building a foundation for sound environmental decisions. The National Academies Press, Washington, D.C., Library of Congress Catalog Card Number 97-67451.

National Research Council (NRC). (2003). Environmental clean up at Navy facilities: Adaptive site management. The National Academies Press, Washington, D.C., Library of Congress Control Number 2003108132.

National Research Council (NRC). (2007). Dredging at Superfund Megasites: Assessing the Effectiveness. The National Academies Press, Washington, D.C., Library of Congress Control Number 2007931836.

Nelson, W.G. & Hansen, D.J. (1991). Development and use of site-specific chemical and biological criteria for assessing New Bedford Harbor pilot dredging project. *Environmental Management* 15, 105 - 112.

Nelson, W.G., Bergen, B.J., & Cobb, D.J. (1995). Comparison of PCB and trace metal bioaccumulation in the blue mussel, *Mytilus edulis*, and the ribbed mussel, *Modiolus demissus*, in New Bedford Harbor, MA. *Environmental Toxicology and Chemistry*, 14, 513 - 521.

Nelson, W.G., Bergen, B.J., Benyi, S.J., Morrison, G., Voyer, R.A., Strobel, C.J., et al. (1996a). *New Bedford Harbor Long-Term Monitoring and Assessment Report: Baseline Sampling*. U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, RI. EPA/600/R-96/097.

Nelson, W.G. & Gleason, T. (1996b). *Deployment and retrieval of caged bivalves for environmental monitoring*. Standard Operating Procedures, U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, RI.

Otis, M.J. & Averett, D.E. (1988). *Pilot study of dredging and dredged material disposal methods: New Bedford, Massachusetts Superfund site*. Superfund '88. Proceedings of the 9th National Conference. Nov 28-30, Washington D.C.

Paul, J.F, Gentile, J.H., Scott, K.J., Schimmel, S.C., Campbell, D.E., & Latimer, R.W. (1999). *EMAP-Virginian Province 4-Year assessment (1990-1993)*. U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, R.I., EPA/620/R-99-004.

Paul, J.F., Scott, K.J., Campbell, D.E., Gentile, J.H., Strobel, C.S., Valente, R.M et al. (2001). Developing and applying a benthic index of estuarine condition for the Virginian Biogeographic Province. *Ecological Indicators*, 1, 83-99.

Reifsteck, D.M, Strobel, C.J., & Keith, D.J. (1993). *EMAP-Estuaries 1993 Virginian Province Field Operations and Safety Manual*, U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, R.I. 172 pp.

Tanabe, S., Tatsukawa, R., Phillips, D.J. (1987). Mussels as bioindicators of PCB pollution: A case study on uptake and release of PCB isomers and congeners in green-lipped mussels *(Perna viridis*) in Hong Kong waters. *Environmental Pollution*, 47, 41-62.

U.S. ACOE. (1999). *Quality Assurance Project Plan for Long-Term Monitoring III at New Bedford Harbor Superfund Site, New Bedford, Massachusetts*. Prepared by ENSR Corporation for the U.S. Army Corps of Engineers, New England Division. Document number 9000-218-103.

U.S. EPA. (1994). *Methods for Assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods*. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, RI., EPA/600/R-94/025.

U.S. EPA. (1995). *EMAP - Estuaries: Laboratory methods Manual - Estuaries. Vol 1. Biological and Physical Analyses*. U.S. Environmental Protection Agency, National Environmental and Ecological Effects Laboratory, Atlantic Ecology Division, Narragansett, R.I., EPA/620/R-95/008.

U.S. EPA. (2004). *Updated report on the incidence and Severity of Sediment contamination in Surface Waters of the United States, National Sediment Quality Survey*. U. S. EPA, Office of Water, Washington, D.C., EPA-823-R-04-007.

U.S. EPA. (2005). *Contaminated Sediment Remediation Guidance for Hazardous Waste Sites*. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, D.C. EPA-540-R-05-012.

Valente, R.M. & Strobel, C.J. (1993). *EMAP-Estuaries Virginian Province 1993 Quality Assurance Project Plan*. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Narragansett, R.I.

Washington, H.G. (1984). Diversity, biotic and similarity indices. *Water Research*, 18, 653-694.

Table 1 Suite of physical, chemical, and biological variables collected at each sediment sampling station during the NBH-LTM program.

Physical	Chemical		Biological
Grain size	PCB Congeners	Metals	Sediment Toxicity
Gravel	CB008	Arsenic	Benthic community
Sand very coarse	CB018	Cadmium	
Sand coarse	CB028	Chromium	
Sand medium	CB044	Copper	
Sand fine	CB052	Lead	
Sand very fine	CB066	Nickel	
Silt and clay	CB101	Mercury	
	CB105	Selenium	
	CB118	Zinc	
	CB128		
	CB138		
	CB153		
	CB170		
	CB180		
	CB187		
	CB195		
	CB206		
	CB209		
	Total organic carbon		
	Acid volatile sulfide		

Table 2 Means and standard deviations (SD) for each of the eight metals quantified during the NBH-LTM program for each harbor segment-collection year combination. *Spatial differences*: means that are not significantly different (p>0.05) among harbor segments within a collection year are connected by solid lines, those that are different have no lines (e.g., in 1993, Cu concententrations were not different between the UH and LH). *Temporal differences*: means that are significantly different (p<0.05) from the 1993 baseline values for each harbor segment are shown in bold italics (e.g., mean Cu concentrations for 1995, 1999, and 2004 in the UH were significantly higher than the 1993 values).

Year	Segment	As		Cd		Cr		Cu		Pb		Hg		Ni		Se		Zn	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1993	UH	5.1	2.9	64.8	138.6	304.2	205.3	611.7	376.3	267.3	158.5	0.4	0.4	33.1	24.6	0.3	0.2	627.9	393.1
	LH	5.3	2.7	12.4	38.6	189.2	175.8	454.2	460.6	129.1	93.2	0.4	0.3	10.6	6.0	0.4	0.4	258.0	141.3
	OH	3.1	1.9	0.3	0.3	19.0	15.4	19.5	19.2	18.2	14.0	0.1	0.1	5.3	3.6	0.2	0.2	42.1	28.6
1995	UH	***7.8***	4.0	9.3	7.1	***428.3***	271.2	***840.5***	464.5	***346.2***	219.6	***0.9***	0.6	***65.2***	65.0	***1.2***	1.0	***850.0***	585.5
	LH	***7.6***	4.0	2.2	1.8	***234.9***	210.1	***574.2***	544.8	***161.0***	105.8	***0.6***	0.4	***20.0***	12.2	***1.1***	1.0	***320.1***	175.5
	OH	***4.6***	2.5	***0.2***	0.3	***26.1***	20.7	26.3	30.3	***23.5***	18.8	***0.1***	0.1	***7.7***	4.5	***0.5***	0.4	***55.7***	37.8
1999	UH	***6.4***	2.9	9.4	6.0	***398.9***	209.7	***760.6***	363.7	280.1	129.4	***0.8***	0.4	***38.5***	23.7	***0.1***	0.1	686.7	358.5
	LH	***7.3***	3.9	1.9	1.2	211.9	175.6	638.5	927.5	138.2	87.8	***0.6***	0.3	***18.3***	21.0	***0.1***	0.0	***372.9***	303.7
	OH	3.1	1.8	0.2	0.3	20.5	16.9	18.7	20.0	18.5	16.9	***0.0***	0.0	5.8	3.7	***0.0***	0.0	42.5	30.7
2004	UH	***8.2***	4.7	6.6	4.7	341.1	207.9	***694.4***	385.3	286.1	158.9	***0.8***	0.4	***56.3***	32.3	***0.8***	0.4	622.1	373.3
	LH	***9.1***	5.1	1.9	1.9	203.0	158.4	***535.6***	614.8	***152.5***	89.1	***0.5***	0.3	***31.3***	31.2	***0.8***	0.4	***366.7***	264.8
	OH	***4.9***	3.5	***0.2***	0.2	***34.9***	22.8	21.6	20.9	***22.0***	15.9	0.1	0.1	***11.1***	7.0	0.3	0.2	***60.4***	37.8

Table 3. Mean amphipod (*Ampelisca abdita*) survival (%) for the 10-day sediment toxicity test. Spatial comparisons: within a year, values connected by a line are not significantly different, which only occurred in 1993. Temporal comparisons: within a segment across years, bolded values are significantly different from 1993.

Year	Segment	Mean %Survival
1993	UH	55
	LH	66
	OH	91
1995	UH	***22***
	LH	***48***
	OH	***84***
1999	UH	***2***
	LH	***32***
	OH	***75***
2004	UH	46
	LH	68
	OH	***81***

Table 4. The means of three benthic metrics (number of taxa, Shannon's H, and EMAP-BI) for each harbor segment-collection year combination. Spatial differences: means that are not significantly different (p>0.05) among harbor segments within a collection year are connected by solid lines (e.g., mean # Taxa in UH and LH in 1995). Temporal differences: means that are significantly different from the 1993 baseline values for each harbor segment are shown in bold italics (e.g., the mean # Taxa in the UH in 1999 and 2004 were significantly lower than 1993).

Year	Segment	Mean # Taxa	Mean Shannon	Mean EMAP-BI
1993	UH	19	0.5	-4.2
	LH	28	0.7	-0.6
	OH	62	1.2	2.5
1995	UH	21	***0.6***	-3.7
	LH	26	0.8	-0.4
	OH	57	***1.1***	2.4
1999	UH	***16***	0.5	-2.8
	LH	30	0.8	-0.5
	OH	***54***	***1.0***	2.2
2004	UH	***14***	***0.6***	-2.3
	LH	***22***	0.8	-0.1
	OH	***42***	***1.1***	2.1
2009	UH	19	0.5	-3.7
	LH	30	***0.9***	***0.2***
	OH	64	1.2	***3.3***

Table 5. Summary table showing sediment contaminants, sediment toxicity, and benthic community indices relative to criterion or threshold values discussed in the text. Boxes colored red indicate that the mean value for each harbor segment-collection year combination is signifcantly (p<0.05) worse than it's criterion or threshold value, yellow indicates no significant difference, and green indicates significantly better. Bolded values within the colored boxes are the percent of stations within each segment-year combination that violate the criterion or threshold value. Values in non-colored boxes (e.g., ERM-Q, # Taxa) are the segment-year means of other variables used to explain the responses relative to the criterion or threshold values.

		Sediment Contaminants				Sediment Toxicity	Benthic Condition		
Segment	Year	PCBs	PCB ERM-Q	SEM-AVS	Metals ERM-Q		EMAP-BI	Shannon's H	#Taxa
UH	1993	**78**	592	**37**	1.7	**74**	**96**	0.5	19
	1995	**89**	769	**19**	1.4	**93**	**96**	0.6	21
	1999	**81**	567	**19**	1.2	**100**	**100**	0.5	16
	2004	**74**	334	**11**	1.2	**85**	**100**	0.6	14
	2009	**74**	393				**93**	0.5	19
LH	1993	**0**	46	**19**	0.7	**52**	**74**	0.7	28
	1995	**0**	41	**7**	0.7	**62**	**66**	0.8	26
	1999	**0**	42	**28**	0.7	**100**	**62**	0.8	30
	2004	**0**	27	**10**	0.7	**69**	**66**	0.8	22
	2009	**0**	29				**45**	0.9	30
OH	1993	**0**	4.6	**13**	0.1	**4**	**0**	1.2	62
	1995	**0**	2.4	**4**	0.1	**4**	**4**	1.1	57
	1999	**0**	2.1	**9**	0.1	**48**	**0**	1.0	54
	2004	**0**	1.0	**9**	0.1	**30**	**0**	1.1	42
	2009	**0**	1.3				**0**	1.2	64

<u>Criterion and Threshold Values</u>
PCBs = 10 ppm for UH and OH, 50 ppm for LH
SEM-AVS > 0
Sediment Toxicity <80%
EMAP-BI < 0

Year	Location	Mean Silt/C	Mean TOC %
1993	Upper	45.9	7.4
	Lower	41.2	4.6
	Outer	42.9	1.3
1995	Upper	30.6	5.8
	Lower	21.2	3.4
	Outer	25.9	1.0
1999	Upper	53.1	6.4
	Lower	42.9	4.1
	Outer	41.6	1.5
2004	Upper	57.8	6.3
	Lower	47.4	4.6
	Outer	46.5	1.2
2009	Upper	47.2	2.6
	Lower	37.9	1.8
	Outer	41.5	0.9

List of Figures

Figure 1. Map of the New Bedford Harbor (NBH) Superfund Site showing the three segments of the harbor (UH, LH, OH), with the hexagonal systematic sampling grid superimposed on it. Sediment samples were collected at the center of each hexagon for each of the five collections (1993, 1995, 1999, 2004, 2009). Also shown are the three mussel deployment stations (NBH-2, NBH-4, NBH-5).

Figure 2. Timeline (not to scale) showing the relationship between the five long-term monitoring collections (red), the remedial dredging activities (black), and the cubic yards (cy^3) of contaminated sediment dredged (brown) during each collection interval. The total amount of material dredged up to, and including, the 2009 collection (approximately 200,000 cubic yards) represents about 28% of the total volume to be removed; another 710,000 cubic yards are scheduled to be remediated.

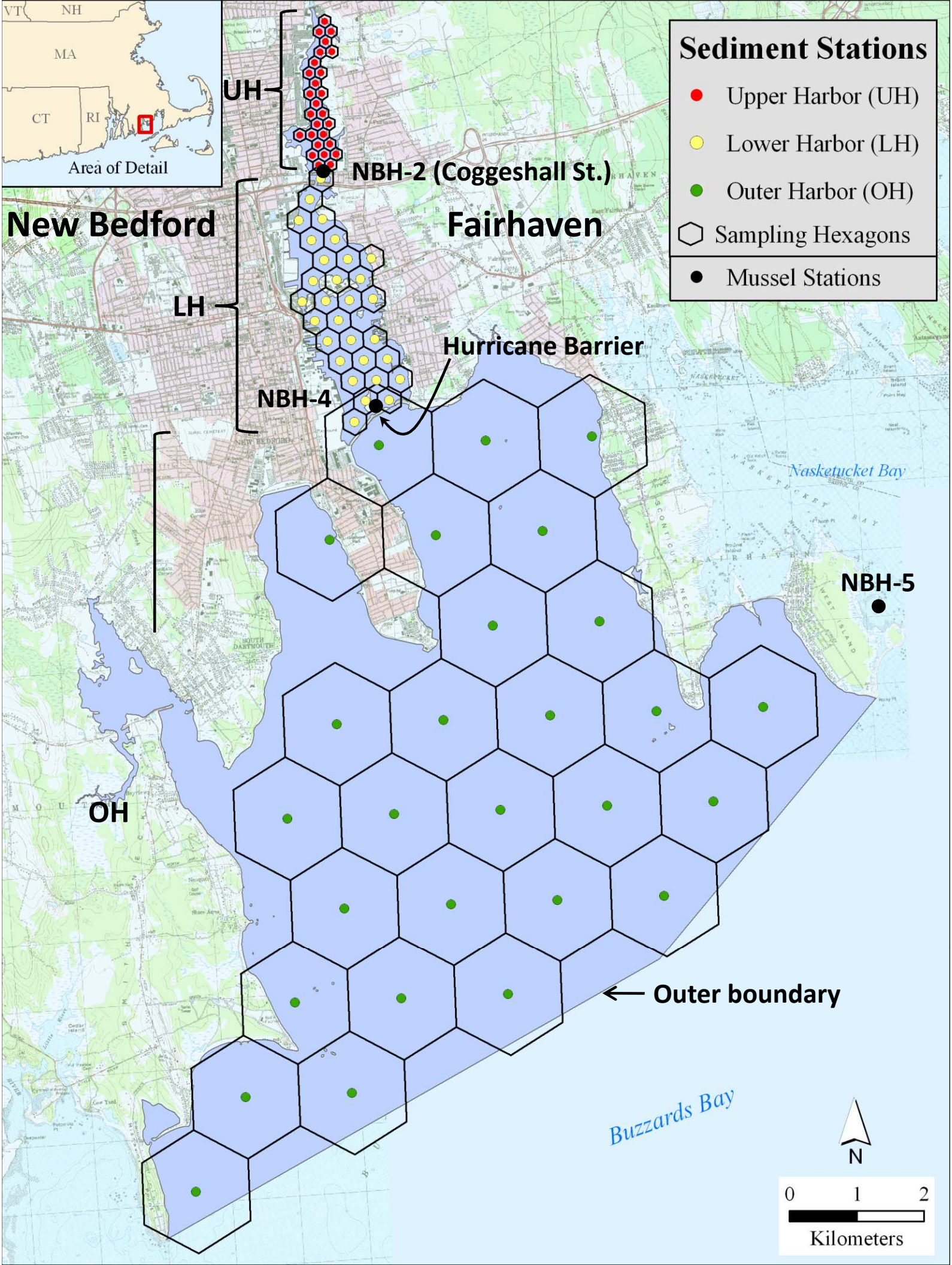
Figure 3. Spatial distribution of the kriged sediment PCB concentration data for the UH, LH, and OH segments, for each of the five long-term monitoring collections. The number to the right of each segment is the surface-weighted average PCB concentration for that segment. Spatially, within a collection year, each segment is significantly different from each other for all five collections. Temporally, within a segment, each mean PCB concentration was compared to the 1993 baseline, with values that are significantly higher shown in red, values not different in black, and significantly lower values shown in green.

Figure 4. Number of dominant taxa (i.e., species or taxa constituting 75% of the total individuals) within each harbor segment for the 1993 baseline sampling. The UH stations had greater abundance of a few small, opportunistic species (3 species), the LH) slightly more taxa (5) with a few larger species, and the OH exhibited a much greater diversity (26 taxa). This pattern, with minor variations among dominant species, was repeated for the 1995, 1999, 2004, and 2009 collections.

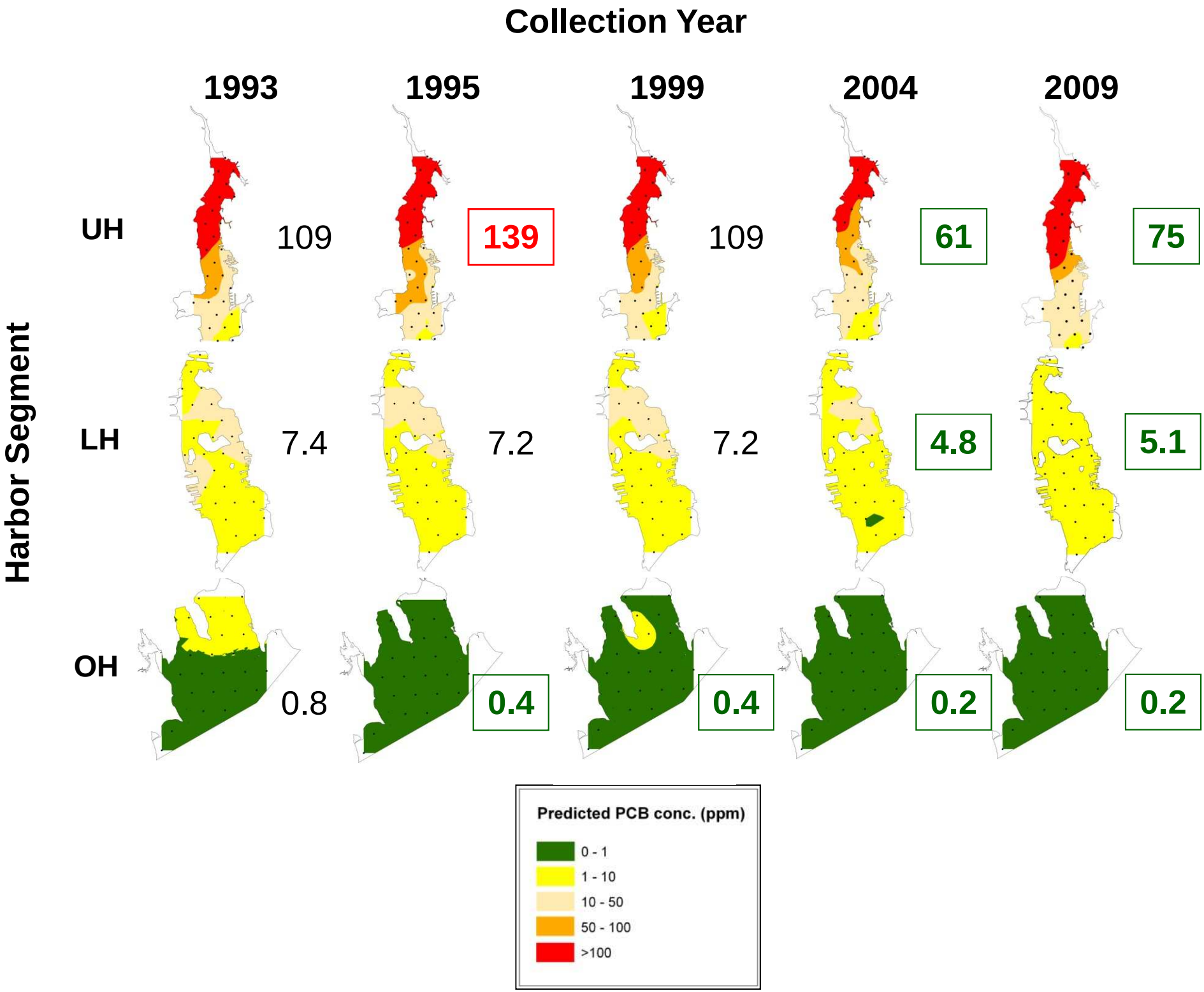
Figure 5. Bar graph showing the mean EMAP Benthic Index value for each harbor segment (UH, LH, OH) and collection year. Positive values indicate a “good” benthic community, while negative values are considered “impaired.” Spatially, within a collection year, there was a significant difference among harbor segments. Temporally, when compared to the 1993 baseline collection, there was a significant increase* in benthic condition in the LH and OH segments in the 2009 collection.

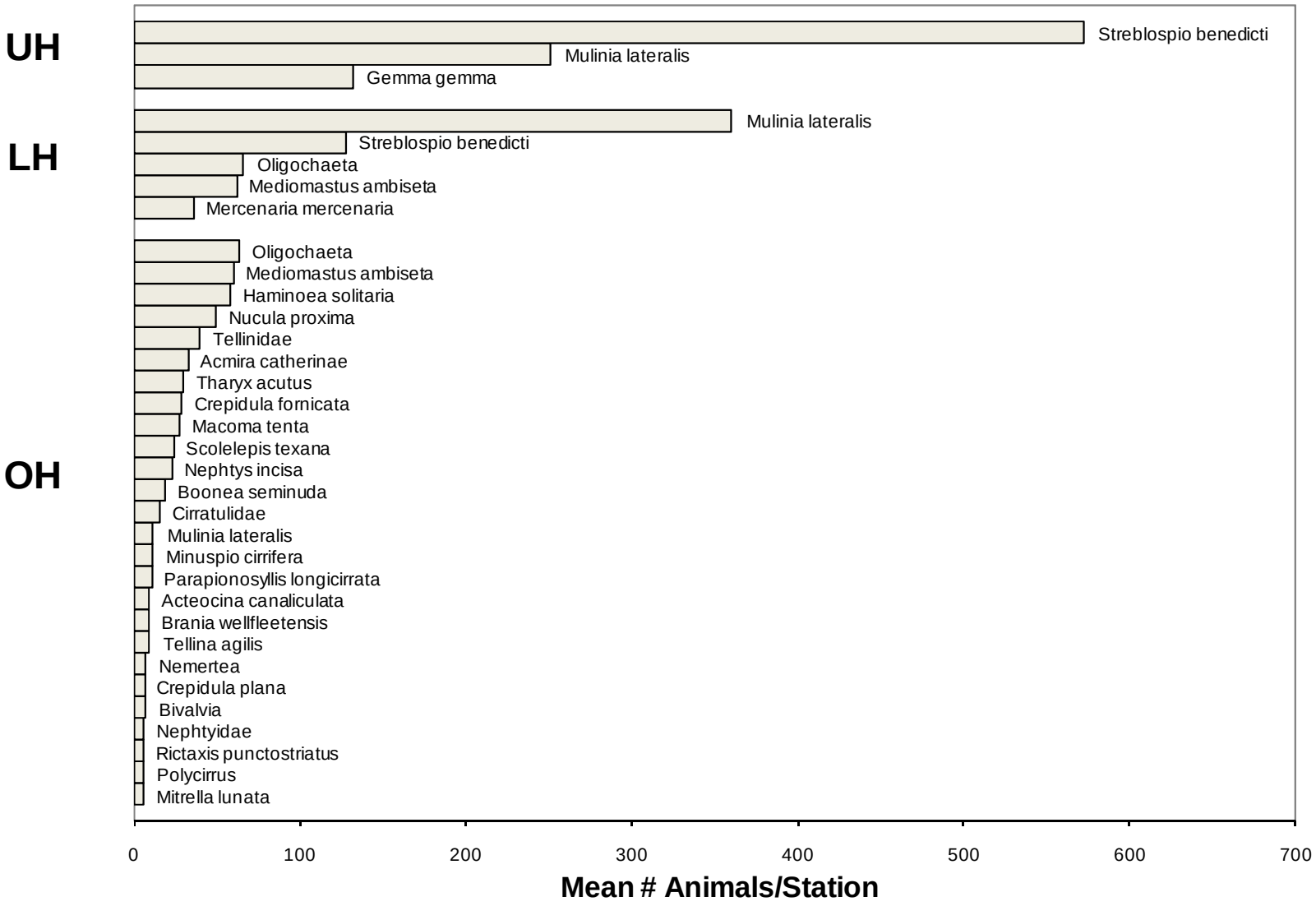
Figure 6. Blue mussel PCB tissue concentrations measured at three stations (NBH-2, NBH-4, NBH-5). The overall means (a) for all deployment (1993-2009) are significantly different among the three areas and reflect the dramatic gradient in water column PCB concentrations. Within a station over time (b), there is similar natural variability pattern at all stations, primarily due to the mussel reproductive cycle; however, the differences in PCB tissue concentration are maintained over the entire period (note difference in scales among stations).

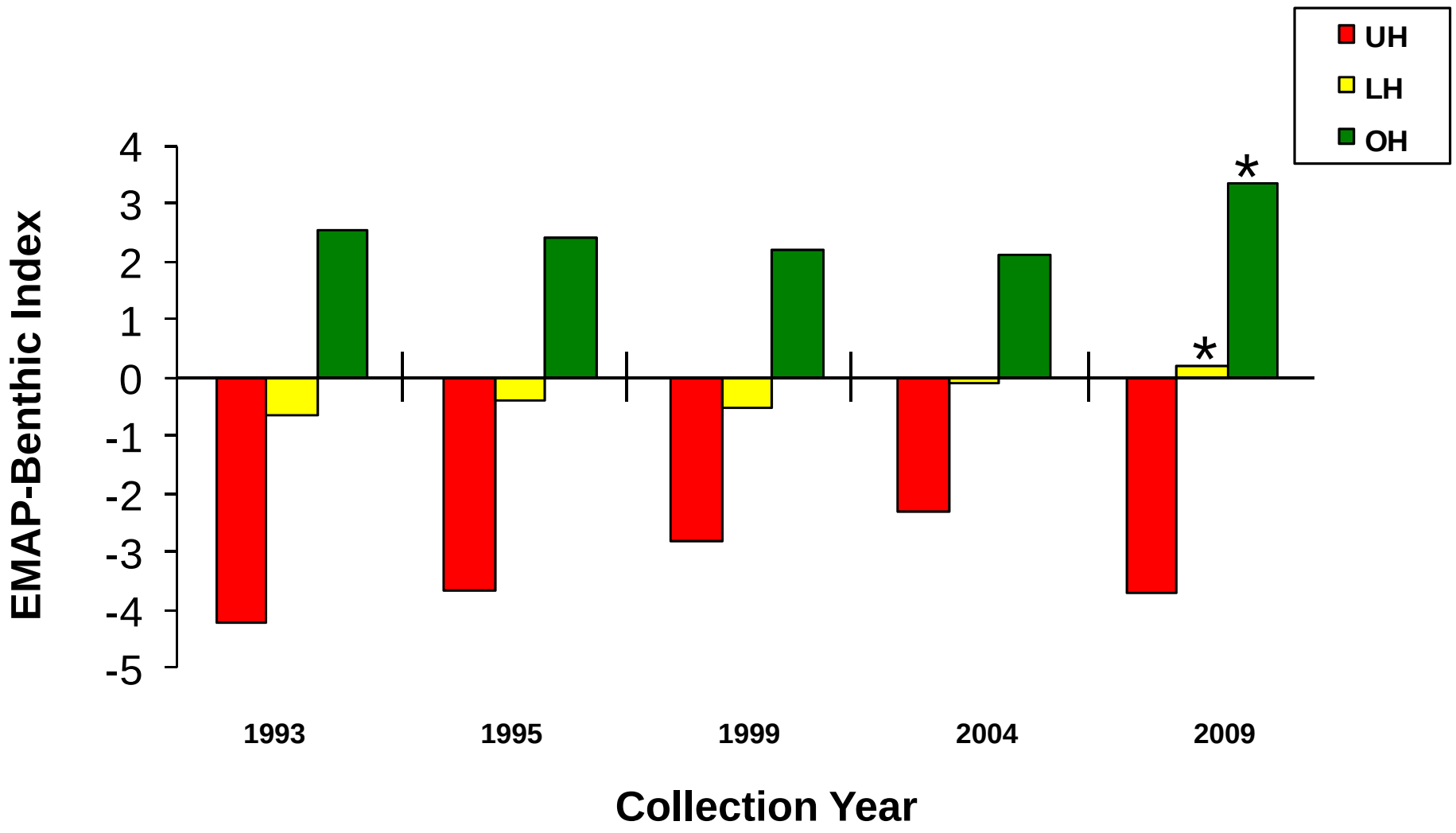
Figure 7. Sediment PCB concentration data from the UH for each station and collection year (a). Also shown is a map of the UH with the location of each station relative to the dredging activities (b). The red numbered dashed lines show the stations located in the northern portion of the UH, in close proximity to the dredging areas, the black lines show those stations in the southern portion of the UH, further away from the remedial activities.

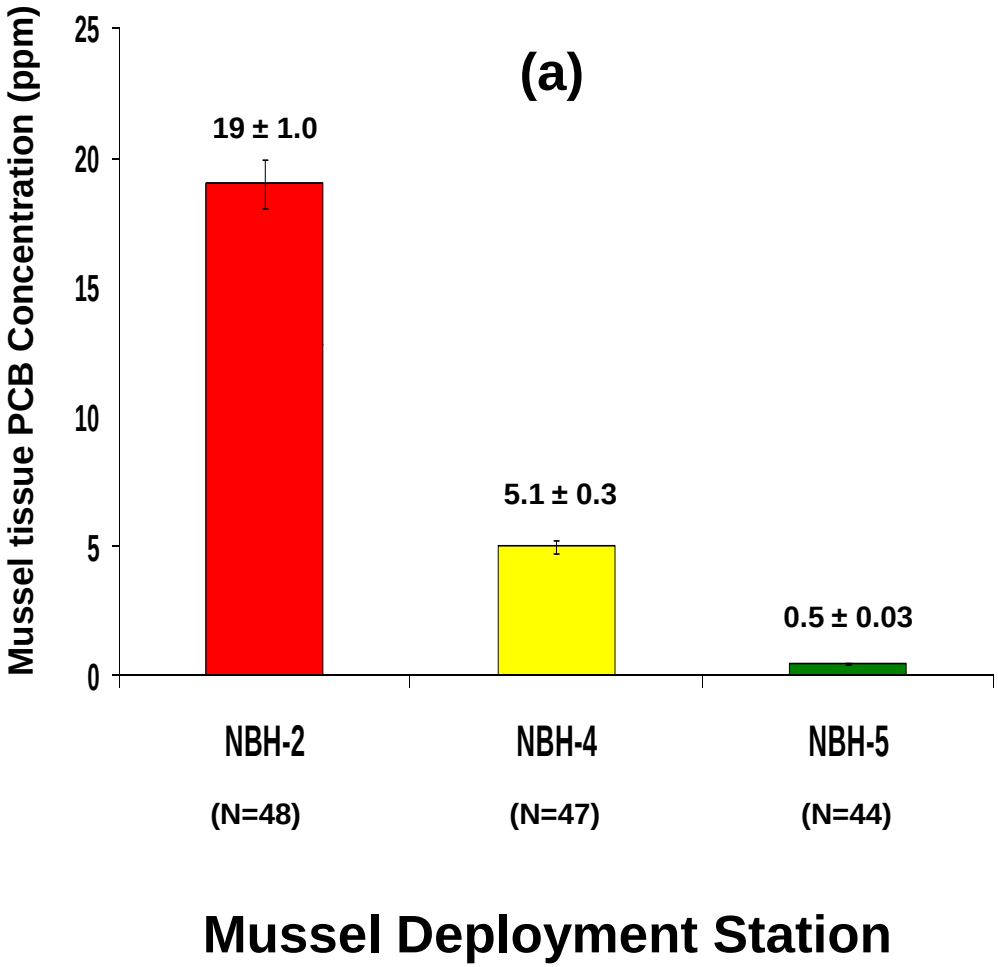

Sediment Stations
Upper Harbor (UH)
Lower Harbor (LH)
Outer Harbor (OH)
Sampling Hexagons
Mussel Stations
VT
NH
MA
CT
RI
Area of Detail
UH
LH
OH
New Bedford
Fairhaven
NBH-2 (Coggeshall St.)
NBH-4
NBH-5
Hurricane Barrier
Nasketucket Bay
Buzzards Bay
Outer boundary
N
0
1
2
Kilometers

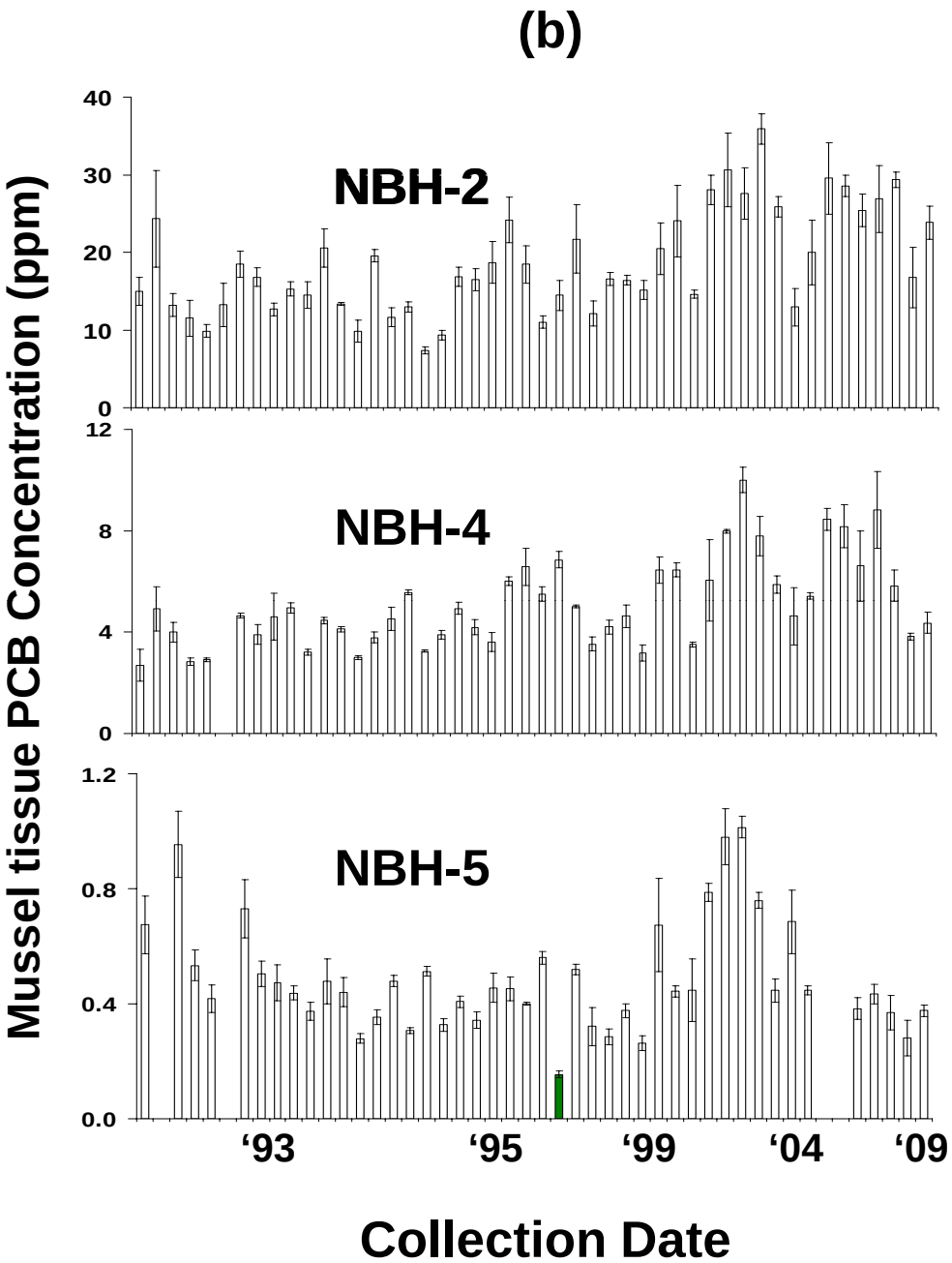
1993	1995	1999	2004	2009
Baseline	"Hot Spot" Dredging	No Dredging	Limited Dredging	Full Scale Dredging
Sediment Dredged (yd^3)	**14,000 (Upper)**	**0**	**20,600 (Upper) 4,000 (Lower)**	**159,000 (Upper)**

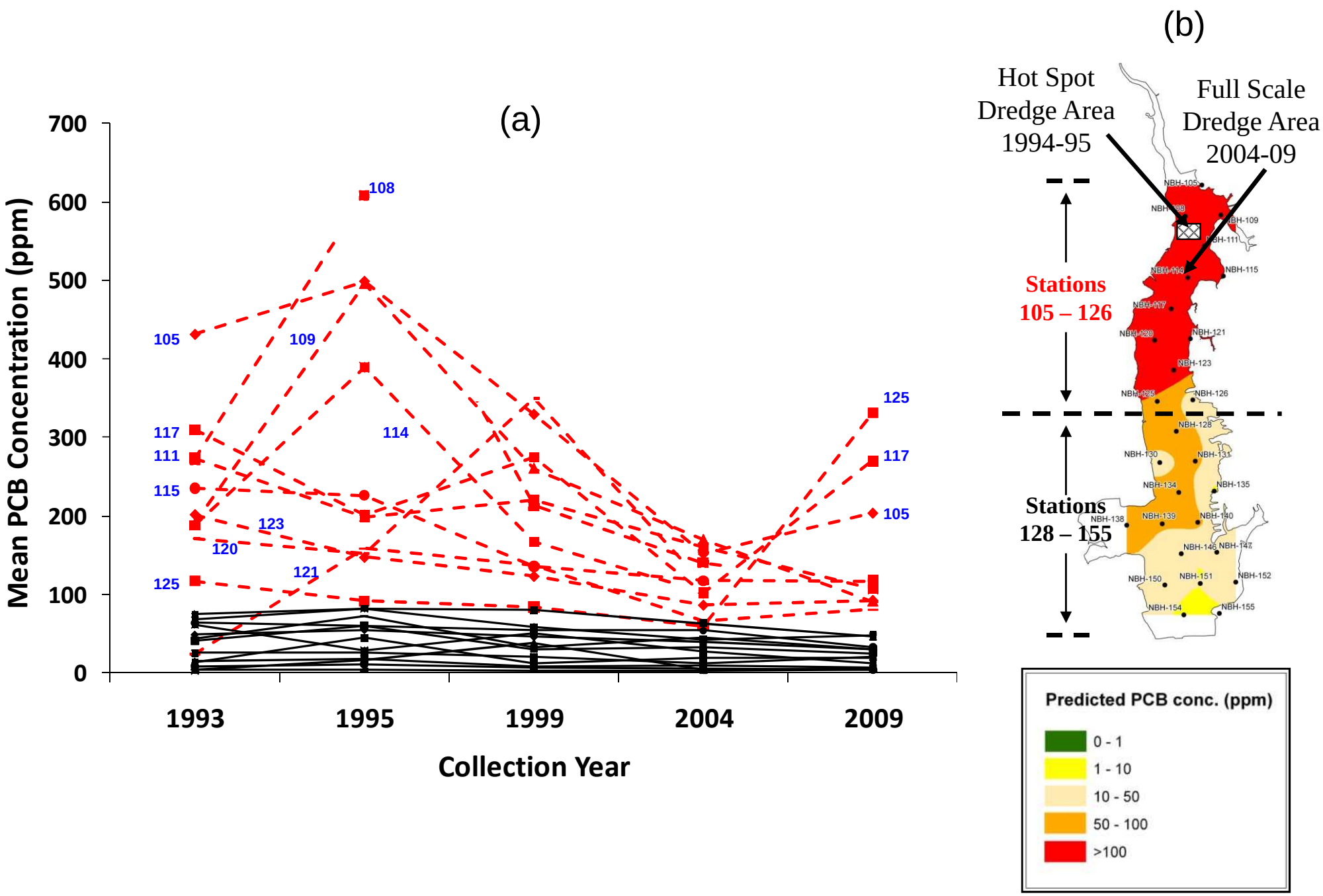

Collection Year
1993
1995
1999
2004
2009
Harbor Segment
UH
LH
OH
109
139
109
61
75
7.4
7.2
7.2
4.8
5.1
0.8
0.4
0.4
0.2
0.2
Predicted PCB conc. (ppm)
0 - 1
1 - 10
10 - 50
50 - 100
>100


UH
LH
OH
Streblospio benedicti
Mulinia lateralis
Gemma gemma
Mulinia lateralis
Streblospio benedicti
Oligochaeta
Mediomastus ambiseta
Mercenaria mercenaria
Oligochaeta
Mediomastus ambiseta
Haminoea solitaria
Nucula proxima
Tellinidae
Acmira catherinae
Tharyx acutus
Crepidula fornicata
Macoma tenta
Scolelepis texana
Nephtys incisa
Boonea seminuda
Cirratulidae
Mulinia lateralis
Minuspio cirrifera
Parapionosyllis longicirrata
Acteocina canaliculata
Brania wellfleetensis
Tellina agilis
Nemertea
Crepidula plana
Bivalvia
Nephtyidae
Rictaxis punctostriatus
Polycirrus
Mitrella lunata
0
100
200
300
400
500
600
700
Mean # Animals/Station


UH
LH
OH
EMAP-Benthic Index
4
3
2
1
0
-1
-2
-3
-4
-5
*
*
1993
1995
1999
2004
2009
Collection Year


(a)
Mussel tissue PCB Concentration (ppm)
25
20
15
10
5
0
19 ± 1.0
5.1 ± 0.3
0.5 ± 0.03
NBH-2
NBH-4
NBH-5
(N=48)
(N=47)
(N=44)
Mussel Deployment Station


(b)
Mussel tissue PCB Concentration (ppm)
NBH-2
40
30
20
10
0
NBH-4
12
8
4
0
NBH-5
1.2
0.8
0.4
0.0
'93
'95
'99
'04
'09
Collection Date


(a)
Mean PCB Concentration (ppm)
700
600
500
400
300
200
100
0
1993
1995
1999
2004
2009
Collection Year
105
109
108
117
111
115
123
120
121
125
114
125
117
105
(b)
Hot Spot
Dredge Area
1994-95
Full Scale
Dredge Area
2004-09
Stations
105 – 126
Stations
128 – 155
Predicted PCB conc. (ppm)
0 - 1
1 - 10
10 - 50
50 - 100
>100